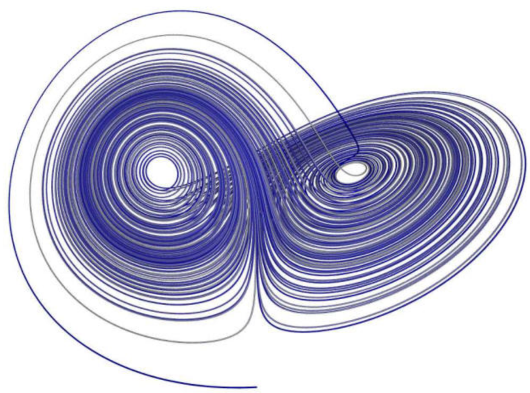
Real Anaylsis

MAT 3120 / MAT 3520

Fall 2016



Alistair Savage

Department of Mathematics and Statistics

University of Ottawa

Contents

Preface

These are notes for the course *Real Analysis* (MAT 3120, or its French version MAT 3520) at the University of Ottawa. Students are expected to have a knowledge of mutlivariable calculus and elementary real analysis, including the concepts of supremum, infimum, the topology of $\mathbb{R}^n$ (including compactness), and uniform continuity. They are also expected to have familiarity with sequences and series of functions, and the notion of uniform convergence.

In *Real Analysis*, we explore analysis over the real numbers in a more general setting. We begin with the concept of a metric space, where we discuss the ideas of convergence, sequences, completeness, sequential compactness, and isometry. We then move to the more general setting of a topological space, where we explore the notions of closed sets, separability, compactness, continuity, path-connectedness, and completions. In the second half of the course, we encounter the concept of a normed vector spaces, something which has both a vector space structure and a topological one. We finish the course with a brief discussion of inner product spaces and Hilbert spaces.

Acknowledgement: Portions of these notes (especially Chapter 4) are based on lecture notes by Vladimir Pestov. Other parts follow the text [Coh03].

Alistair Savage Ottawa, 2016.

Course website: `http://alistairsavage.ca/mat3520/`

Chapter 1

Metric spaces

In this chapter, we encounter the concept of a metric space. Intuitively speaking, this is a set with a good notion of distance. We will see that many of the concepts familiar from your study of analysis in $\mathbb{R}^n$ can be generalized to the setting of metric spaces.

Throughout these notes, we let $\mathbb{N} = \{0, 1, 2, \dots\}$ denote the set of nonnegative integers, and let $\mathbb{N}_+ = \{1, 2, 3, \dots, \}$ denote the set of positive integers.

1.1 Definitions and examples

In this section, we define metric spaces and discuss a number of examples. We begin with some motivation. You have seen in previous courses various notions of *convergence*. For example, you have seen:

- Convergence of sequences of real numbers.
- Convergence of sequences of functions (e.g. Taylor series):

  $$\lim_{n\to\infty} \sum_{k=0}^{n} \frac{x^k}{k!} = e^x;$$

  so the sequence of functions

  $$1,\ 1+x,\ 1+x+\frac{x^2}{2!}, \dots$$

  converges to the function e^x.

Although these two settings seem different (real numbers versus functions), the underlying idea is the same. Namely, elements in the sequence get "closer together". One of the main goals in *Real Analysis* will be to explore an abstract setting in which we can discuss convergence. The two examples above should be special cases of our more general treatment. What we need is some precise notion of *distance*. This is the motivation behind the definition of a metric space.

Definition 1.1.1 (Metric space)**.** A *metric space* is a pair (X, d) where X is a nonempty set and d is a mapping $d\colon X \times X \to \mathbb{R}_{\geq 0} = \{r \in \mathbb{R} \mid r \geq 0\}$ with the following properties:

(M1) for all $x, y \in X$, $d(x, y) = 0$ if and only if $x = y$ (*separation*),

(M2) for all $x, y \in X$, $d(x, y) = d(y, x)$ (*symmetry*),

(M3) for all $x, y, z \in X$, $d(x, z) \leq d(x, y) + d(y, z)$ (*triangle inequality*).

The mapping d is called the *metric* (or *distance function*) for the space. We often call elements of X *points*.

Remark 1.1.2. Sometimes the notation r or ρ (Greek letter "rho") is used instead of d.

Remark 1.1.3. It is possible to define many different metrics on a set X. A metric space is the set X *together with* the metric d. However, if the metric is clear from the context (e.g. if we are working with a set X on which we pick a metric d once and for all, and never work with a different one), then we will sometimes refer to the set X itself as a metric space, leaving the metric d inferred.

Remark 1.1.4. The fact that d takes *non-negative* values actually follows from the axioms. If $d\colon X \times X \to \mathbb{R}$ is *any* mapping satisfying (M1)–(M3), then

$$0 = d(x, x) \leq d(x, y) + d(y, x) = 2d(x, y) \implies d(x, y) \geq 0,$$

for all $x, y \in X$.

Example 1.1.5 (Standard metric on $\mathbb{R}$). If X is any nonempty set of real numbers (for instance, if $X = \mathbb{R}$) then

$$d(x, y) = |x - y|, \quad x, y \in \mathbb{R},$$

defines a metric on X. This is called the *standard*, *usual*, or *natural metric* on such a set X. Whenever we refer to $\mathbb{R}$ as a metric space, without specifying the metric, we mean this standard metric.

Definition 1.1.6 (Metric subspace)**.** Suppose (X, d) is a metric space and Y is a nonempty subset of X. Let $d|_Y$ be the restriction of d to Y (i.e. the restriction of the mapping d to $Y \times Y$). That is, for all $y, z \in Y$,

$$d|_Y(y, z) = d(y, z).$$

Then $(Y, d|_Y)$ is a metric space, called a *metric subspace* of (X, d). The metric $d|_Y$ is called the *induced metric*.

Example 1.1.7. We have already seen an example of induced metrics when we defined the standard metric on *any* subspace of $\mathbb{R}$. In essence, we defined a metric on $\mathbb{R}$ and thus have the induced metric on any nonempty subset of $\mathbb{R}$, such as $\mathbb{N}$, $\mathbb{Q}$, $\{1/n \mid n \in \mathbb{N}_+\}$, etc.

We will return to the subject of metric subspaces in more detail a bit later.

Example 1.1.8 (Discrete metric). For an arbitrary nonempty set X, one can define the *discrete metric* (or *trivial metric*, or *0-1 metric*) by

$$d_{0-1}(x, y) = \begin{cases} 1 & \text{if } x \neq y, \\ 0 & \text{if } x = y. \end{cases}$$

Example 1.1.9 (Euclidean metric). Consider the set $\mathbb{R}^2$ and define d by

$$d(x,y) = \sqrt{(x_1-y_1)^2+(x_2-y_2)^2} \quad \text{for } x=(x_1,x_2),\ y=(y_1,y_2)\in\mathbb{R}^2.$$

This is the usual definition of distance between points in the plane and axiom (M3) is the usual triangle inequality.

More generally, on $\mathbb{R}^n$ define d by

$$d(x,y) = \sqrt{\sum_{k=1}^n (x_k-y_k)^2} \quad \text{for } x=(x_1,\dots,x_n),\ y=(y_1,\dots,y_n)\in\mathbb{R}^n.$$

We will prove that this defines a metric on $\mathbb{R}^n$, called the *Euclidean metric*. Whenever we refer to $\mathbb{R}^n$ as a metric space, we will always imply that the Euclidean metric is used. The metric space $\mathbb{R}^n$ (with this metric) is called *Euclidean space* or *Euclidean n-space*. Note that the $n=1$ case is the metric of Example 1.1.5 and the $n=2$ case is the metric defined above.

It is easy to see that axioms (M1) and (M2) are satisfied by the Euclidean metric. To prove (M3) is also satisfied, we need the following.

Proposition 1.1.10 (Cauchy–Schwarz inequality)**.** *Let* $(a_1,\dots,a_n),(b_1,\dots,b_n)\in\mathbb{R}^n$. *Then*

$$\left(\sum_{k=1}^n a_k b_k\right)^2 \le \left(\sum_{k=1}^n a_k^2\right)\left(\sum_{k=1}^n b_k^2\right). \tag{1.1}$$

Proof. Define the function $\psi\colon \mathbb{R}\to\mathbb{R}$ by

$$\psi(u) = \sum_{k=1}^n (a_k u + b_k)^2, \quad u\in\mathbb{R}.$$

Expanding the squares gives

$$\psi(u) = \left(\sum_{k=1}^n a_k^2\right)u^2 + 2\left(\sum_{k=1}^n a_k b_k\right)u + \sum_{k=1}^n b_k^2.$$

Note that $\psi(u)$ is of the form $Au^2+2Bu+C$, where

$$A=\sum_{k=1}^n a_k^2, \qquad B=\sum_{k=1}^n a_k b_k, \qquad C=\sum_{k=1}^n b_k^2.$$

Since $\psi(u)$ is a sum of squares, we have $\psi(u)\ge 0$ for all $u\in\mathbb{R}$. Therefore, since ψ is a polynomial in u of degree at most two, the graph of ψ is an upwards facing parabola which can have at most one zero. Therefore, the discriminant $(2B)^2-4AC$ cannot be positive. Thus,

$$(2B)^2-4AC\le 0 \implies B^2-AC\le 0 \implies \left(\sum_{k=1}^n a_k b_k\right)^2 - \left(\sum_{k=1}^n a_k^2\right)\left(\sum_{k=1}^n b_k^2\right)\le 0. \qquad \square$$

Proposition 1.1.11. *If* $(a_1, \ldots, a_n), (b_1, \ldots, b_n) \in \mathbb{R}^n$, *then*

$$\sqrt{\sum_{k=1}^{n}(a_k+b_k)^2} \le \sqrt{\sum_{k=1}^{n} a_k^2} + \sqrt{\sum_{k=1}^{n} b_k^2}. \tag{1.2}$$

Proof. First take the square root of both sides of the Cauchy–Schwarz inequality (1.1) to get

$$\sum_{k=1}^{n} a_k b_k \le \left|\sum_{k=1}^{n} a_k b_k\right| \le \sqrt{\sum_{k=1}^{n} a_k^2}\sqrt{\sum_{k=1}^{n} b_k^2}.$$

Then

$$\sum_{k=1}^{n} a_k^2 + 2\sum_{k=1}^{n} a_k b_k + \sum_{k=1}^{n} b_k^2 \le \sum_{k=1}^{n} a_k^2 + 2\sqrt{\sum_{k=1}^{n} a_k^2}\sqrt{\sum_{k=1}^{n} b_k^2} + \sum_{k=1}^{n} b_k^2,$$

and so

$$\sum_{k=1}^{n}(a_k+b_k)^2 \le \left(\sqrt{\sum_{k=1}^{n} a_k^2} + \sqrt{\sum_{k=1}^{n} b_k^2}\right)^2.$$

Taking square roots of both sides gives the inequality (1.2). □

Corollary 1.1.12. *The Euclidean metric satisfies the triangle inequality* (M3) *and so is indeed a metric on* $\mathbb{R}^n$.

Proof. Simply take $a_k = x_k - y_k$ and $b_k = y_k - z_k$ in Proposition 1.1.11. □

Example 1.1.13 (ℓ^p-metric on $\mathbb{R}^n$). It is possible to generalize the above Euclidean metric and define a family of metrics d_p, $p \ge 1$, on $\mathbb{R}^n$ by

$$d_p(x,y) = \left(\sum_{k=1}^{n} |x_k - y_k|^p\right)^{\frac{1}{p}}. \tag{1.3}$$

This is called the *ℓ^p-metric* on $\mathbb{R}^n$. One can prove that (1.3) defines a metric, but the verification of (M3) is more difficult (it is the so-called *Hölder inequality*). Note that d_2 is just the Euclidean metric.

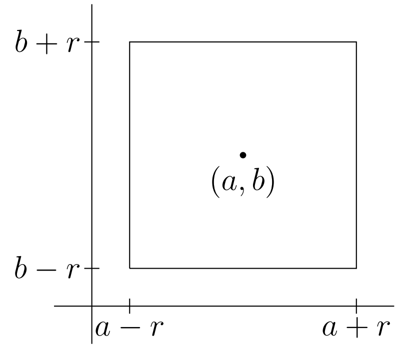
Example 1.1.14 (ℓ^∞-metric on $\mathbb{R}^n$). One can consider the limiting case $p \to \infty$ of the ℓ^p-metric. Namely, define the *ℓ^∞-metric* on $\mathbb{R}^n$ by

$$d_\infty(x,y) = \max_{k=1}^{n} |x_k - y_k|.$$

See Exercises 1.1.8 and 1.1.9.

To emphasize that one must be careful about making assumptions based on intuition in a general metric space, consider the circle in $\mathbb{R}^2$ of radius r with centre (a,b) when one is using the metric d_∞. This is the set

$$\{(x,y) \in \mathbb{R}^2 \mid d_\infty((a,b),(x,y)) = r\} = \{(x,y) \in \mathbb{R}^2 \mid \max\{|x-a|, |y-b|\} = r\}.$$



So our "circle" is actually a square!

Example 1.1.15 (ℓ^p and Euclidean metrics on $\mathbb{C}^n$). We can define the ℓ^p-metric d_p, $p \geq 1$, on $\mathbb{C}^n$ by exactly the same formula (1.3). In the case $p = 2$, we often denote it simply by d,

$$d(x, y) = \sqrt{\sum_{k=1}^{n} |x_k - y_k|^2}, \quad x, y \in \mathbb{C}^n.$$

This is called the *Euclidean metric* on $\mathbb{C}^n$. Whenever we refer to $\mathbb{C}^n$ as a metric space, we imply this metric.

Remark 1.1.16. The term *Euclidean space* typically refers to $\mathbb{R}^n$ with the standard metric and not $\mathbb{C}^n$.

Remark 1.1.17. It is important to remember that in an abstract metric space, we *only* have a notion of distance and there is no additional structure like addition, multiplication, etc. While some of our particular examples (like $\mathbb{C}^n$ and $\mathbb{R}^n$) have such structure, these are special cases. One must resist the urge to add or subtract points in general metric spaces.

We can attempt to generalize our example of $\mathbb{C}^n$ with the Euclidean metric by taking $n \to \infty$ (then n-tuples become infinite sequences). The Euclidean metric then becomes an infinite series and one must be careful about convergence.

Definition 1.1.18 (ℓ^2 space)**.** Denote by ℓ^2 the set of all complex-valued sequences $x_1, x_2, \ldots$ for which the series $\sum_{k=1}^{\infty} |x_k|^2$ converges. Define a metric on ℓ^2 by

$$d(x, y) = \sqrt{\sum_{k=1}^{\infty} |x_k - y_k|^2},$$

where $x = (x_1, x_2, \ldots)$, $y = (y_1, y_2, \ldots)$. We also denote the metric space itself by ℓ^2.

In order for our definition to be valid, we have to show that the infinite sum defining $d(x, y)$ converges for any $x, y \in \ell^2$ and that d satisfies the axioms of a metric.

We first show that $d(x, y)$ is always finite for $x, y \in \ell^2$. We have

$$\sqrt{\sum_{k=1}^{n} |x_k - y_k|^2} \leq \sqrt{\sum_{k=1}^{n} (|x_k| + |y_k|)^2} \leq \sqrt{\sum_{k=1}^{n} |x_k|^2} + \sqrt{\sum_{k=1}^{n} |y_k|^2},$$

where in the first inequality we used the fact that $|x_k - y_k| \leq |x_k| + |y_k|$ for all k (this is the triangle inequality in Euclidean 1-space) and in the second inequality we took $a_k = |x_k|$, $b_k = |y_k|$ in Proposition 1.1.11. On the right-hand side, we have the partial sums for the series $\sum_{k=1}^{\infty} |x_k|^2$ and $\sum_{k=1}^{\infty} |y_k|^2$. These series converge since this is precisely the condition that $x, y \in \ell^2$. Since the terms of these series are nonnegative, we have

$$\sqrt{\sum_{k=1}^{n} |x_k - y_k|^2} \leq \sqrt{\sum_{k=1}^{\infty} |x_k|^2} + \sqrt{\sum_{k=1}^{\infty} |y_k|^2}.$$

This shows that the partial sums of the series $\sum_{k=1}^{\infty} |x_k - y_k|^2$ form a bounded sequence. This sequence of partial sums is increasing since each of the terms $|x_k - y_k| \geq 0$. Thus, the series converges and we have that

$$d(x,y) = \sqrt{\sum_{k=1}^{\infty} (|x_k| + |y_k|)^2} \leq \sqrt{\sum_{k=1}^{\infty} |x_k|^2} + \sqrt{\sum_{k=1}^{\infty} |y_k|^2}.$$

Thus $d(x,y)$ is finite.

We now need to check that d is a metric (i.e. that it satisfies axioms (M1)–(M3)). Axioms (M1) and (M2) are obvious, so it remains to show the triangle inequality. Suppose $x = (x_1, x_2, \dots), y = (y_1, y_2, \dots), z = (z_1, z_2, \dots) \in \ell^2$. By the triangle inequality in $\mathbb{R}$ (Euclidean 1-space), we have

$$|x_k - z_k| = |(x_k - y_k) + (y_k - z_k)| \leq |x_k - y_k| + |y_k - z_k|.$$

Thus,

$$\sqrt{\sum_{k=1}^{n} |x_k - z_k|^2} \leq \sqrt{\sum_{k=1}^{n} (|x_k - y_k| + |y_k - z_k|)^2} \leq \sqrt{\sum_{k=1}^{n} |x_k - y_k|^2} + \sqrt{\sum_{k=1}^{n} |y_k - z_k|^2}$$

by taking $a_k = |x_k - y_k|$ and $b_k = |y_k - z_k|$ in Proposition 1.1.11. Then, by the same argument as above, we have

$$\sqrt{\sum_{k=1}^{\infty} |x_k - z_k|^2} \leq \sqrt{\sum_{k=1}^{\infty} |x_k - y_k|^2} + \sqrt{\sum_{k=1}^{\infty} |y_k - z_k|^2},$$

which is precisely the statement $d(x,z) \leq d(x,y) + d(y,z)$.

Definition 1.1.19 (ℓ^p space)**.** For $p \geq 1$, denote by ℓ^p the set of all complex-valued sequences $x_1, x_2, \dots$ for which the series $\sum_{k=1}^{\infty} |x_k|^p$ converges. Define a metric on ℓ^p by

$$d(x,y) = \left(\sum_{k=1}^{\infty} |x_k - y_k|^p \right)^{1/p},$$

where $x = (x_1, x_2, \dots)$, $y = (y_1, y_2, \dots)$. We also denote the metric space itself by ℓ^p.

One can show that the metric on ℓ^p is well-defined (i.e. the sums involved converge) and satisfies axioms (M1)–(M3) but we will not do so in this course.

Remark 1.1.20. Why do we restrict ourselves to $p \geq 1$? The reason is that the triangle inequality is no longer satisfied for $0 < p < 1$. However, for these values of p,

$$d(x, y) = \sum_{k=1}^{\infty} |x_k - y_k|^p$$

(without the pth root) still defines a metric.

Remark 1.1.21. Fix a natural number n. Note that if $x = (x_1, x_2, \dots)$ is a sequence such that $x_k = 0$ for all $k > n$, then $x \in \ell^p$ for any p. In this way, we can view $\mathbb{R}^n$ as a metric subspace of ℓ^p. The induced metric is simply the ℓ^p-metric we defined on $\mathbb{R}^n$ earlier in Example 1.1.13, hence the notation. Note that [Coh03] uses the notation l_p instead of ℓ^p.

Example 1.1.22 (ℓ^1 space). Taking $p = 1$ in the definition of ℓ^1 space (Definition 1.1.19), we see that ℓ^1 is the set of all complex-valued sequences $x_1, x_2, \dots$ for which the series $\sum_{k=1}^{\infty} |x_k|$ converges. The metric is given by

$$d(x, y) = \sum_{k=1}^{\infty} |x_k - y_k|.$$

As we did for $\mathbb{R}^n$, we can take the limit $p \to \infty$ in the definition of ℓ^p space to obtain the following.

Example 1.1.23 (ℓ^∞ space). Denote by ℓ^∞ the set of all complex-valued bounded sequences. (Recall that a sequence (x_n) is *bounded* if there exists a number $L \geq 0$ such that $|x_n| \leq L$ for all n.) Then

$$d(x, y) = \sup_{n \in \mathbb{N}_+} |x_n - y_n|$$

is well-defined for every $x, y \in \ell^\infty$ (due to the boundedness of both x and y and therefore of their difference) and one can verify that it defines a metric on ℓ^∞. See Exercise 1.1.13.

We now consider some metrics on spaces of functions. For $a, b \in \mathbb{R}$ with $a \leq b$, let $C[a, b]$ denote the space of all continuous functions on the closed interval $[a, b]$.

Example 1.1.24 (Uniform metric on $C[a, b]$). Define

$$d(x, y) = \max_{a \leq t \leq b} |x(t) - y(t)|, \quad x, y \in C[a, b].$$

Note that $|x - y|$ is a continuous function since x and y are (and the absolute value function is continuous). Therefore, it attains a maximum on the closed interval $[a, b]$ and so d is well-defined. It is called the *uniform metric* (or *sup metric*) on $C[a, b]$. We check that it is indeed a metric. It clearly takes nonnegative values. It is also clear that $d(x, x) = 0$. If $x \neq y$, then there exists a $t_0 \in [a, b]$ such that $x(t_0) \neq y(t_0)$, so

$$d(x, y) = \max_{a \leq t \leq b} |x(t) - y(t)| \geq |x(t_0) - y(t_0)| > 0,$$

and so $d(x, y) \neq 0$. Thus (M1) is satisfied. It is clear that (M2) is also satisfied. It remains to show (M3). Suppose $x, y, z \in C[a, b]$ and $s \in [a, b]$. Then, by the triangle inequality for $\mathbb{R}$, we have

$$\begin{aligned} |x(s) - z(s)| &\leq |x(s) - y(s)| + |y(s) - z(s)| \\ &\leq \max_{a \leq t \leq b} |x(t) - y(t)| + \max_{a \leq t \leq b} |y(t) - z(t)| \\ &= d(x, y) + d(y, z). \end{aligned}$$

Hence

$$d(x, z) = \max_{a \leq t \leq b} |x(t) - z(t)| \leq d(x, y) + d(y, z),$$

and so (M3) is satisfied. We will refer to the space $C[a, b]$ with this metric simply by $C[a, b]$.

Example 1.1.25 (The metric space $C_1[a, b]$). We define another metric on the space $C[a, b]$ by

$$d(x, y) = \int_a^b |x(t) - y(t)| \, dt.$$

Since $|x - y|$ is continuous on the closed interval $[a, b]$, it is integrable on that interval. So d is well-defined and clearly takes nonnegative values. This is perhaps our first example where (M1) is not completely obvious. It is clear that $x = y \implies d(x, y) = 0$, but the reverse implication requires some argument. Here we have to use the fact that $|x - y|$ is continuous. Suppose $x \neq y$. Then there exists a $t_0 \in [a, b]$ such that $x(t_0) \neq y(t_0)$. Let $\epsilon = |x(t_0) - y(t_0)| > 0$. Then, since $|x - y|$ is continuous, there exists an interval I containing t_0 of length $\delta > 0$ such that $|x(t) - y(t)| > \epsilon/2$ for all $t \in I$. Then

$$d(x, y) = \int_a^b |x(t) - y(t)| \, dt \geq \delta\epsilon/2 > 0.$$

Thus (M1) is satisfied. Condition (M2) is clear. Now suppose $x, y, z \in C[a, b]$. Then for all $a \leq t \leq b$, we have

$$|x(t) - z(t)| \leq |x(t) - y(t)| + |y(t) - z(y)|.$$

Thus

$$\begin{aligned} d(x, z) = \int_a^b |x(t) - z(t)| \, dt &\leq \int_a^b \Big(|x(t) - y(t)| + |y(t) - z(t)| \Big) \, dt \\ &= \int_a^b |x(t) - y(t)| \, dt + \int_a^b |y(t) - z(t)| \, dt = d(x, y) + d(y, z), \end{aligned}$$

and so (M3) is satisfied. We denote the space $C[a, b]$ with this metric by $C_1[a, b]$.

Example 1.1.26 (The metric space $C_p[a, b]$). For $p > 0$, we can define a metric on $C[a, b]$ by

$$d(x, y) = \left(\int_a^b |x(t) - y(t)|^p \, dt \right)^{1/p}, \quad x, y \in C[a, b].$$

We denote $C[a,b]$ with this metric by $C_p[a,b]$. We will not prove that this is a metric for general p. Note that the $p = 1$ case is simply the metric space of Example 1.1.25. In addition, the metric space of Example 1.1.24 can be viewed as the $p \to \infty$ limit of $C_p[a,b]$. That the $p = 2$ case is a metric space can be proved using an integral version of the Cauchy–Schwarz inequality. See Exercise 1.1.14.

Example 1.1.27 (Baire space). Let $\mathbb{Z}^{\mathbb{N}_+}$ denote the set of all infinite integer sequences $x = (x_1, x_2, \dots)$, $x_i \in \mathbb{Z}$. For $x = (x_1, x_2, \dots), y = (y_1, y_2, \dots) \in \mathbb{Z}^{\mathbb{N}_+}$, define

$$d(x,y) = \begin{cases} 0 & \text{if } x = y, \\ 2^{-\min\{i \,|\, x_i \neq y_i\}} & \text{otherwise.} \end{cases}$$

For example, the distance between

$$(1,2,3,4,5,\dots,) \quad \text{and} \quad (1,2,3,0,0,0,\dots)$$

is $2^{-4} = 1/16$. It is easy to see that axioms (M1) and (M2) are satisfied. The triangle inequality (M3) can be verified as follows. Suppose $x, y, z \in \mathbb{Z}^{\mathbb{N}_+}$. We wish to prove that

$$d(x,z) \le d(x,y) + d(y,z). \tag{1.4}$$

If at least any two of the elements x, y, z are equal, then the inequality (1.4) is trivially satisfied (check this for yourself). Thus we can assume that x, y, z are all different.

Let i be the smallest natural number such that $x_j = y_j = z_j$ for all $j < i$ while x_i, y_i, z_i are not all the same. Such a number exists by our assumption. It follows that $d(x,z) \le 2^{-i}$. However, at least one of the numbers $d(x,y)$ and $d(y,z)$ is exactly equal to 2^{-i} (otherwise we would have $x_i = y_i = z_i$, contradicting our choice of i). Thus the number on the right-hand side of (1.4) is $\ge 2^{-i}$ and so (1.4) is satisfied.

Exercises.

1.1.1. Suppose X is a nonempty set and d is the discrete metric on X. Describe the sets

$$\{x \in X \mid d(x,y) \le r\}, \quad y \in X, \quad r \in \mathbb{R}_{\ge 0}.$$

Hint: Your answer should depend on r.

1.1.2. What is the only possible metric on a singleton (a set with one element)?

1.1.3. (a) Suppose (X,d) is a metric space and $\alpha \in \mathbb{R}$, $\alpha > 0$. Show that αd, defined by

$$(\alpha d)(x,y) = \alpha d(x,y), \quad x, y \in X,$$

is a metric on X.

(b) Show that if X is a set with more than one element, then there are infinitely many metrics on X.

1.1.4 ([Coh03, Ex. 2.4(1)]). If (X, d) is a metric space, and $x, y, z, u \in X$, prove that

$$|d(x, z) - d(y, u)| \leq d(x, y) + d(z, u).$$

1.1.5 ([Coh03, Ex. 2.4(2)]). If (X, d) is a metric space and $x_1, x_2, \ldots, x_n \in X$ $(n \geq 2)$, prove that

$$d(x_1, x_n) \leq d(x_1, x_2) + d(x_2, x_3) + \cdots + d(x_{n-1}, x_n).$$

1.1.6 ([Coh03, Ex. 2.4(3)]). Suppose d_1 and d_2 are two metrics on the same set X. Show that d_3 and d_4, given by

$$d_3(x, y) = d_1(x, y) + d_2(x, y), \quad x, y \in X,$$
$$d_4(x, y) = \max\{d_1(x, y), d_2(x, y)\}, \quad x, y \in X,$$

are also metrics on X.

1.1.7. Note that, when $p = 1$, the metric defined in (1.3) becomes

$$d_1(x, y) = \sum_{k=1}^{n} |x_k - y_k|, \quad x, y \in \mathbb{R}^n.$$

Verify that this does indeed define a metric on $\mathbb{R}^n$.

1.1.8. Justify the limit statement made in Example 1.1.14. More precisely, show that if $(x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$, then

$$\lim_{p \to \infty} \left(\sum_{k=1}^{n} |x_k - y_k| \right)^{1/p} = \max_{k=1}^{n} |x_k - y_k|.$$

1.1.9. Verify that d_∞ (see Example 1.1.14) is indeed a metric on $\mathbb{R}^n$.

1.1.10. Show that $d(x, y)$, as defined in Example 1.1.22, converges for all $x, y \in \ell^1$, and that d defines a metric on ℓ^1.

1.1.11. Draw the sets

$$\{(x, y) \in \mathbb{R}^2 \mid d_p((0, 0), (x, y)) = 1\}$$

for $p = \frac{1}{2}, 1, 2, 3, \infty$.

1.1.12. Show that ℓ^1 is indeed a metric space. Namely, show that its metric is well-defined (i.e. the relevant series converge) and that it satisfies axioms (M1)–(M3).

1.1.13. Verify that the metric of ℓ^∞ (see Example 1.1.23) satisfies axioms (M1)–(M3).

1.1.14 ([Coh03, Ex. 2.4(6)]). Let f and g be continuous functions defined on $[a, b]$.

(a) Derive the integral form of the Cauchy–Schwarz inequality:

$$\left(\int_a^b f(t)g(t)\,dt\right)^2 \le \left(\int_a^b \big(f(t)\big)^2\,dt\right)\left(\int_a^b \big(g(t)\big)^2\,dt\right).$$

(b) Show that there is equality if and only if $f = \beta g$ for some constant β.

(c) Use this Cauchy–Schwarz inequality to deduce the triangle inequality for the mapping d of Example 1.1.26 with $p = 2$.

1.1.15 ([Coh03, Ex. 2.4(7)]). Let X be the set of all continuous functions defined on the whole real line which are zero outside some interval (not necessarily the same interval for different functions). Show that

$$d(x,y) = \max_{t\in\mathbb{R}} |x(t) - y(t)|, \quad x,y \in X,$$

defines a metric on X.

1.1.16 ([Coh03, Ex. 2.4(8)]). Fix $n \in \mathbb{N}_+$, and let X be the set of all $n \times n$ matrices with complex entries. Show that d_1 and d_2, defined by

$$d_1(A,B) = \max_{\substack{1\le j\le n\\ 1\le k\le n}} |a_{j,k} - b_{j,k}|, \qquad d_2(A,B) = \max_{1\le k\le n} \sum_{j=1}^{n} |a_{j,k} - b_{j,k}|,$$

where $A = (a_{j,k}), B = (b_{j,k}) \in X$, are both metrics on X.

1.1.17 ([Coh03, Ex. 2.4(11)]). Let X be any nonempty set and let $\rho\colon X \times X \to \mathbb{R}$ be a mapping such that

(a) for all $x, y \in X$, we have $\rho(x,y) = 0$ if and only if $x = y$, and

(b) $\rho(x,y) \le \rho(z,x) + \rho(z,y)$ for all $x,y,z \in X$.

Show that ρ is a metric on X.

1.1.18 ([Coh03, Ex. 2.4(12)]). We say (X,d) is a *semimetric space* if X is a nonempty set and d is a mapping $d\colon X \times X \to \mathbb{R}_{\ge 0}$ satisfying $d(x,x) = 0$ for all $x \in X$ and (M2) and (M3). (In other words, we drop the "only if" requirement in (M1).) Show that (X,d) is a semimetric space, but not a metric space, when

(a) X is the set of all integrable functions on $[a,b]$ and

$$d(f,g) = \int_a^b |f(t) - g(t)|\,dt, \quad f,g \in X;$$

(b) X is the set of all differentiable functions on $[a,b]$ and

$$d(f,g) = \max_{a\le t\le b} |f'(t) - g'(t)|, \quad f,g \in X.$$

1.2 Convergence in a metric space

Having defined metric spaces and seen a number of examples, we turn our attention to the notion of convergence.

Definition 1.2.1 (Sequence)**.** A *sequence* is a mapping whose domain is the set $\mathbb{N}_+$ of positive integers. If X is any set, a *sequence in* X is a mapping $\mathbb{N}_+ \to X$. If (X, d) is a metric space, a sequence in (X, d) is a sequence in X.

Since many of our examples of metric spaces (such as ℓ^p) involve sequences, we adopt a system of notation to avoid confusing sequences which are elements *of* a metric space with sequences *in* that metric space.

- We use the notation $(x_1, x_2, \dots)$ to denote sequences which are elements of a metric space, such as ℓ^p.
- We use the notation $\{x_n\}_{n=1}^\infty$ or $\{x_n\}$ or $x_1, x_2, \dots$ to denote a sequence of elements of a metric space.

Definition 1.2.2 (Convergence)**.** A sequence $\{x_n\}$ in a metric space (X, d) is said to *converge* to an element $x \in X$ if $\lim_{n\to\infty} d(x_n, x) = 0$. That is, $\{x_n\}$ converges to x if for any $\epsilon > 0$ there exists a positive integer N such that

$$n > N \implies d(x_n, x) < \epsilon.$$

Then x is called the *limit* of the sequence and we write $x_n \to x$, or $\lim_{n\to\infty} x_n = x$, or simply $\lim x_n = x$.

Remark 1.2.3. Note that the point to which a sequence converges must be a point of the metric space. For instance, the sequence $\left\{\frac{1}{n}\right\}$ does *not* converge in $(0, 1]$ (with the usual metric) since $0 \notin (0, 1]$. Similarly, consider the metric space $(\mathbb{Q}, d)$ where d is the natural metric. Then the sequence $\{x_n\}$ where

$$x_n = \sum_{k=0}^{n} \frac{1}{k!}$$

does *not* converge in this metric space (since $e \notin \mathbb{Q}$).

Remark 1.2.4. If d and d' are two metrics on a set X, it is possible for a sequence in X to converge in (X, d) but not in (X, d'). For instance, suppose $X = \mathbb{R}$, d is the natural metric, d' is the discrete metric, and $\{x_n\}$ is either of the sequences of Remark 1.2.3. Then $\{x_n\}$ converges (to e or 0) in (X, d) but does not converge in (X, d'), since $d'(x_n, e) = d'(x_n, 0) = 1$ for all n.

Proposition 1.2.5. *If a sequence in a metric space is convergent, then the limit is unique.*

Proof. Suppose $\{x_n\}$ is a convergent sequence in a metric space (X, d) and both $x_n \to x$ and $x_n \to y$ for some $x, y \in X$. Then, for any $n \in \mathbb{N}_+$,

$$0 \le d(x, y) \le d(x, x_n) + d(x_n, y) = d(x_n, x) + d(x_n, y).$$

Since $d(x_n, x) \to 0$ and $d(x_n, y) \to 0$, we must have $d(x, y) = 0$ and hence $x = y$. □

Let us take a closer look at convergence in $\mathbb{C}^m$ with the usual metric d. Let $\{x_n\}$ be a sequence in this space. Each x_n is an ordered m-tuple. We write

$$x_n = (x_{n,1}, x_{n,2}, \dots, x_{n,m}), \quad n \in \mathbb{N}_+.$$

Suppose that the sequence converges to an element $y = (y_1, \dots, y_n) \in \mathbb{C}^m$. Thus, by definition,

$$d(x_n, y) = \sqrt{\sum_{k=1}^{m} |x_{n,k} - y_k|^2} \to 0.$$

Note that for each $j = 1, 2, \dots, m$, we have

$$0 \le |x_{n,j} - y_j| \le \sqrt{\sum_{k=1}^{m} |x_{n,k} - y_k|^2}.$$

Thus, for each j, we have $x_{n,j} \to y_j$ as $n \to \infty$. In other words, each of the sequences $\{x_{n,j}\}_{n=1}^{\infty}$, $j = 1, 2, \dots, n$, is convergent (and converges to y_j). Conversely, if $x_{n,j} \to y_j$ for each j, then $d(x_n, y) \to 0$. Therefore, convergence of a sequence in $\mathbb{C}^m$ is equivalent to convergence of components. The same is true for $\mathbb{R}^m$, by exactly the same argument.

Now consider the metric space ℓ^2. Suppose $\{x_n\}$ is a convergent sequence in ℓ^2, with limit $y = (y_1, y_2, \dots)$. Say

$$x_n = (x_{n,1}, x_{n,2}, \dots), \quad n \in \mathbb{N}_+.$$

Then, for each $j \in \mathbb{N}_+$,

$$0 \le |x_{n,j} - y_j| \le \sqrt{\sum_{k=1}^{\infty} |x_{n,k} - y_k|^2} = d(x_n, y) \to 0,$$

and hence $x_{n,j} \to y_j$ as $n \to \infty$. Hence convergence of a sequence in ℓ^2 implies convergence of components.

However, unlike in $\mathbb{C}^m$, the converse is not true. Consider the sequence $\{e_n\}$, where

$$\begin{aligned} e_1 &= (1, 0, 0, 0, \dots), \\ e_2 &= (0, 1, 0, 0, \dots), \\ e_3 &= (0, 0, 1, 0, \dots), \\ &\vdots \end{aligned}$$

The sequence of kth components converges to 0 for each k, since every component is eventually zero (more precisely, the kth component of e_j is zero for $j > k$). However, $y = (0, 0, 0, \dots)$ is not $\lim e_n$ since $d(e_n, y) = 1$ for all n. In fact, it will follow from later results that the sequence $\{e_n\}$ does not converge at all. This is essentially because the terms in the sequence do not get closer together; rather they stay at a distance 1 from each other.

