Chapter 8

Bounded Linear Operators on a Hilbert Space

In this chapter we describe some important classes of bounded linear operators on Hilbert spaces, including projections, unitary operators, and self-adjoint operators. We also prove the Riesz representation theorem, which characterizes the bounded linear functionals on a Hilbert space, and discuss weak convergence in Hilbert spaces.

8.1 Orthogonal projections

We begin by describing some algebraic properties of projections. If M and N are subspaces of a linear space X such that every $x \in X$ can be written uniquely as $x = y + z$ with $y \in M$ and $z \in N$, then we say that $X = M \oplus N$ is the *direct sum* of M and N, and we call N a *complementary subspace* of M in X. The decomposition $x = y + z$ with $y \in M$ and $z \in N$ is unique if and only if $M \cap N = \{0\}$. A given subspace M has many complementary subspaces. For example, if $X = \mathbb{R}^3$ and M is a plane through the origin, then any line through the origin that does not lie in M is a complementary subspace. Every complementary subspace of M has the same dimension, and the dimension of a complementary subspace is called the *codimension* of M in X.

If $X = M \oplus N$, then we define the projection $P : X \to X$ of X onto M along N by $Px = y$, where $x = y + z$ with $y \in M$ and $z \in N$. This projection is linear, with $\operatorname{ran} P = M$ and $\ker P = N$, and satisfies $P^2 = P$. As we will show, this property characterizes projections, so we make the following definition.

Definition 8.1 A *projection* on a linear space X is a linear map $P : X \to X$ such that

$$P^2 = P. \tag{8.1}$$

Any projection is associated with a direct sum decomposition.

Theorem 8.2 Let X be a linear space.

(a) If $P : X \to X$ is a projection, then $X = \operatorname{ran} P \oplus \ker P$.
(b) If $X = M \oplus N$, where M and N are linear subpaces of X, then there is a projection $P : X \to X$ with $\operatorname{ran} P = M$ and $\ker P = N$.

Proof. To prove (a), we first show that $x \in \operatorname{ran} P$ if and only if $x = Px$. If $x = Px$, then clearly $x \in \operatorname{ran} P$. If $x \in \operatorname{ran} P$, then $x = Py$ for some $y \in X$, and since $P^2 = P$, it follows that $Px = P^2 y = Py = x$.

If $x \in \operatorname{ran} P \cap \ker P$ then $x = Px$ and $Px = 0$, so $\operatorname{ran} P \cap \ker P = \{0\}$. If $x \in X$, then we have

$$x = Px + (x - Px),$$

where $Px \in \operatorname{ran} P$ and $(x - Px) \in \ker P$, since

$$P(x - Px) = Px - P^2 x = Px - Px = 0.$$

Thus $X = \operatorname{ran} P \oplus \ker P$.

To prove (b), we observe that if $X = M \oplus N$, then $x \in X$ has the unique decomposition $x = y + z$ with $y \in M$ and $z \in N$, and $Px = y$ defines the required projection. □

When using Hilbert spaces, we are particularly interested in orthogonal subspaces. Suppose that $\mathcal{M}$ is a closed subspace of a Hilbert space $\mathcal{H}$. Then, by Corollary 6.15, we have $\mathcal{H} = \mathcal{M} \oplus \mathcal{M}^\perp$. We call the projection of $\mathcal{H}$ onto $\mathcal{M}$ along $\mathcal{M}^\perp$ the *orthogonal projection* of $\mathcal{H}$ onto $\mathcal{M}$. If $x = y + z$ and $x' = y' + z'$, where $y, y' \in \mathcal{M}$ and $z, z' \in \mathcal{M}^\perp$, then the orthogonality of $\mathcal{M}$ and $\mathcal{M}^\perp$ implies that

$$\langle Px, x' \rangle = \langle y, y' + z' \rangle = \langle y, y' \rangle = \langle y + z, y' \rangle = \langle x, Px' \rangle. \tag{8.2}$$

This equation states that an orthogonal projection is self-adjoint (see Section 8.4). As we will show, the properties (8.1) and (8.2) characterize orthogonal projections. We therefore make the following definition.

Definition 8.3 An *orthogonal projection* on a Hilbert space $\mathcal{H}$ is a linear map $P : \mathcal{H} \to \mathcal{H}$ that satisfies

$$P^2 = P, \qquad \langle Px, y \rangle = \langle x, Py \rangle \quad \text{for all } x, y \in \mathcal{H}.$$

An orthogonal projection is necessarily bounded.

Proposition 8.4 If P is a nonzero orthogonal projection, then $\|P\| = 1$.

Proof. If $x \in \mathcal{H}$ and $Px \neq 0$, then the use of the Cauchy-Schwarz inequality implies that

$$\|Px\| = \frac{\langle Px, Px \rangle}{\|Px\|} = \frac{\langle x, P^2 x \rangle}{\|Px\|} = \frac{\langle x, Px \rangle}{\|Px\|} \leq \|x\|.$$

Therefore $\|P\| \leq 1$. If $P \neq 0$, then there is an $x \in \mathcal{H}$ with $Px \neq 0$, and $\|P(Px)\| = \|Px\|$, so that $\|P\| \geq 1$. □

There is a one-to-one correspondence between orthogonal projections P and closed subspaces $\mathcal{M}$ of $\mathcal{H}$ such that $\operatorname{ran} P = \mathcal{M}$. The kernel of the orthogonal projection is the orthogonal complement of $\mathcal{M}$.

Theorem 8.5 Let $\mathcal{H}$ be a Hilbert space.

(a) If P is an orthogonal projection on $\mathcal{H}$, then $\operatorname{ran} P$ is closed, and

$$\mathcal{H} = \operatorname{ran} P \oplus \ker P$$

is the orthogonal direct sum of $\operatorname{ran} P$ and $\ker P$.

(b) If $\mathcal{M}$ is a closed subspace of $\mathcal{H}$, then there is an orthogonal projection P on $\mathcal{H}$ with $\operatorname{ran} P = \mathcal{M}$ and $\ker P = \mathcal{M}^\perp$.

Proof. To prove (a), suppose that P is an orthogonal projection on $\mathcal{H}$. Then, by Theorem 8.2, we have $\mathcal{H} = \operatorname{ran} P \oplus \ker P$. If $x = Py \in \operatorname{ran} P$ and $z \in \ker P$, then

$$\langle x, z\rangle = \langle Py, z\rangle = \langle y, Pz\rangle = 0,$$

so $\operatorname{ran} P \perp \ker P$. Hence, we see that $\mathcal{H}$ is the orthogonal direct sum of $\operatorname{ran} P$ and $\ker P$. It follows that $\operatorname{ran} P = (\ker P)^\perp$, so $\operatorname{ran} P$ is closed.

To prove (b), suppose that $\mathcal{M}$ is a closed subspace of $\mathcal{H}$. Then Corollary 6.15 implies that $\mathcal{H} = \mathcal{M} \oplus \mathcal{M}^\perp$. We define a projection $P : \mathcal{H} \to \mathcal{H}$ by

$$Px = y, \qquad \text{where } x = y + z \text{ with } y \in \mathcal{M} \text{ and } z \in \mathcal{M}^\perp.$$

Then $\operatorname{ran} P = \mathcal{M}$, and $\ker P = \mathcal{M}^\perp$. The orthogonality of P was shown in (8.2) above. □

If P is an orthogonal projection on $\mathcal{H}$, with range $\mathcal{M}$ and associated orthogonal direct sum $\mathcal{H} = \mathcal{M} \oplus \mathcal{N}$, then $I - P$ is the orthogonal projection with range $\mathcal{N}$ and associated orthogonal direct sum $\mathcal{H} = \mathcal{N} \oplus \mathcal{M}$.

Example 8.6 The space $L^2(\mathbb{R})$ is the orthogonal direct sum of the space $\mathcal{M}$ of even functions and the space $\mathcal{N}$ of odd functions. The orthogonal projections P and Q of $\mathcal{H}$ onto $\mathcal{M}$ and $\mathcal{N}$, respectively, are given by

$$Pf(x) = \frac{f(x) + f(-x)}{2}, \qquad Qf(x) = \frac{f(x) - f(-x)}{2}.$$

Note that $I - P = Q$.

Example 8.7 Suppose that A is a measurable subset of $\mathbb{R}$ — for example, an interval — with characteristic function

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

Then

$$P_A f(x) = \chi_A(x) f(x)$$

is an orthogonal projection of $L^2(\mathbb{R})$ onto the subspace of functions with support contained in $\overline{A}$.

A frequently encountered case is that of projections onto a one-dimensional subspace of a Hilbert space $\mathcal{H}$. For any vector $u \in \mathcal{H}$ with $\|u\| = 1$, the map P_u defined by

$$P_u x = \langle u, x \rangle u$$

projects a vector orthogonally onto its component in the direction u. Mathematicians use the tensor product notation $u \otimes u$ to denote this projection. Physicists, on the other hand, often use the "bra-ket" notation introduced by Dirac. In this notation, an element x of a Hilbert space is denoted by a "bra" $\langle x|$ or a "ket" $|x\rangle$, and the inner product of x and y is denoted by $\langle x \mid y \rangle$. The orthogonal projection in the direction u is then denoted by $|u\rangle\langle u|$, so that

$$(|u\rangle\langle u|)\,|x\rangle = \langle u \mid x \rangle |u\rangle.$$

Example 8.8 If $\mathcal{H} = \mathbb{R}^n$, the orthogonal projection $P_{\mathbf{u}}$ in the direction of a unit vector $\mathbf{u}$ has the rank one matrix $\mathbf{u}\mathbf{u}^T$. The component of a vector $\mathbf{x}$ in the direction $\mathbf{u}$ is $P_{\mathbf{u}}\mathbf{x} = (\mathbf{u}^T\mathbf{x})\mathbf{u}$.

