Christopher Heil

A Brief Guide to Metrics, Norms, and Inner Products

An Electronic Manuscript

April 19, 2020

Contents

Chapter 1
Metric Spaces

We will introduce and quickly study the basics of metrics, norms, and inner products in this short manuscript. This manuscript should be accessible to readers who have a background in undergraduate real analysis topics, including sequences, limits, suprema and infima, functions, cardinality, infinite series, Riemann integration, and vector spaces. Assuming that background, the manuscript essentially self-contained, although the presentation is compressed. Most major proofs are included, while certain other proofs are assigned as problems, and references are provided for proofs that are omitted. Explicit proofs of most of the statements made in this appendix, with extensive discussion and motivation, appear in the text [Heil18], and some of these can also be found in the volume [Heil11].

We assume that the reader is familiar with vector spaces (which are also called *linear spaces*). The scalar field associated with the vector spaces in this manuscript will always be either the real line $\mathbb{R}$ or the complex plane $\mathbb{C}$. When we specifically deal with the vector space $\mathbb{R}^d$, we assume that the scalar field is $\mathbb{R}$, and likewise we always assume that the scalar field for $\mathbb{C}^d$ is $\mathbb{C}$. For simplicity of presentation, when dealing with a vector space we often implicitly assume that the scalar field is $\mathbb{C}$, but usually only minor (and obvious) changes are required if the scalar field is $\mathbb{R}$. The elements of the scalar field are often called *scalars*, so for us a scalar will always mean a real or complex number.

1.1 Metrics and Convergence

A metric provides us with a notion of the distance between points in a set.

Definition 1.1.1 (Metric Space). Let X be a nonempty set. A *metric* on X is a function $\mathrm{d}\colon X \times X \to \mathbb{R}$ such that for all x, y, $z \in X$ we have:

(a) Nonnegativity: $0 \le \mathrm{d}(x,y) < \infty$,
(b) Uniqueness: $\mathrm{d}(x,y) = 0$ if and only if $x = y$,
(c) Symmetry: $\mathrm{d}(x,y) = \mathrm{d}(y,x)$, and
(d) The Triangle Inequality: $\mathrm{d}(x,z) \le \mathrm{d}(x,y) + \mathrm{d}(y,z)$.

If these conditions are satisfied, then X is a called a *metric space*. The number $\mathrm{d}(x,y)$ is called the *distance* from x to y. ♢

A metric space X does not have to be a vector space, although most of the metric spaces that we will encounter in this manuscript will be vector spaces (indeed, most are actually normed spaces). If X is a generic metric space, then we often refer to the elements of X as "points," but if we know that X is a vector space, then we may refer to the elements of X as "vectors." We will mostly use letters such as x, y, z to denote elements of a metric or normed space. However, in many cases we know that our space is a space of functions. In such a context we usually use letters such as f, g, h to denote elements of the space, and we may refer to those elements as "functions," "vectors," or "points."

Here is an example of a metric on an infinite-dimensional vector space.

Example 1.1.2. Given a sequence of real or complex scalars

$$x \;=\; (x_k)_{k\in\mathbb{N}} \;=\; (x_1, x_2, \dots),$$

we define the *ℓ^1-norm of x* to be

$$\|x\|_1 \;=\; \|(x_k)_{k\in\mathbb{N}}\|_1 \;=\; \sum_{k=1}^{\infty} |x_k|. \tag{1.1}$$

We say that the sequence x is *absolutely summable* (or just *summable* for short) if $\|x\|_1 < \infty$, and we let ℓ^1 denote the space of all absolutely summable sequences. That is,

$$\ell^1 \;=\; \biggl\{x = (x_k)_{k\in\mathbb{N}} \;:\; \|x\|_1 \;=\; \sum_{k=1}^{\infty} |x_k| \;<\; \infty\biggr\}.$$

This set is a vector space; in particular, it is closed under the operations of addition of sequences and multiplication of a sequence by a scalar. Further, a straightforward calculation shows that the function d defined by

$$\mathrm{d}(x,y) \;=\; \|x-y\|_1 \;=\; \sum_{k=1}^{\infty} |x_k - y_k|, \qquad \text{for } x,\, y \in \ell^1, \tag{1.2}$$

is a metric on ℓ^1, so ℓ^1 is both a vector space and a metric space. ♢

The following particular elements of ℓ^1 appear so often that we introduce a name for them.

Notation 1.1.3 (The Standard Basis Vectors). For each integer $n \in \mathbb{N}$, we let δ_n denote the sequence

$$\delta_n \;=\; (\delta_{nk})_{k\in\mathbb{N}} \;=\; (0,\dots,0,1,0,0,\dots).$$

That is, the nth component of the sequence δ_n is 1, while all other components are zero. We call δ_n the *nth standard basis vector*, and we refer to the family $\{\delta_n\}_{n\in\mathbb{N}}$ as the *sequence of standard basis vectors*, or simply the *standard basis*. $\diamondsuit$

1.2 Convergence and Completeness

Since a metric space has a notion of distance, we can define a corresponding notion of convergence in the space, as follows.

Definition 1.2.1 (Convergent Sequence). Let X be a metric space. A sequence of points $\{x_n\}_{n\in\mathbb{N}}$ in X *converges* to the point $x \in X$ if

$$\lim_{n\to\infty} \mathrm{d}(x_n, x) \;=\; 0.$$

That is, for every $\varepsilon > 0$ there must exist some integer $N > 0$ such that

$$n \,\geq\, N \quad\implies\quad \mathrm{d}(x_n, x) \,<\, \varepsilon.$$

In this case, we write $x_n \to x$. $\diamondsuit$

Convergence implicitly depends on the choice of metric for X, so if we want to emphasize that we are using a particular metric, we may write $x_n \to x$ *with respect to the metric* d.

Closely related to convergence is the idea of a Cauchy sequence, which is defined as follows.

Definition 1.2.2 (Cauchy Sequence). Let X be a metric space. A sequence of points $\{x_n\}_{n\in\mathbb{N}}$ in X is a *Cauchy sequence* if for every $\varepsilon > 0$ there exists an integer $N > 0$ such that

$$m, n \,\geq\, N \quad\implies\quad \mathrm{d}(x_m, x_n) \,<\, \varepsilon. \qquad \diamondsuit$$

By applying the Triangle Inequality, we immediately obtain the following relation between convergent and Cauchy sequences.

Lemma 1.2.3 (Convergent Implies Cauchy). *If $\{x_n\}_{n\in\mathbb{N}}$ is a convergent sequence in a metric space X, then $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in X.*

Proof. Assume that $x_n \to x$ as $n \to \infty$. If we fix $\varepsilon > 0$, then there exists an $N > 0$ such that $\mathrm{d}(x, x_n) < \frac{\varepsilon}{2}$ for all $n \geq N$. Consequently, if $m, n \geq N$ then the Triangle Inequality implies that

$$\mathrm{d}(x_m, x_n) \ \leq\ \mathrm{d}(x_m, x) \ +\ \mathrm{d}(x, x_n) \ <\ \varepsilon. \qquad \diamondsuit$$

However, the converse of Lemma 1.2.3 does not hold in general.

Example 1.2.4. The *standard metric* on the real line $\mathbb{R}$ or the complex plane $\mathbb{C}$ is

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

Using the terminology we will make precise in Definition 1.2.7, each of $\mathbb{R}$ and $\mathbb{C}$ is a *complete metric space* with respect to this metric. This means that every Cauchy sequence of real scalars must converge to a real scalar, and every Cauchy sequence of complex scalars must converge to a complex scalar (for a proof, see [Rud76, Thm. 3.11]).

Now consider the set of rational numbers $\mathbb{Q}$. This is also a metric space with respect to the metric $\mathrm{d}(x, y) = |x - y|$. However, we will show that $\mathbb{Q}$ is not complete. That is, we claim that there exist Cauchy sequences of rational numbers that do not converge to a rational number. For example, let x_n be the following rational numbers based on the decimal expansion of π:

$$x_1 = 3, \quad x_2 = 3.1, \quad x_3 = 3.14, \quad x_4 = 3.141, \quad \dots .$$

Then $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in both $\mathbb{Q}$ and $\mathbb{R}$. Moreover, it is a convergent sequence in $\mathbb{R}$ because $x_n \to \pi \in \mathbb{R}$. However, $\{x_n\}_{n\in\mathbb{N}}$ is not a convergent sequence *in the space* $\mathbb{Q}$, because there is no rational point $r \in \mathbb{Q}$ such that $x_n \to r$. $\diamondsuit$

The metric space $\mathbb{Q}$ in Example 1.2.4 is not a vector space over the real field. Here is an example of an infinite-dimensional vector space that contains Cauchy sequences that do not converge in the space.

Example 1.2.5. Let c_{00} be the space of all sequences of scalars that have only finitely many nonzero components:

$$c_{00} \ =\ \Bigl\{x = (x_1, \dots, x_N, 0, 0, \dots) \ :\ N > 0 \text{ and } x_1, \dots, x_N \in \mathbb{C}\Bigr\}.$$

Since c_{00} is a subset of ℓ^1, it is a metric space with respect to the metric d defined in equation (1.2).

For each $n \in \mathbb{N}$, let x_n be the sequence

$$x_n \ =\ (2^{-1}, \dots, 2^{-n}, 0, 0, 0, \dots).$$

and consider the sequence of vectors $\{x_n\}_{n\in\mathbb{N}}$, which is contained in both c_{00} and ℓ^1. If $m < n$, then

$$\|x_n - x_m\|_1 \;=\; \sum_{k=m+1}^{n} 2^{-k} \;<\; 2^{-m},$$

and it follows from this that $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence. If we consider this sequence to be a sequence in the space ℓ^1, then it does converge; in fact $x_n \to x$ where

$$x \;=\; (2^{-1}, 2^{-2}, \dots) \;=\; (2^{-k})_{k\in\mathbb{N}}.$$

However, this vector x does not belong to c_{00}, and there is no sequence $y \in c_{00}$ such that $x_n \to y$. Therefore c_{00} is not a complete metric space. $\square$

The set of rationals, $\mathbb{Q}$, is not a vector space over the real field or the complex field (which are the only fields we are considering in this manuscript). In contrast, c_{00} is a vector space with respect to the real field (if we are using real scalars), or with respect to the complex field (if we are using complex scalars). In fact, c_{00} is the finite linear span of the set of standard basis vectors $\mathcal{E} = \{\delta_k\}_{k\in\mathbb{N}}$:

$$c_{00} \;=\; \mathrm{span}(\mathcal{E}) \;=\; \biggl\{\sum_{k=1}^{N} x_k\delta_k \;:\; N > 0 \text{ and } x_1, \dots, x_N \in \mathbb{C}\biggr\}.$$

However, at least when we use the metric d defined by equation (1.2), there are Cauchy sequences in c_{00} that do not converge to an element of c_{00}. This fact should be compared to the next theorem, which shows that every Cauchy sequence in ℓ^1 converges to an element of ℓ^1.

Theorem 1.2.6. *If $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in ℓ^1, then there exists a vector $x \in \ell^1$ such that $x_n \to x$.*

Proof. Assume that $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in ℓ^1. Each x_n is a vector in ℓ^1, and for this proof we will write the components of x_n as

$$x_n \;=\; \bigl(x_n(1), x_n(2), \dots\bigr) \;=\; \bigl(x_n(k)\bigr)_{k\in\mathbb{N}}.$$

By the definition of a Cauchy sequence, if we choose an $\varepsilon > 0$, then there is an integer $N > 0$ such that $\mathrm{d}(x_m, x_n) = \|x_m - x_n\|_1 < \varepsilon$ for all $m, n \geq N$. Therefore, if we fix a particular index $k \in \mathbb{N}$ then for all $m, n \geq N$ we have

$$|x_m(k) - x_n(k)| \;\leq\; \sum_{j=1}^{\infty} |x_m(j) - x_n(j)| \;=\; \|x_m - x_n\|_1 \;<\; \varepsilon.$$

Thus, with k fixed, $\bigl(x_n(k)\bigr)_{n\in\mathbb{N}}$ is a Cauchy sequence of *scalars* and therefore must converge. Define

$$x(k) \;=\; \lim_{n\to\infty} x_n(k), \tag{1.3}$$

and set $x = \bigl(x(1), x(2), \dots\bigr)$. For each fixed k, we have by construction that the kth component of x_n converges to the kth component of x as $n \to \infty$.

We therefore say that x_n *converges componentwise to* x. However, this is not enough. We need to show that $x \in \ell^1$, and that x_n *converges to* x *in* ℓ^1*-norm.*

To do this, again choose an $\varepsilon > 0$. Since $\{x_n\}_{n\in\mathbb{N}}$ is Cauchy, there is an $N > 0$ such that $\|x_m - x_n\|_1 < \varepsilon$ for all $m, n \geq N$. Choose any particular $n \geq N$, and fix an integer $M > 0$. Then, since M is finite,

$$\begin{aligned}\sum_{k=1}^{M} |x(k) - x_n(k)| &= \sum_{k=1}^{M} \lim_{m\to\infty} |x_m(k) - x_n(k)| \\ &= \lim_{m\to\infty} \sum_{k=1}^{M} |x_m(k) - x_n(k)| \\ &\leq \lim_{m\to\infty} \|x_m - x_n\|_1 \\ &\leq \varepsilon.\end{aligned}$$

Since this is true for every M, we conclude that

$$\begin{aligned}\mathrm{d}(x_n, x) = \|x - x_n\|_1 &= \sum_{k=1}^{\infty} |x(k) - x_n(k)| \\ &= \lim_{M\to\infty} \sum_{k=1}^{M} |x(k) - x_n(k)| \\ &\leq \varepsilon. \end{aligned} \tag{1.4}$$

Even though we do not know yet that $x \in \ell^1$, this tells us that the vector $x - x_n$ has finite ℓ^1-norm and therefore belongs to ℓ^1. Since ℓ^1 is closed under addition, it follows that $x = (x - x_n) + x_n$ is in ℓ^1. Therefore our "candidate sequence" x does belong to ℓ^1. Further, equation (1.4) establishes that $\mathrm{d}(x_n, x) \leq \varepsilon$ for all $n \geq N$, so we have shown that x_n does indeed converge to x as $n \to \infty$. Hence ℓ^1 is complete. □

In summary, *some* metric spaces have the property that every Cauchy sequence in the space does converge to an element of the space. Since we can test for Cauchyness without having the limit vector x in hand, this is often very useful. We give such spaces the following name.

Definition 1.2.7 (Complete Metric Space). Let X be a metric space. If every Cauchy sequence in X converges to an element of X, then we say that X is *complete.* ◇

Theorem 1.2.6 shows that ℓ^1 is a complete metric space. More precisely, it is complete with respect to the metric $\mathrm{d}(x, y) = \|x - y\|_1$. If we impose a different metric on ℓ^1, then it might not be the case that ℓ^1 is complete with respect to that new metric. In particular, Problem 3.9.10 defines another metric on ℓ^1 and shows that ℓ^1 is not complete with respect to that metric.

Remark 1.2.8. The reader should be aware that the term "complete" is heavily overused and has a number of distinct mathematical meanings. In particular, the notion of a *complete space* as given in Definition 1.2.7 is quite different from the notion of a *complete sequence* that will be introduced in Definition 2.7.1. $\diamondsuit$

1.3 Topology in Metric Spaces

Given a metric space X, we make the following definitions.

- If $x \in X$ and $r > 0$, then the *open ball in X centered at x with radius r* is the set
$$B_r(x) \;=\; \bigl\{y \in X : \mathrm{d}(x,y) < r\bigr\}.$$
- A set $U \subseteq X$ is *open* if for each $x \in U$ there exists a radius $r > 0$ such that $B_r(x) \subseteq U$. Equivalently, U is open if and only if U can be written as a union of open balls.
- The *topology* of X is the collection of all open subsets of X.
- The *interior* of a set $E \subseteq X$ is the largest open set E° that is contained in E. Explicitly, $E^\circ = \cup\bigl\{U : U \text{ is open and } U \subseteq E\bigr\}$.
- A set $E \subseteq X$ is *closed* if its complement $X \backslash E$ is open.
- A set $E \subseteq X$ is *bounded* if it is contained in some open ball, i.e., there exists some $x \in X$ and some $r > 0$ such that $E \subseteq B_r(x)$.
- A point $x \in X$ is an *accumulation point* of a set $E \subseteq X$ if there exist points $x_n \in E$ with all $x_n \neq x$ such that $x_n \to x$.
- A point $x \in X$ is a *boundary point* of a set $E \subseteq X$ if for every $r > 0$ we have both $B_r(x) \cap E \neq \varnothing$ and $B_r(x) \cap E^{\mathrm{C}} \neq \varnothing$. The set of all boundary points of E is called the *boundary* of E, and it is denoted by ∂E.
- The *closure* of a set $E \subseteq X$ is the smallest closed set $\overline{E}$ that contains E. Explicitly, $\overline{E} = \cap\bigl\{F : F \text{ is closed and } E \subseteq F\bigr\}$.
- A set $E \subseteq X$ is *dense* in X if $\overline{E} = X$.
- We say that X is *separable* if there exists a countable subset E that is dense in X.

The following lemma gives useful equivalent reformulations of some of the notions defined above.

Lemma 1.3.1. *If E is a subset of a metric space X, then the following statements hold.*

(a) *E is closed if and only if*

$$x_n \in E \text{ and } x_n \to x \in X \qquad \Longrightarrow \qquad x \in E.$$

(b) *The closure of E satisfies*

$$\begin{aligned}\overline{E} &= E \cup \{y \in X : y \text{ is an accumulation point of } E\} \\ &= \big\{y \in X : \text{there exist } x_n \in E \text{ such that } x_n \to y\big\}. \qquad (1.5)\end{aligned}$$

(c) *E is dense in X if and only if for every point $x \in X$ there exist $x_n \in E$ such that $x_n \to x$.* $\diamondsuit$

Proof. (a) This statement can be deduced from part (b), or it can be proved directly. We assign the proof of statement (a) as Problem 1.5.13.

