CHAPTER 2

Hilbert Spaces

1. Introduction

Fourier series played a significant role in the development of Hilbert spaces and other aspects of abstract analysis. The theory of Hilbert spaces returns the favor by illuminating much of the information about Fourier series. We first develop enough information about Hilbert spaces to allow us to regard Fourier series as orthonormal expansions. We prove that (the symmetric partial sums of) the Fourier series of a square-integrable function converges in L^2. From this basic result we obtain corollaries such as Parseval's formula and the Riemann–Lebesgue lemma. We prove Bernstein's theorem: the Fourier series of a Hölder continuous function (with exponent greater than $\frac{1}{2}$) converges absolutely. We prove the spectral theorem for compact Hermitian operators. We include Sturm–Liouville theory to illustrate orthonormal expansion. We close by discussing spherical harmonics, indicating one way to pass from the circle to the sphere. These results leave one in awe at the strength of nineteenth-century mathematicians.

The ideas of real and complex geometry combine to make Hilbert spaces a beautiful and intuitive topic. A Hilbert space is a complex vector space with a Hermitian inner product and corresponding norm, making it into a complete metric space. Completeness enables a deep connection between analytic and geometric ideas. Polarization, which holds only for complex vector spaces, also plays a significant role.

2. Norms and Inner Products

Let V be a vector space over the complex numbers. In order to discuss convergence in V, it is natural to use *norms* to compute the lengths of vectors in V. In Chap. 3, we will see the more general concept of a *semi-norm*.

Definition 2.1 (Norm). A *norm* on a (real or) complex vector space V is a function $v \mapsto ||v||$ satisfying the following three properties:

(1) $||v|| > 0$ for all nonzero v.
(2) $||cv|| = |c|\ ||v||$ for all $c \in \mathbf{C}$ and all $v \in V$.
(3) (The triangle inequality) $||v + w|| \leq ||v|| + ||w||$ for all $v, w \in V$.

Given a norm $||\ ||$, we define its corresponding *distance function* by

$$d(u, v) = ||u - v||. \tag{1}$$

J.P. D'Angelo, *Hermitian Analysis: From Fourier Series to Cauchy-Riemann Geometry*, Cornerstones, DOI 10.1007/978-1-4614-8526-1_2,

The function d is symmetric in its arguments u and v, its values are nonnegative, and its values are positive when $u \neq v$. The triangle inequality

$$||u - \zeta|| \leq ||u - v|| + ||v - \zeta||$$

follows immediately from the triangle inequality for the norm. Therefore, d defines a distance function in the metric space sense (defined in the appendix) and (V, d) is a metric space.

DEFINITION 2.2. A sequence $\{z_n\}$ in a normed vector space V converges to z if $||z_n - z||$ converges to 0. A series $\sum z_k$ converges to w if the sequence $\{\sum_{k=1}^{n} z_k\}$ of partial sums converges to w.

Many of the proofs from elementary real analysis extend to the setting of metric spaces and even more of them extend to normed vector spaces. The norm in the Hilbert space setting arises from an inner product. The norm is a much more general concept. Before we give the definition of Hermitian inner product, we recall the basic example of complex Euclidean space. Figures. 2.1–2.3 provide geometric intuition.

EXAMPLE 2.1. Let $\mathbf{C}^n$ denote complex Euclidean space of dimension n. As a set, $\mathbf{C}^n$ consists of all n-tuples of complex numbers; we write $z = (z_1, \ldots, z_n)$ for a point in $\mathbf{C}^n$. This set has the structure of a complex vector space with the usual operations of vector addition and scalar multiplication. The notation $\mathbf{C}^n$ includes the vector space structure, the Hermitian inner product defined by (2.1), and the squared norm defined by (2.2). The *Euclidean inner product* is given by

$$\langle z, w \rangle = \sum_{j=1}^{n} z_j \overline{w}_j \tag{2.1}$$

and the *Euclidean squared norm* is given by

$$||z||^2 = \langle z, z \rangle. \tag{2.2}$$

Properties (1) and (2) of a norm are evident. We establish property (3) below.

The Euclidean norm on $\mathbf{C}^n$ determines by (1) the usual Euclidean distance function. A sequence of vectors in $\mathbf{C}^n$ converges if and only if each component sequence converges; hence $\mathbf{C}^n$ is a complete metric space. See Exercise 2.6.

DEFINITION 2.3 (Hermitian inner product). Let V be a complex vector space. A *Hermitian inner product* on V is a function $\langle\ ,\ \rangle$ from $V \times V$ to $\mathbf{C}$ satisfying the following four properties. For all $u, v, w \in V$, and for all $c \in \mathbf{C}$:

(1) $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$.
(2) $\langle cu, v \rangle = c\langle u, v \rangle$.
(3) $\langle u, v \rangle = \overline{\langle v, u \rangle}$. (Hermitian symmetry)
(4) $\langle u, u \rangle > 0$ for $u \neq 0$. (Positive definiteness)

Three additional properties are consequences:

- $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$.
- $\langle u, cv \rangle = \overline{c}\langle u, v \rangle$.
- $\langle 0, w \rangle = 0$ for all $w \in V$. In particular, $\langle 0, 0 \rangle = 0$.

Positive definiteness provides a technique for verifying that a given z equals 0. We see from the above that $z = 0$ if and only if $\langle z, w \rangle = 0$ for all w in V.

DEFINITION 2.4. The *norm* $||\ ||$ corresponding to the Hermitian inner product $\langle\ ,\ \rangle$ is defined by

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

A Hermitian inner product determines a norm, but most norms do not come from inner products. See Exercise 2.5.

EXERCISE 2.1. Verify the three additional properties of the inner product.

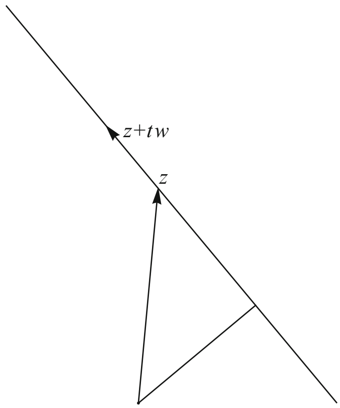

FIGURE 2.1. Proof of the Cauchy–Schwarz inequality

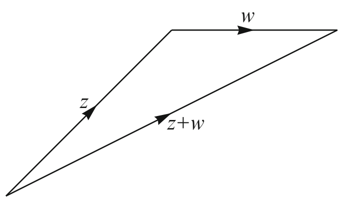

FIGURE 2.2. Triangle inequality

THEOREM 2.1 (The Cauchy–Schwarz and triangle inequalities). *Let V be a complex vector space, let $\langle\ ,\ \rangle$ be a Hermitian inner product on V, and let $||v|| = \sqrt{\langle v, v\rangle}$. The function $||\ ||$ defines a norm on V and the following inequalities hold for all $z, w \in V$:*

$$|\langle z, w\rangle| \leq ||z||\ ||w|| \tag{3}$$

$$||z + w|| \leq ||z|| + ||w||. \tag{4}$$

PROOF. The first two properties of a norm are evident. The first follows from the positive definiteness of the inner product. To prove the second, it suffices to show that $|c|^2||v||^2 = ||cv||^2$. This conclusion follows from

$$||cv||^2 = \langle cv, cv\rangle = c\langle v, cv\rangle = |c|^2\langle v, v\rangle = |c|^2\ ||v||^2.$$

Note that we have used the linearity in the first slot and the conjugate linearity in the second slot. The third property of a norm is the triangle inequality (4).

We first prove the Cauchy–Schwarz inequality (3). For all $t \in \mathbf{C}$, and for all z and w in V,

$$0 \leq ||z + tw||^2 = ||z||^2 + 2\mathrm{Re}\langle z, tw\rangle + |t|^2||w||^2. \tag{5}$$

Think of z and w as fixed, and let ϕ be the quadratic Hermitian polynomial in t and $\bar{t}$ defined by the right-hand side of (5). The values of ϕ are nonnegative; we seek its minimum value by setting its differential equal to 0. (Compare with Exercise 1.13.) We use subscripts to denote the derivatives with respect to t and $\bar{t}$. Since ϕ is real valued, we have $\phi_t = 0$ if and only if $\phi_{\bar{t}} = 0$. From (5) we find

$$\phi_{\bar{t}} = \langle z, w\rangle + t||w||^2.$$

When $w = 0$ we get no useful information, but inequality (3) is true when $w = 0$. To prove (3) when $w \neq 0$, we may set

$$t = \frac{-\langle z, w\rangle}{||w||^2}$$

in (5) and conclude that

$$0 \leq ||z||^2 - 2\frac{|\langle z, w\rangle|^2}{||w||^2} + \frac{|\langle z, w\rangle|^2}{||w||^2} = ||z||^2 - \frac{|\langle z, w\rangle|^2}{||w||^2}. \tag{6}$$

Inequality (6) yields

$$|\langle z, w\rangle|^2 \leq ||z||^2||w||^2,$$

from which (3) follows by taking square roots.

To establish the triangle inequality (4), we begin by squaring its left-hand side:

$$||z + w||^2 = ||z||^2 + 2\mathrm{Re}\langle z, w\rangle + ||w||^2. \tag{7}$$

Since $\mathrm{Re}\langle z, w\rangle \leq |\langle z, w\rangle|$, the Cauchy–Schwarz inequality yields

$$||z + w||^2 = ||z||^2 + 2\mathrm{Re}\langle z, w\rangle + ||w||^2 \leq ||z||^2 + 2||z||\,||w|| + ||w||^2 = (||z|| + ||w||)^2.$$

Taking the square root of each side gives the triangle inequality and completes the proof that $\sqrt{\langle v, v\rangle}$ defines a norm on V. □

In the proof, we noted the identity (7). This (essentially trivial) identity has two significant corollaries.

THEOREM 2.2. *Let V be a complex inner product space. The following hold:*

Pythagorean theorem: $\langle z, w\rangle = 0$ *implies* $||z + w||^2 = ||z||^2 + ||w||^2$.

Parallelogram law: $||z + w||^2 + ||z - w||^2 = 2(||z||^2 + ||w||^2)$.

PROOF. The Pythagorean theorem is immediate from (7), because $\langle z, w\rangle = 0$ implies that $\mathrm{Re}(\langle z, w\rangle) = 0$. The parallelogram law follows from (7) by adding the result in (7) to the result of replacing w by $-w$ in (7). □

The two inequalities from Theorem 2.1 have many consequences. We use them here to show that the inner product and norm on V are (sequentially) continuous functions.

PROPOSITION 2.1 (Continuity of the inner product and the norm). *Let V be a complex vector space with Hermitian inner product and corresponding norm. Let $\{z_n\}$ be a sequence that converges to z in V. Then, for all $w \in V$, the sequence of inner products $\langle z_n, w\rangle$ converges to $\langle z, w\rangle$. Furthermore, $||z_n||$ converges to $||z||$.*

PROOF. By the linearity of the inner product and the Cauchy–Schwarz inequality, we have

$$|\langle z_n, w\rangle - \langle z, w\rangle| = |\langle z_n - z, w\rangle| \leq ||z_n - z||\ ||w||. \tag{8}$$

Thus, when z_n converges to z, the right-hand side of (8) converges to 0, and therefore so does the left-hand side. Thus the inner product (with w) is continuous.

The proof of the second statement uses the triangle inequality. From it we obtain the inequality $||z|| \leq ||z - z_n|| + ||z_n||$ and hence

$$||z|| - ||z_n|| \leq ||z - z_n||.$$

Interchanging the roles of z_n and z gives the same inequality with a negative sign on the left-hand side. Combining these inequalities yields

$$|\ ||z|| - ||z_n||\ | \leq ||z - z_n||,$$

from which the second statement follows. ☐

Suppose that $\sum v_n$ converges in V. For all $w \in V$, we have

$$\langle \sum_n v_n, w\rangle = \sum_n \langle v_n, w\rangle.$$

This conclusion follows by applying Proposition 2.1 to the partial sums of the series. We will often apply this result when working with orthonormal expansions.

Finite-dimensional complex Euclidean spaces are complete in the sense that Cauchy sequences have limits. Infinite-dimensional complex vector spaces with Hermitian inner products need not be complete. By definition, Hilbert spaces are complete.

DEFINITION 2.5. A *Hilbert space* $\mathcal{H}$ is a complex vector space, together with a Hermitian inner product whose corresponding distance function makes $\mathcal{H}$ into a complete metric space.

EXERCISE 2.2. Prove the Cauchy–Schwarz inequality in $\mathbf{R}^n$ by writing $||x||^2||y||^2 - |\langle x, y\rangle|^2$ as a sum of squares. Give the analogous proof in $\mathbf{C}^n$.

EXERCISE 2.3. Prove the Cauchy–Schwarz inequality in $\mathbf{R}^n$ using Lagrange multipliers.

EXERCISE 2.4. Let $\mathcal{H}$ be an inner product space. We showed, for all z and w in $\mathcal{H}$, that (9) holds:

$$||z + w||^2 + ||z - w||^2 = 2||z||^2 + 2||w||^2. \tag{9}$$

Why is this identity called the *parallelogram law*?

EXERCISE 2.5 (Difficult). Let V be a real or complex vector space with a norm. Show that this norm comes from an inner product if and only if the norm satisfies the parallelogram law (9). Comment: Given the norm, one has to define the inner product somehow and then prove that the inner product satisfies all the necessary properties. Use a polarization identity such as (19) to get started.

We give several examples of Hilbert spaces. We cannot verify completeness in the last example without developing the Lebesgue integral. We do, however, make the following remark. Suppose we are given a metric space that is not complete. We may form its completion by considering equivalence classes of Cauchy sequences in a manner similar to defining the real numbers $\mathbf{R}$ as the completion of the rational numbers $\mathbf{Q}$. Given an inner product space, we may complete it into a Hilbert space. The problem is that we wish to have a concrete realization of the limiting objects.

EXAMPLE 2.2. (Hilbert Spaces)

(1) Complex Euclidean space $\mathbf{C}^n$ is a complete metric space with the distance function given by $d(z, w) = ||z - w||$, and hence it is a Hilbert space.

(2) l^2. Let $a = \{a_\nu\}$ denote a sequence of complex numbers. We say that a is *square-summable*, and we write $a \in l^2$, if $||a||_2^2 = \sum_\nu |a_\nu|^2$ is finite. When $a, b \in l^2$ we write

$$\langle a, b\rangle_2 = \sum_\nu a_\nu \overline{b}_\nu$$

for their Hermitian inner product. Exercise 2.6 requests a proof that l^2 is a complete metric space; here $d(a, b) = ||a - b||_2$.

(3) $\mathcal{A}^2(B_1)$. This space consists of all complex analytic functions f on the unit disk B_1 in $\mathbf{C}$ such that $\int_{B_1} |f|^2 dxdy$ is finite. The inner product is given by

$$\langle f, g\rangle = \int_{B_1} f\overline{g}dxdy.$$

(4) $L^2(\Omega)$. Let Ω be an open subset of $\mathbf{R}^n$. Let dV denote Lebesgue measure in $\mathbf{R}^n$. We write $L^2(\Omega)$ for the complex vector space of (equivalence classes of) measurable functions $f : \Omega \to \mathbf{C}$ for which $\int_\Omega |f(x)|^2 dV(x)$ is finite. When f and g are elements of $L^2(\Omega)$, we define their inner product by

$$\langle f, g\rangle = \int_\Omega f(x)\overline{g(x)}dV(x).$$

The corresponding norm and distance function make $L^2(\Omega)$ into a complete metric space, so $L^2(\Omega)$ is a Hilbert space. See [F1] for a proof of completeness.

EXERCISE 2.6. Verify that $\mathbf{C}^n$ and l^2 are complete.

EXERCISE 2.7. Let V be a normed vector space. Show that V is complete if and only if whenever $\sum_n ||v_n||$ converges, then $\sum_n v_n$ converges. Compare with Exercise 1.5.

3. Subspaces and Linear Maps

A *subspace* of a vector space is a subset that is itself a vector space under the same operations of addition and scalar multiplication. A finite-dimensional subspace of a Hilbert space is necessarily closed (in the metric space sense), whereas infinite-dimensional subspaces need not be closed. A closed linear subspace of a Hilbert space is complete and therefore also a Hilbert space. Let B be a bounded domain in $\mathbf{C}^n$. Then $\mathcal{A}^2(B)$ is a closed subspace of $L^2(B)$ and thus a Hilbert space.

Next we define *bounded linear transformations* or *operators*. These mappings are the continuous functions between Hilbert spaces that preserve the vector space structure.

DEFINITION 2.6. Let $\mathcal{H}$ and $\mathcal{H}'$ be Hilbert spaces. A function $L : \mathcal{H} \to \mathcal{H}'$ is called *linear* if it satisfies properties (1) and (2). Also, L is called a *bounded linear transformation* from $\mathcal{H}$ to $\mathcal{H}'$ if L satisfies all three of the following properties:

(1) $L(z_1 + z_2) = L(z_1) + L(z_2)$ for all z_1 and z_2 in $\mathcal{H}$.
(2) $L(cz) = cL(z)$ for all $z \in \mathcal{H}$ and all $c \in \mathbf{C}$.
(3) There is a constant C such that $||L(z)|| \leq C||z||$ for all $z \in \mathcal{H}$.

We write $\mathcal{L}(\mathcal{H}, \mathcal{H}')$ for the collection of bounded linear transformations from $\mathcal{H}$ to $\mathcal{H}'$ and $\mathcal{L}(\mathcal{H})$ for the important special case when $\mathcal{H} = \mathcal{H}'$. In this case, I denotes the identity linear transformation, given by $I(z) = z$. Elements of $\mathcal{L}(\mathcal{H})$ are often called *bounded operators* on $\mathcal{H}$. The collection of bounded operators is an algebra, where composition plays the role of multiplication.

Properties (1) and (2) define the *linearity* of L. Property (3) guarantees the continuity of L; see Lemma 2.1 below. The infimum of the set of constants C that work in (3) provides a measurement of the size of the transformation L; it is called the norm of L and is written $||L||$. Exercise 2.9 justifies the terminology. An equivalent way to define $||L||$ is the formula

$$||L|| = \sup_{\{z \neq 0\}} \frac{||L(z)||}{||z||}.$$

The set $\mathcal{L}(\mathcal{H}, \mathcal{H}')$ becomes a complete normed vector space. See Exercise 2.9.

We next discuss the relationship between boundedness and continuity for linear transformations.

LEMMA 2.1. *Assume $L : \mathcal{H} \to \mathcal{H}'$ is linear. The following three statements are equivalent:*

(1) *There is a constant $C > 0$ such that, for all z,*

$$||Lz|| \leq C||z||.$$

(2) *L is continuous at the origin.*
(3) *L is continuous at every point.*

PROOF. It follows from the ϵ-δ definition of continuity at a point and the linearity of L that statements (1) and (2) are equivalent. Statement (3) implies statement (2). Statement (1) and the linearity of L imply statement (3) because

$$||Lz - Lw|| = ||L(z - w)|| \leq C||z - w||.$$

□

We associate two natural subspaces with a linear mapping.