Example 8.9 If $\mathcal{H} = l^2(\mathbb{Z})$, and $u = e_n$, where

$$e_n = (\delta_{k,n})_{k=-\infty}^{\infty},$$

and $x = (x_k)$, then $P_{e_n} x = x_n e_n$.

Example 8.10 If $\mathcal{H} = L^2(\mathbb{T})$ is the space of 2π-periodic functions and $u = 1/\sqrt{2\pi}$ is the constant function with norm one, then the orthogonal projection P_u maps a function to its mean: $P_u f = \langle f \rangle$, where

$$\langle f \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(x)\, dx.$$

The corresponding orthogonal decomposition,

$$f(x) = \langle f \rangle + f'(x),$$

decomposes a function into a constant mean part $\langle f \rangle$ and a fluctuating part f' with zero mean.

8.2 The dual of a Hilbert space

A *linear functional* on a complex Hilbert space $\mathcal{H}$ is a linear map from $\mathcal{H}$ to $\mathbb{C}$. A linear functional φ is bounded, or continuous, if there exists a constant M such that

$$|\varphi(x)| \leq M\|x\| \quad \text{for all } x \in \mathcal{H}. \tag{8.3}$$

The norm of a bounded linear functional φ is

$$\|\varphi\| = \sup_{\|x\|=1} |\varphi(x)|. \tag{8.4}$$

If $y \in \mathcal{H}$, then

$$\varphi_y(x) = \langle y, x \rangle \tag{8.5}$$

is a bounded linear functional on $\mathcal{H}$, with $\|\varphi_y\| = \|y\|$.

Example 8.11 Suppose that $\mathcal{H} = L^2(\mathbb{T})$. Then, for each $n \in \mathbb{Z}$, the functional $\varphi_n : L^2(\mathbb{T}) \to \mathbb{C}$,

$$\varphi_n(f) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{T}} f(x) e^{-inx}\, dx,$$

that maps a function to its nth Fourier coefficient is a bounded linear functional. We have $\|\varphi_n\| = 1$ for every $n \in \mathbb{Z}$.

One of the fundamental facts about Hilbert spaces is that all bounded linear functionals are of the form (8.5).

Theorem 8.12 (Riesz representation) If φ is a bounded linear functional on a Hilbert space $\mathcal{H}$, then there is a unique vector $y \in \mathcal{H}$ such that

$$\varphi(x) = \langle y, x \rangle \qquad \text{for all } x \in \mathcal{H}. \tag{8.6}$$

Proof. If $\varphi = 0$, then $y = 0$, so we suppose that $\varphi \neq 0$. In that case, $\ker \varphi$ is a proper closed subspace of $\mathcal{H}$, and Theorem 6.13 implies that there is a nonzero vector $z \in \mathcal{H}$ such that $z \perp \ker \varphi$. We define a linear map $P : \mathcal{H} \to \mathcal{H}$ by

$$Px = \frac{\varphi(x)}{\varphi(z)}\, z.$$

Then $P^2 = P$, so Theorem 8.2 implies that $\mathcal{H} = \operatorname{ran} P \oplus \ker P$. Moreover,

$$\operatorname{ran} P = \{\alpha z \mid \alpha \in \mathbb{C}\}, \qquad \ker P = \ker \varphi,$$

so that $\operatorname{ran} P \perp \ker P$. It follows that P is an orthogonal projection, and

$$\mathcal{H} = \{\alpha z \mid \alpha \in \mathbb{C}\} \oplus \ker \varphi$$

is an orthogonal direct sum. We can therefore write $x \in \mathcal{H}$ as

$$x = \alpha z + n, \qquad \alpha \in \mathbb{C} \text{ and } n \in \ker \varphi.$$

Taking the inner product of this decomposition with z, we get

$$\alpha = \frac{\langle z, x \rangle}{\|z\|^2},$$

and evaluating φ on $x = \alpha z + n$, we find that

$$\varphi(x) = \alpha\varphi(z).$$

The elimination of α from these equations, and a rearrangement of the result, yields

$$\varphi(x) = \langle y, x\rangle,$$

where

$$y = \frac{\overline{\varphi(z)}}{\|z\|^2}\, z.$$

Thus, every bounded linear functional is given by the inner product with a fixed vector.

We have already seen that $\varphi_y(x) = \langle y, x\rangle$ defines a bounded linear functional on $\mathcal{H}$ for every $y \in \mathcal{H}$. To prove that there is a unique y in $\mathcal{H}$ associated with a given linear functional, suppose that $\varphi_{y_1} = \varphi_{y_2}$. Then $\varphi_{y_1}(y) = \varphi_{y_2}(y)$ when $y = y_1 - y_2$, which implies that $\|y_1 - y_2\|^2 = 0$, so $y_1 = y_2$. □

The map $J : \mathcal{H} \to \mathcal{H}^*$ given by $Jy = \varphi_y$ therefore identifies a Hilbert space $\mathcal{H}$ with its dual space $\mathcal{H}^*$. The norm of φ_y is equal to the norm of y (see Exercise 8.7), so J is an isometry. In the case of complex Hilbert spaces, J is antilinear, rather than linear, because $\varphi_{\lambda y} = \overline{\lambda}\varphi_y$. Thus, Hilbert spaces are *self-dual*, meaning that $\mathcal{H}$ and $\mathcal{H}^*$ are isomorphic as Banach spaces, and anti-isomorphic as Hilbert spaces. Hilbert spaces are special in this respect. The dual space of an infinite-dimensional Banach space, such as an L^p-space with $p \neq 2$ or $C([a, b])$, is in general not isomorphic to the original space.

Example 8.13 In quantum mechanics, the *observables* of a system are represented by a space $\mathcal{A}$ of linear operators on a Hilbert space $\mathcal{H}$. A *state* ω of a quantum mechanical system is a linear functional ω on the space $\mathcal{A}$ of observables with the following two properties:

$$\omega(A^*A) \geq 0 \quad \text{for all } A \in \mathcal{A}, \tag{8.7}$$
$$\omega(I) = 1. \tag{8.8}$$

The number $\omega(A)$ is the expected value of the observable A when the system is in the state ω. Condition (8.7) is called positivity, and condition (8.8) is called normalization. To be specific, suppose that $\mathcal{H} = \mathbb{C}^n$ and $\mathcal{A}$ is the space of all $n \times n$ complex matrices. Then $\mathcal{A}$ is a Hilbert space with the inner product given by

$$\langle A, B\rangle = \operatorname{tr} A^*B.$$

By the Riesz representation theorem, for each state ω there is a unique $\rho \in \mathcal{A}$ such that

$$\omega(A) = \operatorname{tr} \rho^* A \quad \text{for all } A \in \mathcal{A}.$$

The conditions (8.7) and (8.8) translate into $\rho \geq 0$, and $\operatorname{tr} \rho = 1$, respectively.

Another application of the Riesz representation theorem is given in Section 12.11, where we use it to prove the existence and uniqueness of weak solutions of Laplace's equation.

8.3 The adjoint of an operator

An important consequence of the Riesz representation theorem is the existence of the *adjoint* of a bounded operator on a Hilbert space. The defining property of the adjoint $A^* \in \mathcal{B}(\mathcal{H})$ of an operator $A \in \mathcal{B}(\mathcal{H})$ is that

$$\langle x, Ay \rangle = \langle A^* x, y \rangle \qquad \text{for all } x, y \in \mathcal{H}. \tag{8.9}$$

The uniqueness of A^* follows from Exercise 8.14. The definition implies that

$$(A^*)^* = A, \qquad (AB)^* = B^* A^*.$$

To prove that A^* exists, we have to show that for every $x \in \mathcal{H}$, there is a vector $z \in \mathcal{H}$, depending linearly on x, such that

$$\langle z, y \rangle = \langle x, Ay \rangle \qquad \text{for all } y \in \mathcal{H}. \tag{8.10}$$

For fixed x, the map φ_x defined by

$$\varphi_x(y) = \langle x, Ay \rangle$$

is a bounded linear functional on $\mathcal{H}$, with $\|\varphi_x\| \leq \|A\| \|x\|$. By the Riesz representation theorem, there is a unique $z \in \mathcal{H}$ such that $\varphi_x(y) = \langle z, y \rangle$. This z satisfies (8.10), so we set $A^* x = z$. The linearity of A^* follows from the uniqueness in the Riesz representation theorem and the linearity of the inner product.