(b) Let F be the union of E and the accumulation points of E:

$$F = E \cup \{y \in X : y \text{ is an accumulation point of } E\},$$

and let G be the set of all limits of points of E:

$$G = \big\{y \in X : \text{there exist } x_n \in E \text{ such that } x_n \to y\big\}.$$

We must prove that $F = \overline{E} = G$.

To show that $\overline{E} \subseteq F$, fix any point $x \in F^{\mathrm{C}} = X \backslash F$. Then $x \notin E$ and x is not an accumulation point of E. If every open ball $B_r(x)$ contained a point $y \in E$, then x would be an accumulation point of E (because y cannot equal x). Therefore there must exist some ball $B_r(x)$ that contains no points of E. We claim $B_r(x)$ contains no accumulation points of E either. To see this, suppose that some point $y \in B_r(x)$ is an accumulation point of E. Then there must exist points $x_n \in E$ such that $x_n \to y$. By taking n large enough it follows that $x_n \in B_r(y)$, which is a contradiction. Thus $B_r(x)$ contains no points of E and no accumulation points of E, so $B_r(x) \subseteq F^{\mathrm{C}}$. This shows that F^{C} is open, and therefore F is closed. Since $E \subseteq F$ and $\overline{E}$ is the smallest closed set that contains E, we conclude that $\overline{E} \subseteq F$.

Next, we will prove that $F \subseteq \overline{E}$ by showing that $\overline{E}^{\,\mathrm{C}} \subseteq F^{\mathrm{C}}$. Since $\overline{E}^{\,\mathrm{C}}$ is open, if x belongs to this set then there exists some $r > 0$ such that

$$B_r(x) \subseteq \overline{E}^{\,\mathrm{C}}.$$

Hence $B_r(x)$ contains no points of E, so x cannot be an accumulation point of E. We also have $x \notin E$ since $x \notin \overline{E}$, so it follows that $x \notin F$.

The above work shows that $\overline{E} = F$. Now choose any point $y \in G$. If $y \in E$, then $y \in F$ since $E \subseteq F$. So, suppose that $y \in G \backslash E$. Then, by the definition of G, there exist points $x_n \in E$ such that $x_n \to y$. But $x_n \neq y$ for any n since $x_n \in E$ and $y \notin G$, so y is an accumulation point of E. Therefore $y \in F$. Thus $G \subseteq F$.

Finally, suppose that $y \in F$. If $y \in E$, then y is the limit of the sequence $\{y, y, y, \dots\}$, so $y \in G$. On the other hand, if $y \in F \setminus E$, then y must be an accumulation point of E. Therefore there exist points $x_n \in E$, with $x_n \neq y$ for every n, such that $x_n \to y$. This shows that y is a limit of points from E, so $y \in G$. Therefore $F \subseteq G$.

(c) This follows from equation (1.5) and the fact that E is dense in X if and only if $X = \overline{E}$. $\square$

Restating parts of Lemma 1.3.1, we see that:

- a set E is closed if and only if it contains every limit of points from E,
- the closure of a set E is the set of all limits of points from E, and
- a set E is dense in X if and only if every point in X is a limit of points from E.

For example, the set of rationals $\mathbb{Q}$ is not a closed subset of $\mathbb{R}$ because a limit of rational points need not be rational. The closure of $\mathbb{Q}$ is $\mathbb{R}$ because every point in $\mathbb{R}$ can be written as a limit of rational points. Similarly, $\mathbb{Q}$ is a dense subset of $\mathbb{R}$ because every real number can be written as a limit of rational numbers. Since $\mathbb{Q}$ is both countable and dense, this also shows that $\mathbb{R}$ is a separable space. Problem 2.9.13 shows that the space ℓ^1 defined in Example 1.1.2 is separable.

1.4 Compact Sets in Metric Spaces

Now we define compact sets and discuss their properties.

Definition 1.4.1 (Compact Set). A subset K of a metric space X is *compact* if every covering of K by open sets has a finite subcovering. Stated precisely, K is compact if it is the case that whenever

$$K \subseteq \bigcup_{i \in I} U_i,$$

where $\{U_i\}_{i \in I}$ is any collection of open subsets of X, there exist *finitely* many indices $i_1, \dots, i_N \in I$ such that

$$K \subseteq \bigcup_{k=1}^{N} U_{i_k}. \qquad \diamondsuit$$

First we show that every compact subset of a metric space is both closed and bounded.

Lemma 1.4.2. *If K is a compact subset of a metric space X, then K is closed and there exists some open ball $B_r(x)$ such that $K \subseteq B_r(x)$.*

Proof. Suppose that K is compact, and fix any particular point $x \in X$. Then the union of the open balls $B_n(x)$ over $n \in \mathbb{N}$ is all of X, so $\{B_n(x)\}_{n\in\mathbb{N}}$ an open cover of K. This cover must have a finite subcover, and since the balls are nested it follows that $K \subseteq B_n(x)$ for some single n.

It remains to show that K is closed. If $K = X$ then we are done, so assume that $K \neq X$. Fix any point y in $K^{\mathrm{C}} = X \backslash K$. If $x \in K$, then $x \neq y$, so by the *Hausdorff property* stated in Problem 1.5.7, there exist disjoint open sets U_x and V_x such that $x \in U_x$ and $y \in V_x$. The collection $\{U_x\}_{x\in K}$ is an open cover of K, so it must contain some finite subcover, say

$$K \subseteq U_{x_1} \cup \cdots \cup U_{x_N}. \tag{1.6}$$

Each V_{x_j} is disjoint from U_{x_j}, so it follows from equation (1.6) that the set

$$V = V_{x_1} \cap \cdots \cap V_{x_N}$$

is entirely contained in the complement of K. Thus, V is an open set that satisfies

$$y \in V \subseteq K^{\mathrm{C}}.$$

This shows that K^{C} is open, so we conclude that K is closed. □

If X is a *finite-dimensional normed space*, then a subset of X is compact *if and only if* it closed and bounded (for a proof, see [Kre78, Thm. 2.5-3] or [Heil18, Sec. 3.7]). However, *every infinite-dimensional normed space contains a set that is closed and bounded but not compact.* In fact, if X is an infinite-dimensional normed space then the closed unit ball in X is closed and bounded but not compact. Problem 2.9.12 asks for a proof of this fact for the space ℓ^1; for a proof in the general setting see [Heil11, Exercise 1.44].

We will give several equivalent reformulations of compactness for subsets of metric spaces in terms of the following concepts.

Definition 1.4.3. Let E be a subset of a metric space X.

(a) E is *sequentially compact* if every sequence $\{x_n\}_{n\in\mathbb{N}}$ of points from E contains a convergent subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ whose limit belongs to E.

(b) E is *totally bounded* if for each radius $r > 0$ we can cover E by finitely many open balls of radius r. That is, for each $r > 0$ there must exist finitely many points $x_1, \dots, x_N \in X$ such that

$$E \subseteq \bigcup_{k=1}^{N} B_r(x_k). \qquad \diamondsuit$$

We will need the following lemma.

Lemma 1.4.4. *Let E be a sequentially compact subset of a metric space X. If $\{U_i\}_{i\in I}$ is an open cover of E, then there exists a number $\delta > 0$ such that if B is an open ball of radius δ that intersects E, then there is an index $i \in I$ such that $B \subseteq U_i$.*

Proof. Let $\{U_i\}_{i\in I}$ be an open cover of E. We want to prove that there is a $\delta > 0$ such that

$$B_r(x) \cap E \neq \varnothing \quad \Longrightarrow \quad B_r(x) \subseteq U_i \text{ for some } i \in I.$$

Suppose that no $\delta > 0$ has this property. Then for each positive integer n, there must exist some open ball with radius $\frac{1}{n}$ that intersects E but is not contained in any set U_i. Call this open ball G_n. For each n, choose a point $x_n \in G_n \cap E$. Then since $\{x_n\}_{n\in\mathbb{N}}$ is a sequence in E and E is sequentially compact, there must be a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ that converges to a point $x \in E$. Since $\{U_i\}_{i\in I}$ is a cover of E, we must have $x \in U_i$ for some $i \in I$, and since U_i is open there must exist some $r > 0$ such that $B_r(x) \subseteq U_i$. Now choose k large enough that we have both

$$\frac{1}{n_k} < \frac{r}{3} \qquad \text{and} \qquad \mathrm{d}\big(x, x_{n_k}\big) < \frac{r}{3}.$$

Keeping in mind the facts that G_{n_k} contains x_{n_k}, G_{n_k} is an open ball with radius $1/n_k$, the distance from x to x_{n_k} is less than $r/3$, and $B_r(x)$ has radius r, it follows that $G_{n_k} \subseteq B_r(x) \subseteq U_i$, which is a contradiction. □

Now we prove some reformulations of compactness that hold for subsets of metric spaces.

Theorem 1.4.5. *If K is a subset of a complete metric space X, then the following statements are equivalent.*

(a) *K is compact.*

(b) *K is sequentially compact.*

(c) *K is totally bounded and every Cauchy sequence of points from K converges to a point in K.*

Proof. (a) $\Rightarrow$ (b). Suppose that K is not sequentially compact. Then there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ in K that has no subsequence that converges to an element of K. Choose any point $x \in K$. If every open ball centered at x contains infinitely many of the points x_n, then by considering radii $r = \frac{1}{k}$ we can construct a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ that converges to x. This is a contradiction, so there must exist some open ball centered at x that contains only finitely many x_n. If we call this ball B_x, then $\{B_x\}_{x\in K}$ is an open cover of K that contains no finite subcover. Consequently, K is not compact.

(b) $\Rightarrow$ (c). Suppose that K is sequentially compact, and suppose that $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in K. Then since K is sequentially compact, there is a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ that converges to some point $x \in K$. Hence $\{x_n\}_{n\in\mathbb{N}}$ is Cauchy and has a convergent subsequence. Appealing to Problem 1.5.10, this implies that $x_n \to x$. Therefore K is complete.

Suppose that K were not totally bounded. Then there would be a radius $r > 0$ such that K cannot be covered by finitely many open balls of radius r

centered at points of X. Choose any point $x_1 \in K$. Since K cannot be covered by a single ball of radius r, K cannot be a subset of $B_r(x_1)$. Hence there exists a point $x_2 \in K \setminus B_r(x_1)$. In particular, $\mathrm{d}(x_2, x_1) \ge r$. But K cannot be covered by two balls of radius r, so there must exist a point x_3 that belongs to $K \setminus \big(B_r(x_1) \cup B_r(x_2)\big)$. In particular, we have both $\mathrm{d}(x_3, x_1) \ge r$ and $\mathrm{d}(x_3, x_2) \ge r$. Continuing in this way, we obtain a sequence of points $\{x_n\}_{n\in\mathbb{N}}$ in K that has no convergent subsequence, which is a contradiction.

(c) $\Rightarrow$ (b). Assume that K is complete and totally bounded, and let $\{x_n\}_{n\in\mathbb{N}}$ be any sequence of points in K. Since K is totally bounded, it can be covered by finitely many open balls of radius $\frac{1}{2}$. Each x_n belongs to K and hence must be contained in one or more of these balls. But there are only finitely many of the balls, so at least one ball must contain x_n for infinitely many different indices n. That is, there is some infinite subsequence $\{x_n^{(1)}\}_{n\in\mathbb{N}}$ of $\{x_n\}_{n\in\mathbb{N}}$ that is contained in an open ball of radius $\frac{1}{2}$. The Triangle Inequality therefore implies that

$$\forall\, m, n \in \mathbb{N}, \quad \mathrm{d}\big(x_m^{(1)}, x_n^{(1)}\big) \;<\; 1.$$

Similarly, since K can be covered by finitely many open balls of radius $\frac{1}{4}$, there is some subsequence $\{x_n^{(2)}\}_{n\in\mathbb{N}}$ of $\{x_n^{(1)}\}_{n\in\mathbb{N}}$ such that

$$\forall\, m, n \in \mathbb{N}, \quad \mathrm{d}\big(x_m^{(2)}, x_n^{(2)}\big) \;<\; \frac{1}{2}.$$

Continuing by induction, for each $k > 1$ we find a subsequence $\{x_n^{(k)}\}_{n\in\mathbb{N}}$ of $\{x_n^{(k-1)}\}_{n\in\mathbb{N}}$ such that $\mathrm{d}\big(x_m^{(k)}, x_n^{(k)}\big) < \frac{1}{k}$ for all $m, n \in \mathbb{N}$.

Now consider the diagonal subsequence $\{x_k^{(k)}\}_{k\in\mathbb{N}}$. Given $\varepsilon > 0$, let N be large enough that $\frac{1}{N} < \varepsilon$. If $j \ge k > N$, then $x_j^{(j)}$ is one element of the sequence $\{x_n^{(k)}\}_{n\in\mathbb{N}}$, say $x_j^{(j)} = x_n^{(k)}$. Hence

$$\mathrm{d}\big(x_j^{(j)}, x_k^{(k)}\big) \;=\; \mathrm{d}\big(x_n^{(k)}, x_k^{(k)}\big) \;<\; \frac{1}{k} \;<\; \frac{1}{N} \;<\; \varepsilon.$$

Thus $\{x_k^{(k)}\}_{k\in\mathbb{N}}$ is a Cauchy subsequence of the original sequence $\{x_n\}_{n\in\mathbb{N}}$. Since K is complete, this subsequence must converge to some element of K. Hence K is sequentially compact.

(b) $\Rightarrow$ (a). Assume that K is sequentially compact. Since we have already proved that statement (b) implies statement (c), we know that K is complete and totally bounded.

Suppose that $\{U_i\}_{i\in I}$ is any open cover of K. By Lemma 1.4.4, there exists a $\delta > 0$ such that if B is an open ball of radius δ that intersects K, then there is an $i \in I$ such that $B \subseteq U_i$. However, K is totally bounded, so we can cover K by finitely many open balls of radius δ. Each of these balls is contained in some U_i, so K is covered by finitely many U_i. $\square$

1.5 Continuity for Functions on Metric Spaces

Here is the abstract definition of continuity for functions on metric spaces.

Definition 1.5.1 (Continuous Function). Let X and Y be metric spaces. We say that a function $f\colon X \to Y$ is *continuous* if for every open set $V \subseteq Y$, its inverse image $f^{-1}(V)$ is an open subset of X. $\diamondsuit$

By applying the definition of continuity, the following lemma shows that the *direct image* of a compact set under a continuous function is compact.

Lemma 1.5.2. *Let X and Y be metric spaces. If $f\colon X \to Y$ is continuous and K is a compact subset of X, then $f(K)$ is a compact subset of Y.*

Proof. Let $\{V_i\}_{i\in J}$ be any open cover of $f(K)$. Each set $U_i = f^{-1}(V_i)$ is open, and $\{U_i\}_{i\in J}$ is an open cover of K. Since K is compact, this cover must have a finite subcover $\{U_{i_1}, \dots, U_{i_N}\}$. But then $\{V_{i_1}, \dots, V_{i_N}\}$ is a finite subcover of $f(K)$, so we conclude that $f(K)$ is compact. $\square$

The next lemma gives a useful reformulation of continuity for functions on metric spaces in terms of preservation of limits.

Lemma 1.5.3. *Let X be a metric space with metric d_X, and let Y be a metric space with metric d_Y. If $f\colon X \to Y$, then the following three statements are equivalent.*

(a) *f is continuous.*

(b) *If x is any point in X, then for every $\varepsilon > 0$ there exists a $\delta > 0$ such that*

$$\mathrm{d}_X(x,y) < \delta \quad \Longrightarrow \quad \mathrm{d}_Y\bigl(f(x), f(y)\bigr) < \varepsilon.$$

(c) *If x is any point in X and $\{x_n\}_{n\in\mathbb{N}}$ is any sequence of points from X, then*

$$x_n \to x \text{ in } X \quad \Longrightarrow \quad f(x_n) \to f(x) \text{ in } Y. \qquad \diamondsuit$$

Proof. For this proof we let $B_r^X(x)$ and $B_s^Y(y)$ denote open balls in X and Y, respectively.

(a) $\Rightarrow$ (b). Suppose that f is continuous, and choose any point $x \in X$ and any $\varepsilon > 0$. Then the ball $V = B_\varepsilon^Y\bigl(f(x)\bigr)$ is an open subset of Y, so $U = f^{-1}(V)$ must be an open subset of X. Since $x \in U$, there exists some $\delta > 0$ such that $B_\delta^X(x) \subseteq U$. If $y \in X$ is any point that satisfies $\mathrm{d}_X(x,y) < \delta$, then $y \in B_\delta^X(x) \subseteq U$, and therefore

$$f(y) \in f(U) \subseteq V = B_\varepsilon^Y\bigl(f(x)\bigr).$$

Consequently $\mathrm{d}_Y\bigl(f(x), f(y)\bigr) < \varepsilon$.

(b) $\Rightarrow$ (c). Assume that statement (b) holds, choose any point $x \in X$, and suppose that points $x_n \in X$ are such that $x_n \to x$. Fix any $\varepsilon > 0$, and let $\delta > 0$ be the number whose existence is given by statement (b). Since $x_n \to x$, there must exist some $N > 0$ such that $\mathrm{d}_X(x, x_n) < \delta$ for all $n \geq N$. Statement (b) therefore implies that $\mathrm{d}_Y\bigl(f(x), f(x_n)\bigr) < \varepsilon$ for all $n \geq N$, so we conclude that $f(x_n) \to f(x)$ in Y.

(c) $\Rightarrow$ (a). Suppose that statement (c) holds, and let V be any open subset of Y. Suppose that $f^{-1}(V)$ were not open in X. Then there is some point $x \in f^{-1}(V)$ such that there is no radius $r > 0$ for which the open ball $B_r(x)$ is a subset of $f^{-1}(V)$. In particular, for each $n \in \mathbb{N}$ the ball $B_{1/n}(x)$ is not contained in $f^{-1}(V)$, and therefore there is some point $x_n \in B_{1/n}(x)$ such that $x_n \notin f^{-1}(V)$. As a consequence, $\mathrm{d}(x, x_n) < 1/n$ for every n, but $f(x_n) \notin V$ for any n.

Now, $x \in f^{-1}(V)$, so $f(x)$ does belong to V. Therefore, since V is open, $B_\varepsilon(f(x)) \subseteq V$ for some $\varepsilon > 0$.