DEFINITION 2.7. For $L \in \mathcal{L}(\mathcal{H}, \mathcal{H}')$, the *nullspace* $\mathcal{N}(L)$ is the set of $v \in \mathcal{H}$ for which $L(v) = 0$. The *range* $\mathcal{R}(L)$ is the set of $w \in \mathcal{H}'$ for which there is a $v \in \mathcal{H}$ with $L(v) = w$.

Definition 2.8. An operator $P \in \mathcal{L}(\mathcal{H})$ is a *projection* if $P^2 = P$.

Observe (see Exercise 2.11) that $P^2 = P$ if and only if $(I-P)^2 = I-P$. Thus $I-P$ is also a projection if P is. Furthermore, in this case, $\mathcal{R}(P) = \mathcal{N}(I-P)$ and $\mathcal{H} = \mathcal{R}(P) + \mathcal{N}(P)$.

Bounded linear functionals, that is, elements of $\mathcal{L}(\mathcal{H}, \mathbf{C})$, are especially important. The vector space of bounded linear functionals on $\mathcal{H}$ is called the *dual space* of $\mathcal{H}$. We characterize this space in Theorem 2.4 below.

Definition 2.9. A *bounded linear functional* on a Hilbert space $\mathcal{H}$ is a bounded linear transformation from $\mathcal{H}$ to $\mathbf{C}$.

One of the major results in pure and applied analysis is the Riesz lemma, Theorem 2.4 below. A bounded linear functional on a Hilbert space is always given by an inner product. In order to prove this basic result, we develop material on orthogonality that also particularly illuminates our work on Fourier series.

Exercise 2.8. For $L \in \mathcal{L}(\mathcal{H}, \mathcal{H}')$, verify that $\mathcal{N}(L)$ is a subspace of $\mathcal{H}$ and $\mathcal{R}(L)$ is a subspace of $\mathcal{H}'$.

Exercise 2.9. With $||L||$ defined as above, show that $\mathcal{L}(\mathcal{H})$ is a complete normed vector space.

Exercise 2.10. Show by using a basis that a linear functional on $\mathbf{C}^n$ is given by an inner product.

Exercise 2.11. Let P be a projection. Verify that $I-P$ is a projection, that $\mathcal{R}(P) = \mathcal{N}(I-P)$, and that $\mathcal{H} = \mathcal{R}(P) + \mathcal{N}(P)$.

4. Orthogonality

Let $\mathcal{H}$ be a Hilbert space, and suppose $z, w \in \mathcal{H}$. We say that z and w are *orthogonal* if $\langle z, w \rangle = 0$. The Pythagorean theorem indicates that orthogonality generalizes perpendicularity and provides geometric insight in the general Hilbert space setting. The term "orthogonal" applies also for subspaces. Subspaces V and W of $\mathcal{H}$ are orthogonal if $\langle v, w \rangle = 0$ for all $v \in V$ and $w \in W$. We say that z is orthogonal to V if $\langle z, v \rangle = 0$ for all v in V, or equivalently, if the one-dimensional subspace generated by z is orthogonal to V.

Let V and W be orthogonal closed subspaces of a Hilbert space; $V \oplus W$ denotes their orthogonal sum. It is the subspace of $\mathcal{H}$ consisting of those z that can be written $z = v + w$, where $v \in V$ and $w \in W$. We sometimes write $z = v \oplus w$ in order to emphasize orthogonality. By the Pythagorean theorem, $||v \oplus w||^2 = ||v||^2 + ||w||^2$. Thus $v \oplus w = 0$ if and only if both $v = 0$ and $w = 0$.

We now study the geometric notion of orthogonal projection onto a closed subspace. The next theorem guarantees that we can *project* a vector w in a Hilbert space onto a closed subspace. This existence and uniqueness theorem has diverse corollaries.

Theorem 2.3. *Let V be a closed subspace of a Hilbert space $\mathcal{H}$. For each w in $\mathcal{H}$, there is a unique $z \in V$ that minimizes $||z - w||$. This z is the orthogonal projection of w onto V.*

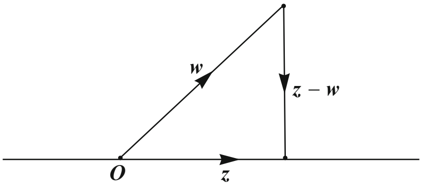

FIGURE 2.3. Orthogonal projection

PROOF. Fix w. If $w \in V$, then the conclusion holds with $z = w$. In general, let $d = \inf_{z \in V} ||z - w||$. Choose a sequence $\{z_n\}$ such that $z_n \in V$ for all n and $||z_n - w||$ tends to d. We will show that $\{z_n\}$ is a Cauchy sequence, and hence it converges to some z. Since V is closed, z is in V. By continuity of the norm (Proposition 2.1), $||z - w|| = d$.

By the parallelogram law, we express $||z_n - z_m||^2$ as follows:

$$||z_n - z_m||^2 = ||(z_n - w) + (w - z_m)||^2 = 2||z_n - w||^2 + 2||w - z_m||^2 - ||(z_n - w) - (w - z_m)||^2.$$

The last term on the right-hand side is

$$4||\frac{z_n + z_m}{2} - w||^2.$$

Since V is a subspace, the midpoint $\frac{z_n + z_m}{2}$ lies in V as well. Therefore, this term is at least $4d^2$, and we obtain

$$0 \le ||z_n - z_m||^2 \le 2||z_n - w||^2 + 2||w - z_m||^2 - 4d^2. \tag{10}$$

As m and n tend to infinity, the right-hand side of (10) tends to $2d^2 + 2d^2 - 4d^2 = 0$. Thus, $\{z_n\}$ is a Cauchy sequence in $\mathcal{H}$ and hence converges to some z in V.

It remains only to show uniqueness. Given a pair of minimizers z and ζ, let d_m^2 denote the squared distance from their midpoint to w. By the parallelogram law, we may write

$$2d^2 = ||z - w||^2 + ||\zeta - w||^2 = 2||\frac{z + \zeta}{2} - w||^2 + 2||\frac{z - \zeta}{2}||^2 = 2d_m^2 + 2||\frac{z - \zeta}{2}||^2.$$

Thus $d^2 \ge d_m^2$. But d is minimal. Hence $d_m = d$ and thus $\zeta = z$. □

COROLLARY 2.1. *Let V be a closed subspace of a Hilbert space $\mathcal{H}$. For each $w \in \mathcal{H}$, there is a unique way to write $w = v + \zeta = v \oplus \zeta$, where $v \in V$ and ζ is orthogonal to V.*

PROOF. Let v be the projection of w onto V guaranteed by Theorem 2.3. Since $w = v + (w - v)$, the existence result follows if we can show that $w - v$ is orthogonal to V. To see the orthogonality, choose $u \in V$. Then consider the function f of one complex variable defined by

$$f(\lambda) = ||v + \lambda u - w||^2.$$

By Theorem 2.3, f achieves its minimum at $\lambda = 0$. Therefore, for all λ,

$$0 \le f(\lambda) - f(0) = 2\mathrm{Re}\langle v - w, \lambda u\rangle + |\lambda|^2||u||^2. \tag{11}$$

We claim that (11) forces $\langle v - w, u\rangle = 0$. Granted the claim, we note that u is an arbitrary element of V. Therefore, $v - w$ is orthogonal to V, as required.

To prove the claim, thereby completing the proof of existence, we note that $\langle v - w, u\rangle$ is the (partial) derivative of f with respect to $\overline{\lambda}$ at 0 and hence vanishes at a minimum of f.

The uniqueness assertion is easy; we use the notation for orthogonal sum. Suppose $w = v \oplus \zeta = v' \oplus \zeta'$, as in the statement of the Corollary. Then

$$0 = w - w = (v - v') \oplus (\zeta - \zeta')$$

from which we obtain $v = v'$ and $\zeta = \zeta'$. □

Corollary 2.2. *Let V be a closed subspace of a Hilbert space $\mathcal{H}$. For each $w \in \mathcal{H}$, let Pw denote the unique $z \in V$ guaranteed by Theorem 2.3; Pw is also the v guaranteed by Corollary 2.1. Then the mapping $w \to P(w)$ is a bounded linear transformation satisfying $P^2 = P$. Thus P is a projection.*

Proof. Both the existence and uniqueness assertions in Corollary 2.1 matter in this proof. Given w_1 and w_2 in $\mathcal{H}$, by existence, we may write $w_1 = Pw_1 \oplus \zeta_1$ and $w_2 = Pw_2 \oplus \zeta_2$. Adding gives

$$w_1 + w_2 = (Pw_1 \oplus \zeta_1) + (Pw_2 \oplus \zeta_2) = (Pw_1 + Pw_2) \oplus (\zeta_1 + \zeta_2). \tag{12}$$

The uniqueness assertion and (12) show that $Pw_1 + Pw_2$ is the unique element of V corresponding to $w_1 + w_2$ guaranteed by Corollary 2.1; by definition this element is $P(w_1 + w_2)$. By uniqueness $Pw_1 + Pw_2 = P(w_1 + w_2)$, and P is additive. In a similar fashion, we write $w = Pw \oplus \zeta$ and hence

$$cw = c(Pw) \oplus c\zeta.$$

Again by uniqueness, $c(Pw)$ must be the unique element corresponding to cw guaranteed by Corollary 2.1; by definition this element is $P(cw)$. Hence $cP(w) = P(cw)$. We have now shown that P is linear.

To show that P is bounded, we note from the Pythagorean theorem that $||w||^2 = ||Pw||^2 + ||\zeta||^2$, and hence $||Pw|| \leq ||w||$.

Finally we show that $P^2 = P$. For $z = v \oplus \zeta$, we have $P(z) = v = v \oplus 0$. Hence

$$P^2(z) = P(P(z)) = P(v \oplus 0) = v = P(z).$$

□

Theorem 2.3 and its consequences are among the most powerful results in the book. The theorem guarantees that we can solve a minimization problem in diverse infinite-dimensional settings, and it implies the Riesz representation lemma.

Fix $w \in \mathcal{H}$, and consider the function from $\mathcal{H}$ to $\mathbf{C}$ defined by $Lz = \langle z, w\rangle$. Then L is a bounded linear functional. The linearity is evident. The boundedness follows from the Cauchy–Schwarz inequality; setting $C = ||w||$ yields $|L(z)| \leq C||z||$ for all $z \in \mathcal{H}$.

The following fundamental result of F. Riesz characterizes bounded linear functionals on a Hilbert space; a bounded linear functional must be given by an inner product. The proof relies on projection onto a closed subspace.

THEOREM 2.4 (Riesz lemma). *Let $\mathcal{H}$ be a Hilbert space and suppose that $L \in \mathcal{L}(\mathcal{H}, \mathbf{C})$. Then there is a unique $w \in \mathcal{H}$ such that*

$$L(z) = \langle z, w \rangle$$

for all $z \in \mathcal{H}$. The norm $||L||$ of the linear transformation L equals $||w||$.

PROOF. Since L is bounded, its nullspace $\mathcal{N}(L)$ is closed. If $\mathcal{N}(L) = \mathcal{H}$, we take $w = 0$, and the result is true.

Suppose that $\mathcal{N}(L)$ is not $\mathcal{H}$. Theorem 2.3 implies that there is a nonzero element w_0 orthogonal to $\mathcal{N}(L)$. To find such a w_0, choose any nonzero element not in $\mathcal{N}(L)$ and subtract its orthogonal projection onto $\mathcal{N}(L)$.

Let z be an arbitrary element of $\mathcal{H}$. For a complex number α, we can write

$$z = (z - \alpha w_0) + \alpha w_0.$$

Note that $L(z - \alpha w_0) = 0$ if and only if $\alpha = \frac{L(z)}{L(w_0)}$. For each z, we therefore let $\alpha_z = \frac{L(z)}{L(w_0)}$.

Since w_0 is orthogonal to $\mathcal{N}(L)$, computing the inner product with w_0 yields

$$\langle z, w_0 \rangle = \alpha_z ||w_0||^2 = \frac{L(z)}{L(w_0)} ||w_0||^2. \tag{13}$$

From (13) we see that

$$L(z) = \langle z, \frac{w_0}{||w_0||^2} \overline{L(w_0)} \rangle$$

and the existence result is proved. An explicit formula for w holds:

$$w = \frac{w_0}{||w_0||^2} \overline{L(w_0)}.$$

The uniqueness for w is immediate from the test we mentioned earlier. If $\langle \zeta, w - w' \rangle$ vanishes for all ζ, then $w - w' = 0$.

It remains to show that $||L|| = ||w||$. The Cauchy–Schwarz inequality yields

$$||L|| = \sup_{||z||=1} |\langle z, w \rangle| \leq ||w||.$$

Choosing $\frac{w}{||w||}$ for z yields

$$||L|| \geq |L(\frac{w}{||w||})| = \frac{\langle w, w \rangle}{||w||} = ||w||.$$

Combining the two inequalities shows that $||L|| = ||w||$. □

EXERCISE 2.12. Fix w with $w \neq 0$. Define $P(v)$ by

$$P(v) = \frac{\langle v, w \rangle}{||w||^2} w.$$

Verify that $P^2 = P$.

EXERCISE 2.13. Let $\mathcal{H} = L^2[-1,1]$. Recall that f is *even* if $f(-x) = f(x)$ and f is *odd* if $f(-x) = -f(x)$. Let V_e be the subspace of even functions and V_o the subspace of odd functions. Show that V_e is orthogonal to V_o.

EXERCISE 2.14. A hyperplane in $\mathcal{H}$ is a level set of a nontrivial linear functional. Assume that $w \neq 0$. Find the distance between the parallel hyperplanes given by $\langle z, w \rangle = c_1$ and $\langle z, w \rangle = c_2$.

EXERCISE 2.15. Let $b = \{b_j\}$ be a sequence of complex numbers, and suppose there is a positive number C such that

$$\left| \sum_{j=1}^{\infty} a_j \overline{b}_j \right| \leq C (\sum_{j=1}^{\infty} |a_j|^2)^{\frac{1}{2}}$$

for all $a \in l^2$. Show that $b \in l^2$ and that $\sum |b_j|^2 \leq C^2$. Suggestion: Consider the map that sends a to $\sum a_j \overline{b}_j$.

5. Orthonormal Expansion

We continue our general discussion of Hilbert spaces by studying orthonormal expansions. The simplest example comes from basic physics. Let $v = (a, b, c)$ be a point or vector in $\mathbf{R}^3$. Physicists write $v = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$, where $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are mutually perpendicular vectors of unit length. Mathematicians write the same equation as $v = ae_1 + be_2 + ce_3$; here $e_1 = (1, 0, 0) = \mathbf{i}$, $e_2 = (0, 1, 0) = \mathbf{j}$, and $e_3 = (0, 0, 1) = \mathbf{k}$. This equation expresses v in terms of an *orthonormal expansion*:

$$ae_1 + be_2 + ce_3 = (a, b, c) = v = \langle v, e_1 \rangle e_1 + \langle v, e_2 \rangle e_2 + \langle v, e_3 \rangle e_3.$$

Orthonormal expansion in a Hilbert space abstracts this idea. Fourier series provide the basic example, where the functions $x \to e^{inx}$ are analogous to mutually perpendicular unit vectors.

We assume here that a Hilbert space is *separable*. This term means that the Hilbert space has a countable dense set; separability implies that the orthonormal systems we are about to define are either finite or countably infinite sets. All the specific Hilbert spaces mentioned or used in this book are separable. Some of the proofs given tacitly use separability even when the result holds more generally.

DEFINITION 2.10. Let $S = \{z_n\}$ be a finite or countably infinite collection of elements in a Hilbert space $\mathcal{H}$. We say that S is an *orthonormal system* in $\mathcal{H}$ if, for each n we have $||z_n||^2 = 1$, and for each n, m with $n \neq m$, we have $\langle z_n, z_m \rangle = 0$. We say that S is a *complete orthonormal system* if, in addition, $\langle z, z_n \rangle = 0$ for all n implies $z = 0$.

PROPOSITION 2.2 (Bessel's inequality). *Let $S = \{z_n\}$ be a countably infinite orthonormal system in $\mathcal{H}$. For each $z \in \mathcal{H}$, we have*

$$\sum_{n=1}^{\infty} |\langle z, z_n \rangle|^2 \leq ||z||^2. \tag{14}$$

PROOF. Choose $z \in \mathcal{H}$. By orthonormality, for each positive integer N, we have

$$0 \leq ||z - \sum_{n=1}^{N} \langle z, z_n \rangle z_n||^2 = ||z||^2 - \sum_{n=1}^{N} |\langle z, z_n \rangle|^2. \tag{15}$$

Define a sequence of real numbers $r_N = r_N(z)$ by

$$r_N = \sum_{n=1}^{N} |\langle z, z_n \rangle|^2.$$

By (15), r_N is bounded above by $||z||^2$ and nondecreasing. Therefore, it has a limit $r = r(z)$. Bessel's inequality follows. □

PROPOSITION 2.3 (Best approximation lemma). *Let $S = \{z_n\}$ be an orthonormal system (finite or countable) in $\mathcal{H}$. Let V be the span of S. Then, for each $z \in \mathcal{H}$ and each $w \in V$,*

$$||z - \sum \langle z, z_n \rangle z_n|| \leq ||z - w||.$$

PROOF. The expression $\sum \langle z, z_n \rangle z_n$ equals the orthogonal projection of z onto V. Hence the result follows from Theorem 2.3. □

The limit $r(z)$ of the sequence in the proof of Bessel's inequality equals $||z||^2$ for each z if and only if the orthonormal system S is *complete*. This statement is the content of the following fundamental theorem. In general, $r(z)$ is the squared norm of the projection of z onto the span of the z_j.

THEOREM 2.5 (Orthonormal expansion). *An orthonormal system $S = \{z_n\}$ is complete if and only if, for each $z \in \mathcal{H}$, we have*

$$z = \sum_{n} \langle z, z_n \rangle z_n. \tag{16}$$

PROOF. The cases where S is a finite set or where $\mathcal{H}$ is finite-dimensional are evident. Assume then that $\mathcal{H}$ is infinite dimensional and S is a countably infinite set. We first verify that the series in (16) converges. Fix $z \in \mathcal{H}$, and put

$$T_N = \sum_{n=1}^{N} \langle z, z_n \rangle z_n.$$

Define r_N as in the proof of Bessel's inequality. For $N > M$, observe that

$$||T_N - T_M||^2 = ||\sum_{n=M+1}^{N} \langle z, z_n \rangle z_n||^2 = \sum_{n=M+1}^{N} |\langle z, z_n \rangle|^2 = r_N - r_M. \tag{17}$$

Since $\{r_N\}$ converges, it is a Cauchy sequence of real numbers. By (17), $\{T_N\}$ is a Cauchy sequence in $\mathcal{H}$. Since $\mathcal{H}$ is complete, T_N converges to some element w of $\mathcal{H}$, and $w = \sum \langle z, z_n \rangle z_n$, the right-hand side of (16). Note that $\langle w, z_n \rangle = \langle z, z_n \rangle$ for each n, so $z - w$ is orthogonal to each z_n.