Example 8.14 The matrix of the adjoint of a linear map on $\mathbb{R}^n$ with matrix A is A^T, since

$$\mathbf{x} \cdot (A\mathbf{y}) = \left(A^T \mathbf{x}\right) \cdot \mathbf{y}.$$

In component notation, we have

$$\sum_{i=1}^n x_i \left(\sum_{j=1}^n a_{ij} y_j \right) = \sum_{j=1}^n \left(\sum_{i=1}^n a_{ij} x_i \right) y_j.$$

The matrix of the adjoint of a linear map on $\mathbb{C}^n$ with complex matrix A is the Hermitian conjugate matrix,

$$A^* = \overline{A^T}.$$

Example 8.15 Suppose that S and T are the right and left shift operators on the sequence space $\ell^2(\mathbb{N})$, defined by

$$S(x_1, x_2, x_3, \ldots) = (0, x_1, x_2, x_3, \ldots), \quad T(x_1, x_2, x_3, \ldots) = (x_2, x_3, x_4, \ldots).$$

Then $T = S^*$, since

$$\langle x, Sy \rangle = \overline{x_2} y_1 + \overline{x_3} y_2 + \overline{x_4} y_3 + \ldots = \langle Tx, y \rangle.$$

Example 8.16 Let $K : L^2([0,1]) \to L^2([0,1])$ be an integral operator of the form

$$Kf(x) = \int_0^1 k(x,y) f(y)\, dy,$$

where $k : [0,1] \times [0,1] \to \mathbb{C}$. Then the adjoint operator

$$K^* f(x) = \int_0^1 \overline{k(y,x)} f(y)\, dy$$

is the integral operator with the complex conjugate, transpose kernel.

The adjoint plays a crucial role in studying the solvability of a linear equation

$$Ax = y, \tag{8.11}$$

where $A : \mathcal{H} \to \mathcal{H}$ is a bounded linear operator. Let $z \in \mathcal{H}$ be any solution of the homogeneous adjoint equation,

$$A^* z = 0.$$

We take the inner product of (8.11) with z. The inner product on the left-hand side vanishes because

$$\langle Ax, z \rangle = \langle x, A^* z \rangle = 0.$$

Hence, a necessary condition for a solution x of (8.11) to exist is that $\langle y, z \rangle = 0$ for all $z \in \ker A^*$, meaning that $y \in (\ker A^*)^\perp$. This condition on y is not always sufficient to guarantee the solvability of (8.11); the most we can say for general bounded operators is the following result.

Theorem 8.17 If $A : \mathcal{H} \to \mathcal{H}$ is a bounded linear operator, then

$$\overline{\operatorname{ran} A} = (\ker A^*)^\perp, \qquad \ker A = (\operatorname{ran} A^*)^\perp. \tag{8.12}$$

Proof. If $x \in \operatorname{ran} A$, there is a $y \in \mathcal{H}$ such that $x = Ay$. For any $z \in \ker A^*$, we then have

$$\langle x, z \rangle = \langle Ay, z \rangle = \langle y, A^* z \rangle = 0.$$

This proves that $\operatorname{ran} A \subset (\ker A^*)^\perp$. Since $(\ker A^*)^\perp$ is closed, it follows that $\overline{\operatorname{ran} A} \subset (\ker A^*)^\perp$. On the other hand, if $x \in (\operatorname{ran} A)^\perp$, then for all $y \in \mathcal{H}$ we have

$$0 = \langle Ay, x \rangle = \langle y, A^* x \rangle.$$

Therefore $A^*x = 0$. This means that $(\operatorname{ran} A)^\perp \subset \ker A^*$. By taking the orthogonal complement of this relation, we get

$$(\ker A^*)^\perp \subset (\operatorname{ran} A)^{\perp\perp} = \overline{\operatorname{ran} A},$$

which proves the first part of (8.12). To prove the second part, we apply the first part to A^*, instead of A, use $A^{**} = A$, and take orthogonal complements. □

An equivalent formulation of this theorem is that if A is a bounded linear operator on $\mathcal{H}$, then $\mathcal{H}$ is the orthogonal direct sum

$$\mathcal{H} = \overline{\operatorname{ran} A} \oplus \ker A^*.$$

If A has closed range, then we obtain the following necessary and sufficient condition for the solvability of (8.11).

Theorem 8.18 Suppose that $A : \mathcal{H} \to \mathcal{H}$ is a bounded linear operator on a Hilbert space $\mathcal{H}$ with closed range. Then the equation $Ax = y$ has a solution for x if and only if y is orthogonal to $\ker A^*$.

This theorem provides a useful general method of proving existence from uniqueness: if A has closed range, and the solution of the adjoint problem $A^*x = y$ is unique, then $\ker A^* = \{0\}$, so every y is orthogonal to $\ker A^*$. Hence, a solution of $Ax = y$ exists for every $y \in \mathcal{H}$. The condition that A has closed range is implied by an estimate of the form $c\|x\| \le \|Ax\|$, as shown in Proposition 5.30.

A commonly occurring dichotomy for the solvability of a linear equation is summarized in the following Fredholm alternative.

Definition 8.19 A bounded linear operator $A : \mathcal{H} \to \mathcal{H}$ on a Hilbert space $\mathcal{H}$ satisfies the *Fredholm alternative* if one of the following two alternatives holds:

(a) either $Ax = 0$, $A^*x = 0$ have only the zero solution, and the equations $Ax = y$, $A^*x = y$ have a unique solution $x \in \mathcal{H}$ for every $y \in \mathcal{H}$;
(b) or $Ax = 0$, $A^*x = 0$ have nontrivial, finite-dimensional solution spaces of the same dimension, $Ax = y$ has a (nonunique) solution if and only if $y \perp z$ for every solution z of $A^*z = 0$, and $A^*x = y$ has a (nonunique) solution if and only if $y \perp z$ for every solution z of $Az = 0$.

Any linear operator $A : \mathbb{C}^n \to \mathbb{C}^n$ on a finite-dimensional space, associated with an $n \times n$ system of linear equations $Ax = y$, satisfies the Fredholm alternative. The ranges of A and A^* are closed because they are finite-dimensional. From linear algebra, the rank of A^* is equal to the rank of A, and therefore the nullity of A is equal to the nullity of A^*. The Fredholm alternative then follows from Theorem 8.18.

Two things can go wrong with the Fredholm alternative in Definition 8.19 for bounded operators A on an infinite-dimensional space. First, $\operatorname{ran} A$ need not be closed; and second, even if $\operatorname{ran} A$ is closed, it is not true, in general, that $\ker A$ and $\ker A^*$ have the same dimension. As a result, the equation $Ax = y$ may be solvable for all $y \in \mathcal{H}$ even though A is not one-to-one, or $Ax = y$ may not be solvable for all $y \in \mathcal{H}$ even though A is one-to-one. We illustrate these possibilities with some examples.

Example 8.20 Consider the multiplication operator $M : L^2([0,1]) \to L^2([0,1])$ defined by

$$Mf(x) = xf(x).$$

Then $M^* = M$, and M is one-to-one, so every $g \in L^2([0,1])$ is orthogonal to $\ker M^*$; but the range of M is a proper dense subspace of $L^2([0,1])$, so $Mf = g$ is not solvable for every $g \in L^2([0,1])$ (see Example 9.5 for more details).

Example 8.21 The range of the right shift operator $S : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N})$, defined in Example 8.15, is closed since it consists of $y = (y_1, y_2, y_3, \ldots) \in \ell^2(\mathbb{N})$ such that $y_1 = 0$. The left shift operator $T = S^*$ is singular since its kernel is the one-dimensional space with basis $\{(1, 0, 0, \ldots)\}$. The equation $Sx = y$, or

$$(0, x_1, x_2, \ldots) = (y_1, y_2, y_3, \ldots),$$

is solvable if and only if $y_1 = 0$, or $y \perp \ker T$, which verifies Theorem 8.18 in this case. If a solution exists, then it is unique. On the other hand, the equation $Tx = y$ is solvable for every $y \in \ell^2(\mathbb{N})$, even though T is not one-to-one, and the solution is not unique.

These examples motivate the following definition.

Definition 8.22 A bounded linear operator A on a Hilbert space is a *Fredholm operator* if:

(a) $\operatorname{ran} A$ is closed;
(b) $\ker A$ and $\ker A^*$ are finite-dimensional.

The *index* of a Fredholm operator A, $\operatorname{ind} A$, is the integer

$$\operatorname{ind} A = \dim \ker A - \dim \ker A^*.$$

For example, a linear operator on a finite-dimensional Hilbert space and the identity operator on an infinite-dimensional Hilbert space are Fredholm operators with index zero. The right and left shift operators S and T in Example 8.21 are Fredholm, but their indices are nonzero. Since $\dim \ker S = 0$, $\dim \ker T = 1$, and

$S = T^*$, we have $\operatorname{ind} S = -1$ and $\operatorname{ind} T = 1$. The multiplication operator in Example 8.20 is not Fredholm because it does not have closed range.

It is possible to prove that if A is Fredholm and K is compact, then $A + K$ is Fredholm, and $\operatorname{ind}(A + K) = \operatorname{ind} A$. Thus the index of a Fredholm operator is unchanged by compact perturbations. In particular, compact perturbations of the identity are Fredholm operators with index zero, so they satisfy the Fredholm alternative in Definition 8.19. We will prove a special case of this result, for compact, self-adjoint perturbations of the identity, in Theorem 9.26.

8.4 Self-adjoint and unitary operators

Two of the most important classes of operators on a Hilbert space are the classes of self-adjoint and unitary operators. We begin by defining self-adjoint operators.

Definition 8.23 A bounded linear operator $A : \mathcal{H} \to \mathcal{H}$ on a Hilbert space $\mathcal{H}$ is *self-adjoint* if $A^* = A$.

Equivalently, a bounded linear operator A on $\mathcal{H}$ is self-adjoint if and only if

$$\langle x, Ay\rangle = \langle Ax, y\rangle \qquad \text{for all } x, y \in \mathcal{H}.$$

Example 8.24 From Example 8.14, a linear map on $\mathbb{R}^n$ with matrix A is self-adjoint if and only if A is *symmetric*, meaning that $A = A^T$, where A^T is the transpose of A. A linear map on $\mathbb{C}^n$ with matrix A is self-adjoint if and only if A is *Hermitian*, meaning that $A = A^*$.