On the other hand, $x_n \to x$, so $f(x_n) \to f(x)$ by statement (c). Consequently, there is some $N > 0$ such that $\mathrm{d}\bigl(f(x), f(x_n)\bigr) < \varepsilon$ for all $n \geq N$. But then $f(x_N) \in B_\varepsilon(f(x)) \subseteq V$, which contradicts the fact that $f(x_N) \notin V$. Therefore $f^{-1}(V)$ must be open, and hence f is continuous. $\square$

The number δ that appears in statement (b) of Lemma 1.5.3 depends both on the point x and the number ε. We say that a function f is *uniformly continuous* if δ can be chosen independently of x. Here is the precise definition.

Definition 1.5.4 (Uniform Continuity). Let X be a metric space with metric d_X, and let Y be a metric space with metric d_Y. If $E \subseteq X$, then we say that a function $f\colon X \to Y$ is *uniformly continuous* on E if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $x, y \in E$ we have

$$\mathrm{d}_X(x,y) < \delta \implies \mathrm{d}_Y\bigl(f(x),\, f(y)\bigr) < \varepsilon. \qquad \diamondsuit$$

The next lemma shows that a continuous function whose domain is a compact set is uniformly continuous on that set.

Lemma 1.5.5. *Let X and Y be metric spaces. If K is a compact subset of X and $f\colon K \to Y$ is continuous, then f is uniformly continuous on K.*

Proof. For this proof, let d_X and d_Y denote the metrics on X and Y, and let $B^X_r(x)$ and $B^Y_s(y)$ denote open balls in X and Y, respectively.

Suppose that $f\colon K \to Y$ is continuous, and fix $\varepsilon > 0$. For each point $z \in Y$, let $U_z = f^{-1}(B^Y_\varepsilon(z))$. Then $\{U_z\}_{z \in Y}$ is an open cover of K. Since K is compact, it is sequentially compact. Therefore Lemma 1.4.4 implies that there exists a number $\delta > 0$ such that if B is any open ball of radius δ in X that intersects K, then $B \subseteq U_z$ for some $z \in Y$.

Now choose any points $x, y \in K$ with $\mathrm{d}_X(x,y) < \delta$. Then $x, y \in B^X_\delta(x)$, and $B^X_\delta(x)$ must be contained in some set U_z, so

$$f(x), f(y) \in f(B_\delta^X(x)) \subseteq f(U_z) \subseteq B_\varepsilon^Y(z).$$

Therefore $\mathrm{d}_Y(f(x),\, f(y)) < 2\varepsilon$, so f is uniformly continuous. $\square$

Problems

1.5.6. Given a set X, show that

$$\mathrm{d}(x,y) = \begin{cases} 1, & x \neq y, \\ 0, & x = y, \end{cases}$$

is a metric on X. This is called the *discrete metric* on X.

1.5.7. Prove that every metric space X is *Hausdorff*, which means that if $x \neq y$ are two distinct elements of X, then there exist *disjoint open sets* U, V such that $x \in U$ and $y \in V$.

1.5.8. Let X be a metric space X. Prove that the following inequality (called the *Reverse Triangle Inequality*) holds for all x, y, $z \in X$:

$$\bigl|\mathrm{d}(x,z) - \mathrm{d}(y,z)\bigr| \leq \mathrm{d}(x,y).$$

1.5.9. Let X be a metric space. Prove that the limit of any convergent sequence in X is unique, i.e., if $x_n \to y$ and $x_n \to z$ then $y = z$.

1.5.10. Suppose that $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in a metric space X, and suppose there exists a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ that converges to $x \in X$, i.e., $x_{n_k} \to x$ as $k \to \infty$. Prove that $x_n \to x$ as $n \to \infty$.

1.5.11. Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in a metric space X, and fix $x \in X$. Suppose that every subsequence $\{y_n\}_{n\in\mathbb{N}}$ of $\{x_n\}_{n\in\mathbb{N}}$ has a subsequence $\{z_n\}_{n\in\mathbb{N}}$ of $\{y_n\}_{n\in\mathbb{N}}$ such that $z_n \to x$. Prove that $x_n \to x$.

1.5.12. Let A be a subset of a metric space X.

(a) Prove that A is open if and only if $A = A^\circ$.

(b) Prove that A is closed if and only if $A = \overline{A}$.

1.5.13. Prove part (a) of Lemma 1.3.1.

1.5.14. Let X be a *complete* metric space, and let A be a subset of X. Prove that the following two statements are equivalent.

(a) A is closed.

(b) A is complete, i.e., every Cauchy sequence of points in A converges to a point of A.

1.5.15. (a) Let Q be the "open first quadrant" in ℓ^1, i.e.,

$$Q \;=\; \bigl\{x = (x_k)_{k\in\mathbb{N}} \in \ell^1 : x_k > 0 \text{ for every } k\bigr\}.$$

Prove that Q is not an open subset of ℓ^1.

(b) Let R be the "closed first quadrant" in ℓ^1, i.e.,

$$R \;=\; \bigl\{x = (x_k)_{k\in\mathbb{N}} \in \ell^1 : x_k \ge 0 \text{ for every } k\bigr\}.$$

Prove that R is a closed subset of ℓ^1. Additionally, prove that R contains no open balls, and therefore its interior is the empty set.

1.5.16. Let X be a metric space. The *diameter* of a subset S of X is $\operatorname{diam}(S) = \sup\bigl\{\mathrm{d}(x,y) : x, y \in S\bigr\}$.

(a) Suppose that X is complete and $F_1 \supseteq F_2 \supseteq \cdots$ is a nested decreasing sequence of closed nonempty subsets of X such that $\operatorname{diam}(F_n) \to 0$. Prove that there exists some point $x \in X$ such that $\cap F_n = \{x\}$.

(b) Show by example that the conclusion of part (a) can fail if X is not complete.

1.5.17. Let E be a subset of a metric space X. Given $x \in X$, prove that the following four statements are equivalent.

(a) x is an accumulation point of E.

(b) If U is an open set that contains x, then $(E \cap U) \setminus \{x\} \neq \varnothing$, i.e., there exists a point $y \in E \cap U$ such that $y \neq x$.

(c) If $r > 0$, then exists a point $y \in E$ such that $0 < \mathrm{d}(x,y) < r$.

(d) Every open set U that contains x also contains *infinitely many* distinct points of E.

1.5.18. Let X be a metric space.

(a) Prove that if A and B are subsets of X, then $\overline{A \cup B} \;=\; \overline{A} \,\cup\, \overline{B}$.

(b) Let I be an arbitrary index set. Given sets $E_i \subseteq X$ for $i \in I$, prove that

$$\overline{\bigcap_{i\in I} E_i} \;\subseteq\; \bigcap_{i\in I} \overline{E_i}.$$

Show by example that equality can fail.

1.5.19. Let K be a compact subset of a metric space X, and let E be a subset of K. Prove that E is closed if and only if E is compact.

1.5.20. Let K be a compact subset of a metric space X, and assume that $f\colon K \to \mathbb{R}$ is continuous. Prove that f achieves its maximum and its minimum on K, i.e., there exist points $x, y \in K$ such that

$$f(x) \;=\; \inf\bigl\{f(t) : t \in K\bigr\} \qquad \text{and} \qquad f(y) \;=\; \sup\bigl\{f(t) : t \in K\bigr\}.$$

1.5.21. (a) Let X and Y be metric spaces, and suppose that $f\colon X \to Y$ is uniformly continuous. Prove that f maps Cauchy sequences to Cauchy sequences, i.e.,

$$\{x_n\}_{n\in\mathbb{N}} \text{ is Cauchy in } X \quad\implies\quad \{f(x_n)\}_{n\in\mathbb{N}} \text{ is Cauchy in } Y.$$

(b) Show by example that part (a) can fail if we only assume that f is continuous instead of uniformly continuous. (Contrast this with Lemma 1.5.3, which shows that every continuous function must map convergent sequences to convergent sequences.)

Chapter 2
Norms and Banach Spaces

While a metric provides us with a notion of the *distance between points* in a space, a norm gives us a notion of the *length* of an individual vector. A norm can only be defined on a vector space, while a metric can be defined on arbitrary sets.

2.1 The Definition of a Norm

Here is the definition of a norm, and the slightly weaker notion of a seminorm.

Definition 2.1.1 (Seminorms and Norms). Let X be a vector space over the real field $\mathbb{R}$ or the complex field $\mathbb{C}$. A *seminorm* on X is a function $\|\cdot\|\colon X \to \mathbb{R}$ such that for all vectors $x, y \in X$ and all scalars c we have:

(a) Nonnegativity: $\|x\| \geq 0$,

(b) Homogeneity: $\|cx\| = |c|\,\|x\|$, and

(c) The Triangle Inequality: $\|x+y\| \leq \|x\| + \|y\|$.

A seminorm is a *norm* if we also have:

(d) Uniqueness: $\|x\| = 0$ if and only if $x = 0$.

A vector space X together with a norm $\|\cdot\|$ is called a *normed vector space*, a *normed linear space*, or simply a *normed space*. $\diamondsuit$

We refer to the number $\|x\|$ as the *length* of a vector x, and we say that $\|x-y\|$ is the *distance* between the vectors x and y. A vector x that has length 1 is called a *unit vector*, or is said to be *normalized*.

We usually use $\|\cdot\|$ to denote a norm or seminorm, sometimes with subscripts to distinguish among different norms (for example, $\|\cdot\|_a$ and $\|\cdot\|_b$). Other common symbols for norms or seminorms are $|\cdot|$, $|||\cdot|||$, or $\rho(\cdot)$. The absolute value function $|x|$ is a norm on the real line $\mathbb{R}$ and on the complex plane $\mathbb{C}$. Likewise, the Euclidean norm is a norm on $\mathbb{R}^d$ and $\mathbb{C}^d$.

Example 2.1.2. The space ℓ^1 introduced in Example 1.1.2 is a vector space, and it is straightforward to check that the "ℓ^1-norm"

$$\|x\|_1 \;=\; \sum_{k=1}^{\infty} |x_k|, \qquad \text{for } x = (x_k)_{k\in\mathbb{N}} \in \ell^1,$$

is indeed a norm on ℓ^1. Therefore ℓ^1 is a normed space. ♢

2.2 The Induced Metric

If X is a normed space, then it follows directly from the definition of a norm that

$$\mathrm{d}(x,y) \;=\; \|x-y\|, \qquad \text{for } x, y \in X,$$

defines a metric on X. This is called the *metric on X induced from* $\|\cdot\|$, or simply the *induced metric* on X.

Consequently, whenever we are given a normed space X, we have a metric on X as well as a norm, and this metric is the induced metric. Therefore all definitions made for metric spaces apply to normed spaces, using the induced norm $\mathrm{d}(x,y) = \|x-y\|$. For example, convergence in a normed space is defined by

$$x_n \to x \quad \iff \quad \lim_{n\to\infty} \|x - x_n\| \;=\; 0.$$

It may be possible to place a metric on X other than the induced metric, but unless we explicitly state otherwise, all metric-related statements on a normed space are taken with respect to the induced metric.

2.3 Properties of Norms

Here are some useful properties of norms (we assign the proof as Problem 2.9.6).

Lemma 2.3.1. *If X is a normed space, then the following statements hold.*

(a) Reverse Triangle Inequality: $\big|\|x\| - \|y\|\big| \le \|x-y\|$.

(b) Convergent implies Cauchy: *If $x_n \to x$, then $\{x_n\}_{n\in\mathbb{N}}$ is Cauchy.*

(c) Boundedness of Cauchy sequences: *If $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence, then* $\sup \|x_n\| < \infty$.

(d) Continuity of the norm: *If* $x_n \to x$, *then* $\|x_n\| \to \|x\|$.

(e) Continuity of vector addition: *If* $x_n \to x$ *and* $y_n \to y$, *then* $x_n + y_n \to x + y$.

(f) Continuity of scalar multiplication: *If* $x_n \to x$ *and* $c_n \to c$, *then* $c_n x_n \to cx$. $\diamondsuit$

2.4 Convexity

If x and y are vectors in a vector space X, then the line segment joining x to y is the set of all points of the form $tx + (1-t)y$ where $0 \le t \le 1$.

A subset K of a vector space is *convex* if given any two points $x, y \in K$, the line segment joining x to y is entirely contained within K. That is, K is convex if

$$x, y \in K,\ 0 \le t \le 1 \implies tx + (1-t)y \in K.$$

If X is a normed space, then every open ball $B_r(x)$ in X is convex (this is Problem 2.9.7). However, if X is a vector space that is merely a metric space but not a normed space then open balls in X need not be convex.

2.5 Banach Spaces

Every convergent sequence in a normed vector space must be Cauchy, but the converse does not hold in general. In *some* normed spaces it is true that every Cauchy sequence in the space is convergent. We give spaces that have this property the following name.

Definition 2.5.1 (Banach Space). A normed space X is a *Banach space* if it is complete, i.e., if every Cauchy sequence in X converges to an element of X. $\diamondsuit$

The terms "Banach space" and "complete normed space" are interchangeable, and we will use whichever is more convenient in a given context.

The set of real numbers $\mathbb{R}$ is complete with respect to absolute value (using real scalars), and likewise the complex plane $\mathbb{C}$ is complete with respect to absolute value (using complex scalars). More generally, $\mathbb{R}^d$ and $\mathbb{C}^d$ are Banach spaces with respect to the Euclidean norm. In fact, every finite-dimensional vector space V is complete with respect to any norm that we place on V (for a proof, see [Con90, Sec. III.3] or [Heil18, Sec. 3.7]).

We proved in Theorem 1.2.6 that the space ℓ^1 is complete. Therefore ℓ^1 is an example of an infinite-dimensional Banach space. One example of an infinite-dimensional normed space that is not a Banach space is the space c_{00} studied in Example 1.2.5. A space that is complete with respect to one norm may be incomplete if replace that norm with another norm. For example, ℓ^1 is no longer complete if we replace the norm $\|\cdot\|_1$ with the "ℓ^2-norm" (see Problem 3.9.10).

2.6 Infinite Series in Normed Spaces

An infinite series is a limit of the partial sums of the series. Hence, the definition of an infinite series involves both addition and limits, each of which are defined in normed spaces. The precise definition of an infinite series in a normed space is as follows.

Definition 2.6.1 (Convergent Series). Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in a normed space X. We say that the series $\sum_{n=1}^{\infty} x_n$ *converges and equals* $x \in X$ if the *partial sums* $s_N = \sum_{n=1}^{N} x_n$ converge to x, i.e., if

$$\lim_{N\to\infty} \|x - s_N\| \;=\; \lim_{N\to\infty} \left\| x - \sum_{n=1}^{N} x_n \right\| \;=\; 0.$$

In this case, we write $x = \sum_{n=1}^{\infty} x_n$, and we also use the shorthands $x = \sum x_n$ or $x = \sum_n x_n$. ♢

In order for an infinite series to converge in X, the *norm of the difference* between x and the partial sum s_N must converge to zero. If we wish to emphasize which norm we are referring to, we may write that $x = \sum x_n$ *converges with respect to* $\|\cdot\|$, or we may say that $x = \sum x_n$ *converges in* X.

Here is another type of convergence notion for series.

Definition 2.6.2. Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in a normed space X. We say that the series $\sum_{n=1}^{\infty} x_n$ is *absolutely convergent* if

$$\sum_{n=1}^{\infty} \|x_n\| \;<\; \infty. \qquad ♢$$

Note that Definition 2.6.2 does not imply that an absolutely convergent series will converge in the sense of Definition 2.6.1. We will prove in the next theorem that if X is complete then every absolutely convergent series in X converges, while if X is not complete then there will exist some vectors $x_n \in X$ such that $\sum \|x_n\| < \infty$ yet $\sum x_n$ does not converge.

Theorem 2.6.3. *If X is a normed space, then the following two statements are equivalent.*

(a) *X is complete (i.e., X is a Banach space).*

(b) *Every absolutely convergent series in X converges in X. That is, if $\{x_n\}_{n\in\mathbb{N}}$ is a sequence in X and $\sum \|x_n\| < \infty$, then the series $\sum x_n$ converges in X.*

Proof. (a) $\Rightarrow$ (b). Assume that X is complete, and suppose that $\sum \|x_n\| < \infty$. Set

$$s_N = \sum_{n=1}^{N} x_n \qquad \text{and} \qquad t_N = \sum_{n=1}^{N} \|x_n\|.$$

If $N > M$, then

$$\|s_N - s_M\| = \left\| \sum_{n=M+1}^{N} x_n \right\| \le \sum_{n=M+1}^{N} \|x_n\| = |t_N - t_M|.$$

Since $\{t_N\}_{N\in\mathbb{N}}$ is a Cauchy sequence of scalars, this implies that $\{s_N\}_{N\in\mathbb{N}}$ is a Cauchy sequence of vectors in X, and hence converges. By definition, this means that $\sum_{n=1}^{\infty} x_n$ converges in X.

(b) $\Rightarrow$ (a). Suppose that every absolutely convergent series in X is convergent. Let $\{x_n\}_{n\in\mathbb{N}}$ be a Cauchy sequence in X. Applying Problem 2.9.9, there exists a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ such that

$$\|x_{n_{k+1}} - x_{n_k}\| < 2^{-k}, \qquad \text{for every } k \in \mathbb{N}.$$

This implies that the series $\sum_{k=1}^{\infty} (x_{n_{k+1}} - x_{n_k})$ is absolutely convergent. Therefore the series $\sum_{k=1}^{\infty} (x_{n_{k+1}} - x_{n_k})$ must converge in X, so we can set

$$x = \sum_{k=1}^{\infty} (x_{n_{k+1}} - x_{n_k}).$$

By definition, this means that the partial sums

$$s_M = \sum_{k=1}^{M} (x_{n_{k+1}} - x_{n_k}) = x_{n_{M+1}} - x_{n_1}$$

converge to x as $M \to \infty$. Setting $y = x + x_{n_1}$, it follows that

$$x_{n_{M+1}} = s_M + x_{n_1} \to x + x_{n_1} = y \qquad \text{as } M \to \infty.$$

Reindexing (replace $M+1$ by k), we conclude that $x_{n_k} \to y$ as $k \to \infty$.