We can now establish both implications. Suppose first that S is a complete system. Since $z - w$ is orthogonal to each z_n, we have $z - w = 0$. Thus (16) holds. Conversely, suppose that (16) holds. To show that S is a complete system, we assume that $\langle z, z_n \rangle = 0$ for all n and hope to show that $z = 0$. This conclusion follows immediately from (16). $\square$

EXERCISE 2.16. Verify (15).

EXERCISE 2.17. Let $\mathcal{H} = L^2([0,1])$ with the usual inner product. Let V be the span of 1 and x. Find the orthogonal projection of x^2 onto V. Do the same problem if $\mathcal{H} = L^2([-1,1])$.

EXERCISE 2.18. Let $\mathcal{H} = L^2([-1,1])$ with the usual inner product. Apply the Gram–Schmidt process (see [G]) to orthonormalize the polynomials $1, x, x^2, x^3$.

EXERCISE 2.19. A sequence $\{f_n\}$ in a Hilbert space $\mathcal{H}$ *converges weakly* to f if, for each $g \in \mathcal{H}$, the sequence $\{\langle f_n, g \rangle\}$ converges to $\langle f, g \rangle$. Put $\mathcal{H} = L^2([0, 2\pi])$. Put $f_n(x) = \sin(nx)$. Show that $\{f_n\}$ converges weakly to 0, but does not converge to 0.

EXERCISE 2.20. Assume $\mathcal{H}$ is infinite dimensional. Show that a sequence of orthonormal vectors does not converge, but does converge weakly to 0.

6. Polarization

In a Hilbert space, we can recover the Hermitian inner product from the squared norm. In addition, for each linear operator L, we can recover $\langle Lz, w \rangle$ for all z, w from knowing $\langle Lz, z \rangle$ for all z. See Theorem 2.6. The corresponding result for real vector spaces with inner products fails.

To introduce these ideas, let m be an integer with $m \geq 2$. Recall, for a complex number $a \neq 1$, the sum of the finite geometric series:

$$1 + a + a^2 + \cdots + a^{m-1} = \frac{1 - a^m}{1 - a}.$$

When a is an m-th root of unity, the sum is zero. A *primitive* m-th root of unity is a complex number ω such that $\omega^m = 1$, but no smaller positive power equals 1. The set of powers ω^j for $j = 0, 1, \ldots, m-1$ forms a cyclic group Γ of order m.

Let z, ζ be elements of a Hilbert space $\mathcal{H}$. Let ω be a primitive m-th root of unity and consider averaging the m complex numbers $\gamma ||z + \gamma\zeta||^2$ as γ varies over Γ. Since each group element is a power of ω, this average equals

$$\frac{1}{m} \sum_{j=0}^{m-1} \omega^j ||z + \omega^j \zeta||^2.$$

The next proposition gives a simple expression for the average.

PROPOSITION 2.4 (Polarization identities). *Let ω be a primitive m-th root of unity. For $m \geq 3$, we have*

$$\langle z, \zeta \rangle = \frac{1}{m} \sum_{j=0}^{m-1} \omega^j ||z + \omega^j \zeta||^2. \tag{18}$$

For $m = 2$, the right-hand side of (18) equals $2\mathrm{Re}\langle z, \zeta \rangle$.

PROOF. We prove (18) below when $m = 4$, leaving the general case to the reader. $\square$

For $m \geq 3$, each identity in (18) expresses the inner product in terms of squared norms. It is both beautiful and useful to recover the inner product from the squared norm. The special case of (18) where $m = 4$, and thus $\omega = i$, arises often. We state it explicitly and prove it:

$$4\langle z, \zeta\rangle = ||z+\zeta||^2 + i||z+i\zeta||^2 - ||z-\zeta||^2 - i||z-i\zeta||^2. \tag{19}$$

To verify (19), observe that expanding the squared norms gives both equations:

$$4\mathrm{Re}\langle z, \zeta\rangle = ||z+\zeta||^2 - ||z-\zeta||^2$$

$$4\mathrm{Re}\langle z, i\zeta\rangle = ||z+i\zeta||^2 - ||z-i\zeta||^2.$$

Observe for $a \in \mathbf{C}$ that $\mathrm{Re}(-ia) = \mathrm{Im}(a)$. Thus, multiplying the second equation by i, using $i(-i) = 1$, and then adding the two equations give (19).

In addition to polarizing the inner product, we often polarize expressions involving linear transformations.

THEOREM 2.6 (Polarization identities for operators). *Let $L \in \mathcal{L}(\mathcal{H})$. Let ω be a primitive m-th root of unity.*

(1) *For $m \geq 3$, we have*

$$\langle Lz, \zeta\rangle = \frac{1}{m}\sum_{j=0}^{m-1} \omega^j \langle L(z+\omega^j\zeta), z+\omega^j\zeta\rangle. \tag{20}$$

(2) *For $m = 2$, we have*

$$\langle Lz, \zeta\rangle + \langle L\zeta, z\rangle = \frac{1}{2}(\langle L(z+\zeta), z+\zeta\rangle - \langle L(z-\zeta), z-\zeta\rangle). \tag{21}$$

(3) *Suppose in addition that $\langle Lv, v\rangle$ is real for all $v \in \mathcal{H}$. Then, for all z and ζ,*

$$\langle Lz, \zeta\rangle = \overline{\langle L\zeta, z\rangle}.$$

(4) *Suppose $\langle Lz, z\rangle = 0$ for all z. Then $L = 0$.*

PROOF. To prove (20) and (21), expand each $\langle L(z+\omega^j\zeta), z+\omega^j\zeta\rangle$ using the linearity of L and the defining properties of the inner product. Collect similar terms, and use the above comment about roots of unity. For $m \geq 3$, all terms inside the sum cancel except for m copies of $\langle Lz, \zeta\rangle$. The result gives (20). For $m = 2$, the coefficient of $\langle L\zeta, z\rangle$ does not vanish, and we obtain (21). Thus statements (1) and (2) hold.

To prove the third statement, we apply the first for some m with $m \geq 3$ and $\omega^m = 1$; the result is

$$\langle Lz, \zeta\rangle = \frac{1}{m}\sum_{j=0}^{m-1} \omega^j \langle L(z+\omega^j\zeta), z+\omega^j\zeta\rangle = \frac{1}{m}\sum_{j=0}^{m-1} \omega^j \langle L(\omega^{m-j}z+\zeta), \omega^{m-j}z+\zeta\rangle. \tag{22}$$

Change the index of summation by setting $l = m - j$. Also observe that $\omega^{-1} = \overline{\omega}$. Combining gives the first equality in (23) below. Finally, because $\langle Lv, v\rangle$ is real and $\omega^0 = \omega^m$, we obtain the second equality in (23):

$$\langle Lz, \zeta\rangle = \frac{1}{m}\sum_{l=1}^{m} \overline{\omega}^l \langle L(\zeta + \omega^l z), \zeta + \omega^l z\rangle = \overline{\langle L\zeta, z\rangle}. \tag{23}$$

We have now proved the third statement.

The fourth statement follows from (20); each term in the sum on the right-hand side of (20) vanishes if $\langle Lw, w\rangle = 0$ for all w. Thus $\langle Lz, \zeta\rangle = 0$ for all ζ. Hence $Lz = 0$ for all z, and thus $L = 0$. □

The reader should compare these results about polarization with our earlier results about Hermitian symmetric polynomials.

EXERCISE 2.21. Give an example of a linear map of $\mathbf{R}^2$ such that $\langle Lu, u\rangle = 0$ for all u but L is not 0.

7. Adjoints and Unitary Operators

Let I denote the identity linear transformation on a Hilbert space $\mathcal{H}$. Let $L \in \mathcal{L}(\mathcal{H})$. Then L is called *invertible* if there is a bounded linear mapping T such that $LT = TL = I$. If such a T exists, then T is unique and written L^{-1}. We warn the reader (see the exercises) that, in infinite dimensions, $LT = I$ does not imply that L is invertible. When L is bounded, injective, and surjective, the usual set-theoretic inverse is also linear and bounded.

Given a bounded linear mapping L, the adjoint of L is written L^*. It is defined as follows. Fix $v \in \mathcal{H}$. Consider the map $u \to \langle Lu, v\rangle = \phi_v(u)$. It is obviously a linear functional. It is also continuous because

$$|\phi_v(u)| = |\langle Lu, v\rangle| \le ||Lu||\ ||v|| \le ||u||\ ||L||\ ||v|| = c||u||, \tag{24}$$

where the constant c is independent of u. By Theorem 2.4, there is a unique $w_v \in \mathcal{H}$ for which $\phi_v(u) = \langle u, w_v\rangle$. We denote w_v by L^*v. It is easy to prove that L^* is itself a bounded linear mapping on $\mathcal{H}$, called the adjoint of L.

The following properties of adjoints are left as exercises:

PROPOSITION 2.5. *Let $L, T \in \mathcal{L}(\mathcal{H})$. The following hold:*

(1) $L^* : \mathcal{H} \to \mathcal{H}$ *is linear.*
(2) L^* *is bounded. (In fact $||L^*|| = ||L||$.)*
(3) $(L^*)^* = L$.
(4) $\langle Lu, v\rangle = \langle u, L^*v\rangle$ *for all u, v.*
(5) $(LT)^* = T^*L^*$.

PROOF. See Exercise 2.22. □

EXERCISE 2.22. Prove Proposition 2.5.

DEFINITION 2.11. A bounded linear transformation L on a Hilbert space $\mathcal{H}$ is called *Hermitian* or *self-adjoint* if $L = L^*$. It is called *unitary* if it is invertible and $L^* = L^{-1}$.

The following simple but beautiful result characterizes unitary transformations:

PROPOSITION 2.6. *The following are equivalent for* $L \in \mathcal{L}(\mathcal{H})$*:*

(1) L *is surjective and preserves norms:* $||Lu||^2 = ||u||^2$ *for all* u*.*
(2) L *is surjective and preserves inner products:* $\langle Lu, Lv\rangle = \langle u, v\rangle$ *for all* u, v*.*
(3) L *is unitary:* $L^* = L^{-1}$*.*

PROOF. If $L \in \mathcal{L}(\mathcal{H})$, then $\langle Lu, Lv\rangle = \langle u, v\rangle$ for all u, v if and only if $\langle u, L^*Lv\rangle = \langle u, v\rangle$ for all u, v and thus if and only if $\langle u, (L^*L - I)v\rangle = 0$ for all u, v. This last statement holds if and only if $(L^*L - I)v = 0$ for all v. Thus $L^*L = I$. If L is also surjective, then $L^* = L^{-1}$, and therefore the second and third statements are equivalent.

The second statement obviously implies the first. It remains to prove the subtle point that the first statement implies the second or third statement. We are given $\langle L^*Lz, z\rangle = \langle z, z\rangle$ for all z. Hence $\langle (L^*L - I)z, z\rangle = 0$. By part 4 of Theorem 2.6, $L^*L - I = 0$, and the second statement holds. If L is also surjective, then L is invertible and hence unitary. $\square$

The equivalence of the first two statements does not require L to be surjective. See the exercises for examples where L preserves inner products, but L is not surjective and hence not unitary.

PROPOSITION 2.7. *Let* $L \in \mathcal{L}(\mathcal{H})$*. Then*

$$\mathcal{N}(L) = \mathcal{R}(L^*)^{\perp}$$
$$\mathcal{N}(L^*) = \mathcal{R}(L)^{\perp}.$$

PROOF. Note that $L^*(z) = 0$ if and only if $\langle L^*z, w\rangle = 0$ for all w, if and only if $\langle z, Lw\rangle = 0$ for all w, and if and only if $z \perp \mathcal{R}(L)$. Thus, the second statement holds. When $L \in \mathcal{L}(\mathcal{H})$, it is easy to check that $(L^*)^* = L$. See Exercise 2.22. The first statement then follows from the second statement by replacing L with L^*. $\square$

EXERCISE 2.23. If $L : \mathbf{C}^n \to \mathbf{C}^n$ and $L = L^*$, what can we conclude about the matrix of L with respect to the usual basis $(1, 0, \dots, 0)$, $\dots$, $(0, 0, \dots, 1)$?

EXERCISE 2.24. Suppose U is unitary and $Uz = \lambda z$ for $z \neq 0$. Prove that $|\lambda| = 1$. Suppose L is Hermitian and $Lz = \lambda z$ for $z \neq 0$. Prove that λ is real.

EXERCISE 2.25. Let $L : l^2 \to l^2$ be defined by

$$L(z_1, z_2, \dots) = (0, z_1, z_2, \dots).$$

Show that $||Lz||_2 = ||z||_2$ for all z but that L is not unitary.

EXERCISE 2.26. Give an example of a bounded linear $L : \mathcal{H} \to \mathcal{H}$ that is injective but not surjective and an example that is surjective but not injective.

EXERCISE 2.27. Let V be the vector space of all polynomials in one variable. Let D denote differentiation and J denote integration (with integration constant 0). Show that $DJ = I$ but that $JD \neq I$. Explain.

EXERCISE 2.28. Give an example of an operator L for which $||L^2|| \neq ||L||^2$. Suppose $L = L^*$; show that $||L^2|| = ||L||^2$.

We close this section with an interesting difference between real and complex vector spaces, related to inverses, polarization, and Exercise 2.21. The formula (*) below interests the author partly because, although no real numbers satisfy the equation, teachers often see it on exams.

DEFINITION 2.12. A real vector space V *admits a complex structure* if there is a linear map $J : V \to V$ such that $J^2 = -I$.

It is easy to show (Exercise 2.30) that a finite-dimensional real vector space admits a complex structure if and only if its dimension is even. The linear transformation $J : \mathbf{R}^2 \to \mathbf{R}^2$ corresponding to the complex structure is given by the matrix

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

PROPOSITION 2.8. *Let V be a vector space over $\mathbf{R}$. Then there are invertible linear transformations A, B on V satisfying*

$$(A + B)^{-1} = A^{-1} + B^{-1} \tag{*}$$

if and only if V admits a complex structure.

PROOF. Invertible A, B satisfying (*) exist if and only if

$$I = (A + B)(A^{-1} + B^{-1}) = I + BA^{-1} + I + AB^{-1}.$$

Put $C = BA^{-1}$. The condition (*) is therefore equivalent to finding C such that $0 = I + C + C^{-1}$, which is equivalent to $0 = I + C + C^2$. Suppose such C exists. Put $J = \frac{1}{\sqrt{3}}(I + 2C)$. Then we have

$$J^2 = \frac{1}{3}(I + 2C)^2 = \frac{1}{3}(I + 4C + 4C^2) = \frac{1}{3}(-3I + 4(I + C + C^2)) = -I.$$

Hence V admits a complex structure. Conversely, if V admits a complex structure, then J exists with $J^2 = -I$. Put $C = \frac{-I+\sqrt{3}J}{2}$; then $I + C + C^2 = 0$. $\square$

COROLLARY 2.3. *There exist n by n matrices satisfying (*) if and only if n is even.*

EXERCISE 2.29. Explain the proof of Proposition 2.8 in terms of cube roots of unity.

EXERCISE 2.30. Prove that a finite-dimensional real vector space with a complex structure must have even dimension. Hint: Consider the determinant of J.

8. A Return to Fourier Series

The specific topic of Fourier series motivated many of the abstract results about Hilbert spaces, and it provides one of the best examples of the general theory. In return, the general theory clarifies the subject of Fourier series.

Let h be (Riemann) integrable on the circle and consider its Fourier series $\sum \hat{h}(n)e^{inx}$. Recall that its symmetric partial sums S_N are given by

$$S_N(h)(x) = \sum_{n=-N}^{N} \hat{h}(n)e^{inx}.$$

When h is sufficiently smooth, $S_N(h)$ converges to h. See, for example, Theorem 2.8. We show next that $S_N(h)$ converges to h in L^2. Rather than attempting to prove convergence at each point, this result considers an integrated form of convergence.

THEOREM 2.7. *Suppose f is integrable on the circle. Then $||S_N(f)-f||_{L^2} \to 0$.*

PROOF. Given $\epsilon > 0$ and an integrable f, we first approximate f to within $\frac{\epsilon}{2}$ in the L^2 norm by a continuous function g. Then we approximate g by a trig polynomial p to within $\frac{\epsilon}{2}$. See below for details. These approximations yield

$$||f-p||_{L^2} \leq ||f-g||_{L^2} + ||g-p||_{L^2} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \tag{25}$$

Once we have found this p, we use orthogonality as in Theorem 2.3. Let N be at least as large as the degree of p. Let V_N denote the $(2N+1)$-dimensional (hence closed) subspace spanned by the functions e^{inx} for $|n| \leq N$. By Theorem 2.3, there is a unique element w of V_N minimizing $||f-w||_{L^2}$. That w is the partial sum $S_N(f)$, namely, the orthogonal projection of f onto V_N.

By Proposition 2.3, we have

$$||f - S_N(f)||_{L^2} \leq ||f-p||_{L^2} \tag{26}$$

for all elements p of V_N. Take p to be the polynomial in (25) and take N at least the degree of p. Combining (26) and (25) then gives

$$||f - S_N(f)||_{L^2} \leq ||f-p||_{L^2} \leq ||f-g||_{L^2} + ||g-p||_{L^2} < \epsilon. \tag{27}$$

It therefore suffices to verify that the two above approximations are valid. Given f integrable, by Lemma 1.6 we can find a continuous g such that $\sup(|g|) \leq \sup(|f|) = M$ and such that $||f-g||_{L^1}$ is as small as we wish. Since

$$||f-g||_{L^2}^2 = \frac{1}{2\pi}\int_0^{2\pi} |f-g|^2 dx \leq \frac{\sup(|f-g|)}{2\pi}\int_0^{2\pi} |f-g| dx \leq 2M||f-g||_{L^1}, \tag{28}$$

we may choose g to bound the expression in (28) by $\frac{\epsilon}{2}$.

Now g is given and continuous on the circle. By Corollary 1.8, there is a trig polynomial p such that $||g-p||_{L^\infty} < \frac{\epsilon}{2}$. Therefore,

$$||g-p||_{L^2}^2 = \frac{1}{2\pi}\int_0^{2\pi} |g(x)-p(x)|^2 dx \leq ||g-p||_{L^\infty}^2.$$

Hence $||g-p||_{L^2} < \frac{\epsilon}{2}$ as well. We have established both approximations used in (25) and hence the conclusion of the theorem. □

COROLLARY 2.4 (Parseval's formula). *If f is integrable on the circle, then*

$$\sum_{-\infty}^{\infty} |\hat{f}(n)|^2 = ||f||_{L^2}^2. \tag{29}$$

PROOF. By the orthonormality properties of the functions $x \to e^{inx}$, $f - S_N(f)$ is orthogonal to V_N. By the Pythagorean theorem, we have

$$||f||_{L^2}^2 = ||f-S_N(f)||_{L^2}^2 + ||S_N(f)||_{L^2}^2 = ||f-S_N(f)||_{L^2}^2 + \sum_{-N}^{N} |\hat{f}(n)|^2. \tag{30}$$

Letting N tend to infinity in (30) and using Theorem 2.7 give (29). □

COROLLARY 2.5 (Riemann–Lebesgue lemma). *If f is integrable on the circle, then* $\lim_{|n|\to\infty} \hat{f}(n) = 0$.

