Notes on metric and Hilbert spaces
An invitation to functional analysis

Alexandru Ghitza*
School of Mathematics and Statistics
University of Melbourne

Version of Fri 18$^{\text{th}}$ Oct, 2024 at 11:13

*(aghitza@alum.mit.edu)

Contents

1. Introduction 5
   1.1. (*) What's up with infinite-dimensional vector spaces? . . . 5
   1.2. Notations and conventions . . . 9

2. Metric and topological spaces 11
   2.1. Metrics . . . 11
   2.2. Open subsets of metric spaces . . . 14
   2.3. Topological spaces . . . 16
   2.4. Connectedness . . . 20
   2.5. Compactness . . . 23
   2.6. (*) A diversion: topological groups . . . 25
   2.7. Sequences in metric spaces . . . 27
   2.8. Cauchy sequences . . . 29
   2.9. Completions . . . 31
   2.10. Banach Fixed Point Theorem . . . 34
   2.11. Boundedness and compactness in metric spaces . . . 36
   2.12. Spaces of bounded continuous functions . . . 39
   2.13. Function spaces: compactness . . . 41

3. Normed and Hilbert spaces 45
   3.1. Norms . . . 45
   3.2. Inner products . . . 49
   3.3. Convexity and inequalities . . . 51
   3.4. Sequence spaces . . . 53
   3.5. Continuous linear transformations . . . 56
   3.6. Series and Schauder bases . . . 60
   3.7. Dual normed spaces and completeness of sequence spaces . . . 63
   3.8. Orthogonality and projections . . . 66
   3.9. Duality in Hilbert spaces . . . 70
   3.10. Adjoint maps . . . 71
   3.11. Orthonormal bases . . . 73
   3.12. Function spaces: the uniform norm . . . 77
   3.13. (*) Function spaces: the L^p-norms . . . 80
   3.14. (*) The Lebesgue integral from scratch . . . 82
   3.15. (*) Some spectral theory . . . 84

A. Appendix 89
   A.1. Set theory . . . 89
   A.2. Linear algebra . . . 90
      A.2.1. Bilinear maps . . . 91
      A.2.2. Dual vector space . . . 91
      A.2.3. Inner products . . . 93

1. Introduction

1.1. (*) What's up with infinite-dimensional vector spaces?

The discussion in this section is heavily inspired by the lecture notes [2] by Karen Smith.

Despite the inevitable ups and downs, linear algebra as seen in a first-year subject is very satisfying. There is one fundamental construct (the linear combination, built out of the two operations defining the vector space structure) that gives rise to all the other abstract concepts (linear transformation, subspace, span, linear independence, etc.). And one of these abstract concepts (the basis) allows us to identify even the most ill-conceived of vector spaces with one of the friendly standard spaces $\mathbf{F}^n$, whereby we can use the concreteness of coordinates and matrices to perform computations that allow us to give explicit answers to many questions about these spaces.

If these vector spaces are finite-dimensional, that is. Once finite-dimensionality goes out the window, it takes much of our clear and satisfying linear-algebraic worldview with it. The purpose of this introduction is to bluntly point out the dangers of the infinite-dimensional landscape, and to take some tentative steps around it to see what tools we might need to use. After all, giving up is not an option: infinite-dimensional vector spaces are everywhere, so we might as well learn how to deal with them.

Let $\mathbf{F}$ be a field and V a vector space over $\mathbf{F}$. As you know, a *linear combination* is a **finite** expression of the form

$$a_1 v_1 + \cdots + a_n v_n \qquad \text{where } n \in \mathbf{N}, \quad a_1, \ldots, a_n \in \mathbf{F}, \quad v_1, \ldots, v_n \in V.$$

Finally, a subset B of V is a *basis* if every vector in V can be written **uniquely** as a **finite** linear combination of vectors in B.

First year linear algebra tells us that every finite-dimensional vector space V has a basis[1].

What happens if V is not finite-dimensional?

Example 1.1. The space of polynomials in one variable $\mathbf{R}[x]$ (sometimes called $\mathcal{P}(\mathbf{R})$ in linear algebra) has basis $B = \{1, x, x^2, \ldots\}$.

Solution. This is really just a restatement of the definition of polynomial: any element f of $\mathbf{R}[x]$ is of the form

$$f = a_0 + a_1 x + \cdots + a_n x^n,$$

thus a linear combination of elements of B.

If we have

$$f = a_0 + a_1 x + \cdots + a_n x^n = b_0 + b_1 x + \cdots + b_m x^m,$$

[1] This statement appears to be circular, as "finite-dimensional" is typically defined as "having a finite basis", but the circularity can be resolved by provisionally defining "finite-dimensional" as "being the span of some finite subset" until the existence of bases is established.

then the second equality is an equality of polynomials, which by definition requires $n = m$ and $a_i = b_i$ for all $i = 0, \dots, n$. $\square$

This first example worked out great: the space has bases, and we can actually write down a basis explicitly. We owe our luck to the fact that, even though the space of polynomials is not finite-dimensional, each element of the space is in some sense "finitely generated".

Something we can try is to start with the standard finite-dimensional spaces we know, namely $\mathbf{R}^n$, and "take the limit as $n \longrightarrow \infty$". This leads us to consider the space $\mathbf{R}^\infty$ of arbitrary real sequences $(x_1, x_2, \dots)$. We may naively hope that, since $\{e_1, e_2, \dots, e_n\}$ is a basis for $\mathbf{R}^n$, and these standard bases nest nicely as n increases, we end up with $\{e_1, e_2, \dots\}$ being a basis for $\mathbf{R}^\infty$, but that is not the case because, for instance, the constant sequence $(1, 1, \dots)$ is not in the span of $\{e_1, e_2, \dots\}$. (See Exercise 1.3 for more details.)

For another example, take $V = \mathbf{R}$ viewed as a vector space over $\mathbf{Q}$. One can show that the set $S = \left\{\sqrt{n} : n \in \mathbf{N} \text{ squarefree}\right\}$ is $\mathbf{Q}$-linearly independent in $\mathbf{R}$, but not a basis. The same is true of the set $T = \left\{\pi^n : n \in \mathbf{N}\right\}$. (See Exercise 1.4.) In fact, $\mathbf{R}$ has no countable basis over $\mathbf{Q}$. (See Exercise 1.5.) It's a sign that it may be rather difficult to write down an explicit $\mathbf{Q}$-basis of $\mathbf{R}$.

This is turning into a very depressing motivating section, so here is some good news:

Theorem 1.2. *Any vector space V has a basis.*

The proof of this theorem requires the (in)famous

Lemma 1.3 (Zorn's Lemma)**.** *Let X be a nonempty poset such that every nonempty chain C in X has an upper bound in X. Then X has a maximal element.*

For an explanation of the terms that appear in the statement of Zorn's Lemma, as well as a proof of Theorem 1.2, see Exercises 1.6 to 1.8.

The result is worth celebrating: we have bases for all vector spaces... but the proof gives absolutely no handle on what a basis looks like or how to compute one explicitly. This severely reduces the usefulness of the notion of a basis for an infinite-dimensional vector space.

And yet, it is hard to ignore the success of Example 1.1, where we saw an explicit, nice basis for the space of polynomials: $\left\{1, x, x^2, \dots\right\}$. We also know that many functions of one real variable can be expressed as Taylor series, for instance

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

This suggests that maybe one should drop the finiteness condition from the definition of linear combination and see where that leads. Consideration of Taylor series also tells us that we need something more than just the algebraic structure of a vector space if we are to make sense of infinite linear combinations. The notion of convergence of infinite series in real analysis is based on the Euclidean distance function on the real line: $d(x, y) = |x - y|$. We know from first year linear algebra that choosing an inner product on a vector space gives rise to a distance function, so that's a possible direction to explore. Before saying more about it though, note that an inner product also gives a concept of orthogonality, and of more general angles; and it is unclear whether angles are needed for what we want to do.

So here is, in rough terms, how we will be spending our time this semester.

The first thing that we will do is axiomatise the essential properties of the Euclidean distance function. We do this on arbitrary sets and obtain the notion of a **metric space**, and see that a surprising amount of results from real analysis carry through to this more general setting. There are certain respects in which metric spaces are not that well-behaved. Slightly counterintuitively, we remedy this by generalising even further to **topological spaces**, where we abandon the idea of distance between points in favour of the notion of neighbourhood of a point.

Once we have a grasp on the behaviour of general metric spaces and their topology, we consider the special case where the underlying set has a vector space structure. These are called **normed vector spaces** (in this setting, it is customary to single out the norm of a vector rather than the distance between two vectors; the two are equivalent).

Finally, because of their importance in many applications, we specialise further to inner product spaces. We could, for instance, consider the space $V = \mathrm{Cts}\big([-\pi,\pi],\mathbf{R}\big)$ of continuous functions $f\colon [-\pi,\pi] \longrightarrow \mathbf{R}$, endowed with the inner product

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(x)g(x)\,dx.$$

(A normalising factor is often placed in front of the integral for convenience, but we'll stick with this definition.)

The distance function is of course

$$d(f,g) = \sqrt{\langle f-g, f-g\rangle}.$$

This allows us to bring rigorous meaning to expressions such as

$$x = \sum_{n=1}^{\infty} \frac{2(-1)^{n+1}}{n}\sin(nx).$$

In our setting, we have

$$f(x) = x, \quad f_n(x) = \frac{2(-1)^{n+1}}{n}\sin(nx), \quad s_N(x) = \sum_{n=1}^{N} f_n(x),$$

all of them elements of V, and the claim is that $d(f, s_N) \longrightarrow 0$ as $N \longrightarrow \infty$.

It turns out that this space V has a maximal orthonormal set B such that every $f \in V$ can be written uniquely as an infinite series of elements of B, as in the example above. One can take B to consist of

$$\frac{1}{\sqrt{2\pi}}, \quad \frac{1}{\sqrt{\pi}}\sin(nx) \text{ for } n \in \mathbf{Z}_{\geqslant 1}, \quad \frac{1}{\sqrt{\pi}}\cos(nx) \text{ for } n \in \mathbf{Z}_{\geqslant 1},$$

and the unique expression of any $f \in V$ in terms of these elements is the Fourier series of f. (Note that the above B is countable, but V has uncountable dimension, a bit like $\mathbf{Q}$ being countable while $\mathbf{R}$ is uncountable.)

A modification of the Zorn Lemma argument in Exercise 1.8 shows that any inner product space V has a maximal orthonormal set. However, it is not true in general that every element of V can be written uniquely as an infinite series in the elements of the maximal orthonormal set. It is also not true in general that arbitrary infinite series give rise to an element of the vector space, even when these series "look like" they are converging.