Thus $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence that has a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ that converges to the vector y. Appealing to Problem 1.5.10, this implies that $x_n \to y$. Hence every Cauchy sequence in X converges, so X is complete. $\square$

The ordering of the vectors in a series may be important. If we reorder a series, or in other words consider a new series $\sum_{n=1}^{\infty} x_{\sigma(n)}$ where $\sigma\colon \mathbb{N} \to \mathbb{N}$ is a bijection, there is no guarantee that this reordered series will still converge. If $\sum_{n=1}^{\infty} x_{\sigma(n)}$ does converge for every bijection σ, then we say that the series $\sum_{n=1}^{\infty} x_n$ *converges unconditionally.* A series that converges but does not converge unconditionally is said to be *conditionally convergent.*

For series of *scalars*, unconditional and absolute convergence are equivalent (for one proof of the following result, see [Heil11, Lemma 3.3]).

Lemma 2.6.4. *If $(c_n)_{n\in\mathbb{N}}$ is a sequence of scalars, then*

$$\sum_{n=1}^{\infty} |c_n| \;<\; \infty \qquad\Longleftrightarrow\qquad \sum_{n=1}^{\infty} c_n \text{ converges unconditionally.} \qquad \diamondsuit$$

More generally, if X is a *finite-dimensional* normed space, then an infinite series converges absolutely if and only if it converges unconditionally. The proof of the following theorem on infinite-dimensional spaces is nontrivial. This result is a consequence of the *Dvoretzky–Rogers Theorem* (for one proof, see Theorem 3.33 in [Heil11]).

Theorem 2.6.5. *If X is an infinite-dimensional Banach space, then there exist vectors $x_n \in X$ such that the infinite series $\sum x_n$ converges* unconditionally *but* not absolutely. $\diamondsuit$

2.7 Span and Closed Span

If A is a subset of a vector space X, then the *finite linear span* of A, denoted by $\operatorname{span}(A)$, is the set of all possible *finite linear combinations* of elements of A. If the scalar field is $\mathbb{C}$, then

$$\operatorname{span}(A) \;=\; \biggl\{\sum_{n=1}^{N} c_n x_n \;:\; N > 0,\, x_n \in A,\, c_n \in \mathbb{C}\biggr\}, \tag{2.1}$$

and if the scalar field is $\mathbb{R}$ then we need only replace $\mathbb{C}$ by $\mathbb{R}$ on the preceding line. We often call $\operatorname{span}(A)$ the *finite span*, the *linear span*, or simply the *span* of A. We say that A *spans* X if $\operatorname{span}(A) = X$.

If X is a normed space, then we can take limits of elements, and consider the closure of the span. The closure of $\operatorname{span}(A)$ is called the *closed linear span* or simply the *closed span* of A. For compactness of notation, usually write the closed span as $\overline{\operatorname{span}}(A)$ instead of $\overline{\operatorname{span}(A)}$.

It follows from part (b) of Lemma 1.3.1 that the closed span consists of all limits of elements of the span:

$$\overline{\operatorname{span}}(A) \;=\; \bigl\{y \in X \;:\; \exists\, y_n \in \operatorname{span}(A) \text{ such that } y_n \to y\bigr\}. \tag{2.2}$$

Suppose that $x_n \in A$ and c_n is a scalar for each $n \in \mathbb{N}$, and $x = \sum c_n x_n$ is a convergent infinite series in X. Then the partial sums

$$s_N \;=\; \sum_{n=1}^{N} c_n x_n$$

belong to $\operatorname{span}(A)$ and $s_N \to x$, so it follows from equation (2.2) that x belongs to $\overline{\operatorname{span}}(A)$. As a consequence, we have the inclusion

$$\biggl\{\sum_{n=1}^{\infty} c_n x_n \,:\, x_n \in A,\, c_n \text{ scalar}, \sum_{n=1}^{\infty} c_n x_n \text{ converges}\biggr\} \;\subseteq\; \overline{\operatorname{span}}(A). \tag{2.3}$$

However, *equality need not hold in equation* (2.3). A specific example of such a situation is presented in Example 2.9.5.

According to the following definition, if the closed span of a sequence is the entire space X, then we say that sequence is *complete*.

Definition 2.7.1 (Complete Sequence). Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence of vectors in a normed vector space X. We say that the sequence $\{x_n\}_{n\in\mathbb{N}}$ is *complete* in X if $\operatorname{span}\{x_n\}_{n\in\mathbb{N}}$ is dense in X, i.e., if

$$\overline{\operatorname{span}}\{x_n\}_{n\in\mathbb{N}} \;=\; X.$$

Complete sequences are also known as *total* or *fundamental* sequences. $\diamondsuit$

If M is a subspace of a normed space X, then, by definition, M is finite-dimensional if it has a finite vector space basis (which we will call a *Hamel basis* in Definition 2.8.1). Equivalently, a subspace M is finite-dimensional if and only if we can write $M = \operatorname{span}\{x_1, \dots, x_n\}$ for some finitely many vectors $x_1, \dots, x_n \in X$. The following result states that every finite-dimensional subspace of a normed space is *closed*. For a proof of this theorem, see [Con90, Prop. III.3.3] or [Heil18, Cor. 3.7.3].

Theorem 2.7.2. *If M is a finite-dimensional subspace of a normed space X, then M is a closed subset of X.* $\diamondsuit$

2.8 Hamel Bases and Schauder Bases

A *Hamel Basis* is simply another name for the usual notion of a basis for a vector space. We formalize this in the next definition.

Definition 2.8.1 (Hamel Basis). Let V be a vector space. A *Hamel basis*, *vector space basis*, or simply a *basis* for V is a set $\mathcal{B} \subseteq V$ such that

(a) $\mathcal{B}$ is linearly independent, and

(b) $\operatorname{span}(\mathcal{B}) = V$. $\diamondsuit$

The span of $\mathcal{B}$ was defined in equation (2.1) to be the set of all finite linear combinations of vectors from $\mathcal{B}$. Likewise, independence is defined in terms of finite linear combinations. Specifically, $\mathcal{B}$ is linearly independent if for each choice of finitely many distinct vectors $x_1, \dots, x_N \in \mathcal{B}$ we have

$$\sum_{n=1}^{N} c_n x_n \;=\; 0 \qquad \Longleftrightarrow \qquad c_1 = \cdots = c_N = 0.$$

Thus the definition of a Hamel basis, as given in Definition 2.8.1, is made entirely in terms of *finite linear combinations* of vectors. This "vector space" notion of a basis is most useful when V is a *finite-dimensional* vector space. One consequence of the *Baire Category Theorem* is that *any Hamel basis for an infinite-dimensional Banach space must be uncountable* (for one proof, see [Heil11, Sec. 4.1]).

In an infinite-dimensional Banach space, instead of restricting to just finite linear combinations, it is usually preferable to employ "infinite linear combinations." This idea leads us to the following definition of a *Schauder basis* for a Banach space.

Definition 2.8.2 (Schauder Basis). Let X be a Banach space. A countably infinite sequence $\{x_n\}_{n\in\mathbb{N}}$ of elements of X is a *Schauder basis* for X if for each vector $x \in X$ there exist *unique* scalars $c_n(x)$ such that

$$x \;=\; \sum_{n=1}^{\infty} c_n(x)\, x_n, \tag{2.4}$$

where this series converges in the norm of X. $\diamondsuit$

Recall from Definition 2.6.1 that equation (2.4) means that the partial sums of the series converge to x in the norm of X, i.e.,

$$\lim_{N\to\infty} \left\| x \;-\; \sum_{n=1}^{N} c_n(x)\, x_n \right\| \;=\; 0.$$

Example 2.8.3. Let $\mathcal{E} = \{\delta_n\}_{n\in\mathbb{N}}$ be the sequence of standard basis vectors, and let x be any element of the space ℓ^1. Then we can write

$$x \;=\; (x_n)_{n\in\mathbb{N}} \;=\; (x_1, x_2, \dots)$$

where the x_n are scalars such that

$$\sum_{n=1}^{\infty} |x_n| \;<\; \infty.$$

We claim that

$$x \;=\; \sum_{n=1}^{\infty} x_n \delta_n, \tag{2.5}$$

where this series converges in ℓ^1-norm. To see why, let

$$s_N \;=\; \sum_{n=1}^{N} x_n \delta_n \;=\; (x_1, \dots, x_N, 0, 0, \dots)$$

be the Nth partial sum of the series. Then

$$x - s_N \;=\; (0, \dots, 0, x_{N+1}, x_{N+2}, \dots).$$

Therefore $\|x - s_N\|_1 = \sum_{n=N+1}^{\infty} |x_n|$, and since $\sum |x_n| < \infty$, it follows that

$$\lim_{N\to\infty} \|x - s_N\|_1 \;=\; \lim_{N\to\infty} \biggl\|x - \sum_{n=1}^{N} x_n \delta_n\biggr\|_1 \;=\; \lim_{N\to\infty} \sum_{n=N+1}^{\infty} |x_n| \;=\; 0.$$

This shows that the partial sums s_N converge to x in ℓ^1-norm, which proves that equation (2.5) holds. Further, the *only* way to write $x = \sum c_n \delta_n$ is with $c_n = x_n$ for every n (consider Problem 2.9.11), so we conclude that $\mathcal{E}$ is a Schauder basis. $\diamondsuit$

The question of whether every separable Banach space has a Schauder basis was a longstanding open problem known as the *Basis Problem.* It was finally shown by Enflo [Enf73] that there exist separable Banach spaces that have no Schauder bases!

For an introduction to the theory of Schauder bases and related topics and generalizations, we refer to [Heil11].

2.9 The Space $C_b(X)$

Let X be a metric space, and let $C_b(X)$ be the space of all bounded, continuous functions $f\colon X \to \mathbb{C}$ (entirely analogous statements if we prefer to restrict our attention to real-valued functions only). The *uniform norm* of a function $f \in C_b(X)$ is

$$\|f\|_{\mathrm{u}} \;=\; \sup_{x\in X} |f(x)|.$$

This is a norm on $C_b(X)$.

Convergence with respect to the norm $\|\cdot\|_{\mathrm{u}}$ is called *uniform convergence*. That is, if $f_n,\ f \in C_b(X)$ and

$$\lim_{n\to\infty} \|f - f_n\|_{\mathrm{u}} \;=\; \lim_{n\to\infty} \left(\sup_{x\in X} |f(x) - f_n(x)| \right) \;=\; 0,$$

then we say that f_n *converges uniformly to* f. Similarly, a sequence $\{f_n\}_{n\in\mathbb{N}}$ that is Cauchy with respect to the uniform norm is said to be a *uniformly Cauchy sequence*.

Suppose that f_n converges uniformly to f. Then, by the definition of a supremum, for each individual point x we have

$$\lim_{n\to\infty} |f(x) - f_n(x)| \;\le\; \lim_{n\to\infty} \|f - f_n\|_{\mathrm{u}} \;=\; 0.$$

Thus $f_n(x) \to f(x)$ for each individual point $x \in X$. We have a name for this latter type of convergence. Specifically, we say that f_n *converges pointwise to* f if $f_n(x) \to f(x)$ for each $x \in X$. Our argument above shows that uniform convergence implies pointwise convergence. However, the converse need not hold, i.e., pointwise convergence does not imply uniform convergence in general. Here is an example.

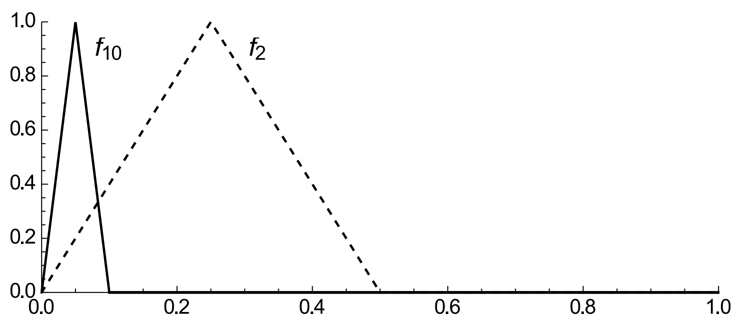

Fig. 2.1 Graphs of the functions f_2 (dashed) and f_{10} (solid) from Example 2.9.1.

Example 2.9.1 (Shrinking Triangles). Set $X = [0,1]$. For each $n \in \mathbb{N}$, let f_n be the continuous function on $[0,1]$ defined by

$$f_n(x) \;=\; \begin{cases} 0, & x = 0, \\ \text{linear}, & 0 < x < \frac{1}{2n}, \\ 1, & x = \frac{1}{2n}, \\ \text{linear}, & \frac{1}{2n} < x < \frac{1}{n}, \\ 0, & \frac{1}{n} \le x \le 1. \end{cases}$$

Then $f_n(x) \to 0$ for each point $x \in [0,1]$ (see the illustration in Figure 2.1). Hence f_n converges pointwise to the zero function. However, f_n does not converge uniformly to the zero function on $[0,1]$ because no matter what $n \in \mathbb{N}$ that we consider, we have

$$\|0 - f_n\|_{\mathrm{u}} \;=\; \sup_{x\in[0,1]} |0 - f_n(x)| \;=\; 1. \qquad \diamondsuit$$

According to the next lemma, the uniform limit of a sequence of continuous functions is continuous.

Lemma 2.9.2. *Let X be a metric space. If we have functions $f_n \in C_b(X)$ for $n \in \mathbb{N}$, and $f\colon X \to \mathbb{C}$ is a function such that f_n converges uniformly to f, then $f \in C_b(X)$.*

Proof. Let x be any fixed point in X, and choose $\varepsilon > 0$. Then, by the definition of uniform convergence, there exists some integer $n > 0$ such that

$$\|f - f_n\|_{\mathrm{u}} \;<\; \varepsilon. \tag{2.6}$$

In fact, equation (2.6) will be satisfied for all large enough n, but we need only one particular n for this proof.

The function f_n is continuous, so there is a $\delta > 0$ such that for all $y \in I$ we have

$$\mathrm{d}(x,y) \;<\; \delta \quad\implies\quad |f_n(x) - f_n(y)| \;<\; \varepsilon.$$

Consequently, if $y \in X$ is any point that satisfies $\mathrm{d}(x,y) < \delta$, then

$$\begin{aligned} |f(x) - f(y)| \;&\le\; |f(x) - f_n(x)| \;+\; |f_n(x) - f_n(y)| \;+\; |f_n(y) - f(y)| \\ &\le\; \|f - f_n\|_{\mathrm{u}} \;+\; \varepsilon \;+\; \|f_n - f\|_{\mathrm{u}} \\ &<\; \varepsilon + \varepsilon + \varepsilon \;=\; 3\varepsilon. \end{aligned} \tag{2.7}$$

This proves that f is continuous. To see that f is bounded, observe that f_n is bounded because it belongs to $C_b(X)$, and $f - f_n$ is bounded because $\|f - f_n\|_{\mathrm{u}} < \varepsilon$. Since the sum of two bounded functions is bounded, we conclude that $f = f_n + (f - f_n)$ is bounded. Thus f is both bounded and continuous, so f belongs to $C_b(X)$. $\square$

Next, we use Lemma 2.9.2 to prove that $C_b(X)$ *is complete with respect to the uniform norm.*

Theorem 2.9.3. *If $\{f_n\}_{n\in\mathbb{N}}$ is a sequence in $C_b(X)$ that is Cauchy with respect to $\|\cdot\|_{\mathrm{u}}$, then there exists a function $f \in C_b(X)$ such that f_n converges to f uniformly. Consequently $C_b(X)$ is a Banach space with respect to the uniform norm.*

Proof. Suppose that $\{f_n\}_{n\in\mathbb{N}}$ is a uniformly Cauchy sequence in $C_b(X)$. If we fix any particular point $x \in X$, then for all m and n we have

$$|f_m(x) - f_n(x)| \;\le\; \|f_m - f_n\|_{\mathrm{u}}.$$

It follows that $\{f_n(x)\}_{n\in\mathbb{N}}$ is a Cauchy sequence of scalars. Since $\mathbb{R}$ and $\mathbb{C}$ are complete, this sequence of scalars must converge. Define

$$f(x) \;=\; \lim_{n\to\infty} f_n(x), \qquad \text{for } x \in X.$$

Now choose $\varepsilon > 0$. Then there exists an N such that $\|f_m - f_n\|_{\mathrm{u}} < \varepsilon$ for all $m, n \ge N$. If $n \ge N$, then we have for every $x \in X$ that

$$|f(x) - f_n(x)| \;=\; \lim_{m\to\infty} |f_m(x) - f_n(x)| \;\le\; \lim_{m\to\infty} \|f_m - f_n\|_{\mathrm{u}} \;\le\; \varepsilon.$$

Hence $\|f - f_n\|_{\mathrm{u}} \le \varepsilon$ whenever $n \ge N$. This shows that $f_n \to f$ uniformly as $n \to \infty$. Consequently Lemma 2.9.2 implies that $f \in C_b(X)$. Thus every uniformly Cauchy sequence in $C_b(X)$ converges uniformly to a function in $C_b(X)$, so $C_b(X)$ is complete with respect to the uniform norm. □

The *Weierstrass Approximation Theorem* is an important result about $C_b(X)$ when $X = [a,b]$ is a finite closed interval in the real line. This theorem, which we state next, tells us that the set of polynomial functions is a dense subset of $C[a,b]$. There are many different proofs of the Weierstrass Approximation Theorem; for example, see [Rud76, Thm. 7.26] or [Heil18, Thm. 4.6.2].

Theorem 2.9.4 (Weierstrass Approximation Theorem). *If f is a function in $C[a,b]$, then for each $\varepsilon > 0$ there exists some polynomial*

$$p(x) \;=\; \sum_{k=0}^{N} c_k x^k$$

such that

$$\|f - p\|_{\mathrm{u}} \;=\; \sup_{x\in[a,b]} |f(x) - p(x)| \;<\; \varepsilon. \qquad \diamondsuit$$

We will use the Weierstrass Approximation Theorem to help illustrate the difference between a Hamel basis, a complete sequence, and a Schauder basis.