Example 8.25 From Example 8.16, an integral operator $K : L^2([0,1]) \to L^2([0,1])$,

$$Kf(x) = \int_0^1 k(x,y) f(y)\, dy,$$

is self-adjoint if and only if $k(x,y) = \overline{k(y,x)}$.

Given a linear operator $A : \mathcal{H} \to \mathcal{H}$, we may define a sesquilinear form

$$a : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$$

by $a(x,y) = \langle x, Ay\rangle$. If A is self-adjoint, then this form is *Hermitian symmetric*, or *symmetric*, meaning that

$$a(x,y) = \overline{a(y,x)}.$$

It follows that the associated quadratic form $q(x) = a(x,x)$, or

$$q(x) = \langle x, Ax\rangle, \tag{8.13}$$

is real-valued. We say that A is *nonnegative* if it is self-adjoint and $\langle x, Ax\rangle \geq 0$ for all $x \in \mathcal{H}$. We say that A is *positive*, or *positive definite*, if it is self-adjoint and $\langle x, Ax\rangle > 0$ for every nonzero $x \in \mathcal{H}$. If A is a positive, bounded operator, then

$$(x, y) = \langle x, Ay\rangle$$

defines an inner product on $\mathcal{H}$. If, in addition, there is a constant $c > 0$ such that

$$\langle x, Ax\rangle \geq c\|x\|^2 \qquad \text{for all } x \in \mathcal{H},$$

then we say that A is *bounded from below*, and the norm associated with $(\cdot, \cdot)$ is equivalent to the norm associated with $\langle \cdot, \cdot\rangle$.

The quadratic form associated with a self-adjoint operator determines the norm of the operator.

Lemma 8.26 If A is a bounded self-adjoint operator on a Hilbert space $\mathcal{H}$, then

$$\|A\| = \sup_{\|x\|=1} |\langle x, Ax\rangle|.$$

Proof. Let

$$\alpha = \sup_{\|x\|=1} |\langle x, Ax\rangle|.$$

The inequality $\alpha \leq \|A\|$ is immediate, since

$$|\langle x, Ax\rangle| \leq \|Ax\|\,\|x\| \leq \|A\|\,\|x\|^2.$$

To prove the reverse inequality, we use the definition of the norm,

$$\|A\| = \sup_{\|x\|=1} \|Ax\|.$$

For any $z \in \mathcal{H}$, we have

$$\|z\| = \sup_{\|y\|=1} |\langle y, z\rangle|.$$

It follows that

$$\|A\| = \sup\left\{|\langle y, Ax\rangle| \mid \|x\| = 1, \|y\| = 1\right\}. \tag{8.14}$$

The polarization formula (6.5) implies that

$$\begin{aligned}\langle y, Ax\rangle \quad = \quad & \frac{1}{4}\left\{\langle x + y, A(x + y)\rangle - \langle x - y, A(x - y)\rangle\right.\\ & \left. - \, i\langle x + iy, A(x + iy)\rangle + i\langle x - iy, A(x - iy)\rangle\right\}.\end{aligned}$$

Since A is self-adjoint, the first two terms are real, and the last two are imaginary. We replace y by $e^{i\varphi}y$, where $\varphi \in \mathbb{R}$ is chosen so that $\langle e^{i\varphi}y, Ax\rangle$ is real. Then the imaginary terms vanish, and we find that

$$\begin{aligned} |\langle y, Ax\rangle|^2 &= \frac{1}{16}|\langle x+y, A(x+y)\rangle - \langle x-y, A(x-y)\rangle|^2 \\ &\leq \frac{1}{16}\alpha^2(\|x+y\|^2 + \|x-y\|^2)^2 \\ &= \frac{1}{4}\alpha^2(\|x\|^2 + \|y\|^2)^2, \end{aligned}$$

where we have used the definition of α and the parallelogram law. Using this result in (8.14), we conclude that $\|A\| \leq \alpha$. □

As a corollary, we have the following result.

Corollary 8.27 If A is a bounded operator on a Hilbert space then $\|A^*A\| = \|A\|^2$. If A is self-adjoint, then $\|A^2\| = \|A\|^2$.

Proof. The definition of $\|A\|$, and an application Lemma 8.26 to the self-adjoint operator A^*A, imply that

$$\|A\|^2 = \sup_{\|x\|=1} |\langle Ax, Ax\rangle| = \sup_{\|x\|=1} |\langle x, A^*Ax\rangle| = \|A^*A\|.$$

Hence, if A is self-adjoint, then $\|A\|^2 = \|A^2\|$. □

Next, we define orthogonal or unitary operators, on real or complex spaces, respectively.

Definition 8.28 A linear map $U : \mathcal{H}_1 \to \mathcal{H}_2$ between real or complex Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ is said to be *orthogonal* or *unitary*, respectively, if it is invertible and if

$$\langle Ux, Uy\rangle_{\mathcal{H}_2} = \langle x, y\rangle_{\mathcal{H}_1} \qquad \text{for all } x, y \in \mathcal{H}_1.$$

Two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ are *isomorphic* as Hilbert spaces if there is a unitary linear map between them.

Thus, a unitary operator is one-to-one and onto, and preserves the inner product. A map $U : \mathcal{H} \to \mathcal{H}$ is unitary if and only if $U^*U = UU^* = I$.

Example 8.29 An $n \times n$ real matrix Q is orthogonal if $Q^T = Q^{-1}$. An $n \times n$ complex matrix U is unitary if $U^* = U^{-1}$.

Example 8.30 If A is a bounded self-adjoint operator, then

$$e^{iA} = \sum_{n=0}^{\infty} \frac{1}{n!}(iA)^n$$

is unitary, since

$$\left(e^{iA}\right)^* = e^{-iA} = \left(e^{iA}\right)^{-1}.$$

A bounded operator S is *skew-adjoint* if $S^* = -S$. Any skew-adjoint operator S on a complex Hilbert space may be written as $S = iA$ where A is a self-adjoint operator. The commutator $[A, B] = AB - BA$ is a Lie bracket on the space of bounded, skew-adjoint operators, and we say that this space is the Lie algebra of the Lie group of unitary operators.

Example 8.31 Let $\mathcal{H}$ be a finite dimensional Hilbert space. If $\{e_1, e_2, \ldots, e_n\}$ is an orthonormal basis of $\mathcal{H}$, then $U : \mathbb{C}^n \to \mathcal{H}$ defined by

$$U(z_1, z_2, \ldots, z_n) = z_1 e_1 + z_2 e_2 + \ldots + z_n e_n$$

is unitary. Thus, any n-dimensional, complex Hilbert space is isomorphic to $\mathbb{C}^n$.

Example 8.32 Suppose that $\mathcal{H}_1$ and $\mathcal{H}_2$ are two Hilbert spaces of the same, possibly infinite, dimension. Let $\{u_\alpha\}$ be an orthonormal basis of $\mathcal{H}_1$ and $\{v_\alpha\}$ an orthonormal basis of $\mathcal{H}_2$. Any $x \in \mathcal{H}_1$ can be written uniquely as

$$x = \sum_\alpha c_\alpha u_\alpha,$$

with coefficients $c_\alpha \in \mathbb{C}$. We define $U : \mathcal{H}_1 \to \mathcal{H}_2$ by

$$U\left(\sum_\alpha c_\alpha u_\alpha\right) = \sum_\alpha c_\alpha v_\alpha.$$

Then U is unitary. Thus, Hilbert spaces of the same dimension are isomorphic.

More generally, if $\lambda_\alpha = e^{i\varphi_\alpha}$ are complex numbers with $|\lambda_\alpha| = 1$, then $U : \mathcal{H}_1 \to \mathcal{H}_2$ defined by

$$Ux = \sum_\alpha \lambda_\alpha \langle u_\alpha, x\rangle v_\alpha$$

is unitary. For example, the periodic *Hilbert transform* $\mathbb{H} : L^2(\mathbb{T}) \to L^2(\mathbb{T})$ is defined by

$$\mathbb{H}\left(\sum_{n=-\infty}^{\infty} \hat{f}_n e^{inx}\right) = \sum_{n=-\infty}^{\infty} i\,(\operatorname{sgn} n)\,\hat{f}_n e^{inx},$$

where sgn is the sign function, defined in (5.8). The Hilbert transform is not a unitary mapping on $L^2(\mathbb{T})$ because $\mathbb{H}(1) = 0$; however, Parseval's theorem implies that it is a unitary mapping on the subspace $\mathcal{H}$ of square-integrable periodic functions with zero mean,

$$\mathcal{H} = \left\{ f \in L^2(\mathbb{T}) \;\middle|\; \int_{\mathbb{T}} f(x)\,dx = 0 \right\}.$$

Example 8.33 The operator $U : L^2(\mathbb{T}) \to \ell^2(\mathbb{Z})$ that maps a function to its Fourier coefficents is unitary. Explicitly, we have

$$Uf = (c_n)_{n\in\mathbb{Z}}, \qquad c_n = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x) e^{-inx}\, dx.$$

Thus, the Hilbert space of square integrable functions on the circle is isomorphic to the Hilbert space of sequences on $\mathbb{Z}$. As this example illustrates, isomorphic Hilbert spaces may be given concretely in forms that, at first sight, do not appear to be the same.