A Hilbert space is an inner product space V that is complete: every Cauchy sequence converges to an element of V. This is certainly a desirable feature. But note that $\mathrm{Cts}\big([-\pi,\pi],\mathbf{R}\big)$ lacks it:

Example 1.4. Consider, for $n \geqslant 1$:

$$f_n(x) = \begin{cases} 0 & \text{if } x \leqslant 0, \\ x^{1/n} & \text{otherwise.} \end{cases}$$

The sequence (f_n) is Cauchy in $V = \mathrm{Cts}\big([-\pi,\pi],\mathbf{R}\big)$ with the distance function

$$d(f,g) = \sqrt{\int_{-\pi}^{\pi} (f-g)^2(x)\,dx}.$$

There is a pointwise limit given by

$$f(x) = \begin{cases} 0 & \text{if } x \leqslant 0, \\ 1 & \text{otherwise,} \end{cases}$$

that is, for any $x \in [-\pi,\pi]$ we have $f_n(x) \longrightarrow f(x)$ as $n \longrightarrow \infty$; but $f \notin V$, so V is not complete.

We will see that we can complete inner product spaces to obtain Hilbert spaces: in the example above, the completion is $L^2\big([-\pi,\pi],\mathbf{R}\big)$ consisting of (certain equivalence classes of) functions $f\colon [-\pi,\pi] \longrightarrow \mathbf{R}$ such that

$$\int_{-\pi}^{\pi} f^2(x)\,dx$$

exists and is finite.

Example 1.5. The function defined in Example 1.4

$$f(x) = \begin{cases} 0 & \text{if } x \leqslant 0, \\ 1 & \text{otherwise} \end{cases}$$

defines an element of $L^2\big([-\pi,\pi],\mathbf{R}\big)$ and the sequence (f_n) defined in Example 1.4 converges to f with respect to the given distance function.

Solution. We haven't discussed the Lebesgue integral but the function $f^2 = f$ is Lebesgue integrable and its Lebesgue integral is the sum of the Riemann integrals on the two intervals on which f is continuous:

$$\int_{-\pi}^{\pi} f^2(x)\,dx = \int_{-\pi}^{0} 0\,dx + \int_{0}^{\pi} 1\,dx = 0 + \pi = \pi.$$

For the statement about convergence we have

$$d(f,f_n)^2 = \int_{-\pi}^{0} (0-0)^2\,dx + \int_{0}^{\pi} (1-x^{1/n})^2\,dx = \pi - 2\frac{\pi^{1+1/n}}{1+1/n} + \frac{\pi^{1+2/n}}{1+2/n},$$

so $d(f,f_n) \longrightarrow 0$ as $n \longrightarrow \infty$. □

Of course, one cannot study mathematical structures without studying the maps between them. For topological spaces, this will mean continuous functions. For metric spaces, depending on what we are trying to do, it could be continuous functions, or distance-preserving functions, or contractions. For normed vector spaces, we will mostly work with continuous linear transformations; this naturally leads to questions about eigenvalues and eigenvectors, and ultimately to spectral theory, which is much richer than in the finite-dimensional setting.

1.2. Notations and conventions

Set inclusions are denoted $S \subseteq T$ (nonstrict inclusion: equality is possible) or $S \subsetneq T$ (strict inclusion: equality is ruled out). I will definitely avoid using $S \subset T$ (as it is ambiguous), and will try to avoid $S \nsubseteq T$ (not ambiguous, but too easily confused with $S \subsetneq T$). While we're at it, the power set of a set X, that is, the set of all subsets of X, is denoted $\mathcal{P}(X)$.

The symbols $|z|$ will always denote the usual absolute value (or modulus) function on $\mathbf{C}$:

$$|z| = \sqrt{x^2 + y^2}, \qquad \text{where } z = x + iy.$$

It, of course, defines a restricted function $|\cdot|\colon S \longrightarrow \mathbf{R}_{\geqslant 0}$ for any subset $S \subseteq \mathbf{C}$, which is the same as the real absolute value function when $S = \mathbf{R}$.

For better or worse, the natural numbers

$$\mathbf{N} = \{0, 1, 2, 3, \dots\}$$

start at 0. The variant starting at 1 is

$$\mathbf{Z}_{\geqslant 1} = \{1, 2, 3, \dots\}.$$

I use the term countable to mean what is more precisely called countably infinite, that is, a set in bijection with $\mathbf{N}$.

A Hermitian inner product is linear in the first variable and conjugate-linear in the second variable:

$$\langle \lambda x, y\rangle = \lambda\langle x, y\rangle, \qquad \langle x, \lambda y\rangle = \overline{\lambda}\langle x, y\rangle \qquad \text{for all } \lambda \in \mathbf{C}.$$

Unless otherwise specified, $\mathbf{F}$ denotes an arbitrary field.

I am not the right person to ask about foundational questions of logic or set theory: I neither know enough or care sufficiently about the topic. It's of course okay if you care and (want to) know more about these things. I am happy to spend my mathematical life in ZFC (Zermelo–Fraenkel set theory plus the Axiom of Choice), and these notes are part of my life so they are also hanging out in ZFC. In particular, I am very likely to use the Axiom of Choice without comment (and sometimes without noticing); I may occasionally point it out if someone brings my attention to it.

Acknowledgements

Thanks go to Thomas Black, Stephanie Carroll, Isaac Doosey-Shaw, Jack Gardiner, Leigh Greville, Ethan Husband, Peter Karapalidis, Rose-Maree Locsei, Quan Nguyen, Quang Ong, Hai Ou, Joshua Pearson, Lucas Pedersson, Kashma Pillay, Guozhen Wu, Corey Zelez, and Chengjing Zhang for corrections and suggestions on various incarnations of these notes.

2. Metric and topological spaces

2.1. Metrics

Think of Euclidean distance in $\mathbf{R}$:

$$d(x, y) = |x - y|.$$

What properties does it have? Well, certainly distances are non-negative, and two points are at distance zero from each other only if they are equal. The distance from x to y is equal to the distance from y to x. And we all love the triangle inequality: if you want to get from x to y, adding an intermediate stopover point t will not make the journey shorter.

We already know of other spaces where such functions exist ($\mathbf{R}^n$ comes to mind). So let's formalise these properties and see what we get.

Let X be a set. A *metric* (or *distance*) on X is a function

$$d \colon X \times X \longrightarrow \mathbf{R}_{\geqslant 0}$$

such that:

(a) $d(x, y) = d(y, x)$ for all $x, y \in X$;

(b) $d(x, y) \leqslant d(x, t) + d(t, y)$ for all $x, y, t \in X$;

(c) $d(x, y) = 0$ with $x, y \in X$ if and only if $x = y$.

The pair (X, d) is called a *metric space*; when the choice of metric is understood, we may drop it from the notation and simply write X.

Of course, the simplest example of a metric space is $\mathbf{R}$ with the Euclidean distance. But there are many other examples, some of which are quite exotic:

Example 2.1. (*) Let $X = \mathbf{Q}$ and fix a prime number p. We define a metric d_p on X that, in some sense, measures the distance between rational numbers from the point of view of divisibility by p. The definition proceeds in several stages:

(i) Define the *p-adic valuation* $v_p \colon \mathbf{Z} \longrightarrow \mathbf{Z}_{\geqslant 0} \cup \{\infty\}$ by:

$$v_p(n) = \text{the largest power of } p \text{ that divides } n,$$

with the convention that $v_p(0) = \infty$.

Show that $v_p(mn) = v_p(m) + v_p(n)$ for all $m, n \in \mathbf{Z}$.

(ii) Extend to the *p-adic valuation* $v_p \colon \mathbf{Q} \longrightarrow \mathbf{Z} \cup \{\infty\}$ by defining

$$v_p\left(\frac{m}{n}\right) = v_p(m) - v_p(n).$$

Show that for all $x, y \in \mathbf{Q}$ we have

$$v_p(xy) = v_p(x) + v_p(y)$$

and

$$v_p(x+y) \geqslant \min\left\{v_p(x), v_p(y)\right\},$$

with equality holding if $v_p(x) \neq v_p(y)$.

(iii) Next define the *p-adic absolute value* $|\cdot|_p \colon \mathbf{Q} \longrightarrow \mathbf{Q}_{\geqslant 0}$ by:

$$|x|_p = p^{-v_p(x)},$$

with the convention that $|0|_p = p^{-\infty} = 0$.

Show that for all $x, y \in \mathbf{Q}$ we have

$$|xy|_p = |x|_p\,|y|_p$$

and

$$|x+y|_p \leqslant \max\left\{|x|_p, |y|_p\right\},$$

with equality if $|x|_p \neq |y|_p$.

(iv) Finally define the *p-adic metric* on $\mathbf{Q}$ by

$$d_p(x, y) = |x - y|_p.$$

Show that $(\mathbf{Q}, d_p)$ is indeed a metric space.

Solution.

(i) Using the fundamental theorem of arithmetic (the existence of a unique prime factorisation of any natural number $\geqslant 2$), we have $m = p^{v_p(m)}m'$ and $n = p^{v_p(n)}n'$ with $p \nmid m'$ and $p \nmid n'$. Then

$$mn = p^{v_p(m)+v_p(n)}m'n' \qquad \text{and } p \nmid m'n',$$

so that $v_p(m) + v_p(n)$ is indeed the same as $v_p(mn)$.

(ii) Write $x = \frac{m}{n}$, $y = \frac{a}{b}$, then

$$v_p(xy) = v_p\left(\frac{ma}{nb}\right) = v_p(ma) - v_p(nb) = v_p(m) + v_p(a) - v_p(n) - v_p(b) = v_p(x) + v_p(y).$$

For $v_p(x+y)$, without loss of generality assume $v \coloneqq v_p(x) \leqslant v_p(y) =: u$ and write $x = p^v\frac{m'}{n'}$, $y = p^u\frac{a'}{b'}$. Then

$$x + y = p^v\,\frac{m'}{n'} + p^u\,\frac{a'}{b'} = p^v\left(\frac{m'}{n'} + p^{u-v}\,\frac{a'}{b'}\right) = p^v\left(\frac{m'b' + p^{u-v}a'n'}{n'b'}\right),$$

so that (since p does not divide $n'b'$)

$$v_p(x+y) = v + v_p\big(m'b' + p^{u-v}a'n'\big).$$

Since v_p of the quantity in parentheses is non-negative, we conclude that $v_p(x+y) \geqslant v = \min\big\{v_p(x), v_p(y)\big\}$.