Example 2.9.5. (a) Let

$$\mathcal{M} \;=\; \{x^k\}_{k=0}^{\infty} \;=\; \{1, x, x^2, \dots\}.$$

By definition, a polynomial is a finite linear combination of the monomials $1, x, x^2, \dots$, so the set of polynomials is the finite linear span of $\mathcal{M}$:

$$\mathcal{P} \;=\; \mathrm{span}(\mathcal{M}).$$

Since

$$\sum_{k=0}^{N} c_k x^k \;=\; 0 \qquad \Longleftrightarrow \qquad c_0 = \cdots = c_N = 0,$$

the set of monomials $\mathcal{M}$ is finitely linearly independent. Hence $\mathcal{M}$ is a Hamel basis for $\mathcal{P}$.

(b) Restricting our attention to functions on the domain $[a,b]$, we have that $\operatorname{span}(\mathcal{M}) = \mathcal{P}$ is a proper subset of $C[a,b]$. Therefore $\mathcal{M}$ is not a Hamel basis for $C[a,b]$. On the other hand, by combining Theorem 2.9.4 with Lemma 1.3.1 we see that $\operatorname{span}(\mathcal{M})$ is dense in $C[a,b]$, i.e.,

$$\overline{\operatorname{span}}(\mathcal{M}) \;=\; C[a,b].$$

Thus $\mathcal{M}$ is a *complete sequence* in $C[a,b]$ in the sense of Definition 2.7.1.

(c) Even though $\mathcal{M}$ is complete and finitely linearly independent, not every function $f \in C[a,b]$ can be written as

$$f(x) \;=\; \sum_{k=0}^{\infty} \alpha_k x^k. \tag{2.8}$$

A series of this form is called a *power series*, and if it converges at some point x, then it converges absolutely for all points t with $|t| < r$ where $r = |x|$. In fact (see Problem 2.9.18), the function f so defined is infinitely differentiable on $(-r,r)$. Therefore a continuous function that is not infinitely differentiable cannot be written as a power series. Although such a function belongs to the closed span of $\mathcal{M}$, it cannot be written in the form given in equation (2.8). Consequently $\mathcal{M}$ is *not* a Schauder basis for $C[a,b]$. $\diamondsuit$

Problems

2.9.6. Prove Lemma 2.3.1.

2.9.7. Prove that every open ball $B_r(x)$ in a normed space X is convex.

2.9.8. Let Y be a subspace of a Banach space X, and let the norm on Y be the restriction of the norm on X to the set Y. Prove that Y is a Banach space with respect to this norm if and only if Y is a closed subset of X.

2.9.9. Suppose that $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in a normed space X. Show that there exists a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ such that $\|x_{n_{k+1}} - x_{n_k}\| < 2^{-k}$ for every $k \in \mathbb{N}$.

2.9.10. Suppose that $A = \{x_n\}_{n\in\mathbb{N}}$ is a Schauder basis for a Banach space X. Prove that equality holds in equation (2.3).

2.9.11. Suppose that $\{x_n\}_{n\in\mathbb{N}}$ is a sequence of vectors in ℓ^1. For each n, write the components of x_n as

$$x_n \;=\; \bigl(x_n(k)\bigr)_{k\in\mathbb{N}} \;=\; \bigl(x_n(1), x_n(2), \dots\bigr).$$

Suppose further that

$$x \;=\; \bigl(x(k)\bigr)_{k\in\mathbb{N}} \;=\; \bigl(x(1), x(2), \dots\bigr)$$

is such that x_n converges to x in ℓ^1-norm, i.e., $\|x - x_n\|_1 \to 0$ as $n \to \infty$. Prove that for each fixed k, the kth component of x_n converges to the kth component of x, i.e.,

$$\lim_{k\to\infty} x_n(k) \;=\; x(k), \qquad k \in \mathbb{N}.$$

Thus, convergence in ℓ^1-norm implies *componentwise convergence*.

2.9.12. The closed unit disk in ℓ^1 is $D = \bigl\{x \in \ell^1 : \|x\|_1 \le 1\bigr\}$.

(a) Prove that D is a closed and bounded subset of ℓ^1.

(b) Prove that the set of standard basis vectors $\{\delta_n\}_{n\in\mathbb{N}}$ contains no convergent subsequences. Conclude that D is not sequentially compact, and therefore (by Theorem 1.4.5) D is not compact.

2.9.13. Let S be the set of all "finite sequences" whose components are rational, i.e.,

$$S \;=\; \Bigl\{(r_1, \dots, r_N, 0, 0, \dots) : N > 0,\ r_1, \dots, r_N \text{ rational}\Bigr\},$$

where we say that a complex number is rational if both its real and imaginary parts are rational. Prove that S is a countable, dense subset of ℓ^1. Conclude that ℓ^1 is separable.

2.9.14. Suppose that X is a Banach space that has a Schauder basis $\{x_n\}_{n\in\mathbb{N}}$. Prove that X must be separable.

2.9.15. For the following choices of functions g_n, determine whether the sequence $\{g_n\}_{n\in\mathbb{N}}$ is: pointwise convergent, uniformly convergent, uniformly Cauchy, or bounded with respect to the uniform norm.

(a) $g_n(x) = n f_n(x)$, $x \in [0,1]$, where f_n is the Shrinking Triangle from Example 2.9.1.

(b) $g_n(x) = e^{-n|x|}$, $x \in \mathbb{R}$.

(c) $g_n(x) = x e^{-n|x|}$, $x \in \mathbb{R}$.

(d) $g_n(x) = \dfrac{nx}{1 + n^2 x^2}$, $x \in \mathbb{R}$.

2.9.16. The space of continuous functions that "vanish at infinity" is

$$C_0(\mathbb{R}) \;=\; \Bigl\{f \in C(\mathbb{R}) \,:\, \lim_{x\to\pm\infty} f(x) = 0\Bigr\}.$$

Prove the following statements.

(a) $C_0(\mathbb{R})$ is a closed subspace of $C_b(\mathbb{R})$ with respect to the uniform norm, and is therefore a Banach space with respect to $\|\cdot\|_{\mathrm{u}}$.

(b) Every function in $C_0(\mathbb{R})$ is uniformly continuous on $\mathbb{R}$, but there exist functions in $C_b(\mathbb{R})$ that are not uniformly continuous.

(c) If $f \in C_0(\mathbb{R})$ and $g(x) = f(x-k)$ where k is fixed, then $f \in C_0(\mathbb{R})$ and $\|g\|_{\mathrm{u}} = \|f\|_{\mathrm{u}}$.

(d) The "closed unit ball" $D = \{f \in C_0(\mathbb{R}) : \|f\|_{\mathrm{u}} \le 1\}$ in $C_0(\mathbb{R})$ is closed and bounded, but it is not sequentially compact.

2.9.17. Let

$$f_n(x) \;=\; \frac{e^{-n^2x^2}}{n^2}, \qquad \text{for } x \in \mathbb{R}.$$

(a) Prove that the series $\sum_{n=1}^{\infty} f_n$ converges absolutely in $C_b(\mathbb{R})$ with respect to the uniform norm, and therefore converges since $C_b(\mathbb{R})$ is a Banach space.

(b) Prove that the series $\sum_{n=1}^{\infty} f_n'$ converges in $C_b(\mathbb{R})$ with respect to the uniform norm, but it does not converge absolutely with respect to that norm.

2.9.18. Let $(c_k)_{k\ge 0}$ be a fixed sequence of scalars. Suppose that the series $\sum_{k=0}^{\infty} c_k x^k$ converges for some number $x \in \mathbb{R}$, and set $r = |x|$. Prove the following statements.

(a) The series $f(t) = \sum_{k=0}^{\infty} c_k t^k$ converges absolutely for all $t \in (-r,r)$. That is,

$$\sum_{k=0}^{\infty} |c_k|\,|t|^k \;<\; \infty, \qquad \text{for } |t| < r.$$

(b) f is infinitely differentiable on the interval $(-r,r)$.

Chapter 3
Inner Products and Hilbert Spaces

In a normed vector space, each vector has an assigned *length*, and from this we obtain the *distance* from x to y as the length of the vector $x - y$. For vectors in $\mathbb{R}^d$ and $\mathbb{C}^d$ we also know how to measure the *angle* between vectors. In particular, two vectors x and y in $\mathbb{R}^d$ or $\mathbb{C}^d$ are perpendicular, or *orthogonal*, if their dot product is zero. An inner product is a generalization of the dot product, and it provides us with a generalization of the notion of orthogonality to vector spaces other than $\mathbb{R}^d$ or $\mathbb{C}^d$.

3.1 The Definition of an Inner Product

Here are the defining properties of an inner product.

Definition 3.1.1 (Semi-Inner Product, Inner Product). Let H be a vector space over either the real field $\mathbb{R}$ or the complex field $\mathbb{C}$. A *semi-inner product* on H is a scalar-valued function $\langle \cdot, \cdot \rangle$ on $H \times H$ such that for all vectors $x, y, z \in H$ and all scalars a, b we have:

(a) Nonnegativity: $\langle x, x \rangle \ge 0$,

(b) Conjugate Symmetry: $\langle x, y \rangle = \overline{\langle y, x \rangle}$, and

(c) Linearity in the First Variable: $\langle ax + by, z \rangle = a\langle x, z \rangle + b\langle y, z \rangle$.

If a semi-inner product $\langle \cdot, \cdot \rangle$ also satisfies:

(d) Uniqueness: $\langle x, x \rangle = 0$ if and only if $x = 0$,

then it is called an *inner product* on H. In this case, H is called an *inner product space* or a *pre-Hilbert space*. $\diamondsuit$

The usual dot product

$$u \cdot v \;=\; u_1\overline{v_1} + \cdots + u_n\overline{v_n} \tag{3.1}$$

is an inner product on $\mathbb{R}^d$ (using real scalars) and on $\mathbb{C}^n$ (using complex scalars). Of course, if the scalar field is $\mathbb{R}$, then the complex conjugate in equation (3.1) is superfluous. Similarly, if H is a real vector space then the complex conjugate in the definition of conjugate symmetry is irrelevant since $\langle x, y\rangle$ will be real.

If $\langle\cdot,\cdot\rangle$ is a semi-inner product on a vector space H, then we define

$$\|x\| \;=\; \langle x, x\rangle^{1/2}, \quad \text{for } x \in H.$$

We will prove in Lemma 3.2.3 that $\|\cdot\|$ is a seminorm on H, and therefore we refer to $\|\cdot\|$ as the *seminorm induced by* $\langle\cdot,\cdot\rangle$. Likewise, if $\langle\cdot,\cdot\rangle$ is an inner product, then we will prove that $\|\cdot\|$ is a norm on H, and in this case we refer to $\|\cdot\|$ as the *norm induced by* $\langle\cdot,\cdot\rangle$. It may be possible to place other norms on H, but unless we explicitly state otherwise, we assume that all norm-related statements on an inner product space are taken with respect to the induced norm.

3.2 Properties of an Inner Product

Here are some properties of inner products (the proof is assigned as Problem 3.9.4).

Lemma 3.2.1. *If $\langle\cdot,\cdot\rangle$ is a semi-inner product on a vector space H, then the following statements hold for all vectors x, y, $z \in H$ and all scalars a and b.*

(a) Antilinearity in the Second Variable: $\langle x, ay + bz\rangle = \overline{a}\,\langle x, y\rangle + \overline{b}\,\langle x, z\rangle$.

(b) Polar Identity: $\|x+y\|^2 = \|x\|^2 + 2\,\mathrm{Re}(\langle x, y\rangle) + \|y\|^2$.

(c) Pythagorean Theorem: *If* $\langle x, y\rangle = 0$, *then* $\|x \pm y\|^2 = \|x\|^2 + \|y\|^2$.

(d) Parallelogram Law: $\|x+y\|^2 + \|x-y\|^2 = 2\big(\|x\|^2 + \|y\|^2\big)$. $\diamondsuit$

The inequality that we prove next is variously known as the *Schwarz*, *Cauchy–Schwarz*, *Cauchy–Bunyakovski–Schwarz*, or *CBS Inequality*.

Theorem 3.2.2 (Cauchy–Bunyakovski–Schwarz Inequality). *If $\langle\cdot,\cdot\rangle$ is a semi-inner product on a vector space H, then*

$$|\langle x, y\rangle| \;\le\; \|x\|\,\|y\|, \qquad \textit{for all } x, y \in H.$$

Proof. If $x = 0$ or $y = 0$ then there is nothing to prove, so suppose that x and y are both nonzero. Let $\alpha \in \mathbb{C}$ be a scalar such that $|\alpha| = 1$ and

$$\langle x, y\rangle \;=\; \alpha\,|\langle x, y\rangle|.$$

Then for each $t \in \mathbb{R}$, the Polar Identity implies that

$$\begin{aligned} 0 \ \le \ \|x-\alpha t y\|^2 \ &= \ \|x\|^2 \ - \ 2\,\mathrm{Re}\bigl(\langle x, \alpha t y\rangle\bigr) \ + \ t^2\,\|y\|^2 \\ &= \ \|x\|^2 \ - \ 2t\,\mathrm{Re}\bigl(\overline{\alpha}\,\langle x, y\rangle\bigr) \ + \ t^2\,\|y\|^2 \\ &= \ \|x\|^2 \ - \ 2t\,|\langle x, y\rangle| \ + \ t^2\,\|y\|^2. \end{aligned}$$

This is a real-valued quadratic polynomial in the variable t. Since this polynomial is nonnegative, it can have at most one real root. This requires that the discriminant of the polynomial be at most zero, so we must have

$$\bigl(-2\,|\langle x, y\rangle|\bigr)^2 \ - \ 4\,\|x\|^2\,\|y\|^2 \ \le \ 0.$$

The desired result follows by rearranging this inequality. □

By combining the Polar Identity with the Cauchy–Bunyakovski–Schwarz Inequality, we can now prove that the induced seminorm $\|\cdot\|$ satisfies the Triangle Inequality.

Lemma 3.2.3. *Let H be a vector space. If $\langle\cdot,\cdot\rangle$ is a semi-inner product on H, then $\|\cdot\|$ is a seminorm on H, and if $\langle\cdot,\cdot\rangle$ is an inner product on H, then $\|\cdot\|$ is a norm on H.*

Proof. The only property of a seminorm that is not obvious is the Triangle Inequality. To prove this, we use the Polar Identity and the Cauchy–Bunyakovski–Schwarz Inequality to compute that

$$\begin{aligned} \|x+y\|^2 \ &= \ \|x\|^2 \ + \ 2\,\mathrm{Re}\langle x, y\rangle \ + \ \|y\|^2 \\ &\le \ \|x\|^2 \ + \ 2\,|\langle x, y\rangle| \ + \ \|y\|^2 \\ &\le \ \|x\|^2 \ + \ 2\,\|x\|\,\|y\| \ + \ \|y\|^2 \\ &= \ \bigl(\|x\| \ + \ \|y\|\bigr)^2. \end{aligned}$$

The Triangle Inequality follows by taking square roots. □

3.3 Hilbert Spaces

The question of whether every Cauchy sequence in a given inner product space must converge is very important, just as it is in a metric or normed space. We give the following name to those inner product spaces that have this property.

Definition 3.3.1 (Hilbert Space). An inner product space H is called a *Hilbert space* if it is complete with respect to the induced norm. ♢

Thus, an inner product space is a Hilbert space if and only if every Cauchy sequence in H converges to an element of H. Equivalently, a Hilbert space is an inner product space that is a Banach space with respect to the induced norm.

Using real scalars, $\mathbb{R}^d$ is a Hilbert space with respect to the usual dot product. Likewise, $\mathbb{C}^d$ is a Hilbert space with respect to the dot product if we use complex scalars.

Here is an example of an infinite-dimensional Hilbert space (compare this to Example 1.1.2, which introduced the space ℓ^1).

Example 3.3.2. For each sequence of real or complex scalars

$$x = (x_k)_{k\in\mathbb{N}} = (x_1, x_2, \dots),$$

we define the ℓ^2-*norm of* x to be

$$\|x\|_2 = \|(x_k)_{k\in\mathbb{N}}\|_2 = \left(\sum_{k=1}^{\infty} |x_k|^2\right)^{1/2}. \tag{3.2}$$

Note that we have not yet proved that $\|\cdot\|_2$ is a norm in any sense; we will address this issue below.

We say that a sequence $x = (x_k)_{k\in\mathbb{N}}$ is *square-summable* if $\|x\|_2 < \infty$, and we let ℓ^2 denote the space of all square summable sequences, i.e.,

$$\ell^2 = \left\{ x = (x_k)_{k\in\mathbb{N}} \;:\; \|x\|_2^2 = \sum_{k=1}^{\infty} |x_k|^2 < \infty \right\}.$$

We will define an inner product on ℓ^2. First we recall that the arithmetic-geometric mean inequality implies that if $a, b \ge 0$, then

$$ab \le \frac{a^2}{2} + \frac{b^2}{2}.$$

Consequently, if we choose any two vectors $x = (x_k)_{k\in\mathbb{N}}$ and $y = (y_k)_{k\in\mathbb{N}}$ in ℓ^2, then

$$\sum_{k=1}^{\infty} |x_k y_k| \le \sum_{k=1}^{\infty} \left(\frac{|x_k|^2}{2} + \frac{|y_k|^2}{2}\right) = \frac{\|x\|_2^2}{2} + \frac{\|y\|_2^2}{2} < \infty. \tag{3.3}$$

Therefore we can define

$$\langle x, y\rangle = \sum_{k=1}^{\infty} x_k \overline{y_k}, \qquad \text{for } x, y \in \ell^2, \tag{3.4}$$

because equation (3.3) tells us that this series of scalars converges absolutely. In particular, for $x = y$ we have

$$\langle x, x\rangle \;=\; \sum_{k=1}^{\infty} |x_k|^2 \;=\; \|x\|_2^2. \tag{3.5}$$

We can check that the function $\langle\cdot,\cdot\rangle$ satisfies all of the requirements of an inner product. For example, simply by definition we have $0 \le \langle x,x\rangle < \infty$ for every $x \in \ell^2$. Further, if $\langle x,x\rangle = 0$ then it follows from equation (3.5) that $x_k = 0$ for every k, so $x = 0$. This establishes the nonnegativity and uniqueness requirements of an inner product, and the conjugate symmetry and linearity in the first variable requirements are easily checked as well.