PROOF. The series in (29) converges; hence its terms tend to 0. □

Polarization has several applications to Fourier series. By (29), if f and g are integrable on the circle S^1, then $\sum |\hat{f}|^2 = ||f||_{L^2}^2$ and similarly for g. It follows by polarization that

$$\langle \hat{f}, \hat{g} \rangle_2 = \sum_{-\infty}^{\infty} \hat{f}(n)\overline{\hat{g}(n)} = \frac{1}{2\pi}\int_0^{2\pi} f(x)\overline{g(x)}dx = \langle f, g\rangle_{L^2}. \tag{31}$$

COROLLARY 2.6. *If f and g are integrable on the circle, then (31) holds.*

COROLLARY 2.7. *The map $f \to \mathcal{F}(f)$ from $L^2(S^1)$ to l^2 satisfies the relation*

$$\langle \mathcal{F}f, \mathcal{F}g\rangle_2 = \langle f, g\rangle_{L^2}.$$

The analogue of this corollary holds for Fourier transforms on $\mathbf{R}$, $\mathbf{R}^n$, or in even more abstract settings. Such results, called Plancherel theorems, play a crucial role in extending the definition of Fourier transform to objects (called distributions) more general than functions. See Chap. 3.

THEOREM 2.8. *Suppose f is continuously differentiable on the circle. Then its Fourier series converges absolutely to f.*

PROOF. By Lemma 1.8, we have $\hat{f}(n) = \frac{\hat{f'}(n)}{in}$ for $n \neq 0$. We first apply the Parseval identity to the Fourier series for f', getting

$$\frac{1}{2\pi}\int |f'(x)|^2 dx = \sum |\hat{f'}(n)|^2 = \sum n^2 |\hat{f}(n)|^2. \tag{32}$$

Then we use the Cauchy–Schwarz inequality on $\sum |\hat{f}(n)|$ to get

$$\sum |\hat{f}(n)| = |\hat{f}(0)| + \sum \frac{1}{n}\, n|\hat{f}(n)| \le |\hat{f}(0)| + (\sum \frac{1}{n^2})^{\frac{1}{2}}\ (\sum n^2 |\hat{f}(n)|^2)^{\frac{1}{2}}. \tag{33}$$

By (32), the second sum on the right-hand side of (33) converges. The sum $\sum_{n\neq 0} \frac{1}{n^2}$ also converges and can be determined exactly using Fourier series. See Exercise 2.31.

Since each partial sum is continuous and the partial sums converge uniformly, the limit is continuous. By Corollary 1.10, the Fourier series converges absolutely to f. □

EXERCISE 2.31. Compute the Fourier series for the function f defined by $f(x) = (\pi - x)^2$ on $(0, 2\pi)$. Use this series to show that $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$.

EXERCISE 2.32. Find $\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2}$. Suggestion: Find the Fourier series for x^2 on $(-\pi, \pi)$.

9. Bernstein's Theorem

We continue by proving a fairly difficult result. We include it to illustrate circumstances more general than Theorem 2.8 in which Fourier series converge absolutely and uniformly.

Definition 2.13. Let $f : S^1 \to \mathbf{C}$ be a function and suppose $\alpha > 0$. We say that f satisfies a Hölder condition of order α if there is a constant C such that

$$|f(x) - f(y)| \leq C|x-y|^{\alpha} \tag{34}$$

for all x, y. Sometimes we say f is Hölder continuous of order α.

By the mean value theorem from calculus, a differentiable function satisfies the inequality

$$|f(x) - f(y)| \leq \sup|f'(t)|\ |x-y|.$$

Hence, if f' is bounded, f satisfies a Hölder condition with $\alpha = 1$. Note also that a function satisfying (34) must be uniformly continuous.

Theorem 2.9. *Suppose f is Hölder continuous on the circle of order α and $\alpha > \frac{1}{2}$. Then the Fourier series for f converges absolutely and uniformly.*

Proof. The Hölder condition means that there is a constant C such that inequality (34) holds. We must somehow use this condition to study

$$\sum_{n \in \mathbf{Z}} |\hat{f}(n)|.$$

The remarkable idea here is to break up this sum into dyadic parts and estimate differently in different parts. For p a natural number, let R_p denote the set of $n \in \mathbf{Z}$ for which $2^{p-1} \leq |n| < 2^p$. Note that there are 2^p integers in R_p. We have

$$\sum_{n \in \mathbf{Z}} |\hat{f}(n)| = |\hat{f}(0)| + \sum_p \sum_{n \in R_p} |\hat{f}(n)|. \tag{35}$$

In each R_p, we can use the Cauchy–Schwarz inequality to write

$$\sum_{n \in R_p} |\hat{f}(n)| \leq \left(\sum_{n \in R_p} |\hat{f}(n)|^2 \right)^{\frac{1}{2}} (2^p)^{\frac{1}{2}}. \tag{36}$$

At first glance the factor $2^{\frac{p}{2}}$ looks troublesome, but we will nonetheless verify convergence of the Fourier series.

Let g_h be defined by $g_h(x) = f(x+h) - f(x-h)$. The Hölder condition gives

$$|g_h(x)|^2 \leq C^2 |2h|^{2\alpha} = C'|h|^{2\alpha},$$

and integrating we obtain

$$||g_h||_{L^2}^2 \leq C'|h|^{2\alpha}.$$

By the Parseval–Plancherel theorem (Corollary 2.7), for any h, we have

$$\sum_{n \in \mathbf{Z}} |\hat{g_h}(n)|^2 = ||g_h||_{L^2}^2 \leq C'|h|^{2\alpha}. \tag{37}$$

Now we compute the Fourier coefficients of g_h, relating them to f. Using the definition directly, we get

$$\hat{g_h}(n) = \frac{1}{2\pi}\int_0^{2\pi} \left(f(x+h) - f(x-h)\right) e^{-inx} dx.$$

Changing variables in each term and recollecting give

$$\hat{g_h}(n) = \frac{1}{2\pi}\int_0^{2\pi} f(y)e^{-iny}e^{inh}dy - \frac{1}{2\pi}\int_0^{2\pi} f(y)e^{-iny}e^{-inh}dy = 2i\sin(nh)\hat{f}(n).$$

Hence, we have

$$|\hat{g_h}(n)|^2 = 4\sin^2(nh)|\hat{f}(n)|^2.$$

Putting things together, we obtain, with a new constant c,

$$\sin^2(nh)|\hat{f}(n)|^2 = \frac{1}{4}|\hat{g_h}(n)|^2 \leq \frac{1}{4}\sum_n |\hat{g_h}(n)|^2 \leq c|h|^{2\alpha}. \tag{38}$$

Also we have

$$\sum_{n\in R_p} |\hat{f}(n)|^2 = \sum_{n\in R_p} |\hat{f}(n)|^2 \sin^2(nh)\frac{1}{\sin^2(nh)} = \sum_{n\in R_p} |\hat{g_h}(n)|^2 \frac{1}{4\sin^2(nh)}. \tag{39}$$

Put $h = \frac{\pi}{2^{p+1}}$. Then $\frac{\pi}{4} \leq |n|h \leq \frac{\pi}{2}$ and hence $\frac{1}{2} \leq \sin^2(nh) \leq 1$. Using $\sin^2(nh) \geq \frac{1}{2}$ in (39), we get

$$\sum_{n\in R_p} |\hat{f}(n)|^2 \leq \frac{1}{2}\sum_{n\in R_p} |\hat{g_h}(n)|^2. \tag{40}$$

For $h = \frac{\pi}{2^{p+1}}$, we have

$$|\hat{g_h}(n)|^2 \leq C_1 |\frac{\pi}{2^{p+1}}|^{2\alpha} \leq C_2 2^{-2\alpha p}. \tag{41}$$

Combining (40), (41), and (36) (note the exponent $\frac{1}{2}$ there) gives

$$\sum_{n\in\mathbf{Z}} |\hat{f}(n)| = |\hat{f}(0)| + \sum_p \sum_{n\in R_p} |\hat{f}(n)| \leq |\hat{f}(0)| + C_2 \sum_p 2^{-\alpha p} 2^{\frac{p}{2}}. \tag{42}$$

The series on the right-hand side of (42) is of the form $\sum x^p$, where $x = 2^{\frac{1-2\alpha}{2}}$. If $\alpha > \frac{1}{2}$, then $|x| < 1$, and this series converges. $\square$

The conclusion of the theorem fails if f satisfies a Hölder condition of order $\frac{1}{2}$. See [K].

10. Compact Hermitian Operators

Fourier series give but one of many examples of orthonormal expansions. In this section, we establish the spectral theorem for compact Hermitian operators. Such operators determine complete orthonormal systems consisting of eigenvectors. In the next section, we apply this result to Sturm–Liouville equations. These second-order ordinary differential equations with homogeneous boundary conditions played a major role in the historical development of operator theory and remain significant in current applied mathematics, engineering, and physics.

An operator on a Hilbert space is *compact* if it can be approximated (in norm) arbitrarily well by operators with finite-dimensional range. We mention this characterization for the intuition it provides. The precise definition, which also applies in the context of complete normed vector spaces, involves subsequences. In older literature, compact operators are called *completely continuous*.

DEFINITION 2.14. Suppose $L \in \mathcal{L}(\mathcal{H})$. Then L is *compact* if whenever $\{z_n\}$ is a bounded sequence in $\mathcal{H}$, then $\{L(z_n)\}$ has a convergent subsequence.

By the Bolzano–Weierstrass theorem (see Theorem 5.2), each bounded sequence in $\mathbf{C}^d$ has a convergent subsequence. Hence an operator with finite-dimensional range must be compact. A constant multiple of a compact operator is compact. The sum of two compact operators is compact. We check in Proposition 2.10 that the composition (on either side) of a compact operator with a bounded operator is compact. On the other hand, the identity operator is compact only when the Hilbert space is finite dimensional. Proposition 2.13 gives one of many possible proofs of this last statement.

We will use the following simple characterization of compact operators. See [D1] for many uses of the method. The two statements in the proof are equivalent, with different values of ϵ. In the statement, we write f for an element of $\mathcal{H}$ to remind us that we are typically working on function spaces.

PROPOSITION 2.9. *Suppose $L \in \mathcal{L}(\mathcal{H})$. Then L is compact if (and only if), for each $\epsilon > 0$, there are compact operators K_ϵ and T_ϵ such that either of the following (equivalent) statements holds:*

$$||Lf|| \leq \epsilon||f|| + ||K_\epsilon f||$$
$$||Lf||^2 \leq \epsilon||f||^2 + ||T_\epsilon f||^2.$$

PROOF. Assuming the first inequality, we prove that L is compact. The proof assuming the second inequality is similar. Let $\{f_n\}$ be a bounded sequence; we may assume that $||f_n|| \leq 1$. For each positive integer m, we set $\epsilon = \frac{1}{m}$ in the inequality. We obtain a sequence $\{L_m\}$ of compact operators. Thus each sequence $\{L_m(f_n)\}$ has a convergent subsequence. By the Cantor diagonalization trick, there is a single subsequence $\{f_{n_k}\}$ such that $\{L_m(f_{n_k})\}$ converges for all m. By the inequality, for each m, we have

$$||L(f_{n_k}) - L(f_{n_l})|| = ||L(f_{n_k} - f_{n_l})|| \leq (\frac{1}{m})||f_{n_k} - f_{n_l}|| + ||L_m(f_{n_k} - f_{n_l})||.$$

Given $\delta > 0$, we can bound the first term by $\frac{\delta}{2}$ by choosing $\frac{1}{m} < \frac{\delta}{4}$. Since $\{L_m(f_{n_k})\}$ converges, it is Cauchy; we can therefore bound the second term by $\frac{\delta}{2}$ by picking n_k and n_l sufficiently large. Therefore, the sequence $\{L(f_{n_k})\}$ is also Cauchy in $\mathcal{H}$. Since $\mathcal{H}$ is complete, $\{L(f_{n_k})\}$ converges, and thus L is compact. □

If we know that L is compact, then we may choose K_ϵ or T_ϵ equal to L. The point of Proposition 2.9 is the converse statement. We can often prove compactness by proving an inequality, instead of dealing with subsequences. We illustrate with several examples, which can of course also be proved using subsequences.

PROPOSITION 2.10. *Suppose $L \in \mathcal{L}(\mathcal{H})$ and L is compact. If $M, T \in \mathcal{L}(\mathcal{H})$, then ML and LT are compact.*

PROOF. That LT is compact follows directly from the definition of compactness. If $\{z_n\}$ is a bounded sequence, then $\{Tz_n\}$ also is, and hence $\{L(Tz_n)\}$ has a convergent subsequence. Similarly, ML is compact.

That ML is compact can also be proved using Proposition 2.9 as follows. Given $\epsilon > 0$, put $\epsilon' = \frac{\epsilon}{1+||M||}$. Put $K = ||M||\ L$; then K is compact. We have

$$||MLz|| \leq ||M||\ ||Lz|| \leq ||M||(\epsilon'||z|| + ||Lz||) \leq \epsilon||z|| + ||Kz||.$$

By Proposition 2.9, ML is also compact. □

PROPOSITION 2.11. *Let $\{L_n\}$ be a sequence of operators with* $\lim_n ||L_n - L|| = 0$. *If each L_n is compact, then L is also compact.*

PROOF. Given $\epsilon > 0$, we can find an n such that $||L - L_n|| < \epsilon$. Then we write

$$||Lf|| \leq ||(L - L_n)f|| + ||L_n(f)|| \leq \epsilon||f|| + ||L_n(f)||.$$

The result therefore follows from Proposition 2.9. □

A converse of Proposition 2.11 also holds; each compact operator is the limit in norm of a sequence of operators with finite-dimensional ranges. We can also use Proposition 2.9 to prove the following result.

THEOREM 2.10. *Assume $L \in \mathcal{L}(\mathcal{H})$. If L is compact, then L^* is compact. Furthermore, L is compact if and only if L^*L is compact.*

PROOF. See Exercise 2.35. □

EXERCISE 2.33 (Small constant large constant trick). Given $\epsilon > 0$, prove that there is a $C_\epsilon > 0$ such that

$$|\langle x, y\rangle| \leq \epsilon||x||^2 + C_\epsilon||y||^2.$$

EXERCISE 2.34. Prove that the second inequality in Proposition 2.9 implies compactness.

EXERCISE 2.35. Prove Theorem 2.10. Use Proposition 2.9 and Exercise 2.23 to verify the *if* part of the implication.

Before turning to the spectral theorem for compact Hermitian operators, we give one of the classical types of examples. The function K in this example is called the *integral kernel* of the operator T. Such integral operators arise in the solutions of differential equations such as the Sturm–Liouville equation.

PROPOSITION 2.12. *Let $\mathcal{H} = L^2([a,b])$. Assume that $(x,t) \to K(x,t)$ is continuous on $[a,b] \times [a,b]$. Define an operator T on $\mathcal{H}$ by*

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

Then T is compact. (The conclusion holds under weaker assumptions on K.)

PROOF. Let $\{f_n\}$ be a bounded sequence in $L^2([a,b])$. The following estimate follows from the Cauchy–Schwarz inequality:

$$|T(f_n)(x) - T(f_n)(y)|^2 \leq \sup|K(x,t) - K(y,t)|^2||f_n||^2_{L^2}.$$

Since K is continuous on the compact set $[a, b] \times [a, b]$, it is uniformly continuous. It follows that the sequence $\{T(f_n)\}$ is equi-continuous and uniformly bounded. By the Arzela–Ascoli theorem, there is a subsequence of $\{T(f_n)\}$ that converges uniformly. In particular, this subsequence converges in L^2. Hence, $\{T(f_n)\}$ has a convergent subsequence, and thus T is compact. $\square$

EXERCISE 2.36. Suppose that the integral kernel in Proposition 2.12 satisfies $\int_a^b |K(x,t)|dt \le C$ and $\int_a^b |K(x,t)|dx \le C$. Show that $T \in \mathcal{L}(\mathcal{H})$ and that $||T|| \le C$.

A compact operator need not have any eigenvalues or eigenvectors.

EXAMPLE 2.3. Let $L : l^2 \to l^2$ be defined by

$$L(z_1, z_2, \dots) = (0, z_1, \frac{z_2}{2}, \frac{z_3}{3}, \dots).$$

Think of L as given by an infinite matrix with sub-diagonal entries $1, \frac{1}{2}, \frac{1}{3}, \dots$. Then L is compact but has no eigenvalues.

EXERCISE 2.37. Verify the conclusions of Example 2.3.

Compact Hermitian operators, however, have many eigenvectors. In fact, by the spectral theorem, there is a complete orthonormal system of eigenvectors. Before proving the spectral theorem, we note two easy results about eigenvectors and eigenvalues.

PROPOSITION 2.13. *An eigenspace of a compact operator corresponding to a nonzero eigenvalue must be finite dimensional.*

PROOF. Assume that L is compact and $L(z_j) = \lambda z_j$ for a sequence of orthogonal unit vectors z_j. Since L is compact, $L(z_j) = \lambda z_j$ has a convergent subsequence. If $\lambda \ne 0$, then z_j has a convergent subsequence. But no sequence of orthogonal unit vectors can converge. Thus $\lambda = 0$. $\square$

PROPOSITION 2.14. *The eigenvalues of a Hermitian operator are real, and the eigenvectors corresponding to distinct eigenvalues are orthogonal.*

PROOF. Assume $Lf = \lambda f$ and $f \ne 0$. We then have

$$\lambda ||f||^2 = \langle Lf, f \rangle = \langle f, L^* f \rangle = \langle f, Lf \rangle = \langle f, \lambda f \rangle = \overline{\lambda} ||f||^2.$$

Since $||f||^2 \ne 0$, we conclude that $\lambda = \overline{\lambda}$.