Exercises.

1.2.1. Show that if we replace the usual metric on $\mathbb{C}^m$ (or $\mathbb{R}^m$) by the ℓ^p-metric, it remains true that convergence in this metric space is equivalent to convergence of components.

1.2.2 ([Coh03, Ex. 2.9(14)]). In a semimetric space (see Exercise 1.1.18), convergence of a sequence is defined as it is in a metric space. Let (X, d) be the semimetric space of Exercise 1.1.18(b), with $a = 0$, $b = \frac{1}{2}$. Show that the sequence $\{x_n\}$, where $x_n(t) = t^n$ $(0 \le t \le \frac{1}{2})$ is convergent and that any constant function on $[0, \frac{1}{2}]$ serves as its limit. (Hence, convergent sequences in semimetric spaces need not have unique limits.)

1.3 Sequences of functions

In this section we discuss sequences of functions and various types of convergence for such sequences.

Suppose $\{f_n\}$ is a sequence of real-valued (or complex-valued) functions, all having the same domain D.

Definition 1.3.1 (Pointwise convergence)**.** We say that the sequence $\{f_n\}$ *converges pointwise* if, for all $x \in D$, the sequence $\{f_n(x)\}$ of real (or complex) numbers converges. In this case, the function f with domain D defined by $f(x) = \lim f_n(x)$, $x \in D$, is called the *pointwise limit* of $\{f_n\}$. We write $\lim f_n = f$ or $f_n \to f$.

Remark 1.3.2. Another way of writing the definition of pointwise convergence is that $\{f_n\}$ converges pointwise to f if for all $\epsilon > 0$ and $x \in D$, there exists a positive integer $N(x)$ such that

$$n > N(x) \implies |f_n(x) - f(x)| < \epsilon.$$

It is important to note that $N(x)$ depends on x. That is, we can choose a different value of $N(x)$ for each x (and ϵ). It is for this reason that we use the notation $N(x)$, instead of simply N.

Example 1.3.3. For $n \in \mathbb{N}_+$, let $f_n(x) = x/n$, $x \in \mathbb{R}$. Then $\{f_n\}$ is a sequence of functions with domain $\mathbb{R}$. This sequence converges pointwise to $f(x) = 0$, the zero function.

Definition 1.3.4 (Uniform convergence)**.** Let $\{f_n\}$ be a sequence of real-valued (or complex-valued) functions with domain D and f be another function with the same domain. We say that the sequence $\{f_n\}$ *converges uniformly* to f if for all $\epsilon > 0$, there exists a positive integer N such that

$$n > N \implies |f_n(x) - f(x)| < \epsilon \quad \forall\, x \in D.$$

We write $f_n \rightrightarrows f$ and call f the *uniform limit* of the sequence $\{f_n\}$.

Remark 1.3.5. Note that, for each ϵ, we need to find *one* N that works for *all* x. This is different from the situation for pointwise convergence, where N can depend on x.

Examples 1.3.6. (a) Consider the sequence of Example 1.3.3. This sequence does *not* converge uniformly to the zero function. This can be seen as follows. For any choice of positive integer n, we have

$$|f_n(x) - 0| = \frac{|x|}{n} \geq 1 \text{ for all } x \geq n.$$

Thus, the condition in Definition 1.3.4 cannot be satisfied for any $\epsilon \leq 1$.

(b) For all $n \in \mathbb{N}_+$, define a function $f_n \colon \mathbb{R} \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} 2 + 1/n, & x \in \mathbb{Q}, \\ 2, & x \notin \mathbb{Q}. \end{cases}$$

Then $\{f_n\}$ converges uniformly to the constant function 2.

Proposition 1.3.7. *A sequence of functions $\{f_n\}$ in $C[a,b]$ (with the uniform metric) converges in this metric space if and only if it converges uniformly.*

Proof. Suppose $\{f_n\}$ is convergent in $C[a,b]$ (with the uniform metric) with limit f. Then

$$d(f_n, f) = \max_{a \leq t \leq b} |f_n(t) - f(t)| \to 0.$$

Therefore, by the definition of a limit, for every $\epsilon > 0$, we can find a positive integer N such that

$$n > N \implies \max_{a \leq t \leq b} |f_n(t) - f(t)| < \epsilon.$$

Thus

$$n > N \implies |f_n(t) - f(t)| < \epsilon \ \forall \ t \in [a,b].$$

Thus $f_n \rightrightarrows f$. The converse is left as an (easy) exercise. □

Remark 1.3.8. Proposition 1.3.7 justifies the name *uniform metric.*

Exercises.

1.3.1. Prove the converse statement whose proof was omitted in the proof of Proposition 1.3.7.

1.4 Complete metric spaces

We would like to be able to determine that a sequence converges without having to know the limit in advance. This should have something to do with the terms in a sequence getting closer and closer together.

Definition 1.4.1 (Cauchy sequence)**.** A sequence $\{x_n\}$ in a metric space (X, d) is called a *Cauchy sequence* if, for all $\epsilon > 0$, there exists a positive integer N such that

$$m, n > N \implies d(x_n, x_m) < \epsilon.$$

Definition 1.4.2 (Complete metric space)**.** A metric space is said to be *complete* if every Cauchy sequence in the space converges.

Theorem 1.4.3 (Completeness of the real numbers)**.** *The metric space $\mathbb{R}$ (with the usual metric) is complete.*

Proof. You have seen this result in previous analysis courses. It can also be found, for example, in [Coh03, Th. 1.7.12]. □

Example 1.4.4. The set $\mathbb{Q}$ of rational numbers is not complete. For example, consider the sequence $\{x_n\}$ where x_n is equal to the decimal expansion of π truncated after the nth decimal place:

$$\begin{aligned} x_1 &= 3.1 \\ x_2 &= 3.14 \\ x_3 &= 3.141 \\ &\vdots \end{aligned}$$

Then $\{x_n\}$ is a Cauchy sequence since $|x_n - x_m| \leq 10^{-N}$ for $n, m > N$. However, this sequence does not converge in $\mathbb{Q}$, since π is not a rational number.

Theorem 1.4.5. *Every convergent sequence in a metric space is a Cauchy sequence.*

Proof. Suppose $\{x_n\}$ is a convergent sequence in a metric space (X, d), with $\lim x_n = x$. Let $\epsilon > 0$. By definition of convergence of a sequence, there exists an integer N such that

$$n > N \implies d(x_n, x) < \frac{\epsilon}{2}.$$

Then, for all $n, m > N$, we have

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) = d(x_n, x) + d(x_m, x) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus $\{x_n\}$ is a Cauchy sequence. □

So in a metric space, every convergent sequence is a Cauchy sequence, but Cauchy sequences do not necessarily converge (unless the metric space is complete).

At the end of Section 1.2, we considered the sequence $e_1 = (1, 0, 0, \dots)$, $e_2 = (0, 1, 0, 0, \dots)$, $\dots$ in ℓ^2 and argued that this sequence did not converge to $(0, 0, 0, \dots)$ in ℓ^2 even though it did converge to this element componentwise. We can now prove that the sequence $\{e_n\}$ in fact does not converge at all. This is because when $n \neq m$, we have $d(e_n, e_m) = \sqrt{2}$. Thus $\{e_n\}$ is not a Cauchy sequence and so, by Theorem 1.4.5, it does not converge. Proving that a sequence does not converge is one important use of Theorem 1.4.5.

Example 1.4.6 ($\mathbb{C}$ is complete). We will show that the metric space $\mathbb{C}$, with the natural metric, is complete. Let $\{x_n = u_n + iv_n\}$ be a Cauchy sequence in $\mathbb{C}$, where $u_n, v_n \in \mathbb{R}$ for all $n \in \mathbb{N}_+$. Since $\{x_n\}$ is a Cauchy sequence, for all $\epsilon > 0$ there exists an N such that $m, n > N \implies |x_n - x_m| < \epsilon$. Now,

$$\begin{aligned} |x_n - x_m| &= |(u_n + iv_n) - (u_m + iv_m)| = |(u_n - u_m) + i(v_n - v_m)| \\ &= \sqrt{(u_n - u_m)^2 + (v_n - v_m)^2}. \end{aligned}$$

Therefore,

$$n, m > N \implies |u_n - u_m| \leq \sqrt{(u_n - u_m)^2 + (v_n - v_m)^2} = |x_n - x_m| < \epsilon.$$

Similarly,

$$n, m > N \implies |v_n - v_m| < \epsilon.$$

Thus $\{u_n\}$ and $\{v_n\}$ are Cauchy sequences in $\mathbb{R}$. Since $\mathbb{R}$ is complete, these sequences converge to some real numbers u and v (respectively). Let $x = u + iv \in \mathbb{C}$. Now, for any $\epsilon > 0$, there exists an N such that $|u_n - u| < \epsilon/2$ and $|v_n - v| < \epsilon/2$ for all $n > N$. Then

$$0 \leq d(x_n, x) = |x_n - x| = |(u_n + iv_n) - (u + iv)| = |(u_n - u) + i(v_n - v)| \leq |u_n - u| + |v_n - v| < \epsilon,$$

for all $n > N$. Thus, the sequence $\{x_n\}$ is convergent. Therefore, every Cauchy sequence in $\mathbb{C}$ converges and so $\mathbb{C}$ is complete.

Example 1.4.7 (ℓ^2 is complete). Let $\{x_n\}$ be a Cauchy sequence in ℓ^2, where $x_n = (x_{n,1}, x_{n,2}, \dots)$. By the definition of the space ℓ^2, the series $\sum_{k=1}^{\infty} |x_{n,k}|^2$ converges for all n. Since $\{x_n\}$ is a Cauchy sequence, for any $\epsilon > 0$, there exists an N such that

$$m, n > N \implies d(x_n, x_m) = \sqrt{\sum_{k=1}^{\infty} |x_{n,k} - x_{m,k}|^2} < \epsilon.$$

This implies that

$$m, n > N \implies \sum_{k=1}^{\infty} |x_{n,k} - x_{m,k}|^2 < \epsilon^2$$

and hence

$$m, n > N \implies |x_{n,k} - x_{m,k}| < \epsilon \quad \forall\, k \in \mathbb{N}_+.$$

Thus, for each $k \in \mathbb{N}_+$, $\{x_{n,k}\}$ is a Cauchy sequence in $\mathbb{C}$. Since $\mathbb{C}$ is complete, this sequence converges. Let $y_k = \lim_{n\to\infty} x_{n,k}$ and set $y = (y_1, y_2, \dots)$. We will prove that $x_n \to y$ as $n \to \infty$.

For any $r = 1, 2, \dots,$

$$\sum_{k=1}^{r} |x_{n,k} - x_{m,k}|^2 \le \sum_{k=1}^{\infty} |x_{n,k} - x_{m,k}|^2 < \epsilon^2 \text{ for all } m, n > N.$$

Thus, since $x_{m,k} \to y_k$ as $m \to \infty$, we have

$$\sum_{k=1}^{r} |x_{n,k} - y_k|^2 \le \epsilon^2 \text{ for all } n > N. \tag{1.5}$$

By the triangle inequality in $\mathbb{C}^r$, we have

$$\sqrt{\sum_{k=1}^{r} |y_k|^2} \le \sqrt{\sum_{k=1}^{r} |y_k - x_{n,k}|^2} + \sqrt{\sum_{k=1}^{r} |x_{n,k}|^2} \le \epsilon + \sqrt{\sum_{k=1}^{r} |x_{n,k}|^2} \le \epsilon + \sqrt{\sum_{k=1}^{\infty} |x_{n,k}|^2},$$

for all $n > N$. Now, the series $\sum_{k=1}^{\infty} |x_{n,k}|^2$ converges since $x_n \in \ell^2$. Therefore, the series $\sum_{k=1}^{r} |y_k|^2$ converges and hence $y \in \ell^2$. Then, by (1.5), we have

$$d(x_n, y) = \sqrt{\sum_{k=1}^{\infty} |x_{n,k} - y_k|^2} \le \epsilon \text{ for all } n > N.$$

Therefore $\{x_n\}$ converges to y in ℓ^2.

Remark 1.4.8 (ℓ^p is complete). An argument similar to the one above shows that in fact ℓ^p is complete for $p \ge 1$. See Exercise 1.4.3.

Example 1.4.9 ($\mathbb{R}^n$ and $\mathbb{C}^n$ are complete). The argument for ℓ^2 can easily be adapted to show that $\mathbb{R}^n$ and $\mathbb{C}^n$ are complete. Or, one can use the fact that $\mathbb{R}^n$ and $\mathbb{C}^n$ can be viewed as subspaces of ℓ^2. Then consider a Cauchy sequence $\{x_i\}$ in $\mathbb{C}^n$. Since this is also a sequence in ℓ^2, which is complete, this sequence converges to some point x in ℓ^2. We have seen that convergence in ℓ^2 implies convergence of components. But since the kth component of each x_i is zero for $k > n$, the same must be true of the limiting point x. Thus $x \in \mathbb{C}^n$. Since the standard metric on $\mathbb{C}^n$ is just the restriction of the metric of ℓ^2, we see that $\{x_i\}$ converges to x in $\mathbb{C}^n$. A similar argument works for $\mathbb{R}^n$.

Example 1.4.10. The argument of Example 1.4.9 also shows that $\mathbb{R}^n$ and $\mathbb{C}^n$ are complete in the ℓ^p- and ℓ^∞-metrics.

Example 1.4.11 ($C[a, b]$ is complete). Suppose $\{x_n\}$ is a Cauchy sequence in $C[a, b]$. Then for all $\epsilon > 0$, we can find an $N > 0$ such that

$$m, n > N \implies \max_{a \le t \le b} |x_n(t) - x_m(t)| < \epsilon.$$

Thus, for each $t \in [a,b]$, we have

$$m, n > N \implies |x_n(t) - x_m(t)| < \epsilon, \tag{1.6}$$

and so $\{x_n(t)\}$ is a Cauchy sequence in $\mathbb{R}$. Since $\mathbb{R}$ is complete, we know that the sequence $\{x_n(t)\}$ converges to some real number, which we denote by $x(t)$. This defines a function x on $[a,b]$. If we take the limit $m \to \infty$ in (1.6), we see that for all $t \in [a,b]$,

$$n > N \implies |x_n(t) - x(t)| \le \epsilon.$$

This implies that the sequence $\{x_n\}$ converges uniformly to x.

It remains to show that x is continuous. Fix $s \in [a,b]$. We will show that x is continuous at s. Choose $\epsilon > 0$. Since $\{x_n\}$ converges uniformly to x and $\{x_n(s)\}$ converges to $x(s)$, we can find $N > 0$ such that

$$n > N,\ t \in [a,b] \implies |x(t) - x_n(t)| < \epsilon/3,$$

and

$$n > N \implies |x_n(s) - x(s)| < \epsilon/3.$$

Now fix an $n > N$. Since x_n is continuous, we can choose a $\delta > 0$ such that

$$|t - s| < \delta,\ t \in [a,b] \implies |x_n(t) - x_n(s)| < \epsilon/3.$$

Then, whenever $|t - s| < \delta$, we have

$$\begin{aligned} |x(t) - x(s)| &= |(x(t) - x_n(t)) + (x_n(t) - x_n(s)) + (x_n(s) - x(s))| \\ &\le |x(t) - x_n(t)| + |x_n(t) - x_n(s)| + |x_n(s) - x(s)| < \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon. \end{aligned}$$

Thus, x is continuous at s. Since s was arbitrary, x is continuous.

Example 1.4.12 ($C_1[a,b]$ is not complete ([Coh03, 2.4(6)])). In order to show that $C_1[a,b]$ is not complete it suffices to find an example of a Cauchy sequence in this space that does not converge. We will consider the case where $a < 0$ and $b > 1$, but our example can easily be modified for other values of a and b. Define the sequence of functions $\{x_n\}$ in $C_1[a,b]$ by

$$x_n(t) = \begin{cases} 0, & a \le t \le 0, \\ nt, & 0 < t < \frac{1}{n}, \\ 1, & \frac{1}{n} \le t \le b. \end{cases}$$

Then

$$d(x_n, x_m) = \left(\int_a^b |x_n(t) - x_m(t)|^p dt \right)^{1/p} \le \left(\frac{1}{\min\{m,n\}} \right)^{1/p}.$$

Therefore, for any $\epsilon > 0$,

$$n, m > \frac{1}{\epsilon^p} \implies d(x_n, x_m) < \epsilon.$$

Thus $\{x_n\}$ is a Cauchy sequence. We show that this sequence does not converge. Let g be the function defined on $[a, b]$ by

$$g(t) = \begin{cases} 0, & a \le t \le 0, \\ 1, & 0 < t \le b, \end{cases}$$

and let f be any continuous function on $[a, b]$.

For all $n \in \mathbb{N}_+$, we have

$$|g(t) - f(t)| = |(g(t) - x_n(t)) - (x_n(t) - f(t))| \le |g(t) - x_n(t)| + |x_n(t) - f(t)|.$$

Thus

$$\int_a^b |g(t) - f(t)| dt \le \int_a^b |g(t) - x_n(t)| dt + \int_a^b |x_n(t) - f(t)| dt.$$

The integral on the left is $\int_a^0 |f(t)| dt + \int_0^b |1 - f(t)| dt$. Since f is continuous, at least one of these terms must be positive (since we cannot have both $f(0) = 0$ and $f(0) = 1$). The first integral on the right approaches zero as $n \to \infty$ (by an inequality similar to the one we used to show the sequence $\{x_n\}$ is Cauchy). Thus, we cannot have

$$d(x_n, f) = \int_a^b |x_n(t) - f(t)| dt \to 0,$$

for any continuous function f. Therefore $\{x_n\}$ does not converge in $C_1[a, b]$.

Remark 1.4.13. The previous example shows that a set can be complete when equipped with one metric and not complete with others.

Remark 1.4.14. In fact, Example 1.4.12 can be generalized, to show that $C_p[a, b]$, $p > 0$, is not complete. The proof uses Hölder's inequality.

Recall (from previous analysis courses) that any subsequence of a convergent sequence is itself convergent. However, a sequence which does *not* converge can still have convergent subsequences. For example

$$1, \frac{1}{2}, 2, \frac{1}{3}, 3, \frac{1}{4}, \ldots$$

does not converge but has the convergent subsequence $\frac{1}{2}, \frac{1}{3}, \ldots$. The next result shows us that this cannot happen for Cauchy sequences.

Proposition 1.4.15. *In a metric space, any Cauchy sequence having a convergent subsequence is itself convergent, with the same limit.*

Remark 1.4.16. Of course, the theorem is trivial if the metric space is complete, but it holds in *any* metric space.

Proof. Suppose $\{x_n\}$ is a Cauchy sequence in a metric space (X, d). Let $\{x_{n_k}\}$ be a convergent subsequence of $\{x_n\}$ with limit x. Then for any $\epsilon > 0$, there exists $N_1 > 0$ such that

$$k > N_1 \implies d(x_{n_k}, x) < \frac{\epsilon}{2}.$$

Since $\{x_n\}$ is a Cauchy sequence, there also exists $N_2 > 0$ such that

$$m, n > N_2 \implies d(x_n, x_m) < \frac{\epsilon}{2}.$$

Let $N = \max\{N_1, N_2\}$ and choose $k > N$. Then $n_k \geq k > N$ and so

$$n > N \implies d(x_n, x) \leq d(x_n, x_{n_k}) + d(x_{n_k}, x) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus $\{x_n\}$ converges to x. □

Exercises.

1.4.1 ([Coh03, Ex. 2.9(1)]). Use the inequality of Exercise 1.1.4 to prove that if $\{x_n\}$ and $\{y_n\}$ are convergent sequences in a metric space and $\lim x_n = x$, $\lim y_n = y$, then $d(x_n, y_n) \to d(x, y)$, where d is the metric for the space.

1.4.2 ([Coh03, Ex. 2.9(2)]). Let $\{x_n\}$ and $\{y_n\}$ be two Cauchy sequences in a complete metric space, with metric d. Prove that they have the same limit if and only if $d(x_n, y_n) \to 0$.

1.4.3. Show that ℓ^1 is complete (see Example 1.1.22).

1.4.4. Show that ℓ^∞ is complete (see Example 1.1.23).

1.4.5. Is a nonempty set with the discrete metric complete?

1.4.6 ([Coh03, Ex. 2.9(3)]). Show that the sequence $\{x_n\}$ defined in Example 1.4.12 is not a Cauchy sequence in $C[a, b]$, where $a < 0$ and $1 < b$.

1.4.7 ([Coh03, Ex. 2.9(12)]). Why does the counterexample in Example 1.4.12 fail when $a = 0$?

1.5 Subspaces of metric spaces

Recall the definition of a metric subspace and the induced metric from Definition 1.1.6. Note that if Y is a metric subspace of X, then it is possible to have a sequence in (Y, d) which converges to a point in X that is not in Y. For instance, see Example 1.4.4 (there $X = \mathbb{R}$ and $Y = \mathbb{Q}$).

Definition 1.5.1 (Closed subspace)**.** A subspace S of a metric space X is said to be *(sequentially) closed* if it contains the limits of all the sequences in S that converge in X.

Remark 1.5.2. We will sometimes use the more precise term ‘sequentially closed’, instead of just ‘closed’, later on when we discuss topology and consider closed sets, where ‘closed’ has a different meaning.

Examples 1.5.3. (a) A closed interval $[a, b]$, $a < b$ is a closed subspace of $\mathbb{R}$ (with the usual metric).

(b) $\{(x, 0) \mid x \in \mathbb{R}\}$ is a closed subspace of $\mathbb{R}^2$ (with the usual metric).

(c) For a fixed positive number c, $\{z \mid z \in \mathbb{C},\ |z| \leq c\}$ is a closed subspace of $\mathbb{C}$.

(d) For a fixed positive number c, $\{z \mid z \in \mathbb{C},\ |z| < c\}$ is not a closed subspace of $\mathbb{C}$.

(e) In a metric space with the discrete metric, all subspaces are closed.

(f) The argument in Example 1.4.9 shows that $\mathbb{R}^n$ and $\mathbb{C}^n$ are closed subspaces of ℓ^2 (in fact, of ℓ^p for $1 \leq p \leq \infty$).

(g) Every metric space is a closed subspace of itself. Note that it need not be complete. In the definition of closed subspace, we only require that the limit of any sequence *that converges* in the larger metric space be in the subspace. For instance for $a < b$, (a, b) is a closed subspace of itself, but not a closed subspace of $\mathbb{R}$.

Proposition 1.5.4. *A subspace of a complete metric space is complete if and only if it is closed.*

Proof. Suppose S is a closed subspace of a complete metric space X. Let $\{x_n\}$ be a Cauchy sequence in S. Then it is also a Cauchy sequence in X. Since X is complete, $\{x_n\}$ converges to some point $x \in X$. Then, since S is closed, $x \in S$. Thus every Cauchy sequence in S converges in S, and so S is complete.

Now assume that S is a complete subspace of a complete metric space X. Let $\{x_n\}$ be a sequence in S that, as a sequence in X, converges to some point $x \in X$. Then $\{x_n\}$ is a Cauchy sequence in X (by Theorem 1.4.5) and hence also in S. Since S is complete, we have $x \in S$. Thus S is closed. □

Definition 1.5.5 (Diameter)**.** Let (X, d) be a metric space and let S be a nonempty subset of X. The number

$$\delta(S) = \sup\{d(x, y) \mid x, y \in S\}$$

is called the *diameter* of the set S.

Definition 1.5.6 (Bounded subset of a metric space)**.** A subset of a metric space is said to be *bounded* if it is empty or if it has a finite diameter.

Example 1.5.7. Any Cauchy sequence in a metric space is bounded.

Proof. Let $\{x_n\}$ be a Cauchy sequence in a metric space (X, d). Fix $\epsilon > 0$. Then we can find $N > 0$ such that

$$n, m > N \implies d(x_n, x_m) < \epsilon.$$

In particular,

$$n > N \implies d(x_n, x_{N+1}) < \epsilon.$$

Let

$$K = \max\{d(x_n, x_{N+1}) \mid n = 1, 2, \ldots, N\}.$$

So

$$d(x_n, x_{N+1}) < K + \epsilon \quad \text{for all } n \in \mathbb{N}_+.$$

Then, for all $n, m \in \mathbb{N}_+$, we have, by the triangle inequality,

$$d(x_n, x_m) \leq d(x_n, x_{N+1}) + d(x_m, x_{N+1}) < 2(K + \epsilon).$$

Thus we have an upper bound on the set $\{d(x_n, x_m) \mid n, m \in \mathbb{N}_+\}$. Hence the supremum of this set is finite and so $\{x_n\}$ is bounded. □

Exercises.

1.5.1. Suppose that S is a nonempty subset of a metric space (X, d). Show that S is bounded if and only if there exists a point $x \in X$ and a number $K > 0$ such that $d(x, y) < K$ for all $y \in S$. Loosely speaking, this says that a subset is bounded if and only if it is contained in some ball in X of finite radius.

1.5.2. Show that any convergent sequence in a metric space is bounded.

1.5.3 ([Coh03, Ex. 2.9(10)]). If $\{z_n\}$ is a complex-valued sequence, and $z_n \to z$, prove that $|z_n| \to |z|$. Hence show that the subset $\{w \mid w \in \mathbb{C},\ |w| \leq c\}$ of $\mathbb{C}$ is closed for any positive real number c.

1.5.4 ([Coh03, Ex. 2.9(11)]). Let Y be the set of all complex-valued sequences $(y_1, y_2, \dots)$ for which $|y_k| \leq 1/k$, $k \in \mathbb{N}_+$. Define d by

$$d(x, y) = \sqrt{\sum_{k=1}^{\infty} |x_k - y_k|^2}, \quad x, y \in Y.$$

Prove that (Y, d) is a subspace of ℓ^2, and that it is closed.

1.6 Mappings between metric spaces

In this section, we discuss mappings between metric spaces.

Definition 1.6.1. A *mapping* from the metric space (X, d) to the metric space (Y, d') is simply a map of sets $A \colon X \to Y$.

We will often omit parentheses when dealing with mappings of metric spaces. Hence we write Ax instead of $A(x)$. If $A \colon X \to Y$ and $B \colon Y \to Z$, we denote their composition by BA, which is a map from X to Z. Recall that composition of maps is associative, so if $A \colon W \to X$, $B \colon X \to Y$, and $C \colon Y \to Z$, then $C(BA) = (CB)A$. Thus we can write CBA without any risk of ambiguity.

If $A\colon X \to X$, then we can compose A with itself and for n a positive integer, we define

$$A^n = AA\cdots A \qquad (n \text{ factors}).$$

By convention, A^0 is the identity mapping $I\colon X \to X$ defined by $Ix = x$ for all $x \in X$.

Example 1.6.2. Define the mapping $A\colon C[a,b] \to \mathbb{R}$ by

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

Note that the mapping A is well defined because any continuous function on the interval $[a,b]$ is integrable over that interval.

Example 1.6.3. Suppose we view elements of $\mathbb{C}^n$ as column vectors (i.e. $n \times 1$ matrices with complex entries). If B is an $m \times n$ complex matrix, then B defines a mapping $\mathbb{C}^n \to \mathbb{C}^m$ by sending $x \in \mathbb{C}^n$ to $Bx \in \mathbb{C}^m$. In such examples, we will often denote the matrix and the corresponding map by the same letter.

Since in a metric space we have a notion of distance, it should make sense to speak of the continuity (or lack of continuity) of a mapping between metric spaces.

Definition 1.6.4 (Sequentially continuous mapping)**.** Suppose X and Y are metric spaces. We say a mapping $A\colon X \to Y$ is *(sequentially) continuous at* $x \in X$ if for every convergent sequence $\{x_n\}$ in X with limit x, $\{Ax_n\}$ is a convergent sequence in Y with limit Ax. The mapping A is said to be *(sequentially) continuous on* X, or simply *(sequentially) continuous*, if it is (sequentially) continuous at every point of X.

Remark 1.6.5. Later, in Section 4.6, we will see another definition of continuity, which is why we sometimes insert the word 'sequentially' above. However, we will see that for metric spaces, the two definitions are the same and so there is no risk of confusion.

Example 1.6.6. Consider the mapping A of Example 1.6.2. Suppose $\{x_n\}$ is a convergent sequence in $C[a,b]$ with limit x. Choose $\epsilon > 0$. Then, by the definition of the uniform metric on $C[a,b]$, there exists an $N > 0$ such that

$$n > N \implies \max_{a \le t \le b} |x_n(t) - x(t)| < \frac{\epsilon}{b-a}.$$

Thus

$$n > N \implies |x_n(t) - x(t)| < \frac{\epsilon}{b-a} \ \forall\ t \in [a,b].$$

Then

$$\begin{aligned} d(Ax_n, Ax) &= |Ax_n - Ax| \\ &= \left| \int_a^b x_n(t)\,dt - \int_a^b x(t)\,dt \right| \\ &= \left| \int_a^b \left(x_n(t) - x(t)\right)\,dt \right| \end{aligned}$$

$$\begin{aligned}
&\leq \int_a^b |x_n(t) - x(t)| \, dt \\
&\leq \int_a^b \frac{\epsilon}{b-a} \, dt \\
&= \frac{\epsilon}{b-a}(b-a) \\
&= \epsilon.
\end{aligned}$$

Therefore $Ax_n \to Ax$. Hence A is continuous.

Exercises.

1.6.1. Show that the mappings described in Example 1.6.3 are continuous.

Chapter 2

The Fixed Point Theorem

In this chapter we prove the Fixed Point Theorem for metric spaces and discuss some applications of this important result.

2.1 The Fixed Point Theorem

In this section, we will only consider mappings from a metric space into itself.

Definition 2.1.1 (Fixed points and contractions)**.** Suppose A is a mapping from a metric space (X, d) into itself.

(a) A point $x \in X$ such that $Ax = x$ is called a *fixed point* of the mapping A.

(b) The mapping A is called a *contraction mapping* (or a *contraction*) if there exists a real number α, with $0 < \alpha < 1$, such that

$$d(Ax, Ay) \leq \alpha d(x, y) \quad \text{for all } x, y \in X.$$

Such an α is called a *contraction constant* for A.

Proposition 2.1.2. *Contraction mappings are continuous.*

Proof. Suppose A is a contraction mapping on a metric space (X, d) and that $\{x_n\}$ is a convergent sequence in X with limit x. Let α be a contraction constant for A. Then

$$0 \leq d(Ax_n, Ax) \leq \alpha d(x_n, x) < d(x_n, x) \to 0,$$

and so $d(Ax_n, Ax) \to 0$ (in other words, $Ax_n \to Ax$). □

Remark 2.1.3. Note that the converse of the Proposition 2.1.2 is not true. For instance, the identity mapping is not a contraction mapping but it is certainly continuous. Note also that the proof did not use the fact that $0 < \alpha < 1$, and so could be somewhat generalized.

Theorem 2.1.4 (Fixed Point Theorem)**.** *Every contraction mapping on a complete metric space has exactly one fixed point.*

Proof. Suppose (X, d) is a complete metric space and A is a contraction mapping on X with contraction constant α. Take any point $x_0 \in X$ and define a sequence $\{x_n\}$ in X by

$$x_n = Ax_{n-1} = A^n x_0, \quad n \in \mathbb{N}_+.$$

We show that $\{x_n\}$ is a Cauchy sequence. For $k > 1$, we have

$$\begin{aligned} d(x_k, x_{k-1}) &= d(A^k x_0, A^{k-1}x_0) \\ &= d(A(A^{k-1}x_0), A(A^{k-2}x_0)) \\ &\le \alpha d(A^{k-1}x_0, A^{k-2}x_0) \\ &\le \alpha^2 d(A^{k-2}x_0, A^{k-3}x_0) \\ &\vdots \\ &\le \alpha^{k-1} d(Ax_0, x_0). \end{aligned}$$

Now, to show that $\{x_n\}$ is a Cauchy sequence, we need to consider $d(x_n, x_m)$ for large values of m and n. Since $d(x_n, x_m) = d(x_m, x_n)$, we may assume that $1 \le m < n$. Then

$$\begin{aligned} d(x_n, x_m) &= d(A^n x_0, A^m x_0) \\ &\le d(A^n x_0, A^{n-1}x_0) + d(A^{n-1}x_0, A^{n-2}x_0) + \cdots + d(A^{m+1}x_0, A^m x_0) \quad (\Delta \text{ inequality}) \\ &\le \alpha^{n-1} d(Ax_0, x)) + \alpha^{n-2} d(Ax_0, x_0) + \cdots + \alpha^m d(Ax_0, x_0) \\ &= \alpha^m (1 + \alpha + \alpha^2 + \cdots + \alpha^{n-m-1}) d(x_1, x_0) \\ &= \frac{\alpha^m (1 - \alpha^{n-m})}{1 - \alpha} d(x_1, x_0) \quad \text{(sum of a geometric series)} \\ &\le \frac{\alpha^m}{1 - \alpha} d(x_1, x_0) \quad (\text{since } 0 < \alpha < 1). \end{aligned}$$

Since $0 < \alpha < 1$, we have $\alpha^m \to 0$ as $m \to \infty$. Thus, for any $\epsilon > 0$, we have $d(x_n, x_m) < \epsilon$ for m and n sufficiently large. Therefore $\{x_n\}$ is a Cauchy sequence. Since X is complete, this sequence converges to some point $x \in X$.

We will show that x is a fixed point of A. For any positive integer n, we have

$$0 \le d(Ax, x) \le d(Ax, x_n) + d(x_n, x) = d(Ax, Ax_{n-1}) + d(x_n, x) \le \alpha d(x, x_{n-1}) + d(x_n, x).$$

Since $d(x_n, x) \to 0$ and $d(x, x_{n-1}) \to 0$, we have $d(Ax, x) = 0$ and so $Ax = x$. Thus x is a fixed point of A. Now suppose y is another fixed point of A. Then

$$d(x, y) = d(Ax, Ay) \le \alpha d(x, y).$$

Since $\alpha < 1$, this can only be true if $d(x, y) = 0$. So $x = y$. Therefore, x is the unique fixed point of A. □

We can in fact prove a slight generalization of the fixed point theorem that is useful in some applications.

Theorem 2.1.5. *Suppose A is a mapping on a complete metric space such that A^n is a contraction for some $n \in \mathbb{N}_+$. Then A has a unique fixed point.*

Proof. By the Fixed Point Theorem, A^n has a unique fixed point x. So $A^n x = x$. Since

$$A^n(Ax) = A^{n+1}x = A(A^n x) = Ax,$$

we have that Ax is also a fixed point of A^n. But by the Fixed Point Theorem, A^n has a *unique* fixed point. Thus $Ax = x$. In other words, x is also a fixed point of A. Now suppose that y is another fixed point of A. Then

$$A^n y = A^{n-1}(Ay) = A^{n-1}y = \cdots = Ay = y,$$

and so y is also a fixed point of A^n. Thus $x = y$ by the uniqueness in the Fixed Point Theorem. □

Exercises.

2.1.1 ([Coh03, Ex. 3.5(3)]). Consider $\mathbb{R}^n$ with the metric

$$d_1(x, y) = \sum_{k=1}^{n} |x_k - y_k|, \quad x, y \in \mathbb{R}^n.$$

(See Exercise 1.1.7.) Show that this metric is complete. (*Hint*: One approach is to use Exercise 1.4.3.) Call this metric space X. Define a mapping $M\colon X \to X$ by $y = Mx$, where $x \in X$, and

$$y_j = \sum_{k=1}^{n} c_{j,k}x_k + b_j, \quad j = 1, 2, \ldots, n,$$

with all $c_{j,k}, b_j \in \mathbb{R}$. Prove that M is a contraction mapping on X if

$$0 < \max_{k=1}^{n} \sum_{j=1}^{n} |c_{j,k}| < 1.$$

2.1.2 ([Coh03, Ex. 3.5(8)]). Let A be a mapping from a complete metric space (X, d) into itself. Prove that if the contraction condition is weakened to

$$d(Ax, Ay) < d(x, y) \ \forall\, x, y \in X, \ x \neq y,$$

then the existence of a fixed point is no longer assured.

2.1.3 ([Coh03, Ex. 3.5(10)]). Let c be the set of all convergent complex-valued sequences and define a mapping $d\colon c \times c \to \mathbb{R}_{\geq 0}$ by

$$d(x, y) = \sup_{k=1}^{\infty} |x_k - y_k|, \quad x = (x_1, x_2, \ldots), \ y = (y_1, y_2, \ldots) \in c.$$

(a) Prove that (c, d) is a metric space and that it is complete.

(b) Define a mapping A on c by

$$A(x_1, x_2, x_3, \dots,) = \left(\frac{1}{2}x_2, \frac{1}{3}x_3, \frac{1}{4}x_4, \dots, \right).$$

Prove that A is a contraction and hence that A has a unique fixed point (immediately obtained by inspection). Suppose this point is to be obtained by iteration and let $x^{(0)}, x^{(1)}, x^{(2)}, \dots$ denote the successive iterates. Taking $x^{(0)} = (1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots)$, show that $x^{(n)}$ has kth component

$$\frac{k!}{(n+k-1)!(n+k)^2}, \quad n = 0, 1, 2, \dots, \ k = 1, 2, \dots .$$

(c) Define a mapping B on c by

$$B(x_1, x_2, x_3, \dots) = \left(1 + \frac{1}{2}x_2 + \frac{1}{3}x_3, 1 + \frac{1}{2}x_3 + \frac{1}{3}x_4, 1 + \frac{1}{2}x_4 + \frac{1}{3}x_5, \dots\right).$$

Prove that B is a contraction. Find the fixed point of B (by any means).