Thus $\langle\cdot,\cdot\rangle$ is an inner product on ℓ^2, and therefore ℓ^2 is an inner product space with respect to this inner product. Further, equation (3.5) shows that the norm induced from this inner product is precisely the ℓ^2-norm $\|\cdot\|_2$, so Lemma 3.2.3 implies that $\|\cdot\|_2$ really is a norm on ℓ^2.

Is ℓ^2 a Hilbert space? This is slightly more difficult to check, but an argument very similar to the one used in Theorem 1.2.6 shows that every Cauchy sequence in ℓ^2 converges to an element of ℓ^2 (we assign the details as Problem 3.9.9). Consequently ℓ^2 is complete, so it is a Hilbert space. $\diamondsuit$

Observe that ℓ^1 is a proper subspace of ℓ^2. For, if $x = (x_k)_{k\in\mathbb{N}}$ belongs to ℓ^1 then we must have $|x_k| < 1$ for all large enough k, say $k \ge N$, and therefore

$$\sum_{k=N}^{\infty} |x_k|^2 \;\le\; \sum_{k=N}^{\infty} |x_k| \;\le\; \|x\|_1 \;<\; \infty.$$

Hence $x \in \ell^2$. On the other hand, the sequence

$$x \;=\; \bigl(\tfrac{1}{k}\bigr)_{k\in\mathbb{N}} \;=\; \bigl(1, \tfrac{1}{2}, \tfrac{1}{3}, \dots\bigr)$$

belongs to ℓ^2, but it does not belong to ℓ^1. Now, since ℓ^1 is a subspace of ℓ^2, we can define an inner product on ℓ^1 simply by restricting equation (3.4) to vectors in ℓ^1. However, the norm induced from this inner product is $\|\cdot\|_2$ rather than $\|\cdot\|_1$. Although ℓ^1 is complete with respect to the norm $\|\cdot\|_1$, it is not complete with respect to the norm $\|\cdot\|_2$ (see Problem 3.9.10). Therefore ℓ^1 is an inner product space with respect to the inner product defined in equation (3.4), *but it is not a Hilbert space.*

Here is a different example of an inner product space.

Example 3.3.3. Let $C[a,b]$ be the space of all continuous functions of the form $f\colon [a,b] \to \mathbb{C}$. All functions in $C[a,b]$ are bounded and Riemann integrable. If we choose $f, g \in C[a,b]$, then the product $f(x)\,\overline{g(x)}$ is continuous and Riemann integrable, so we can define

$$\langle f, g\rangle \;=\; \int_a^b f(x)\,\overline{g(x)}\,dx, \qquad \text{for } f, g \in C[a,b]. \tag{3.6}$$

We can easily see that $\langle\cdot,\cdot\rangle$ satisfies the nonnegativity, conjugate symmetry, and linearity in the first variable requirements stated in Definition 3.1.1, and hence is at least a semi-inner product on $C[a,b]$. According to Problem 3.9.8, the uniqueness requirement also holds, and therefore $\langle\cdot,\cdot\rangle$ is an inner product on $C[a,b]$. The norm induced from this inner product is

$$\|f\|_2 \;=\; \langle f,f\rangle^{1/2} \;=\; \left(\int_a^b |f(x)|^2\,dx\right)^{1/2}, \qquad \text{for } f \in C[a,b].$$

We call $\|f\|_2$ the L^2-*norm* of the function f. $\diamondsuit$

Although we have defined an inner product on $C[a,b]$, we will prove that $C[a,b]$ is not complete with respect to the norm $\|\cdot\|_2$ that is induced from that inner product (in contrast, note that Theorem 2.9.3 implies that $C[a,b]$ is complete with respect to a different norm—the *uniform norm* $\|\cdot\|_{\mathrm{u}}$).

Lemma 3.3.4. *$C[a,b]$ is not a Hilbert space with respect to $\langle\cdot,\cdot\rangle$.*

Proof. For simplicity of presentation we will take $a=-1$ and $b=1$. For each $n\in\mathbb{N}$ let f_n be the continuous function (illustrated in Figure 3.1 for the case $n=5$) defined by

$$f_n(x) \;=\; \begin{cases} -1, & -1\le x\le -\frac{1}{n},\\ \text{linear}, & -\frac{1}{n}<x<\frac{1}{n},\\ 1, & \frac{1}{n}\le x\le 1; \end{cases}$$

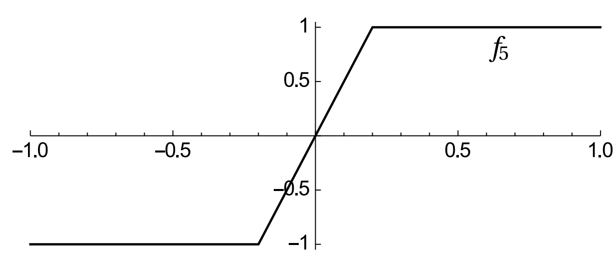

Fig. 3.1 Graph of the function f_5.

For each $m<n$, we have that

$$\begin{aligned} \|f_m-f_n\|_2^2 \;&=\; \int_{-1}^{1} |f_m(x)-f_n(x)|^2\,dx\\ &=\; \int_{-1/m}^{1/m} |g(x)-f_n(x)|^2\,dx\\ &\le\; \int_{-1/m}^{1/m} 1\,dx \;=\; \frac{2}{m}. \end{aligned}$$

Therefore, if we fix $\varepsilon > 0$ then for all large enough m and n we will have $\|f_m - f_n\|_2 < \varepsilon$. This shows that $\{f_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $C[-1,1]$. However, we will prove that there is *no function* $g \in C[-1,1]$ such that $\|g - f_n\|_2 \to 0$ as $n \to \infty$.

Suppose that there were some $g \in C[-1,1]$ such that $\|g - f_n\|_2 \to 0$. Fix $0 < c < 1$, and suppose that g is not identically 1 on $[c,1]$. Then

$$C \;=\; \int_c^1 |g(x) - 1|\,dx \;>\; 0.$$

On the other hand, f_n is identically 1 on $[c,1]$ for all $n > 1/c$, so for all large enough n we have

$$\begin{aligned}\|g - f_n\|_2^2 \;&=\; \int_{-1}^1 |g(x) - f_n(x)|^2\,dx \\ &\geq\; \int_c^1 |g(x) - f_n(x)|^2\,dx \\ &=\; \int_c^1 |g(x) - 1|\,dx \;\geq\; C.\end{aligned}$$

Since $C > 0$, it follows that $\|g - f_n\|_2 \not\to 0$, which is a contradiction. Therefore g must be identically 1 on $[c,1]$. This is true for every $c > 1$, so $g(x) = 1$ for all $0 < x \leq 1$. A similar argument shows that $g(x) = -1$ for all $-1 \leq x < 0$. However, there is no continuous function that takes these values, so we have obtained a contradiction.

Thus, although $\{f_n\}_{n\in\mathbb{N}}$ is Cauchy in $C[-1,1]$, there is no function in $C[-1,1]$ that this sequence can converge to with respect to the induced norm $\|\cdot\|_2$. Therefore $C[-1,1]$ is not complete with respect to this norm. □

Every incomplete normed space that we have encountered prior to $C[a,b]$ has been contained in some larger complete space. For example, c_{00} is incomplete with respect to the norm $\|\cdot\|_1$, but it is a subset of ℓ^1, which is complete with respect to $\|\cdot\|_1$. Likewise, $C_c(\mathbb{R})$ is incomplete with respect to the uniform norm, yet it is contained in $C_0(\mathbb{R})$, which is a Banach space with respect to that norm. It is likewise true that the inner product space $C[a,b]$ is contained in a larger Hilbert space, the *Lebesgue space* $L^2[a,b]$. However, defining $L^2[a,b]$ is beyond the scope of this manuscript, as it requires the theory of the *Lebesgue integral*. We refer to texts such as [Fol99], [SS05], [WZ77], or [Heil19] for discussion of the Lebesgue integral and the Hilbert space $L^2[a,b]$.

3.4 Orthogonal and Orthonormal Sets

Orthogonality plays a important role in the analysis of inner product spaces. We declare that two vectors are orthogonal if their inner product is zero (consequently, the zero vector is orthogonal to every other vector). We say that a collection of vectors is orthogonal if every pair of vectors from the collection is orthogonal. Often it is convenient to work with orthogonal unit vectors; we refer to these as orthonormal vectors. Here are the precise definitions (the "Kronecker delta" is defined by $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ if $i \neq j$).

Definition 3.4.1. Let H be an inner product space, and let I be an arbitrary index set.

(a) Two vectors $x, y \in H$ are *orthogonal*, denoted $x \perp y$, if $\langle x, y\rangle = 0$.

(b) A set of vectors $\{x_i\}_{i\in I}$ is *orthogonal* if $\langle x_i, x_j\rangle = 0$ whenever $i \neq j$.

(c) A set of vectors $\{x_i\}_{i\in I}$ is *orthonormal* if it is orthogonal and each vector x_i is a unit vector. Using the Kronecker delta notation, $\{x_i\}_{i\in I}$ is orthonormal if for all $i, j \in I$ we have

$$\langle x_i, x_j\rangle \;=\; \delta_{ij} \;=\; \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \qquad \diamondsuit$$

The zero vector may be an element of a set of orthogonal vectors. Any orthogonal set $\{x_i\}_{i\in I}$ of *nonzero* vectors can be rescaled to form an orthonormal set. That is, if $\{x_i\}_{i\in I}$ is orthogonal and $x_i \neq 0$ for every $i \in I$, then we can set

$$y_i \;=\; \frac{x_i}{\|x_i\|},$$

and $\{y_i\}_{i\in I}$ will be an orthonormal set.

Example 3.4.2. The sequence of standard basis vectors $\{\delta_n\}_{n\in\mathbb{N}}$ is an orthonormal sequence in ℓ^2. $\diamondsuit$

3.5 Orthogonal Complements

We define the orthogonal complement of a set A to be the largest set B that is "orthogonal to A" in the sense that every vector in B is orthogonal to every vector in A.

Definition 3.5.1 (Orthogonal Complement). Let A be a subset of an inner product space H. The *orthogonal complement* of A is

$$A^\perp \;=\; \bigl\{x \in H \,:\, \langle x, y\rangle = 0 \text{ for all } y \in A\bigr\}. \qquad \diamondsuit$$

Here are some properties of orthogonal complements (the proof is assigned as Problem 3.9.12).

Lemma 3.5.2. *If A is a subset of an inner product space H, then the following statements hold.*

(a) *$A^\perp$ is a closed subspace of H.*

(b) *$H^\perp = \{0\}$ and $\{0\}^\perp = H$.*

(c) *If $A \subseteq B$, then $B^\perp \subseteq A^\perp$.*

(d) *$A \subseteq (A^\perp)^\perp$.* $\diamondsuit$

Later we will prove that *if M is a closed subspace of a Hilbert space, then* $(M^\perp)^\perp = M$ (see Lemma 3.6.4).

3.6 Orthogonal Projections

The following theorem states that if S is a *closed and convex subset of a Hilbert space* H, then for each vector $x \in H$ there exists is a *unique* vector $y \in S$ that is closest to x.

Theorem 3.6.1 (Closest Point Theorem). *Let H be a Hilbert space, and let S be a nonempty closed, convex subset of H. Then for each vector $x \in H$ there exists a unique vector $y \in S$ that is closest to x. That is, there is a unique vector $y \in S$ that satisfies*

$$\|x - y\| \;=\; \operatorname{dist}(x, S) \;=\; \inf\bigl\{\|x - k\| : k \in S\bigr\}.$$

Proof. Set $d = \operatorname{dist}(x, S)$. Then, by the definition of an infimum, there exist vectors $y_n \in S$ such that

$$\lim_{n\to\infty} \|x - y_n\| \;=\; d.$$

Further, we have $\|x - y_n\| \ge d$ for every n. Therefore, if we fix $\varepsilon > 0$ then we can find an integer $N > 0$ such that

$$d^2 \;\le\; \|h - y_n\|^2 \;\le\; d^2 + \varepsilon^2, \qquad \text{for all } n \ge N.$$

Set

$$p \;=\; \frac{y_m + y_n}{2}.$$

Since p is the midpoint of the line segment joining y_m to y_n we have $p \in S$, and therefore

$$\|x - p\| \;\ge\; \operatorname{dist}(x, S) \;=\; d.$$

Using the Parallelogram Law, it follows that if $m, n \ge N$, then

$$\begin{aligned}
\|y_n - y_m\|^2 + 4d^2 &\le \|y_n - y_m\|^2 + 4\,\|x-p\|^2 \\
&= \|(x-y_n) - (x-y_m)\|^2 + \|(x-y_n) + (x-y_m)\|^2 \\
&= 2\left(\|x-y_n\|^2 + \|x-y_m\|^2\right) \\
&\le 4\,(d^2 + \varepsilon^2).
\end{aligned}$$

Rearranging, we see that $\|y_m - y_n\| \le 2\varepsilon$ for all $m, n \ge N$. Therefore $\{y_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in H. Since H is complete, this sequence must converge, say to y. Since S is closed and $y_n \in S$ for every n, the vector y must belong to S. Also, since $x - y_n \to x - y$, it follows from the continuity of the norm that

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

Hence y is a point in S that is closest to x.

It only remains to show that y is the unique point in S that is closest to x. If $z \in S$ is also a closest point, then $\|x-y\| = d = \|x-z\|$. Further, the midpoint $p = (y+z)/2$ belongs to S, so $\|x-p\| \ge d$. Applying the Parallelogram Law again, we see that

$$\begin{aligned}
4d^2 &= 2\left(\|x-y\|^2 + \|x-z\|^2\right) \\
&= \|(x-y) - (x-z)\|^2 + \|(x-y) + (x-z)\|^2 \\
&= \|y-z\|^2 + 4\,\|x-p\|^2 \\
&\ge \|y-z\|^2 + 4d^2.
\end{aligned}$$

Rearranging this yields $\|y-z\| \le 0$, which implies that $y = z$. □

In particular, every closed subspace M of H is nonempty, closed, and convex. For this setting we introduce a name for the point in M that is closest to a given vector x. We also use the same name to denote the function that maps x to the point in M that is closest to x.

Definition 3.6.2 (Orthogonal Projection). Let M be a closed subspace of a Hilbert space H.

(a) If $x \in H$, then the unique vector $p \in M$ that is closest to x is called the *orthogonal projection of x onto M*.

(b) The function $P\colon H \to H$ defined by $Px = p$, where p is the orthogonal projection of x onto M, is called the *orthogonal projection of H onto M*. ♢

Since the orthogonal projection p is the vector in M that is closest to x, we can think of p as being the best approximation to x by vectors from M. The difference vector $e = x - p$ is the error in this approximation. The following lemma states that the orthogonal projection of x is the unique vector $p \in H$ such that the error vector $e = x - p$ is orthogonal to M.

Lemma 3.6.3. *Let M be a closed subspace of a Hilbert space H. If x and p are vectors in H, then the following four statements are equivalent.*

(a) *p is the orthogonal projection of x onto M, i.e., p is the unique point in M that is closest to h.*

(b) *$p \in M$ and $x - p \perp M$.*

(c) *$x = p + e$ where $p \in M$ and $e \in M^{\perp}$.*

(d) *$e = x - p$ is the orthogonal projection of x onto $M^{\perp}$.*

Proof. We will prove one implication, and assign the task of proving the remaining (easier) implications as Problem 3.9.13. We assume that scalars in this problem are complex, but the proof remains valid if we assume that scalars are real.

(a) $\Rightarrow$ (b). Let p be the (unique) point in M closest to x, and let $e = p - x$. Choose any vector $y \in M$. We must show that $\langle y, e\rangle = 0$. Since M is a subspace, $p + \lambda y \in M$ for every scalar $\lambda \in \mathbb{C}$. Hence,

$$\begin{aligned} \|x-p\|^2 \ \le\ \|x-(p+\lambda y)\|^2 \ &=\ \|(x-p)-\lambda y\|^2 \\ &=\ \|x-p\|^2 \ -\ 2\,\mathrm{Re}\langle \lambda y, x-p\rangle \ +\ |\lambda|^2\,\|y\|^2 \\ &=\ \|x-p\|^2 \ -\ 2\,\mathrm{Re}(\lambda\langle y, e\rangle) \ +\ |\lambda|^2\,\|y\|^2. \end{aligned}$$

Therefore,

$$\forall\, \lambda \in \mathbb{C}, \quad 2\,\mathrm{Re}(\lambda\langle y,e\rangle) \ \le\ |\lambda|^2\,\|y\|^2.$$

If we consider $\lambda = t > 0$, then we can divide through by t to get

$$\forall\, t > 0, \quad 2\,\mathrm{Re}\langle y,e\rangle \ \le\ t\,\|y\|^2.$$

Letting $t \to 0^+$, we conclude that $\mathrm{Re}\langle y,e\rangle \le 0$. If we similarly take $\lambda = t < 0$ and let $t \to 0^-$, we obtain $\mathrm{Re}\langle y,e\rangle \ge 0$, so $\mathrm{Re}\langle y,e\rangle = 0$.

Finally, by taking $\lambda = it$ with $t > 0$ and then $\lambda = it$ with $t < 0$, it follows that $\mathrm{Im}\langle y,e\rangle = 0$ as well. □

We will use Lemma 3.6.3 to compute the orthogonal complement of the orthogonal complement of a set.

Lemma 3.6.4. *Let H be a Hilbert space.*

(a) *If M is a closed subspace of H, then $(M^{\perp})^{\perp} = M$.*

(b) *If A is any subset of H, then*

$$A^{\perp} \ =\ \mathrm{span}(A)^{\perp} \ =\ \overline{\mathrm{span}}(A)^{\perp} \qquad \textit{and} \qquad (A^{\perp})^{\perp} \ =\ \overline{\mathrm{span}}(A).$$

Proof. (a) We are given a closed subspace M in H. If $x \in M$ then $\langle x, y\rangle = 0$ for every $y \in M^{\perp}$, so $x \in (M^{\perp})^{\perp}$. Hence $M \subseteq (M^{\perp})^{\perp}$.