Moreover, if $v < u$ then the quantity in parentheses has valuation zero, so that $v_p(x+y) = v = \min\big\{v_p(x), v_p(y)\big\}$.

(iii) Direct from the previous part and $|x|_p = p^{-v_p(x)}$.

(iv) We have

(a) Clearly $v_p(y-x) = v_p(-1) + v_p(x-y) = v_p(x-y)$, so $d_p(y,x) = d_p(x,y)$.

(b) Letting $u = x - t$ and $v = t - y$, we want to prove that $|u+v|_p \leqslant |u|_p + |v|_p$. But we have already seen that

$$|u+v|_p \leqslant \max\big\{|x|_p, |y|_p\big\},$$

and the latter is clearly $\leqslant |x|_p + |y|_p$.

(c) If $x \in \mathbf{Q} \neq 0$, then $v_p(x) \in \mathbf{Z}$ so $|x|_p = p^{-v_p(x)} \in \mathbf{Q} \setminus \{0\}$. Hence $|x|_p = 0$ iff $x = 0$, which implies that $d_p(x,y) = 0$ iff $x = y$. □

Example 2.2. Let Γ be a finite connected undirected simple graph (finitely many vertices, each pair of which are joined by at most one undirected edge; no loops). Given vertices x and y, we let $d(x,y)$ denote the minimum length of any path joining x and y.

Then d is a metric on the set of vertices of Γ.

Solution.

(a) Symmetry follows directly from the fact that Γ is undirected.

(b) Let $x, y, t \in \Gamma$, let p_1 be a shortest path (of length $d(x,t)$) joining x and t, and p_2 a shortest path (of length $d(t,y)$) joining t and y. Concatenating p_1 and p_2 we get a path of length $d(x,t) + d(t,y)$ from x to y, therefore $d(x,y)$ is at most equal to this length.

(c) Clear (if $x = y$ then the empty path goes from x to y; conversely, if $d(x,y) = 0$ then there is an empty path joining x to y, forcing $x = y$). □

Given a metric space, we can obtain other metric spaces by considering subsets:

Example 2.3. If (X,d) is a metric space, then for any subset S of X, the restriction of d to S gives a metric on S. (This is called the *induced metric.*)

Solution. Straightforward (follows immediately from the definitions). □

Or we can construct metric spaces as Cartesian products of other metric spaces. There are many ways of doing this, none of which is particularly canonical.

Example 2.4. Let (X_1, d_{X_1}) and (X_2, d_{X_2}) denote two metric spaces. Prove that the function d_1 defined by

$$d_1\big((x_1, x_2), (y_1, y_2)\big) = d_{X_1}(x_1, y_1) + d_{X_2}(x_2, y_2)$$

is a metric on the Cartesian product $X_1 \times X_2$.

The definition extends in the obvious manner to the Cartesian product of finitely many metric spaces $(X_1, d_{X_1}), \dots, (X_n, d_{X_n})$.

(This is sometimes called the *Manhattan metric* or *taxicab metric*. In the context of $\mathbf{R}^n = \mathbf{R} \times \cdots \times \mathbf{R}$, it is called the ℓ^1 *metric*.)

Solution. Straightforward. □

Example 2.5. Same setup as Example 2.4, but with the function

$$d_\infty\big((x_1, x_2), (y_1, y_2)\big) = \max\big(d_{X_1}(x_1, y_1), d_{X_2}(x_2, y_2)\big).$$

The definition extends in the obvious manner to the Cartesian product of finitely many metric spaces $(X_1, d_{X_1}), \dots, (X_n, d_{X_n})$.

(This is called the *sup norm metric* or *uniform norm metric*. In the context of $\mathbf{R}^n$, it is called the ℓ^∞ *metric*.)

Solution. Straightforward; proving the triangle inequality uses

$$\max\{a + b, c + d\} \leqslant \max\{a, c\} + \max\{b, d\}.$$ □

Example 2.6. Take $X_1 = X_2 = \mathbf{R}$ with the Euclidean metric and convince yourself that neither d_1 from Example 2.4 nor d_∞ from Example 2.5 is the Euclidean metric on $\mathbf{R}^2$.

Solution. Consider $(1, 2)$ and $(0, 0)$, then the distances are:

$$\begin{aligned} d_1\big((1,2),(0,0)\big) &= 1 + 2 = 3 \\ d_\infty\big((1,2),(0,0)\big) &= \max\{1, 2\} = 2 \\ d_2\big((1,2),(0,0)\big) &= \sqrt{1^2 + 2^2} = \sqrt{5}. \end{aligned}$$ □

Not every metric has to do with lengths and geometry in an obvious way. The p-adic metric in Example 2.1 is an example of something a little different. For another example, let $n \in \mathbf{Z}_{\geqslant 1}$, $X = \mathbf{F}_2^n$, and let $d(x, y)$ be the number of indices $i \in \{1, \dots, n\}$ such that $x_i \neq y_i$. Then d is a metric on X; it is called the *Hamming metric*. See Exercise 2.7 for more details.

2.2. Open subsets of metric spaces

A metric on a set X gives us a precise notion of distance between elements of the set. We use familiar geometric language to refer to the set of points within a fixed distance $r \in \mathbf{R}_{\geqslant 0}$ of a fixed point $c \in X$: the *open ball* of radius r and centre c is

$$\mathbf{B}_r(c) = \{x \in X : d(x, c) < r\}.$$

There is also, of course, a corresponding *closed ball*

$$\mathbf{D}_r(c) = \{x \in X : d(x, c) \leqslant r\}$$

and a corresponding *sphere*

$$\mathbf{S}_r(c) = \{x \in X : d(x, c) = r\}.$$

The familiar names are useful for guiding our intuition, but beware of the temptation to assume things about the shapes of balls in general metric spaces:

Example 2.7. Describe the Euclidean open balls centred at 0 in $\mathbf{Z}$ (endowed with the metric induced from the Euclidean metric on $\mathbf{R}$).

Solution. In addition to the empty set $\varnothing = \mathbf{B}_0(0)$, we have for all $n \in \mathbf{N}$ the set

$$\{-n, -n+1, \dots, -1, 0, 1, \dots, n-1, n\} = \mathbf{B}_{n+1}(0) = \mathbf{B}_r(0) \quad \text{for any } r \in (n, n+1]. \qquad \square$$

For more intuition-challenging examples, see Exercises 2.3 and 2.5.

We are now ready for a simple yet fundamental concept: a subset $U \subseteq X$ of a metric space (X, d) is an *open set* if, for every $u \in U$, there exists $r \in \mathbf{R}_{>0}$ such that $\mathbf{B}_r(u) \subseteq U$.

Example 2.8. Prove that $\varnothing$ and X are open sets.

Solution. The first statement is vacuously true; the second follows directly from the definition of $\mathbf{B}_r(x)$. $\square$

Example 2.9. Fix $x \in X$ and let $U = X \setminus \{x\}$; prove that U is an open set.

Solution. Let $u \in U$, then $u \neq x$ so $r := d(u, x) > 0$. Then $x \notin \mathbf{B}_r(u)$, so $\mathbf{B}_r(u) \subseteq U$. $\square$

Example 2.10. Prove that any open ball is an open set.

Solution. Let $U = \mathbf{B}_r(x)$. If $r = 0$ then $U = \varnothing$, an open set. Otherwise, let $u \in U$ and let $t = r - d(u, x)$. Since $d(u, x) < r$ we have $t > 0$.

I claim that $\mathbf{B}_t(u) \subseteq U$. Let $w \in \mathbf{B}_t(u)$, so that $d(w, u) < t$. Then

$$d(w, x) \leqslant d(w, u) + d(u, x) < t + r - t = r. \qquad \square$$

What happens if we combine open sets using set operations?

Proposition 2.11. *Let X be a metric space. The union of an arbitrary collection of open sets is an open set.*

Proof. Let I be an arbitrary set and, for each $i \in I$, let $U_i \subseteq X$ be an open set. We want to prove that

$$U = \bigcup_{i \in I} U_i$$

is open. Let $u \in U$, then there exists $i \in I$ such that $u \in U_i$. But $U_i \subseteq X$ is open, so there exists an open ball $\mathbf{B}_r(u) \subseteq U_i$. Since $U_i \subseteq U$, we have $\mathbf{B}_r(u) \subseteq U$. $\square$

Intersections are a bit more delicate:

Proposition 2.12. *Let X be a metric space. The intersection of a finite collection of open sets is an open set.*

Proof. Let $n \in \mathbf{N}$ and, for $i = 1, \dots, n$, let $U_i \subseteq X$ be an open set. We want to prove that

$$U = \bigcap_{i=1}^{n} U_i$$

is open. Let $u \in U$, then $u \in U_i$ for all $i = 1, \dots, n$. Since U_i is open, there exists an open ball $\mathbf{B}_{r_i}(u) \subseteq U_i$. Let $r = \min\{r_1, \dots, r_n\}$, then $\mathbf{B}_r(u) \subseteq \mathbf{B}_{r_i}(u) \subseteq U_i$ for each $i = 1, \dots, n$. Therefore $\mathbf{B}_r(u) \subseteq U$. □

Wondering about the necessity of the word "finite" in the statement of the proposition? See Tutorial Question 2.2.

2.3. Topological spaces

Given a set X, a *topology* on X is a subset $\mathcal{T} \subseteq \mathcal{P}(X)$ (in other words, $\mathcal{T}$ is a collection of subsets of X) such that

(a) $\varnothing \in \mathcal{T}$ and $X \in \mathcal{T}$;

(b) if $\{U_i : i \in I\}$ is an arbitrary collection of elements of $\mathcal{T}$, then $\bigcup_{i \in I} U_i \in \mathcal{T}$;

(c) if $\{U_1, \dots, U_n\}$ is a finite collection of elements of $\mathcal{T}$, then $\bigcap_{j=1}^{n} U_j \in \mathcal{T}$.

The elements of $\mathcal{T}$ are called *open sets* in X, and $(X, \mathcal{T})$ is called a *topological space*. A *closed set* of a topological space $(X, \mathcal{T})$ is a set whose complement is open.

Putting together Example 2.8 and Propositions 2.11 and 2.12, we see that metric spaces are topological spaces. (If (X, d) is a metric space, we call the topology defined by d the *metric topology* on X.)

Topological spaces are a very general concept encompassing much more than metric spaces[1]. We will not place a heavy focus on them in this subject, using them mostly to separate those properties of metric spaces that actually depend on the metric from those that depend only on the configuration of open subsets.