2.2 Applications

We now turn to some applications of the Fixed Point Theorem. Suppose $f\colon [a, b] \to [a, b]$ is a function for which there exists a constant $0 < K < 1$ such that

$$|f(x_1) - f(x_2)| \le K|x_1 - x_2| \quad \forall\ x_1, x_2 \in [a, b].$$

We say f satisfies the *Lipschitz condition* with *Lipschitz constant* K (see Definition 4.6.9). Note that the interval $[a, b]$ with the usual metric is complete (it is a closed subset of $\mathbb{R}$, which is complete), and that f is a contraction mapping. So the Fixed Point Theorem tells us that the equation $f(x) = x$ has a unique solution.

In particular, if $f\colon [a, b] \to [a, b]$ is differentiable and there exists a $0 < K < 1$ such that

$$|f'(x)| \le K \quad \forall\ x \in [a, b],$$

then $f(x) = x$ has a unique solution. This is because the Mean Value Theorem of differential calculus tell us that, for any $x_1, x_2 \in [a, b]$, $x_1 < x_2$, there exists a point c, $x_1 < c < x_2$, such that

$$|f(x_1) - f(x_2)| = |f'(c)(x_1 - x_2)| = |f'(c)||x_1 - x_2| \le K|x_1 - x_2|,$$

and so f satisfies the Lipschitz condition with constant $K < 1$.

Example 2.2.1. Let us show that the equation

$$x^4 - 2x^3 - 12x + 1 = 0$$

has a unique solution in the interval $[0, 1]$. Note that the equation is equivalent to

$$x^4 - 2x^3 + 1 = 12x \iff \frac{1}{12}x^4 - \frac{1}{6}x^3 + \frac{1}{12} = x.$$

Define a function f by

$$f(x) = \frac{1}{12}x^4 - \frac{1}{6}x^3 + \frac{1}{12}.$$

Then solving the original equation is equivalent to solving the equation $f(x) = x$. We first check that the image of f is contained in $[0, 1]$. To see this, note that

$$f'(x) = \frac{1}{3}x^3 - \frac{1}{2}x^2 = 0 \implies x^3(2x - 3) = 0 \implies x = 0 \text{ or } \frac{3}{2}.$$

So f has no critical points in the interval $[0, 1]$. Since $f(0) = \frac{1}{12}$ and $f(1) = 0$, we have $f([0, 1]) \subseteq [0, 1]$.

Now, for $x \in [0, 1]$,

$$|f'(x)| = \left|\frac{1}{3}x^3 - \frac{1}{2}x^2\right| \leq \frac{1}{3} + \frac{1}{2} < 1.$$

Therefore, our conditions are met and so f has a unique fixed point. To find it, we can take $x_0 = 0$ and begin applying f. We get

$$x_1 = f(0) \approx 0.0833, \quad x_2 = f(x_1) \approx 0.0832, \quad x_3 = f(x_2) \approx 0.08324, \ldots.$$

To 5 decimal places, the root is 0.08324.

As another example, let us consider systems of linear equations. Suppose we want to solve the linear system

$$Ax = b$$

for some $n \times n$ matrix A and $b \in \mathbb{R}^n$ (viewed as a column vector). In order to apply the Fixed Point Theorem, we need the equation in the form $f(x) = x$. So we write

$$Ax = b \iff x - Ax + b = x \iff (I - A)x + b = x \iff Cx + b = x, \quad \text{where } C = I - A.$$

So we define a mapping $M : \mathbb{R}^n \to \mathbb{R}^n$ by $Mx = Cx + b$ and we are looking to solve the equation $Mx = x$. Since $\mathbb{R}^n$ is a complete metric space, there will be a unique solution if M is a contraction mapping. Let us find some conditions that will imply that M is a contraction mapping. For $y, z \in \mathbb{R}^n$, we have

$$\begin{aligned} d(My, Mz) &= \sqrt{\sum_{j=1}^{n} \left(\left(\sum_{k=1}^{n} c_{jk} y_k + b_j \right) - \left(\sum_{k=1}^{n} c_{jk} z_k + b_j \right) \right)^2} \\ &= \sqrt{\sum_{j=1}^{n} \left(\sum_{k=1}^{n} c_{jk}(y_k - z_k) \right)^2} \\ &\leq \sqrt{\sum_{j=1}^{n} \left(\left(\sum_{k=1}^{n} c_{jk}^2 \right) \left(\sum_{k=1}^{n} (y_k - z_k)^2 \right) \right)}, \end{aligned}$$

by the Cauchy–Schwarz inequality. Here $C = (c_{jk})$. Therefore

$$d(My, Mz) \leq \sqrt{\sum_{j=1}^{n} \sum_{k=1}^{n} c_{jk}^2} \cdot d(y, z),$$

and so M is a contraction if

$$0 < \sum_{j=1}^{n} \sum_{k=1}^{n} c_{jk}^2 < 1.$$

(Note that this is a sufficient condition, but not necessary). If this condition is satisfied, then one can solve the equation in question by iteration.

Note that here we used the standard metric on $\mathbb{R}^n$. However, we are free to use any metric we like as long as it makes $\mathbb{R}^n$ a complete metric space. For instance, we could use the metric d_∞ (see Example 1.1.14). Recall that we know $(\mathbb{R}^n, d_\infty)$ is complete. We have

$$\begin{aligned}
d_\infty(My, Mz) &= \max_{1\le j\le n} \left| \left(\sum_{k=1}^{n} c_{jk} y_k + b_j \right) - \left(\sum_{k=1}^{n} c_{jk} z_k + b_j \right) \right| \\
&= \max_{1\le j\le n} \left| \sum_{k=1}^{n} c_{jk}(y_k - z_k) \right| \\
&\le \max_{1\le j\le n} \sum_{k=1}^{n} |c_{jk}||y_k - z_k| \\
&\le \max_{1\le j\le n} \sum_{k=1}^{n} |c_{jk}| \cdot \max_{1\le k\le n} |y_k - z_k| \\
&= \left(\max_{1\le j\le n} \sum_{k=1}^{n} |c_{jk}| \right) \cdot d_\infty(y, z).
\end{aligned}$$

So M is a contraction in this metric if

$$0 < \max_{1\le j\le n} \sum_{k=1}^{n} |c_{jk}| < 1.$$

So we see that we might choose to work with different metrics depending on our particular application. For instance, it is possible for a matrix to satisfy one of the above conditions and not the other.

Remark 2.2.2. For further applications of the Fixed Point Theorem to differential and integral equations, see [Coh03, §3.3].

Exercises.

2.2.1 ([Coh03, Ex. 3.5(2a)]). Show that

$$x^4 + 8x^3 + 32x - 32 = 0$$

has a unique solution in the interval [0,1].

2.2.2 ([Coh03, Ex. 3.5(4)]). Use the Fixed Point Theorem to show that the following systems of equations have unique solutions. (No adjustment of the coefficients, by dividing an equation through by some number, for example, should be necessary.)

(a)

$$\begin{array}{rcrcrcr} \frac{3}{4}x & - & \frac{1}{2}y & + & \frac{1}{8}z & = & 2 \\ \frac{1}{6}x & + & \frac{1}{3}y & & & = & -1 \\ -\frac{2}{5}x & + & \frac{1}{4}y & + & \frac{5}{4}z & = & 1 \end{array}$$

(b)

$$\begin{array}{rcrcrcl} \frac{1}{4}x & + & \frac{1}{2}y & - & \frac{1}{8}z & = & x-3 \\ -\frac{1}{6}x & + & \frac{2}{3}y & & & = & y+1 \\ \frac{2}{5}x & - & \frac{1}{4}y & - & \frac{1}{4}z & = & z+5 \end{array}$$

Chapter 3

Compactness

In previous courses, you have encountered the notion of a compact subset of $\mathbb{R}^n$. We now discuss compactness in the more general setting of metric spaces.

3.1 Sequentially compact sets

Recall, from Proposition 1.4.15 that, in a metric space, any Cauchy sequence having a convergent subsequence is itself convergent, with the same limit. We will use this fact several times in this section.

Definition 3.1.1 (Sequentially compact)**.** A subset of a metric space is called *(sequentially) compact* if every sequence in the subset has a convergent subsequence (that converges to a point in the subset).

Remarks 3.1.2.

(a) The use of the word 'sequentially' is to distinguish this notion of compactness from another (involving open covers). We will examine the relationship between the two properties a bit later in the course (see Theorem 4.5.9).

(b) We can apply Definition 3.1.1 to the entire metric space. So a metric space is compact if every sequence in it has a convergent subsequence.

(c) We consider the empty set to be a compact subset of any metric space.

Proposition 3.1.3. *A compact metric space is complete.*

Proof. Let $\{x_n\}$ be a Cauchy sequence in a compact metric space. By the definition of compactness, $\{x_n\}$ has a convergent subsequence. Then, by Proposition 1.4.15, $\{x_n\}$ is itself convergent. □

Remarks 3.1.4.

(a) We will often use the contrapositive form of this theorem: If a metric space is not complete, then it is not compact.

(b) It is possible for a metric space to be complete but not compact. For instance, we know that $\mathbb{R}$ is complete but $1, 2, 3, \dots$ is a sequence with no convergent subsequence. Hence $\mathbb{R}$ is not compact.

Theorem 3.1.5. *Every compact subset of a metric space is closed.*

Proof. Suppose S is a compact subset of a metric space X. Let $\{x_n\}$ be a sequence in S that converges to a point $x \in X$. This implies, in particular, that $\{x_n\}$ is a Cauchy sequence. Since S is compact, $\{x_n\}$ has a subsequence which converges to a point $y \in S$. Then, by Proposition 1.4.15, we know that $\{x_n\}$ is convergent in S and so $x = y \in S$, by uniqueness of limits. Hence S is closed. □

Remark 3.1.6. Note that the converse of the above theorem is not true. For instance, $\mathbb{R}$ is a closed subset of itself but is not compact.

Theorem 3.1.7. *Every compact subset of a metric space is bounded.*

Proof. We will prove the result by contradiction. Since the theorem is clear for the empty set, we assume S is a nonempty compact subset of a metric space (X, d) and that S is not bounded. We will construct a sequence in S with no convergent subsequence.

Choose any element $x_1 \in S$. We cannot have $d(x, x_1) < 1$ for all $x \in S$ since this would imply (by the triangle inequality) that $\delta(S) \leq 2$ (recall that $\delta(S)$ is the diameter of the subset S). Therefore, we can find a point $x_2 \in S$ with $d(x_2, x_1) \geq 1$. Let

$$\lambda_1 = 1, \quad \lambda_2 = \lambda_1 + d(x_2, x_1) = 1 + d(x_2, x_1).$$

We cannot have $d(x, x_1) < \lambda_2$ for all $x \in S$ since this would imply that $\delta(S) \leq 2\lambda_2$. Therefore, we can find a point $x_3 \in S$ with $d(x_3, x_1) \geq \lambda_2$. Let

$$\lambda_3 = \lambda_1 + d(x_3, x_1) = 1 + d(x_3, x_1).$$

Continuing in this manner, we construct a sequence $\{x_n\}$ in S and an increasing sequence of numbers $\{\lambda_n\}$ such that

$$d(x_n, x_1) = \lambda_n - 1 \geq \lambda_{n-1}, \quad n = 2, 3, \dots .$$

Then, for any $n > m \geq 2$, we have

$$\lambda_m \leq \lambda_{n-1} \leq d(x_n, x_1) \leq d(x_n, x_m) + d(x_m, x_1) = d(x_n, x_m) + \lambda_m - 1,$$

and so $d(x_n, x_m) \geq 1$. Therefore, the sequence $\{x_n\}$ cannot have a convergent subsequence (since no subsequence can be a Cauchy sequence). This contradicts the assumption that S is a compact subset of X. Therefore, S is bounded. □

Remark 3.1.8. The converse to the above theorem is false. For instance, $(0, 1)$ is a bounded subset of $\mathbb{R}$ but it is not compact by Theorem 3.1.5, since it is not closed.

Theorem 3.1.9. *A subset of $\mathbb{R}^n$ is compact if and only if it is both closed and bounded.*

Proof. We already know from Theorems 3.1.5 and 3.1.7 that if a subset of $\mathbb{R}^n$ is compact, it is closed and bounded. The converse (in fact, both implications) was done in previous courses. The proof can also be found, for example, in [Coh03, Th. 4.1.6]. For the case $n = 1$ (and, in some references, in the general case), this is known as the Bolzano–Weierstrass theorem. □

Example 3.1.10. Theorem 3.1.9 is not true for arbitrary metric spaces. A useful example to keep in mind is the subset

$$S = \{e_i \mid i = 1, 2, \dots\}$$

of the metric space ℓ^2. Here e_i is the usual sequence with a 1 in the ith position and a 0 in every other position. Since

$$d(e_m, e_n) = \sqrt{2} \quad \text{for } m \neq n,$$

we see that $\delta(S) = \sqrt{2}$ and so S is bounded. Since the sequence $e_1, e_2, e_3, \dots$ has no convergent subsequence, the set S is not compact. Note that S is closed since any sequence in S which converges in ℓ^2 must be eventually constant (i.e. the range of the sequence is finite) and thus convergent in S. So S is an example of a subset which is closed and bounded, but not compact. So

$$\text{compact} \implies (\text{closed and bounded})$$

but the converse is not true for an arbitrary metric space. However, in *certain* metric spaces, such as $\mathbb{R}^n$, the converse is true.

In a *compact* metric spaces, we have the following partial converse to Theorem 3.1.9.

Lemma 3.1.11. *Every closed subset of a compact metric space is compact.*

Proof. The proof of this lemma is left as an exercise (Exercise 3.1.2). □

Exercises.

3.1.1 ([Coh03, Ex. 4.5(2)]). (a) Prove that any finite subset of a metric space is compact.

(b) Let x be the limit of a convergent sequence $\{x_n\}$ in a metric space. Prove that the set $\{x, x_1, x_2, x_3, \dots\}$ is compact.

3.1.2. Prove Lemma 3.1.11.

3.1.3. Suppose X is a nonempty set with the discrete metric. Is X compact? Which subsets of X are compact?

3.1.4. Show that Baire space $\mathbb{Z}^{\mathbb{N}_+}$ (see Example 1.1.27) is not compact.

3.2 Sequences of functions and the Arzelà–Ascoli Theorem

Definition 3.2.1. Suppose F is a family of real-valued (or complex-valued) functions on a fixed metric space (X, d) and D is a subset of X.

(a) We say the family F is *uniformly bounded* on D if there exists an $M > 0$ such that

$$f \in F,\ x \in D \implies |f(x)| \leq M.$$

(b) We say F is *equicontinuous* at a point $x \in D$ if, for any $\epsilon > 0$, there exists a $\delta > 0$ such that

$$f \in F,\ x' \in D,\ d(x, x') < \delta \implies |f(x) - f(x')| < \epsilon.$$

(c) We say F is *uniformly equicontinuous* on D if, for any $\epsilon > 0$, there exists a $\delta > 0$ such that

$$f \in F,\ x, x' \in D,\ d(x, x') < \delta \implies |f(x) - f(x')| < \epsilon.$$

Remarks 3.2.2. (a) Note that the definition of 'uniformly bounded' does not depend on any metric on F. However, if $F \subseteq C[a, b]$ has the uniform metric, then 'uniformly bounded' is the same as 'bounded'. However, for subsets of $C_1[a, b]$ the two notions are different (Exercise 3.2.1).

(b) Similarly, the definition of (uniform) equicontinuity is independent of any metric on F. However, if the domain of the functions in an equicontinuous family F is $[a, b]$ and F is given the uniform metric, then F is clearly a subspace of $C[a, b]$.

(c) The definition of uniform equicontinuity could be extended to the case where the functions in F take values in any fixed metric space, instead of $\mathbb{R}$ or $\mathbb{C}$. Then the condition becomes that for any $\epsilon > 0$, there exists a $\delta > 0$ such that

$$f \in F,\ x, x' \in D,\ d(x, x') < \delta \implies d(f(x), f(x')) < \epsilon.$$

Similarly, one can define equicontinuity in this more general situation.

(d) The definition of uniformly bounded could be extended to the case where the functions in F takes values in any 'normed space' (see Chapter 5).

(e) The reference [Coh03] uses the term 'equicontinuous' to mean what we call 'uniformly equicontinuous'.

Example 3.2.3. The set of functions $\{x, x^2, x^3, \dots\}$ on $[0, 1]$ is not uniformly equicontinuous, even though each individual function is continuous. To see this, note that for $0 < \epsilon < 1$, in order to ensure that

$$|x^n - 1^n| = 1 - x^n < \epsilon \quad \text{for } 0 \leq x \leq 1,$$

we need $x > (1 - \epsilon)^{1/n}$, or $1 - x < 1 - (1 - \epsilon)^{1/n}$. Since $(1 - \epsilon)^{1/n} \to 1$ as $n \to \infty$, there is no $\delta > 0$ such that

$$|1 - x| < \delta \implies |x^n - 1^n| < \epsilon \quad \forall\, n \in \mathbb{N}_+.$$

For $\mathbb{R}^n$ (or $\mathbb{C}^n$), we have nice criteria for when a set is compact. Namely, the compact sets are precisely the closed and bounded sets. It turns out that we can also develop criteria for when subsets of $C[a, b]$ are compact.

Theorem 3.2.4 (Arzelà–Ascoli Theorem)**.** *A subset F of the metric space $C[a,b]$ is compact if and only if F is closed, uniformly bounded, and uniformly equicontinuous.*

Proof. First suppose that F is closed, uniformly bounded, and uniformly equicontinuous. We will show that F is compact. Suppose $\{f_n\}$ is a sequence in F. We prove that $\{f_n\}$ has a convergent subsequence. The proof involves six steps.

Step 1. We know that $\mathbb{Q} \cap [a,b]$ is countable. Let $\{x_1, x_2, \dots\}$ be an enumeration of the elements of this set.

Step 2. Since F is uniformly bounded, we can find an $M > 0$ such that

$$|f_n(x)| \le M \quad \forall\, x \in [a,b],\ n \in \mathbb{N}_+.$$

In particular, we have

$$|f_n(x_1)| \le M \quad \forall n \in \mathbb{N}_+.$$

Thus the sequence $\{f_n(x_1)\}_{n=1}^\infty$ is contained in the interval $[-M, M]$. Since $[-M, M]$ is compact, this sequence has a convergent subsequence $\{f_{n_k}(x_1)\}$. So we have a subsequence $\{f_{n_k}\}$ of the sequence $\{f_n\}$ which converges pointwise at x_1.

Let us write this subsequence as $\{f_n^{(1)}\}_{n=1}^\infty$, instead of $\{f_{n_k}\}_{k=1}^\infty$ in an effort to make the notation somewhat easier to follow. We now apply the same reasoning as above to the sequence $\{f^{(1)}\}$. We have

$$|f_n^{(1)}(x_2)| \le M \quad \forall\, n \in \mathbb{N}_+,$$

and so the sequence $\{f_n^{(1)}(x_2)\}$ of real numbers has a convergent subsequence $\{f_{n_k}^{(1)}(x_2)\}$. So we have chosen a subsequence $\{f_{n_k}^{(1)}\}$ of $\{f_n^{(1)}\}$ which converges pointwise at x_1 and x_2. We write this new sequence as $\{f_n^{(2)}\}$.

Continuing in this manner, we construct sequences $\{f_n^{(m)}\}_{n=1}^\infty$ for $m \in \mathbb{N}_+$ such that for each m, the sequence $\{f_n^{(m)}\}_{n=1}^\infty$ is a subsequence of $\{f_n\}$ converging pointwise at $x_1, x_2, \dots, x_m$. Furthermore, each sequence is a subsequence of the one before it.

Step 3. We have constructed sequences

$$\begin{gathered} f_1^{(1)},\ f_2^{(1)},\ f_3^{(1)},\ \dots, \\ f_1^{(2)},\ f_2^{(2)},\ f_3^{(2)},\ \dots, \\ f_1^{(3)},\ f_2^{(3)},\ f_3^{(3)},\ \dots, \\ \vdots \end{gathered}$$

Let $f^n = f_n^{(n)}$. Thus $\{f^n\}$ is the 'diagonal' sequence. For each $m \ge n$, the sequence $f^m, f^{m+1}, \dots$ is a subsequence of $\{f_n^{(m)}\}_{n=1}^\infty$. Thus $f^m, f^{m+1}, \dots$ converges pointwise at $x_1, x_2, \dots, x_m$. Since the addition of points at the beginning of a sequence does not alter its convergence, the sequence $\{f^n\}_{n=1}^\infty$ also converges at $x_1, x_2, \dots, x_m$. Since this is true for all m, we see that $\{f^n\}$ converges at all points $x_1, x_2, \dots$ (i.e. all points of $\mathbb{Q} \cap [a,b]$).

Step 4. We now show that $\{f^n\}$ is a convergent sequence. Choose $\epsilon > 0$. Since the sequence $\{f^n\}$ is a subset of F, which is uniformly equicontinuous, there exists $\delta > 0$ such that

$$n \in \mathbb{N}_+,\ x', x'' \in [a,b],\ |x' - x''| < \delta \implies |f^n(x') - f^n(x'')| < \frac{1}{3}\epsilon. \tag{3.1}$$

Choose K rational points in $[a, b]$ such that every point of $[a, b]$ lies within δ of one of those rational point (we can do this since $\delta > 0$ and $b - a < \infty$). Renumbering if necessary, we can assume these rational points are $x_1, \dots, x_K$. Thus

$$\forall x \in [a, b]\ \exists\ i \in \{1, 2, \dots, K\} \text{ such that } |x - x_i| < \delta.$$

Step 5. Since $\{f^n(x_i)\}_{n=1}^{\infty}$ is a convergent sequence for each $i = 1, 2, \dots, K$ (note that the number of values of i is finite), we can choose a $N > 0$ such that

$$m, n > N \implies |f^n(x_i) - f^m(x_i)| < \frac{1}{3}\epsilon \quad \forall\ i = 1, 2, \dots, K.$$

Step 6. Let $x \in [a, b]$. As in Step 4, we can choose a point x_i, $i = 1, 2, \dots, K$, such that $|x - x_i| < \delta$. Then, by (3.1),

$$|f^n(x) - f^n(x_i)| < \frac{1}{3}\epsilon \quad \forall\ n \in \mathbb{N}_+.$$

Therefore

$$\begin{aligned} |f^n(x) - f^m(x)| \le |f^n(x) - f^n(x_i)| + |f^n(x_i) - f^m(x_i)| + |f^m(x_i) - f^m(x)| \\ < \frac{1}{3}\epsilon + \frac{1}{3}\epsilon + \frac{1}{3}\epsilon = \epsilon, \end{aligned}$$

for all $m, n > N$. Since our choice of $x \in [a, b]$ was arbitrary, it follows that

$$m, n > N \implies d(f^n, f^m) = \max_{a \le x \le b} |f^n(x) - f^m(x)| < \epsilon.$$

Therefore $\{f^n\}$ is a Cauchy sequence in F. Since F is a closed subset of the complete metric space $C[a, b]$, it is complete by Proposition 1.5.4. Thus $\{f^n\}$ converges. Hence F is compact.

It remains to show the other implication asserted in the statement of the theorem. Suppose F is a compact subset of $C[a, b]$. We already know from Theorem 3.1.5 that F is closed. By Theorem 3.1.7, F is bounded in the uniform metric and hence uniformly bounded. It remains to show that F is uniformly equicontinuous. For this, we will use the following two facts.

(a) A metric space is sequentially compact if and only if it is compact in the sense you learned in previous courses (any open cover has a finite subcover). We will return to this issue later in the course (see Theorem 4.5.9).

(b) If a real-valued function f is continuous on a closed interval $[a, b]$ then it is *uniformly continuous* on that interval. This means that for all $\epsilon > 0$, there exists a $\delta > 0$ such that

$$x_1, x_2 \in [a, b],\ |x_1 - x_2| < \delta \implies |f(x_1) - f(x_2)| < \epsilon.$$

(See Definition 5.5.1.) You may have seen this in previous courses. In any case, this is left as an exercise for you if you have not seen it (Exercise 3.2.4). See also Theorem 5.5.2, which is more general.

Assuming these facts, we continue with the proof. By the above, all elements of F are uniformly continuous on $[a, b]$. Therefore, for all $f \in F$, we can choose a δ_f such that

$$x_1, x_2 \in [a, b], \; |x_1 - x_2| < \delta_f \implies |f(x_1) - f(x_2)| < \epsilon/3.$$

Let

$$B_f = \{g \in F \mid d(f, g) < \epsilon/3\}.$$

Here $d(f, g)$ is the distance in the uniform metric. Then for all $g \in B_f$, we have

$$\begin{gathered} x_1, x_2 \in [a, b], \; |x_1 - x_2| < \delta_f \\ \implies |g(x_1) - g(x_2)| \le |g(x_1) - f(x_1)| + |f(x_1) - f(x_2)| + |f(x_2) - g(x_2)| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{gathered}$$

The $\{B_f\}$ form an open cover of F (since $f \in B_f$ for all $f \in F$). Strictly speaking, we have not yet discussed open sets in metric spaces, but we will see that 'balls' of the form B_f are open sets in metric spaces. Since F is compact, there exists a finite subcover $\{B_{f_1}, \dots, B_{f_n}\}$. Let $\delta = \min\{\delta_{f_1}, \dots, \delta_{f_n}\}$. Then

$$x_1, x_2 \in [a, b], \; |x_1 - x_2| < \delta, \; g \in F \implies |g(x_1) - g(x_2)| < \epsilon.$$

Hence F is uniformly equicontinuous. □

Remark 3.2.5. Some references (including [Coh03]), refer to Theorem 3.2.4 as the Ascoli Theorem and only include the 'if' part.

In light of Theorem 3.2.4, is it useful to have easy-to-check criteria for a family of functions to be uniformly equicontinuous.

Lemma 3.2.6. *Suppose F is a family of differentiable functions defined on the same interval $[a, b]$ and there exists a number K for which*

$$|f'(x)| \le K \quad \forall \; f \in F, \; x \in [a, b].$$

Then F is uniformly equicontinuous.

Proof. Suppose $\epsilon > 0$ and let $\delta = \epsilon/K$. Then

$$|x_1 - x_2| < \delta \implies |f(x_1) - f(x_2)| \le K|x_1 - x_2| < K\delta = \epsilon,$$

for all $f \in F$. (Here we are using the mean value theorem in calculus.) Hence F is uniformly equicontinuous. □

Exercises.

3.2.1 ([Coh03, Ex. 4.5(7)]). Let F be a subset of $C[a,b]$. Prove that F is a uniformly bounded family if and only if it is bounded. Show however that if F is considered as a subset of $C_1[a,b]$, then F may be bounded but not uniformly bounded. (See Remark 3.2.2(a).)

3.2.2 ([Coh03, Ex. 4.5(8)]). Let K and α be fixed positive numbers and let F be a subset of $C[a,b]$ such that, for all $f \in F$ and any points $x', x'' \in [a,b]$,

$$|f(x') - f(x'')| \leq K|x' - x''|^{\alpha}.$$

Show that F is equicontinuous.

3.2.3 ([Coh03, Ex. 4.5(9)]). Let F be a bounded subset of $C[a,b]$. Prove that the set of all functions g, where

$$g(x) = \int_a^x f(t)\,dt, \quad f \in F,\ a \leq x \leq b,$$

is uniformly bounded and equicontinuous.

3.2.4. Prove that if a real-valued function f is continuous on a closed interval $[a,b]$ then it is uniformly continuous on that interval.

3.3 Compactness and mappings

In this section we look at the interplay between compact sets and mappings between metric spaces.

Theorem 3.3.1. *Suppose X, Y are metric spaces and $A\colon X \to Y$ is a continuous mapping. If S is a compact subset of X, then its image $A(S)$ is a compact subset of Y.*

Proof. If S is empty, the result is trivial. So assume S is not empty. Suppose $\{y_n\}$ is a sequence in $A(S)$. We wish to show that $\{y_n\}$ has a convergent subsequence. For each n, we can find $x_n \in S$ such that $Ax_n = y_n$. Since S is compact, the sequence $\{x_n\}$ has a convergent subsequence $\{x_{n_k}\}$ with some limit $x \in S$. Then $Ax \in A(S)$. Since A is continuous, we have $y_{n_k} = Ax_{n_k} \to Ax$. Thus $\{y_{n_k}\}$ is a convergent subsequence. Therefore $A(S)$ is compact. □

Corollary 3.3.2. *If X is a metric space, $f\colon X \to \mathbb{R}$ is a continuous mapping, and S is any nonempty compact set in X, then f attains its maximum and minimum on S. In other words, there exist points $x_{\max}$ and $x_{\min}$ in S such that*

$$f(x_{\min}) \leq f(x) \leq f(x_{\max}) \quad \forall\ x \in S.$$

Proof. By Theorem 3.3.1, $f(S)$ is compact. Thus, by Theorem 3.1.9, $f(S)$ is closed and bounded. But we know from previous courses that closed and bounded subsets of $\mathbb{R}$ contain a maximum and minimum. □

Theorem 3.3.3. *Suppose S is a nonempty compact subset of a metric space (X,d) and $x \in X$. Then there exists a point $p \in S$ such that $d(p,x)$ is a minimum.*

Proof. Define $f\colon X \to \mathbb{R}$ by $f(y) = d(y, x)$ for $y \in X$. So we wish to minimize f. First, we claim that f is continuous. If $\{y_n\}$ is a sequence in X and $y_n \to y$, then

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

The last inequality is true since

$$d(y_n, x) \le d(y_n, y) + d(y, x) \implies d(y_n, x) - d(y, x) \le d(y_n, y),$$

and

$$d(y, x) \le d(y, y_n) + d(y_n, x) \implies d(y_n, x) - d(y, x) \ge -d(y_n, y).$$

Therefore $f(y_n) \to f(y)$ since $d(y_n, y) \to 0$. This shows that f is continuous on X. Therefore, by Corollary 3.3.2, f attains its minimum at some point $p \in S$. □

Remarks 3.3.4. (a) Theorem 3.3.3 does not imply that the point p is the *unique* point in S minimizing the distance to x.

(b) Theorem 3.3.3 gives no method for actually finding such a point p.

Example 3.3.5. Fix $n \in \mathbb{N}_+$ and $M > 0$. Let

$$F = \{a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0 \mid a_i \in [-M, M],\ 0 \le x \le 1\}.$$

Then F is a uniformly bounded and uniformly equicontinuous family of functions on the interval $[0, 1]$. To see that F is uniformly bounded, note that for $f \in F$, we have

$$|f(x)| \le |a_n| + |a_{n-1}| + \cdots + |a_0| \le nM.$$

To see that F is uniformly equicontinuous, note that for all $f \in F$, we have

$$|f'(x)| = |n a_n x^{n-1} + \cdots + a_1| \le n|a_n| + (n-1)|a_{n-1}| + \cdots + |a_1| \le \frac{n(n+1)}{2} M,$$

and so we can apply Lemma 3.2.6. One can also show that F is a closed subset of $C[0, 1]$. Thus, by the Arzelà–Ascoli Theorem (Theorem 3.2.4), F is compact in $C[0, 1]$. Therefore, by Theorem 3.3.3, for any continuous function f on the interval $[0, 1]$, there exist values of $a_0, a_1, \ldots, a_n$ in $[-M, M]$ such that

$$\max_{0 \le x \le 1} |f(x) - (a_n x^n + \cdots + a_0)|$$

is a minimum. This is called a *minimax approximation* of f. As noted above, it may not be unique.

Typically the advantage of minimax approximations lies in the fact that we are approximating an arbitrary continuous functions by 'nice' ones. This has many applications including image compression (where one approximates the data in some image by simpler data taking less room to store).

You were asked to show in Exercise 2.1.2 that if the contraction condition is weakened to $d(Ax, Ay) < d(x, y)$ for all x, y, the existence of a fixed point is no longer assured. However, in a *compact* metric space, this condition is enough.

Proposition 3.3.6. *Suppose (X, d) is a compact metric space and $A\colon X \to X$ is a mapping such that*

$$d(Ax, Ay) < d(x, y) \quad \forall\ x, y \in X, x \neq y.$$

Then A has a unique fixed point in X.

Proof. Suppose $\{x_n\}$ is a convergent sequence in X with $x_n \to x$. Then

$$0 \leq d(Ax_n, Ax) < d(x_n, x) \to 0,$$

and so A is continuous. (This is essentially the same argument we used to show that a contraction is continuous in the proof of Proposition 2.1.2.) Define

$$B\colon X \to \mathbb{R}, \quad B(x) = d(x, Ax), \quad x \in X.$$

By Exercise 1.4.1, we have

$$Bx_n = d(x_n, Ax_n) \to d(x, Ax) = Bx,$$

and so B is continuous. Since X is compact, it follows from Corollary 3.3.2 that B attains a minimum. So $By = \min_{x\in X} Bx$ for some $y \in X$. Thus

$$d(y, Ay) \leq d(x, Ax) \quad \forall\ x \in X.$$

Suppose $d(y, Ay) > 0$. Then

$$B(Ay) = d(Ay, A(Ay)) < d(y, Ay) = By.$$

But this contradicts our choice of y. Therefore $d(y, Ay) = 0$ and so $Ay = y$. Thus y is a fixed point of A.

Now suppose $Az = z$ for some $z \in X$, $z \neq y$. Then

$$d(y, z) = d(Ay, Az) < d(y, z),$$

which is a contradiction. Thus y is the unique fixed point of A. □

Chapter 4

Topological spaces

In this chapter, we move from metric spaces to an even more general setting: topological spaces. Topological spaces have no notion of distance, but do have open and closed subsets. We will see that all metric spaces can naturally be given the structure of a topological space, but the reverse is not true.

4.1 Definitions

We have already seen some topological terms—namely 'compact' and 'closed'.

Definition 4.1.1 (Topology)**.** A *topology* on a nonempty set X is a collection $\mathcal{T}$ of subsets of X with the following properties:

(T1) $\varnothing, X \in \mathcal{T}$,

(T2) $\mathcal{T}$ is closed under arbitrary unions: $\bigcup_{T\in\mathcal{S}} T \in \mathcal{T}$ for any $\mathcal{S} \subseteq \mathcal{T}$,

(T3) $\mathcal{T}$ is closed under finite intersections: $T_1 \cap T_2 \in \mathcal{T}$ for all $T_1, T_2 \in \mathcal{T}$.

The pair $(X, \mathcal{T})$ is called a *topological space* and the sets $T \in \mathcal{T}$ are called the *open sets* in $(X, \mathcal{T})$. A subset S of X is said to be *closed* in $(X, \mathcal{T})$ if its complement $S^{\complement} = X \setminus S$ is an open set in $(X, \mathcal{T})$. We sometimes refer to open and closed subsets of X (instead of $(X, \mathcal{T})$) when the topology is clear from the context.

Remarks 4.1.2. (a) It follows from (T3) that the intersection of any finite number of open sets in X is again an open set in X.

(b) The case where X has only one element will sometimes be an exception to various theorems we will prove. Thus we will often assume X has at least two elements.

Example 4.1.3 (Discrete and indiscrete topologies). For any (nonempty) set X, we have the *discrete topology*

$$\mathcal{T}_{\max} = \{A \mid A \subseteq X\}$$

consisting of *all* subsets of X and the *indiscrete topology* (or *trivial topology*)

$$\mathcal{T}_{\min} = \{\varnothing, X\}.$$

Definition 4.1.4 (Weaker/stronger topology)**.** If $\mathcal{T}_1$ and $\mathcal{T}_2$ are two topologies on a set X such that $\mathcal{T}_1 \subseteq \mathcal{T}_2$, then we say $\mathcal{T}_1$ is *weaker* (or *coarser*) than $\mathcal{T}_2$, and $\mathcal{T}_2$ is *stronger* (or *finer*) than $\mathcal{T}_1$.

Therefore, on any given set, the discrete topology is the strongest topology of all and the indiscrete topology is the weakest topology.

Example 4.1.5. Let $X = \{1, 2, 3, 4\}$ and

$$\begin{aligned}
\mathcal{T}_1 &= \{\varnothing, \{1\}, \{1,2\}, \{2,3\}, \{1,2,3\}, X\}, \\
\mathcal{T}_2 &= \{\varnothing, \{2\}, \{3\}, \{2,3\}, X\}, \\
\mathcal{T}_3 &= \{\varnothing, \{1\}, \{2\}, \{1,2\}, \{1,2,3\}, X\}, \\
\mathcal{T}_4 &= \{\varnothing, \{1\}, \{2\}, \{1,2\}, \{1,2,3\}, \{1,2,3,4\}, X\}.
\end{aligned}$$

Then

(a) $\mathcal{T}_1$ is *not* a topology since $\{1,2\}, \{2,3\} \in \mathcal{T}_1$ but $\{1,2\} \cap \{2,3\} = \{2\} \notin \mathcal{T}_1$.

(b) $\mathcal{T}_2$, $\mathcal{T}_3$, and $\mathcal{T}_4$ are all topologies since they all contain $\varnothing$ and X and are closed under arbitrary unions and finite intersections.

(c) $\mathcal{T}_3$ is weaker than $\mathcal{T}_4$ since $\mathcal{T}_3 \subseteq \mathcal{T}_4$.

(d) $\mathcal{T}_2$ is neither weaker nor stronger than $\mathcal{T}_3$ or $\mathcal{T}_4$.