Conversely, suppose that $x \in (M^{\perp})^{\perp}$. Let p be the orthogonal projection of x onto M. Since M is a closed subspace, we have $x = p + e$ where $p \in M$

and $e \in M^\perp$. Since p belongs to M and we have seen that $M \subseteq (M^\perp)^\perp$, it follows that $e = x - p \in (M^\perp)^\perp$. However, we also know that $e \in M^\perp$, so e is orthogonal to itself and therefore is zero. Hence $x = p + 0 \in M$. This shows that $(M^\perp)^\perp \subseteq M$.

(b) Now we are given an arbitrary subset A of H. Let $M = \overline{\text{span}}(A)$. We must show that $A^\perp = M^\perp$. Since $A \subseteq M$, we have $M^\perp \subseteq A^\perp$.

Suppose that $x \in A^\perp$. Then $x \perp A$, i.e., x is orthogonal to every vector in A. By forming linear combinations, it follows that $x \perp \text{span}(A)$. By taking limits, it follows from this that $x \perp \overline{\text{span}}(A) = M$. Hence $x \in M^\perp$ and therefore $A^\perp \subseteq M^\perp$.

Thus, we have shown that $A^\perp = M^\perp$. Applying part (a), we conclude that $(A^\perp)^\perp = (M^\perp)^\perp = M$. $\square$

By taking A to be a sequence, we obtain the following corollary.

Corollary 3.6.5. *Given a sequence $\{x_n\}_{n\in\mathbb{N}}$ in a Hilbert space H, the following two statements are equivalent.*

(a) *$\{x_n\}_{n\in\mathbb{N}}$ is a complete sequence, i.e.,* $\text{span}\{x_n\}_{n\in\mathbb{N}}$ *is dense in H.*

(b) *The only vector in H that is orthogonal to every x_n is the zero vector, i.e.,*

$$x \in H \text{ and } \langle x, x_n \rangle = 0 \text{ for every } n \quad \Longrightarrow \quad x = 0.$$

Proof. By definition, $\{x_n\}_{n\in\mathbb{N}}$ is complete if and only if its finite linear span is dense in H. Lemma 3.6.4 tells us that $\overline{\text{span}}\{x_n\}_{n\in\mathbb{N}} = \{x_n\}_{n\in\mathbb{N}}^\perp$, so we conclude that $\{x_n\}_{n\in\mathbb{N}}$ is complete if only if $\{x_n\}_{n\in\mathbb{N}}^\perp = \{0\}$. $\square$

3.7 Orthonormal Sequences

If $\{e_n\}_{n\in\mathbb{N}}$ is any sequence of orthonormal vectors in a Hilbert space, then its closed span $M = \overline{\text{span}}\big(\{e_n\}_{n\in\mathbb{N}}\big)$ is a closed subspace of H. The following lemma gives an explicit formula for the orthogonal projection of a vector onto M, along with other useful properties of $\{e_n\}_{n\in\mathbb{N}}$. Entirely similar results hold for a finite orthonormal sequence $\{e_1, \dots, e_N\}$ by replacing $n = 1, 2, \dots$ by $n = 1, \dots, N$ in Theorem 3.7.1. In fact, the proof is easier for finite sequences, since in that case there are no issues about convergence of infinite series.

Theorem 3.7.1. *Let H be a Hilbert space, let $\{e_n\}_{n\in\mathbb{N}}$ be an orthonormal sequence in H, and set*

$$M = \overline{\text{span}}\big(\{e_n\}_{n\in\mathbb{N}}\big). \tag{3.7}$$

Then the following statements hold.

(a) Bessel's Inequality:

$$\sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2 \;\le\; \|x\|^2, \qquad \textit{for every } x \in H.$$

(b) *If* c_n *is a scalar for* $n \in \mathbb{N}$ *and the series* $x = \sum_{n=1}^{\infty} c_n e_n$ *converges, then*

$$c_n \;=\; \langle x, e_n\rangle, \qquad \textit{for every } n \in \mathbb{N}.$$

(c) *If* c_n *is a scalar for* $n \in \mathbb{N}$, *then*

$$\sum_{n=1}^{\infty} c_n e_n \;\textit{converges} \qquad \Longleftrightarrow \qquad \sum_{n=1}^{\infty} |c_n|^2 < \infty.$$

Further, in this case the series $\sum_{n=1}^{\infty} c_n e_n$ *converges* unconditionally, *i.e., it converges regardless of the ordering of the index set.*

(d) *If* $x \in H$, *then*

$$p \;=\; \sum_{n=1}^{\infty} \langle x, e_n\rangle\, e_n$$

is the orthogonal projection of x *onto* M, *and*

$$\|p\|^2 \;=\; \sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2.$$

(e) *If* $x \in H$, *then*

$$x \in M \iff x = \sum_{n=1}^{\infty} \langle x, e_n\rangle\, e_n \iff \|x\|^2 = \sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2. \tag{3.8}$$

Proof. (a) Choose $x \in H$. For each $N \in \mathbb{N}$ define

$$y_N \;=\; x \;-\; \sum_{n=1}^{N} \langle x, x_n\rangle\, x_n.$$

If $1 \le m \le N$, then

$$\langle y_N, x_m\rangle \;=\; \langle x, x_m\rangle \;-\; \sum_{n=1}^{N} \langle x, x_n\rangle\, \langle x_n, x_m\rangle \;=\; \langle x, x_m\rangle - \langle x, x_m\rangle \;=\; 0.$$

Thus $\{x_1, \dots, x_N, y_N\}$ is an orthogonal sequence in H (though not necessarily orthonormal, since we do not know the norm of y_N). Therefore, by the Pythagorean Theorem (see Problem 3.9.7),

$$\begin{aligned}
\|x\|^2 &= \Big\| y_N + \sum_{n=1}^{N} \langle x, x_n \rangle \, x_n \Big\|^2 \\
&= \|y_N\|^2 + \sum_{n=1}^{N} \| \langle x, x_n \rangle \, x_n \|^2 \\
&= \|y_N\|^2 + \sum_{n=1}^{N} |\langle x, x_n \rangle|^2 \\
&\geq \sum_{n=1}^{N} |\langle x, x_n \rangle|^2.
\end{aligned}$$

Letting $N \to \infty$, we obtain Bessel's Inequality.

(b) If $x = \sum c_n x_n$ converges, then for each fixed m we have by the continuity of the inner product that

$$\langle x, x_m \rangle = \sum_{n=1}^{\infty} c_n \, \langle x_n, x_m \rangle = \sum_{n=1}^{\infty} c_n \delta_{mn} = c_m.$$

(c) If $\sum c_n x_n$ converges, then part (b) implies that $c_n = \langle x, x_n \rangle$, and hence $\sum |c_n|^2 < \infty$ by Bessel's Inequality.

For the converse direction, suppose that $\sum |c_n|^2 < \infty$. Set

$$s_N = \sum_{n=1}^{N} c_n x_n \qquad \text{and} \qquad t_N = \sum_{n=1}^{N} |c_n|^2.$$

We know that $\{t_N\}_{N \in \mathbb{N}}$ is a convergent (hence Cauchy) sequence of scalars, and we must show that $\{s_N\}_{N \in \mathbb{N}}$ is a convergent sequence of vectors. For $N > M$ we have that

$$\begin{aligned}
\|s_N - s_M\|^2 &= \Big\| \sum_{n=M+1}^{N} c_n x_n \Big\|^2 \\
&= \sum_{n=M+1}^{N} \|c_n x_n\|^2 \\
&= \sum_{n=M+1}^{N} |c_n|^2 \\
&= |t_N - t_M|.
\end{aligned}$$

Since $\{t_N\}_{N \in \mathbb{N}}$ is Cauchy, we conclude that $\{s_N\}_{N \in \mathbb{N}}$ is a Cauchy sequence in H and hence converges.

(d) By Bessel's Inequality and part (c), we know that the series

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

converges. Our task is to show that this vector p is the orthogonal projection of x onto M.

Given any fixed integer $k \in \mathbb{N}$, we compute that

$$\begin{aligned}
\langle x-p, x_k\rangle \;&=\; \langle x, x_k\rangle \;-\; \Bigl\langle \sum_{n=1}^{\infty} \langle x, x_n\rangle\, x_n,\; x_k \Bigr\rangle \\
&=\; \langle x, x_k\rangle \;-\; \sum_{n=1}^{\infty} \langle x, x_n\rangle\, \langle x_n, x_k\rangle \\
&=\; \langle x, x_k\rangle \;-\; \langle x, x_k\rangle \;=\; 0,
\end{aligned}$$

where we have used Problem 3.9.6 to move the infinite series to the outside of the inner product. Thus $x-p$ is orthogonal to each vector x_k. Since the inner product is antilinear in the second variable, this implies that $x-p$ is orthogonal to every finite linear combination of the x_k. Taking limits and applying the continuity of the inner product again, it follows that $x-p$ is orthogonal to every vector in M, i.e., $x-p \perp M$. Lemma 3.6.3 therefore implies that that p is the orthogonal projection of x onto M.

(e) Statement (d) tells us that $p = \sum \langle x, e_n\rangle\, e_n$ is the orthogonal projection of x onto M, and that

$$\|p\|^2 \;=\; \langle p, p\rangle \;=\; \sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2.$$

Let i, ii, iii denote the three statements that appear in equation (3.8). We must prove that these statements i, ii, and iii are equivalent.

i $\Rightarrow$ ii. If $x \in M$, then the orthogonal projection of x onto M is x itself. Thus $x = p = \sum \langle x, x_n\rangle\, x_n$.

ii $\Rightarrow$ iii. If $x = p$ then $\|x\|^2 = \|p\|^2 = \sum |\langle x, e_n\rangle|^2$.

iii $\Rightarrow$ i. Suppose $\|x\|^2 = \sum |\langle x, e_n\rangle|^2$. Then since $x-p \perp p$, we can use the Pythagorean Theorem to compute that

$$\begin{aligned}
\|x\|^2 \;=\; \|(x-p)+p\|^2 \;&=\; \|x-p\|^2 \;+\; \|p\|^2 \\
&=\; \|x-p\|^2 \;+\; \sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2 \\
&=\; \|x-p\|^2 \;+\; \|x\|^2.
\end{aligned}$$

Hence $\|x-p\| = 0$, so $x = p \in M$. $\square$

3.8 Orthonormal Bases

If the closed span of an orthonormal sequence $\{e_n\}_{n\in\mathbb{N}}$ is $M = H$, then we say that the sequence $\{e_n\}_{n\in\mathbb{N}}$ is *complete*, *total*, or *fundamental* (see Definition 2.7.1). Part (e) of Theorem 3.7.1 implies that if $\{e_n\}_{n\in\mathbb{N}}$ is both orthonormal and complete, then *every* vector $x \in H$ can be written as $x = \sum \langle x, e_n\rangle\, e_n$. The following theorem states that this property characterizes completeness (assuming that our sequence $\{e_n\}_{n\in\mathbb{N}}$ is orthonormal), and gives several other characterizations of complete orthonormal sequences. A completely analogous theorem holds for finite orthonormal sequence $\{e_1, \dots, e_d\}$ (see Problem 3.9.15).

Theorem 3.8.1. *If H is a Hilbert space and $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal sequence in H, then the following five statements are equivalent.*

(a) $\{e_n\}_{n\in\mathbb{N}}$ *is* complete, *i.e.,* $\overline{\mathrm{span}}\{e_n\}_{n\in\mathbb{N}} = H$.

(b) $\{e_n\}_{n\in\mathbb{N}}$ *is a* Schauder basis *for H, i.e., for each $x \in H$ there exists a* unique *sequence of scalars $(c_n)_{n\in\mathbb{N}}$ such that $x = \sum c_n e_n$.*

(c) *For each $x \in H$ we have*

$$x \;=\; \sum_{n=1}^{\infty} \langle x, e_n\rangle\, e_n. \tag{3.9}$$

(d) Plancherel's Equality:

$$\|x\|^2 \;=\; \sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2, \qquad \textit{for all } x \in H.$$

(e) Parseval's Equality:

$$\langle x, y\rangle \;=\; \sum_{n=1}^{\infty} \langle x, e_n\rangle\, \langle e_n, y\rangle, \qquad \textit{for all } x, y \in H.$$

Proof. (a) $\Rightarrow$ (b). If $\{e_n\}$ is complete, then its closed span is all of H, so Theorem 3.7.1(c) implies that $x = \sum \langle x, e_n\rangle\, e_n$ for every $x \in H$.

(b) $\Rightarrow$ (c). If statement (b) holds, then we must have $c_n = \langle x, x_n\rangle$ by Theorem 3.7.1(b).

(c) $\Rightarrow$ (e). Suppose that statement (c) holds, and fix any vectors $x, y \in H$. Then

$$\langle x, y\rangle \;=\; \Big\langle \sum_{n=1}^{\infty} \langle x, e_n\rangle\, e_n,\, y \Big\rangle \;=\; \sum_{n=1}^{\infty} \big\langle \langle x, e_n\rangle\, e_n,\, y \big\rangle \;=\; \sum_{n=1}^{\infty} \langle x, e_n\rangle\, \langle e_n, y\rangle,$$

where we have used Problem 3.9.6 to move the infinite series outside of the inner product.

(e) $\Rightarrow$ (d). This follows by taking $x = y$.

(d) $\Rightarrow$ (c). Suppose that $\|x\|^2 = \sum_{n=1}^{\infty} |\langle x, e_n \rangle|^2$ for every $x \in H$. Fix x, and define $s_N = \sum_{n=1}^{N} \langle x, e_n \rangle \, e_n$. Then, by direct calculation,

$$\begin{aligned}
\|x - s_N\|^2 &= \|x\|^2 - \langle x, s_N \rangle - \langle s_N, x \rangle + \|s_N\|^2 \\
&= \|x\|^2 - \sum_{n=1}^{N} |\langle x, e_n \rangle|^2 - \sum_{n=1}^{N} |\langle x, e_n \rangle|^2 + \sum_{n=1}^{N} |\langle x, e_n \rangle|^2 \\
&= \|x\|^2 - \sum_{n=1}^{N} |\langle x, e_n \rangle|^2 \\
&\to 0 \quad \text{as } N \to \infty.
\end{aligned}$$

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

(d) $\Rightarrow$ (a). Suppose that Plancherel's Equality holds, and $\langle x, e_n \rangle = 0$ for every $n \in \mathbb{N}$. Then

$$\|x\|^2 = \sum_{n=1}^{\infty} |\langle x, e_n \rangle|^2 = 0,$$

so $x = 0$. Hence $\{e_n\}$ is complete. $\square$

We will refer to a sequence that satisfies the equivalent conditions in Theorem 3.8.1 as an *orthonormal basis*.

Definition 3.8.2 (Orthonormal Basis). Let H be a Hilbert space. A countably infinite orthonormal sequence $\{e_n\}_{n \in \mathbb{N}}$ that is complete in H is called an *orthonormal basis* for H. $\diamondsuit$

We use similar terminology for finite-dimensional Hilbert spaces. Specifically, if $\{e_1, \dots, e_d\}$ is a complete orthonormal sequence in a Hilbert space H, then $\{e_1, \dots, e_d\}$ is a basis for H in the usual vector space sense (i.e., it is a Hamel basis), and we call it an *orthonormal basis* for H.

Example 3.8.3. The sequence of standard basis vectors $\{\delta_n\}_{n \in \mathbb{N}}$ is both complete and orthonormal in ℓ^2, so it is an orthonormal basis for ℓ^2. $\diamondsuit$

By Theorem 3.8.1, a countable sequence that is both *complete and orthonormal* is a Schauder basis for H. However, we emphasize that *a complete sequence that is not orthonormal need not be a Schauder basis*, even if it is finitely linearly independent. An example in the Hilbert space ℓ^2 is given in Problem 3.9.21. A similar phenomenon in the Banach space $C[a, b]$ was discussed earlier in Example 2.9.5, where it was shown that the sequence of monomials $\mathcal{M} = \{1, x, x^2, \dots\}$ is complete and finitely linearly independent in $C[a, b]$ but is not a Schauder basis for $C[a, b]$.

3.9 Existence of an Orthonormal Basis

A Hilbert space is *separable* if it contains a countable dense subset. All finite-dimensional Hilbert spaces are separable, and Problem 3.9.20 shows that the space ℓ^2 is separable. However, not every Hilbert space is separable; an example is given in Problem 3.9.22.

We will show that every separable Hilbert space contains an orthonormal basis. We begin with finite-dimensional spaces, where we can use the same *Gram–Schmidt* procedure that is employed to construct orthonormal sequences in $\mathbb{R}^d$ or $\mathbb{C}^d$.

Theorem 3.9.1. *If H is a finite-dimensional Hilbert space and d is the vector space dimension of H, then H contains an orthonormal basis of the form $\{e_1, \dots, e_d\}$.*

Proof. Since H is a d-dimensional vector space, it has a Hamel basis $\mathcal{B}$ that consists of d vectors, say $\mathcal{B} = \{x_1, \dots, x_d\}$.

Set $y_1 = x_1$, and note that $y_1 \neq 0$ since $x_1, \dots, x_d$ are linearly independent and therefore nonzero. Define

$$M_1 \;=\; \operatorname{span}\{x_1\} \;=\; \operatorname{span}\{y_1\}.$$

Note that M_1 is closed since it is finite-dimensional.

If $d = 1$ then we stop here. Otherwise M_1 is a proper subspace of H, and $x_2 \notin M_1$ (because $\{x_1, \dots, x_d\}$ is linearly independent). Let p_2 be the orthogonal projection of x_2 onto M_1. Then the vector $y_2 = x_2 - p_2$ is orthogonal to x_1, and $y_2 \neq 0$ since $x_2 \notin M_1$. Therefore we can define

$$M_2 \;=\; \operatorname{span}\{x_1, x_2\} \;=\; \operatorname{span}\{y_1, y_2\},$$

where the second equality follows from the fact that y_1, y_2 are linear combinations of x_1, x_2, and vice versa.