The proof of the second statement amounts to polarizing the first. Thus we suppose $Lf = \lambda f$ and $Lg = \mu g$ where $\lambda \ne \mu$. We have, as μ is real,

$$\lambda \langle f, g \rangle = \langle Lf, g \rangle = \langle f, Lg \rangle = \mu \langle f, g \rangle.$$

Hence $0 = (\lambda - \mu)\langle f, g \rangle$ and the second conclusion follows. $\square$

PROPOSITION 2.15. *Suppose $L \in \mathcal{L}(\mathcal{H})$ is Hermitian. Then*

$$||L|| = \sup_{||z||=1} |\langle Lz, z \rangle|. \tag{43}$$

PROOF. Let α equal the right-hand side of (43). We prove both inequalities: $\alpha \leq ||L||$ and $||L|| \leq \alpha$. Since $|\langle Lz, z\rangle| \leq ||Lz||\,||z||$, we see that

$$\alpha = \sup_{||z||=1} |\langle Lz, z\rangle| \leq \sup_{||z||=1} ||Lz|| = ||L||.$$

The opposite inequality is harder. It uses the polarization identity (21) and the parallelogram law (9). We first note, by Theorem 2.6, that $\alpha = 0$ implies $L = 0$. Hence we may assume $\alpha \neq 0$. Since L is Hermitian, it follows that

$$\langle Lz, w\rangle = \langle z, Lw\rangle = \overline{\langle Lw, z\rangle}.$$

Applying this equality in (21), we obtain, for all z, w,

$$4\mathrm{Re}\langle Lz, w\rangle = \langle L(z+w), z+w\rangle - \langle L(z-w), z-w\rangle.$$

Using $\langle L\zeta, \zeta\rangle \leq \alpha||\zeta||^2$ and the parallelogram law, we obtain

$$4\mathrm{Re}\langle Lz, w\rangle \leq \alpha(||z+w||^2 + ||z-w||^2) = 2\alpha(||z||^2 + ||w||^2). \tag{44}$$

Set $w = \frac{Lz}{\alpha}$ in (44) to get

$$\frac{4||Lz||^2}{\alpha} \leq 2\alpha(||z||^2 + \frac{||Lz||^2}{\alpha^2}).$$

Simplifying shows that this inequality is equivalent to $2\frac{||Lz||^2}{\alpha} \leq 2\alpha||z||^2$, which implies $||Lz||^2 \leq \alpha^2||z||^2$. Hence $||L|| \leq \alpha$. ☐

THEOREM 2.11 (Spectral theorem). *Suppose $L \in \mathcal{L}(\mathcal{H})$ is compact and Hermitian. Then there is a complete orthonormal system consisting of eigenvectors of L. Each eigenspace corresponding to a nonzero eigenvalue is finite dimensional.*

PROOF. The conclusion holds if L is the zero operator; we therefore ignore this case and assume $||L|| > 0$.

The first fact needed is that there is an eigenvalue λ with $|\lambda| = ||L||$. Note also, since L is Hermitian, that in this case λ is real and thus $\lambda = \pm||L||$. In the proof, we write α for $\pm||L||$; in general only one of the two values works.

Because L is Hermitian, the subtle formula (43) for the norm of L holds. We let $\{z_\nu\}$ be a sequence on the unit sphere such that $|\langle Lz_\nu, z_\nu\rangle|$ converges to $||L||$. Since L is compact, we can find a subsequence (still labeled $\{z_\nu\}$) such that $L(z_\nu)$ converges to some w.

We will show that $||w|| = ||L||$ and also that αz_ν converges to w. It follows that z_ν converges to $z = \frac{w}{\alpha}$. Then we have a unit vector z for which $Lz = w = \alpha z$, and hence the first required fact will hold.

To see that $||w|| = ||L||$, we prove both inequalities. Since the norm is continuous and $||z_\nu|| = 1$, we obtain

$$||w|| = \lim_\nu ||Lz_\nu|| \leq ||L||.$$

To see the other inequality, note that $|\langle Lz_\nu, z_\nu\rangle|$ is converging to $||L||$ and $L(z_\nu)$ is converging to w. Hence $|\langle w, z_\nu\rangle|$ is converging to $||L||$ as well. We then have

$$||L|| = \lim_\nu |\langle w, z_\nu\rangle| \leq ||w||.$$

Thus $||w|| = ||L||$.

Next we show that αz_ν converges to w. Consider the squared norm

$$||L(z_\nu) - \alpha z_\nu||^2 = ||L(z_\nu)||^2 - \alpha 2\mathrm{Re}\langle L z_\nu, z_\nu\rangle + ||L||^2.$$

The right-hand side converges to $||w||^2 - 2||L||^2 + ||L||^2 = 0$. Therefore, the left-hand side converges to 0 as well, and hence $w = \lim(\alpha z_\nu)$. Thus z_ν itself converges to $z = \frac{w}{\alpha}$. Finally

$$L(z) = \lim(L(z_\nu)) = w = \alpha z.$$

We have found an eigenvector z with eigenvalue $\alpha = \pm||L||$. By Proposition 2.13, the eigenspace E_α corresponding to α is finite dimensional and thus a closed subspace of $\mathcal{H}$.

Once we have found one eigenvalue λ_1, we consider the orthogonal complement $E_{\lambda_1}^\perp$ of the eigenspace E_{λ_1}. Then $E_{\lambda_1}^\perp$ is invariant under L, and the restriction of L to this subspace remains compact and Hermitian. We repeat the procedure, obtaining an eigenvalue λ_2. The eigenspaces E_{λ_1} and E_{λ_2} are orthogonal. Continuing in this fashion, we obtain a nonincreasing sequence of (absolute values of) eigenvalues and corresponding eigenvectors. Each eigenspace is finite dimensional, and the eigenspaces are orthogonal. We normalize the eigenvectors to have norm 1; hence there is a bounded sequence $\{z_j\}$ of eigenvectors. By compactness, $\{L(z_j)\}$ has a convergent subsequence. Since $L(z_j) = \lambda_j z_j$, also $\{\lambda_j z_j\}$ has a convergent subsequence. A sequence of orthonormal vectors cannot converge; the subsequence cannot be eventually constant because each eigenspace is of finite dimension. The only possibilities are that there are only finitely many nonzero eigenvalues, or that the eigenvalues λ_j tend to 0.

Finally we establish completeness. Let M denote a maximal collection of orthonormal eigenvectors, including those with eigenvalue 0. Since we are assuming $\mathcal{H}$ is separable, we may assume the eigenvectors are indexed by the positive integers. Let P_n denote the projection onto the span of the first n eigenvectors. We obtain

$$P_n(\zeta) = \sum_{j=1}^{n} \langle \zeta, z_j\rangle z_j.$$

Therefore

$$||L(P_n(\zeta)) - L(\zeta)|| \leq \max_{(j\geq n+1)}|\lambda_j|\ ||\zeta||. \tag{45}$$

Since the eigenvalues tend to zero, (45) shows that $L(P_n(\zeta))$ converges to $L(\zeta)$. Hence we obtain the orthonormal expansion for w in the range $\mathcal{R}(L)$ of L:

$$w = L(\zeta) = \sum_{j=1}^{\infty} \langle \zeta, z_j\rangle \lambda_j z_j. \tag{46}$$

The nullspace $\mathcal{N}(L)$ is the eigenspace corresponding to eigenvalue 0, and hence any element of $\mathcal{N}(L)$ has an expansion in terms of vectors in M. Finally, for any bounded linear map L, Proposition 2.7 guarantees that $\mathcal{N}(L) \oplus \mathcal{R}(L^*) = \mathcal{H}$. If also $L = L^*$, then $\mathcal{N}(L) \oplus \mathcal{R}(L) = \mathcal{H}$. Therefore, 0 is the only vector orthogonal to M, and M is complete. □

EXERCISE 2.38. Try to give a different proof of (43). (In finite dimensions, one can use Lagrange multipliers.)

EXERCISE 2.39. Show that L^*L is compact and Hermitian if L is compact.

REMARK 2.1. The next several exercises concern *commutators* of operators.

DEFINITION 2.15. Let A, B be bounded operators. Their commutator $[A, B]$ is defined by $AB - BA$.

EXERCISE 2.40. Let A, B, C be bounded operators, and assume that $[C, A]$ and $[C, B]$ are compact. Prove that $[C, AB]$ is also compact. Suggestion: Do some easy algebra and then use Proposition 2.10.

EXERCISE 2.41. For a positive integer n, express $[A, B^n]$ as a sum of n terms involving $[A, B]$. What is the result when $[A, B] = I$?

EXERCISE 2.42. Use the previous exercise to show that there are no bounded operators satisfying $[A, B] = I$. Suggestion: Compute the norm of $[A, B^n]$ in two ways and let n tend to infinity.

EXERCISE 2.43. Suppose that $\langle Lz, z\rangle \geq 0$ for all z and that $||L|| \leq 1$. Show that $||I - L|| \leq 1$.

EXERCISE 2.44. Assume $L \in \mathcal{L}(\mathcal{H})$. Show that L is a linear combination of two Hermitian operators.

EXERCISE 2.45. Fill in the following outline to show that a Hermitian operator A is a linear combination of two unitary operators. Without loss of generality, we may assume $||A|| \leq 1$. If $-1 \leq a \leq 1$, put $b = \sqrt{1 - a^2}$. Then $a = \frac{1}{2}((a + ib) + (a - ib))$ is the average of two points on the unit circle. We can analogously write the operator A as the average of unitary operators $A + iB$ and $A - iB$, if we can find a square root of $I - A^2$. Put $L = I - A^2$. We can find a square root of L as follows. We consider the power series expansion for $\sqrt{1 - z}$ and replace z by A^2. In other words, $\sqrt{I - C}$ makes sense if $||C|| \leq 1$. You will need to know the sign of the coefficients in the expansion to verify convergence. Hence $\sqrt{L} = \sqrt{I - (I - L)}$ makes sense.

We close this section with a few words about *unbounded operators*. This term refers to linear mappings, defined on dense subsets of a Hilbert space, but not continuous.

Suppose $\mathcal{D}$ is a dense subset of a Hilbert space $\mathcal{H}$ and L is defined and linear on $\mathcal{D}$. If L were continuous, then L would extend to a linear mapping on $\mathcal{H}$. Many important operators are not continuous. Differentiation $\frac{d}{dx}$ is defined and linear on a dense set in $L^2([0, 2\pi])$, but it is certainly not continuous. For example, $\{\frac{e^{inx}}{in}\}$ converges to 0 in L^2, but $\frac{d}{dx}(\frac{e^{inx}}{in}) = e^{inx}$, whose L^2 norm equals 1 for each n. To apply Hilbert space methods to differential operators, we must be careful.

Let $L : \mathcal{D}(L) \subseteq \mathcal{H} \to \mathcal{H}$ be an unbounded operator. The domain $\mathcal{D}(L^*)$ of the adjoint of L is the set of $v \in \mathcal{H}$ such that the mapping $u \to \langle Lu, v\rangle$ is a continuous linear functional. By the Riesz lemma, there is then a unique w such that $\langle Lu, v\rangle = \langle u, w\rangle$. We then put $L^*(v) = w$. It can happen that the domain of L^* is not dense in $\mathcal{H}$.

We say that an unbounded (but densely defined) operator L is *Hermitian* if

$$\langle Lz, w\rangle = \langle z, Lw\rangle$$

for all z and w in the domain of L. We say that L is *self-adjoint* if $\mathcal{D}(L) = \mathcal{D}(L^*)$ and the two maps agree there. Thus, L is Hermitian if $Lz = L^*z$ when both are defined and self-adjoint if also $\mathcal{D}(L) = \mathcal{D}(L^*)$. It often happens, with a given definition of $\mathcal{D}(L)$, that L^* agrees with L on $\mathcal{D}(L)$, but L is not self-adjoint. One must increase the domain of L, thereby decreasing the domain of L^*, until these domains are equal, before one can use without qualification the term self-adjoint.

EXERCISE 2.46 (Subtle). Put $L = i\frac{d}{dx}$ on the subspace of differentiable functions f in $L^2([0,1])$ for which $f(0) = f(1) = 0$. Show that $\langle Lf, g\rangle = \langle f, Lg\rangle$, but that L is not self-adjoint. Can you state precisely a domain for L making it self-adjoint? Comment: Look up the term *absolutely continuous* and weaken the boundary condition.

11. Sturm–Liouville Theory

Fourier series provide the most famous example of orthonormal expansion, but many other orthonormal systems arise in applied mathematics and engineering. We illustrate by considering certain differential equations known as Sturm–Liouville equations. Mathematicians from the nineteenth century were well aware that many properties of the functions sine and cosine have analogues when these functions are replaced by linearly independent solutions of a second-order linear ordinary differential equation. In addition to orthonormal expansions, certain oscillation issues generalize as well. We prove the Sturm separation theorem, an easy result, to illustrate this sort of generalization, before we turn to the more difficult matter of orthonormal expansion.

Consider a second-order linear ordinary differential equation $y'' + qy' + ry = 0$. Here q and r are continuous functions of x. What can we say about the zeroes of solutions? Figure 2.4 illustrates the situation for cosine and sine. Theorem 2.12 provides a general result.

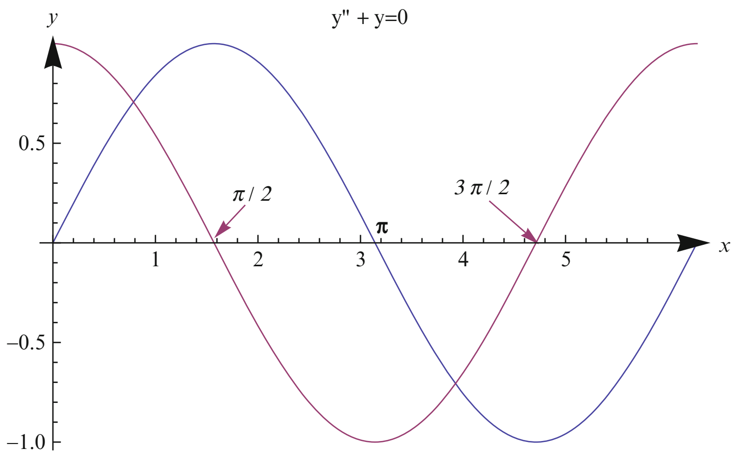

FIGURE 2.4. Sturm separation

THEOREM 2.12 (Sturm separation theorem). *Let y_1 and y_2 be linearly independent (twice differentiable) solutions of $y'' + qy' + ry = 0$. Suppose that $\alpha < \beta$ and α, β are consecutive zeroes of y_1. Then there is a unique x in the interval (α, β) with $y_2(x) = 0$. Hence the zeroes of y_1 and y_2 alternate.*

PROOF. Consider the expression $W(x) = y_1(x)y_2'(x) - y_2(x)y_1'(x)$, called the Wronskian. We claim that it does not vanish. Assuming the claim, W has only one sign. We evaluate W at α and β, obtaining $-y_2(\alpha)y_1'(\alpha)$ and $-y_2(\beta)y_1'(\beta)$; these expressions must have the same sign. In particular, y_1' does not vanish at these points. Also, the values $y_1'(\alpha)$ and $y_1'(\beta)$ must have opposite signs because α and β are consecutive zeroes of y_1. Hence the values of $y_2(\alpha)$ and $y_2(\beta)$ have opposite signs. By the intermediate value theorem, there is an x in between α and β with $y_2(x) = 0$. This x must be unique, because otherwise the same reasoning would find a zero of y_1 in between the two zeroes of y_2. Since α and β are consecutive zeroes of y_1, we would get a contradiction.

It remains to show that W is of one sign. We show more in Lemma 2.2. □

LEMMA 2.2. *Suppose y_1 and y_2 both solve $L(y) = y'' + qy' + ry = 0$. Then y_1 and y_2 are linearly dependent if and only if W vanishes identically. Also y_1 and y_2 are linearly independent if and only if W vanishes nowhere.*

PROOF. Suppose first that $W(x_0) = 0$. Since $W(x_0)$ is the determinant of the matrix of coefficients, the system of equations

$$\begin{pmatrix} y_1(x_0) & y_2(x_0) \\ y_1'(x_0) & y_2'(x_0) \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

has a nontrivial solution (c_1, c_2). Since L is linear, the function $y = c_1y_1 + c_2y_2$ also satisfies $L(y) = 0$. Since $y(x_0) = y'(x_0) = 0$, this solution y is identically 0. (See the paragraph after the proof.) Therefore, the matrix equation holds at all x, the functions y_1 and y_2 are linearly dependent, and W is identically 0.

Suppose next that W is never zero. Consider a linear combination $c_1y_1 + c_2y_2$ that vanishes identically. Then also $c_1y_1' + c_2y_2'$ vanishes identically, and hence

$$\begin{pmatrix} y_1 & y_2 \\ y_1' & y_2' \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Since W is the determinant of the matrix here and $W(x) \neq 0$ for all x, the only solution is $c_1 = c_2 = 0$. Therefore, y_1 and y_2 are linearly independent. □

In the proof of Lemma 2.2, we used the following standard fact. The second-order linear equation $Ly = 0$, together with initial conditions $y(x_0)$ and $y'(x_0)$, has a unique solution. This result can be proved by reducing the second-order equation to a first-order system. Uniqueness for the first-order system can be proved using the contraction mapping principle in metric spaces. See [Ro].

We now turn to the more sophisticated Sturm–Liouville theory. Consider the following second-order differential equation on a real interval $[a, b]$. Here y is the unknown function; p, q, w are fixed real-valued functions, and the α_j and β_j are real constants. These constants are subject only to the constraint that both (*SL.1*) and (*SL.2*) are nontrivial. In other words, neither $\alpha_1^2 + \alpha_2^2$ nor $\beta_1^2 + \beta_2^2$ is 0. This condition makes the equation into a *boundary value problem*. Both endpoints of the interval $[a, b]$ matter. The functions p', q, w are assumed to be continuous, and the functions p and w are assumed positive:

$$(py')' + qy + \lambda wy = 0 \tag{SL}$$

$$\alpha_1 y(a) + \alpha_2 y'(a) = 0 \tag{SL.1}$$

$$\beta_1 y(b) + \beta_2 y'(b) = 0. \tag{SL.2}$$

REMARK 2.2. It is natural to ask how general the Sturm–Liouville equation is among second-order linear equations. Consider any second-order ODE of the form $Py'' + Qy' + Ry = 0$, where $P \neq 0$. We can always put it into the Sturm–Liouville form by the following typical trick from ODE, called an *integrating factor*. We multiply the equation by an unknown function u and figure out what u must be to put the equation in Sturm–Liouville form:

$$0 = uPy'' + uQy' + uRy = (py')' + ry.$$

To make this equation hold, we need $uP = p$ and $uQ = p'$. Hence we require $\frac{p'}{p} = \frac{Q}{P}$, which yields $p = e^{\int \frac{Q}{P}}$. Hence, if we choose $u = \frac{1}{P} e^{\int \frac{Q}{P}}$, we succeed in putting the equation in the form (*SL*).

The following lemma involving the Wronskian gets used in an important integration by parts below, and it also implies that each eigenspace is one dimensional. Note that the conclusion also holds if we replace g by $\overline{g}$, because all the parameters in (*SL*), (*SL.1*), and (*SL.2*) are real.