Proposition 4.1.6. *In any topological space $(X, \mathcal{T})$, the following properties hold.*

(a) The empty set, $\varnothing$, and the whole space, X, are closed.

(b) The union of finitely many closed sets is closed.

(c) The intersection of any family of closed sets is closed

Proof. (a) This follows from the fact that $\varnothing^{\complement} = X$, $X^{\complement} = \varnothing$, and both X and $\varnothing$ are open.

(b) Suppose $F_1, \ldots, F_n$ are closed sets. According to De Morgan's laws,

$$\left(\bigcup_{i=1}^{n} F_i\right)^{\complement} = \bigcap_{i=1}^{n} F_i^{\complement}.$$

Since the F_i are closed, the $F_i^{\complement}$ are open and hence so is their intersection $\bigcap_{i=1}^{n} F_i^{\complement}$.

(c) This proof is left as an exercise (Exercise 4.1.1).

□

Definition 4.1.7 (Interior and closure)**.** Suppose $(X, \mathcal{T})$ is a topological space and S is a subset of X.

(a) The *interior* of S is the union of all open sets contained in S and is denoted by $\operatorname{int} S$ or S°. Thus

$$S^\circ = \cup\{U \in \mathcal{T} \mid U \subseteq S\}.$$

If we wish to emphasize the ambient metric space, we write $\operatorname{int}_X S$.

(b) The *closure* of S is the intersection of all closed sets containing S and is denoted by $\operatorname{cl} S$ or $\bar{S}$. Thus

$$\bar{S} = \cap\{Z \subseteq X \mid Z^\complement \in \mathcal{T},\ S \subseteq Z\}.$$

Therefore, S° is the largest open set contained in S and $\bar{S}$ is the smallest closed set containing S.

Now that we have defined topological spaces in general, we would like to restrict our attention to metric spaces. We would like to be able to define a natural topology on *any* metric space. We can do this as follows.

Definition 4.1.8 (Metric topology)**.** Suppose (X, d) is a metric space.

(a) For $x_0 \in X$ and $r > 0$, the set

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

is called the *open ball* in X with centre x_0 and radius r.

(b) A subset U of X is *open* if $U = \varnothing$ or if for all $x \in U$ there exists an $\epsilon > 0$ such that $B(x, \epsilon) \subseteq U$.

(c) The *metric topology* on X is the collection of open sets as defined above. This topology is denoted $\mathcal{T}_d$. In other words, we define

$$\mathcal{T}_d = \{\varnothing\} \cup \{U \subseteq X \mid \forall\, x \in U,\ \exists\, \epsilon > 0 \text{ tel que } B(x, \epsilon) \subseteq U\}.$$

Of course, we should justify that $\mathcal{T}_d$ is actually a topology. It is trivial to see that $\varnothing, X \in \mathcal{T}_d$. Now suppose that $U_i \in \mathcal{T}_d$ for all i in some indexing set I and let $U = \bigcup_{i \in I} U_i$. We want to show that $U \in \mathcal{T}_d$ (that is, U is open in the sense defined above). Let $x \in U$. Then $x \in U_i$ for some $i \in I$. But since $U_i \in \mathcal{T}_d$, we can find an $r > 0$ such that $B(x, r) \subseteq U_i \subseteq U$. Thus $U \in \mathcal{T}_d$.

Now suppose that $U_1, U_2 \in \mathcal{T}_d$. We want to show that $U_1 \cap U_2 \in \mathcal{T}_d$. Let $x \in U_1 \cap U_2$. Then we can find ϵ_1 and ϵ_2 such that $B(x, \epsilon_1) \subseteq U_1$ and $B(x, \epsilon_2) \subseteq U_2$. Let $\epsilon = \min\{\epsilon_1, \epsilon_2\}$. Then

$$B(x, \epsilon) \subseteq B(x, \epsilon_1) \subseteq U_1 \quad \text{and} \quad B(x, \epsilon) \subseteq B(x, \epsilon_2) \subseteq U_2.$$

Thus $B(x, \epsilon) \subseteq U_1 \cap U_2$. Hence $U_1 \cap U_2 \in \mathcal{T}_d$.

Whenever we refer to a metric space as a topological space, we will assume we are working with the metric topology.

Recall that a *singleton* is a set with one element.

Example 4.1.9. Suppose X is a metric space with the discrete metric. Then for all $x \in X$, $\{x\} = B(x, 1)$ is an open set in X. Since every subset $A \subseteq X$ is a union of singletons (since $A = \bigcup_{a \in A} \{a\}$), we see that all subsets of X are open. Thus, the metric topology is the discrete topology (which is good since we use the word 'discrete' for both).

Example 4.1.10. Consider the metric space $C[a, b]$ and $f \in C[a, b]$. Then, for $r > 0$,

$$B(f, r) = \{g \in C[a, b] \mid f(x) - r \leq g(x) \leq f(x) + r \ \forall \ x \in [a, b]\}.$$

Thus $B(f, r)$ consists of all the continuous functions on $[a, b]$ that differ from f at any point by at most r.

Examples 4.1.11. (a) $\operatorname{int}_{\mathbb{Q}} \mathbb{Q} = \mathbb{Q}$, since every metric space is an open subset of itself.

(b) $\operatorname{int}_{\mathbb{R}} \mathbb{Q} = \varnothing$, since every nonempty open interval in $\mathbb{R}$ (note that the open balls of $\mathbb{R}$ are the open intervals) contains an irrational number and so cannot be a subset of $\mathbb{Q}$.

(c) $\operatorname{int}_{\mathbb{R}}[a, b] = (a, b)$.

(d) $\operatorname{int}_{\mathbb{R}}\{0\} = \varnothing$.

(e) $\operatorname{int}_{\{0\}}\{0\} = \{0\}$.

(f) In a metric space equipped with the discrete topology, every set is open and hence the interior of any subset A is A itself.

Exercises.

4.1.1. Prove Proposition 4.1.6(c).

4.1.2 ([Coh03, Ex. 5.7(1)]). Let $X = \{a, b, c, d\}$,

$$\mathcal{T}_1 = \{\varnothing, \{a\}, \{b\}, \{a, b\}, \{a, b, c\}, \{a, b, d\}, X\}, \quad \text{and}$$
$$\mathcal{T}_2 = \{\varnothing, \{a\}, \{a, b\}, \{a, c\}, \{a, b, c\}, X\}.$$

(a) Verify that $\mathcal{T}_1$ and $\mathcal{T}_2$ are topologies on X.

(b) In $(X, \mathcal{T}_1)$, find the closed sets, and find the interiors and closures of $\{a\}$, $\{c\}$, and $\{a, c\}$.

(c) Do the same for $(X, \mathcal{T}_2)$.

4.1.3. (a) Suppose X is a nonempty set with the discrete topology $\mathcal{T}_{\max}$. Find $\operatorname{int} S$ and $\bar{S}$ for any subset S of X.

(b) Suppose X is a nonempty set with the indiscrete topology $\mathcal{T}_{\min}$. Find $\operatorname{int} S$ and $\bar{S}$ for any subset S of X.

4.1.4. Prove that every open ball in a metric space is an open set.

4.2 Closed sets

We have now given two definitions of closed sets in the context of metric spaces. We defined a sequentially closed set in Definition 1.5.1 and we have now defined a closed set coming from the metric topology. This would be very confusing were it not for the fact that these two definitions actually coincide (as we will now see).

Theorem 4.2.1. *Suppose (X, d) is a metric space with the metric topology $\mathcal{T}_d$. A subset of X is closed in $(X, \mathcal{T}_d)$ if and only if it is sequentially closed in (X, d).*

Proof. Suppose that S is a closed set in $(X, \mathcal{T}_d)$. Since the case $S = \varnothing$ is trivial, we assume $S \neq \varnothing$. We wish to show that S is sequentially closed. Let $\{x_n\}$ be a sequence in S which converges to a point $x \in X$. If $x \notin S$, then $x \in S^\complement$. Since S is closed, $S^\complement$ is open. Thus there exists an open ball $B(x, \epsilon) \subseteq S^\complement$. But this implies that $d(x_n, x) \geq \epsilon$ for all n, which contradicts the fact that $x_n \to x$. Thus $x \in S$ and so S is sequentially closed.

Now assume that S is sequentially closed. Again we may assume $S \neq \varnothing$. We wish to show that S is a closed set, i.e. that $S^\complement$ is open. If this is not the case, then there exists a point $x \in S^\complement$ such that no open ball centred at x is contained in $S^\complement$. In other words, every open ball centred at x contains a point of S. Thus, for each $n \in \mathbb{N}_+$, we can choose a point $x_n \in S$ with $x_n \in B(x, \frac{1}{n})$. Then $\{x_n\}$ is a sequence in S with $d(x_n, x) = \frac{1}{n} \to 0$. Hence $\lim x_n = x$. Since S is sequentially closed, $x \in S$. But this contradicts the fact that $x \in S^\complement$. Hence $S^\complement$ is open and so S is closed. □

Remarks 4.2.2. (a) Theorem 4.2.1 only holds for the metric topology. There are topological spaces in which 'closed' is *not* the same as 'sequentially closed' (they have sequentially closed sets that are not closed).

(b) We have seen that we can turn any metric space into a topological space using the metric topology. It is natural to ask if one can reverse this procedure in some way. More precisely, given a topological space $(X, \mathcal{T})$, can one always define a metric d on X such that $\mathcal{T}_d = \mathcal{T}$? The answer is no. If it is possible to define such a metric, we say the space is *metrizable*. A simple example of a topological space that is not metrizable is $X = \{1, 2\}$ with the indiscrete topology (see Exercise 4.2.1). There are theorems which state that all topological spaces satisfying certain additional conditions are metrizable (these are known as *metrization theorems*).

Remarks 4.2.3. (a) A subset A of a metric space (X, d) can be both closed *and* open, so one property does not exclude the other. For instance, X and $\varnothing$ are both open and closed. In a discrete metric space, every subset is open and hence every subset is also closed. (Sometimes sets that are both open and closed are called *clopen*.)

(b) It is also possible for a subset A of a metric space (X, d) to be neither open nor closed. For example, consider any interval of the form $[a, b)$, $a < b$, in $\mathbb{R}$, with the usual metric. It is not open because it does not contain any open ball centred at a and it is not closed because $A^\complement = (\infty, a) \cup [b, \infty)$ is not open (it does not contain any open ball centred at b). Therefore, if you have proven that a subset of X is not open, that does *not* mean that it is closed (and vice versa).

(c) Of course, there exist sets which are open but not closed (for instance, (a,b), $a<b$, in $\mathbb{R}$) and sets are that are closed but not open (for instance, $[a,b]$, $a<b$, in $\mathbb{R}$).

Proposition 4.2.4. *Every singleton in a metric space is closed in it.*

Proof. Let (X,d) be a metric space and $x \in X$. We must show that $\{x\}^{\complement}$ is open. Choose any $y \in \{x\}^{\complement}$. Then $y \neq x$ and so $d(x,y) > 0$. Let $\epsilon = d(x,y)$. Then the open ball $B(y,\epsilon)$ does not contain x and so $B(y,\epsilon) \subseteq \{x\}^{\complement}$. Since y was arbitrary, $\{x\}^{\complement}$ is open. □

Definition 4.2.5 (Neighbourhood, cluster point, derived set)**.** Suppose X is a topological space.

(a) A *neighbourhood* of a point $x \in X$ is any open set in X containing x.

(b) A point $x \in X$ is called a *cluster point* (or *limit point*) for $S \subseteq X$ if every neighbourhood of x contains a point of S other than x.

(c) The set of all cluster points of a subset S of X is called the *derived set* of S and is denoted by S'.

(d) A point $x \in X$ is called a *closure point* (or *adherent point*) for $S \subseteq X$ if every neighbourhood of x contains a point of S.

Remarks 4.2.6. (a) Note that in the definitions of cluster points and closure points, x need not be a point of S.

(b) Note the difference between a cluster point and a closure point. In the definition of a cluster point, we require the neighbourhood of x to contain a point of S *other than* x itself, whereas in the definition of a closure point, we allow x itself. Thus, a closure point is either a cluster point for S or a point of S.

(c) A point $x \in X$ is a closure point for $S \subseteq X$ if and only if $x \in \bar{S}$. In other words, the closure $\bar{S}$ of a subset S is precisely the set of its closure points (hence the name 'closure point').

(d) A point $x \in X$ which is a closure point for $S \subseteq X$ but not a cluster point, is called an *isolated point* of S.

(e) Sometimes the term 'neighbourhood of x' is used to mean any subset containing some ball centred at x. We then use the term *open neighbourhood* if we require that this set be open.

Proposition 4.2.7. *Suppose S is a subset of a metric space X. Then $x \in X$ is a closure point for S if and only if some sequence in S converges to x.*

Proof. This proof is left as an exercise (Exercise 4.2.2). □

Example 4.2.8. Let $S = \{0\} \cup (1,2]$, considered as a subset of $\mathbb{R}$.

(a) 0, 1 and 2 are closure points of S (of course, they are not the only closure points).

(b) 1 and 2 are cluster points of S, but 0 is not.

(c) $\bar{S} = \{0\} \cup [1, 2]$.

Example 4.2.9. Suppose X is a nonempty set with the discrete topology $\mathcal{T}_{\max}$. Then $\{x\}$ is a neighbourhood of x that does not contain any other point of X. Hence

(a) no point can be a cluster point for any subset of X,

(b) the only closure points of a subset S of X are the points of S itself, and

(c) the closure of any subset S of X is S itself.

Example 4.2.10. Suppose X is a nonempty set with the indiscrete topology $\mathcal{T}_{\min}$. Then for any $x \in X$, X is the only neighbourhood of x. Thus every point $x \in X$ is a cluster point for every subset of X, except $\{x\}$ and $\varnothing$. Furthermore, for $A \subseteq X$, we have

$$\bar{A} = \begin{cases} \varnothing & \text{if } A = \varnothing, \\ X & \text{if } A \neq \varnothing. \end{cases}$$

Proposition 4.2.11. *A set S in a topological space is closed if and only if it contains its cluster points, that is, $S' \subseteq S$.*

Proof. Let X be a topological space and $S \subseteq X$. Assume first that S is closed. If $S = X$, then $S' \subseteq S$ and we are done. Otherwise $S^{\complement}$ is a nonempty open subset of X. For any $x \in S^{\complement}$, the set $S^{\complement}$ is itself a neighbourhood of x containing no point of S. So $x \notin S'$. Thus, taking the contrapositive, $x \in S' \implies x \in S$. Hence $S' \subseteq S$.

Now assume that $S' \subseteq S$. We want to show that $S^{\complement}$ is open. If $S^{\complement} = \varnothing$, we are done. So assume $S^{\complement} \neq \varnothing$ and choose $x \in S^{\complement}$. This means that $x \notin S'$ (since $S' \subseteq S$). Hence x is not a cluster point for S and so there exists some neighbourhood U_x of x with $U \subseteq S^{\complement}$. Let

$$V = \bigcup_{x \in S^{\complement}} U_x.$$

Then $V \subseteq S^{\complement}$ and V is open since it is a union of open sets. Since every point x of $S^{\complement}$ is contained in $U_x \subseteq V$, we have that $S^{\complement} = V$. Hence $S^{\complement}$ is open and so S is closed. □

Remark 4.2.12. Since a closure point for S is either a cluster point for S or a point of S itself, another way of stating Proposition 4.2.11 is that a set S in a topological space is closed if and only if it contains its closure points (and hence is equal to its set of closure points, since every point of S is a closure point by definition).

Corollary 4.2.13. *A subset S of a metric space is closed if and only if it is equal to its closure: $S = \bar{S}$.*

Proof. This follows immediately from Remark 4.2.12 since any subset S is contained in its set of closure points (since points of S are closure points of S). □

We now summarize some useful properties of closures.

Theorem 4.2.14. *Let (X, d) be a metric space and let $A, B \subseteq X$. Then the following statements are true:*

(a) $\mathrm{cl}\,\varnothing = \varnothing$*,*

(b) $A \subseteq \mathrm{cl}\,A$*,*

(c) if $A \subseteq B$*, then* $\mathrm{cl}\,A \subseteq \mathrm{cl}\,B$ *(we say that closure is* monotonic*)*

(d) $\mathrm{cl}(\mathrm{cl}\,A) = \mathrm{cl}\,A$*.*

Proof. The proofs of these statements are left as an exercise (Exercise 4.2.3). □

Definition 4.2.15 (Closed ball)**.** For a point x_0 of a metric space X and $r > 0$, the *closed ball* with centre x_0 and radius r is the set

$$B^{\mathrm{cl}}(x_0, r) = \{x \in X \mid d(x, x_0) \le r\}.$$

One can prove that $B^{\mathrm{cl}}(x_0, r)$ is a closed subset of X (see Exercise 4.2.4). In addition, it is clear that the closed ball $B^{\mathrm{cl}}(x, r)$ contains the corresponding open ball $B(x, r)$. However, in general, the closure $\mathrm{cl}\,B(x, r)$ is *not* equal to the closed ball $B^{\mathrm{cl}}(x, r)$.

Example 4.2.16. Let X be a metric space with the discrete metric containing at least two points and choose $x \in X$. Then

$$\mathrm{cl}\,B(x, 1) = \mathrm{cl}\{x\} = \{x\},$$

while

$$B^{\mathrm{cl}}(x, 1) = X \ne \{x\},$$

since $|X| \ge 2$.

Definition 4.2.17 (Boundary point)**.** A point $x \in X$ of a metric space is called a *boundary point* of a subset $A \subseteq X$ if for all $\epsilon > 0$, the open ball $B(x, \epsilon)$ contains points from both A and $A^{\complement}$.

Equivalently, x is a boundary point for A if and only if x is a closure point for both A and $A^{\complement}$.

Still equivalently: x is a boundary point for A if and only if x is a closure point for A and not an interior point of A.

Definition 4.2.18 (Boundary)**.** Suppose A is a subset of a metric space X. The *boundary* of A (in X) is the set of all boundary points of A and is denoted $\mathrm{bd}\,A$. When we wish to emphasize the ambient metric space, we write $\mathrm{bd}_X\,A$.

By the above remarks, we have

$$\mathrm{bd}\,A = \mathrm{cl}\,A \setminus \mathrm{int}\,A.$$

Example 4.2.19. We have $\mathrm{bd}_{\mathbb{R}}(0, 1) = \mathrm{bd}_{\mathbb{R}}[0, 1] = \mathrm{bd}_{\mathbb{R}}[0, 1) = \{0, 1\}$.

Example 4.2.20. We have $\operatorname{bd}_{\mathbb{R}} \mathbb{Q} = \mathbb{R}$. This is because every nonempty interval in $\mathbb{R}$ contains both rational points (points of $\mathbb{Q}$) and irrational points (points of $\mathbb{Q}^{\complement}$). However, $\operatorname{bd}_{\mathbb{Q}} \mathbb{Q} = \varnothing$, because the intersection of any neighbourhood of a point in $\mathbb{Q}$ with the complement of $\mathbb{Q}$ in $\mathbb{Q}$ (which is the empty set) is empty. We see here that the boundary points of a subset (in this case $\mathbb{Q}$) depend on the ambient metric space (in this case, $\mathbb{R}$ versus $\mathbb{Q}$).

Example 4.2.21. We have $\operatorname{bd}_{\mathbb{R}} \varnothing = \varnothing$. The empty set has no boundary points.

Theorem 4.2.22. *Let A be a subset of a metric space X. The following statements are equivalent.*

(a) $\operatorname{cl}_X A = X$.

(b) Every point of X is a closure point for A.

(c) For every $x \in X$ and $\epsilon > 0$, the open ball $B(x, \epsilon)$ meets A.

(d) Every nonempty open subset of X contains a point of A.

Proof. (a) $\implies$ (d): Let V be a nonempty open subset of X. Since V is nonempty, it contains a point x. By (a) and Remark 4.2.12, every point of X is a closure point of A. Hence x is a closure point of A and so V, being a neighbourhood of x, contains a point of A.

(d) $\implies$ (c): This is trivial, since every open ball is a nonempty open set.

(c) $\implies$ (b): Let $x \in X$. For all $\epsilon > 0$, the ball $B(x, \epsilon)$ meets A. Thus x is a closure point for A.

(b) $\implies$ (a): This follows from Remark 4.2.6(c). □

Definition 4.2.23 (Everywhere dense set)**.** A subset A of a metric space X is *everywhere dense* (or simply *dense*) in X if it satisfies at least one (therefore all) of the equivalent conditions in Theorem 4.2.22.

Theorem 4.2.24 (Density theorem)**.** *The set $\mathbb{Q}$ is everywhere dense in $\mathbb{R}$ (with the usual metric). Equivalently, every real number is the limit of a sequence of rational numbers; or, every nonempty open interval in $\mathbb{R}$ contains at least one rational number; or, the closure of $\mathbb{Q}$ in $\mathbb{R}$ is $\mathbb{R}$ itself; or, every element of $\mathbb{R}$ is a closure point for $\mathbb{Q}$.*

Proof. It suffices to prove that there is a rational number between any two distinct real numbers. Let $a, b \in \mathbb{R}$, and $a < b$. We will find a rational number between a and b.

To begin with, assume that $a \geq 0$. By Archimedes' Axiom, there is a natural number n such that $n(b - a) > 1$, that is,

$$\frac{1}{n} < b - a.$$

This n will be the denominator of a rational point inside (a, b). To find the numerator, set

$$A = \left\{ k \in \mathbb{N}_+ \;\middle|\; \frac{k}{n} < b \right\}.$$

This set A is nonempty, because it contains 1: indeed, $1/n < b - a \leq b$. (Here we use the assumption that a is non-negative.) By applying Archimedes' Axiom again, we see that A

is finite: indeed, for some N one has $N/n > b$, and every element from A is less than N. Therefore, there exists the maximal element in A:

$$m = \max A.$$

We claim that $m/n \in (a, b)$. One has $m/n < b$ by the choice of A, and so it remains to verify that $m/n > a$. Assume the contrary: $m/n \leq a$. Since $1/n < b - a$, one can see that $(m + 1)/n < b$ and so $m + 1 \in A$, in contradiction with the maximality of m.

Now let us consider the remaining cases. If $a < 0$ and $b > 0$, then we apply the previous conclusion to the interval $(0, b)$. If $b \leq 0$, then we choose a rational point r in the (positive) interval $(-b, -a)$, then the element $-r$ is rational and contained in (a, b). □

Example 4.2.25. A subset A of a metric space X equipped with the discrete metric is everywhere dense in X if and only if $A = X$. To see this, let $A \subsetneq X$ be a proper subset of X. Then we can choose $x \in X \setminus A$. The open ball $B(x, 1)$ is just the singleton $\{x\}$ and so does not meet A. Hence A is not dense in X.

Exercises.

4.2.1. Let $X = \{x_1, \dots, x_n\}$ be a finite set with a metric d. Show that the metric topology is the discrete topology. In particular, this shows that any finite set with a topology other than the discrete topology is not metrizable.

4.2.2. Prove Proposition 4.2.7.

4.2.3. Prove Theorem 4.2.14.

4.2.4. Suppose (X, d) is a metric space. Show that for all $x_0 \in X$ and $r > 0$, the closed ball $B^{\text{cl}}(x_0, r)$ is a closed subset of X.

4.2.5. (a) Show that a point x of a metric space X is an isolated point if and only if $\{x\}$ is an open set. (Don't over think this question—it's almost a tautology.)

(b) Show that Baire space $\mathbb{Z}^{\mathbb{N}_+}$ (see Example 1.1.27) has no isolated points.

4.3 Separable metric spaces

Definition 4.3.1 (Separable)**.** A metric space X is *separable* if it contains a countable everywhere dense subset. In other words, there exists $A \subseteq X$ such that

(a) the elements of A can be arranged in a sequence $A = \{a_n\}_{n=1}^{\infty}$, and

(b) A is everywhere dense in X.

Remark 4.3.2. One can actually define the term 'separable' for any topological space, but we will be concentrating on metric spaces in this course.

Examples 4.3.3. (a) By The Density Theorem (Theorem 4.2.24), the real line is separable since $\mathbb{Q}$ is countable and dense in $\mathbb{R}$.

(b) Similarly, the Euclidean space $\mathbb{R}^n$ is separable since the subset $\mathbb{Q}^n$ of all points whose n coordinates are all rational is both countable and dense in $\mathbb{R}^n$.

(c) A space X equipped with the discrete metric is separable if and only if X is itself countable.

(d) The space $\mathbb{N}_+$ is separable because it is countable. In fact, any countable metric space is separable.

Theorem 4.3.4. *Every metric subspace of a separable metric space is separable.*

Proof. Suppose Y is a nonempty subset of a separable metric space (X, d). Fix a countable dense subset of X:

$$A = \{a_n \mid n = 1, 2, 3, \dots\}.$$

We wish to find a countable dense subset of Y. The difficulty is that $Y \cap A$ need not be dense in Y. In fact, Y might not meet A at all (for instance $X = \mathbb{R}$, $A = \mathbb{Q}$, $Y = \mathbb{Q}^{\complement}$).

Fix an element $y' \in Y$. For every two natural numbers $n, k \geq 1$, consider the ball $B(a_n, 1/k)$. If this ball meets Y, then choose a point in the intersection $B(a_n, 1/k) \cap Y$ and denote it $y_{n,k}$. If this intersection is empty, then let $y_{n,k} = y'$. In other words,

$$y_{n,k} = \begin{cases} \text{any point in } B(a_n, 1/k) \cap Y, & \text{if } B(a_n, 1/k) \cap Y \neq \varnothing, \\ y', & \text{otherwise.} \end{cases}$$

Let

$$B = \{y_{n,k} \mid n, k \in \mathbb{N}_+\}.$$

Clearly, B is a countable subset of Y. It remains to show that B is dense in Y. Let $y \in Y$ and $\epsilon > 0$. Choose a natural number $k \geq 1$ such that

$$\frac{1}{k} < \frac{\epsilon}{2}.$$

Since A is dense in X, there is a point in A contained in $B(y, 1/k)$. Fix $n \in \mathbb{N}_+$ such that $a_n \in B(y, 1/k)$. In other words, choose n such that

$$d(a_n, y) < \frac{1}{k}.$$

The ball $B(a_n, 1/k)$ meets Y (in particular, it contains $y \in Y$). Thus the point $y_{n,k} \in B$ assigned to this particular pair of natural numbers is contained in the intersection $B(a_n, 1/k) \cap Y$. Now, by the triangle inequality,

$$d(y, y_{n,k}) \leq d(y, a_n) + d(a_n, y_{n,k}) < \frac{1}{k} + \frac{1}{k} = \frac{2}{k} < \epsilon.$$

Therefore $y_{n,k}$ is contained in $B(y, \epsilon)$. Since $y \in Y$ and $\epsilon > 0$ were arbitrary, the subset B is indeed dense in Y, as desired. $\square$

Proposition 4.3.5. *The metric space ℓ^∞ is not separable.*

Proof. By Theorem 4.3.4, it is enough to find a subspace of ℓ^∞ that is not separable. Let X be the set of all sequences whose entries are either 0 or 1. Obviously such sequences are bounded and so $X \subseteq \ell^\infty$. We claim that the metric induced on X is the discrete metric. Obviously, if $x, y \in X$ and $x = y$, then $d_\infty(x, y) = 0$. Suppose $x \neq y$. Then $x_i \neq y_i$ for some i. For every n, x_n and y_n are either 0 or 1 and so $|x_n - y_n|$ is either 0 or 1. Since $|x_i - y_1| = 1$, we have

$$d_\infty(x, y) = \sup_{n=1}^{\infty} |x_n - y_n| = 1.$$

We will now show that X is uncountable, which will mean, by Example 4.3.3(c), that the space X is not separable, finishing the argument. It will follow that X is uncountable if we can show that for any sequence $\{x_n\}$ in X, there is an element $x' \in X$, which is not in this sequence.

Write the elements of the sequence $\{x_n\}$ as an array:

$$\begin{aligned} x_1 &= (x_{1,1}, x_{1,2}, x_{1,3}, \dots, x_{1,n}, \dots) \\ x_2 &= (x_{2,1}, x_{2,2}, x_{2,3}, \dots, x_{2,n}, \dots) \\ x_3 &= (x_{3,1}, x_{3,2}, x_{3,3}, \dots, x_{3,n}, \dots) \\ &\vdots \\ x_n &= (x_{n,1}, x_{n,2}, x_{n,3}, \dots, x_{n,n}, \dots) \\ &\vdots \end{aligned}$$

Note that each entry $x_{n,k}$ is either 0 or 1. Now form a new sequence x' by moving along the main diagonal and switching the entry $x_{n,n}$ each time. In other words, define the sequence $x' = (x'_1, x'_2, \dots)$ by

$$x'_n = 1 - x_{n,n}.$$

Then x' and x_n differ in at least the nth coordinate and so

$$x' \neq x_n \text{ for all } n.$$

Thus X is uncountable. Hence X is a non-separable metric subspace of ℓ^∞ and so ℓ^∞ is not separable. □

The above argument used to establish that X is not countable is known as the *Cantor diagonal process*.

Exercises.

4.3.1. Show that $\mathbb{Q}^n$ is dense in $\mathbb{R}^n$.

4.3.2. Show that Baire space $\mathbb{Z}^{\mathbb{N}_+}$ (see Example 1.1.27) is separable. *Hint:* Think about finite sequences.

4.4 Connectedness

Definition 4.4.1 (Disjoint)**.** Two sets A and B are said to be *disjoint* if $A \cap B = \varnothing$. We sometimes write $A \sqcup B$ for the union of A and B when these sets are disjoint. So the statement '$X = A \sqcup B$' means that $X = A \cup B$ and $A \cap B = \varnothing$.

Definition 4.4.2 (Connected/disconnected)**.** A topological space X is called *disconnected* if $X = U \sqcup V$ for some open nonempty subsets $U, V \subseteq X$. If X is not disconnected, we say it is *connected*.

Remarks 4.4.3.

(a) Note that, in the above definition, $U^\complement = V$ and $V^\complement = U$. Thus U and V are both open and closed. In the above definition, we could replace the word 'open' by 'closed' everywhere.

(b) When we speak of connected and disconnected metric spaces, we are always referring to the metric topology. When we say that Y is a connected/disconnected subset of a metric space (X, d), we mean that Y is connected/disconnected as a metric space on its own (i.e. with the induced metric).

(c) Here is an equivalent definition of connected that we will typically use in proofs: A topological space X is connected if whenever $X = U \sqcup V$ for open sets U, V, and U is nonempty, then $U = X$ and so V is empty.

Examples 4.4.4.

(a) Any set equipped with the discrete metric is disconnected as long as it contains at least two points. To see this, suppose that X is a set equipped with the discrete metric and X has at least two points. Choose $x \in X$ and set $U = \{x\}$, $V = X \setminus \{x\}$. Since all subsets of a metric space with the discrete metric are open, U and V are both open. They are also both nonempty (U is nonempty since X has at least two points). They are also clearly disjoint.

(b) Any singleton $X = \{x\}$ is connected.

(c) The set of rational numbers $\mathbb{Q}$ with the usual metric is disconnected. To see this, let

$$U = \{x \in \mathbb{Q} \mid x < \pi\}, \quad V = \{x \in \mathbb{Q} \mid x > \pi\}.$$

(We could use any irrational number in place of π.) It is clear that $\mathbb{Q} = U \sqcup V$. Also, each of U and V is nonempty since, for instance, $0 \in U$ and $4 \in V$.

It remains to show that U and V are open. We will show that U is open. The proof that V is open is similar. Choose $x \in U$ and set $\epsilon = \pi - x$, which is positive since $x < \pi$. Let us show that $B(x, \epsilon) \subseteq U$. Suppose $z \in B(x, \epsilon)$. Thus, by definition, $z \in \mathbb{Q}$ and $|z - x| < \epsilon$. If $z - x < 0$, then $z < x < \epsilon$ and so $z \in U$. On the other hand, if $z - x > 0$ then

$$z - x = |z - x| < \epsilon \implies z < x + \epsilon = \pi \implies z \in U.$$

Thus $B(x, \epsilon) \subseteq U$ and so (since our choice of $x \in U$ was arbitrary), U is open.

Theorem 4.4.5. *Every closed interval $[a, b]$ of $\mathbb{R}$ is connected.*

Proof. We will prove the result for the interval $I = [0, 1]$, just to avoid unnecessary notational complications. Let I be decomposed into a disjoint union $U \sqcup V$ of two open subsets. Since 0 belongs to exactly one of them, we may assume without loss of generality, and renaming U and V if necessary, that $0 \in U$. Now we shall show that $V = \varnothing$.

Define a set

$$A = \{a \in I \mid [0, a] \subseteq U\}.$$

The set A is nonempty, because it contains zero: $[0, 0] = 0 \subseteq U$. Also, A is bounded above (say, by 1). Dedekind's completeness axiom for the real line implies that the set A has a least upper bound. Denote it by b:

$$b = \sup A.$$

We are going to prove that $1 \in A$. This will finish the proof immediately, because it would mean $[0, 1] \subseteq U$, that is, $[0, 1] = U$ and $V = \varnothing$. With this purpose, we will establish, first, that $b \in A$, and second, that $b = 1$.

Since for every $a \in A$ one has $[0, a] \subseteq U$ and obviously

$$[0, b) \subseteq \bigcup_{a \in A} [0, a],$$

we have $[0, b) \subseteq U$. By the monotonicity of closure, $\mathrm{cl}_I [0, b) \subseteq \mathrm{cl}_I U$. Notice that $\mathrm{cl}_I [0, b) = [0, b]$, and $\mathrm{cl}_I U = U$ (U is closed in I because its complement, $U^{\complement} = V$, is open). These facts imply that $[0, b] \subseteq U$ or, equivalently, $b \in A$.

It remains to prove that $b = 1$. Suppose, towards a contradiction, that $b < 1$. Since $b \in U$ and U is open, there is an $\epsilon > 0$ such that $B(b, \epsilon) \subseteq U$, where the open ball is formed in I. Without loss of generality and decreasing ϵ if necessary, we can assume that $\epsilon \leq 1 - b$. Now one has

$$[0, b + \epsilon/2] = [0, b] \cup [b, b + \epsilon/2] \subseteq [0, b] \cup B(b, \epsilon) \subseteq U,$$

from where one concludes that $b + \epsilon/2 \in A$. But this is a contradiction, because b is, by its choice, an upper bound for A. We conclude: $b = 1$, therefore $1 \in A$ and the proof is finished. □

Lemma 4.4.6. *Suppose Y is a subspace of a metric space X. If U is an open subset of X, then $U \cap Y$ is an open subset of Y.*

Proof. Choose an arbitrary point $y \in U \cap Y$. Since $y \in U$ and U is open in X, there exists an $\epsilon > 0$ such that $B_X(y, \epsilon) \subseteq U$, where we use the subscript X to denote that

$$B_X(y, \epsilon) = \{x \in X \mid d(y, x) < \epsilon\}$$

is the ball formed in X. Now consider the ball with the same centre and radius, but formed in Y:

$$B_Y(y, \epsilon) = \{x \in Y \mid d(y, x) < \epsilon\}.$$

It is clear that $B_Y(y, \epsilon) = B_X(y, \epsilon) \cap Y$. Thus, $B_Y(y, \epsilon) \subseteq U \cap Y$ and so y is an interior point of $U \cap Y$. Since y was an arbitrary point of $U \cap Y$, this set is open. □

Before stating the next theorem, we should introduce (or recall) some notation from set theory. If $\mathcal{C}$ is a collection of sets, then

$$\cup\mathcal{C} = \bigcup_{A\in\mathcal{C}} A = \{x \mid x \in A \text{ for some } A \in \mathcal{C}\}, \quad \text{and}$$
$$\cap\mathcal{C} = \bigcap_{A\in\mathcal{C}} A = \{x \mid x \in A \text{ for all } A \in \mathcal{C}\}.$$

Theorem 4.4.7. *If X is a metric space and $\mathcal{C}$ is a collection of connected subsets of X such that*

(a) $\cup\mathcal{C} = X$, *and*

(b) $\cap\mathcal{C} \neq \varnothing$,

then X is connected.

Proof. Suppose $X = U \sqcup V$ for open sets U, V. Since $\cap\mathcal{C} \neq \varnothing$, there is a common point x_0 contained in all elements of $\mathcal{C}$. Then x_0 is contained in exactly one of U or V. By exchanging the names of U and V if necessary, we can assume $x_0 \in U$. Now choose an arbitrary point $x \in X$. Since $\cup\mathcal{C} = X$, we have $x \in C$ for some $C \in \mathcal{C}$. By Lemma 4.4.6, the sets $U \cap C$ and $V \cap C$ are open in C. Since $X = U \sqcup V$, we have

$$C = (U \cap C) \sqcup (V \cap C).$$

Since C is connected, one of them must be empty. Since $x_0 \in U \cap C$, we have that $V \cap C = \varnothing$. Thus $x \in U \cap C$ and so $x \in U$. Since x was arbitrary, we have $X = U$. □

Corollary 4.4.8. *Every interval of the real line, that is, every set of the form*

$$[a,b],\ (a,b),\ [a,b),\ (a,d],\ (-\infty,b),\ (-\infty,b],\ (a,\infty),\ [a,\infty),\ (-\infty,\infty),$$

is connected.

Proof. For each of these intervals, we can find a collection $\mathcal{C}$ satisfying the hypotheses of Theorem 4.4.7. For instance,

$$(-\infty,\infty) = \mathbb{R} = \bigcup_{n=1}^{\infty} [-n,n].$$

The closed intervals $[-n,n]$ are connected by Theorem 4.4.5 and they all contain the point 0.