If $d = 2$ then we stop here. Otherwise we continue in the same manner. At stage k we have constructed orthogonal vectors $y_1, \dots, y_k$ such that

$$M_k \;=\; \operatorname{span}\{x_1, \dots, x_k\} \;=\; \operatorname{span}\{y_1, \dots, y_k\}.$$

The subspace M_k has dimension k. Consequently, when we reach $k = d$ we will have $M_d = H$, and therefore

$$H \;=\; M_d \;=\; \operatorname{span}\{x_1, \dots, x_d\} \;=\; \operatorname{span}\{y_1, \dots, y_d\}.$$

The vectors $y_1, \dots, y_d$ are orthogonal and nonzero, so by setting

$$e_k \;=\; \frac{y_k}{\|y_k\|}, \qquad k = 1, \dots, d,$$

we obtain an orthonormal basis $\{e_1, \dots, e_d\}$ for H. $\square$

Next we consider infinite-dimensional, but still separable, Hilbert spaces.

Theorem 3.9.2. *If H is a infinite-dimensional, separable Hilbert space, then H contains an orthonormal basis of the form $\{e_n\}_{n\in\mathbb{N}}$.*

Proof. Since H is separable, it contains a countable dense subset. This subset must be infinite, so let us say that it is $\mathcal{Z} = \{z_n\}_{n\in\mathbb{N}}$. However, $\mathcal{Z}$ need not be linearly independent, so we will extract a linearly independent subsequence as follows.

Let k_1 be the first index such that $z_{k_1} \neq 0$, and set $x_1 = z_{k_1}$. Then let k_2 be the first index larger than k_1 such that $z_{k_2} \notin \operatorname{span}\{x_1\}$, and set $x_2 = z_{k_2}$. Then let k_3 be the first index larger than k_2 such that $z_{k_3} \notin \operatorname{span}\{x_1, x_2\}$, and so forth. In this way we obtain vectors $x_1, x_2, \dots$ such that $x_1 \neq 0$ and for each $n > 1$ we have

$$x_n \notin \operatorname{span}\{x_1, \dots, x_{n-1}\} \quad \text{and} \quad \operatorname{span}\{x_1, \dots, x_n\} \;=\; \operatorname{span}\{z_1, \dots, z_{k_n}\}.$$

Therefore $\{x_n\}_{n\in\mathbb{N}}$ is linearly independent, and furthermore $\operatorname{span}\bigl(\{x_n\}_{n\in\mathbb{N}}\bigr)$ is dense in H.

Now we apply the Gram–Schmidt procedure utilized in the proof of Theorem 3.9.1, but without stopping. This gives us orthonormal vectors $e_1, e_2, \dots$ such that for every n we have

$$\operatorname{span}\{e_1, \dots, e_n\} \;=\; \operatorname{span}\{x_1, \dots, x_n\}.$$

Consequently $\operatorname{span}\bigl(\{e_n\}_{n\in\mathbb{N}}\bigr)$ is dense in H. Therefore $\{e_n\}_{n\in\mathbb{N}}$ is a complete orthonormal sequence in H, and hence it is an orthonormal basis for H. $\square$

The following result gives a converse to Theorems 3.9.1 and 3.9.2: *only* a separable Hilbert space can contain an orthonormal basis. This is a consequence of the fact that we declared in Definition 3.8.2 that an orthonormal basis must be a *countable sequence.*

Theorem 3.9.3. *If a Hilbert space H is not separable, then H does not contain an orthonormal basis.*

Proof. We will prove the contrapositive statement. Suppose that H contains an orthonormal basis. Such a basis is either finite or countably infinite; since both cases are similar let us assume that $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal basis for H.

Say that a complex number is *rational* if both its real and imaginary parts are rational, and let

$$S \;=\; \left\{\sum_{n=1}^{N} r_n e_n : N > 0,\ r_1, \dots, r_N \text{ rational}\right\}.$$

This is a countable subset of H, and we will show that it is dense.

Choose any $x \in H$ and fix $\varepsilon > 0$. Since $\|x\|^2 = \sum |\langle x, e_n\rangle|^2$, we can choose N large enough that

$$\sum_{n=N+1}^{\infty} |\langle x, e_n\rangle|^2 \;<\; \frac{\varepsilon^2}{2}.$$

For each $n = 1, \dots, N$, choose a scalar r_n that has real and imaginary parts and satisfies

$$|\langle x, e_n\rangle - r_n|^2 \;<\; \frac{\varepsilon^2}{2N}.$$

Then the vector

$$z \;=\; \sum_{n=1}^{N} r_n e_n$$

belongs to S, and by the Plancherel Equality we have

$$\|x - z\|^2 \;=\; \sum_{n=1}^{N} |\langle x, e_n\rangle - r_n|^2 + \sum_{n=N+1}^{\infty} |\langle x, e_n\rangle|^2 \;<\; N\frac{\varepsilon^2}{2N} + \frac{\varepsilon^2}{2} \;=\; \varepsilon^2.$$

Thus $\|x - z\| < \varepsilon$, so S is dense in H. Since S is also countable, it follows that H is separable. $\square$

Nonseparable Hilbert spaces do exist (see Problem 3.9.22 for one example). An argument based on the Axiom of Choice in the form of Zorn's Lemma shows that every Hilbert space, including nonseparable Hilbert spaces, contains a complete orthonormal set (for one proof, see [Heil11, Thm. 1.56]). However, if H is nonseparable, then such a complete orthonormal set must be uncountable. An uncountable complete orthonormal set does have certain basis-like properties (e.g., see [Heil11, Exer. 3.6]), and for this reason some authors refer to a complete orthonormal set of any cardinality as an orthonormal basis. In keeping with the majority of the Banach space literature, we prefer to reserve the word "basis" for use in conjunction with countable sequences only.

Problems

3.9.4. Prove Lemma 3.2.1.

3.9.5. Let H be an inner product space. Prove the *continuity of the inner product*: If $x_n \to x$ and $y_n \to y$ in H, then $\langle x_n, y_n\rangle \to \langle x, y\rangle$.

3.9.6. Prove that if a series $\sum_{n=1}^{\infty} x_n$ converges in an inner product space H, then

$$\Big\langle \sum_{n=1}^{\infty} x_n,\, y \Big\rangle \;=\; \sum_{n=1}^{\infty} \langle x_n, y\rangle, \qquad \text{for every } y \in H.$$

Note that this is not merely a consequence of the linearity of the inner product in the first variable; the continuity of the inner product is also needed.

3.9.7. Extend the Pythagorean Theorem to finite orthogonal sets of vectors. That is, prove that if $x_1, \dots, x_N \in H$ are orthogonal vectors in an inner product space H, then

$$\left\| \sum_{n=1}^{N} x_n \right\|^2 \;=\; \sum_{n=1}^{N} \|x_n\|^2.$$

Does the result still hold if we only assume that $\langle \cdot, \cdot \rangle$ is a semi-inner product?

3.9.8. Prove that the function $\langle \cdot, \cdot \rangle$ defined in equation (3.6) is an inner product on $C[a,b]$.

3.9.9. Prove that ℓ^2 is complete with respect to the norm $\|\cdot\|_2$ defined in equation (3.2).

3.9.10. For each $n \in \mathbb{N}$, let

$$y_n \;=\; \bigl(1, \tfrac{1}{2}, \dots, \tfrac{1}{n}, 0, 0, \dots\bigr).$$

Note that $y_n \in \ell^1$ for every n.

(a) Assume that the norm on ℓ^1 is its usual norm $\|\cdot\|_1$. Prove that $\{y_n\}_{n\in\mathbb{N}}$ is not a Cauchy sequence in ℓ^1 with respect to this norm.

(b) Now assume that the norm on ℓ^1 is $\|\cdot\|_2$. Prove that $\{y_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in ℓ^1 with respect to $\|\cdot\|_2$. Even so, prove that there is no vector $y \in \ell^1$ such that $\|y - y_n\|_2 \to 0$. Conclude that ℓ^1 is not complete with respect to the norm $\|\cdot\|_2$.

3.9.11. Suppose that $\{x_n\}_{n\in\mathbb{N}}$ is a sequence in a Hilbert space H, and $y \in H$ is orthogonal to x_n for every n. Prove the following statements.

(a) y is orthogonal to every vector in $\mathrm{span}\{x_n\}_{n\in\mathbb{N}}$, and therefore $y \in \mathrm{span}\{x_n\}_{n\in\mathbb{N}}^{\perp}$.

(b) y is orthogonal to every vector in $\overline{\mathrm{span}}\{x_n\}_{n\in\mathbb{N}}$, and therefore $y \in \overline{\mathrm{span}}\{x_n\}_{n\in\mathbb{N}}^{\perp}$.

3.9.12. Prove Lemma 3.5.2.

3.9.13. Prove the remaining implications in Lemma 3.6.3.

3.9.14. Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in a Hilbert space H. Prove that the following two statements are equivalent.

(a) For each $m \in \mathbb{N}$ we have $x_m \notin \overline{\mathrm{span}}\bigl(\{x_n\}_{n\neq m}\bigr)$ (such a sequence is said to be *minimal*).

(b) There exists a sequence $\{y_n\}_{n\in\mathbb{N}}$ in H such that $\langle x_m, y_n\rangle = \delta_{mn}$ for all $m, n \in \mathbb{N}$ (we say that sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{y_n\}_{n\in\mathbb{N}}$ satisfying this condition are *biorthogonal*).

Show further that in case statements (a) and (b) hold, the sequence $\{y_n\}_{n\in\mathbb{N}}$ is unique if and only if $\{x_n\}_{n\in\mathbb{N}}$ is complete.

3.9.15. Formulate and prove analogues of Theorems 3.7.1 and 3.8.1 for finite orthonormal sequences.

3.9.16. A $d \times d$ matrix A with scalar entries is said to be *positive definite* if $Ax \cdot x > 0$ for all nonzero vectors $x \in \mathbb{C}^d$, where $x \cdot y$ denotes the usual dot product of vectors in $\mathbb{C}^d$.

(a) Let S be an invertible $d \times d$ matrix, and suppose that Λ is a diagonal matrix whose diagonal entries are all positive. Prove that $A = S\Lambda S^{\mathrm{H}}$ is a positive definite matrix, where $S^{\mathrm{H}} = \overline{S^{\mathrm{T}}}$ is the complex conjugate of the transpose of S (usually referred to as the *Hermitian* of S).

Remark: In fact, it can be shown that every positive definite matrix has this form, but that is not needed for this problem.

(b) Show that if A is a positive definite $d \times d$ matrix, then

$$\langle x, y\rangle_A \;=\; Ax \cdot y, \qquad \text{for } x, y \in \mathbb{C}^d,$$

defines an inner product on $\mathbb{C}^d$.

(c) Show that if $\langle\cdot,\cdot\rangle$ is an inner product on $\mathbb{C}^d$, then there exists some positive definite $d \times d$ matrix A such that $\langle\cdot,\cdot\rangle = \langle\cdot,\cdot\rangle_A$.

3.9.17. Let $\langle\cdot,\cdot\rangle$ be a semi-inner product on a vector space H. Show that equality holds in the Cauchy–Bunyakovski–Schwarz Inequality if and only if there exist scalars α, β, not both zero, such that $\|\alpha x + \beta y\| = 0$. In particular, if $\langle\cdot,\cdot\rangle$ is an inner product, then either $x = cy$ or $y = cx$ where c is a scalar.

3.9.18. Let M be a closed subspace of a Hilbert space H, and let P be the orthogonal projection of H onto M. Show that $I - P$ is the orthogonal projection of H onto $M^\perp$.

3.9.19. Assume $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal basis for a Hilbert space H.

(a) Suppose that vectors $y_n \in H$ satisfy $\sum \|e_n - y_n\|^2 < 1$. Prove that $\{y_n\}_{n\in\mathbb{N}}$ is a complete sequence in H.

(b) Show that part (a) can fail if we only have $\sum \|x_n - y_n\|^2 = 1$.

3.9.20. Prove that the set S defined in Problem 2.9.13 is a countable, dense subset of ℓ^2. Conclude that ℓ^2 is separable.

3.9.21. We say that a sequence $\{x_n\}_{n\in\mathbb{N}}$ in a Banach space X is ω-*dependent* if there exist scalars c_n, not all zero, such that $\sum_{n=1}^{\infty} c_n x_n = 0$, where the series converges in the norm of X. A sequence is ω-*independent* if it is not ω-dependent.

(a) Prove that every Schauder basis is both complete and ω-independent.

(b) Let $\alpha, \beta \in \mathbb{C}$ be fixed nonzero scalars such that $|\alpha| > |\beta|$. Let $\{\delta_n\}_{n\in\mathbb{N}}$ be the sequence of standard basis vectors, and define

$$x_0 = \delta_1 \qquad \text{and} \qquad x_n = \alpha\delta_n + \beta\delta_{n+1}, \; n \in \mathbb{N}.$$

Prove that the sequence $\{x_n\}_{n\geq 0}$ is complete and finitely linearly independent in ℓ^2, but it is not ω-independent and therefore is not a Schauder basis for ℓ^2.

3.9.22. Let I be an uncountable index set I, and let $\ell^2(I)$ consist of all sequences $x = (x_i)_{i\in I}$ *with at most countably many nonzero components* such that

$$\|x\|_2^2 = \sum_{i\in I} |x_i|^2 < \infty.$$

(a) Prove that $\ell^2(I)$ is a Hilbert space with respect to the inner product

$$\langle x, y\rangle = \sum_{i\in I} x_i \, \overline{y_k}.$$

(b) For each $i \in I$, define $\delta_i = (\delta_{ij})_{j\in I}$, where δ_{ij} is the Kronecker delta. Show that $\{\delta_i\}_{i\in I}$ is a complete orthonormal sequence in $\ell^2(I)$.

(c) Prove that $\ell^2(I)$ is not separable.

Index of Symbols

Sets

Symbol	Description
$\varnothing$	Empty set
$B_r(x)$	Open ball of radius r centered at x
$\mathbb{C}$	Complex plane
$\mathbb{N}$	Natural numbers, $\{1, 2, 3, \dots\}$
$\mathcal{P}$	Set of all polynomials
$\mathbb{Q}$	Rational numbers
$\mathbb{R}$	Real line
$\mathbb{Z}$	Integers, $\{\dots, -1, 0, 1, \dots\}$

Sets

Symbol	Description
$A^{\mathrm{C}} = X \backslash A$	Complement of a set $A \subseteq X$
A°	Interior of a set A
$\overline{A}$	Closure of a set A
∂A	Boundary of a set A
$A \times B$	Cartesian product of A and B
$\mathrm{diam}(A)$	Diameter of a set
$\mathrm{dist}(x, E)$	Distance from a point to a set
$\inf(S)$	Infimum of a set of real numbers S
$\mathrm{span}(A)$	Finite linear span of a set A
$\overline{\mathrm{span}}(A)$	Closure of the finite linear span of A
$\sup(S)$	Supremum of a set of real numbers S

Sequences

Symbol	Description
$\{x_n\}_{n\in\mathbb{N}}$	A sequence of points $x_1, x_2, \dots$
$(x_k)_{k\in\mathbb{N}}$	A sequence of scalars $x_1, x_2, \dots$
δ_n	nth standard basis vector

Functions

Symbol	Description
$f\colon A \to B$	A function from A to B
$f(A)$	Direct image of A under f
$f^{-1}(B)$	Inverse image of B under f

Vector Spaces

Symbol	Description
c_{00}	Set of all "finite sequences"
$C_b(X)$	Set of bounded continuous functions on X
$C[a,b]$	Set of continuous functions on $[a,b]$
ℓ^1	Set of all absolutely summable sequences
ℓ^2	Set of all square-summable sequences

Metrics, Norms, and Inner Products

Symbol	Description
$A^\perp$	Orthogonal complement of a set A
$\mathrm{d}(\cdot,\cdot)$	Generic metric
$\langle\cdot,\cdot\rangle$	Generic inner product
$\|\cdot\|$	Generic norm
$\|f\|_{\mathrm{u}}$	Uniform norm of a function f
$x \perp y$	Orthogonal vectors
$\|x\|_1$	ℓ^1-norm of a sequence x
$\|x\|_2$	ℓ^2-norm of a sequence x

References

[Ben97] J. J. Benedetto, *Harmonic Analysis and Applications*, CRC Press, Boca Raton, FL, 1997.

[Con90] J. B. Conway, *A Course in Functional Analysis*, Second Edition, Springer-Verlag, New York, 1990.

[DM72] H. Dym and H. P. McKean, *Fourier Series and Integrals*, Academic Press, New York–London, 1972.

[Enf73] P. Enflo, A counterexample to the approximation problem in Banach spaces, *Acta Math.*, **130** (1973), pp. 309–317.

[Fol99] G. B. Folland, *Real Analysis*, Second Edition, Wiley, New York, 1999.

[Heil11] C. Heil, *A Basis Theory Primer*, Expanded Edition, Birkhäuser, Boston, 2011.

[Heil18] C. Heil, *Metrics, Norms, Inner Products, and Operator Theory*, Birkhäuser/ Springer, Cham, 2018.

[Heil19] C. Heil, *Introduction to Real Analysis*, Springer, Cham 2019.

[Kat04] Y. Katznelson, *An Introduction to Harmonic Analysis*, Third Edition, Cambridge University Press, Cambridge, UK, 2004.

[Kre78] E. Kreyszig, *Introductory Functional Analysis with Applications*, Wiley, New York, 1978.

[Mun75] J. R. Munkres, *Topology: A First Course*, Prentice-Hall, Englewood Cliffs, NJ, 1975.

[Rud76] W. Rudin, *Principles of Mathematical Analysis*, Third Edition. McGraw-Hill, New York, 1976.

[ST76] I. M. Singer and J. A. Thorpe, *Lecture Notes on Elementary Topology and Geometry*, Springer-Verlag, New York-Heidelberg, 1976 (reprint of the 1967 edition).

[SS05] E. M. Stein and R. Shakarchi, *Real Analysis*, Princeton University Press, Princeton, NJ, 2005.

[WZ77] R. L. Wheeden and A. Zygmund, *Measure and Integral*, Marcel Dekker, New York-Basel, 1977.

Index