Example 8.34 For $a \in \mathbb{T}$, we define the translation operator $T_a : L^2(\mathbb{T}) \to L^2(\mathbb{T})$ by

$$(T_a f)(x) = f(x - a).$$

Then T_a is unitary, and

$$T_{a+b} = T_a T_b.$$

We say that $\{T_a \mid a \in \mathbb{T}\}$ is a unitary representation of the additive group $\mathbb{R}/(2\pi\mathbb{Z})$ on the linear space $L^2(\mathbb{T})$.

An operator $T : \mathcal{H} \to \mathcal{H}$ is said to be *normal* if it commutes with its adjoint, meaning that $TT^* = T^*T$. Both self-adjoint and unitary operators are normal. An important feature of normal operators is that they have a nice spectral theory. We will discuss the spectral theory of compact, self-adjoint operators in detail in the next chapter.

8.5 The mean ergodic theorem

Ergodic theorems equate time averages with probabilistic averages, and they are important, for example, in understanding the statistical behavior of deterministic dynamical systems.

The proof of the following ergodic theorem, due to von Neumann, is a good example of Hilbert space methods.

Theorem 8.35 (von Neumann ergodic) Suppose that U is a unitary operator on a Hilbert space $\mathcal{H}$. Let $\mathcal{M} = \{x \in \mathcal{H} \mid Ux = x\}$ be the subspace of vectors that are invariant under U, and P the orthogonal projection onto $\mathcal{M}$. Then, for all $x \in \mathcal{H}$, we have

$$\lim_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} U^n x = Px. \tag{8.15}$$

That is, the averages of U^n converge strongly to P.

Proof. It is sufficient to prove (8.15) for $x \in \ker P$ and $x \in \operatorname{ran} P$, because then the orthogonal decomposition $\mathcal{H} = \ker P \oplus \operatorname{ran} P$ implies that (8.15) holds for all $x \in \mathcal{H}$. Equation (8.15) is trivial when $x \in \operatorname{ran} P = \mathcal{M}$, since then $Ux = x$ and $Px = x$.

To complete the proof, we show that (8.15) holds when $x \in \ker P$. From the definition of P, we have $\operatorname{ran} P = \ker(I - U)$. If U is unitary, then $Ux = x$ if and only if $U^*x = x$. Hence, using Theorem 8.17, we find that

$$\ker P = \ker(I - U)^{\perp} = \ker(I - U^*)^{\perp} = \overline{\operatorname{ran}(I - U)}.$$

Therefore every $x \in \ker P$ may be approximated by vectors of the form $(I - U)y$. If $x = (I - U)y$, then

$$\begin{aligned} \frac{1}{N+1}\sum_{n=0}^{N} U^n x &= \frac{1}{N+1}\sum_{n=0}^{N}\left(U^n - U^{n+1}\right) y \\ &= \frac{1}{N+1}\left(y - U^{N+1}y\right) \\ &\to 0 \qquad \text{as } N \to \infty. \end{aligned}$$

If $x \in \ker P$, then there is a sequence of elements $x_k = (I - U)y_k$ with $x_k \to x$. Hence,

$$\begin{aligned} \lim_{N\to\infty}\left\|\frac{1}{N+1}\sum_{n=0}^{N} U^n x\right\| &\leq \limsup_{N\to\infty}\left\|\frac{1}{N+1}\sum_{n=0}^{N} U^n (x - x_k)\right\| \\ &\quad + \limsup_{N\to\infty}\left\|\frac{1}{N+1}\sum_{n=0}^{N} U^n x_k\right\| \\ &\leq \|x - x_k\|. \end{aligned}$$

Since k is arbitrary and $x_k \to x$, it follows that (8.15) holds for every $x \in \ker P$. □

Next, we explain the implications of this result in probability theory. Suppose that P is a probability measure on a probability space Ω, as described in Section 6.4. A one-to-one, onto, measurable map $T : \Omega \to \Omega$ is said to be *measure preserving* if $P\left(T^{-1}(A)\right) = P(A)$ for all measurable subsets A of Ω. Here,

$$T^{-1}(A) = \{\omega \in \Omega \mid T(\omega) \in A\}.$$

The rotations of the unit circle, studied in Theorem 7.11, are a helpful example to keep in mind here. In that case, $\Omega = \mathbb{T}$, and P is the measure which assigns a probability of $\theta/2\pi$ to an interval on $\mathbb{T}$ of length θ. Any rotation of the circle is a measure preserving map.

If f is a random variable (that is, a measurable real- or complex-valued function on Ω) then the composition of T and f, defined by $f \circ T(\omega) = f\left(T(\omega)\right)$, is also a random variable. Since T is measure preserving, we have $\mathbb{E}f = \mathbb{E}f \circ T$, or

$$\int_\Omega f\, dP = \int_\Omega f \circ T\, dP.$$

If $f = f \circ T$, then we say that f is invariant under T. This is always true if f is a constant function. If these are the only invariant functions, then we say that T is ergodic.

Definition 8.36 A one-to-one, onto, measure preserving map T on a probability space (Ω, P) is *ergodic* if the only functions $f \in L^2(\Omega, P)$ such that $f = f \circ T$ are the constant functions.

For example, rotations of the circle through an irrational multiple of 2π are ergodic, but rotations through a rational multiple of 2π are not. To make the connection between ergodic maps and Theorem 8.35 above, we define an operator

$$U : L^2(\Omega, P) \to L^2(\Omega, P)$$

on the Hilbert space $L^2(\Omega, P)$ of second-order random variables on Ω by

$$Uf = f \circ T. \tag{8.16}$$

Suppose that $f, g \in L^2(\Omega, P)$. Then, since T is measure preserving, we have

$$\langle Uf, Ug \rangle = \int_\Omega \overline{f\left(T(\omega)\right)} g\left(T(\omega)\right)\, dP(\omega) = \int_\Omega \overline{f(\omega)} g(\omega)\, dP(\omega) = \langle f, g \rangle,$$

so the map U is unitary. The subspace of functions invariant under U consists of the functions that are invariant under T. Thus, if T is ergodic, the invariant subspace of U consists of the constant functions, and the orthogonal projection onto the invariant subspace maps a random variable to its expected value. An application of the von Neumann ergodic theorem to the map U defined in (8.16) then gives the following result.

Theorem 8.37 A one-to-one, onto, measure preserving map $T : \Omega \to \Omega$ on a probability space (Ω, P) is ergodic if and only if for every $f \in L^2(\Omega, P)$

$$\lim_{N \to \infty} \frac{1}{N+1} \sum_{n=0}^{N} f \circ T^n = \int_\Omega f dP, \tag{8.17}$$

where the convergence is in the L^2-norm.

If we think of $T : \Omega \to \Omega$ as defining a discrete dynamical system $x_{n+1} = Tx_n$ on the state space Ω, as described in Section 3.2, then the left-hand side of (8.17) is the time average of f, while the right-hand side is the probabilistic (or "ensemble") average of f. Thus, the theorem states that time averages and probabilistic averages coincide for ergodic maps.

There is a second ergodic theorem, called the *Birkhoff ergodic theorem*, which states that the averages on the left-hand side of equation (8.17) converge almost surely to the constant on the right-hand side for every f in $L^1(\Omega, P)$.

8.6 Weak convergence in a Hilbert space

A sequence (x_n) in a Hilbert space $\mathcal{H}$ converges *weakly* to $x \in \mathcal{H}$ if

$$\lim_{n\to\infty} \langle x_n, y\rangle = \langle x, y\rangle \qquad \text{for all } y \in \mathcal{H}.$$

Weak convergence is usually written as

$$x_n \rightharpoonup x \qquad \text{as } n \to \infty,$$

to distinguish it from strong, or norm, convergence. From the Riesz representation theorem, this definition of weak convergence for sequences in a Hilbert space is a special case of Definition 5.59 of weak convergence in a Banach space. Strong convergence implies weak convergence, but the converse is not true on infinite-dimensional spaces.

Example 8.38 Suppose that $\mathcal{H} = \ell^2(\mathbb{N})$. Let

$$e_n = (0, 0, \ldots, 0, 1, 0, \ldots)$$

be the standard basis vector whose nth term is 1 and whose other terms are 0. If $y = (y_1, y_2, y_3, \ldots) \in \ell^2(\mathbb{N})$, then

$$\langle e_n, y\rangle = y_n \to 0 \qquad \text{as } n \to \infty,$$

since $\sum |y_n|^2$ converges. Hence $e_n \rightharpoonup 0$ as $n \to \infty$. On the other hand, $\|e_n - e_m\| = \sqrt{2}$ for all $n \neq m$, so the sequence (e_n) does not converge strongly.

It is a nontrivial fact that a weakly convergent sequence is bounded. This is a consequence of the uniform boundedness theorem, or Banach-Steinhaus theorem, which we prove next.

Theorem 8.39 (Uniform boundedness) Suppose that

$$\{\varphi_n : X \to \mathbb{C} \mid n \in \mathbb{N}\}$$

is a set of linear functionals on a Banach space X such that the set of complex numbers $\{\varphi_n(x) \mid n \in \mathbb{N}\}$ is bounded for each $x \in X$. Then $\{\|\varphi_n\| \mid n \in \mathbb{N}\}$ is bounded.