For the interval $[a,b)$, one has

$$[a,b) = \bigcup_{n=1}^{\infty} \left[a, b - \frac{b-a}{n}\right].$$

We leave the remaining cases as an exercise (Exercise 4.4.2). □

Exercises.

4.4.1. Show that Baire space $\mathbb{Z}^{\mathbb{N}_+}$ (see Example 1.1.27) is disconnected.

4.4.2. Complete the proof of Corollary 4.4.8.

4.5 Compact sets

In Section 3.1, we discussed sequentially compact metric spaces. We now discuss the notion of a compact set further. We will give the definition of a compact topological space and then see that, in the case of a metric space, this definition agrees with our earlier one (of sequential compactness).

Metric spaces have a nice 'separability' property. Namely, in any metric space (X, d) containing at least two points x and y, we can find open balls (which are open sets in the metric topology) centred at x and y and not intersecting. For instance, we can take the open balls $B(x, \epsilon)$ and $B(y, \epsilon)$ with $\epsilon = d(x, y)/2$. We have a special name for topological spaces which have this property.

Definition 4.5.1 (Hausdorff space)**.** A topological space $(X, \mathcal{T})$ is called a *Hausdorff space*, and $\mathcal{T}$ is called a *Hausdorff topology*, if for all $x, y \in X$, $x \neq y$, there is a neighbourhood U_x of x and a neighbourhood U_y of y such that $U_x \cap U_y = \varnothing$.

As noted above, metric spaces are Hausdorff. Any metric space $(X, \mathcal{T}_{\max})$ with the discrete topology is Hausdorff. For any distinct points $x, y \in X$, the sets $\{x\}$ and $\{y\}$ are disjoint neighbourhoods containing them. However the indiscrete topology $\mathcal{T}_{\min}$ is never Hausdorff (if the space has at least two points) because any nonempty open set is the entire space, and so it is impossible to find disjoint neighbourhoods containing any two distinct points.

Definition 4.5.2 (Open covering)**.** Suppose S is a subset of a topological space X. An *open covering* of S is a collection of open subsets of X whose union contains S. An open covering is *finite* if it is a finite collection. If $\mathcal{C}_1$ and $\mathcal{C}_2$ are open covers, then $\mathcal{C}_1$ is a *subcover* of $\mathcal{C}_2$ if every open subset in $\mathcal{C}_1$ is also an element of $\mathcal{C}_2$ (i.e. $\mathcal{C}_1 \subseteq \mathcal{C}_2$).

Definition 4.5.3 (Compact set)**.** A subset S of a topological space X is called *compact* if every open cover of S contains a finite subcover of S. In other words, S is compact if every collection of open subsets of X whose union contains S has a finite subcollection whose union contains S.

Note that this definition is different from the definition of *sequentially* compact. We will soon see that for metric spaces, the two definitions agree (see Theorem 4.5.9).

Recall that we showed that sequentially compact subsets of metric spaces are closed (Theorem 3.1.5). We now see that there is an analogous theorem in the setting of topological spaces.

Theorem 4.5.4. *Every compact subset of a Hausdorff space is closed.*

Proof. Let S be a compact subset of a Hausdorff space. If $S = X$, then S is obviously closed in X. So we assume $S \neq X$. Recall from Proposition 4.2.11 that S is closed if it contains its cluster points, that is, if $S' \subseteq S$. We will prove $S' \subseteq S$ by showing that if $x \in S^{\complement}$, then $x \notin S'$ (i.e. x is not a cluster point of S). Choose an arbitrary $x \in S^{\complement}$. Since X is Hausdorff, for each $y \in S$, there are disjoint neighbourhoods U_y of x and V_y of y. Then the collection $\{V_y \mid y \in S\}$ is an open covering of S. Since S is compact, there is a finite subcovering. In other words, there is a finite collection

$$\{V_{y_1},\ V_{y_2},\ \dots,\ V_{y_n}\}$$

which is a covering of S. Let $U = \bigcap_{i=1}^n U_{y_i}$. Since U is a finite intersection of open sets, it is itself open. (Here is where we needed a *finite* subcovering.) Since $x \in U_{y_i}$ for $i = 1, 2, \dots, n$, we have $x \in U$. Thus U is a neighbourhood of x. Also, U is disjoint from $V = \bigcup_{i=1}^n V_{y_i}$ and hence from S. Therefore, x is not a cluster point of S. □

In Theorem 4.5.4, we need the assumption that the space is Hausdorff. To see this, consider the following example.

Example 4.5.5. Let X be an infinite set and define

$$\mathcal{T} = \{T \subseteq X \mid T = \varnothing \text{ or } T^{\complement} \text{ is finite}\}.$$

One can show that $\mathcal{T}$ is a topology (Exercise 4.5.1). We will show that every subset of X is compact. Let $S \subseteq X$ and let $\mathcal{C}$ be an open cover of S. Choose one (nonempty) $U \in \mathcal{C}$. Then U is open and hence $U^{\complement}$ is finite. Thus $S \cap U^{\complement}$ is also finite. Say $S \cap U^{\complement} = \{x_1, \dots, x_n\}$. For each $i = 1, \dots, n$, there is a $U_i \in \mathcal{C}$ such that $x_i \in U_i$ (since $\mathcal{C}$ is a cover of S). Then $\{U, U_1, \dots, U_n\}$ is a finite subcover of S. However, X certainly contains subsets that are not closed (for instance, any infinite proper subset of X). Thus X contains compact subsets that are not closed. Note that X is not Hausdorff since any two nonempty opens sets have nontrivial intersection. Indeed, if U and V are nonempty open subsets of X, then $U^{\complement}$ and $V^{\complement}$ are finite. Thus $(U \cap V)^{\complement} = U^{\complement} \cup V^{\complement}$ is finite. Hence $U \cap V$ must be infinite (since X is).

Definition 4.5.6 (Bolzano–Weierstrass property, countably compact)**.** We say that a metric space X has the *Bolzano–Weierstrass property* if every infinite subset of X has a cluster point. We say X is *countably compact* if every countable open covering of X (i.e. countable collection of open sets whose union is X) has a finite subcovering.

Before proving the equivalence of 'compact' and 'sequentially compact', we will need one technical result.

Definition 4.5.7. Suppose S is a nonempty subset of a metric space (X, d) and $\epsilon > 0$. A subset Z of X is called an *ϵ-net* for S if

$$S \subseteq \bigcup_{z \in Z} B(z, \epsilon).$$

In other words, for all $x \in S$, there exists $z \in Z$ such that $d(x, z) < \epsilon$. An ϵ-net is *finite* if it consists of a finite number of points.

Lemma 4.5.8. *Every nonempty sequentially compact subset of a metric space has a finite ϵ-net for every $\epsilon > 0$.*

Proof. Suppose S is a nonempty sequentially compact subset of a metric space (X, d) and suppose that for some $\epsilon > 0$, S does not have a finite ϵ-net. Choose any point $x_1 \in S$. Since the set $\{x_1\}$ cannot be an ϵ-net for S (because it is finite), there is some point $x_2 \in S$ such that $d(x_2, x_1) \geq \epsilon$. Again, the set $\{x_1, x_2\}$ cannot be an ϵ-net for S (because it is finite), and so there is some point $x_3 \in S$ such that $d(x_3, x_1) \geq \epsilon$ and $d(x_3, x_2) \geq \epsilon$. Continuing in this manner, we construct a sequence

$$x_1, x_2, x_3, x_4, \ldots \quad \text{such that} \quad d(x_m, x_n) \geq \epsilon \ \forall\ m \neq n.$$

This sequence cannot have any convergent subsequence, contradicting the sequential compactness of S. Therefore S has a finite ϵ-net for all $\epsilon > 0$. □

Theorem 4.5.9 (Equivalence of two notions of compactness)**.** *Suppose (X, d) is a metric space. Then the following statements are equivalent:*

(a) X is compact,

(b) X is countably compact,

(c) X has the Bolzano–Weierstrass property,

(d) X is sequentially compact.

Proof. We will show

$$(a) \implies (b) \implies (c) \implies (d) \implies (b) \implies (a).$$

Statement (a) clearly implies (b).

(b) $\implies$ (c): We prove this by contradiction. Suppose X is countably compact, but does not have the Bolzano–Weierstrass property. Then X has an infinite subset Y that does not have a cluster point. Let S be any countably infinite subset of Y. Since any cluster point of S would be a cluster point of Y, S has no cluster points. Therefore, each point $x \in S$ has a neighbourhood U_x containing no other point of Y. Recall (Proposition 4.2.11) that a topological space is closed if and only if it contains its cluster points. Therefore (in a trivial way) S is closed, since it has no cluster points. Thus $S^{\complement}$ is open. Then

$$\left\{S^{\complement}\right\} \cup \{U_x \mid x \in S\}$$

is a countable open covering of X. Since X is countably compact, some finite subcovering must cover S. But this is impossible, since each $x \in S$ is contained only in U_x (and not in any U_y, $y \neq x$, nor in $S^{\complement}$) and S is infinite. Therefore (b) $\implies$ (c).

(c) $\implies$ (d): Now suppose X has the Bolzano–Weierstrass property. Let $\{x_n\}$ be a sequence in X. If the range of this sequence is finite, then there must be at least one point that appears infinitely many times in the sequence and thus there is a convergent subsequence (simply take the subsequence whose terms are all this point). Otherwise, the range of the sequence $\{x_n\}$ is infinite and so has a cluster point x. For all $k \in \mathbb{N}_+$, the open ball $B(x, 1/k)$ is a neighbourhood of x and so contains a point x_{n_k} in the sequence, with $x_{n_k} \neq x$. (Since every cluster point of $\{x_n\}_{n=1}^\infty$ is also a cluster point of $\{x_n\}_{n=m}^\infty$ for all $m \in \mathbb{N}_+$, we may assume that the function $k \mapsto n_k$ is increasing.) Then $d(x_{n_k}, x) < 1/k$ for all k and hence $\{x_{n_k}\}_{k=1}^\infty$ is a convergent subsequence of $\{x_n\}$. Therefore X is sequentially compact. Hence (*c*) $\implies$ (*d*).

(*d*) $\implies$ (*b*): We prove this by contradiction. Suppose X is sequentially compact and has a countable open covering $\{T_1, T_2, \dots\}$ with no finite subcovering. Then none of the finite unions $\bigcup_{k=1}^n T_k$, $n \in \mathbb{N}_+$, can be all of X. Hence all of the sets

$$U_n = \left(\bigcup_{k=1}^n T_k \right)^{\complement}, \ n \in \mathbb{N}_+,$$

are nonempty. For each $n \in \mathbb{N}_+$, choose a point $x_n \in U$. Thus $x_n \notin T_k$ for $k = 1, 2, \dots, n$. We will show that $\{x_n\}$ has no convergent subsequence, contradicting the fact that X is sequentially compact. Suppose there is subsequence $\{x_{n_k}\}_{k=1}^\infty$ converging to some limit x. Then, since $\{T_n \mid n \in \mathbb{N}_+\}$ is a cover of X, we must have $x \in T_N$ for some $N \in \mathbb{N}_+$. Since $x_{n_k} \to x$ as $k \to \infty$, there is a K such that $x_{n_k} \in T_N$ for all $k > K$ (this is because T_N is a neighbourhood of x, hence contains some ball $B(x, \epsilon)$ and by the definition of convergence, there is some K such that $k > K$ implies $d(x_{n_k}, x) < \epsilon$). Any larger value of K will satisfy the same condition and so we may assume $K > N$. Since $n_k \geq k$ for all k, we have $x_{n_{K+1}} \in T_N$, which is a contradiction.

(*b*) $\implies$ (*a*): Our strategy to prove this will be to reduce any open cover to a countable cover. Suppose X is countably compact. Then, as we have already proven, it is sequentially compact and hence, by Lemma 4.5.8, it has a finite ϵ-net for all $\epsilon > 0$. In particular, for each $n \in \mathbb{N}_+$, X has a finite $\left(\frac{1}{n}\right)$-net $E\left(\frac{1}{n}\right)$. Then

$$F = \bigcup_{n=1}^\infty E\left(\frac{1}{n}\right)$$

is countable (since it is a countable union of finite sets). Therefore,

$$\mathcal{V} = \{B(u, 1/n) \mid u \in F, \ n \in \mathbb{N}_+\}$$

is a countable set of open balls in X. Now, let x be any point of X and U any neighbourhood of x. Then we can find $m \in \mathbb{N}_+$ such that $B(x, 1/m) \subseteq U$ and $u \in E(1/2m) \subseteq F$ such that $d(u, x) < 1/2m$. Therefore,

$$x \in B(u, 1/2m) \subseteq B(x, 1/m).$$

Thus there is an open ball $B \in \mathcal{V}$ such that $x \in B \subseteq U$.

Now, suppose $\mathcal{U}$ is an arbitrary (i.e. not necessarily countable) open covering of X and let

$$\mathcal{V}_0 = \{B \in \mathcal{V} \mid \exists\ U \in \mathcal{U} \text{ such that } B \subseteq U\}.$$

For each $B \in \mathcal{V}_0$, choose a $U_B \in \mathcal{U}$ such that $B \subseteq U_B$ (we can do this by the definition of $\mathcal{V}_0$). Then

$$\mathcal{U}_0 = \{U_B \mid B \in \mathcal{V}_0\}$$

is a countable subcollection of $\mathcal{U}$. We claim that it is also a covering of X. To see this, suppose $x \in X$. Then $x \in U$ for some $U \in \mathcal{U}$ (since $\mathcal{U}$ is a covering of X). As above, there exists $B \in \mathcal{V}$ such that $x \in B \subseteq U$. Thus $x \in B \subseteq U_B$. Hence $\mathcal{U}_0$ is an open covering of X. Since X is countably compact, $\mathcal{U}_0$ has a finite subcovering of X, which is also a finite subcovering of $\mathcal{U}$. Hence X is compact. □

Remarks 4.5.10. (a) The method of proof of the above theorem shows that any subset of a metric space is compact if and only if it is sequentially compact.

(b) *Warning:* We have only proven that 'compact' is the same as 'sequentially compact' for metric spaces. This is *not* true for arbitrary topological spaces. In fact, in general, neither property implies the other.

Exercises.

4.5.1. Show that $\mathcal{T}$, as defined in Example 4.5.5, is a topology.

4.5.2. Suppose X is a Hausdorff space and $x \in X$. Show that $\{x\}$ is closed.

4.5.3. Show that the discrete topology is the only Hausdorff topology on a finite set.

4.5.4. (a) Show that the intersection of an arbitrary family of compact subsets of a metric space is compact.

(b) Show that the union of two (and therefore, finitely many) compact subsets of a metric space is compact.

(c) Is the union of an arbitrary family of compact subsets of a metric space compact? Prove or give a counterexample.

4.5.5 ([Coh03, Ex. 5.7(15)]). A topological space X is called a T_1-*space* if, given any two distinct points of X, each has a neighbourhood that does not contain the other.

(a) Show that every Hausdorff space is a T_1-space.

(b) Prove that a topological space is a T_1-space if and only if $\{x\}$ is a closed set for every $x \in X$.

(c) Show that every finite T_1-space has the discrete topology.

4.6 Continuity

You have encountered the concept of continuous functions in previous courses. In this section, we will extend this notion to arbitrary metric spaces.

Definition 4.6.1 (Continuous)**.** A mapping f between two metric spaces, (X, d_X) and (Y, d_Y), is called *continuous at a point* $x_0 \in X$, if

$$\forall\, \epsilon > 0\ \exists\, \delta > 0\ \forall\, x \in X,\ d_X(x, x_0) < \delta \implies d_Y(f(x), f(x_0)) < \epsilon,$$

equivalently,

$$\forall\, \epsilon > 0\ \exists\, \delta > 0\ f\big(B_X(x_0, \delta)\big) \subseteq B_Y\big(f(x_0), \epsilon\big).$$

We say f is *continuous on* X if it is continuous at every point of X.

Remark 4.6.2. Note that if X and Y are subsets of $\mathbb{R}$ (with the usual metric), then this is the standard definition of continuity you learned in calculus.

Recall (Definition 1.6.4) that we defined 'sequentially continuous' for mappings between metric spaces. The next result shows that the two notions are equivalent.

Proposition 4.6.3. *A mapping f between two metric spaces, (X, d_X) and (Y, d_Y), is continuous at a point $x_0 \in X$ if and only if it is sequentially continuous at x_0.*

Proof. $\Rightarrow$: Let f be continuous at a point $x_0 \in X$, and let $\{x_n\}$ be a sequence of elements of X such that $x_n \to x_0$. We will show that $f(x_n) \to f(x_0)$. Let $\epsilon > 0$. One can find, due to continuity of f at x_0, a number $\delta > 0$ with $f(B(x_0, \delta)) \subseteq B(f(x_0), \epsilon)$. Since $x_n \to x_0$, for some $N \in \mathbb{N}_+$, one has:

$$\forall\, n \geq N,\ x_n \in B(x_0, \delta),$$

which implies

$$\forall\, n \geq N,\ f(x_n) \in B(f(x), \epsilon),$$

that is, $f(x_n) \to f(x)$, as desired.

$\Leftarrow$: We will prove the contrapositive. Suppose that f is discontinuous (i.e., not continuous) at x_0. By writing down the negation of the definition, we have:

$$\exists\, \epsilon_0\ \forall \delta > 0\ f\big(B(x, \delta)\big) \nsubseteq B\big(f(x_0), \epsilon_0\big).$$

In particular, this is true for all δ of the form $1/n$, $n = 1, 2, 3, \ldots$, so we have

$$\exists\, \epsilon_0 > 0\ \forall\, n \in \mathbb{N}_+\ f\big(B(x_0, 1/n)\big) \nsubseteq B\big(f(x_0), \epsilon_0\big).$$

Fix such an $\epsilon_0 > 0$ and select for any $n \in \mathbb{N}_+$, an element $x_n \in B(x_0, \delta)$ with the property $f(x_n) \notin B(f(x_0), \epsilon_0)$.

Then $x_n \to x_0$ in X, while $f(x_n) \nrightarrow f(x_0)$ in Y, because none of the points $f(x_n)$ is contained in the ball $B(f(x), \epsilon_0)$. ☐

Example 4.6.4. The function $f \colon \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 0, & \text{if } x \text{ is irrational}, \\ x, & \text{if } x \text{ is rational}, \end{cases}$$

is continuous at $x = 0$ and discontinuous at every other point of $\mathbb{R}$.

Definition 4.6.5 (Image and preimage)**.** Let f be a mapping from a set X to another set Y. If $A \subseteq X$, then

$$f(A) = \{f(a) \mid a \in A\}$$

is the *image* of A under the map f.

If $B \subseteq Y$, then

$$f^{-1}(B) = \{x \in X \mid f(x) \in B\}$$

is the *preimage* or *inverse image* of B under f.

Lemma 4.6.6. *Let $f\colon X \to Y$ be a mapping of sets and $B \subseteq Y$. Then*

(a) $f^{-1}(B^{\complement}) = (f^{-1}(B))^{\complement}$, and

(b) if f is surjective, then $f(f^{-1}(B)) = B$.

Proof. This proof is left as an exercise (Exercise 4.6.1). □

Theorem 4.6.7 (Equivalent definitions of continuity)**.** *For a mapping f between metric spaces (X, d_X) and (Y, d_Y), the following conditions are equivalent.*

(a) The mapping f is continuous.

(b) For every $x \in X$ and every (not necessarily open) neighbourhood V of $f(x)$, the inverse image $f^{-1}(V)$ is a (not necessarily open) neighbourhood of x.

(c) The inverse image $f^{-1}(V)$ of every open set $V \subseteq Y$ is open in X.

(d) The inverse image $f^{-1}(F)$ of every closed set $F \subseteq Y$ is closed in X.

(e) For every $A \subseteq X$, one has $f(\operatorname{cl}_X A) \subseteq \operatorname{cl}_Y f(A)$.

Proof. (a) ⇒ (b): If V is a neighbourhood of $f(x)$, then for some $\epsilon > 0$ one has $B(f(x), \epsilon) \subseteq V$. Because of the continuity of f at x, there is a $\delta > 0$ with $f(B(x, \delta)) \subseteq B(f(x), \epsilon)$. Now if follows that $B(x, \delta) \subseteq f^{-1}(V)$, which means that the latter set is a neighbourhood for x.

(b) ⇒ (c): Let V be open in Y and let $x \in f^{-1}(V)$. Then $f(x) \in V$ and, since V is open, $f(x)$ is an interior point of V. Thus V is a neighbourhood of $f(x)$ and so, by assumption, $f^{-1}(V)$ is a neighbourhood of x. So there exists an $\epsilon > 0$ such that $B(x, \epsilon) \subseteq f^{-1}(V)$. Since $x \in f^{-1}(V)$ was arbitrary, $f^{-1}(V)$ is open.

(c) ⇒ (d): Suppose $F \subseteq Y$ is closed. Then $(f^{-1}(F))^{\complement} = f^{-1}(F^{\complement})$ is open (since $F^{\complement}$ is open) and so $f^{-1}(F)$ is closed.

(d) ⇒ (e): Since the set $\operatorname{cl}_Y f(A)$ is closed in Y, the inverse image $f^{-1}(\operatorname{cl}_Y f(A))$ is closed in X. Clearly, this inverse image contains A. Since $\operatorname{cl}_X A$ is the smallest closed set containing A, we have

$$\operatorname{cl}_X A \subseteq f^{-1}(\operatorname{cl}_Y f(A)).$$

This implies $f(\operatorname{cl}_X A) \subseteq \operatorname{cl}_Y f(A)$.

(e) ⇒ (a): Let $x \in X$ and $\epsilon > 0$. Let $A = (f^{-1}(B(f(x), \epsilon)))^{\complement}$. By assumption, we have $f(\operatorname{cl}_X A) \subseteq \operatorname{cl}_Y f(A)$. Since $f(A) \cap B(f(x), \epsilon) = \varnothing$, we have $\operatorname{cl} f(A) \cap B(f(x), \epsilon) = \varnothing$ (because $B(f(x), \epsilon)$ is open and so its complement is closed). Therefore, $f(\operatorname{cl} A) \cap B(f(x), \epsilon) = \varnothing$.

Thus $x \notin \operatorname{cl} A$. Thus, there is some neighbourhood of x which does not intersect A. Hence x is an interior point for $A^{\complement}$. Thus, there is a $\delta > 0$ such that $B(x, \delta) \subseteq f^{-1}(B(f(x), \epsilon))$. Hence $f(B(x, \delta)) \subseteq B(f(x), \epsilon)$. This means that f is continuous at x. Since x was arbitrary, f is continuous. □

Example 4.6.8. If X is equipped with the discrete metric, then any mapping from X to another metric space is continuous. Indeed, if $x \in X$ and $\epsilon > 0$, take $\delta = 1$. Then

$$f(B(x, 1)) = f(\{x\}) = \{f(x)\} \subseteq B(f(x), \epsilon).$$

Definition 4.6.9 (Lipschitz mapping)**.** A mapping $f\colon X \to Y$ between two metric spaces is called *Lipschitz* (*with Lipschitz constant* L) if there exists $L > 0$ such that

$$\forall\ x, y \in X,\ d_Y(f(x), f(y)) \leq L \cdot d_X(x, y).$$

Exercises.

4.6.1. Prove Lemma 4.6.6.

4.6.2. Show that every Lipschitz mapping is uniformly continuous.

4.7 Path-connectedness

Recall that a topological space (in particular, a metric space, given the metric topology) is *connected* if it cannot be written as a disjoint union $U \sqcup V$ of two nonempty open subsets.

Proposition 4.7.1. *Suppose $f\colon X \to Y$ is a surjective continuous mapping of metric spaces. If X is connected, then Y is connected.*

Proof. Let $Y = U \sqcup V$ for some open sets U and V with $U \neq \varnothing$. We want to show that $V = \varnothing$. Let

$$U' = f^{-1}(U), \quad V' = f^{-1}(V).$$

Since f is continuous, U' and V' are open. They are also disjoint since U and V are. Furthermore $U' \cup V' = f^{-1}(U \cup V) = f^{-1}(Y) = X$.

Now, since U is nonempty, we can choose $y \in U$. Because f is surjective, there exists $x \in X$ such that $f(x) = y$. Then, by definition, $x \in U'$ and so U' is nonempty. Since X is connected, V' must be empty. Therefore,

$$V = f(V') = f(\varnothing) = \varnothing,$$

where in the first equality we used that f is surjective and $V' = f^{-1}(V)$. □

Remark 4.7.2. Proposition 4.7.1 says that the continuous image of a connected metric space is connected.

Definition 4.7.3 (Path-connected)**.** A metric space X is called *path-connected* (or *arcwise-connected*) if for every $x, y \in X$, there exists a continuous mapping $p\colon [0,1] \to X$, such that $p(0) = x$ and $p(1) = y$. Such a p is called a *path between x and y*.

Theorem 4.7.4. *Every path-connected metric space is connected.*

Proof. Suppose X is a path-connected metric space. Let $x_0 \in X$ be an arbitrary point and let

$$\mathcal{C} = \{p([0,1]) \mid p \text{ is a continuous path starting at } x_0\}.$$

The family $\mathcal{C}$ consists of connected subsets of X by Proposition 4.7.1, and each one of them contains the point x_0. Furthermore, the union $\bigcup \mathcal{C}$ is all of X since if $x \in X$, then there is a continuous path p' from x_0 to x, and so

$$x \in p'([0,1]) \subseteq \bigcup \mathcal{C}.$$

Thus, X is connected by Theorem 4.4.7. □

Example 4.7.5. Every interval of the real line is connected (one can easily construct a path between any two points).

Example 4.7.6. Euclidean space $\mathbb{R}^n$ is path-connected. To see this, let $x, y \in \mathbb{R}^n$. Define a path by

$$p(t) = (1-t)x + ty, \quad t \in [0,1].$$

Then $p\colon [0,1] \to \mathbb{R}^n$ is continuous since every coordinate component is continuous. Since $p(0) = x$ and $p(1) = y$, p is a path from x to y.

Remark 4.7.7. One can extend the definition of a continuous map to the setting of topological spaces and then define path-connectedness is this setting. It remains true that path-connected spaces are connected (so this is not a special property of metric spaces).

Although path-connected implies connected, the converse is not true. Before giving a counter-example, we need two results.

Lemma 4.7.8. *Let X be a (nonempty) path-connected subset of $\mathbb{R}$ and let $f\colon X \to \mathbb{R}$ be a continuous function. Then the graph of f,*

$$\operatorname{gr} f = \{(x, f(x)) \mid x \in X\},$$

is a connected metric subspace of the Euclidean plane $\mathbb{R}^2$.

Proof. The proof of this lemma is left as an exercise (Exercise 4.7.4). □

Remark 4.7.9. The above lemma can actually be generalized to the case where f is a continuous mapping between two metric spaces X and Y, and the graph $\operatorname{gr} f$ is viewed as a metric subspace of $X \times Y$ equipped, for instance, with the metric

$$d((x_1, y_1), (x_2, y_2)) = d_X(x_1, x_2) + d_Y(y_1, y_2).$$

Lemma 4.7.10. *Suppose a metric space X contains a connected everywhere dense subset Z. Then X is connected.*

Proof. The proof of this lemma is left as an exercise (Exercise 4.7.5). □

We now give a classic example of a space that is connected but not path-connected.

Example 4.7.11 (Closed topologist's sine curve). Consider the following two subsets of $\mathbb{R}^2$:

$$X_1 = \left\{ \left(x, \sin \frac{1}{x} \right) \;\middle|\; x > 0 \right\} \quad \text{and} \quad X_2 = \{0\} \times [-1, 1].$$

Then X_1 is connected by Lemma 4.7.8 and it is easy to see that X_2 is connected (see Exercise 4.7.6). Now let

$$X = X_1 \cup X_2,$$

with the induced metric from $\mathbb{R}^2$.

Every point of X_2 is a closure point for X_1 (draw a picture, and then for each point $x \in X_2$, find a sequence of elements of X_1 that converges to x). Thus X_1 is everywhere dense in X. Therefore, by Lemma 4.7.10, X is connected.

We claim that X is not path-connected. Let p be a continuous path in X with $p(0) \in X_1$, $p(1) = (0, 0) \in X_2$. We will obtain a contradiction, showing that such a path does not exist.

The inverse image $p^{-1}(X_2)$ is a nonempty closed subset of the interval $[0, 1]$ and so $r = \inf p^{-1}(X_2) \in p^{-1}(X_2)$. Since $0 \notin p^{-1}(X_2)$, we have $r > 0$. We have $p(r) \in X_2$, and so r is the smallest element of $[0, 1]$ whose image under p is contained in X_2.

Since p is continuous, there exists a $\delta > 0$ such that

$$t \in [0, 1], \; |t - r| < \delta \implies d(p(t), p(r)) < 1.$$

(Here d is the usual Euclidean distance in the plane.)

The intersection

$$B(p(r), 1) \cap X_1$$

is the union of countably many disjoint open intervals. We claim that $p((r - \delta, r]) \subseteq X_2$. Indeed, $p((r - \delta, r])$ is connected by Proposition 4.7.1. So if it contained a point of X_1, it would have to be contained in one of the above-mentioned open intervals and thus would not meet X_2. This would contradict the fact that $p(r) \in X_2$. This proves our claim. But now $p(r - \delta/2) \in X_2$, contradicting the choice of r.

These concepts have a familiar application.

Theorem 4.7.12 (Intermediate Value Theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. If r is a real number between $f(a)$ and $f(b)$, then there exists $\xi \in [a, b]$ such that $f(\xi) = r$.*

Proof. Without loss of generality, assume $f(a) < f(b)$. (If $f(a) = f(b)$, there is nothing to prove. If $f(a) > f(b)$, replace f by $-f$.) Let

$$f(a) < r < f(b).$$

Suppose, towards a contradiction, that $r \notin f([a,b])$. Let

$$U = \{f(t) \mid t \in [a,b],\ f(t) < r\} = f([a,b]) \cap (-\infty, r),$$
$$V = \{f(t) \mid t \in [a,b],\ f(t) > r\} = f([a,b]) \cap (r, \infty).$$

Since each of U and V is the intersection of $f([a,b])$ with an open set, both U and V are open in $f([a,b])$ by Lemma 4.4.6. We also have $U \cup V = f([a,b])$ because we assumed $r \notin f([a,b])$. Since $f(a) \in U$ and $f(b) \in V$, both U and V are nonempty. We have thus shown that the image of the (connected) interval $[a,b]$ under the continuous mapping f is disconnected. But this contradicts Proposition 4.7.1. □

Exercises.

4.7.1. Show that ℓ^∞ is path-connected using an argument similar to that of Example 4.7.6.

4.7.2. Let X be a metric space, and let $\mathcal{C}$ be family of path-connected subsets of X with the following properties:

(a) $\bigcup \mathcal{C} = X$, and

(b) for all $C, D \in \mathcal{C}$, $C \cap D \neq \varnothing$.

Prove that X is path-connected.

4.7.3. Suppose $n \geq 2$, and let

$$X = \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_i \in \mathbb{Q} \text{ for some } i = 1, \ldots, n\}.$$

Show that X is path-connected.

4.7.4. Prove Lemma 4.7.8

4.7.5. Prove Lemma 4.7.10.

4.7.6. Prove that the subset X_2 of $\mathbb{R}^2$ defined in Example 4.7.11 is connected.

4.8 Isometries, isometric embeddings, and completions

In this section we make precise the notion that two metric spaces are "the same".

Definition 4.8.1 (Isometry)**.** A mapping $f\colon X \to Y$ between two metric spaces is called an *isometry*, or an *isometric isomorphism*, if

(a) f is a bijection, and

(b) for all $x, y \in X$, we have $d_X(x,y) = d_Y(f(x), f(y))$.

We say that two metric spaces are *isometric*, or *isometrically isomorphic*, if there is an isometry between them.

The relation of being isometric is an equivalence relation (easy exercise). Two isometric spaces will have the same properties. That is, if X and Y are isometric and one of them is complete, then the other is complete, and so on.

Definition 4.8.2 (Isometric embedding)**.** A mapping $j\colon X \to Y$ is called an *isometric embedding* if

$$d_X(x,y) = d_Y(j(x), j(y)) \quad \forall\, x, y \in X.$$

Remark 4.8.3. We will often use the notation $j\colon X \hookrightarrow Y$ for an isometric embedding. Note that any isometric embedding $j\colon X \hookrightarrow Y$ is injective since

$$j(x) = j(y) \implies d(x,y) = d(j(x), j(y)) = 0 \implies x = y.$$

However, isometric embeddings need not be surjective. We can think of them as isometries from X to some subspace of Y.

Remark 4.8.4. It follows easily from the definition that every isometric embedding is continuous (see Exercise 4.8.2).

Recall from Section 1.4 that a metric space is complete if every Cauchy sequence converges in that space.

Theorem 4.8.5. *For a metric space X, the following conditions are equivalent.*

(a) X is complete.

(b) Every decreasing chain of closed balls whose radii converge to zero has a nonempty intersection. In other words, if $x_i \in X$ and $\epsilon_i > 0$, $i \in \mathbb{N}_+$, are such that $\epsilon_i \to 0$ and

$$B^{\mathrm{cl}}(x_1, \epsilon_1) \supseteq B^{\mathrm{cl}}(x_2, \epsilon_2) \supseteq B^{\mathrm{cl}}(x_3, \epsilon_3) \supseteq \cdots,$$

then $\bigcap_{i=1}^{\infty} B^{\mathrm{cl}}(x_i, \epsilon_i) \neq \varnothing$.

Proof. (a) ⇒ (b): Assume X is complete, and let $x_i \in X$ and $\epsilon_i > 0$ be as in condition (b). We claim that the centers $\{x_i\}$ of the balls form a Cauchy sequence. Indeed, given an $\epsilon > 0$, choose N so that for all $n \geq N$ one has $\epsilon_n < \epsilon/2$. Then for every such n, since the closed ball $B^{\mathrm{cl}}(x_n, \epsilon_n)$ is contained in the closed ball $B^{\mathrm{cl}}(x_N, \epsilon_N)$ and in particular $x_n \in B^{\mathrm{cl}}(x_N, \epsilon_N)$, we have

$$d(x_N, x_n) \leq \epsilon_N < \epsilon/2.$$

Then, for all $m, n \geq N$, we have

$$d(x_n, x_m) \leq d(x_n, x_N) + d(x_N, x_m) < \epsilon/2 + \epsilon/2 = \epsilon.$$

Therefore $\{x_n\}$ is a Cauchy sequence. Since X is complete, this sequence converges to some $x \in X$. We claim that $x \in \bigcap_{i=1}^{\infty} B^{\mathrm{cl}}(x_i, \epsilon_i)$ and hence this intersection is nonempty. Indeed, for all $i \in \mathbb{N}_+$, the sequence $\{x_n\}_{n=i}^\infty$ is contained in $B^{\text{cl}}(x_i, \epsilon_i)$ and converges to x. Since $B^{\text{cl}}(x_i, \epsilon_i)$ is closed, $x \in B^{\text{cl}}(x_i, \epsilon_i)$.

(b) $\Rightarrow$ (a): Assume the statement in (b) holds and let $\{x_n\}$ be a Cauchy sequence in X. For each $i \in \mathbb{N}_+$, choose N_i such that $N_i > N_{i-1}$ and

$$n > N_i \implies d(x_{N_i}, x_n) < 2^{-N_{i-1}}.$$

We claim that the sequence of closed balls

$$B_i = B^{\text{cl}}\left(x_{N_i}, 2^{-N_{i-1}+1}\right), \; i \in \mathbb{N}_+,$$

is decreasing. Indeed, for $i \in \mathbb{N}_+$, let $y \in B_{i+1}$. Then

$$d(x_{N_i}, y) \le d(x_{N_i}, x_{N_{i+1}}) + d(x_{N_{i+1}}, y) \le 2^{-N_{i-1}} + 2^{-N_i+1} \le 2^{-N_{i-1}} + 2^{-N_{i-1}} = 2^{-N_{i-1}+1},$$

and so $y \in B_i$. Hence $B_{i+1} \subseteq B_i$. By assumption, there is a point

$$x \in \bigcap_{i=1}^{\infty} B_i.$$

We claim that $x_i \to x$. First note that

$$x \in B_i \implies d(x_{N_i}, x) \le 2^{-N_{i-1}},$$

and so $x_{N_i} \to x$. Thus $\{x_{N_i}\}$ is a subsequence of $\{x_n\}$, converging to x. The result then follows from Proposition 1.4.15. □

Example 4.8.6 (Cantor intersection property). In the case $X = \mathbb{R}$, a closed ball is just a closed interval. Therefore, condition (b) in Theorem 4.8.5 becomes:

Every decreasing sequence of closed intervals whose length goes to zero has a common point.

This property is known as the *Cantor intersection property.*

We know that not all metric spaces are complete. Given a metric space that is not complete, is there a way to "fill the gaps"? That is, is there a way to embed it inside a complete metric space? The answer is yes.

Theorem 4.8.7. *Let (X, d) be a metric space. There exists a complete metric space $(\hat{X}, \hat{d})$ containing (X, d) as an everywhere dense metric subspace.*

Remark 4.8.8. In the above theorem, by 'contains', we mean that X embeds isometrically into $\hat{X}$ and not that X is necessarily actually a metric subspace of $\hat{X}$ itself.

Sketch of proof. Let $\mathcal{N}$ be the set of all Cauchy sequences in X (with or without limit). Define an equivalence relation $\sim$ on $\mathcal{N}$ by saying

$$\{x_n\} \sim \{y_n\} \iff d(x_n, y_n) \to 0, \quad \{x_n\}, \{y_n\} \in \mathcal{N}. \tag{4.1}$$

We leave it as an exercise to verify that this is an equivalence relation (Exercise 4.8.1).