Example 2.13. Let X be an arbitrary set and let $\mathcal{T} = \{\varnothing, X\}$. This is called the *trivial topology* on X.

Example 2.14. Let X be an arbitrary set and let $\mathcal{T} = \mathcal{P}(X)$. (Every subset is an open subset.) This is called the *discrete topology* on X.

Example 2.15. Let X be an arbitrary set and let

$$\mathcal{T} = \{S \in \mathcal{P}(X) : X \setminus S \text{ is finite}\} \cup \{\varnothing\}.$$

This is called the *cofinite topology* on X.

[1] We say that a topological space $(X, \mathcal{T})$ is *metrisable* if there exists a metric d on X such that the resulting open sets are precisely $\mathcal{T}$. For an example of a non-metrisable space, see Tutorial Question 2.3.

In Tutorial Question 2.3 you will find all possible topologies on a set with two elements.

This game quickly becomes complicated as the size of the set increases, for instance a set of three elements has 29 distinct topologies.

Here is an easy way to produce many topologies on a set:

Example 2.16. Let X be a set and $S \subseteq \mathcal{P}(X)$. The *topology generated by* S is obtained by letting S' consist of all finite intersections of elements of S, then letting $\mathcal{T}$ consist of all arbitrary unions of elements of S'.

For instance, the discrete topology on X is generated by the set of singletons.

If (X, d) is a metric space, then the metric topology on X is generated by the set of open balls, see Exercise 2.8.

If $\mathcal{T}_1$ and $\mathcal{T}_2$ are two topologies on the same set X and $\mathcal{T}_1 \subseteq \mathcal{T}_2$ we say that $\mathcal{T}_1$ is *coarser* than $\mathcal{T}_2$ and $\mathcal{T}_2$ is *finer* than $\mathcal{T}_1$.

If d_1 and d_2 are two metrics on the same set X, we say that d_1 is *coarser* (resp. *finer*) than d_2 if the topology defined by d_1 is coarser (resp. finer) than the topology defined by d_2. We say that the metrics d_1 and d_2 are *(topologically) equivalent* if d_1 is both finer and coarser than d_2, simply put that d_1 and d_2 define precisely the same topology on X.

The appropriate notion of morphism for topological spaces is that of continuous function: if $f\colon X \longrightarrow Y$ is a function from one topological space to another, we say that f is *continuous* if, for any open subset $V \subseteq Y$, its inverse image $f^{-1}(V)$ is an open subset of X. The corresponding notion of isomorphism of topological spaces has a special name: a *homeomorphism* is a bijective continuous function $f\colon X \longrightarrow Y$ such that $f^{-1}\colon Y \longrightarrow X$ is continuous. In this case, X and Y are said to be *homeomorphic* topological spaces. It is easy to see (with the help of Tutorial Question 2.9) that this is an equivalence relation. (As an example, the 29 distinct topologies on a set with three elements fall into 9 homeomorphism classes.)

In the important special case of a metric space, the concept of continuous function has equivalent formulations that are more familiar from calculus and analysis. For example, the equivalence to the ε-δ definition is in Tutorial Question 2.8.

Example 2.17. Let (X, d) be a metric space and fix a point $t \in X$. Define $f\colon X \longrightarrow \mathbf{R}_{\geqslant 0}$ by

$$f(x) = d(x, t).$$

Then f is a continuous function.

Solution. Here is a proof that pretends to avoid the ε-δ formalism. By Tutorial Question 2.6 it suffices to consider opens $U \subseteq \mathbf{R}_{\geqslant 0}$ in a set that generates the topology on $\mathbf{R}_{\geqslant 0} \subseteq \mathbf{R}$; from real analysis, or a special case of Exercise 2.8, we can take $U = (a, b) \subseteq \mathbf{R}_{\geqslant 0}$ to be an open interval of finite length. Then

$$\begin{aligned} f^{-1}(U) &= f^{-1}\big((a, b)\big) \\ &= \{x \in X : a < d(x, t) < b\} \\ &= \big\{x \in X : a < d(x, t)\big\} \cap \{x \in X : d(x, t) < b\} \\ &= \big(X \setminus \mathbf{D}_a(t)\big) \cap \mathbf{B}_b(t), \end{aligned}$$

which is open in X as it is the intersection of two open sets. (Here we also used Exercise 2.10 to deduce that $\mathbf{D}_a(t)$ is a closed set.) □

If $(X, \mathcal{T})$ is a topological space and Y is any subset of X, we define

$$\mathcal{T}|_Y = \{U \cap Y : U \in \mathcal{T}\} \subseteq \mathcal{P}(Y).$$

Then $\mathcal{T}|_Y$ is a topology on Y, called the *induced (or subspace) topology*. On a metric space, this is compatible with the concept of induced metric, as you can see in Exercise 2.9.

If X_1 and X_2 are topological spaces, the *product topology* on $X_1 \times X_2$ is generated by the set

$$\mathcal{R} = \{U_1 \times U_2 \colon U_1 \subseteq X_1 \text{ open}, U_2 \subseteq X_2 \text{ open}\}.$$

(We might refer to the elements of $\mathcal{R}$ as *(open) rectangles*.)

Example 2.18. Show that $\mathcal{R}$ is closed under finite intersections, so that the product topology consists of arbitrary unions of rectangles.

Solution. By induction, we can reduce to checking that the intersection of two rectangles is again a rectangle. (Take a moment to appreciate the power and the danger of names.)

Let $R = U_1 \times U_2$, $R' = U_1' \times U_2'$ be two rectangles. Then

$$\begin{aligned} R \cap R' &= \{(x_1,x_2) \in X_1 \times X_2 \colon x_1 \in U_1, x_2 \in U_2\} \cap \{(x_1,x_2) \in X_1 \times X_2 \colon x_1 \in U_1', x_2 \in U_2'\} \\ &= \{(x_1,x_2) \in X_1 \times X_2 \colon x_1 \in U_1 \cap U_1', x_2 \in U_2 \cap U_2'\} \\ &= \big(U_1 \cap U_1'\big) \times \big(U_2 \cap U_2'\big). \end{aligned}$$ □

Proposition 2.19. *Let X_1, X_2 be topological spaces and endow $X_1 \times X_2$ with the product topology. Then the two projection maps $\pi_1 \colon X_1 \times X_2 \longrightarrow X_1$, $\pi_1(x_1,x_2) = x_1$, and $\pi_2 \colon X_1 \times X_2 \longrightarrow X_2$, $\pi_2(x_1,x_2) = x_2$, are continuous.*

The product topology is the coarsest topology on $X_1 \times X_2$ such that both π_1 and π_2 are continuous.

Proof. Straightforward: if $U_1 \subseteq X_1$ is open, then $\pi_1^{-1}(U_1) = U_1 \times X_2$ is an open rectangle in $X_1 \times X_2$.

For the minimality statement, suppose $\mathcal{T}$ is a topology on $X_1 \times X_2$ such that π_1 and π_2 are continuous. Let $U_1 \subseteq X_1$ and $U_2 \subseteq X_2$ be arbitrary opens. By continuity, $U_1 \times X_2 = \pi_1^{-1}(U_1)$ and $X_1 \times U_2 = \pi_2^{-1}(U_2)$ must be in $\mathcal{T}$, therefore so must their intersection

$$\big(U_1 \times X_2\big) \cap \big(X_1 \times U_2\big) = U_1 \times U_2.$$

We conclude that $\mathcal{T}$ contains all rectangles $U_1 \times U_2$, so the coarsest such topology is the topology generated by the rectangles (see Tutorial Question 2.4), that is the product topology. □

Let's go back to an example of the notion of metric on a product of metric spaces:

Example 2.20. In Exercise 2.5 we considered $X = \mathbf{R}$ and $X \times X = \mathbf{R}^2$ endowed with three different metrics:

$$\begin{aligned} d_1\big((x_1,x_2),(y_1,y_2)\big) &= |x_1 - y_1| + |x_2 - y_2| \\ d_\infty\big((x_1,x_2),(y_1,y_2)\big) &= \max\big\{|x_1 - y_1|, |x_2 - y_2|\big\} \\ d_2\big((x_1,x_2),(y_1,y_2)\big) &= \sqrt{|x_1 - y_1|^2 + |x_2 - y_2|^2}. \end{aligned}$$

These three different metrics give rise to the same topology on $\mathbf{R}^2$ (which is the same as the product topology); this is an easy application of the following criterion (Proposition 2.21).

Let X be a topological space. An *open neighbourhood* of $x \in X$ is an open set $U \subseteq X$ such that $x \in U$. A *neighbourhood* of $x \in X$ is a set $V \subseteq X$ containing an open neighbourhood of x.

Proposition 2.21. *Let X be a set and $\mathcal{T}_1, \mathcal{T}_2$ two topologies on X. The following statements are equivalent:*

(a) $\mathcal{T}_2$ is coarser than $\mathcal{T}_1$ (that is, $\mathcal{T}_2 \subseteq \mathcal{T}_1$);

(b) for any $x \in X$ and any $\mathcal{T}_2$-open neighbourhood U_x^2 of x, there exists a $\mathcal{T}_1$-open neighbourhood U_x^1 of x such that $U_x^1 \subseteq U_x^2$;

(c) the function $f\colon (X, \mathcal{T}_1) \longrightarrow (X, \mathcal{T}_2)$ given by $f(x) = x$ is continuous.

Proof. See Exercise 2.16. □

Topological spaces are sometimes *too* general. Life is a little easier given some basic amenities; here is a simple property that can make things more comfortable: we say that a topological space X is *Hausdorff* if given any distinct points $x \neq y$ of X, there exist open neighbourhoods U of x and V of y such that $U \cap V = \varnothing$. (We sometimes say that x and y are *separated* by opens, and refer to the Hausdorff condition as a *separation property*; there are others, weaker or stronger than this.)

Example 2.22. Any metric space (X, d) is Hausdorff.

Solution. If X is empty or a singleton, the statement is vacuously true.

Now suppose $x \neq y$, so that $d(x,y) > 0$. Let $2r = d(x,y)$, $U = \mathbf{B}_r(x)$, $V = \mathbf{B}_r(y)$, then $r > 0$ so U and V are nonempty opens, $x \in U$, $y \in V$, and $U \cap V = \varnothing$. □

Recall that a subset $C \subseteq X$ is *closed* if $X \setminus C$ is an open set. Beware: as opposed to their English language counterparts, the terms "open" and "closed" do not indicate a dichotomy! All four possibilities can be realised: you can have (a) sets that are both open and closed, (b) sets that are open but not closed, (c) sets that are closed but not open, (d) sets that are neither open nor closed.