Proof. First, we show that the functionals are uniformly bounded if they are uniformly bounded on any ball. Suppose that there is a ball

$$B(x_0, r) = \{x \in X \mid \|x - x_0\| < r\},$$

with $r > 0$, and a constant M such that

$$|\varphi_n(x)| \leq M \qquad \text{for all } x \in B(x_0, r) \text{ and all } n \in \mathbb{N}.$$

Then, for any $x \in X$ with $x \neq x_0$, the linearity of φ_n implies that

$$|\varphi_n(x)| \leq \frac{\|x - x_0\|}{r} \left| \varphi_n \left(r \frac{x - x_0}{\|x - x_0\|} \right) \right| + |\varphi_n(x_0)| \leq \frac{M}{r} \|x - x_0\| + |\varphi_n(x_0)|.$$

Hence, if $\|x\| \leq 1$, we have

$$|\varphi_n(x)| \leq \frac{M}{r} (1 + \|x_0\|) + |\varphi_n(x_0)|.$$

Thus, the set of norms $\{\|\varphi_n\| \mid n \in \mathbb{N}\}$ is bounded, because $\{|\varphi_n(x_0)| \mid n \in \mathbb{N}\}$ is bounded.

We now assume for contradiction that $\{\|\varphi_n\|\}$ is unbounded. It follows from what we have just shown that for every open ball $B(x_0, r)$ in X with $r > 0$, the set

$$\{|\varphi_n(x)| \mid x \in B(x_0, r) \text{ and } n \in \mathbb{N}\}$$

is unbounded. We may therefore pick $n_1 \in \mathbb{N}$ and $x_1 \in B(x_0, r)$ such that $|\varphi_{n_1}(x_1)| > 1$. By the continuity of φ_{n_1}, there is an $0 < r_1 < 1$ such that $|\varphi_{n_1}(x)| > 1$ for all $x \in B(x_1, r_1)$. Next, we pick $n_2 > n_1$ and $x_2 \in B(x_1, r_1)$ such that $|\varphi_{n_2}(x_2)| > 2$. We choose a sufficiently small $0 < r_2 < 1/2$ such that $B(x_2, r_2)$ is contained in $B(x_1, r_1)$ and $|\varphi_{n_2}(x)| > 2$ for all $x \in B(x_2, r_2)$. Continuing in this way, we obtain a subsequence (φ_{n_k}) of linear functionals, and a nested sequence of balls $B(x_k, r_k)$ such that $0 < r_k < 1/k$ and

$$|\varphi_{n_k}(x)| > k \qquad \text{for all } x \in B(x_k, r_k).$$

The sequence (x_k) is Cauchy, and hence $x_k \to \overline{x}$ since X is complete. But $\overline{x} \in B(x_k, r_k)$ for all $k \in \mathbb{N}$ so that $|\varphi_{n_k}(\overline{x})| \to \infty$ as $k \to \infty$, which contradicts the pointwise boundedness of $\{\varphi_n(\overline{x})\}$. $\square$

Thus, the boundedness of the pointwise values of a family of linear functional implies the boundedness of their norms. Next, we prove that a weakly convergent sequence is bounded, and give a useful necessary and sufficient condition for weak convergence.

Theorem 8.40 Suppose that (x_n) is a sequence in a Hilbert space $\mathcal{H}$ and D is a dense subset of $\mathcal{H}$. Then (x_n) converges weakly to x if and only if:

(a) $\|x_n\| \leq M$ for some constant M;
(b) $\langle x_n, y\rangle \to \langle x, y\rangle$ as $n \to \infty$ for all $y \in D$.

Proof. Suppose that (x_n) is a weakly convergent sequence. We define the bounded linear functionals φ_n by $\varphi_n(x) = \langle x_n, x\rangle$. Then $\|\varphi_n\| = \|x_n\|$. Since $(\varphi_n(x))$ converges for each $x \in \mathcal{H}$, it is a bounded sequence, and the uniform boundedness theorem implies that $\{\|\varphi_n\|\}$ is bounded. It follows that a weakly convergent sequence satisfies (a). Part (b) is trivial.

Conversely, suppose that (x_n) satisfies (a) and (b). If $z \in \mathcal{H}$, then for any $\epsilon > 0$ there is a $y \in D$ such that $\|z - y\| < \epsilon$, and there is an N such that $|\langle x_n - x, y\rangle| < \epsilon$ for $n \geq N$. Since $\|x_n\| \leq M$, it follows from the Cauchy-Schwarz inequality that for $n \geq N$

$$\begin{aligned} |\langle x_n - x, z\rangle| &\leq |\langle x_n - x, y\rangle| + |\langle x_n - x, z - y\rangle| \\ &\leq \epsilon + \|x_n - x\|\|z - y\| \\ &\leq (1 + M + \|x\|)\,\epsilon. \end{aligned}$$

Thus, $\langle x_n - x, z\rangle \to 0$ as $n \to \infty$ for every $z \in \mathcal{H}$, so $x_n \rightharpoonup x$. □

Example 8.41 Suppose that $\{e_\alpha\}_{\alpha\in I}$ is an orthonormal basis of a Hilbert space. Then a sequence (x_n) converges weakly to x if and only if it is bounded and its coordinates converge, meaning that $\langle x_n, e_\alpha\rangle \to \langle x, e_\alpha\rangle$ for each $\alpha \in I$.

The boundedness of the sequence is essential to ensure weak convergence, as the following example shows.

Example 8.42 In Example 8.38, we saw that the bounded sequence (e_n) of standard basis elements in $\ell^2(\mathbb{N})$ converges weakly to zero. The unbounded sequence (ne_n), where

$$ne_n = (0, 0, \ldots, 0, n, 0, \ldots),$$

does not converge weakly, however, even though the coordinate sequences with respect to the basis (e_n) converge to zero. For example,

$$x = \left(n^{-3/4}\right)_{n=1}^{\infty}$$

belongs to $\ell^2(\mathbb{N})$, but $\langle ne_n, x\rangle = n^{1/4}$ does not converge as $n \to \infty$.

The next example illustrates *oscillation*, *concentration*, and *escape to infinity*, which are typical ways that a weakly convergent sequence of functions fails to converge strongly.

Example 8.43 The sequence $(\sin n\pi x)$ converges weakly to zero in $L^2([0, 1])$ because

$$\int_0^1 f(x) \sin n\pi x\, dx \to 0 \qquad \text{as } n \to \infty$$

for all $f \in L^2([0, 1])$ (see Example 5.47). The sequence cannot converge strongly to zero since $\|\sin n\pi x\| = 1/\sqrt{2}$ is bounded away from 0. In this case, the functions oscillate more and more rapidly as $n \to \infty$. If a function

$$f(x) = \sum_{n=1}^{\infty} a_n \sin n\pi x$$

in $L^2([0,1])$ is represented by its sequence (a_n) of Fourier sine coefficients, then this example is exactly the same as Example 8.38.

The sequence (f_n) defined by

$$f_n(x) = \begin{cases} \sqrt{n} & \text{if } 0 \leq x \leq 1/n, \\ 0 & \text{if } 1/n \leq x \leq 1, \end{cases}$$

converges weakly to zero in $L^2([0,1])$. To prove this fact, we observe that, for any polynomial p,

$$\begin{aligned} \left| \int_0^1 p(x) f_n(x)\, dx \right| &= \sqrt{n} \left| \int_0^{1/n} p(x)\, dx \right| \\ &\leq \frac{1}{\sqrt{n}} \left| n \int_0^{1/n} p(x)\, dx \right| \\ &\to 0 \qquad \text{as } n \to \infty \end{aligned}$$

since, by the continuity of p,

$$n \int_0^{1/n} p(x)\, dx = p(0) + n \int_0^{1/n} \{p(x) - p(0)\}\, dx \to p(0) \qquad \text{as } n \to \infty.$$

Thus, $\langle p, f_n \rangle \to 0$ as $n \to \infty$ for every polynomial p. Since the polynomials are dense in $L^2([0,1])$ and $\|f_n\| = 1$ for all n, Theorem 8.40 implies that $f_n \rightharpoonup 0$. The norms of the f_n are bounded away from 0, so they cannot converge strongly to zero. In this case the functions f_n have a singularity that concentrates at a point.

The sequence (f_n) defined by

$$f_n(x) = \begin{cases} 1 & \text{if } n < x < n+1, \\ 0 & \text{otherwise,} \end{cases}$$

converges weakly, but not strongly, to zero in $L^2(\mathbb{R})$. In this case, the functions f_n escape to infinity. The proof follows from the density of functions with compact support in $L^2(\mathbb{R})$.

As the above examples show, the norm of the limit of a weakly convergent sequence may be strictly less than the norms of the terms in the sequence, corresponding to a loss of "energy" in oscillations, at a singularity, or by escape to infinity in the weak limit. In each case, the expansion of f_n in any orthonormal basis contains coefficients that wander off to infinity. If the norms of a weakly convergent sequence converge to the norm of the weak limit, then the sequence converges strongly.

Proposition 8.44 If (x_n) converges weakly to x, then

$$\|x\| \leq \liminf_{n\to\infty} \|x_n\|. \tag{8.18}$$

If, in addition,

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

then (x_n) converges strongly to x.

Proof. Using the weak convergence of (x_n) and the Cauchy-Schwarz inequality, we find that

$$\|x\|^2 = \langle x, x\rangle = \lim_{n\to\infty} \langle x, x_n\rangle \leq \|x\| \liminf_{n\to\infty} \|x_n\|,$$

which proves (8.18). Expansion of the inner product gives

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

If $x_n \rightharpoonup x$, then $\langle x_n, x\rangle \to \langle x, x\rangle$. Hence, if we also have $\|x_n\| \to \|x\|$, then $\|x_n - x\|^2 \to 0$, meaning that $x_n \to x$ strongly. □

One reason for the utility of weak convergence is that it is much easier for sets to be compact in the weak topology than in the strong topology; in fact, a set is weakly precompact if and only if it is bounded. This result provides a partial analog of the Heine-Borel theorem for infinite-dimensional spaces, and is illustrated by the orthonormal sequence of vectors in Example 8.38. The sequence converges weakly, but no subsequence converges strongly, so the terms of the sequence form a weakly precompact, but not a strongly precompact, set.