Let $\hat{X} = \mathcal{N}/\sim$ be the set of equivalence classes. So elements of $\mathcal{N}$ are equivalence classes $[\{x_n\}]$ of Cauchy sequences $\{x_n\}$ in X.

Define a metric on $\hat{X}$ by

$$\hat{d}\big([\{x_n\}],[\{y_n\}]\big) \stackrel{\text{def}}{=} \lim_{n\to\infty} d(x_n,y_n).$$

Some technical work is needed to verify that the above distance is well-defined, that is, the limit exists, the value of the limit does not depend on a choice of representatives of equivalence classes, and that all three axioms of metric are fulfilled.

Let us embed X as a metric subspace into $\hat{X}$. This is done by identifying every element $x \in X$ with the equivalence class of the corresponding constant sequence,

$$[\{x,x,x,\dots\}].$$

In other words, we associate to an $x \in X$ the collection of all sequences in X that converge to x. It is straightforward to verify that such an identification is an isometric embedding.

To show that X is everywhere dense in $\hat{X}$, take an element $z = [\{x_n\}] \in \hat{X}$ and $\epsilon > 0$. We will find an element $x \in X$ in the ball $B(z,\epsilon)$. Pick a number N with the property

$$\forall\ m,n \in \mathbb{N}_+,\ d(x_m,x_n) \le \epsilon/2,$$

and put $x = x_{N+1}$. We then have

$$\hat{d}([\{x,x,x\dots\}],z) < \epsilon,$$

as desired.

Finally, let us show that $(\hat{X},\hat{d})$ is a complete metric space. Suppose $\{z_n\}$ is a Cauchy sequence in $\hat{X}$. For every $n \in \mathbb{N}_+$, let

$$z_n = [\{z_{n,m}\}_{m\in\mathbb{N}_+}],$$

where $\{z_{n,m}\}_{m\in\mathbb{N}_+}$ is a Cauchy sequence in X (hence its equivalence class is a point in $\hat{X}$) for every $n \in \mathbb{N}_+$.

For each $k \in \mathbb{N}_+$, find a number $N(k)$ such that

$$p,q > N(k) \implies d(z_{k,p},z_{k,q}) < 1/k.$$

One can verify that the sequence $\{z_{k,N(k)}\}_{k\in\mathbb{N}_+}$ is Cauchy. Now let $z = [\{z_{k,N(k)}\}_{k\in\mathbb{N}_+}]$. One can check from the definition that $z_n \to z$ in the metric space $(\hat{X},\hat{d})$. □

Definition 4.8.9 (Completion)**.** A *completion* of a metric space X is a complete metric space containing X as an everywhere dense subspace.

For example, the space $(\hat{X},\hat{d})$ constructed in Theorem 4.8.7 is a completion of (X,d).

Exercises.

4.8.1. Prove that (4.1) is an equivalence relation.

4.8.2. Prove that every isometric embedding is continuous.

4.8.3. Let X and Y be two isometric metric spaces. Suppose that X is path-connected. Prove that Y is path-connected as well.

4.8.4. Consider the set $\mathbb{N}_+$ with the metric

$$d(m,n) = \begin{cases} \frac{1}{m+n} + 1 & \text{if } m \neq n, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Prove that d is indeed a metric on $\mathbb{N}_+$.

(b) Show that the metric topology is the discrete topology.

(c) Show that $(\mathbb{N}_+, d)$ is complete.

(d) Let

$$S_n = B^{\mathrm{cl}}(n, 1 + 1/2n) = \left\{ m \in \mathbb{N}_+ \;\middle|\; d(m,n) \leq 1 + \frac{1}{2n} \right\}.$$

Show that $S_n = \{n, n+1, n+2, \dots\}$.

(e) Show that $\{S_n\}$ is a descending sequence of closed balls whose intersection is empty. That is, show that

$$S_1 \supseteq S_2 \supseteq S_3 \supseteq \cdots$$

and $\bigcap_{n=1}^{\infty} S_n = \varnothing$.

(f) Why does this example not contradict Theorem 4.8.5?

Chapter 5

Normed vector spaces

A metric space is simply a set with a metric (which must satisfy some axioms). In general, in a metric space, there is no notion of addition or scalar multiplication. However, many of our *examples* of metric spaces, naturally have such structures, because they are also vector spaces (for instance $\mathbb{R}^n$, $\mathbb{C}^n$, $C[0,1]$). We now turn out attention to combining these two topics. That is, we want to consider mathematical objects which are both vector spaces and metric spaces, and with some sort of compatibility between the two structures.

5.1 Definitions and examples

We will use the boldface notation $\mathbf{0}$ to denote the zero of an arbitrary vector space, and 0 to denote the scalar zero.

Definition 5.1.1 (Normed vector space)**.** A *normed vector space* (or simply a *normed space*) is a vector space X over the field $F = \mathbb{R}$ or $\mathbb{C}$, together with a mapping $\|\cdot\|\colon X \to \mathbb{R}_+$ having the following properties:

(N1) For $x \in X$, $\|x\| = 0$ if and only if $x = \mathbf{0}$.

(N2) $\|\alpha x\| = |\alpha| \cdot \|x\|$ for all $x \in X$ and $\alpha \in F$.

(N3) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in X$ (*triangle inequality*).

The map $\|\cdot\|$ is called a *norm*. If we wish to emphasize the space in question, we write $\|\cdot\|_X$ for the norm on X.

Remark 5.1.2. Note that the definition of a norm uses the absolute value of the field. This is why we work with the real or complex numbers. However, there are other fields with absolute values. The study of normed vector spaces over these other fields is called *non-archimedean functional analysis*.

We have formulated the definition of a normed vector space, having in mind that

$$\|x\| = d(x, \mathbf{0})$$

should be thought of as the distance of the vector x from the origin. Then, as in the vector spaces $\mathbb{R}^n$, we should be able to introduce a metric by setting

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

So axiom (N1) enures that this metric will satisfy axiom (M1). Axiom (N3) ensures that this metric satisfies the triangle inequality:

$$\begin{aligned} d(x, z) &= \|x - z\| \\ &= \|x - y + y - z\| \\ &\leq \|x - y\| + \|y - x\| \\ &= d(x, y) + d(y, z). \end{aligned}$$

Finally, axiom (N2) ensures that the distance function is symmetric:

$$\begin{aligned} d(x, y) &= \|x - y\| \\ &= \|(-1)(y - x)\| \\ &= |-1| \cdot \|y - z\| \\ &= \|y - x\| \\ &= d(y, x). \end{aligned}$$

Of course, axiom (N2) is stronger than what we need to ensure symmetry of the metric. But it matches with our intuition that

$$\|\alpha x\| = d(\alpha x, \mathbf{0}) = |\alpha| \cdot d(x, \mathbf{0}) = |\alpha| \cdot \|x\|.$$

Definition 5.1.3 (Metric associated with a norm)**.** Let $X = (X, \| \cdot \|)$ be a normed vector space. The *metric associated with* X (or with $\| \cdot \|$) is the metric given by

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

Example 5.1.4. The real vector space $\mathbb{R}^1$ admits only one norm (up to a scalar factor): $\|x\| = |x|$. The same is true for the complex vector space $\mathbb{C}^1$.

Example 5.1.5 (Euclidean norm). The *Euclidean norm* on $\mathbb{R}^n$ is given by

$$\|x\|_2 = \sqrt{\sum_{k=1}^{n} x_k^2}.$$

Axioms (N1) and (N2) are easy to check. Axiom (N3) follows from the Cauchy–Schwarz inequality (this is Proposition 1.1.11). If we refer to $\mathbb{R}^n$ as a normed vector space, we assume this is the norm. The associated metric is the Euclidean metric on $\mathbb{R}^n$.

The Euclidean norm on $\mathbb{C}^n$ is given by

$$\|x\|_2 = \sqrt{\sum_{k=1}^{n} |x_k|^2}.$$

Again, this is our default norm on $\mathbb{C}^n$ and the associated metric is the Euclidean metric on $\mathbb{C}^n$.

Example 5.1.6 (ℓ^∞-norm on $\mathbb{R}^n$ and $\mathbb{C}^n$). The ℓ^∞-norm on $\mathbb{R}^n$ or $\mathbb{C}^n$ is given by

$$\|x\|_\infty = \max_{k=1}^{n} |x_k|.$$

We leave it as an exercise to prove that this defines a norm (Exercise 5.1.3). The associated metric is the ℓ^∞-metric.

Example 5.1.7 (ℓ^p-norm on $\mathbb{R}^n$ and $\mathbb{C}^n$). The ℓ^p-norm, $p \geq 1$, on $\mathbb{R}^n$ or $\mathbb{C}^n$ is given by

$$\|x\|_p = \left(\sum_{k=1}^{n} |x_k|^p \right)^{1/p}.$$

The triangle inequality for this norm is called the *Minkowski inequality*. We will not prove it in this course. (You can find a proof by looking up the Minkowski inequality on Wikipedia (or in other references) if you like.) The ℓ^2-norm is just the Euclidean norm. The metric associated to the ℓ^p-norm is the ℓ^p metric.

Example 5.1.8 (Norm on ℓ^∞). Recall that ℓ^∞ is the space of all (real or complex) infinite bounded sequences. We can define a norm on ℓ^∞ by

$$\|x\|_\infty = \sup_{k=1}^{\infty} |x_k|, \quad x = (x_1, x_2, \dots) \in \ell^\infty.$$

We leave it as an exercise to check that this defines a norm (Exercise 5.1.5). It is easy to see that the associated metric is the usual metric on ℓ^∞.

Example 5.1.9 (Norm on ℓ^p). Recall that ℓ^p, $p \geq 1$, is the space of all (real or complex) infinite sequences $x = (x_1, x_2, \dots)$ such that

$$\sum_{k=1}^{\infty} |x_k|^p$$

converges. We can define a norm on ℓ^p by

$$\|x\|_p = \left(\sum_{k=1}^{\infty} |x_k|^p \right)^{1/p}.$$

Again, the proof of the triangle inequality (Minkowski's inequality) will not be done in this course (the other axioms of a norm are easy to check). It is easy to check that the associated metric is the usual metric on ℓ^p. However, we should check that ℓ^p is indeed a vector space. We do this for $p = 2$ (although it is true in general). Let c_0 be the set of all complex-valued sequences that converge to zero. It is an easy exercise to check that this is a vector space (under component-wise addition and scalar multiplication). Since the terms of any convergent series approach zero, ℓ^p is a subset of c_0. So we can prove that ℓ^p is a vector space by showing that it is a subspace of c_0. It is easy to see that the zero sequence is in ℓ^p and that ℓ^p is closed under scalar multiplication. Suppose $x = (x_1, x_2, \dots), y = (y_1, y_2, \dots) \in \ell^p$. Then

$$(|x_k| - |y_k|)^2 \geq 0 \implies |x_k|^2 + |y_k|^2 \geq 2|x_k| \cdot |y_k|$$

and so

$$|x_k + y_k|^2 \leq (|x_k| + |y_k|)^2 = |x_k|^2 + |y_k|^2 + 2|x_k| \cdot |y_k| \leq 2(|x_k|^2 + |y_k|^2).$$

Therefore, convergence of $\sum |x_k|^2$ and $\sum |y_k|^2$ implies convergence of $\sum |x_k + y_k|^2$.

Remark 5.1.10. As in Remark 1.1.20, we require $p \geq 1$ since the triangle inequality is not satisfied for $0 < p < 1$. However, unlike in the setting of metric spaces, we cannot fix this in a natural way since $\|x\| = \sum_{k=1}^{\infty} |x_k|^p$ does *not* define a norm (see Exercise 5.1.6).

Remark 5.1.11. As before, when we considered them only as metric spaces, we can think of $\mathbb{R}^n$ and $\mathbb{C}^n$ as subspaces of ℓ^p and ℓ^∞ by considering sequences that are zero after the nth term.

Example 5.1.12 (Uniform norm on $C[a, b]$). The *uniform norm* on $C[a, b]$ is given by

$$\|x\|_\infty = \max_{a \leq t \leq b} |x(t)|, \quad x \in C[a, b].$$

When we refer to $C[a, b]$ as a normed space, we imply this norm. The associated metric is the uniform metric.

Example 5.1.13 (The normed space $C_p[a, b]$). For $p \geq 1$, we can define a norm on $C[a, b]$ by

$$\|x\|_p = \left(\int_a^b |x(t)|^p \, dt \right)^{1/p}, \quad x \in C[a, b].$$

We denote the resulting normed space by $C_p[a, b]$. Since the associated metric is the metric on $C_p[a, b]$, this matches our previous notation.

Exercises.

5.1.1 ([Coh03, Ex. 6.4(1)]). Suppose X is a normed vector space and let d be the associated metric (see Definition 5.1.3). Show that

$$d(x + z, y + z) = d(x, y) \quad \text{and} \quad d(\alpha x, \alpha y) = |\alpha| d(x, y),$$

for any $x, y, z \in X$ and any scalar α. We say that the metric is *translation invariant* and *homogeneous*.

5.1.2 ([Coh03, Ex. 6.4(2)]). Suppose X is a normed space.

(a) Show that $\|x - y\| \geq |\, \|x\| - \|y\| \,|$ for all $x, y \in X$.

(b) Show that, if $x \neq \mathbf{0}$, then $\|(1/\alpha)x\| = 1$ when $\alpha = \|x\|$.

5.1.3. Prove that $\|x\|_\infty$, as defined in Example 5.1.6 is a norm.

5.1.4. (a) Prove that

$$\|x\|_o = |x_1| + \cdots + |x_n|, \quad x = (x_1, \ldots, x_n) \in \mathbb{C}^n,$$

defines a norm on $\mathbb{C}^n$.

(b) If $\| \ \|$ is the Euclidean norm on $\mathbb{C}^n$ and $\| \ \|_\infty$ is the norm on $\mathbb{C}^n$ defined in Example 5.1.6, prove that, for all $x \in \mathbb{C}^n$.

(i) $\|x\|_\infty \le \|x\| \le \sqrt{n}\|x\|_\infty$,

(ii) $\|x\| \le \|x\|_o \le \sqrt{n}\|x\|$,

(iii) $\frac{1}{n}\|x\|_o \le \|x\|_c \le \|x\|_o$.

5.1.5. With notation as in Example 5.1.8, show that ℓ^∞ is a vector space and that $\|\cdot\|_\infty$ defines a norm on this space.

5.1.6. In the setup of Remark 5.1.10, which axiom in the definition of a norm is violated if we define $\|x\| = \sum_{k=1}^\infty |x_k|^p$?

5.1.7. For $p = 1$ and $p = 2$, verify that $\|x\|_p$, as defined in Example 5.1.13, is a norm on $C_p[a, b]$. *Hint*: Use Exercise 1.1.14 for the $p = 2$ case.

5.1.8. Let S be a nonempty subset of a normed space. Recall the definition of bounded subset of a metric space (Definition 1.5.6). Prove that S is bounded (in the metric associated with X) if and only if there is a positive number M such that $\|x\| \le M$ for all $x \in S$.

5.1.9 ([Coh03, Ex 6.5(12)]). If X is a vector space, a *seminorm* on X is a mapping $\nu \colon X \to \mathbb{R}_{\ge 0}$ satisfying $\nu(\mathbf{0}) = 0$, $\nu(\alpha x) = |\alpha|\nu(x)$, $\nu(x + y) \le \nu(x) + \nu(y)$, for all $x, y \in X$ and all scalars α. (The second requirement of (N1) is omitted; compare this to the definition of a semimetric in Exercise 1.1.18.)

Let P be the real vector space of all polynomial functions. Prove that

$$\nu(p) = |p(0)| + |p'(0)| + |p''(0)|, \quad p \in P,$$

defines a seminorm for P, but not a norm. Determine all polynomial functions $p \in P$ for which $\nu(p) = 0$ and show that they form a subspace of P.

5.2 Convergence in normed vector spaces

Since to any norm we have an associated metric, we can speak of convergence in normed spaces (since we have a definition of convergence in a metric space). Let us formulate the definition of convergence in the setting of normed spaces.

A sequence $\{x_n\}$ in a normed space X is *convergent* if there exists an $x \in X$ such that for all $\epsilon > 0$, there exists $N > 0$ such that

$$n > N \implies \|x_n - x\| < \epsilon.$$

The sequence $\{x_n\}$ is a *Cauchy sequence* if for all $\epsilon > 0$, there exists an $N > 0$ such that

$$m, n > N \implies \|x_n - x_m\| < \epsilon.$$

We say a normed space is *complete* if every Cauchy sequence in it converges.

Definition 5.2.1 (Banach space)**.** A *Banach space* is a complete normed vector space.

Example 5.2.2. The following are Banach spaces:

- $\mathbb{R}^n$, $\mathbb{C}^n$ with the Euclidean, ℓ^p ($p \geq 1$), or ℓ^∞ norms,
- ℓ^p ($p \geq 1$),
- ℓ^∞,
- $C[a, b]$.

Indeed, we defined norms on them and showed in Section 1.4 that the associated metric spaces are complete. However $C_1[a, b]$ is *not* a Banach space since we showed that it is not complete in Example 1.4.12.

We can apply all of the terms we learned for metric spaces to normed spaces by considering the associated metric. For instance, we can speak of normed spaces being compact, sequentially compact, etc., and we can talk about contraction mappings between normed spaces. Additionally, on any normed space, we have the topology induced by the associated metric.

We need to be a bit careful when discussing *subspaces* of normed spaces. Whereas any subset of a metric space was a metric space itself under the induced metric, an arbitrary subset of a normed space may not be a vector space and hence not a normed space. However, we can still talk of compact subsets, closed subsets, etc. of normed spaces (and we do not require them to be vector subspaces). Since the topology on a normed space is the one coming from the metric topology, we know that any subset is compact if and only if it is sequentially compact. So we simply use the term *compact*.

Definition 5.2.3 (Series)**.** Suppose $\{x_n\}$ is a sequence in a normed space X and let $s_n = \sum_{k=1}^n x_k$ be the *partial sum*. Then $\{s_n\}$ is also a sequence in X. We say the *series* $\sum_{k=1}^\infty x_k$ *converges* if $\lim_{n\to\infty} s_n$ exists, in which case we call $\lim s_n$ the *sum* of the series. We say the series $\sum x_k$ is *absolutely convergent* if the series $\sum \|x_k\|$ of real numbers is convergent (so here we are talking about convergence in the normed space $\mathbb{R}$).

Remark 5.2.4. Note that we need the setting of normed spaces in order for series to be defined. Series make no sense in general metric spaces because we have no notion of addition.

You learned in calculus that any absolutely convergent sequence of real numbers is convergent. However, this is not true for all normed spaces. As we will now see, it is true for real numbers since $\mathbb{R}$ is a Banach space.

Theorem 5.2.5. *A normed vector space X is a Banach space if and only if every absolutely convergent series in X is convergent.*

Proof. First suppose X is a Banach space and $\sum x_k$ is an absolutely convergent series in X. So $\sum \|x_k\|$ converges. We wish to show that $\sum x_k$ converges. Choose $\epsilon > 0$. Recall from previous courses (see, for example, [Coh03, Th. 1.8.2]) that a series $\sum y_k$ of real numbers converges if and only if for all $\epsilon' > 0$, there exists an $N > 0$ such that

$$n \geq m > N \implies \left| \sum_{k=m}^{n} y_k \right| < \epsilon.$$

(In other words, the series $\sum y_k$ converges if and only if the sequence of partial sums is Cauchy.) So, taking $y_k = \|x_k\|$, we can find an $N > 0$ such that

$$n \geq m > N \implies \left\| \sum_{k=m}^{n} x_k \right\| \leq \sum_{k=m}^{n} \|x_k\| < \epsilon$$

(the first inequality after the '$\implies$' is the triangle inequality). This means that

$$n \geq m > N \implies \|s_n - s_{m-1}\| < \epsilon,$$

and so the partial sums $s_n = \sum_{k=1}^{n} x_k$ form a Cauchy sequence in X. Since X is Banach space (hence complete), the partial sums converge.

Now suppose that every absolutely convergent sequence in X converges. We wish to show that X is complete, that is, every Cauchy sequence converges. Let $\{x_n\}$ be a Cauchy sequence in X. Recall (Proposition 1.4.15) that if $\{x_n\}$ contains some convergent subsequence, then $\{x_n\}$ converges. Our goal is to find a convergent subsequence of $\{x_n\}$.

For each $k \in \mathbb{N}_+$, choose $N_k > 0$ such that

$$m, n \geq N_k \implies \|x_n - x_m\| < 1/2^k.$$

By increasing them if necessary, we may assume $N_1 < N_2 < N_3 < \dots$. Then for each k we have

$$\|x_{N_{k+1}} - x_{N_k}\| < \frac{1}{2^k}.$$

Thus

$$\sum_{k=1}^{\infty} \|x_{N_{k+1}} - x_{N_k}\| \leq \sum_{k=1}^{\infty} \frac{1}{2^k} = 1,$$

and so the series $\sum \|x_{N_{k+1}} - x_{N_k}\|$ is convergent. Therefore, the series

$$\sum_{k=1}^{\infty} (x_{N_{k+1}} - x_{N_k})$$

is absolutely convergent and hence convergent by our assumption. By definition, this means that its sequence $\{s_m\}_{m=1}^{\infty}$ of partial sums converges. In other words, there exists a point $s \in X$ such that $\|s_m - s\| \to 0$. Now,

$$s_m = \sum_{k=1}^{m} (x_{N_{k+1}} - x_{N_k}) = x_{N_{m+1}} - x_{N_1}.$$

Let $x = s + x_{N_1}$. Then

$$\|x_{N_{m+1}} - x\| = \|s_m - s\| \to 0.$$

Hence the subsequence $\{x_{N_k}\}$ converges (to x) as desired. □

In its contrapositive form, Theorem 5.2.5 says that in any normed vector space that is not a Banach space, there are series that are absolutely convergent but not convergent. This may seem odd since the term 'absolutely convergent' seems to imply a special type of convergence. However, looking closely at the definition, the absolute convergence of a series $\sum x_k$ says that the (different) series $\sum \|x_k\|$ converges, but does not say anything directly about the series $\sum x_k$ itself.

We have not seen many examples of normed spaces that are not Banach spaces. The only major example we have at our disposable is $C_p[a, b]$. One can find explicit examples of series of functions in $C_p[a, b]$ that are absolutely convergent, but not convergent. See, for example, [Coh03, §6.2]. We will give a different example here.

Lemma 5.2.6. *The space of all polynomial functions p on $[0, 1]$ is a normed space with norm*

$$\|p\|_\infty = \max_{0 \le t \le 1} |p(t)|.$$

(Note that this is a vector subspace of $C[0, 1]$ with the induced norm.)

Proof. We leave the prove of this lemma as an exercise (Exercise 5.2.2). □

Example 5.2.7 (An absolutely convergent series that is not convergent). Consider the series

$$\sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Since all the partial sums are polynomials, this is a series in the normed space of Lemma 5.2.6. We have

$$\sum_{n=0}^{\infty} \left\| \frac{x^n}{n!} \right\|_\infty = \sum_{n=0}^{\infty} \frac{1}{n!},$$

which converges (in $\mathbb{R}$) to e. So this series is absolutely convergent. However, it is not convergent in the space of polynomial functions (since it converges to the exponential function). Thus this space of polynomial functions is not a Banach space.

Exercises.

5.2.1 ([Coh03, Ex. 6.4(3)]). (a) Let $\{x_n\}$ and $\{y_n\}$ be convergent sequences in a normed space, with $\lim x_n = x$ and $\lim y_n = y$. Prove that $x_n + y_n \to x + y$.

(b) Let $\{x_n\}$ be a convergent sequence in a normed space, with $\lim x_n = x$, and let $\{\alpha_n\}$ be a convergent sequence of scalars, with $\lim \alpha_n = \alpha$. Prove that $\alpha_n x_n \to \alpha x$.

(c) Let $\{x_n\}$ be a convergent sequence in a normed space, with $\lim x_n = x$. Prove that $\|x_n\| \to \|x\|$. (Thus, $\| \ \|$ is a continuous mapping on a normed space.)

5.2.2. Prove Lemma 5.2.6.

5.2.3 ([Coh03, Ex. 6.4(6)]). Let P be the set of all polynomial functions. Show that

$$\|p\| = |a_0| + |a_1| + \cdots + |a_n|, \quad p(t) = a_0 + a_1 t + \cdots + a_n t^n \in P,$$

defines a norm on P. Show, however, that P is not a Banach space.

5.2.4 ([Coh03, Ex. 6.4(7)]). Let V be a vector space of dimension n and let $\{v_1, \dots, v_n\}$ be a basis for V. Prove that

$$\|x\| = \sum_{k=1}^{n} |\alpha_k|, \quad x = \sum_{k=1}^{n} \alpha_k v_k \in V,$$

defines a norm on V. Furthermore, show that convergence with respect to this norm is equivalent to componentwise convergence.

5.2.5 ([Coh03, Ex. 6.4(8)]). Define a sequence $\{x_n\}$ of functions continuous on $[0,1]$ by

$$x_n(t) = \begin{cases} nt, & 0 \le t < \frac{1}{n}, \\ 1, & \frac{1}{n} \le t \le 1. \end{cases}$$

Show that $\{x_n\}$ is convergent (with limit x, where $x(t) = 1$, $0 \le t \le 1$) when considered as a sequence in $C_1[0,1]$, but not convergent when considered as a sequence in $C[0,1]$.

5.2.6 ([Coh03, Ex. 6.4(11)]). Let c_0 be the vector space of all sequences $(x_1, x_2, \dots)$ of real numbers for which $x_n \to 0$.

(a) Show that c_0 is a Banach space under the norm

$$\|x\| = \max_{k \ge 1} |x_k|, \quad x = (x_1, x_2, \dots) \in c_0.$$

(b) Let $e_1 = (1,0,0,\dots)$, $e_2 = (0,1,0,\dots)$, $e_3 = (0,0,1,\dots)$, ... Prove that the series $\sum_{k=1}^{\infty} e_k/k$ is convergent but not absolutely convergent in c_0. Find the sum of the series.

5.3 Finite-dimensional normed spaces

We will see that finite-dimensional vector spaces are quite well-behaved (from the point of view of norms). First, we show that *every* finite-dimensional vector space (over $\mathbb{R}$ or $\mathbb{C}$) can be given a norm.

For this entire section, let V be an n-dimensional vector space with basis $\{v_1, \dots, v_n\}$. The letter α (with or without a subscript) will always denote a scalar.

Theorem 5.3.1. *The map* $\|\cdot\|_\infty \colon V \to \mathbb{R}_+$ *given by*

$$\|x\|_\infty = \max_{1\le k\le n} |\alpha_k|, \quad \textit{where} \quad x = \sum_{k=1}^{n} \alpha_k v_k \in V,$$

is a norm on V.

Proof. We need to check axioms (N1)–(N3) in the definition of a normed space. Since $\{v_1, \dots, v_n\}$ is a basis of V, $\|\cdot\|_\infty$ is well-defined (since the α_k are uniquely determined). For $x = \sum \alpha_k v_k$, we clearly have

$$x = \mathbf{0} \iff (\alpha_k = 0\ \forall\ k) \iff \|x\|_\infty = 0.$$

So (N1) is satisfied.

Now, for any scalar α, we have

$$\|\alpha x\|_\infty = \left\| \sum_{k=1}^{n} (\alpha\alpha_k) v_k \right\|_\infty = \max_{1\le k\le n} |\alpha\alpha_k| = |\alpha| \max_{1\le k\le n} |\alpha_k| = |\alpha| \cdot \|x\|_\infty.$$

Thus, (N2) is satisfied.

Finally, we must prove the triangle inequality. Let $y = \sum \beta_k v_k$ be a second vector in V. For $k = 1, \dots, n$, we have

$$|\alpha_k + \beta_k| \le |\alpha_k| + |\beta_k| \le \max_{1\le \ell\le n} |\alpha_\ell| + \max_{1\le \ell\le n} |\beta_\ell| = \|x\| + \|y\|.$$

Therefore,

$$\|x + y\|_\infty = \left\| \sum_{k=1}^{n} (\alpha_k + \beta_k) v_k \right\|_\infty = \max_{1\le k\le n} |\alpha_k + \beta_k| \le \|x\| + \|y\|.$$

So (N3) is satisfied. □

Theorem 5.3.2. *Convergence in a finite-dimensional vector space with the norm* $\|\cdot\|_\infty$ *is equivalent to componentwise convergence. In other words, if* $\{x_m\}$ *is a sequence in a finite-dimensional vector space with this norm, and* $x_m = \sum_{k=1}^{n} \alpha_{mk} v_k$, *then*

$$\{x_m\}_{m=1}^{\infty} \text{ converges} \iff \{\alpha_{mk}\}_{m=1}^{\infty} \text{ converges for each } k = 1, 2, \dots, n.$$

(Here the convergence of $\{\alpha_{mk}\}_{m=1}^{\infty}$ *is in* $\mathbb{R}$ *or* $\mathbb{C}$*.)*

Proof. Suppose $x_m \to x$, with $x = \sum_{k=1}^{n} \alpha_k v_k$. Then for all $\epsilon > 0$, there exists an $N > 0$ such that

$$m > N \implies \|x_m - x\|_\infty = \max_{1\le k\le n} |\alpha_{mk} - \alpha_k| < \epsilon.$$

Therefore, for each $k = 1, \dots, n$,

$$m > N \implies |\alpha_{mk} - \alpha_k| < \epsilon.$$

Therefore, $\{\alpha_{mk}\}_{m=1}^{\infty}$ converges (to α_k) for each $k = 1, 2, \ldots, n$.

Now suppose that for each $k = 1, 2, \ldots, n$, the sequence $\{\alpha_{mk}\}_{m=1}^{\infty}$ converges to, say, α_k. Then for all $\epsilon > 0$ and $k = 1, 2, \ldots, n$, there exists an N_k such that

$$m > N_k \implies |\alpha_{mk} - \alpha_k| < \epsilon.$$

Let $N = \max\{N_1, \ldots, N_n\}$ and $x = \sum_{k=1}^{n} \alpha_k v_k$. Then

$$m > N \implies \|x_m - x\|_\infty = \max_{1 \le k \le n} |\alpha_{mk} - \alpha_k| < \epsilon.$$

Thus $\{x_m\}$ converges to x. □

Proposition 5.3.3. *The subset*

$$\{x \in V \mid \|x\|_\infty \le 1\} = B_V^{\text{cl}}(0, 1) \subseteq V$$

is compact.

Proof. We assume V is a complex vector space (the case of a real vector space is almost identical). We prove the result by induction on the dimension n of V. First suppose that $n = 1$ and that $\{v\}$ is a basis of V. Let

$$Z = \{\alpha \in \mathbb{C} \mid |\alpha| \le 1\} = B_{\mathbb{C}}^{\text{cl}}(0, 1),$$

which we know is compact in $\mathbb{C}$ (it is a closed and bounded subset of $\mathbb{C}$, which is the same as $\mathbb{R}^2$ topologically). Define

$$A \colon Z \to V, \quad A\alpha = \alpha v.$$

Then $A(Z) = B_V^{\text{cl}}(0, 1)$. Since the continuous image of a compact set is compact (Theorem 4.4.7), it suffices to show that A is continuous. Suppose $\{\alpha_m\}$ is a sequence in Z converging to $\alpha \in Z$. Then

$$\|A\alpha_m - A\alpha\|_\infty = \|\alpha_m v - \alpha v\|_\infty = \|(\alpha_m - \alpha)v\|_\infty = |\alpha_m - \alpha| \to 0.$$

Hence $A\alpha_m \to A\alpha$ and so A is continuous.

Now assume the proposition is true for $n < h - 1$ for some $h > 1$. We want to show that it is true for $n = h$. Let

$$B_i = B_V^{\text{cl}}(0, 1)$$

when $n = i$, $i \in \mathbb{N}_+$. So our inductive assumption is that B_n is compact for $n < h - 1$ and we wish to show that B_h is compact.

Let $\{x_m\}$ be a sequence in B_h and let

$$x_m = \sum_{j=1}^{h} \alpha_{mj} v_j$$

for $m \in \mathbb{N}_+$. The sequence $\{\alpha_{m1} v_1\}_{m=1}^{\infty}$ is a sequence in B_1, which is compact. Therefore, it has a convergent subsequence $\{\alpha_{m_k 1} v_1\}_{k=1}^{\infty}$. So

$$\{x_{m_k}\}_{k=1}^{\infty}$$

is a subsequence of $\{x_m\}$ such that the sequence of coefficients of v_1 converges. Then

$$\left\{ \sum_{j=2}^{h} \alpha_{m_k,j} v_j \right\}_{k=1}^{\infty} \tag{5.1}$$

is a sequence in $\text{Span}\{v_2, \dots, v_h\}$, which is a vector space of dimension $h-1$. Since, for $k \in \mathbb{N}_+$,

$$\left\| \sum_{j=2}^{h} \alpha_{m_k j} v_j \right\|_{\infty} = \max_{2 \le j \le h} |\alpha_{m_k j}| \le \max_{1 \le j \le h} |\alpha_{m_k j}| = \|x_{m_k}\|_\infty \le 1,$$

it is a sequence in B_{h-1}. Since B_{h-1} is compact by the inductive hypothesis, the sequence (5.1) has a convergent subsequence. By Theorem 5.3.2, this subsequence converges componentwise. Therefore the corresponding subsequence of $\{x_m\}$ also converges componentwise (since we had already chosen a subsequence in which the first component converged) and hence converges. □

Lemma 5.3.4. *Any norm is a continuous mapping.*

Proof. The proof of this lemma is left as an exercise (Exercise 5.2.1(c)). □

Lemma 5.3.5. *If V is a finite-dimensional vector space, then the set*

$$S = \{x \in V \mid \|x\|_\infty = 1\}$$

is compact.

Proof. The proof of this lemma is left as an exercise (Exercise 5.3.1). □

We have seen that every finite-dimensional vector space can be given a norm. We now show that, in this setting, any two norms are "equivalent". First we need to define what we mean by "equivalent."

Definition 5.3.6 (Equivalent norms)**.** Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on a vector space X are said to be *equivalent* if there exists positive numbers a and b such that

$$a\|x\|_1 \le \|x\|_2 \le b\|x\|_1 \quad \forall\, x \in X.$$

Remark 5.3.7. It follows from Definition 5.3.6 that

$$\frac{1}{b}\|x\|_2 \le \|x\|_1 \le \frac{1}{a}\|x\|_2 \quad \forall\, x \in X.$$

Thus, the definition is symmetric.

Lemma 5.3.8. *Equivalence of norms is an equivalence relation on the set of all norms on a fixed vector space X.*

Proof. The proof of this lemma is left as an exercise (Exercise 5.3.2). □

Lemma 5.3.9. *If $\|\cdot\|_1$ and $\|\cdot\|_2$ are two equivalent norms on X, then a sequence $\{x_n\}$ in X converges with respect to $\|\cdot\|_1$ if and only if it converges with respect to $\|\cdot\|_2$. Similarly, $\{x_n\}$ is Cauchy with respect to $\|\cdot\|_1$ if and only if it is Cauchy with respect to $\|\cdot\|_2$.*

Proof. If $\{x_n\}$ converges with respect to $\|\cdot\|_1$, then there exists $x \in X$ such that $\|x_n - x\|_1 \to 0$. But then

$$\|x_n - x\|_2 \leq b\|x_n - x\|_1 \to 0$$

(where b is some fixed positive number) and so $\{x_n\}$ converges with respect to $\|\cdot\|_2$. The converse holds by the symmetry of the definition of equivalent norms.

The second part of the lemma, concerning Cauchy sequences, is left as an exercise (Exercise 5.3.3). □

Lemma 5.3.10. *Suppose $\|\cdot\|_1$ and $\|\cdot\|_2$ are two equivalent norms on X and Y is a subset of X. Then*

(a) Y is closed with respect to $\|\cdot\|_1$ if and only if it is closed with respect to $\|\cdot\|_2$, and

(b) Y is bounded with respect to $\|\cdot\|_1$ if and only if it is bounded with respect to $\|\cdot\|_2$.

Proof. The proof of this lemma is left as an exercise (Exercise 5.3.4). □

Theorem 5.3.11 (Equivalence of norms on finite-dimensional vector spaces)**.** *Any two norms on a finite-dimensional vector space are equivalent.*

Proof. Suppose V is a finite-dimensional vector space. By Lemma 5.3.8, it suffices to show that any norm on V is equivalent to the norm $\|\cdot\|_\infty$. Let $\|\cdot\|$ be a norm on V. Since the set

$$S = \{x \in V \mid \|x\|_\infty = 1\}$$

is compact by Lemma 5.3.5, we can conclude from Lemma 5.3.4 and Corollary 3.3.2 that $\|\cdot\|$ attains its maximum and minimum on S. In other words, there are points $x_M, x_m \in S$ such that

$$\|x_M\| = \max_{x\in S} \|x\|, \quad \|x_m\| = \min_{x\in S} \|x\|.$$

So

$$\|x_m\| \leq \|x\| \leq \|x_M\| \quad \forall\ x \in S.$$

Since $\|x_m\|_\infty = 1$, we have $x_m \neq 0$, and so $\|x_m\| > 0$.