Because of the interplay between open and closed sets, collections of closed sets have properties that are complementary to those of collections of open sets, see Exercise 2.18.

Given a topological space X and a subset $A \subseteq X$, we define

(a) the *interior* A° of A to be the union of all open subsets of A, equivalently the largest open subset of A;

(b) the *closure* $\overline{A}$ of A to be the intersection of all closed sets that contain A, equivalently the smallest closed set that contains A;

(c) the *boundary* ∂A of A to be $\partial A = \overline{A} \cap \overline{X \setminus A}$.

Proposition 2.23. *If A is a subset of a topological space X, then $x \in \overline{A}$ if and only if every open neighbourhood of x intersects A nontrivially.*

Proof. We prove the equivalent statement: $x \in X \setminus \overline{A}$ if and only if there exists an open neighbourhood U_x of x such that $U_x \cap A = \varnothing$.

Suppose $x \in X \setminus \overline{A}$. Letting $U_x = X \setminus \overline{A}$, we get an open neighbourhood of x with the property that $U_x \cap \overline{A} = \varnothing$, so a fortiori $U_x \cap A = \varnothing$.

Conversely, given U_x open and disjoint to A, $X \setminus U_x$ is closed and contains A, so it contains the closure $\overline{A}$. Hence $x \in X \setminus \overline{A}$. □

Proposition 2.24. *For any subset A of a topological space X we have:*

(a) $\partial A \cap A^\circ = \varnothing$;

(b) $\overline{A} = A^\circ \cup \partial A$;

(c) $A^\circ = A \setminus \partial A$.

Proof.

(a) $\partial A \cap A^\circ = \overline{A} \cap \overline{(X \setminus A)} \cap A^\circ = \overline{(X \setminus A)} \cap A^\circ$ since $A^\circ \subseteq A \subseteq \overline{A}$. Suppose $x \in \overline{(X \setminus A)} \cap A^\circ$. By Proposition 2.23 every open neighbourhood of x intersects $X \setminus A$ nontrivially; in particular A° intersects $X \setminus A$ nontrivially, contradiction.

(b) Since $A^\circ \subseteq A \subseteq \overline{A}$ and $\partial A = \overline{A} \cap \overline{(X \setminus A)} \subseteq \overline{A}$, the inclusion $A^\circ \cup \partial A \subseteq \overline{A}$ is clear.

In the other direction, let $x \in \overline{A}$ and suppose $x \notin \partial A$, which forces $x \notin \overline{(X \setminus A)}$. By Proposition 2.23 there exists an open neighbourhood U_x of x such that $U_x \cap (X \setminus A) = \varnothing$, that is $U_x \subseteq A$. Therefore $x \in A^\circ$.

(c) Since $A^\circ \subseteq A$ and $A^\circ \cap \partial A = \varnothing$ we have $A^\circ \subseteq A \setminus \partial A$.

From parts (a) and (b) we see that $\overline{A}$ is the disjoint union of A° and ∂A; in addition $A \subseteq \overline{A}$ so

$$A \setminus \partial A \subseteq \overline{A} \setminus \partial A = A^\circ.$$ □

We say that A is *nowhere dense* in X if $\left(\overline{A}\right)^\circ = \varnothing$. A simple example of this is $\mathbf{Z}$ as a subset of $\mathbf{R}$, see Tutorial Question 3.5.

We say that A is *dense* in X if $\overline{A} = X$.

Proposition 2.25. *If A is a subset of a topological space X, then A is dense in X if and only if every nonempty open subset of X intersects A nontrivially.*

Proof. Suppose A is dense in X and U is a nonempty open subset. Assume, by contradiction, that $A \cap U = \varnothing$, then $A \subseteq \left(X \setminus U\right)$. The latter is a closed set containing A, so by the definition of the closure we have $\overline{A} \subseteq \left(X \setminus U\right) \subsetneq X$, contradicting $\overline{A} = X$.

In the other direction, suppose A intersects all nonempty open subsets nontrivially. Assume, by contradiction, that $\overline{A} \neq X$, so that $U := X \setminus \overline{A}$ is a nonempty open set. Then it intersects A nontrivially: there exists $a \in A$ such that $a \in U$. But then $a \notin \overline{A}$, contradicting $a \in A \subseteq \overline{A}$. □

Example 2.26. Consider $\mathbf{R}$ with its usual topology. Both $\mathbf{Q}$ and $\mathbf{R} \setminus \mathbf{Q}$ are dense in $\mathbf{R}$.

Solution. Let $(a,b) \subseteq \mathbf{R}$ be a finite length interval with $a < b$. Let $n \in \mathbf{Z}_{\geqslant 1}$ be such that $n > 1/(b-a)$, then $nb - na > 1$. This means that there exists $m \in \mathbf{Z}$ such that $nb > m > na$. Hence the rational number $m/n \in (a,b)$.

Now (a,b) is uncountable and $\mathbf{Q}$ is countable, so (a,b) must also contain some irrational number. □

So we have two disjoint sets, each of which is dense in $\mathbf{R}$. The situation is very different if we ask for the sets to be both dense and open, which we do in Exercise 2.25.

2.4. Connectedness

We say that a topological space X is *disconnected* if there exist open subsets $U, V \subseteq X$ such that

$$X = U \cup V, \qquad U \cap V = \varnothing, \qquad U \neq \varnothing, \qquad V \neq \varnothing.$$

Note that this forces both U and V to be both closed and open.

We may sometimes refer to the above condition as expressing X as a nontrivial disjoint union of open subsets. If no such expressions for X exist, we say that X is *connected*.

More generally, a subset $D \subseteq X$ is said to be disconnected (resp. connected) if D is disconnected (resp. connected) with respect to the induced topology.

Spelling this out:

Proposition 2.27. *A subset D of a topological space X is disconnected if and only if there exist open subsets $U, V \subseteq X$ such that*

$$D \subseteq U \cup V, \qquad D \cap U \cap V = \varnothing, \qquad D \cap U \neq \varnothing, \qquad D \cap V \neq \varnothing.$$

Proof. See Exercise 2.19. □

Example 2.28. In any topological space X, $\varnothing$ and the singletons $\{x\}$, $x \in X$, are (vacuously) connected.

The set $\{0,1\} = \{0\} \cup \{1\}$ with the discrete topology is clearly disconnected. Unless we specify otherwise, we'll always endow $\{0,1\}$ with the discrete topology.

We say that a topological space X is *totally disconnected* if the only connected subsets of X are the empty set and the singletons.

Proposition 2.29. *A topological space X is disconnected if and only if there exists a non-constant continuous function $g\colon X \longrightarrow \{0,1\}$.*

(Of course a non-constant function with codomain $\{0,1\}$ is automatically surjective.)

Proof. Suppose there exists a non-constant continuous function $g\colon X \longrightarrow \{0,1\}$. Let $U = g^{-1}(0)$ and $V = g^{-1}(1)$, then $U \neq \varnothing$, $V \neq \varnothing$. Since $\{0\} \cap \{1\} = \varnothing$, we have $U \cap V = \varnothing$. Clearly $X = U \cup V$, and both U and V are open since $\{0\}$ and $\{1\}$ are open. This implies that X is disconnected.

For the other direction, suppose that X is disconnected and write $X = U \cup V$ with U, V open nonempty and $U \cap V = \varnothing$. Define $g\colon X \longrightarrow \{0,1\}$ by

$$g(x) = \begin{cases} 0 & \text{if } x \in U \\ 1 & \text{if } x \in V. \end{cases}$$

This is well-defined since $U \cap V = \varnothing$. It is continuous as $g^{-1}(0) = U$ and $g^{-1}(1) = V$ are open. It is not constant since it takes both values 0 and 1 (as both U and V are nonempty). □

Proposition 2.30. *If $f\colon X \longrightarrow Y$ is a continuous function between topological spaces and X is connected, then $f(X)$ is connected.*

Proof. Suppose $f(X)$ is disconnected, then by Proposition 2.29 there exists a non-constant continuous function $g\colon f(X) \longrightarrow \{0,1\}$. In particular, $f(X)$ has at least two elements. Then the composition $g \circ f\colon X \longrightarrow f(X) \longrightarrow \{0,1\}$ is a non-constant continuous function, implying that X is disconnected. □

Proposition 2.31. *Let X be a topological space.*

(a) A subset A of X is both closed and open if and only if $\partial A = \varnothing$.

(b) X is disconnected if and only if it has a nonempty subset $U \subsetneq X$ with $\partial U = \varnothing$.

Proof.

(a) By definition $\partial A = \overline{A} \cap \overline{(X \setminus A)}$.

If A is open then $X \setminus A$ is closed so $\overline{(X \setminus A)} = X \setminus A$. If A is closed then $\overline{A} = A$. So if A is both open and closed then $\partial A = \overline{A} \cap \overline{(X \setminus A)} = A \cap (X \setminus A) = \varnothing$.

Conversely, suppose $\partial A = \varnothing$. By Proposition 2.24 we have $\overline{A} = A^\circ \cup \partial A$, so in our case $\overline{A} = A^\circ$, but also $A^\circ \subseteq A \subseteq \overline{A}$. We conclude that $A^\circ = A = \overline{A}$, which implies that A is both an open set and a closed set.

(b) Suppose there exists a nonempty subset $U \subsetneq X$ with $\partial U = \varnothing$, and let $V := X \setminus U$. By part (a), U is both closed and open, so its complement V is both closed and open.

In the other direction, suppose X is disconnected and write $X = U \cup V$, $U \cap V = \varnothing$, both U and V open nonempty. Then U is both open and closed, so by part (a), $\partial U = \varnothing$. □

Example 2.32. $\mathbf{R}$ is connected.

Solution. Recall the notion of supremum of a subset $S \subseteq \mathbf{R}$: $M \in \mathbf{R}$ is a *supremum* of S if it is an upper bound for S (that is, $s \leqslant M$ for all $s \in S$), and if $x \in \mathbf{R}$ is any upper bound for S then $M \leqslant x$.

$\mathbf{R}$ has the property that every nonempty bounded above subset has a (unique) supremum.

There is a similar notion of *infimum*.

We will abuse this notation/terminology and say that a subset $S \subseteq \mathbf{R}$ that is not bounded above has $\sup(S)$ equal to $+\infty$, and a subset that is not bounded below has $\inf(S)$ equal to $-\infty$.

With this convention, an *interval* in $\mathbf{R}$ is a subset I with the property that for any $x \in \mathbf{R}$ with $\inf(I) < x < \sup(I)$, we have $x \in I$.