Theorem 8.45 (Banach-Alaoglu) The closed unit ball of a Hilbert space is weakly compact.

Proof. We will prove the result for a separable Hilbert space. The result remains true for nonseparable spaces, but the proof requires deeper topological arguments and we will not give it here. We will use a diagonal argument to show that any sequence in the unit ball of a separable, infinite-dimensional Hilbert space has a convergent subsequence. Sequential weak compactness implies weak compactness, although this fact is not obvious because the weak topology is not metrizable.

Suppose that (x_n) is a sequence in the unit ball of a Hilbert space $\mathcal{H}$. Let $D = \{y_n \mid n \in \mathbb{N}\}$ be a dense subset of $\mathcal{H}$. Then $(\langle x_n, y_1\rangle)$ is a bounded sequence in $\mathbb{C}$, since

$$|\langle x_n, y_1\rangle| \leq \|x_n\|\|y_1\| \leq \|y_1\|.$$

By the Heine-Borel theorem, there is a subsequence of (x_n), which we denote by $(x_{1,k})$, such that $(\langle x_{1,k}, y_1\rangle)$ converges as $k \to \infty$. In a similar way, there is a subsequence $(x_{2,k})$ of $(x_{1,k})$ such that $(\langle x_{2,k}, y_2\rangle)$ converges. Continuing in this way, we obtain successive subsequences $(x_{j,k})$ such that $(\langle x_{j,k}, y_i\rangle)$ converges as $k \to \infty$ for each $1 \le i \le j$. Taking the diagonal subsequence $(x_{k,k})$ of (x_n), we see that $(\langle x_{k,k}, y\rangle)$ converges as $k \to \infty$ for every $y \in D$. We define the linear functional $\varphi : D \subset \mathcal{H} \to \mathbb{C}$ by

$$\varphi(y) = \lim_{k\to\infty} \langle x_{k,k}, y\rangle.$$

Then $|\varphi(y)| \le \|y\|$ since $\|x_{k,k}\| \le 1$, so φ is bounded on D. It therefore has a unique extension to a bounded linear functional on $\mathcal{H}$, and the Riesz representation theorem implies that there is an $x \in \mathcal{H}$ such that $\varphi(y) = \langle x, y\rangle$. It follows from Theorem 8.40 that $x_{k,k} \rightharpoonup x$ as $k \to \infty$. Moreover, from Proposition 8.44,

$$\|x\| \le \liminf_{k\to\infty} \|x_{k,k}\| \le 1,$$

so x belongs to the closed unit ball of $\mathcal{H}$. Thus every sequence in the ball has a weakly convergent subsequence whose limit belongs to the ball, so the ball is weakly sequentially compact. $\square$

An important application of Theorem 8.45 is to minimization problems. A function $f : K \to \mathbb{R}$ on a weakly closed set K is said to be weakly sequentially lower semicontinuous, or weakly lower semicontinuous for short, if

$$f(x) \le \liminf_{n\to\infty} f(x_n)$$

for every sequence (x_n) in K such that $x_n \rightharpoonup x$. For example, from Proposition 8.44, the norm $\|\cdot\|$ is weakly lower semicontinuous.

Theorem 8.46 Suppose that $f : K \to \mathbb{R}$ is a weakly lower semicontinuous function on a weakly closed, bounded subset K of a Hilbert space. Then f is bounded from below and attains its infimum.

The proof of this theorem is exactly the same as the proof of Theorem 1.72. Weak precompactness is a less stringent condition than strong precompactness, but weak closure and weak lower semicontinuity are more stringent conditions than their strong counterparts because there are many more weakly convergent sequences than strongly convergent sequences in an infinite-dimensional space.

A useful sufficient condition that allows one to deduce weak lower semicontinuity, or closure, from strong lower semicontinuity, or closure, is convexity. Convex sets were defined in (1.3). Convex functions are defined as follows.

Definition 8.47 Let $f : C \to \mathbb{R}$ be a real-valued function on a convex subset C of a real or complex linear space. Then f is *convex* if

$$f(tx + (1-t)y) \le tf(x) + (1-t)f(y)$$

for all $x, y \in C$ and $0 \le t \le 1$. If we have strict inequality in this equation whenever $x \ne y$ and $0 < t < 1$, then f is *strictly convex*.

The following result, called *Mazur's theorem*, explains the connection between convexity and weak convergence, and gives additional insight into weak convergence. We say that a vector y, in a real or complex linear space, is a *convex combination* of the vectors $\{x_1, x_2, \dots, x_n\}$ if there are nonnegative real numbers $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ such that

$$y = \sum_{k=1}^{n} \lambda_k x_k, \qquad \sum_{k=1}^{n} \lambda_k = 1.$$

Theorem 8.48 (Mazur) If (x_n) converges weakly to x in a Hilbert space, then there is a sequence (y_n) of finite convex combinations of $\{x_n\}$ such that (y_n) converges strongly to x.

Proof. Replacing x_n by $x_n - x$, we may assume that $x_n \rightharpoonup 0$. We will construct y_n as a mean of almost orthogonal terms of a subsequence of (x_n). We pick $n_1 = 1$, and choose $n_2 > n_1$ such that $\langle x_{n_1}, x_{n_2}\rangle \leq 1$. Given $n_1, \dots, n_k$, we pick $n_{k+1} > n_k$ such that

$$\left|\langle x_{n_1}, x_{n_{k+1}}\rangle\right| \leq \frac{1}{k}, \dots, \left|\langle x_{n_k}, x_{n_{k+1}}\rangle\right| \leq \frac{1}{k}. \tag{8.19}$$

This is possible because, by the weak convergence of (x_n), we have $\langle x_{n_i}, x_n\rangle \to 0$ as $n \to \infty$ for $1 \leq i \leq k$. Let

$$y_k = \frac{1}{k}\left(x_{n_1} + x_{n_2} + \ldots + x_{n_k}\right).$$

Then

$$\|y_k\|^2 = \frac{1}{k^2}\sum_{i=1}^{k} \|x_{n_k}\|^2 + \frac{2}{k^2}\mathrm{Re}\sum_{j=1}^{k}\sum_{i=1}^{j-1}\langle x_{n_i}, x_{n_j}\rangle.$$

Since (x_n) converges weakly, it is bounded, and there is a constant M such that $\|x_n\| \leq M$. Using (8.19), we obtain that

$$\|y_k\|^2 \leq \frac{M^2}{k} + \frac{2}{k^2}\sum_{j=1}^{k}\sum_{i=1}^{j-1}\frac{1}{j-1} \leq \frac{M^2+2}{k}.$$

Hence, $y_k \to 0$ as $k \to \infty$. □

It follows immediately from this result that a strongly closed, convex set is weakly closed. This need not be true without convexity; for example, the closed unit ball $\{x \in \ell^2(\mathbb{N}) \mid \|x\| \leq 1\}$ is weakly closed, but the closed unit sphere $\{x \in \ell^2(\mathbb{N}) \mid \|x\| = 1\}$ is not. It also follows from Exercise 8.19 that a strongly lower semicontinuous, convex function is weakly lower semicontinuous. We therefore have the following basic result concerning the existence of a minimizer for a convex optimization problem.

Theorem 8.49 Suppose that $f : C \to \mathbb{R}$, is a strongly lower semicontinuous, convex function on a strongly closed, convex, bounded subset C of a Hilbert space. Then f is bounded from below and attains its infimum. If f is strictly convex, then the minimizer is unique.

For example, the norm on a Hilbert space is strictly convex, as well as weakly lower semicontinuous, so it follows that every convex subset of a Hilbert space has a unique point with minimum norm. The existence of a minimizer for a nonconvex variational problem is usually much harder to establish, if one exists at all (see Exercise 8.22).

As in the finite dimensional case (see Exercise 1.25), a similar result holds if $f : \mathcal{H} \to \mathbb{R}$ and f is *coercive*, meaning that

$$\lim_{\|x\|\to\infty} f(x) = \infty.$$

Theorem 8.50 Suppose that $f : \mathcal{H} \to \mathbb{R}$, is a coercive, strongly lower semicontinuous, convex function on a Hilbert space $\mathcal{H}$. Then f is bounded from below and attains its infimum.

Proof. Since f is coercive, there is an $R > 0$ such that

$$f(x) > \inf_{y\in\mathcal{H}} f(y) + 1 \qquad \text{for all } x \in \mathcal{H} \text{ with } \|x\| > R.$$

We may therefore restrict f to the closed, convex ball $\{x \in \mathcal{H} \mid \|x\| \leq R\}$, and apply Theorem 8.49. □

The same theorems hold, with the same proofs, when C is a convex subset of a reflexive Banach space. We will use these abstract existence results to obtain a solution of Laplace's equation in Section 13.7.

8.7 References

For more about convex analysis, see Rockafellar [46]. For bounded linear operators in Hilbert spaces see, for example, Kato [26], Lusternik and Sobolev [33], Naylor and Sell [40], and Reed and Simon [45].