Now, for any $x \in V$, $x \neq \mathbf{0}$, we have

$$\left\| \frac{1}{\|x\|_\infty} x \right\|_\infty = \frac{1}{\|x\|_\infty}\|x\|_\infty = 1 \implies \frac{1}{\|x\|_\infty} x \in S.$$

Thus, for all $x \neq \mathbf{0}$, we have

$$\|x_m\| \leq \left\| \frac{1}{\|x\|_\infty} x \right\| \leq \|x_M\|,$$

and so

$$\|x_m\| \cdot \|x\|_\infty \leq \|x\| \leq \|x_M\| \cdot \|x\|_\infty.$$

Since this is also clearly true when $x = \mathbf{0}$, we see that the norms $\|\cdot\|$ and $\|\cdot\|_\infty$ are equivalent. □

Remark 5.3.12. Since we have shown (Theorem 5.3.2) that convergence with respect to the norm $\|\cdot\|_\infty$ is equivalent to componentwise convergence, we see from Theorem 5.3.11 that convergence with respect to *any* norm on a finite-dimensional vector space is equivalent to componentwise convergence. Note, however, that this is *not* true for infinite-dimensional normed spaces. See, for example, Exercise 5.2.5.

Theorem 5.3.13. *Every finite-dimensional normed space is a Banach space.*

Proof. Let V be a finite-dimensional vector space. By Lemma 5.3.9 and Theorem 5.3.11, it suffices to prove that the normed space $(V, \|\cdot\|_\infty)$ is complete. Let $\{x_n\}$ be a Cauchy sequence in this space and write $x_m = \sum_{k=1}^n \alpha_{mk} v_k$ for $m \in \mathbb{N}_+$. Fix $\epsilon > 0$. Then we can choose $N > 0$ such that

$$j, m > N \implies \|x_m - x_j\|_\infty = \max_{1 \le k \le n} |\alpha_{mk} - \alpha_{jk}| < \epsilon.$$

Then for all $k = 1, 2, \ldots, n$, we have

$$j, m > N \implies |\alpha_{mk} - \alpha_j k| < \epsilon,$$

and so $\{\alpha_{mk}\}_{m=1}^\infty$ is a Cauchy sequence in $\mathbb{C}$. Since $\mathbb{C}$ is complete, each of these sequences converges. Thus, by Theorem 5.3.2, the sequence $\{x_m\}$ converges. □

You learned in previous analysis courses that a subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded (Theorem 3.1.9). We can now extend this result to arbitrary finite-dimensional normed spaces.

Theorem 5.3.14. *A subset of a finite-dimensional normed space is compact if and only if it is closed and bounded.*

Proof. Let $(V, \|\cdot\|)$ be a finite-dimensional normed space. We already know that any compact subset is closed and bounded (Theorems 3.1.5 and 3.1.7). It remains to show that any closed and bounded subset is also compact. We first do this for the norm $\|\cdot\|_\infty$. Let S be a closed and bounded subset of V. The proof of Proposition 5.3.3 can easily be modified to show that

$$B^{(L)} = B_V^{\text{cl}}(0, L) = \{x \in V \mid \|x\|_\infty \le L\}$$

is compact for any positive number L. Since S is bounded, there exists an L such that $S \subseteq B^{(L)}$. Since S is closed, it is compact by Lemma 3.1.11.

Now let $\|\cdot\|$ be any norm on V and let S be a closed and bounded subset of V with respect to this norm. By Lemma 5.3.10, S is also closed and bounded with respect to $\|\cdot\|_\infty$ and thus, by the above, is compact with respect to $\|\cdot\|_\infty$. Now let $\{x_m\}$ be a sequence in S. Then there is a subsequence $\{x_{m_k}\}$ that converges with respect to $\|\cdot\|_\infty$. By Lemma 5.3.9, this subsequence also converges with respect to $\|\cdot\|$. Hence S is compact with respect to $\|\cdot\|$. □

Exercises.

5.3.1. Prove Lemma 5.3.5.

5.3.2. Prove Lemma 5.3.8.

5.3.3. Complete the proof of Lemma 5.3.9.

5.3.4. Prove Lemma 5.3.10.

5.3.5. Prove that, for any norm on a finite-dimensional vector space, convergence of a sequence is equivalent to convergence of the sequences of coefficients (with respect to some basis).

5.4 Approximation theory

By Theorem 3.3.3, if S is a nonempty compact subset of a normed space X, then there is a best approximation of any point of X by a point of S. More precisely, for every $x \in X$, there is a $p \in S$ such that $\|p - x\|$ is a minimum.

Theorem 5.4.1. *Suppose S is a finite-dimensional subspace of a normed space X and $x \in X$. Then there is a point $p \in S$ such that $\|x - p\|$ is a minimum.*

Proof. Fix $p_0 \in S$ and let

$$Y = \{y \in S \mid \|y - x\| \leq \|p_0 - x\|\}.$$

Any point $p \in S$ minimizing $\|x - p\|$ certainly lies in Y. Hence it suffices to show, by Theorem 3.3.3, that Y is compact. Since S is finite-dimensional, any closed and bounded subset of S is compact. So we will show that Y is a closed and bounded subset of S.

For all $y \in Y$, we have

$$\|y\| = \|(y - x) + x\| \leq \|y - x\| + \|x\| \leq \|p_0 - x\| + \|x\|.$$

Thus Y is bounded. Now suppose $\{y_n\}$ is a sequence in Y that converges to a point $y \in S$. For any $\epsilon > 0$, we can find a large enough value of n such that

$$\|y - x\| \leq \|y - y_n\| + \|y_n - x\| < \epsilon + \|p_0 - x\|.$$

Since $\epsilon > 0$ is arbitrary, this implies that

$$\|y - x\| \leq \|p_0 - x\|.$$

Therefore $y \in Y$ and so Y is closed. □

Definition 5.4.2. Let $\mathcal{P}_n([a, b])$ be the set of polynomial functions on $[a, b]$ with degree at most n.

Since $\mathcal{P}_n([a, b])$ has basis $\{1, t, t^2, \ldots, t^n\}$, it is a finite-dimensional subspace of $C[a, b]$. By Theorem 5.4.1, for any function $f \in C[a, b]$, there is a polynomial function $p \in \mathcal{P}_n([a, b])$ that best approximates f, that is, such that

$$\|p - f\| = \max_{a \leq t \leq b} |p(t) - f(t)|$$

is a minimum.

Remark 5.4.3. (a) We will see in the Weierstrass Approximation Theorem (Theorem 5.5.5) that if we allow the polynomials to have arbitrarily large degree, we can approximate any element of $C[a,b]$ as closely as we like.

(b) Theorems 3.3.3 and 5.4.1 guarantee a best approximation but do not tell us if such a best approximation is unique, nor do they tell us how to find this best approximation.

We now see an additional condition which will ensure that best approximations are unique.

Definition 5.4.4 (Strictly convex)**.** A normed space X is *strictly convex* if, for $x, y \in X$, $x, y \neq \mathbf{0}$,

$$\|x+y\| = \|x\| + \|y\| \iff x = \beta y \text{ for some } \beta \in \mathbb{R},\ \beta > 0.$$

Remark 5.4.5. In any normed space X, the triangle inequality ensures that

$$\|x+y\| \leq \|x\| + \|y\| \ \forall\ x, y \in X.$$

If $x = \beta y$ for some $\beta \in \mathbb{R}$, $\beta > 0$, then

$$\|x+y\| = \|(\beta+1)y\| = (\beta+1)\|y\| = \beta\|y\| + \|y\| = \|\beta y\| + \|y\| = \|x\| + \|y\|.$$

However, it is possible, in some normed spaces, to have $x, y \in X$ satisfying $\|x+y\| = \|x\|+\|y\|$ but for which $x \neq \beta y$ for any positive real number β. For instance, this is possible in $C[a,b]$.

Theorem 5.4.6. *Suppose X is a strictly convex normed space and S is a finite-dimensional subspace of X. Then every point of X has a unique best approximation by a point of S.*

Proof. By Theorem 5.4.1, a best approximation exists. It remains to show it is unique. Let $x \in X$. If $x \in S$, then clearly x is its own unique best approximation (since $\|x-p\| > 0$ for any $p \neq x$). Thus we assume $x \notin S$. By Theorem 5.4.1, we can find a point $p \in S$ such that $\|x-p\|$ is a minimum. Suppose $p' \in S$ is another best approximation. Let

$$d = \|x-p\| = \|x-p'\|.$$

Since S is a subspace, we have $\frac{1}{2}(p+p) \in S$. Thus

$$d \leq \left\| x - \frac{1}{2}(p+p') \right\| = \left\| \frac{1}{2}(x-p) + \frac{1}{2}(x-p') \right\| \leq \frac{1}{2}\|x-p\| + \frac{1}{2}\|x-p'\| = d.$$

Therefore

$$\left\| x - \frac{1}{2}(p+p') \right\| = d = \frac{1}{2}\|x-p\| + \frac{1}{2}\left\|x-p'\right\|.$$

Since X is strictly convex, this implies that, for some $\beta > 0$,

$$x - p = \beta(x-p') \implies x(1-\beta) = p - \beta p'.$$

If $\beta \neq 1$, then

$$x = \frac{1}{1-\beta}p - \frac{\beta}{1-\beta}p'.$$

But this contradicts the fact that $p, p' \in S$ and $x \notin S$ (since S is a subspace). Hence, $\beta = 1$. This implies that

$$x - p = x - p' \implies p = p'$$

□

Example 5.4.7 ($C[a,b]$ is not strictly convex). The space $C[a,b]$ is not strictly convex. For instance, if $b > |a|$, let

$$f(t) = b^2, \quad g(t) = tb, \quad a \le t \le b.$$

Then $\|f\| = \|g\| = b^2$, but $\|f+g\| = 2b^2 = \|f\| + \|g\|$. But clearly $f \neq \beta g$ for any $\beta \in \mathbb{R}$.

One can show that the normed space $C_2[a,b]$ is strictly convex (see Exercise 5.4.1). Therefore, for any $f \in C_2[a,b]$ and positive integer n, there exists a unique polynomial function p of degree at most n such that

$$\|f - p\| = \sqrt{\int_a^b (f(x) - p(x))^2 \, dx}$$

is a minimum. This p is called the *best least squares polynomial approximation* of f.

Exercises.

5.4.1 ([Coh03, Ex. 6.10(5)]). Show that the normed space $C_2[a,b]$ is strictly convex. *Hint*: Use Exercise 1.1.14.

5.5 Weierstrass approximation

Our goal is to approximate any continuous function on a closed interval, say $[0,1]$, by a polynomial function. We will see that it is possible to do this with arbitrary precision (in the uniform norm).

We first extend the definition of uniform continuity to a more general setting.

Definition 5.5.1 (Uniformly continuous)**.** Suppose X and Y are normed spaces and S is a subset of X. A mapping $A\colon S \to Y$ is said to be *uniformly continuous* on S if, for all $\epsilon > 0$, there exists a $\delta > 0$ such that

$$x', x'' \in S, \ \|x' - x''\| < \delta \implies \|Ax' - Ax''\| < \epsilon.$$

It is immediate from the ϵ-δ definition of continuity (Definition 4.6.1) that if a mapping is uniformly continuous on its entire domain, then it is continuous.

Theorem 5.5.2. *A continuous mapping on a compact set is uniformly continuous. In other words, if $A\colon X \to Y$ is a mapping between normed spaces that is continuous on a nonempty compact subset S of X, then A is uniformly continuous on S.*

Proof. Fix $\epsilon > 0$. For each $x \in S$, we can choose a δ_x such that

$$x' \in S, \ \|x - x'\| < \delta_x \implies \|Ax - Ax'\| < \epsilon/2.$$

Let

$$B_x = B_S(x, \delta_x/2) = \{x' \in S \mid \|x - x'\| < \delta_x/2\}.$$

Since $x \in B_x$, we have that $\{B_x\}_{x \in S}$ is an open cover of S. Since S is compact, there is some finite set of points $\{x_1, \ldots, x_n\}$ such that $\bigcup_{i=1}^n B_{x_i} = S$. Now take $\delta = \min_{i=1}^n \delta_{x_i}/2$ and suppose y, y' are two arbitrary ponts of S such that $\|y - y'\| < \delta$. Then we must have $y \in B_{x_j}$ for some $j = 1, \ldots, n$. Thus, we also have

$$\|x_j - y'\| \le \|x_j - y\| + \|y - y'\| < \frac{\delta_{x_j}}{2} + \delta \le \delta_{x_j}.$$

Therefore,

$$\|y - x_j\| < \frac{\delta_{x_j}}{2} < \delta_{x_j} \quad \text{and} \quad \|x_j - y'\| < \delta_{x_j}.$$

So

$$\|Ay - Ay'\| \le \|Ay - Ax_j\| + \|Ax_j - Ay'\| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$ □

Remark 5.5.3. See [Coh03, Th. 6.8.2] for an alternative proof using sequential compactness.

Recall that the binomial theorem states

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k.$$

Lemma 5.5.4. *We have*

(a) $\sum_{k=0}^n \binom{n}{k} x^k (1 - x)^{n-k} = 1.$

(b) $\sum_{k=0}^n (k - nx)^2 \binom{n}{k} x^k (1 - x)^{n-k} = nx(1 - x).$

Proof. The first equation is obtained by setting $a = 1 - x$ and $b = x$ in the binomial theorem. The second follows from manipulating the first. We will not discuss the details. The proof can be found, for example, in [Coh03, p. 201]. □

Theorem 5.5.5 (Weierstrass Approximation Theorem)**.** *The polynomial functions are dense in $C[a, b]$. In other words, for any $f \in C[a, b]$ and $\epsilon > 0$, there exists a polynomial function p such that $\|p - f\| < \epsilon$.*

Proof. In order to simplify notation, we will consider only the case $a = 0$, $b = 1$. The general case is analogous. Fix $f \in C[0, 1]$ and $\epsilon > 0$. For $n \in \mathbb{N}_+$, define

$$p_n(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1 - x)^{n-k}.$$

These are called the *Bernstein polynomials* for f. By Lemma 5.5.4(a), we have

$$|f(x) - p_n(x)| = \left| \sum_{k=0}^n \left(f(x) - f\left(\frac{k}{n}\right) \right) \binom{n}{k} x^k (1 - x)^{n-k} \right|$$

$$\leq \sum_{k=0}^{n} \left| f(x) - f\left(\frac{k}{n}\right) \right| \binom{n}{k} x^k (1-x)^{n-k}.$$

Since f is continuous on the interval $[0,1]$, which is compact, f is uniformly continuous on this interval. Therefore, there exists a δ such that

$$x', x'' \in [0,1],\ |x' - x''| < \delta \implies |f(x') - f(x'')| < \epsilon/2.$$

Fix a point $x_0 \in [0,1]$ and partition the set $S = \{0, 1, \ldots, n\}$ into two disjoint sets:

$$S_1 = \left\{ k \;\middle|\; k \in S,\ \left| \frac{k}{n} - x_0 \right| < \delta \right\},$$

$$S_2 = \left\{ k \;\middle|\; k \in S,\ \left| \frac{k}{n} - x_0 \right| \geq \delta \right\}.$$

Then $|f(x_0) - f(k/n)| < \epsilon/2$ when $k \in S_1$ and

$$\left| f(x_0) - f\left(\frac{k}{n}\right) \right| \leq |f(x_0)| + \left| f\left(\frac{k}{n}\right) \right| \leq 2 \max_{0 \leq x \leq 1} |f(x)| = 2\|f\|.$$

Therefore

$$\begin{aligned} |f(x_0) - p_n(x_0)| &\leq \sum_{k \in S_1} \left| f(x) - f\left(\frac{k}{n}\right) \right| \binom{n}{k} x^k (1-x)^{n-k} + \sum_{k \in S_2} \left| f(x) - f\left(\frac{k}{n}\right) \right| \binom{n}{k} x^k (1-x)^{n-k} \\ &< \frac{\epsilon}{2} \sum_{k \in S_1} \binom{n}{k} x_0^k (1-x_0)^{n-k} + 2\|f\| \sum_{k \in S_2} \binom{n}{k} x_0^k (1-x_0)^{n-k}. \end{aligned}$$

By Lemma 5.5.4(a), we have

$$\sum_{k \in S_1} \binom{n}{k} x_0^k (1-x_0)^{n-k} \leq \sum_{k=0}^{n} \binom{n}{k} x_0^k (1-x_0)^{n-k} = 1,$$

and by Lemma 5.5.4(b), we have

$$\begin{aligned} n x_0 (1-x_0) &= \sum_{k=0}^{n} (k - n x_0)^2 \binom{n}{k} x_0^k (1-x_0)^{n-k} \\ &= n^2 \sum_{k=0}^{n} \binom{n}{k} \left(\frac{k}{n} - x_0 \right)^2 x_0^k (1-x_0)^{n-k} \\ &\geq n^2 \sum_{k \in S_2} \binom{n}{k} \left(\frac{k}{n} - x_0 \right)^2 x_0^k (1-x_0)^{n-k} \\ &\geq n^2 \delta^2 \sum_{k \in S_2} \binom{n}{k} x_0^k (1-x_0)^{n-k}. \end{aligned}$$

Now,

$$x_0(1-x_0) = \frac{1}{4} - \left(x_0 - \frac{1}{2}\right)^2 \le \frac{1}{4},$$

and so

$$\sum_{k\in S_2} \binom{n}{k} x_0^k(1-x_0)^{n-k} \le \frac{x_0(1-x_0)}{n\delta^2} \le \frac{1}{4n\delta^2}.$$

Therefore,

$$|f(x_0) - p_n(x_0)| < \frac{\epsilon}{2} + \frac{\|f\|}{2n\delta^2} \quad \forall\, x_0 \in [0,1],\ n \in \mathbb{N}_+.$$

Now take $n > \|f\|/\epsilon\delta^2$, so $\|f\|/n\delta^2 < \epsilon$. Then

$$|f(x_0) - p_n(x_0)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Since this holds for all $x_0 \in [0,1]$, we have

$$\|f - p_n\| = \max_{0\le x\le 1} |f(x) - p_n(x)| < \epsilon. \qquad \square$$

Remark 5.5.6. Note that the proof of Theorem 5.5.5 even tells us how to find the polynomial that approximates the given function—it is not an abstract existence proof.

Theorem 5.5.7 ($C[a,b]$ is separable)**.** *The metric space $C[a,b]$ is separable. More precisely, the set of polynomials functions with rational coefficients is dense in $C[a,b]$. In other words, for any $f \in C[a,b]$ and any $\epsilon > 0$, there exists a polynomial function p with rational coefficients such that $\|p - f\| < \epsilon$.*

Proof. Again, we will consider the case $a = 0$, $b = 1$. Suppose $f \in C[0,1]$ and $\epsilon > 0$. By The Weierstrass Approximation Theorem (Theorem 5.5.5), we can find a polynomial function q such that $\|q - f\| < \epsilon/2$. The only problem is that this polynomial function may have some irrational coefficients. So we will approximate it by one that has only rational coefficients. Suppose

$$q(x) = \sum_{k=0}^{r} a_k x^k, \quad 0 \le x \le 1.$$

Since the rationals are dense in the reals, for each $k = 0, \dots, r$, we choose a rational number b_k such that

$$|b_k - a_k| \le \frac{\epsilon}{2(r+1)}.$$

Define

$$p(x) = \sum_{k=0}^{r} b_k x_k, \quad 0 \le x \le 1.$$

Then for all $x \in [0,1]$, we have

$$|p(x) - q(x)| = \left|\sum_{k=0}^{r} (b_k - a_k)x^k\right| \le \sum_{k=0}^{r} |b_k - a_k| \cdot |x|^k \le \sum_{k=0}^{r} \frac{\epsilon}{2(r+1)} = \frac{\epsilon}{2}.$$

Thus

$$\|p - f\| \le \|p - q\| + \|q - f\| < \epsilon. \qquad \square$$

Exercises.

5.5.1. Suppose X and Y are normed spaces and $A\colon X \to Y$ is a mapping that is uniformly continuous on X. Show that if $\{x_n\}$ is a Cauchy sequence in X, then $\{Ax_n\}$ is a Cauchy sequence in Y.

5.5.2. Prove that contraction mappings are uniformly continuous.

5.5.3 ([Coh03, Ex. 6.10(14)]). Generalize the Weierstrass Theorem to show that, given $\epsilon > 0$, for any function $f \in C[a,b]$, there is a polynomial function p such that $\|p - f\| < \epsilon$. *Hint*: Define a function g by $g(y) = f(a + (b-a)y)$. Then $g \in C[0,1]$, so there is a polynomial function q such that $\|q - g\| < \epsilon$. Set $p(x) = q\big((x-a)/(b-a)\big)$, so p is a polynomial function with the desired property.

5.5.4 ([Coh03, Ex. 6.10(14)]). Let $f \in C^{(1)}[a,b]$, which is the space of all differentiable functions defined on $[a,b]$, with the uniform norm. Show that, if $\epsilon > 0$ is given and p is a polynomial function such that $\|p - f'\| < \epsilon$, then $\|q - f\| < \epsilon(b-a)$, where q is the polynomial function defined by $q(x) = \int_a^x p(t)\,dt + f(a)$.

Chapter 6

Mappings on normed vector spaces

Having introduced normed vector spaces in Chapter 5, we now turn our attention to mappings on such spaces.

6.1 Bounded linear mappings

Since normed spaces have the structure of a vector space (unlike metric spaces in general), we can talk of *linear* maps between normed spaces. We will see that these maps have several nice properties.

Theorem 6.1.1. *For a linear map $A\colon X \to Y$ between two normed spaces, the following conditions are equivalent.*

(a) A is continuous.

(b) The restriction of A to the closed unit ball $B = B_X^{\mathrm{cl}}(0,1)$ around zero is continuous.

(c) A is continuous at zero.

(d) The restriction of A to the ball B is bounded.

(e) There exists a constant $K \geq 0$ such that

$$\|Ax\|_Y \leq K \cdot \|x\|_X, \quad \forall\, x \in X.$$

(f) A is Lipschitz.

(g) A is uniformly continuous.

Proof. (a) $\Rightarrow$ (b): This follows from the fact that the restriction of a continuous function to a metric subspace is always continuous.

(b) $\Rightarrow$ (c): This follows from the fact that 0 is an interior point of B.

(c) $\Rightarrow$ (d): If A is continuous at zero, then we can choose $\delta > 0$ such that

$$\|x - \mathbf{0}\| = \|x\| < \delta \implies \|Ax - A\mathbf{0}\| = \|Ax\| < 1.$$

Let $K = 2/\delta$. Then for all $x \in B$, we have

$$\|Ax\| = \left\| A\left(\frac{2}{\delta}\frac{\delta x}{2}\right)\right\| = \frac{2}{\delta}\left\| A\left(\frac{\delta x}{2}\right)\right\| < \frac{2}{\delta} \cdot 1 = K,$$

where the last inequality holds since $\|\delta x/2\| = (\delta/2)\|x\| \le \delta/2 < \delta$. Thus A is bounded on B.

(d) $\Rightarrow$ (e): Suppose the restriction of A to B is bounded. Then there exists a $K \ge 0$ such that for all $x \in B$, we have $\|Ax\| \le K$. If $x = \mathbf{0}$, then $Ax = \mathbf{0}$ and the conclusion follows. So suppose $x \neq \mathbf{0}$. Then $\|x\| \neq 0$. Since $\|\, x/\|x\|\, \| = 1$, we have $x/\|x\| \in B$ and so

$$\|Ax\| = \left\| A\left(\|x\| \cdot \frac{x}{\|x\|}\right)\right\| = \|x\| \cdot \left\| A\left(\frac{x}{\|x\|}\right)\right\| < \|x\| \cdot K,$$

as desired.

(e) $\Rightarrow$ (f): Fix a $K \ge 0$ as in condition (e). Then for all $x, y \in X$, we have

$$d_Y(Ax, Ay) = \|Ax - Ay\|_Y = \|A(x-y)\|_Y \le K \cdot \|x - y\|_X = K \cdot d_X(x, y).$$

Thus A is Lipschitz with Lipschitz constant K.

(f) $\Rightarrow$ (g): Every Lipschitz function between two metric spaces is uniformly continuous (see Exercise 4.6.2).

(g) $\Rightarrow$ (a): Every uniformly continuous mapping is continuous. □

Definition 6.1.2 (Bounded linear operator)**.** A linear mapping between two normed spaces is called a *bounded (linear) operator* if is satisfies any one of the equivalent conditions of Theorem 6.1.1.

Remark 6.1.3. The text [Coh03] uses the term *operator* to mean bounded linear mapping.

Definition 6.1.4 (Norm of a bounded operator)**.** Suppose A is a bounded operator between two normed spaces. The number

$$\inf\{K \ge 0 \mid \|Ax\| \le K\|x\|,\ x \in X\}$$

is called the *norm* (or *operator norm*) of A, and is denoted $\|A\|$.

Example 6.1.5. Suppose X is a normed space and $A\colon X \to X$ is the mapping defined by $Ax = \beta x$, $x \in X$, for some fixed scalar β. This is clearly a linear map and it is bounded since

$$\|Ax\| = \|\beta x\| = |\beta|\, \|x\|, \quad \forall\, x \in X.$$

If X is not the zero vector space, then $\|A\| = |\beta|$.

Example 6.1.6. Let $X = C[0,1]$ (with the uniform norm) and $Y = \mathbb{R}$ (with the usual norm). Define the *evaluation functional*, or *Dirac delta functional*, $\delta\colon C[0,1] \to \mathbb{R}$, by

$$\delta(g) = g(0), \quad \forall\, g \in C[0,1].$$

It is easy to check that δ is linear. We claim that it is bounded, with norm one. Indeed,

$$|\delta(g)| = |g(0)| \le \max_{0\le t\le 1} |g(t)| = \|g\| = 1 \cdot \|g\|.$$

So δ is bounded and $\|\delta\| \le 1$. Taking the function $g(x) \equiv 1$, we have

$$\|\delta(g)\| = |g(0)| = 1 = 1 \cdot \|g\|,$$

and so $\|\delta\| \ge 1$. Thus $\|\delta\| = 1$.

Example 6.1.7. Again, take $X = C[0,1]$ and $Y = \mathbb{R}$. Define $T\colon C[0,1] \to \mathbb{R}$ by

$$Tg = \int_0^1 g(x)\,dx - g(0).$$

It is easy to check that T is linear. Now

$$\begin{aligned} |Tg| = \left|\int_0^1 g(x)\,dx - g(0)\right| &\le \left|\int_0^1 g(x)\,dx\right| + |g(0)| \\ &\le \int_0^1 |g(x)|\,dx + |g(0)| \le 2\max_{0\le t\le 1}|g(t)| = 2\|g\|. \end{aligned}$$

Thus T is bounded and $\|T\| \le 2$. We claim that $\|T\| = 2$. For any $0 < \epsilon < 1$, define $g_\epsilon \in C[0,1]$ by

$$g_\epsilon(x) = \begin{cases} -1 + 2x/\epsilon, & 0 \le x \le \epsilon, \\ 1, & \epsilon < x \le 1. \end{cases}$$

Then $\|g\| = 1$ and

$$|T(g)| = \left|\int_0^1 g(x)\,dx - g(0)\right| = 2 - \epsilon = (2-\epsilon)\|g\|.$$

Since we can take ϵ as close to zero as we like, this shows that $\|T\| \ge 2$. Hence $\|T\| = 2$.

Theorem 6.1.8. *Let $A\colon X \to Y$ be a bounded operator between normed spaces. Then*

$$\begin{aligned} \|A\| &= \sup\left\{ \frac{\|Ax\|}{\|x\|} \;\middle|\; x \in X,\ x \ne \mathbf{0} \right\} && (= b) \\ &= \sup\{\|Ax\| \mid x \in X,\ \|x\| = 1\} && (= c) \\ &= \sup\{\|Ax\| \mid x \in X,\ \|x\| \le 1\} && (= d). \end{aligned}$$

Proof. We will show $\|A\| \le b \le c \le d \le \|A\|$. For $x \in X$, $x \ne \mathbf{0}$, we have

$$b \ge \frac{\|Ax\|}{\|x\|} \implies \|Ax\| \le b\|x\|,$$

which is also true when $x = \mathbf{0}$. Thus $\|A\| \le b$.

Now, for $x \in X$, $x \neq \mathbf{0}$, we have

$$\frac{\|Ax\|}{\|x\|} = \left\| \frac{1}{\|x\|} Ax \right\| = \left\| A\left(\frac{x}{\|x\|}\right) \right\| \leq c,$$

since $x/\|x\|$ has norm 1. Therefore $b \leq c$.

Since

$$\{x \mid x \in X,\ \|x\| = 1\} \subseteq \{x \mid x \in X,\ \|x\| \leq 1\},$$

we have $c \leq d$.

Finally, if $\|x\| \leq 1$, then $\|Ax\| \leq \|A\|\,\|x\| \leq \|A\|$, and so $d \leq \|A\|$. □

Theorem 6.1.9. *Suppose X and Y are normed spaces. The set $B(X,Y)$ of all bounded operators from X to Y is a vector space under the usual addition and scalar multiplication for linear maps.*

Proof. We know from linear algebra that the set of all linear maps between two fixed vector spaces is itself a vector space. It remains to prove that a linear combination of *bounded* operators is bounded. Suppose $A, B \colon X \to Y$ are bounded operators and α, β are scalars. Then, for all $x \in X$, we have

$$\begin{aligned}
\|(\alpha A + \beta B)x\| &= \|\alpha Ax + \beta Bx\| \\
&\leq \|\alpha Ax\| + \|\beta Bx\| \\
&= |\alpha|\,\|Ax\| + |\beta|\,\|Bx\| \\
&\leq |\alpha|\,\|A\|\,\|x\| + |\beta|\,\|B\|\,\|x\| \\
&= (|\alpha|\,\|A\| + |\beta|\,\|B\|)\|x\|
\end{aligned}$$

Thus $\alpha A + \beta B$ is bounded, as required. □

Theorem 6.1.10. *Suppose X and Y are normed spaces. Then $B(X,Y)$, with the operator norm, is a normed space.*

Proof. If A is the zero map, then $\|Ax\| = 0 = 0 \cdot \|x\|$ for all $x \in X$ and so $\|A\| = 0$. On the other hand, if $\|A\| = 0$, then $\|Ax\| \leq \|A\|\,\|x\| = 0$ for all $x \in X$. Then $Ax = \mathbf{0}$ for all $x \in X$ (by the property (N1) of the norm on Y). Thus A is the zero map. Hence the operator norm satisfies property (N1).

Now suppose $A \in B(X,Y)$ and α is a scalar. To prove (N2), we need to show $\|\alpha A\| = |\alpha|\,\|A\|$. For all $x \in X$,

$$\|(\alpha A)x\| = \|\alpha Ax\| = |\alpha|\,\|Ax\| \leq (|\alpha|\,\|A\|)\,\|x\|.$$

Thus $\|\alpha A\| \leq |\alpha|\,\|A\|$. It remains to show that $\|\alpha A\| \geq |\alpha|\,\|A\|$. We have

$$\|Ax\| = \|(\alpha^{-1}\alpha)Ax\| = \|\alpha^{-1}(\alpha A)x\| = |\alpha|^{-1}\|(\alpha A)x\| \leq |\alpha|^{-1}\|\alpha A\|\,\|x\|.$$

Therefore

$$\|A\| \leq |\alpha|^{-1}\|\alpha A\| \implies \|\alpha A\| \geq |\alpha|\,\|A\|,$$

as desired.

Finally, we must show that (N3) is satisfied. Suppose $A_1, A_2 \in B(X,Y)$. Then, for all $x \in X$,

$$\begin{aligned}\|(A_1 + A_2)x\| &= \|A_1 x + A_2 x\| \le \|A_1 x\| + \|A_2 x\| \\ &\le \|A_1\|\,\|x\| + \|A_2\|\,\|x\| = (\|A_1\| + \|A_2\|)\,\|x\|.\end{aligned}$$

Hence $\|A_1 + A_2\| \le \|A_1\| + \|A_2\|$, which is (N3). □

Theorem 6.1.11. *If Y is a Banach space, then $B(X,Y)$ is a Banach space.*

Proof. Assume Y is a Banach space. We need to show that $B(X,Y)$ is a complete. Let $\{A_n\}$ be a Cauchy sequence in $B(X,Y)$. Choose $x \in X$, $x \neq \mathbf{0}$, and $\epsilon > 0$. Since $\{A_n\}$ is a Cauchy sequence, we can find $N > 0$ such that

$$n, m > N \implies \|A_n - A_m\| < \epsilon/\|x\|.$$

Then

$$n, m > N \implies \|A_n x - A_m x\| = \|(A_n - A_m)x\| \le \|A_n - A_m\|\,\|x\| < \epsilon.$$

Therefore $\{A_n x\}$ is a Cauchy sequence in Y (this is clearly also true for $x = \mathbf{0}$ since then $A_n x = \mathbf{0}$ for all n). Since Y is complete, this sequence converges to some $y \in Y$. We define $Ax = y$. This defines a map $A\colon X \to Y$. We want to show that $A \in B(X,Y)$ and that $A_n \to A$.

For $x_1, x_2 \in X$ and scalars α_1, α_2, we have

$$\begin{aligned}A(\alpha_1 x_1 + \alpha_2 x_2) &= \lim_{n\to\infty} A_n(\alpha_1 x_1 + \alpha_2 x_2) \\ &= \lim_{n\to\infty} (\alpha_1 A_n x_1 + \alpha_2 A_n x_2) \\ &= \alpha_1 \left(\lim_{n\to\infty} A_n x_1\right) + \alpha_2 \left(\lim_{n\to\infty} A_n x_2\right) \\ &= \alpha_1 A x_1 + \alpha_2 A x_2.\end{aligned}$$

Thus A is linear. Suppose $x \in X$, $x \neq \mathbf{0}$. Since $A_n x \to Ax$, we can choose an $N_1 > 0$ such that

$$n > N_1 \implies \|A_n x - Ax\| < \|x\|.$$

Then, for $n > N_1$, we have

$$\|Ax\| \le \|Ax - A_n x\| + \|A_n x\| < \|x\| + \|A_n\|\,\|x\| = (1 + \|A_n\|)\|x\|.$$

Since $\{A_n\}$ is Cauchy, we can choose an $N_2 > 0$ such that

$$n, m > N_2 \implies \|A_n - A_m\| < 1,$$

and thus

$$n > N_2 \implies \|A_n - A_{N_2+1}\| < 1 \implies \|A_n\| \le \|A_{N_2+1}\| + 1.$$

Therefore, if $N = \max\{N_1, N_2\}$, we have

$$n > N \implies \|Ax\| < (1 + \|A_n\|)\|x\| < (2 + \|A_{N_2+1}\|)\|x\|.$$

Since N_2 does not depend on n or x, we see that A is bounded. So $A \in B(X, Y)$.

It remains to prove that $A_n \to A$ (in the operator norm). Fix $\epsilon > 0$. Since $\{A_n\}$ is Cauchy, we can choose $N_1 > 0$ such that

$$m, n \geq N_1 \implies \|A_n - A_m\| < \epsilon/2.$$

Now, fix $x \in X$. Since $A_n x \to Ax$ by the definition of A, we can choose $N_2 > 0$ such that

$$n \geq N_2 \implies \|Ax - A_n x\| < \frac{\epsilon\|x\|}{2}.$$

Let $k = \max\{N_1, N_2\}$. Then for any $n \geq N_1$, we have

$$\begin{aligned}
\|(A - A_n)x\| &= \|Ax - A_k x\| \\
&\leq \|Ax - A_k x\| + \|A_k x - A_n x\| \\
&\leq \|Ax - A_k x\| + \|(A_k - A_n)x\| \\
&\leq \frac{\epsilon\|x\|}{2} + \|A_k - A_n\| \, \|x\| \\
&\leq \frac{\epsilon\|x\|}{2} + \frac{\epsilon}{2} \, \|x\| \\
&= \epsilon\|x\|.
\end{aligned}$$

Note that the choice of N_1 did not depend on x (only on ϵ). Thus

$$n \geq N_1 \implies \|A - A_n\| < \epsilon,$$

and so $A_n \to A$ in the operator norm. □

Remark 6.1.12. See [Coh03, Th. 7.2.5] for an alternative proof that uses the fact that a normed space is a Banach space if and only if every absolutely convergent sequence is convergent.

Theorem 6.1.13. *Every linear mapping from a finite-dimensional normed space to another normed space is bounded.*

Proof. We will work over the real numbers (the proof in the case of complex vector spaces is analogous). Since any finite-dimensional vector space can be identified with $\mathbb{R}^n$ after choosing a basis, it suffice to prove the result for $\mathbb{R}^n$. We first prove the theorem for the norm $\|\cdot\|_\infty$. Let $\{e_1, \ldots, e_n\}$ denote the standard basis of $\mathbb{R}^n$ and let $A \colon \mathbb{R}^n \to X$ be a linear map, where X is any normed space. Then, for all $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, we have

$$\begin{aligned}
\|Ax\| &= \left\| A \left(\sum_{i=1}^n x_i e_i \right) \right\| \\
&= \left\| \sum_{i=1}^n x_i A e_i \right\| \leq \sum_{i=1}^n |x_i| \, \|Ae_i\| \\
&\leq \sum_{i=1}^n \left(\max_{j=1}^n |x_i| \right) \, \|Ae_i\|
\end{aligned}$$

$$= \|x\|_\infty \left(\sum_{i=1}^n \|Ae_i\| \right).$$

Thus A is bounded and

$$\|A\| \le \sum_{i=1}^n \|Ae_i\|.$$

Now, let $\|\cdot\|$ be an arbitrary norm on $\mathbb{R}^n$. Since we know all norms on a fixed finite-dimensional vector space are equivalent, there exist $a, b \ge 0$ such that

$$a\|x\| \le \|x\|_\infty \le b\|x\|, \quad \forall\, x \in \mathbb{R}^n.$$

Then, for all $x \in \mathbb{R}^n$, we have

$$\|Ax\| \le \left(\sum_{i=1}^n \|Ae_i\| \right) \|x\|_\infty \le b \left(\sum_{i=1}^n \|Ae_i\| \right) \|x\|,$$

and so A is bounded. □

Remark 6.1.14. Note that the operator norm of a bounded linear mapping depends on the norms on the domain and codomain. For instance, in the above example, the norm of A when $\mathbb{R}^n$ is given the norm $\|\cdot\|_\infty$ may be different than the norm of A when $\mathbb{R}^n$ is given some other norm.