LEMMA 2.3. *If f and g both satisfy (SL.1) and (SL.2), then*

$$f(a)g'(a) - f'(a)g(a) = f(b)g'(b) - f'(b)g(b) = 0. \tag{47}$$

PROOF. Assume both f and g satisfy the conditions in (*SL*). We then can write

$$\begin{pmatrix} f(b) & f'(b) \\ g(b) & g'(b) \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{48}$$

and similarly for the values at a and the α_j. Equations (*SL.1*) and (*SL.2*) are nontrivial; hence (48) and its analogue for a have nontrivial solutions, and each of the matrices

$$\begin{pmatrix} f(a) & f'(a) \\ g(a) & g'(a) \end{pmatrix}$$

$$\begin{pmatrix} f(b) & f'(b) \\ g(b) & g'(b) \end{pmatrix}$$

has a nontrivial nullspace. Hence, each determinant vanishes. □

COROLLARY 2.8. *Suppose f and g both solve the same (SL) equation. Then f and g are linearly dependent.*

PROOF. By Lemma 2.3, the two expressions in (47) vanish. But these expressions are Wronskian determinants. By Lemma 2.2, the two solutions are linearly independent if and only if their Wronskian determinant is (everywhere) nonzero. □

Later we use one more fact about the Wronskian.

LEMMA 2.4. *Assume* u, v *both solve the Sturm–Liouville equation* $(py')' + qy = 0$. *Let* $W = uv' - u'v$. *Then* pW *is constant. If* u, v *are linearly independent, then this constant is nonzero.*

PROOF. We want to show that $(p(uv' - u'v))' = 0$. Computing the expression, without any assumptions on u, v, gives

$$p(uv'' - u''v) + p'(uv' - u'v).$$

Since u and v satisfy the equation, we also have

$$pu'' + p'u' + qu = 0$$

$$pv'' + p'v' + qv = 0.$$

Multiply the first equation by v, the second by u, and then subtract. We get

$$p(u''v - uv'') + p'(u'v - uv') = 0,$$

which is what we need. The last statement follows immediately from Lemma 2.2. □

Each λ for which (*SL*) admits a nonzero solution is called an eigenvalue of the problem, and each nonzero solution is called an eigenfunction corresponding to this eigenvalue. The terminology is consistent with the standard notions of eigenvalue and eigenvector, as noted in Lemma 2.5 below. In general, when the elements of a vector space are functions, one often says *eigenfunction* instead of *eigenvector*. Corollary 2.8 thus says that the eigenspace corresponding to each eigenvalue is one dimensional.

To connect the Sturm–Liouville setting with Fourier series, take $p = 1$, $q = 0$, and $w = 1$. We get the familiar equation

$$y'' + \lambda y = 0,$$

whose solutions are sines and cosines. For example, if the interval is $[0, \pi]$, and we assume that (*SL.1*) and (*SL.2*) give $y(0) = y(\pi) = 0$, then the eigenvalues are m^2 for positive integers m. The solutions are $y_m(x) = \sin(mx)$.

Sturm–Liouville theory uses the Hilbert space $\mathcal{H} = (L^2([a, b]), w)$, consisting of (equivalence classes of) square-integrable measurable functions with respect to the weight function w. The inner product is defined by

$$\langle f, g \rangle_w = \int_a^b f(x)\overline{g(x)}w(x)dx.$$

Although the Sturm–Liouville situation is much more general than the equation $y'' + \lambda y = 0$, the conclusions in the following theorem are remarkably similar to the results we have proved about Fourier series:

THEOREM 2.13. *Consider the Sturm–Liouville equation* (*SL*) *with boundary conditions* (*SL.1*) *and* (*SL.2*). *There is a countable collection of real eigenvalues* λ_j *tending to* ∞ *with* $\lambda_1 < \lambda_2 < \dots$. *For each eigenvalue, the corresponding eigenspace is one dimensional. The corresponding eigenfunctions* ϕ_j *are orthogonal. After dividing each* ϕ_j *by a constant, we assume that these eigenfunctions* *are orthonormal. These eigenfunctions form a complete orthonormal system for* $\mathcal{H}$*. If* f *is continuously differentiable on* $[a, b]$*, then the series*

$$\sum_{j=1}^{\infty} \langle f, \phi_j \rangle_w \phi_j(x) \tag{49}$$

converges to $f(x)$ *at each point of* (a, b)*.*

Proving this theorem is not easy, but we will give a fairly complete proof. We begin by rephrasing everything in terms of an unbounded operator L on $\mathcal{H}$. On an appropriate domain, L is defined by

$$L = \frac{-1}{w} \left(\frac{d}{dx} \left(p \frac{d}{dx} \right) + q \right). \tag{50}$$

The domain $\mathcal{D}(L)$ contains all twice continuously differentiable functions satisfying the (*SL*) boundary conditions. Eigenvalues of the Sturm–Liouville problem correspond to eigenvalues of this operator L.

Lemma 2.5. *Equation (SL) is equivalent to* $Ly = \lambda y$*.*

Proof. Left to the reader. □

Proposition 2.16. *The operator L is Hermitian. In other words, if* f *and* g *are twice continuously differentiable functions on* $[a, b]$ *and satisfy (SL.1) and (SL.2), then*

$$\langle Lf, g \rangle_w = \langle f, Lg \rangle_w. \tag{51}$$

Proof. The proof amounts to integrating by parts twice and using the boundary conditions. One integration by parts gives

$$\begin{aligned}
\langle Lf, g \rangle_w &= \int_a^b \frac{-1}{w(x)} \left(\frac{d}{dx} \left(p(x) f'(x) \right) + q(x) f(x) \right) \overline{g(x)} w(x) dx \\
&= - \int_a^b \left(\frac{d}{dx} (p(x) f'(x)) + q(x) f(x) \right) \overline{g(x)} dx \\
&= -p(x) f'(x) \overline{g(x)} \big|_a^b + \int_a^b (p(x) f'(x)) \overline{g'(x)} dx - \int_a^b q(x) f(x) \overline{g(x)} dx.
\end{aligned} \tag{52}$$

We integrate the middle term by parts, and stop writing the variable x, to obtain

$$\langle Lf, g \rangle_w = -p\ f' \overline{g} \big|_a^b + p\ f \overline{g'} \big|_a^b - \int_a^b f \frac{d}{dx} (p \overline{g}')\ dx - \int_a^b q f \overline{g}\ dx. \tag{53}$$

After multiplying and dividing by w, the integrals in (53) become

$$\int_a^b \left(\frac{-f}{w} \left(\frac{d}{dx} (p \overline{g}') + q \overline{g} \right) \right) w\ dx = \langle f, Lg \rangle_w. \tag{54}$$

The boundary terms in (53) become

$$p(x) \left(f(x) \overline{g'(x)} - f'(x) \overline{g(x)} \right) \big|_a^b. \tag{55}$$

Since both f and g satisfy the homogeneous boundary conditions, the term in (55) vanishes by Lemma 2.3 (using $\overline{g}$ instead of g). Hence $\langle Lf, g \rangle_w = \langle f, Lg \rangle_w$. □

In order to proceed with Sturm–Liouville theory, we must introduce some standard ideas in operator theory. These ideas are needed because differential operators such as L are defined on only a dense subspace of the Hilbert space and they cannot be extended continuously to the whole space.

Let $\mathcal{H}$ be a Hilbert space and let $L : \mathcal{D}(L) \subseteq \mathcal{H} \to \mathcal{H}$ be a densely defined linear operator. For each complex number z, consider the operator $L - zI$.

DEFINITION 2.16. The complex number z is said to be in the *spectrum* of L if $(L - zI)^{-1}$ does not exist as a bounded linear operator. Otherwise z is said to be in the *resolvent set* of L, and $(L - zI)^{-1}$ is called the *resolvent* of L at z.

Thus, when z is in the resolvent set, $(L - zI)^{-1}$ exists and is bounded. The equation $(L - zI)^{-1} f = \mu f$ is then equivalent to $f = (L - zI)(\mu f)$ and hence also to $Lf = (z + \frac{1}{\mu})f$. Thus, to find the eigenvalues of L, we can study the resolvent $(L - zI)^{-1}$. If L is Hermitian and we choose a real k in the resolvent set for L, then $(L - kI)^{-1}$ is Hermitian. For L as in the Sturm–Liouville setup, the resolvent is a *compact* operator. In general, an unbounded operator L on a Hilbert space has *compact resolvent* if there is a z for which $(L - zI)^{-1}$ is compact. A generalization of Theorem 2.13 holds when L is self-adjoint and has compact resolvent.

In order to prove Theorem 2.13, we need to know that the resolvent $(L - kI)^{-1}$ is compact. We will use Green's functions.

11.1. The Green's Function. In this subsection, we construct the Green's function G in a fashion often used in physics and engineering. It will follow that a complete orthonormal system exists in the Sturm–Liouville setting. Let L be the operator defined in (50).

First we find a solution u to $Lu = 0$ that satisfies the boundary condition at a. Then we find a solution v to $Lv = 0$ that satisfies the boundary condition at b. We put

$$c = p(x)W(x) = p(x)(u(x)v'(x) - u'(x)v(x)). \tag{56}$$

By Lemma 2.4, when u and v are linearly independent, c is a nonzero constant.

We then define the Green's function as follows. Put $G(x,t) = \frac{1}{c}u(t)v(x)$ for $t < x$ and $G(x,t) = \frac{1}{c}u(x)v(t)$ for $t > x$. Then G extends to be continuous when $x = t$. Thus $Lu = 0$ and $Lv = 0$. The following important theorem and its proof illustrate the importance of the Green's function:

THEOREM 2.14. *Consider the Sturm–Liouville equation (SL). Let L be the Hermitian operator defined by (50). Let u be a solution to $Lu = 0$ satisfying boundary condition (SL.1) and v a solution to $Lv = 0$ with boundary condition (SL.2). Assume u and v are linearly independent, and define c by (56). Given f continuous, define y by*

$$y(x) = \frac{1}{c}\int_a^x u(x)(vfw)(t)dt + \frac{1}{c}\int_x^b v(x)(ufw)(t)dt = \int_a^b G(x,t)f(t)\ dt. \tag{57}$$

Then y is twice differentiable and $Ly = f$.

PROOF. We start with (57) and the formula (58) for L:

$$Ly = \frac{-p}{w}y'' - \frac{p'}{w}y' - \frac{q}{w}y. \tag{58}$$

We apply L to (57) using the fundamental theorem of calculus and compute. The collection of terms obtained where we differentiate past the integral must vanish because u, v satisfy $Lu = Lv = 0$. The remaining terms arise because of the fundamental theorem of calculus. The first time we differentiate we get

$$\frac{1}{c}(uvp)(x) - \frac{1}{c}(uvp)(x) = 0.$$

The minus sign arises because the second integral goes from x to b, rather than from b to x.

The next time we differentiate we obtain the term

$$\frac{p}{c}(u_x v - u v_x) f w,$$

with all terms evaluated at x. The term in parentheses is minus the Wronskian. By Lemma 2.4, the entire expression simplifies to $-(fw)(x)$. When we multiply by $\frac{-1}{w}$, from formula (58) of L, this expression becomes $f(x)$. We conclude, as desired, that $Ly = f$. Since u, v are twice differentiable, p is continuously differentiable, and w, f are continuous, it follows that y is twice differentiable. $\square$

Things break down when we cannot find linearly independent u and v, and the Green's function need not exist. In that case, we must replace L by $L - kI$ for a suitable constant k. The following example illustrates several crucial points:

EXAMPLE 2.4. Consider the equation $Ly = y'' = 0$ with $y'(0) = y'(1) = 0$. The only solutions to $Lu = 0$ are constants and hence linearly dependent. If c satisfies (56), then $c = 0$. We cannot solve $Ly = f$ for general f. Suppose that $y'(0) = y'(1) = 0$ and that $y'' = f$. Integrating twice, we then must have

$$y(x) = y(0) + \int_0^x \int_0^t f(s) ds dt.$$

By the fundamental theorem of calculus, $y'(0) = 0$ and $y'(1) = \int_0^1 f(s)ds$. If $\int_0^1 f$ is not 0, then we cannot solve the equation $Ly = f$. In this case, 0 is an eigenvalue for L, and hence, L^{-1} does not exist. The condition $\int_0^1 f = 0$ means that the function f must be orthogonal to the constants.

To finish the proof of the Sturm–Liouville theorem, we need to show that there is a real k such that $(L - kI)^{-1}$ exists as a bounded operator. This statement holds for all k sufficiently negative, but we omit the proof. Assuming this point, we can find linearly independent u and v satisfying the equation, with u satisfying the boundary condition at a and v satisfying it at b. We construct the Green's function for $L - kI$ as above. We write $(L - kI)^{-1} f(x) = \int_a^b f(t)G(x,t)dt$. Since G is continuous on the rectangle $[a, b] \times [a, b]$, $(L - kI)^{-1}$ is compact, by Proposition 2.12. Theorem 2.11 then yields the desired conclusions.

We can express things in terms of orthonormal expansion. Let L be the operator defined in (50). Given f, we wish to solve the equation $Lg = f$. Let $\{\phi_j\}$ be the complete orthonormal system of eigenfunctions for $(L - kI)^{-1}$. This system exists because $(L - kI)^{-1}$ is compact and Hermitian. We expand g in an orthonormal series as in (49), obtaining

$$g(x) = \sum_{j=1}^{\infty} \int_a^b g(t)\overline{\phi_j(t)} w(t) dt \ \phi_j(x).$$

Differentiating term by term yields

$$(Lg)(x) = f(x) = \sum_{j=1}^{\infty} \left(\int_a^b g(t)\overline{\phi_j(t)}w(t)dt \right) \lambda_j \phi_j(x).$$

The function f also has an orthonormal expansion:

$$f(x) = \sum_{j=1}^{\infty} \left(\int_a^b f(t)\overline{\phi_j(t)}w(t)dt \right) \phi_j(x).$$

We equate coefficients to obtain

$$g(x) = \int_a^b \sum_{j=1}^{\infty} \frac{\phi_j(x)\overline{\phi_j(t)}}{\lambda_j} w(t)f(t)\; dt = \int_a^b G(x,t)f(t)w(t)\; dt. \tag{59}$$

We summarize the story. Assume that $(L - kI)^{-1}$ has a continuous Green's function. Then $(L - kI)^{-1}$ is compact and Hermitian, and a complete orthonormal system of eigenfunctions exists. Decompose the Hilbert space into eigenspaces E_{λ_j}. If $h \in E_{\lambda_j}$ we have $(L - kI)h = \lambda_j h$. Note that no λ_j equals 0. Thus, restricted to E_{λ_j}, we can invert $L - kI$ by

$$(L - kI)^{-1}(h) = \frac{1}{\lambda_j} h.$$

We invert in general by inverting on each eigenspace and adding up the results. Things are essentially the same as in Sect. 4 of Chap. 1, where we solved a linear system when there was an orthonormal basis of eigenvectors. In this setting, we see that the Green's function is given by

$$G(x,t) = \sum_{j=1}^{\infty} \frac{\phi_j(x)\overline{\phi_j(t)}}{\lambda_j}.$$

We consider the simple special case where $Ly = -y''$ on the interval $[0, 1]$ with boundary conditions $y(0) = y(1) = 0$. For each positive integer m, there is an eigenvalue $\pi^2 m^2$, corresponding to the normalized eigenfunction $\sqrt{2}\sin(m\pi x)$. In this case, $G(x, t)$ has the following expression:

$$G(x,t) = \begin{Bmatrix} x(1-t) & x < t \\ t(1-x) & x > t \end{Bmatrix}. \tag{60}$$

We can check this formula directly by differentiating twice the relation

$$y(x) = (1-x)\int_0^x tf(t)\; dt + x\int_x^1 (1-t)f(t)\; dt.$$

Of course, we discovered this formula by the prescription from Theorem 2.14. The function x is the solution vanishing at 0. The function $1-x$ is the solution vanishing at 1. See Fig. 2.5. Using orthonormal expansion, we have another expression for $G(x, t)$:

$$G(x,t) = 2\sum_{m=1}^{\infty} \frac{\sin(m\pi x)\,\sin(m\pi t)}{\pi^2 m^2}.$$

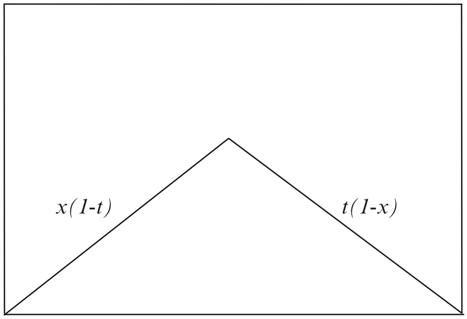

FIGURE 2.5. Green's function for the second derivative

See [F2, G] for many computational exercises involving Green's functions for Sturm–Liouville equations and generalizations. See also [GS] for excellent intuitive discussion concerning the construction of the Green's function and its connections with the Dirac delta function.

EXERCISE 2.47. Assume $0 \leq x < \frac{1}{2}$. Put $L = -(\frac{d}{dx})^2$ on $[0, 1]$ with boundary conditions $y(0) = y(1) = 0$. Equate the two expressions for the Green's function to establish the identity

$$x = \frac{4}{\pi^2} \sum_{r=0}^{\infty} \frac{(-1)^r \sin((2r+1)\pi x)}{(2r+1)^2}.$$

Prove that this identity remains true at $x = \frac{1}{2}$.

EXERCISE 2.48. Consider the equation $y'' + \lambda y = 0$ with boundary conditions $y(0) - y(1) = 0$ and $y'(0) + y'(1) = 0$. Show that every λ is an eigenvalue. Why doesn't this example contradict Theorem 2.13? Hint: Look carefully at (*SL.1*) and (*SL.2*).

EXERCISE 2.49. Suppose $L \in \mathcal{L}(\mathcal{H})$ is Hermitian. Find $\lim_{n\to\infty} ||L^n||^{\frac{1}{n}}$. Suggestion: If $L = L^*$, then $||L^2|| = ||L||^2$.

EXERCISE 2.50. Put the Bessel equation $x^2y'' + xy' + (\lambda^2x^2 - \nu^2)y = 0$ into Sturm–Liouville form.

EXERCISE 2.51. Find the Green's function for the equation $Ly = x^2y'' - 2xy' + 2y = f$ on the interval $[1, 2]$ with $y(1) = y(2) = 0$. (First put the equation in Sturm–Liouville form.) How does the answer change if the boundary condition is replaced by $y'(1) = y'(2) = 0$?

11.2. Exercises on Legendre Polynomials. The next several exercises involve the Legendre polynomials. These polynomials arise throughout pure and applied mathematics. We will return to them in Sect. 13.

We first remind the reader of a method for finding solutions to linear ordinary differential equations, called *reduction of order*. Consider a linear differential operator L of order m. Suppose we know one solution f to $Ly = g$. We then seek a solution of the form $y = uf$ for some unknown function u. The function u' will then satisfy a homogeneous linear differential equation of order $m-1$. We used a similar idea in Sect. 4.1 of Chap. 1, where we replaced a constant c with a function $c(x)$ when solving an inhomogeneous equation. We note, when $m = 2$, that the method of reduction of order yields a first-order equation for u' which can often be solved explicitly.

EXERCISE 2.52. Verify that the method of reduction of order works as described above.