We use the criterion from Proposition 2.31, so we need to show that every nonempty subset $A \subsetneq \mathbf{R}$ has nonempty boundary.

Let $x \in \mathbf{R} \setminus A$. We have two possibilities:

- $S := (-\infty, x) \cap A \neq \varnothing$. Since $S \subseteq \mathbf{R}$ is nonempty and bounded above, it has a supremum $M \in \overline{S} \subseteq \overline{A}$. If $M = x$ then $M \notin A$ so $M \in \partial A$.

  If $M < x$ then $(M, x] \subseteq \mathbf{R} \setminus A$, therefore $M \in \overline{\mathbf{R} \setminus A}$ but $M \notin \big(\mathbf{R} \setminus A\big)^\circ$, hence $M \in \partial\big(\mathbf{R} \setminus A\big) = \partial A$.

- $S := (x, \infty) \cap A \neq \varnothing$, which is considered similarly by interchanging supremum and infimum. □

Example 2.33. The nonempty connected subsets of $\mathbf{R}$ are the intervals.

Solution. Let $S \subseteq \mathbf{R}$ be a nonempty subset that is not an interval. Then there exists $x \in \mathbf{R} \setminus S$ such that $\inf(S) < x < \sup(S)$ (where the infimum and supremum can be infinite). In that case $U := S \cap (-\infty, x)$ and $V := S \cap (x, \infty)$ show that S is disconnected.

Conversely, suppose I is an interval in $\mathbf{R}$. Then (Exercise 2.26) there exists a surjective continuous function $f \colon \mathbf{R} \longrightarrow I$, hence I is connected because $\mathbf{R}$ is connected. □

Theorem 2.34 (Intermediate Value Theorem)**.** *Let $f \colon X \longrightarrow \mathbf{R}$ be a continuous function, with X a connected topological space. For any $x, y \in X$ and any $r \in \mathbf{R}$ such that $f(x) < r < f(y)$, there exists $\xi \in X$ such that $f(\xi) = r$.*

Proof. The image $f(X)$ is a connected subset of $\mathbf{R}$, hence an interval, from which the conclusion follows. ☐

2.5. Compactness

Let X be a topological space. If K is a subset of X, an *open cover* of K is a collection $\{U_i : i \in I\}$ of open sets $U_i \subseteq X$ such that

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

We say that $K \subseteq X$ is *compact* if any open cover $\{U_i : i \in I\}$ of K has a finite *subcover*, that is there exist $n \in \mathbf{N}$ and $i_1, \ldots, i_n \in I$ such that

$$K \subseteq U_{i_1} \cup \cdots \cup U_{i_n}.$$

Proposition 2.35. *If X is a Hausdorff topological space and $K \subseteq X$ is a compact subset, then K is closed.*

Proof. We show that $X \smallsetminus K$ is open. Let $x \in X \smallsetminus K$. For each $k \in K$, since $k \neq x$ there exist open neighbourhoods U_k of k and V_k of x such that $U_k \cap V_k = \varnothing$. Putting it together we get an open cover

$$K \subseteq \bigcup_{k \in K} U_k,$$

which by compactness has a finite subcover

$$K \subseteq U_{k_1} \cup \cdots \cup U_{k_n} =: U.$$

Consider

$$V := V_{k_1} \cap \cdots \cap V_{k_n},$$

which is an open neighbourhood of x. We have $U \cap V = \varnothing$, therefore $V \subseteq X \smallsetminus U \subseteq X \smallsetminus K$ is an open neighbourhood of x contained in $X \smallsetminus K$. By Exercise 2.15, $X \smallsetminus K$ is open. ☐

Proposition 2.36. *If X is a compact topological space and $K \subseteq X$ is a closed subset, then K is compact.*

Proof. Consider an open cover of K:

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

We can turn this into an open cover of X:

$$X = (X \smallsetminus K) \cup K \subseteq (X \smallsetminus K) \cup \bigcup_{i \in I} U_i.$$

As X is compact, there is a finite subcover

$$X \subseteq (X \smallsetminus K) \cup U_{i_1} \cup \cdots \cup U_{i_n}.$$

As $K \subseteq X$ but $K \cap (X \smallsetminus K) = \varnothing$, we must have

$$K \subseteq U_{i_1} \cup \cdots \cup U_{i_n}.$$ ☐

Proposition 2.37. *If $f : X \longrightarrow Y$ is a continuous function between topological spaces and X is compact, then $f(X)$ is compact.*

Proof. Consider an arbitrary open cover of $f(X)$:

$$f(X) \subseteq \bigcup_{i \in I} V_i, \qquad V_i \subseteq Y \text{ open.}$$

Then

$$X \subseteq \bigcup_{i \in I} f^{-1}\big(V_i\big),$$

which is an open cover of X as f is continuous. By the compactness of X there is a finite subcover

$$X \subseteq f^{-1}(V_{i_1}) \cup \cdots \cup f^{-1}(V_{i_n}),$$

therefore

$$f(X) \subseteq V_{i_1} \cup \cdots \cup V_{i_n}. \qquad \square$$

A map $f\colon X \longrightarrow Y$ between topological spaces is *closed* if for any closed subset $C \subseteq X$, the image $f(C) \subseteq Y$ is closed. A map $f\colon X \longrightarrow Y$ between topological spaces is *proper* if for any compact subset $K \subseteq Y$, the inverse image $f^{-1}(K) \subseteq X$ is compact.

Proposition 2.38. *Let $f\colon X \longrightarrow Y$ be a closed map between topological spaces such that $f^{-1}(y) \subseteq X$ is compact for all $y \in Y$. Then f is proper.*

Proof. Take a compact subset $K \subseteq Y$ and consider the inverse image $f^{-1}(K)$. Take an arbitrary open cover

$$f^{-1}(K) \subseteq \bigcup_{i \in I} U_i.$$

Fix for the moment $k \in K$, then certainly

$$f^{-1}(k) \subseteq f^{-1}(K) \subseteq \bigcup_{i \in I} U_i,$$

but $f^{-1}(k)$ is compact by assumption, so there is a finite subcover

$$f^{-1}(k) \subseteq \bigcup_{i \in I_k} U_i =: \tilde{V}_k,$$

where $I_k \subseteq I$ is a finite subset.

Since $\tilde{V}_k$ is open in X, its complement $X \smallsetminus \tilde{V}_k$ is closed in X, so $f(X \smallsetminus \tilde{V}_k)$ is closed in Y (because f is a closed map). Letting $V_k = Y \smallsetminus f(X \smallsetminus \tilde{V}_k)$, we get an open neighbourhood V_k of k in Y such that $f^{-1}(V_k) \subseteq \tilde{V}_k$.

Now we vary $k \in K$ and get an open cover

$$K \subseteq \bigcup_{k \in K} V_k,$$

which by the compactness of K has a finite subcover

$$K \subseteq V_{k_1} \cup \cdots \cup V_{k_n}.$$

Then

$$\begin{aligned} f^{-1}(K) &\subseteq f^{-1}(V_{k_1}) \cup \cdots \cup f^{-1}(V_{k_n}) \\ &\subseteq \tilde{V}_{k_1} \cup \cdots \cup \tilde{V}_{k_n} \\ &= \bigcup_{i \in I_{k_1}} U_i \cup \cdots \cup \bigcup_{i \in I_{k_n}} U_i \\ &= \bigcup_{i \in I_{k_1} \cup \cdots \cup I_{k_n}} U_i, \end{aligned}$$

which is a finite subcover of the original

$$f^{-1}(K) \subseteq \bigcup_{i \in I} U_i. \qquad \square$$

Theorem 2.39. *Let X_1, X_2 be topological spaces.*

(a) If X_1 is compact then the map $\pi_2 \colon X_1 \times X_2 \longrightarrow X_2$ is closed and proper.

(b) If X_1 and X_2 are compact topological spaces, then their product $X_1 \times X_2$ is compact.

Proof.

(a) To show that π_2 is closed, let $C \subseteq X_1 \times X_2$ be a closed subset. Let $U = X_2 \setminus \pi_2(C)$ and let $u \in U$. Then $u \notin \pi_2(C)$; so for any $x \in X_1$, we have that $(x, u) \in (X_1 \times X_2) \setminus C$. As the latter set is open, there is an open neighbourhood of (x, u) that is an open rectangle $V_x^1 \times V_x^2$ with the property that $V_x^1 \times V_x^2 \cap C = \varnothing$. Then $\{V_x^1 \colon x \in X_1\}$ is an open cover of X_1, which is compact, so there is a finite cover

$$V_{x_1}^1 \cup \cdots \cup V_{x_n}^1 = X_1.$$

Setting

$$V = V_{x_1}^2 \cap \cdots \cap V_{x_n}^2,$$

we get an open neighbourhood $V \subseteq X_2$ of u such that

$$X_1 \times V \cap C = \left(V_{x_1}^1 \cup \cdots \cup V_{x_n}^1\right) \times \left(V_{x_1}^2 \cap \cdots \cap V_{x_n}^2\right) \cap C = \varnothing.$$

This means that $V \subseteq X_2 \setminus \pi_2(C) = U$, so that U is open.

The fact that π_2 is proper now follows from Proposition 2.38, since for any $x_2 \in X_2$ we have $\pi_2^{-1}(x_2) = X_1 \times \{x_2\}$, which is homeomorphic to X_1 by Exercise 2.20, hence compact.

(b) Follows directly from part (a) since $X_1 \times X_2 = \pi_2^{-1}(X_2)$. □

2.6. (*) A diversion: topological groups

A *topological group* is a topological space G that is also a group and such that the multiplication map

$$G \times G \longrightarrow G, \qquad (g, h) \longmapsto gh$$

and the inverse map

$$G \longrightarrow G, \qquad g \longmapsto g^{-1}$$

are both continuous.

Obviously, this makes the inverse map into a homeomorphism.

Note that some authors require topological groups G to be Hausdorff. We do not.

Example 2.40. Any group G endowed with the discrete topology (or with the trivial topology) is a topological group.

Example 2.41. Consider $\mathbf{R}$ with the Euclidean topology, under the addition operation on $\mathbf{R}$.

More generally, $V = \mathbf{R}^n$ with the Euclidean topology, under addition of vectors.

Example 2.42 (The circle group). Let

$$\mathbf{S}^1 = \{z \in \mathbf{C} \colon |z| = 1\}.$$

Give this the subspace topology coming from the usual topology on $\mathbf{C}$, and let the group operation be complex multiplication.