Exercises.

6.1.1 ([Coh03, Ex 7.5(2)]). Fix real numbers $a < b$ and let k be a function of two variables, which is continuous on the square $[a, b] \times [a, b]$. Fix a scalar λ, and define a mapping $A \colon C[a, b] \to C[a, b]$ by $Ax = y$, where

$$y(s) = \lambda \int_a^b k(s, t)x(t)\, dt, \quad x \in C[a, b],\ a \le s \le t.$$

Here we consider $C[a, b]$ as a vector space, with the usual uniform norm.

(a) Show that $\|A\| \le |\lambda| M(b-a)$, where M is the maximum value of $|k(s, t)|$ for $a \le s \le b$ and $a \le t \le b$.

(b) Show that the mapping A is still bounded when considered as a mapping from the normed space $C_1[a, b]$ into itself, and from the normed space $C_2[a, b]$ into itself. That is, consider the effects of the different norms. In each chase, show also that the same bound on $\|A\|$ as that above may be obtained.

6.1.2 ([Coh03, Ex 7.5(3)]). Let g be a fixed continuous function on $[a, b]$ and let A be the mapping of $C[a, b]$ into itself defined by $Ax = y$, where $y(t) = g(t)x(t)$, $a \leq t \leq b$. Show that A is an bounded linear mapping. Do the same when A is considered as a mapping from $C_1[a, b]$ into itself.

6.1.3 ([Coh03, Ex 7.5(9)]). Suppose X and Y are normed vector spaces.

(a) Show that
$$\|(x, y)\| = \|x\| + \|y\|, \quad x \in X,\ y \in Y,$$
defines a norm on $X \times Y$.

(b) Show that
$$\|(x, y)\|' = \max\{\|x\|, \|y\|\}, \quad x \in X,\ y \in Y,$$
defines a norm on $X \times Y$.

(c) Show that the two norms on $X \times Y$ defined above are equivalent.

6.1.4 ([Coh03, Ex 7.5(10)]). If X and Y are Banach spaces, show that $X \times Y$ is also a Banach space, under either of the norms on $X \times Y$ defined in Exercise 6.1.3.

6.1.5. Suppose $A\colon X \to X$ is a bounded linear mapping on a normed space X. Show that all eigenvalues λ of A satisfy $|\lambda| \leq \|A\|$. More precisely, show that if there exists $x \in X$, $x \neq \mathbf{0}$, and a scalar λ such that $Ax = \lambda x$, then $|\lambda| \leq \|A\|$.

6.1.6 ([Coh03, Ex. 7.5(14)]). Let A be a linear mapping from a normed space X into a normed space Y. Prove that A is bounded if and only if A maps bounded sets in X into bounded sets in Y.

6.2 Linear functionals and the Hahn–Banach Theorem

Definition 6.2.1. Let K be $\mathbb{R}$ or $\mathbb{C}$ and let X be a normed space over K. A *functional* on X is a mapping $f\colon X \to K$.

Since we have the standard norm on $\mathbb{R}$ and $\mathbb{C}$, functionals are maps between normed spaces. Thus we can speak of functionals that are linear, bounded, continuous, etc. If a functional is bounded, it has a norm (the operator norm).

Recall that $B(X, K)$ is the set of all bounded linear mappings from X to K. Since K is a finite-dimensional normed space, it is a Banach space (Theorem 5.3.13). Hence, by Theorem 6.1.11, $B(X, K)$ is a Banach space.

Definition 6.2.2 (Dual space)**.** If X is a normed space over the field K ($= \mathbb{R}$ or $\mathbb{C}$), then $B(X, K)$ is called the *dual space* of X and is denoted by X^*.

Our above discussion tells us that the dual of a normed vector space is always a Banach space. This is very useful result, although the applications are beyond the scope of this course.

Examples 6.2.3. (a) Fix $(a_1, \dots, a_n) \in \mathbb{C}^n$. Then the map

$$f\colon \mathbb{C}^n \to \mathbb{C}, \quad f(x) = \sum_{k=1}^n a_k x_k, \quad x = (x_1, \dots, x_n) \in \mathbb{C}^n,$$

is a linear functional on $\mathbb{C}^n$.

(b) The map

$$f\colon C[a,b] \to \mathbb{R}, \quad f(x) = \int_a^b x(t)\, dt, \quad x \in C[a,b],$$

is a linear functional on $C[a,b]$.

(c) For a fixed c with $a \le c \le b$,

$$\delta_c(x) = x(c), \quad x \in C[a,b],$$

is a linear functional on $C[a,b]$. This functional is called the *Dirac delta functional*.

(d) If X is a normed space, then

$$f\colon X \to \mathbb{R}, \quad f(x) = \|x\|, \quad x \in X,$$

is a function on X (but is not linear in general).

Proposition 6.2.4. *A linear functional on a normed vector space X is continuous on X if and only if* $\ker f$ *is closed (in X).*

Proof. Suppose f is continuous on X. Suppose $\{x_n\}$ is a sequence in $\ker f$ that converges in X to some point x. We wish to show that $x \in \ker f$. Since f is continuous, we have

$$f(x) = f\left(\lim_{n\to\infty} x_n\right) = \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} 0 = 0.$$

Thus $x \in \ker f$, as desired.

The proof of the converse statement will not be given in class but can be found in [Coh03, Th. 7.3.2]. □

Suppose X is some type of mathematical object (e.g. a set, metric space, vector space, manifold,...). What does it mean to say that X admits "enough" functions of a given type (e.g. continuous functions, bounded operators,...)? One interpretation is that this class of functions *separates points* of X in the sense that for any two distinct points of X, there is a function (of the given type) that "tells them apart" by taking different values at them. More precisely, for all $x, y \in X$, $x \neq y$, there is a function on X of the given type such that $f(x) \neq f(y)$.

Here is an example of this idea.

Proposition 6.2.5. *Lipschitz (real-valued) functions separate points on every metric space.*

Proof. Let X be a metric space, and let $x, y \in X$, $x \neq y$. Define

$$f\colon X \to \mathbb{R}, \quad f(z) = d(x, z).$$

Then f is Lipschitz with Lipschitz constant 1, since for all $z_1, z_2 \in X$, we have

$$d(f(z_1), f(z_2)) = |f(z_1) - f(z_2)| = |d(x, z_1) - d(x, z_2)| \leq d(z_1, z_2)$$

(the inequality follows from the triangle inequality). Since

$$f(x) = d(x, x) = 0 \neq d(x, y) = f(y),$$

we are done. □

Corollary 6.2.6. *Continuous functions separate points on every metric space.*

We now consider the question: Do bounded linear functionals separate points on every normed space?

Example 6.2.7. Bounded linear functionals separate elements of the normed space $C[0, 1]$. To see this, suppose that $f, g \in C[0, 1]$, $f \neq g$. Then $f(x) \neq g(x)$ for some $x \in [0, 1]$. Then

$$\delta_x(f) = f(x) \neq g(x) = \delta_x(g).$$

Example 6.2.8. On finite-dimensional normed spaces, every linear functional is bounded. One can easily see that linear functionals separate points in *any* normed space (see Exercise 6.2.4). Hence bounded linear functionals separate elements of every finite-dimensional normed space.

Definition 6.2.9 (Normed subspace)**.** A *normed subspace* of a normed space X is a vector subspace of X equipped with the restriction of the norm.

Theorem 6.2.10 (Hahn–Banach Theorem)**.** *Suppose Y is a normed subspace of a normed space X and ϕ is a bounded linear functional on Y. Then ϕ can be extended to a bounded linear functional $\tilde{\phi}$ on X (i.e. $\tilde{\phi}|_Y = \phi$) with the property that $\|\tilde{\phi}\| = \|\phi\|$.*

Before turning to the proof of this theorem, let us see how it answers our question.

Corollary 6.2.11. *Bounded linear functionals separate points of every normed space.*

Proof. Suppose X is a normed space over K ($= \mathbb{R}$ or $\mathbb{C}$) and $x, y \in X$, $x \neq y$. Let $z = x - y$ and define

$$Y = \{tz \mid t \in K\}$$

to be the one-dimensional linear space spanned by z. Then define a linear functional ϕ on Y by

$$\phi(tz) = t\|z\|.$$

Since

$$|\phi(tz)| = |t|\,\|z\| = 1 \cdot \|tz\|,$$

ϕ is bounded with norm $\|\phi\| = 1$. By the Hahn–Banach Theorem (Theorem 6.2.10), we can extend ϕ to a linear functional $\tilde{\phi}$ of norm 1 on all of X. Now,

$$\tilde{\phi}(x) - \tilde{\phi}(y) = \tilde{\phi}(z) = \phi(z) = \|z\| \neq 0.$$

Thus $\tilde{\phi}(x) \neq \tilde{\phi}(y)$. □

We will only give a proof of the Hahn–Banach Theorem (Theorem 6.2.10) in the important case where Y has codimension one, that is, $\dim X/Y = 1$. The general case is reduced to this particular case by means of a suitable recursion. Intuitively, we repeat the extension by adding more and more vectors to Y, one at a time, until we arrive at X.

Lemma 6.2.12. *Let Y be a codimension one subsspace of a normed space X, and let $\phi\colon Y \to \mathbb{R}$ be a bounded linear functional. Then there is a bounded linear functional $\tilde{\phi}\colon X \to \mathbb{R}$ with $\|\tilde{\phi}\| = \|\phi\|$, extending ϕ (in the sense that $\tilde{\phi}|_F = \phi$).*

Proof. Choose $x \in X \setminus Y$. Then $X = Y \oplus \mathbb{R}x$ and so every element $w \in X$ can be written uniquely in the form

$$w = y + tx,$$

for some $y \in Y$ and $t \in \mathbb{R}$.

Define (nonlinear) maps $\psi_1, \psi_2 \colon Y \to \mathbb{R}$ by

$$\begin{aligned} \psi_1(y) &= \|\phi\| \, \|x + y\| - \phi(y), \quad y \in Y, \\ \psi_2(y) &= -\|\phi\| \, \|x + y\| - \phi(y), \quad y \in Y. \end{aligned}$$

For all $y, z \in Y$, we have

$$\begin{aligned} \psi_1(y) - \psi_2(z) &= \|\phi\| \left(\|x + y\| + \| - x - z\|\right) - \phi(y - z) \\ &\geq \|\psi\| \, \|x + y - x - z\| - \phi(y - z) \qquad \text{(by (N3))} \\ &= \|\psi\| \, \|y - z\| - \phi(y - z) \\ &\geq 0 \quad \text{(by the definition of the operator norm).} \end{aligned}$$

Therefore

$$\inf_{y \in Y} \psi_1(y) \geq \sup_{z \in Y} \psi_2(z).$$

Thus we can choose an $a \in \mathbb{R}$ such that

$$\psi_1(y) \geq a \geq \psi_2(y) \quad \forall\, y \in Y.$$

We then define $\tilde{\phi}$ by

$$\tilde{\phi}(y + tx) = \tilde{\phi}(y) + t\tilde{\phi}(x) = \phi(y) + ta.$$

This map is clearly linear and extends ϕ. It remains to find its norm.

By the definition of the operator norm, there exists $y \in Y$ such that

$$|\phi(y)| = \|\phi\| \, \|y\|.$$

Thus

$$\|\tilde{\phi}(y)\| = \|\phi(y)\| = \|\phi\| \, \|y\|,$$

and so $\|\tilde{\phi}\| \geq \|\phi\|$. So we only need to show that $\|\tilde{\phi}\| \leq \|\phi\|$. In other words, we want to show that

$$|\tilde{\phi}(w)| \leq \|\phi\| \, \|w\|, \quad \forall\, w \in X.$$

Let $w \in X$ be an arbitrary element and write $w = y + tx$ for a unique $y \in Y$ and $t \in \mathbb{R}$. If $t = 0$, then

$$\tilde{\phi}(w) = \phi(y) \leq \|\phi\| \, \|y\| = \|\phi\| \, \|w\|,$$

and we are done. So assume $t \neq 0$. Note that the choice of a implies that for all $y \in Y$, we have

$$-\|\phi\| \, \|x+y\| \leq a + \phi(y) \leq \|\phi\| \, \|x+y\| \implies |a + \phi(y)| \leq \|\phi\| \, \|x+y\|.$$

Thus

$$\begin{aligned} \|\tilde{\phi}(w)\| &= |t| \cdot \left| \phi\left(\frac{y}{t}\right) + a \right| \\ &\leq |t| \cdot \|\phi\| \cdot \left\| \frac{y}{t} + x \right\| \\ &= \|\phi\| \cdot \|y + tx\| \\ &= \|\phi\| \cdot \|w\|. \end{aligned}$$

□

Exercises.

6.2.1. Show that, in Example 6.2.3, the functionals defined in (a), (b), and (c) are linear, while that defined in (d) is not.

6.2.2 ([Coh03, Ex. 7.5(7)]). Fix $j \in \mathbb{N}_+$ and define $f \colon \ell^2 \to \mathbb{C}$ by $f(x) = x_j$, $x = (x_1, x_2, \dots) \in \ell^2$. Show that f is a linear functional and $\|f\| = 1$.

6.2.3. Show that on every metric space Lipschitz functions separate points from closed subsets.

6.2.4. Show that linear functionals separate points in any normed space.

Chapter 7

Inner product spaces and Hilbert spaces

In this final chapter we look at inner product spaces. You have seen these in previous courses. However, we will discuss them in some more depth here, and tie them into the other concepts we have seen in this course.

7.1 Inner product spaces

In this section, we will assume all vector spaces are over the complex numbers. The topic of this section (inner product spaces) is something you have seen in previous linear algebra courses. Therefore, we will omit some details of basic facts that you have seen before.

Definition 7.1.1 (Inner product space)**.** A (complex) *inner product space* is a vector space X together with a mapping $\langle \cdot, \cdot \rangle \colon X \times X \to \mathbb{C}$ with the following properties:

(IP1) $\langle x, x \rangle > 0$ for all $x \in X$, $x \neq \mathbf{0}$,

(IP2) $\langle x, y \rangle = \overline{\langle y, x \rangle}$ for all $x, y \in X$ (here $\bar{z}$ is the complex conjugate of $z \in \mathbb{C}$),

(IP3) $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$ for all $x, y \in X$ and every scalar α,

(IP4) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ for all $x, y, z \in X$.

The mapping $\langle \cdot, \cdot \rangle$ is called the *inner product*.

Remarks 7.1.2. (a) An inner product is sometimes also called a *scalar product* and an inner product space is sometimes called a *unitary space* or *pre-Hilbert space* (we will see the reason for this last term soon). The inner product is sometimes denoted by $\langle \cdot \mid \cdot \rangle$, $(\cdot, \cdot)$, or $(\cdot \mid \cdot)$.

(b) It follows from (IP2) that, for all $x \in X$, $\overline{\langle x, x \rangle} = \langle x, x \rangle$ and so $\langle x, x \rangle \in \mathbb{R}$.

(c) We will generally use the term inner product space to mean *complex* inner product space, as above. However, one can also consider *real* inner product spaces, where $\mathbb{C}$ is replaced by $\mathbb{R}$ in the definition. In this case, (IP2) becomes $\langle x, y \rangle = \langle y, x \rangle$.

Lemma 7.1.3. *If* $(X, \langle \cdot, \cdot \rangle)$ *is an inner product space, then*

(a) $\langle x, \mathbf{0} \rangle = \langle \mathbf{0}, y \rangle = 0$ *for all* $x, y \in X$*, and*

(b) we have

$$\left\langle \sum_{k=1}^{n} \alpha_k x_k, \sum_{j=1}^{m} \beta_j y_j \right\rangle = \sum_{k=1}^{n} \sum_{k=1}^{m} \alpha_k \bar{\beta}_j \langle x_k, y_j \rangle.$$

Proof. The proof of this lemma is left as an exercise (Exercise 7.1.1). □

Example 7.1.4. We can define inner product on $\mathbb{C}^n$ by

$$\langle x, y \rangle = \sum_{k=1}^{n} x_k \bar{y}_k, \quad x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \mathbb{C}^n.$$

(See Exercise 7.1.4.) We denote this inner product space again by $\mathbb{C}^n$.

Example 7.1.5. We can define an inner product on $\mathbb{R}^n$ by

$$\langle x, y \rangle = \sum_{k=1}^{n} x_k y_k, \quad x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \mathbb{R}^n.$$

This makes $\mathbb{R}^n$ a real inner product space, which we again denote by $\mathbb{R}^n$. The inner product is the usual dot product.

Example 7.1.6. Recall that ℓ^2 is the vector space of all complex-valued sequences $(x_1, x_2, \ldots)$ for which the series $\sum_{k=1}^{\infty} |x_k|^2$ converges. On ℓ^2 we define an inner product by

$$\langle x, y \rangle = \sum_{k=1}^{\infty} x_k \bar{y}_k, \quad x = (x_1, x_2, \ldots), y = (y_1, y_2, \ldots) \in \ell^2.$$

Of course, we need to check that this inner product is well-defined (i.e. that the series in the definition converges). By the Cauchy–Schwarz inequality, if $m \leq n$, we have

$$\left| \sum_{k=m}^{n} x_k \bar{y}_k \right| \leq \sum_{k=m}^{n} |x_k \bar{y}_k| = \sum_{k=m}^{n} |x_k| \, |y_k| \leq \sqrt{\sum_{k=m}^{n} |x_k|^2} \sqrt{\sum_{k=m}^{n} |y_k|^2}.$$

Since $x, y \in \ell^2$, we can make the series on the right as small as we like by requiring m, n to be large. Thus, for all $\epsilon > 0$, we can choose $N > 0$ such that

$$m, n > N \implies \left| \sum_{k=m}^{n} x_k \bar{y}_k \right| < \epsilon.$$

Therefore, $\sum_{k=1}^{\infty} x_k \bar{y}_k$ converges. (See [Coh03, Th. 1.8.2] if you forget this fact about series.)

Definition 7.1.7 (Norm on an inner product space)**.** If $(X, \langle \cdot, \cdot \rangle)$ is an inner product space, then, for $x \in X$, define

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

We will soon see that this is indeed a norm. But first we need to generalize the Cauchy–Schwarz inequality.

Theorem 7.1.8 (General Cauchy–Schwarz inequality)**.** *For any points x, y in an inner product space,*

$$|\langle x, y \rangle| \leq \|x\| \, \|y\|.$$

Proof. If $y = \mathbf{0}$, then the inequality is clear. So we assume $y \neq \mathbf{0}$, so $\|y\| > 0$. For any scalar α, we have

$$\begin{aligned} 0 \leq \|x + \alpha y\|^2 &= \langle x + \alpha y, x + \alpha y \rangle \\ &= \langle x, x \rangle + \langle x, \alpha y \rangle + \langle \alpha y, x \rangle + \langle \alpha y, \alpha y \rangle \\ &= \langle x, x \rangle + \bar{\alpha} \langle x, y \rangle + \alpha \langle y, x \rangle + \alpha \bar{\alpha} \langle y, y \rangle \\ &= \|x\|^2 + \alpha \overline{\langle x, y \rangle} + \bar{\alpha}(\langle x, y \rangle + \alpha \|y\|^2). \end{aligned}$$

Now take $\alpha = -\langle x, y \rangle / \|y\|^2$. Then $\langle x, y \rangle + \alpha \|y\|^2 = 0$ and thus

$$0 \leq \|x\|^2 - \frac{\langle x, y \rangle \overline{\langle x, y \rangle}}{\|y\|^2} = \|x\|^2 - \frac{|\langle x, y \rangle|^2}{\|y\|^2}.$$

The inequality follows. □

Corollary 7.1.9. *If $(X, \langle \cdot, \cdot \rangle)$ is an inner product space, then $(X, \| \cdot \|)$ is a normed space, where $\| \cdot \|$ is defined as in Definition 7.1.7.*

Proof. The verification of (N1) and (N2) are left as exercises (Exercise 7.1.6). To see (N3), note that for $x, y \in X$, we have

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

Taking square roots gives (N3). □

Thus, every inner product space is also a normed space, hence a metric space, hence a topological space.

Example 7.1.10. For continuous functions $x, y \colon [a, b] \to \mathbb{R}$, define

$$\langle x, y \rangle = \int_a^b x(t)y(t)\, dt.$$

One can check that this does indeed define an inner product on $C_2[a, b]$ (see Exercise 7.1.8). We denote this inner product space again by $C_2[a, b]$. Note that

$$\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{\int_a^b x(t)^2\, dt}$$

is the norm on $C_2[a, b]$.

We have seen examples of

- topological spaces that are not metric spaces (i.e. topological spaces that are not metrizable), and
- metric spaces that are not normed spaces (for instance, any metric space that cannot be made into a vector space).

Are there normed spaces that cannot be made into inner product spaces (in such a way that we recover the norm through Definition 7.1.7). Yes! In fact, $C[a, b]$ (with the uniform norm) is one. To prove this, we will prove a property of inner product spaces that is not satisfied by $C[a, b]$.

Proposition 7.1.11 (Parallelogram law)**.** *For any points x, y in an inner product space,*

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

Proof. This proof, which is by direct computation, is left as an exercise. Or see [Coh03, Th. 8.1.3]. □

Proposition 7.1.12. *There is no inner product on $C[a, b]$ that recovers the uniform norm.*

Proof. We will prove the special case $a = 0$, $b = 1$. It suffices to find two points that do not satisfy the parallelogram law. Define

$$x(t) = t, \quad y(t) = 1 - t, \quad 0 \le t \le 1.$$

Then

$$(x + y)(t) = 1, \quad (x - y)(t) = 2t - 1.$$

We easily check that

$$\|x\| = \max_{0 \le t \le 1} |t| = 1, \quad \|y\| = \|x + y\| = \|x - y\| = 1.$$

Therefore,

$$\|x + y\|^2 + \|x - y\|^2 = 1 \neq 4 = 2\|x\|^2 + 2\|y\|^2.$$ □

Definition 7.1.13 (Orthogonal)**.** Two vectors x, y in an inner product space X are called *orthogonal* if $\langle x, y\rangle = 0$, and we write $x \perp y$. A subset S of X is called an *orthogonal set* in X if $x \perp y$ for all $x, y \in S$, $x \neq y$. If, in addition, $\|x\| = 1$ for all $x \in S$, then S is called an *orthonormal set*.

Remark 7.1.14. Some references require the elements of an orthogonal set to be nonzero.

Proposition 7.1.15. *An orthogonal set of nonzero vectors in an inner product space is linearly independent.*

Proof. You saw the proof of this theorem in previous courses. □

Proposition 7.1.16. *Any vector subspace of countable dimension of an inner product space has an orthogonal basis.*

Proof. The proof of this theorem involves the *Gram-Schmidt algorithm*, which you saw in previous courses. □

Exercises.

7.1.1. Prove Lemma 7.1.3.

7.1.2. Suppose X is an inner product space. Show that, for all $x, y, z \in X$, we have

$$\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle.$$

7.1.3. Suppose X is an inner product space and $x, y \in X$ with $x \perp y$ and $\|x\| = \|y\|$. Show that $x + y$ is orthogonal to $x - y$.

7.1.4. Verify that $\langle \cdot, \cdot\rangle$, as defined in Example 7.1.4 does indeed define an inner product.

7.1.5. Verify that for the inner products defined on $\mathbb{C}^n$, $\mathbb{R}^n$ and ℓ^2 in this section, Definition 7.1.7 recovers the norms we defined on these spaces in Section 5.1.

7.1.6. Complete the proof of Corollary 7.1.9 by verifying (N1) and (N2).

7.1.7 ([Coh03, Ex. 8.6(3)]). For vectors in a complex inner product space, prove that

$$\langle x, y\rangle + \langle y, x\rangle = \frac{1}{2}\left(\|x + y\|^2 - \|x - y\|^2\right)$$

and

$$\langle x, y\rangle - \langle y, x\rangle = \frac{i}{2}\left(\|x + iy\|^2 - \|x - iy\|^2\right),$$

and hence deduce the *polarization identity*

$$\langle x, y\rangle = \frac{1}{4}\sum_{k=1}^{4} i^k \|x + i^k y\|^2.$$

(This shows that one can recover the inner product from the norm.)

7.1.8. Show that $\|x\|$, as defined in Example 7.1.10 does indeed define an inner product on $C_2[a, b]$.

7.1.9. Show that any finite-dimensional vector space can be equipped with an inner product.

7.1.10 ([Coh03, Ex. 8.6(5)]). Suppose $\{x_n\}$ and $\{y_n\}$ are Cauchy sequences in an inner product space. Prove that $\{\langle x_n, y_n\rangle\}$ is a convergent sequence in $\mathbb{C}$.

7.1.11 ([Coh03, Ex. 8.6(6)]). Let $\{y_1, y_2, \dots, y_n\}$ be a subset of an inner product space X and suppose $x \perp y_k$ for some $x \in X$ and all $k = 1, \dots, n$. Prove that $x \perp \sum_{k=1}^{n} \alpha_k y_k$ for any scalars $\alpha_1, \dots, \alpha_n$.

7.1.12 ([Coh03, Ex. 8.6(7)]). Suppose $\{x_n\}$ is a convergent sequence in an inner product space X, with $\lim x_n = x$. If there exists $y \in X$ such that $\langle x_n, y\rangle = 0$ for all $n \in \mathbb{N}_+$, prove that $\langle x, y\rangle = 0$.

7.1.13 ([Coh03, Ex. 8.6(8)]). Suppose that $\{x_n\}$ and $\{y_n\}$ are sequences in an inner product space such that $x_n \to \mathbf{0}$ and $\{y_n\}$ is bounded. Show that $\langle x_n, y_n\rangle \to 0$.

7.1.14. Let X be an inner product space and suppose $\{x_1, \dots, x_n\}$ is an orthogonal set in X. Prove that

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

This is a generalization of Pythagoras' Theorem.

7.2 Hilbert spaces

Since any inner product space is a normed space, we can ask if an inner product space is complete or not. Recall that a complete normed space is called a Banach space.

Definition 7.2.1 (Hilbert space)**.** A complete inner product space is called a *Hilbert space*.

Remark 7.2.2. By definition, every Hilbert space is a Banach space. However, the converse is not true. For example, we have seen that $C[a, b]$ is a Banach space, but it is not an inner product space (since it does not satisfy the parallelogram law) and is thus not a Hilbert space.

Example 7.2.3. By Theorem 5.3.13, any finite-dimensional inner product space is complete. Therefore, all finite-dimensional inner product spaces are Hilbert spaces.

Example 7.2.4. The metric space $C_2[a, b]$ is not complete (see Remark 1.4.14). Hence it cannot be complete as an inner product space. So $C_2[a, b]$ is not a Hilbert space. However, if one replaces the Riemann integral by the Lebesgue integral, there is a related space that *is* complete.

Example 7.2.5. The space ℓ^2 is complete, as noted in Remark 1.4.8. Thus ℓ^2 is a Banach space.

Theorem 7.2.6. *If f is a bounded linear functional on a Hilbert space X, then there exists a unique $v \in X$ such that $f(x) = \langle x, v\rangle$ for all $x \in X$.*

Proof. We will not prove this theorem in class. See, for instance, [Coh03, Th. 9.2.1]. □

Remark 7.2.7. Recall that the space of all bounded linear functionals on a normed space X is called the dual space of X and is denoted X^*. The above theorem states

$$f \in X^* \iff \exists\, v \in X \text{ such that } f(x) = \langle x, v\rangle \ \forall\, x \in X.$$

One can easily show that the vector v is unique.

We now develop the notion of the *adjoint* of a bounded linear mapping. Suppose X, Y are Hilbert spaces and $A \in B(X,Y)$ (i.e. $A\colon X \to Y$ is a bounded linear mapping). We define a map $Y \to X$ as follows. Choose an arbitrary $y \in Y$ and define a functional f on X by

$$f(x) = \langle Ax, y\rangle_Y, \quad x \in X. \tag{7.1}$$

One can check that f is linear (see Exercise 7.2.1). We have

$$|f(x)| = |\langle Ax, y\rangle_Y| \le \|Ax\|\,\|y\| \le (\|A\|\,\|y\|)\|x\|$$

where in the first inequality, we have used the Cauchy–Schwarz inequality. Thus f is bounded. Therefore, by Theorem 7.2.6, there exists a unique $v \in X$ such that

$$f(x) = \langle x, v\rangle_X \quad \forall\, x \in X.$$

Thus

$$\langle Ax, y\rangle_Y = \langle x, v\rangle_X \quad \forall x \in X.$$

Note that v depended on the choice of y. We define a mapping $A^*\colon Y \to X$ that sends v to this y. In other words A^* is defined by

$$\langle Ax, y\rangle_Y = \langle x, A^*y\rangle_X \quad \forall\, x \in X,\ y \in Y.$$

We now show that A^* is linear and bounded.

Let $y_1, y_2 \in Y$ and $\alpha_1, \alpha_2 \in \mathbb{C}$. Then, for all $x \in X$,

$$\begin{aligned}
\langle x, A^*(\alpha_1 y_1 + \alpha_2 y_2)\rangle_X &= \langle Ax, \alpha_1 y_1 + \alpha_2 y_2\rangle_Y \\
&= \bar\alpha_1\langle Ax, y_1\rangle_Y + \bar\alpha_2\langle Ax, y_2\rangle_Y \\
&= \bar\alpha_1\langle x, A^*y_1\rangle_X + \bar\alpha_2\langle x, A^*y_2\rangle_X \\
&= \langle x, \alpha_1 A^*y_1 + \alpha_2 A^*y_2\rangle_X.
\end{aligned}$$

Thus

$$\langle x, A^*(\alpha_1 y_1 + \alpha_2 y_2) - (\alpha_1 A^*y_1 + \alpha_2 A^*y_2)\rangle_X = 0 \quad \forall\, x \in X.$$

In particular, this holds for $x = A^*(\alpha_1 y_1 + \alpha_2 y_2) - (\alpha_1 A^*y_1 + \alpha_2 A^*y_2)$ and so

$$A^*(\alpha_1 y_1 + \alpha_2 y_2) = \alpha_1 A^*y_1 + \alpha_2 A^*y_2.$$

So A^* is linear.

It remains to show that A^* is bounded. For f as defined in (7.1), we have

$$f(x) = \langle Ax, y\rangle_Y = \langle x, A^* y\rangle_X \quad \forall\ x \in X,$$

and so

$$|f(x)| = |\langle x, A^* y\rangle_X| \le \|x\|\,\|A^* y\| = \|A^* y\|\,\|x\| \quad \forall\ x \in X,$$

and so f is bounded and $\|f\| \le \|A^* y\|$. Since

$$|f(v)| = |\langle v, v\rangle| = \|v\|\,\|v\|,$$

we also have $\|f\| \ge \|v\|$. Hence $\|f\| = \|v\|$.

Also, for all $x \in X$,

$$|f(x)| = |\langle Ax, y\rangle_Y| \le \|Ax\|\,\|y\| \le \|A\|\,\|x\|\,\|y\| = (\|A\|\,\|y\|)\|x\|,$$

and so

$$\|f\| \le \|A\|\,\|y\|.$$

Therefore, for $y \in Y$, we have

$$\|A^* y\| = \|f\| \le \|A\|\,\|y\|.$$

Thus A^* is bounded and $\|A^*\| \le \|A\|$.

Definition 7.2.8 (Adjoint)**.** If X and Y are Hilbert spaces, the *adjoint* of $A \in B(X, Y)$ is the $A^* \in B(Y, X)$ determined by

$$\langle Ax, y\rangle_Y = \langle x, A^* y\rangle_X, \quad x \in X,\ y \in Y.$$

If $Y = X$ and $A^* = A$, we say A is *self-adjoint*.

Theorem 7.2.9. *If X, Y are Hilbert spaces and $A \in B(X, Y)$, then $A^{**} = A$ and $\|A^*\| = \|A\|$.*

Proof. By the definition of the adjoint, we have

$$\langle y, Ax\rangle_Y = \overline{\langle Ax, y\rangle}_Y = \overline{\langle x, A^* y\rangle}_X = \langle A^* y, x\rangle_X = \langle y, A^{**} x\rangle_Y, \quad \forall\ x \in X,\ y \in Y.$$

Thus $Ax = A^{**}x$ for all $x \in X$, so $A^{**} = A$. We know that $\|A^*\| \le \|A\|$, so we now also have

$$\|A\| = \|A^{**}\| \le \|A^*\|.$$

Thus $\|A^*\| = \|A\|$. □

Self-adjoint operators are very important. For instance, in quantum mechanics, measurable quantities such as position, energy, and momentum are actually operators on an appropriate Hilbert space—the space of wave functions representing the state of a system. In fact, such observable quantities correspond to self-adjoint operators. Since these quantities are real numbers, the next result should not come as a surprise.

Lemma 7.2.10. *If A is a self-adjoint operator on a Hilbert space, then its eigenvalues are all real numbers.*

Proof. Suppose λ is an eigenvalue of A. Then there exists a nonzero vector x such that $Ax = \lambda x$. Then

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

Since $x \neq \mathbf{0}$, we have $\langle x, x\rangle \neq 0$, and so $\lambda = \bar{\lambda}$. Hence $\lambda \in \mathbb{R}$. □

We conclude this section (and the course) with a couple interesting results/applications.

Definition 7.2.11 (Complete set in an inner product space)**.** Suppose T is an orthonormal set in an inner product space X. If $\operatorname{Span} T$ is dense in X, then the set T is said to be *complete*.

Proposition 7.2.12. *A finite complete orthonormal set in an inner product space is a basis for the space.*

Finite complete orthonormal sets are not so interesting. What are more interesting are infinite complete orthonormal sets. They are somewhat like bases except that we need infinite sums (i.e. series) of multiples of the elements.

Proposition 7.2.13. *An inner product space is separable if and only if it contains a complete orthonormal set that is countable.*

Theorem 7.2.14 (Generalized Fourier Series Theorem)**.** *Suppose X is a separable inner product space and $T = \{x_1, x_2, \ldots,\}$ is a complete orthonormal set in X. Then the following are true.*

(a) For any $u \in X$, we have

$$u = \sum_{k=1}^{\infty} \langle u, x_k\rangle x_k.$$

(b) For all $u, v \in X$, we have

$$\langle u, v\rangle = \sum_{k=1}^{\infty} \langle u, x_k\rangle \langle x_k, v\rangle.$$

(c) For all $u \in X$, we have

$$\|u\|^2 = \sum_{k=1}^{\infty} |\langle u, x_k\rangle|^2.$$

The series in (a) is called the *Fourier series* for u, and the numbers $\langle u, x_n\rangle$ are called the *Fourier coefficients* of u.

Example 7.2.15. The more familiar theory of Fourier series is a special case of the above theorem. We take $X = C_2[-\pi, \pi]$ and $T = \left\{\frac{1}{\pi}, \frac{1}{\pi}\sin t, \frac{1}{\pi}\cos t, \frac{1}{\pi}\sin 2t, \frac{1}{\pi}\cos 2t, \dots\right\}$. Then any $f \in C_2[-\pi, \pi]$ can be written in the form

$$f(t) = \frac{1}{2}a_0 + \sum_{k=1}^{\infty}(a_k \cos kt + b_k \sin kt),$$

where

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos kt\, dt, \quad k = 0, 1, 2, \dots,$$
$$b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin kt\, dt, \quad k = 1, 2, \dots.$$

This has applications to music—we are writing an arbitrary function (or 'sound wave') in terms of sine and cosine functions ('pure tones').

Definition 7.2.16 (Isomorphism of inner product spaces)**.** An inner product space X is *isomorphic* to an inner product space Y if there is a bijective linear map $A\colon X \to Y$ such that

$$\langle x_1, x_2\rangle = \langle Ax_1, Ax_2\rangle \quad \forall\ x_1, x_2 \in X.$$

Theorem 7.2.17. *Any infinite-dimensional separable Hilbert space is isomorphic to* ℓ^2.

Exercises.

7.2.1. Prove that f defined by (7.1) is linear.

7.2.2 ([Coh03, Ex. 9.8(1)]). (a) Consider ℓ^2 as a normed space and, for $k \in \mathbb{N}_+$, let e_k be the point in ℓ^2 with all components 0 except for the kth, which is 1. Show that, if $x = (x_1, x_2, \dots,) \in \ell^2$, then $x = \sum_{k=1}^{\infty} x_k e_k$.

(b) Give an example in which this series for x is not absolutely convergent.

7.2.3. Suppose X, Y, and Z are inner product spaces, and that $A\colon X \to Y$ and $B\colon Y \to Z$ are isomorphisms. Show that the composition $BA\colon X \to Z$ is an isomorphism (of inner product spaces).

7.2.4. Prove that any n-dimensional complex inner product space is isomorphic (as an inner product space) to $\mathbb{C}^n$.

Index

Bibliography

[Coh03] Graeme L. Cohen. *A course in modern analysis and its applications*, volume 17 of *Australian Mathematical Society Lecture Series*. Cambridge University Press, Cambridge, 2003. http://dx.doi.org/10.1017/CBO9780511755125.