EXERCISE 2.53. The Legendre equation (in Sturm–Liouville form) is

$$((1-x^2)y')' + n(n+1)y = 0. \tag{61}$$

Find all solutions to (61) when $n = 0$ and when $n = 1$. Comment: When $n = 1$, finding one solution is easy. The method of reduction of order can be used to find an independent solution.

EXERCISE 2.54. Let n be a nonnegative integer. Show that there is a polynomial solution P_n to (61) of degree n. Normalize to make $P_n(1) = 1$. This P_n is called the n-th Legendre polynomial. Show that an alternative definition of P_n is given for $|x| \leq 1$ and $|t| < 1$ by the generating function

$$\frac{1}{\sqrt{1-2xt+t^2}} = \sum_{n=0}^{\infty} P_n(x)t^n.$$

Show that the collection of these polynomials forms a complete orthogonal system for $L^2([-1,1], dx)$. Show that $||P_n||^2 = \frac{2}{2n+1}$. If needed, look ahead to the next section for one method to compute these norms.

EXERCISE 2.55. Obtain the first few Legendre polynomials by applying the Gram–Schmidt process to the monomials $1, x, x^2, x^3, x^4$.

EXAMPLE 2.5. The first few Legendre polynomials (See Fig. 2.6):

- $P_0(x) = 1.$
- $P_1(x) = x.$
- $P_2(x) = \frac{3x^2-1}{2}$
- $P_3(x) = \frac{5x^3-3x}{2}.$
- $P_4(x) = \frac{35x^4-30x^2+3}{8}.$

EXERCISE 2.56. Let P_n be the n-th Legendre polynomial. Show that

$$(n+1)P_{n+1}(x) - (2n+1)xP_n(x) + nP_{n-1}(x) = 0.$$

Use the method of difference equations to find constants a_k such that

$$P_n(x) = \sum_{k=0}^{n} a_k(1+x)^k(1-x)^{n-k}.$$

EXERCISE 2.57. Here is an alternative proof that the Legendre polynomials are orthogonal. First show that $P_n = c_n(\frac{d}{dx})^n(x^2-1)^n$. Then integrate by parts to show that

$$\langle P_n, f\rangle = c_n(-1)^n\langle (x^2-1)^n, (\frac{d}{dx})^n f\rangle.$$

In other words, f is orthogonal to P_n if f is a polynomial of degree less than n.

EXERCISE 2.58. Let P_l denote a Legendre polynomial. Define the *associated Legendre functions* with parameters l and m by

$$P_l^m(x) = (1-x^2)^{\frac{m}{2}} (\frac{d}{dx})^m P_l(x).$$

- Show when m is even that P_l^m is a polynomial.
- Obtain a differential equation satisfied by P_l^m by differentiating m-times the Sturm–Liouville equation (61) defining P_l.
- Show that $P_l^m(x)$ is a constant times a power of $(1-x^2)$ times a derivative of a power of $(1-x^2)$.

The associated Legendre functions arise in Sect. 13 on spherical harmonics.

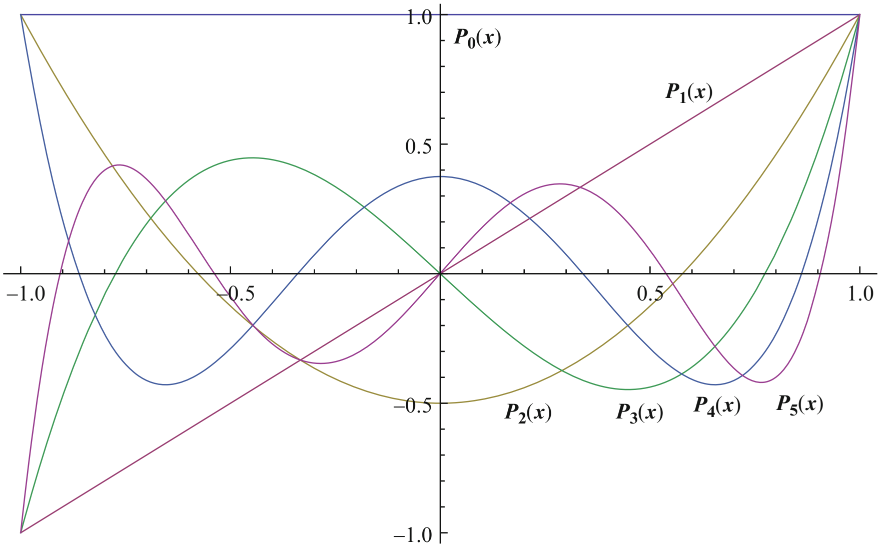

FIGURE 2.6. Legendre polynomials

12. Generating Functions and Orthonormal Systems

Many of the complete orthonormal systems used in physics and engineering are defined via the *Gram–Schmidt process*. Consider an interval I in $\mathbf{R}$ and the Hilbert space $L^2(I, w(x)dx)$ of square-integrable functions with respect to some weight function w. Starting with a nice class of functions, such as the monomials, and then orthonormalizing them, one obtains various *special functions*. The Gram–Schmidt process often leads to tedious computation.

Following the method of Exercise 2.54, we use *generating functions* to investigate orthonormal systems. In addition to the Legendre polynomials, we give two examples of importance in physics, the Laguerre polynomials and the Hermite polynomials. We return to the Hermite polynomials in Chap. 3, where we relate them to eigenfunctions of the Fourier transform.

We will use a simple proposition relating orthonormal systems and generating functions. We then show how the technique works for the Laguerre and Hermite polynomials.

Before stating and proving this proposition, we discuss vector-valued convergent power series. Let $\mathbf{B}$ denote the open unit disk in $\mathbf{C}$. Let $\mathcal{H}$ be a Hilbert space; it is often useful to consider complex analytic functions $f : \mathbf{B} \to \mathcal{H}$.

Consider a power series $A(z) = \sum A_n z^n$, where the coefficients A_n lie in $\mathcal{H}$. This series converges at the complex number z if its partial sums there form a Cauchy sequence in $\mathcal{H}$. We define a function $A : \mathbf{B} \to \mathcal{H}$ to be complex analytic if there is a sequence $\{A_n\}$ in $\mathcal{H}$ such that the series

$$\sum_{n=0}^{\infty} A_n z^n$$

converges to $A(z)$ for all z in $\mathbf{B}$. On compact subsets of $\mathbf{B}$, the series converges in norm, and we may therefore rearrange the order of summation at will.

Proposition 2.17. *Let $\mathcal{H}$ be a Hilbert space, and suppose $A : \mathbf{B} \to \mathcal{H}$ is complex analytic with $A(t) = \sum_{n=0}^{\infty} A_n t^n$. Then the collection of vectors $\{A_n\}$ forms an orthonormal system in $\mathcal{H}$ if and only if, for all $t \in \mathbf{B}$,*

$$||A(t)||^2 = \frac{1}{1 - |t|^2}.$$

Proof. Using the absolute convergence on compact subsets to order the summation as we wish, we obtain

$$||A(t)||^2 = \sum_{m,n=0}^{\infty} \langle A_n, A_m \rangle t^n \bar{t}^m. \tag{62}$$

Comparison with the geometric series yields the result: the right-hand side of (62) equals $\frac{1}{1-|t|^2}$ if and only if $\langle A_n, A_m \rangle$ equals 0 for $n \neq m$ and equals 1 for $n = m$. $\square$

Definition 2.17. The formal series

$$\sum_{n=0}^{\infty} L_n t^n$$

is the *ordinary generating function* for the sequence $\{L_n\}$. The formal series

$$\sum_{n=0}^{\infty} L_n \frac{t^n}{n!}$$

is the *exponential generating function* for the sequence $\{L_n\}$.

Explicit formulas for these generating functions often provide powerful insight as well as simple proofs of orthogonality relations.

Example 2.6 (Laguerre polynomials). Let $\mathcal{H} = L^2\left([0,\infty), e^{-x}dx\right)$ be the Hilbert space of square-integrable functions on $[0,\infty)$ with respect to the measure $e^{-x}dx$. Consider functions L_n defined via their generating function by

$$A(x,t) = \sum_{n=0}^{\infty} L_n(x) t^n = (1-t)^{-1} \exp\left(\frac{-xt}{1-t}\right).$$

Note that $x \geq 0$ and $|t| < 1$. In order to study the inner products $\langle L_n, L_m \rangle$, we compute $||A(x,t)||^2$. We will find an explicit formula for this squared norm; Proposition 2.17 implies that the L_n form an orthonormal system.

We have

$$|A(x,t)|^2 = (1-t)^{-1} \exp\left(\frac{-xt}{1-t}\right) (1-\bar{t})^{-1} \exp\left(\frac{-x\bar{t}}{1-\bar{t}}\right).$$

Multiplying by the weight function e^{-x} and integrating, we obtain

$$||A(x,t)||^2 = (1-t)^{-1}(1-\bar{t})^{-1} \int_0^\infty \exp\left(-x(1+\frac{t}{1-t}+\frac{\bar{t}}{1-\bar{t}})\right) dx.$$

Computing the integral on the right-hand side and simplifying shows that

$$||A(x,t)||^2 = \frac{1}{(1-t)(1-\bar{t})} \frac{1}{1+\frac{t}{1-t}+\frac{\bar{t}}{1-\bar{t}}} = \frac{1}{1-|t|^2}.$$

From Proposition 2.17, we see that $\{L_n\}$ forms an orthonormal system in $\mathcal{H}$.

The series defining the generating function converges for $|t| < 1$, and each L_n is real valued. In Exercise 2.60, we ask the reader to show that the functions L_n satisfy the Rodrigues formula

$$L_n(x) = \frac{e^x}{n!} \left(\frac{d}{dx}\right)^n (x^n e^{-x}) \tag{63}$$

and hence are polynomials of degree n. They are called the Laguerre polynomials, and they form a *complete* orthonormal system for $L^2([0,\infty), e^{-x}dx)$. Laguerre polynomials arise in solving the Schrödinger equation for a hydrogen atom (Fig. 2.7).

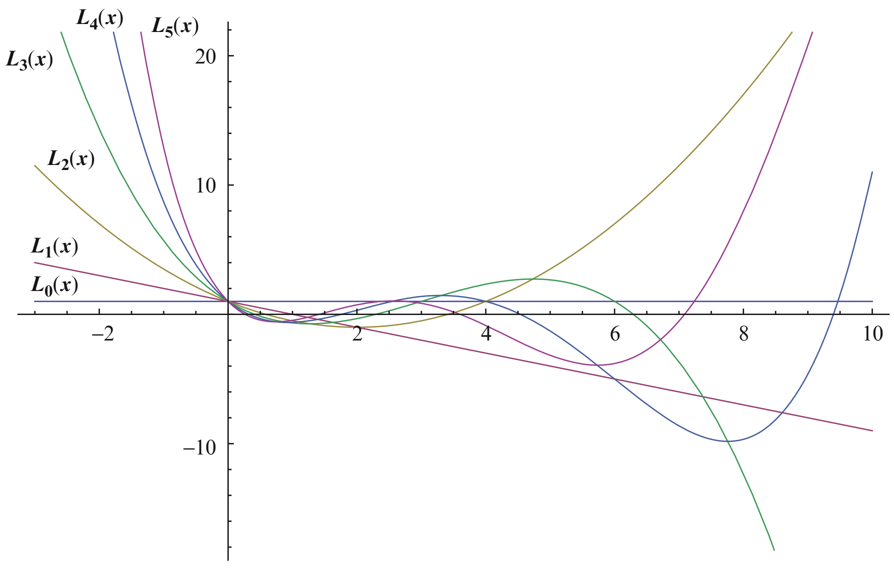

FIGURE 2.7. Laguerre polynomials

A similar technique works for the Hermite polynomials, which arise in many problems in physics, such as the quantum harmonic oscillator. See pp. 120–122 in [GS]. We discuss these polynomials at the end of Chap. 3. One way to define the Hermite polynomials is via the exponential generating function

$$\exp(2xt - t^2) = \sum H_n(x)\frac{t^n}{n!}. \tag{64}$$

The functions H_n are polynomials and form an orthogonal set for $\mathcal{H}=L^2(\mathbf{R}, e^{-x^2}dx)$. With this normalization, the norms are not equal to unity. In Exercise 2.62, the reader is asked to study the Hermite polynomials by mimicking the computations for the Laguerre polynomials. Other normalizations of these polynomials are also common. Sometimes the weight function used is $e^{\frac{-x^2}{2}}$. The advantage of our normalization is Theorem 3.9.

The technique of generating functions can also be used to find normalizing coefficients. Suppose, such as in the Sturm–Liouville setting, that the collection $\{f_n\}$ for $n \geq 0$ forms a complete orthogonal system. We wish to find $||f_n||_{L^2}$. Assume that we have found the generating function

$$B(x,t) = \sum_{n=0}^{\infty} f_n(x)t^n$$

explicitly. We may assume t is real. Taking L^2 norms (in x), we discover that $||f_n||^2$ must be the coefficient of t^{2n} in the series expansion of $||B(x,t)||_{L^2}^2$.

We illustrate this result by solving part of Exercise 2.54. The generating function for the Legendre polynomials is known to be

$$B(x,t) = \frac{1}{\sqrt{1-2xt+t^2}}.$$

By elementary calculus, its L^2 norm on $[-1,1]$ is found to satisfy

$$||B(x,t)||_{L^2}^2 = \frac{1}{t}\left(\log(1+t) - \log(1-t)\right).$$

Expanding $\log(1 \pm t)$ in a Taylor series shows that

$$||B(x,t)||_{L^2}^2 = 2\sum_{n=0}^{\infty} \frac{t^{2n}}{2n+1}.$$

Hence $||P_n||_{L^2}^2 = \frac{2}{2n+1}$.

Exercise 2.59. Fill in the details from the previous paragraph.

Exercise 2.60.

(1) With L_n as in Example 2.6, verify the Rodrigues formula (63). Suggestion: Write the power series of the exponential on the right-hand side of (63) and interchange the order of summation.
(2) Show that each L_n is a polynomial in x. Hint: The easiest way is to use (1).
(3) Prove that $\{L_n\}$ forms a *complete* system in $L^2\left([0,\infty), e^{-x}dx\right)$.

EXERCISE 2.61. For $x > 0$, verify that

$$\sum_{n=0}^{\infty} \frac{L_n(x)}{n+1} = \int_0^{\infty} \frac{e^{-xt}}{t+1} dt.$$

Suggestion: Integrate the relation

$$\sum_{n=0}^{\infty} L_n(x) s^n = (1-s)^{-1} \exp(\frac{-xs}{1-s})$$

over the interval $[0, 1]$ and then change variables in the integral.

EXERCISE 2.62 (Hermite polynomials). Here H_n is defined by (64).

(1) Use (64) to find a simple expression for

$$\sum_{n=0}^{\infty} H_n(x) t^n \sum_{m=0}^{\infty} H_m(x) s^m.$$

(2) Integrate the result in (1) over $\mathbf{R}$ with respect to the measure $e^{-x^2} dx$.
(3) Use (2) to show that the Hermite polynomials form an orthogonal system with

$$||H_n||^2 = 2^n n! \sqrt{\pi}.$$

(4) Prove that the system of Hermite polynomials is complete in $L^2(\mathbf{R}, e^{-x^2} dx)$.

EXERCISE 2.63. Replace the generating function used for the Legendre polynomials by $(1 - 2xt + t^2)^{-\lambda}$ for $\lambda > -\frac{1}{2}$ and carry out the same steps. The resulting polynomials are the *ultraspherical* or *Gegenbauer* polynomials. Note that the Legendre polynomials are the special case when $\lambda = \frac{1}{2}$. See how many properties of the Legendre polynomials you can generalize.

13. Spherical Harmonics

We close this chapter by discussing spherical harmonics. This topic provides one method to generalize Fourier series on the unit circle to orthonormal expansions on the unit sphere. One approach to spherical harmonics follows a thread of history, based on the work of Legendre. This approach relates the exercises from Sect. 11 on Legendre polynomials to elementary physics and relies on spherical coordinates from calculus. Perhaps the most elegant approach, given in Theorems 2.15 and 2.16, uses spaces of homogeneous polynomials. We discuss both approaches.

Let S^2 denote the unit sphere in real Euclidean space $\mathbf{R}^3$. Let Δ denote the Laplace operator $\sum_{j=1}^{3} \frac{\partial^2}{\partial x_j^2}$. We would like to find a complete orthonormal system for $L^2(S^2)$ whose properties are analogous to those of the exponentials e^{inx} on the unit circle. Doing so is not simple.

Recall that Newton's law of gravitation and Coulomb's law of electric charge both begin with a potential function. Imagine a mass or charge placed at a single point p in real Euclidean space $\mathbf{R}^3$. The potential at $\mathbf{x}$ due to this mass or charge is then a constant times the reciprocal of the distance from $\mathbf{x}$ to p. Let us suppose that the mass or charge is located at the point $(0, 0, 1)$. The potential at the point $\mathbf{x} = (x_1, x_2, x_3)$ is then (See Fig. 2.8)

$$\frac{c}{||\mathbf{x} - p||} = \frac{c}{\sqrt{(x_1^2 + x_2^2 + (x_3 - 1)^2)}}. \tag{65}$$

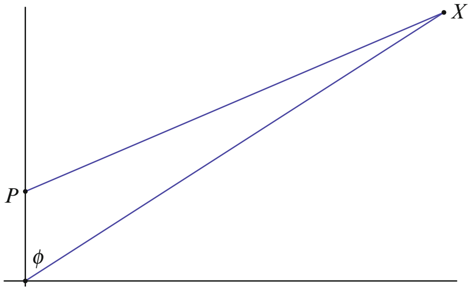

FIGURE 2.8. The colatitude ϕ

We wish to express (65) in spherical coordinates. We write

$$\mathbf{x} = (x_1, x_2, x_3) = (\rho\cos(\theta)\sin(\phi), \rho\sin(\theta)\sin(\phi), \rho\cos(\phi))$$

where ρ is the distance to the origin, θ is the usual polar coordinate angle in the (x_1, x_2) plane measuring longitude, and ϕ is the colatitude. Thus, $0 \leq \theta \leq 2\pi$, whereas $0 \leq \phi \leq \pi$. These conventions are common in calculus books, but the physics literature often interchanges θ and ϕ. Also, sometimes r is used instead of ρ. In many sources, however, r is reserved for its role in cylindrical coordinates, and thus $r^2 = x^2 + y^2$.