8.8 Exercises

Exercise 8.1 If M is a linear subspace of a linear space X, then the *quotient space* X/M is the set $\{x + M \mid x \in X\}$ of affine spaces

$$x + M = \{x + y \mid y \in M\}$$

parallel to M.

(a) Show that X/M is a linear space with respect to the operations

$$\lambda(x + M) = \lambda x + M, \qquad (x + M) + (y + M) = (x + y) + M.$$

(b) Suppose that $X = M \oplus N$. Show that N is linearly isomorphic to X/M.

(c) The *codimension* of M in X is the dimension of X/M. Is a subspace of a Banach space with finite codimension necessarily closed?

Exercise 8.2 If $\mathcal{H} = \mathcal{M} \oplus \mathcal{N}$ is an orthogonal direct sum, show that $\mathcal{M}^\perp = \mathcal{N}$ and $\mathcal{N}^\perp = \mathcal{M}$.

Exercise 8.3 Let $\mathcal{M}$, $\mathcal{N}$ be closed subspaces of a Hilbert space $\mathcal{H}$ and P, Q the orthogonal projections with $\operatorname{ran} P = \mathcal{M}$, $\operatorname{ran} Q = \mathcal{N}$. Prove that the following conditions are equivalent: (a) $\mathcal{M} \subset \mathcal{N}$; (b) $QP = P$; (c) $PQ = P$; (d) $\|Px\| \le \|Qx\|$ for all $x \in \mathcal{H}$; (e) $\langle x, Px\rangle \le \langle x, Qx\rangle$ for all $x \in \mathcal{H}$.

Exercise 8.4 Suppose that (P_n) is a sequence of orthogonal projections on a Hilbert space $\mathcal{H}$ such that

$$\operatorname{ran} P_{n+1} \supset \operatorname{ran} P_n, \qquad \bigcup_{n=1}^{\infty} \operatorname{ran} P_n = \mathcal{H}.$$

Prove that (P_n) converges strongly to the identity operator I as $n \to \infty$. Show that (P_n) does not converge to the identity operator with respect to the operator norm unless $P_n = I$ for all sufficiently large n.

Exercise 8.5 Let $\mathcal{H} = L^2(\mathbb{T}^3; \mathbb{R}^3)$ be the Hilbert space of 2π-periodic, square-integrable, vector-valued functions $\mathbf{u} : \mathbb{T}^3 \to \mathbb{R}^3$, with the inner product

$$\langle \mathbf{u}, \mathbf{v}\rangle = \int_{\mathbb{T}^3} \mathbf{u}(\mathbf{x}) \cdot \mathbf{v}(\mathbf{x})\, d\mathbf{x}.$$

We define subspaces $\mathcal{V}$ and $\mathcal{W}$ of $\mathcal{H}$ by

$$\begin{aligned} \mathcal{V} &= \left\{\mathbf{v} \in C^\infty(\mathbb{T}^3; \mathbb{R}^3) \mid \nabla \cdot \mathbf{v} = 0\right\}, \\ \mathcal{W} &= \left\{\mathbf{w} \in C^\infty(\mathbb{T}^3; \mathbb{R}^3) \mid \mathbf{w} = \nabla\varphi \text{ for some } \varphi : \mathbb{T}^3 \to \mathbb{R}\right\}. \end{aligned}$$

Show that $\mathcal{H} = \mathcal{M} \oplus \mathcal{N}$ is the orthogonal direct sum of $\mathcal{M} = \overline{\mathcal{V}}$ and $\mathcal{N} = \overline{\mathcal{W}}$.

Let P be the orthogonal projection onto $\mathcal{M}$. The velocity $\mathbf{v}(\mathbf{x}, t) \in \mathbb{R}^3$ and pressure $p(\mathbf{x}, t) \in \mathbb{R}$ of an incompressible, viscous fluid satisfy the *Navier-Stokes equations*

$$\begin{aligned} &\mathbf{v}_t + \mathbf{v} \cdot \nabla \mathbf{v} + \nabla p = \nu \Delta \mathbf{v}, \\ &\nabla \cdot \mathbf{v} = 0. \end{aligned}$$

Show that the velocity $\mathbf{v}$ satisfies the nonlocal equation

$$\mathbf{v}_t + P\left[\mathbf{v} \cdot \nabla \mathbf{v}\right] = \nu \Delta \mathbf{v}.$$

Exercise 8.6 Show that a linear operator $U : \mathcal{H}_1 \to \mathcal{H}_2$ is unitary if and only if it is an isometric isomorphism of normed linear spaces. Show that an invertible linear map is unitary if and only if its inverse is.

Exercise 8.7 If φ_y is the bounded linear functional defined in (8.5), prove that $\|\varphi_y\| = \|y\|$.

Exercise 8.8 Prove that $\mathcal{H}^*$ is a Hilbert space with the inner product defined by

$$\langle \varphi_x, \varphi_y \rangle_{\mathcal{H}^*} = \langle y, x \rangle_{\mathcal{H}}.$$

Exercise 8.9 Let $A \subset \mathcal{H}$ be such that

$$\mathcal{M} = \{x \in \mathcal{H} \mid x \text{ is a finite linear combination of elements in } A\}$$

is a dense linear subspace of $\mathcal{H}$. Prove that any bounded linear functional on $\mathcal{H}$ is uniquely determined by its values on A. If $\{u_\alpha\}$ is an orthonormal basis, find a necessary and sufficient condition on a family of complex numbers c_α for there to be a bounded linear functional φ such that $\varphi(u_\alpha) = c_\alpha$.

Exercise 8.10 Let $\{u_\alpha\}$ be an orthonormal basis of $\mathcal{H}$. Prove that $\{\varphi_{u_\alpha}\}$ is an orthonormal basis of $\mathcal{H}^*$.

Exercise 8.11 Prove that if $A : \mathcal{H} \to \mathcal{H}$ is a linear map and $\dim \mathcal{H} < \infty$, then

$$\dim \ker A + \dim \operatorname{ran} A = \dim \mathcal{H}.$$

Prove that, if $\dim \mathcal{H} < \infty$, then $\dim \ker A = \dim \ker A^*$. In particular, $\ker A = \{0\}$ if and only if $\ker A^* = \{0\}$.

Exercise 8.12 Suppose that $A : \mathcal{H} \to \mathcal{H}$ is a bounded, self-adjoint linear operator such that there is a constant $c > 0$ with

$$c\|x\| \le \|Ax\| \qquad \text{for all } x \in \mathcal{H}.$$

Prove that there is a unique solution x of the equation $Ax = y$ for every $y \in \mathcal{H}$.

Exercise 8.13 Prove that an orthogonal set of vectors $\{u_\alpha \mid \alpha \in \mathcal{A}\}$ in a Hilbert space $\mathcal{H}$ is an orthonormal basis if and only if

$$\sum_{\alpha \in \mathcal{A}} u_\alpha \otimes u_\alpha = I.$$

Exercise 8.14 Suppose that $A, B \in \mathcal{B}(\mathcal{H})$ satisfy

$$\langle x, Ay \rangle = \langle x, By \rangle \qquad \text{for all } x, y \in \mathcal{H}.$$

Prove that $A = B$. Use a polarization-type identity to prove that if $\mathcal{H}$ is a complex Hilbert space and

$$\langle x, Ax \rangle = \langle x, Bx \rangle \qquad \text{for all } x \in \mathcal{H},$$

then $A = B$. What can you say about A and B for real Hilbert spaces?

Exercise 8.15 Prove that for all $A, B \in \mathcal{B}(\mathcal{H})$, and $\lambda \in \mathbb{C}$, we have: (a) $A^{**} = A$; (b) $(AB)^* = B^* A^*$; (c) $(\lambda A)^* = \overline{\lambda} A^*$; (d) $(A + B)^* = A^* + B^*$; (e) $\|A^*\| = \|A\|$.

Exercise 8.16 Prove that the operator U defined in (8.16) is unitary.

Exercise 8.17 Prove that strong convergence implies weak convergence. Also prove that strong and weak convergence are equivalent in a finite-dimensional Hilbert space.

Exercise 8.18 Let (u_n) be a sequence of orthonormal vectors in a Hilbert space. Prove that $u_n \rightharpoonup 0$ weakly.

Exercise 8.19 Prove that a strongly lower-semicontinuous convex function $f : \mathcal{H} \to \mathbb{R}$ on a Hilbert space $\mathcal{H}$ is weakly lower-semicontinuous.

Exercise 8.20 Let $\mathcal{H}$ be a real Hilbert space, and $\varphi \in \mathcal{H}^*$. Define the quadratic functional $f : \mathcal{H} \to \mathbb{R}$ by

$$f(x) = \frac{1}{2}\|x\|^2 - \varphi(y).$$

Prove that there is a unique element $\overline{x} \in \mathcal{H}$ such that

$$f(\overline{x}) = \inf_{x \in \mathcal{H}} f(x).$$

Exercise 8.21 Show that a function is convex if and only if its epigraph, defined in Exercise 1.24, is a convex set.

Exercise 8.22 Consider the nonconvex functional

$$f : W^{1,4}([0,1]) \to \mathbb{R},$$

defined by

$$f(u) = \int_0^1 \left\{ u^2 + \left[1 - (u')^2\right]^2 \right\} dx,$$

where $W^{1,4}([0,1])$ is the Sobolev space of functions that belong to $L^4([0,1])$ and whose weak derivatives belong to $L^4([0,1])$. Show that the infimum of f on $W^{1,4}([0,1])$ is equal to zero, but that the infimum is not attained.