Example 2.43 (The general linear groups). Let $n \in \mathbf{Z}_{\geqslant 1}$ and

$$\mathrm{GL}_n(\mathbf{R}) = \{M \in M_{n\times n}(\mathbf{R}) \colon M \text{ is invertible } \}.$$

Give $M_{n\times n}(\mathbf{R}) \equiv \mathbf{R}^{n^2}$ the Euclidean topology and $\mathrm{GL}_n(\mathbf{R})$ the subspace topology.

Matrix multiplication is continuous in the matrix entries. (One should also check that matrix inversion is continuous.)

Proposition 2.44. *Let G be a topological group and $g \in G$. The* left translation *map $L_g \colon G \longrightarrow G$ given by $L_g(x) = gx$ is a homeomorphism. So is the* right translation *map R_g.*

Proof. The map L_g is the composition of the continuous map $G \longrightarrow G\times G$ given by $x \longmapsto (g,x)$ and the multiplication map of G, hence is continuous. It is clear that $L_{g^{-1}}$ is the inverse of L_g, and that it is also continuous. □

Corollary 2.45. *Any topological group G is a* homogeneous topological space*, that is: for every $x, y \in G$ there exists a homeomorphism $f \colon G \longrightarrow G$ such that $f(x) = y$.*

Proof. Let $f = L_{yx^{-1}}$. □

A topological group homomorphism $f \colon G \longrightarrow H$ is a group homomorphism that is continuous with respect to the topologies on G and H.

Example 2.46. We know that the inverse map $G \longrightarrow G$, $g \longmapsto g^{-1}$ is continuous (in fact, a homeomorphism). But it is a group homomorphism (and hence a topological group homomorphism) if and only if G is abelian.

On the other hand, for any topological group G and any $g \in G$, *conjugation* by g given by $c_g \colon G \longrightarrow G$, $c_g(x) = g^{-1}xg$ is a topological group isomorphism, that is a group isomorphism that is also a homeomorphism. (This follows simply from $c_g = R_g \circ L_{g^{-1}}$.)

Example 2.47. The map $\exp \colon \mathbf{R} \longrightarrow \mathbf{R}^\times$ is a topological group homomorphism, where $\mathbf{R}$ has the Euclidean topology and the addition operation, and $\mathbf{R}^\times$ has the subspace topology and the multiplication operation.

Example 2.48. The determinant map $\det \colon \mathrm{GL}_n(\mathbf{R}) \longrightarrow \mathbf{R}^\times$ is a topological group homomorphism.

Proposition 2.49. *Let G be a topological group and H a subgroup. Then the closure $\overline{H}$ is a subgroup of G. Moreover, if H is normal, then so is $\overline{H}$.*

Proof. Clearly the identity element $e \in H \subseteq \overline{H}$.

In the rest of the proof, we will repeatedly use Proposition 2.23: if $A \subseteq X$, then $x \in \overline{A}$ if and only if every open neighbourhood of x intersects A nontrivially.

Suppose $g \in \overline{H}$; we want to show that $g^{-1} \in \overline{H}$. Let $U \subseteq G$ be an open neighbourhood of g^{-1}. Then (since inversion is a homeomorphism) U^{-1} is an open neighbourhood of $g \in \overline{H}$, so let $h \in U^{-1} \cap H$. Then $h^{-1} \in U \cap H^{-1} = U \cap H$ since H is a subgroup; we conclude that U intersects H nontrivially, so $g^{-1} \in \overline{H}$.

Now suppose $g_1, g_2 \in \overline{H}$; we want to show that $g_1 g_2 \in \overline{H}$. Let $U \subseteq G$ be an open neighbourhood of $g_1 g_2$. Then $m^{-1}(U) \subseteq G \times G$ is an open neighbourhood of (g_1, g_2) (since the multiplication map m is continuous), therefore it contains an open rectangle $U_1 \times U_2$ that is an open neighbourhood of (g_1, g_2). There exist $h_1 \in U_1 \cap H$ and $h_2 \in U_2 \cap H$. Let $U' = m(U_1, U_2)$, then $g_1 g_2 \in U' \subseteq U$. Moreover, $(h_1, h_2) \in (U_1 \times U_2) \cap (H \times H)$, therefore $h_1 h_2 \in U' \cap H \subseteq U \cap H$. We conclude that the latter intersection is nonempty, so that $g_1 g_2 \in \overline{H}$.

So $\overline{H}$ is a subgroup of G.

Assume finally that H is a normal subgroup. Let $g \in G$ and $x \in \overline{H}$; we want to show that $gxg^{-1} \in \overline{H}$. Let U be an open neighbourhood of gxg^{-1}. Then $g^{-1}Ug$ is an open neighbourhood of $x \in \overline{H}$, so there exists $h \in H$ such that $h \in g^{-1}Ug \cap H$. Then $ghg^{-1} \in U \cap gHg^{-1} = U \cap H$. □

There is much more to say about topological groups (quotients, action on a topological space, structure, representations, etc.) And there are topological rings, topological fields, topological vector spaces. We will see an important class of the latter in the next chapter, but for now we leave this topic and the generality of topological spaces, and return to the case of metric spaces.

2.7. Sequences in metric spaces

Let (X, d) be a metric space.

A *sequence* in X is a function $\mathbf{N} \longrightarrow X$, commonly denoted as (x_n), meaning that $n \longmapsto x_n$. We say that (x_n) *converges* to a *limit* $x \in X$ if for any $\varepsilon \in \mathbf{R}_{>0}$ there exists $N \in \mathbf{N}$ such that

$$x_n \in \mathbf{B}_\varepsilon(x) \qquad \text{for all } n \geqslant N.$$

The next result describes the relationship between limits and sets that are open or closed.

Proposition 2.50. *Let (X, d) be a metric space and let (x_n) be a sequence that converges to $x \in X$.*

(a) If $U \subseteq X$ is an open subset such that $x \in U$, then there exists $N \in \mathbf{N}$ such that $x_n \in U$ for all $n \geqslant N$.

(We sometimes refer to this situation as: $x_n \in U$ for sufficiently large n.)

(b) If $A \subseteq X$ is an arbitrary subset such that $x_n \in A$ for all $n \in \mathbf{N}$, then $x \in \overline{A}$.

Conversely, given any $y \in \overline{A}$ there exists a sequence (y_n) in A that converges to y.

(c) A is closed if and only if for every sequence $(x_n) \longrightarrow x \in X$ with $x_n \in A$, we have $x \in A$.

Proof.

(a) As $x \in U$ and U is open, there exists $\varepsilon > 0$ such that $\mathbf{B}_\varepsilon(x) \subseteq U$. But as $(x_n) \longrightarrow x$, there exists $N \in \mathbf{N}$ such that $x_n \in \mathbf{B}_\varepsilon(x) \subseteq U$ for all $n \geqslant N$.

(b) Let $U \subseteq X$ be an open neighbourhood of x. By part (a), there exists $N \in \mathbf{N}$ such that $x_n \in U$ for all $n \geqslant N$. In particular, U intersects A nontrivially. By Proposition 2.23, we conclude that $x \in \overline{A}$.

For the converse statement: let $y \in \overline{A}$. Let $y_0 \in A$ be arbitrary, then for any $n \in \mathbf{Z}_{\geqslant 1}$ consider the open neighbourhood $\mathbf{B}_{1/n}(y)$ of y. It must intersect A nontrivially, so let $y_n \in \mathbf{B}_{1/n}(y) \cap A$.

The result is a sequence (y_n) of elements of A that converges to y. (For any $\varepsilon > 0$, take $N \in \mathbf{N}$ such that $1/N < \varepsilon$, etc.)

(c) Follows immediately from (b). □

Suppose (x_n) and (y_n) are two sequences in a metric space (X, d). We say that

$$(x_n) \sim (y_n) \qquad \text{if } \big(d(x_n, y_n)\big) \longrightarrow 0 \text{ as } n \longrightarrow \infty.$$

By Exercise 2.29, this is an equivalence relation on the set of sequences in (X, d).

Proposition 2.51. *Let (x_n) and (y_n) be equivalent sequences in a metric space (X, d) and let $x \in X$. Then (x_n) converges to x if and only if (y_n) converges to x.*

Proof. As equivalence is symmetric, it suffices to prove that if $(x_n) \longrightarrow x$ then $(y_n) \longrightarrow x$.

Let $\varepsilon \in \mathbf{R}_{>0}$. Let $N_1 \in \mathbf{N}$ be such that $d(x_n, y_n) < \varepsilon/2$ for all $n \geqslant N_1$, and let $N_2 \in \mathbf{N}$ be such that $d(x_n, x) < \varepsilon/2$ for all $n \geqslant N_2$. Setting $N = \max\{N_1, N_2\}$, for all $n \geqslant N$ we have

$$d(y_n, x) \leqslant d(y_n, x_n) + d(x_n, x) < \varepsilon.$$ □

Recall (Tutorial Question 2.8) that for metric spaces we have an ε-δ description of continuity. There is also a sequential criterion for continuity:

Theorem 2.52. *Let $f\colon X \longrightarrow Y$ be a function between metric spaces and let $x \in X$. Then f is continuous at x if and only if for all sequences $(x_n) \longrightarrow x$, the sequence $\big(f(x_n)\big) \longrightarrow f(x)$.*

Proof. Suppose f is continuous; let (x_n) be a sequence converging to x in X and let $y = f(x)$.

Let $\varepsilon \in \mathbf{R}_{>0}$. There exists $\delta \in \mathbf{R}_{>0}$ such that if $x' \in \mathbf{B}_\delta(x)$ then $f(x') \in \mathbf{B}_\varepsilon(y)$. On the other hand, since (x_n) converges to x, given the above δ, there exists $N \in \mathbf{N}$ such that $x_n \in \mathbf{B}_\delta(x)$ for all $n \geqslant N$. We conclude that $f(x_n) \in \mathbf{B}_\varepsilon(y)$ for all $n \geqslant N$, so that $\big(f(x_n)\big)$ converges to y.

Conversely, suppose the statement about convergence of sequences holds. We use a proof by contradiction to show that f must be continuous at x.