Writing (65) in spherical coordinates we obtain

$$\frac{c}{||\mathbf{x} - p||} = \frac{c}{\sqrt{1 + \rho^2 - 2\rho\cos(\phi)}}. \tag{66}$$

The denominator in (66) is the same expression as in the generating function for the Legendre polynomials P_n from Exercise 2.54, with t replaced by ρ and x replaced by $\cos(\phi)$. Therefore, we can rewrite (66) as follows:

$$\frac{c}{||\mathbf{x} - p||} = c\sum_{n=0}^{\infty} P_n(\cos(\phi))||\mathbf{x}||^n. \tag{67}$$

The potential function from (65) is harmonic away from p. We leave the computation to Exercise 2.64. We write the Laplace operator in spherical coordinates:

$$\Delta(f) = \frac{1}{\rho^2}\frac{\partial}{\partial\rho}(\rho^2 f_\rho) + \frac{1}{\rho^2\sin(\phi)}\frac{\partial}{\partial\phi}(\sin(\phi)f_\phi) + \frac{1}{\rho^2\sin^2(\phi)}f_{\theta\theta}. \tag{68}$$

We attempt to solve the Laplace equation $\Delta(f) = 0$ using separation of variables, generalizing Exercise 1.57. Thus we assume that

$$f(\rho, \theta, \phi) = A(\rho)B(\phi)C(\theta). \tag{69}$$

Plugging (69) into the Laplace equation yields the equation

$$0 = (\rho^2 A' BC)_\rho + \frac{1}{\sin(\phi)} (\sin(\phi) AB'C)_\phi + \frac{1}{\sin^2(\phi)} ABC''. \tag{70}$$

After dividing by ABC, we obtain

$$0 = (\frac{\rho^2 A'' + 2\rho A'}{A}) + \frac{B''}{B} + \cot(\phi)\frac{B'}{B} + \frac{1}{\sin^2(\phi)} \frac{C''}{C}. \tag{71}$$

The first fraction in (71) depends on ρ; the other terms do not. Hence there is a constant λ such that

$$\rho^2 A'' + 2\rho A' = \lambda A. \tag{72}$$

Furthermore, we also have

$$\left(\frac{B''}{B} + \cot(\phi)\frac{B'}{B}\right) \sin^2(\phi) + \frac{C''}{C} = -\lambda \sin^2(\phi). \tag{73}$$

The only solutions to the (72) for A that are continuous at zero are $A(\rho) = c\rho^l$ for nonnegative integers l. It follows that $\lambda = l(l+1)$.

Now we look at (73). Again by grouping the θ and ϕ terms separately, we obtain two equations:

$$\frac{C''}{C} = -\mu \tag{74}$$

$$\sin^2(\phi) \left(\frac{B''}{B} + \cot(\phi)\frac{B'}{B} + \lambda\right) = \mu. \tag{75}$$

Now (74) must be periodic in θ. Hence μ is the square of an integer k. We see that $C(\theta) = ce^{ik\theta}$. Also (75) becomes

$$\sin^2(\phi) \left(\frac{B''}{B} + \cot(\phi)\frac{B'}{B} + \lambda\right) = k^2. \tag{76}$$

Simplifying (76) leads to the equation

$$B'' + \cot(\phi)B' + (l(l+1) - \frac{k^2}{\sin^2(\phi)})B = 0. \tag{77}$$

Equation (77) evokes the differential equation defining the Legendre polynomials. In fact, if we make the substitution $x = \cos(\phi)$, then (77) is precisely equivalent (See Exercise 2.66) to the equation

$$(1-x^2)B_{xx} - 2xB_x + \left(l(l+1) - \frac{k^2}{(1-x^2)}\right) B = 0. \tag{78}$$

The solutions P_l^k to (78) are the associated Legendre functions from Exercise 2.58 when $k \geq 0$ and related expressions when $k < 0$. The function $e^{ik\theta} P_l^k(\cos(\phi))$ is the spherical harmonic $Y_l^k(\theta, \phi)$. The integer parameter k varies from $-l$ to l, yielding $2l+1$ independent functions. The functions $\rho^l e^{ik\theta} P_l^k(\cos(\phi))$ are harmonic. The functions Y_l^k are not themselves harmonic in general; on the sphere each Y_l^k is an eigenfunction of the Laplacian with eigenvalue $-l(l+1)$.

A Wikipedia page called *Table of spherical harmonics* lists these Y_l^k, including the normalizing constants, for $0 \leq l \leq 10$ and all corresponding k . The functions Y_l^k and Y_b^a are orthogonal, on $L^2(S^2)$, unless $k = a$ and $l = b$. These functions form a complete orthogonal system for $L^2(S^2)$. Remarkable additional properties whose discussion is beyond the scope of this book hold as well.

We next approach spherical harmonics via homogeneous polynomials. Things are simpler this way but perhaps less useful in applied mathematics.

We will work in $\mathbf{R}^n$, although we will write some formulas explicitly when $n = 3$. Let $\mathbf{x} = (x_1, \ldots, x_n)$ denote the variables. A polynomial $p(\mathbf{x})$ is homogeneous of degree k if $p(t\mathbf{x}) = t^k p(\mathbf{x})$. Homogeneous polynomials are therefore determined by their values on the unit sphere. It is often useful to identify a homogeneous polynomial $p(\mathbf{x})$ with the function

$$P(\mathbf{x}) = \frac{p(\mathbf{x})}{||\mathbf{x}||^k},$$

which is defined in the complement of the origin, agrees with p on the sphere, and is homogeneous of degree 0. See Proposition 2.18. For each m, we write $\mathbf{H}_m$ for the vector space of homogeneous harmonic polynomials of degree m. In Theorem 2.16, we will compute the dimension of $\mathbf{H}_m$. When $n = 3$, its dimension turns out be $2m + 1$. We obtain spherical harmonics by restricting harmonic homogeneous polynomials to the unit sphere.

EXAMPLE 2.7. Put $n = 3$. When $m = 1$, the harmonic polynomials x, y, z form a basis for $\mathbf{H}_1$. For $m = 2$, the following five polynomials form a basis for $\mathbf{H}_2$:

- xy
- xz
- yz
- $x^2 + y^2 - 2z^2$
- $x^2 - 2y^2 + z^2$.

Note that the harmonic polynomial $-2x^2 + y^2 + z^2$ is linearly dependent on the last two items in the list.

It will be as easy to work in $\mathbf{R}^n$ as it is in $\mathbf{R}^3$. We write $v \cdot w$ for the usual inner product of v, w in $\mathbf{R}^n$. We assume $n \geq 2$.

Let V_m denote the vector space of homogeneous polynomials of degree m in the variable $\mathbf{x}$ in $\mathbf{R}^n$. We regard $\mathbf{H}_m$ as a subspace of V_m. The dimension of V_m is the binomial coefficient $\binom{m+n-1}{n-1}$. We have a map $M : V_m \to V_{m+2}$ given by multiplication by $||\mathbf{x}||^2$. The Laplace operator Δ maps the other direction. These operators turn out to be adjoints. See Theorem 2.16.

We begin with a remarkable formula involving the Laplacian on harmonic, homogeneous polynomials on $\mathbf{R}^n$. The function P in Proposition 2.18 below is homogeneous of degree 0, and hence its Laplacian is homogeneous of degree -2. This observation explains why we must divide by $||\mathbf{x}||^2$ in (79).

PROPOSITION 2.18. *Let p be a harmonic, homogeneous polynomial of degree l on $\mathbf{R}^n$. Outside the origin, consider the function P defined by*

$$P(\mathbf{x}) = \frac{p(\mathbf{x})}{||\mathbf{x}||^l}.$$

Then we have

$$\Delta(P) = -l(l + n - 2)\frac{P(\mathbf{x})}{||\mathbf{x}||^2}. \tag{79}$$

Restricted to the sphere, P defines an eigenfunction of the Laplacian with eigenvalue $-l(l+n-2)$. When $n=3$, P is therefore a linear combination of the spherical harmonics Y_l^k with $-l \leq k \leq l$.

PROOF. See Exercise 2.72 for the computation yielding (79). The second statement follows from (79) by putting $||\mathbf{x}||^2$ equal to 1. The last statement follows from the discussion just after (78). $\square$

Consider the Hilbert space $L^2(S^{n-1})$, where S^{n-1} is the unit sphere in n-dimensions, and $n \geq 2$. In order to integrate over the unit sphere, we use n-dimensional spherical coordinates. We put $\mathbf{x} = \rho\mathbf{v}$, where $\rho = ||\mathbf{x}||$ and $\mathbf{v}$ lies on the unit sphere. We then can write the volume form dV on $\mathbf{R}^n$ as

$$dV(\mathbf{x}) = \rho^{n-1} d\rho \; d\sigma(\mathbf{v}).$$

Let f be a function on $\mathbf{R}^n$. Away from $\mathbf{0}$, we define a function F by

$$F(\mathbf{x}) = f(\frac{\mathbf{x}}{||\mathbf{x}||}) = f(\mathbf{v}).$$

The function F satisfies $F(t\mathbf{x}) = F(\mathbf{x})$ when $t > 0$. Such a function is called positive homogeneous of degree 0. We note a special case of Euler's formula for such functions, when F is differentiable. See Exercise 2.71 for a more general statement.

PROPOSITION 2.19. *Assume F is differentiable and $F(t\mathbf{x}) = F(\mathbf{x})$ for $t > 0$ and all $\mathbf{x}$. Then $dF(\mathbf{x}) \cdot \mathbf{x} = 0$.*

PROOF. Apply $\frac{d}{dt}$ to the equation $F(t\mathbf{x}) = F(\mathbf{x})$ and set $t = 1$. $\square$

Let χ be a smooth function on $\mathbf{R}$ with the following properties:

(1) $\chi(0) = 0$.
(2) $\chi(t)$ tends to 0 as t tends to infinity.
(3) $\int_0^\infty \chi(t^2) t^{n-1} dt = 1$. (Here n is the dimension.)

Given a smooth function w, we wish to compute $\int_{S^{n-1}} w d\sigma$. Because of property (3) of χ, the integration formula (80) holds. It allows us to express integrals over the sphere as integrals over Euclidean space:

$$\int_{\mathbf{R}^n} \chi(||\mathbf{x}||^2) w(\frac{\mathbf{x}}{||\mathbf{x}||}) dV = \int_{S^{n-1}} \int_0^\infty \chi(\rho^2) \rho^{n-1} d\rho \; w(\mathbf{v}) d\sigma(\mathbf{v}) = \int_{S^{n-1}} w d\sigma. \quad (80)$$

The other two properties of χ will be useful in an integration by parts.

THEOREM 2.15. *For $k \neq l$, the subspaces $\mathbf{H}_k$ and $\mathbf{H}_l$ are orthogonal in $L^2(S^{n-1})$.*

PROOF. Given harmonic homogeneous polynomials f of degree k and g of degree l, let F and G be the corresponding homogeneous functions of degree 0 defined above. By Proposition 2.18, these functions are eigenfunctions of the Laplacian on the sphere, with distinct eigenvalues. We claim that the Laplacian is Hermitian:

$$\int_{S^{n-1}} \Delta F \; \overline{G} \; d\sigma = \int_{S^{n-1}} F \; \overline{\Delta G} \; d\sigma. \quad (81)$$

Given the claim, eigenfunctions corresponding to distinct eigenvalues are orthogonal. Thus harmonic, homogeneous polynomials of different degrees are orthogonal on the unit sphere.

It remains to prove (81). We may assume that G is real. Let

$$W = \Delta F\ G - F\ \Delta G = \sum_{j=1}^{n} (\frac{\partial}{dx_j})(F_{x_j} G - F G_{x_j}).$$

We integrate by parts in (80), moving each $\frac{\partial}{\partial x_j}$. Note that $\frac{\partial}{\partial x_j}(||\mathbf{x}||^2) = 2x_j$.

$$\int_{S^{n-1}} W\ d\sigma = \int_{\mathbf{R}^n} \chi(||\mathbf{x}||^2) W(\frac{\mathbf{x}}{||\mathbf{x}||})\ dV$$

$$= - \int_{\mathbf{R}^n} \sum_{j=1}^{n} (F_{x_j} G - F G_{x_j}) \chi'(||\mathbf{x}||^2) 2x_j\ dV. \tag{82}$$

The last term in (82) is zero by Proposition 2.19, because F and G are positive homogeneous of degree 0. Thus Δ is Hermitian. $\square$

It is convenient to define particular inner products on the spaces V_m, which differ from the usual inner product given by integration. By linearity, to define the inner product on V_m, it suffices to define the inner product of monomials. We illustrate for $n = 3$. Put

$$\langle x^a y^b z^c, x^A y^B z^C \rangle_{V_m} = 0 \tag{83}$$

unless $a = A$, $b = B$, and $c = C$. In this case, we put $||x^a y^b z^c||^2_{V_m} = a!b!c!$. The generalization to other dimensions is evident:

$$|| \prod_{j=1}^{n} x_j^{a_j} ||^2_{V_m} = \prod_{j=1}^{n} a_j!.$$

Thus distinct monomials are decreed to be orthogonal.

Theorem 2.16. *The mapping $M : V_m \to V_{m+2}$ is the adjoint of the mapping $\Delta : V_{m+2} \to V_m$. In other words,*

$$\langle Mf, g \rangle_{V_{m+2}} = \langle f, \Delta g \rangle_{V_m}. \tag{84}$$

Hence the image of M is orthogonal to the harmonic space $\mathbf{H}_{m+2}$ and

$$V_{m+2} = M(V_m) \oplus \mathbf{H}_{m+2}.$$

Furthermore, $\mathbf{H}_m$ is of dimension $\binom{m+n-1}{n-1} - \binom{m+n-3}{n-1}$. When $n = 3$, this dimension is $2m+1$.

Proof. To be concrete, we write out the proof when $n = 3$. By linearity, it suffices to check (84) on monomials $f = x^a y^b z^c$ and $g = x^A y^B z^C$, where it follows by computing both sides of (84) in terms of factorials. There are three possible circumstances in which the inner product is not zero:

- $(a, b, c) = (A - 2, B, C)$
- $(a, b, c) = (A, B - 2, C)$
- $(a, b, c) = (A, B, C - 2)$.

In the first case, we must check that $(a + 2)!b!c! = A(A - 1)(A - 2)!B!C!$, which holds. The other two cases are similarly easy, and hence (84) holds.

Next, suppose that h is in the image of M and that g is in the nullspace of Δ. Then (84) gives

$$\langle h, g \rangle_{V_{m+2}} = \langle Mf, g \rangle_{V_{m+2}} = \langle f, \Delta g \rangle_{V_m} = 0.$$

The desired orthogonality thus holds and the direct sum decomposition follows. Finally, the dimension of V_m is $\binom{m+n-1}{n-1}$. Since M is injective, the dimension of the image of M is the dimension of V_m. The dimension of $\mathbf{H}_{m+2}$ is therefore

$$\binom{m+n+1}{n-1} - \binom{m+n-1}{n-1}.$$

When $n = 3$, the dimension of $\mathbf{H}_{m+2}$ therefore is

$$\frac{(m+4)(m+3)}{2} - \frac{(m+2)(m+1)}{2} = 2m+5,$$

and hence the dimension of $\mathbf{H}_m$ is $2m+1$. □

REMARK 2.3. The formula in Theorem 2.16 for the dimension of $\mathbf{H}_m$ defines a polynomial of degree $n-2$ in m. See Exercise 2.75.

COROLLARY 2.9. *On the sphere, we have* $V_m = \mathbf{H}_m \oplus \mathbf{H}_{m-2} \oplus \dots$.

PROOF. The formula follows by iterating the equality $V_m = M(V_{m-2}) \oplus \mathbf{H}_m$ and noting that $||\mathbf{x}||^2 = 1$ on the sphere. □

COROLLARY 2.10. *Suppose f is continuous on the unit sphere. Then there is a sequence of harmonic polynomials converging uniformly to f.*

PROOF. This proof assumes the Stone–Weierstrass theorem to the effect that a continuous function on a compact subset S of $\mathbf{R}^n$ is the uniform limit on S of a sequence of polynomials. We proved this result in Corollary 1.8 when S is the circle. Given this theorem, the result follows from Corollary 2.9, because each polynomial can be decomposed on the sphere in terms of harmonic polynomials. □

COROLLARY 2.11. *The spherical harmonics form a complete orthogonal system for $L^2(S^2)$.*

We illustrate Corollary 2.11 for $m = 0$ and $m = 1$, when $n = 3$. Of course V_0 is the span of the constant 1. Its image under M is the span of $x^2 + y^2 + z^2$. The space $\mathbf{H}_2$ is spanned by the five functions $xy, xz, yz, x^2 + y^2 - 2z^2, x^2 - 2y^2 + z^2$. Each of these is orthogonal to $x^2 + y^2 + z^2$, which spans the orthogonal complement of $\mathbf{H}_2$. Next, V_1 is spanned by x, y, z. Its image under M is the span of $x(x^2 + y^2 + z^2), y(x^2 + y^2 + z^2), z(x^2 + y^2 + z^2)$. The space V_3 has dimension ten. The seven-dimensional space $\mathbf{H}_3$ is the orthogonal complement of the span of $M(V_1)$.

EXERCISE 2.64. Show that (65) defines a harmonic function away from $(0, 0, 1)$. Use both Euclidean coordinates and spherical coordinates.

EXERCISE 2.65. Verify formula (68).

EXERCISE 2.66. Use the chain rule (and some computation) to show that (77) and (78) are equivalent. Suggestion: First show that

$$B_{\phi\phi} = B_{xx} x_\phi^2 + B_x x_{\phi\phi}.$$

EXERCISE 2.67. For $n = 3$, express the harmonic polynomials of degree two using spherical coordinates.

EXERCISE 2.68. For $n = 3$, find seven linearly independent harmonic polynomials of degree three.

EXERCISE 2.69 (Difficult). Analyze (78) fully in terms of Legendre polynomials.

EXERCISE 2.70. Verify (79) if $p(x, y, z) = x^2 - y^2$.

EXERCISE 2.71. Verify Euler's identity: If f is differentiable and homogeneous of degree k on $\mathbf{R}^n$, then

$$df(\mathbf{x}) \cdot \mathbf{x} = kf(\mathbf{x}).$$

Proposition 2.19 was the case $k = 0$. What is the geometric interpretation of the result in this case?

EXERCISE 2.72. Verify (79). Euler's identity is useful.

EXERCISE 2.73. Take $n = 2$, and regard $\mathbf{R}^2$ as $\mathbf{C}$. Consider the harmonic polynomial $\mathrm{Re}(z^{2m})$. Give a much simpler proof of the analogue of formula (79) using the formula $\Delta(u) = 4u_{z\overline{z}}$ from Sect. 11 of Chap. 1.

EXERCISE 2.74. Again regard $\mathbf{R}^2$ as $\mathbf{C}$. Write down a basis for the homogeneous harmonic polynomials of degree m in terms of z and $\overline{z}$. Comment: The answer is obvious!

EXERCISE 2.75. For $n \geq 2$, simplify the formula in Theorem 2.16 to show that $\dim(\mathbf{H}_m)$ is a polynomial of degree $n - 2$ in m.

http://www.springer.com/978-1-4614-8525-4

Hermitian Analysis
From Fourier Series to Cauchy-Riemann Geometry
D'Angelo, J.
2013, X, 203 p. 27 illus., 19 illus. in color., Hardcover
ISBN: 978-1-4614-8525-4
A product of Birkhäuser Basel