Suppose there exists $\varepsilon \in \mathbf{R}_{>0}$ such that for all $\delta \in \mathbf{R}_{>0}$, $f\big(\mathbf{B}_\delta(x)\big) \setminus \mathbf{B}_\varepsilon(f(x)) \neq \varnothing$. In particular, for any $n \in \mathbf{Z}_{\geqslant 1}$ we can take $\delta = \frac{1}{n}$ and find some element $x_n \in \mathbf{B}_{1/n}(x)$ such that $f(x_n) \notin \mathbf{B}_\varepsilon(f(x))$. This gives us a sequence (x_n) that converges to x, but $\big(f(x_n)\big)$ does not converge to $f(x)$. □

There is a notion of map between metric spaces that is stricter than continuity, in that it preserves the full metric structure: we say that a function $f\colon (X, d_X) \longrightarrow (Y, d_Y)$ is *distance-preserving* if

$$d_Y(f(x_1), f(x_2)) = d_X(x_1, x_2) \qquad \text{for all } x_1, x_2 \in X.$$

Note that a distance-preserving function must be injective, as well as continuous.

An *isometry*[2] is a bijective distance-preserving map whose inverse is also distance-preserving (you should check that this last condition is in fact unnecessary: the inverse of a bijective distance-preserving map is automatically distance-preserving). If an isometry exists we say that X and Y are *isometric*.

Whether continuous or distance-preserving functions are the right tool depends on whether you are concerned only with topological properties, or with the metric structure. There are other useful flavours of maps that we will see soon.

[2]Warning: many authors use the term *isometry* to denote a distance-preserving map.

2.8. Cauchy sequences

Here is something that you know from real analysis and follows easily from the definition of sequential convergence:

Proposition 2.53. *Let (X, d) be a metric space and suppose $(x_n) \longrightarrow x \in X$. Then, given $\varepsilon > 0$, there exists $N \in \mathbf{N}$ such that $d(x_n, x_m) < \varepsilon$ for all $n, m \geqslant N$.*

Proof. Since $(x_n) \longrightarrow x$, there exists $N \in \mathbf{N}$ such that $d(x_n, x) < \varepsilon/2$ for all $n \geqslant N$. Therefore, for all $n, m \geqslant N$ we have

$$d(x_n, x_m) \leqslant d(x_n, x) + d(x, x_m) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$ □

A sequence (x_n) that satisfies the conclusion of Proposition 2.53 is said to be *Cauchy*.

A natural question is whether the converse of Proposition 2.53 holds: does every Cauchy sequence converge? In an arbitrary metric space, the answer is no. We say that a metric space X is *complete* if every Cauchy sequence converges to an element of X.

> **Example 2.54.** (I hope) we know from real analysis that $\mathbf{R}$ is a complete metric space. However, $\mathbf{Q}$ is not complete, as you can see in Exercise 2.35.

Proposition 2.55. *If X is a complete metric space and $S \subseteq X$, then S is complete if and only if S is closed.*

Proof. Suppose S is complete and let $x \in \overline{S}$. Then there exists a sequence (s_n) in S such that $(s_n) \longrightarrow x \in X$; by Proposition 2.53 we know that (s_n) is Cauchy, so by the completeness of S we have $x \in S$. Therefore $\overline{S} = S$.

Conversely, suppose S is closed in X. Let (s_n) be a Cauchy sequence in S, then (s_n) is a Cauchy sequence in X, which is complete, so $(s_n) \longrightarrow x \in X$. By Proposition 2.50 we have $x \in \overline{S} = S$ since S is closed. □

Proposition 2.56. *If (x_n) and (y_n) are Cauchy sequences in a metric space (X, d), then $\big(d(x_n, y_n)\big)$ is a Cauchy sequence in $\mathbf{R}$.*

Solution. First note that for any n, m we have by the triangle inequality:

$$d(x_n, y_n) \leqslant d(x_n, x_m) + d(x_m, y_n) \leqslant d(x_n, x_m) + d(x_m, y_m) + d(y_m, y_n),$$

so

$$d(x_n, y_n) - d(x_m, y_m) \leqslant d(x_n, x_m) + d(y_m, y_n).$$

Similarly:

$$d(x_m, y_m) \leqslant d(x_m, x_n) + d(x_n, y_n) + d(y_n, y_m)$$

so that

$$-\big(d(x_m, x_n) + d(y_n, y_m)\big) \leqslant d(x_n, y_n) - d(x_m, y_m).$$

We can summarise this as

$$\big|d(x_n, y_n) - d(x_m, y_m)\big| \leqslant d(x_m, x_n) + d(y_n, y_m).$$

Let $\varepsilon > 0$. There exists $N_1 \in \mathbf{N}$ such that $d(x_n, x_m) < \varepsilon/2$ for all $m, n \geqslant N_1$. There exists $N_2 \in \mathbf{N}$ such that $d(y_n, y_m) < \varepsilon/2$ for all $m, n \geqslant N_2$. Let $N = \max\{N_1, N_2\}$, then for all $n, m \geqslant N$ we have:

$$\big|d(x_n, y_n) - d(x_m, y_m)\big| \leqslant d(x_n, x_m) + d(y_m, y_n) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

So $\big(d(x_n, y_n)\big)$ is a Cauchy sequence in $\mathbf{R}$. □

The equivalence relation on sequences preserves the Cauchy property:

Proposition 2.57. *Let (x_n) and (y_n) be equivalent sequences in a metric space (X,d). Then (x_n) is Cauchy if and only if (y_n) is Cauchy.*

Solution. It suffices to prove that (x_n) being Cauchy implies (y_n) is Cauchy.

Let $\varepsilon > 0$. As $(y_n) \sim (x_n)$, there exists $N_1 \in \mathbf{N}$ such that $d(y_n, x_n) < \varepsilon/3$ for all $n \geqslant N_1$. As (x_n) is Cauchy, there exists $N_2 \in \mathbf{N}$ such that $d(x_n, x_m) < \varepsilon/3$ for all $n, m \geqslant N_2$. Let $N = \max\{N_1, N_2\}$, then for all $n, m \geqslant N$ we have

$$d(y_n, y_m) \leqslant d(y_n, x_n) + d(x_n, x_m) + d(x_m, y_m) < \varepsilon. \qquad \square$$

However, continuous functions do not necessarily preserve the Cauchy property:

Example 2.58. Take $X = Y = \mathbf{R}_{>0}$ with the induced metric from $\mathbf{R}$, and $f\colon X \longrightarrow Y$ given by $f(x) = \frac{1}{x}$. The function f is continuous on X. Take the sequence (x_n) with $x_n = \frac{1}{n}$ for all $n \in \mathbf{N}$. Then (x_n) is Cauchy, but $\big(f(x_n)\big) = (n)$ is most certainly not Cauchy.

If you want your functions to preserve the Cauchy property, you need a stronger condition than continuity: a function $f\colon X \longrightarrow Y$ between metric spaces is *uniformly continuous* if for all $\varepsilon > 0$ there exists $\delta > 0$ such that for all $x \in X$ we have $f\big(\mathbf{B}_\delta(x)\big) \subseteq \mathbf{B}_\varepsilon\big(f(x)\big)$.

The last part of the definition is equivalent to: for all $x, x' \in X$ we have

$$d_X(x, x') < \delta \quad \Rightarrow \quad d_Y\big(f(x), f(x')\big) < \varepsilon.$$

(You may have to read the definition more than once, and compare it symbol by symbol with the definition of continuity, to see what the difference is: here δ depends only on the given ε, not on $x \in X$. Hence its choice is *uniform over* X.)

Example 2.59. Any distance-preserving function is uniformly continuous. This is immediate from the definitions (can take $\delta = \varepsilon$).

Proposition 2.60. *Any uniformly continuous function maps Cauchy sequences to Cauchy sequences.*

Proof. Let $f\colon X \longrightarrow Y$ be uniformly continuous and let (x_n) be a Cauchy sequence in X. For all $n \in \mathbf{N}$, set $y_n = f(x_n)$.

Let $\varepsilon > 0$. As f is uniformly continuous, there exists $\delta > 0$ such that for all $x, x' \in X$, if $d_X(x, x') < \delta$ then $d_Y\big(f(x), f(x')\big) < \varepsilon$.

But (x_n) is Cauchy in X, so given this δ there exists $N \in \mathbf{N}$ such that $d_X(x_n, x_m) < \delta$ for all $n, m \geqslant N$. Therefore $d_Y\big(y_n, y_m\big) < \varepsilon$ for all $n, m \geqslant N$. $\square$

Proposition 2.61. *Let $f\colon X \longrightarrow Y$ be a continuous function between metric spaces. If X is compact, then f is uniformly continuous.*

Proof. Let $\varepsilon > 0$.

Given $x \in X$, there exists $\delta(x) > 0$ such that $f\big(\mathbf{B}_{\delta(x)}(x)\big) \subseteq \mathbf{B}_{\varepsilon/2}\big(f(x)\big)$. We get an open cover of X:

$$X \subseteq \bigcup_{x \in X} \mathbf{B}_{\delta(x)/2}(x),$$

which therefore has a finite subcover

$$X \subseteq \bigcup_{n=1}^{N} \mathbf{B}_{\delta(x_n)/2}(x_n).$$

Let $\delta = \min\{\delta(x_n)/2\colon n = 1, \dots, N\}$.

Suppose $s, t \in X$ are such that $d_X(s,t) < \delta$. We have $s \in \mathbf{B}_{\delta(x_n)/2}(x_n)$ for some $n \in \{1, \dots, N\}$. I claim that $t \in \mathbf{B}_{\delta(x_n)}(x_n)$:

$$d_X(t, x_n) \leqslant d_X(t,s) + d_X(s, x_n) < \delta + \frac{\delta(x_n)}{2} \leqslant \delta(x_n).$$

Therefore $f(s), f(t) \in \mathbf{B}_{\varepsilon/2}\big(f(x_n)\big)$, hence $d_Y(f(s), f(t)) < \varepsilon$. □

2.9. Completions

Any metric space can be embedded into a complete metric space. To make this precise, we say that a complete metric space $\big(\widehat{X}, \widehat{d}\,\big)$ is a *completion* of a metric space (X,d) if there exists a distance-preserving function $\iota\colon X \longrightarrow \widehat{X}$ such that $\iota(X)$ is a dense subset of $\widehat{X}$. (In particular, this implies that $\big(\iota(X), \widehat{d}\,\big)$ is isometric to (X,d).)

Theorem 2.62. *Any metric space (X,d) has a completion.*

We will see later (Corollary 2.64) that any two completions of (X,d) are isometric.

Proof. Given (X,d), consider the set $\mathcal{C}$ of all Cauchy sequences, equipped with the equivalence relation defined above Proposition 2.51.

Let $\widehat{X}$ be the resulting set of equivalence classes $[(x_n)]$. Define $\widehat{d}\colon \widehat{X} \times \widehat{X} \longrightarrow \mathbf{R}_{\geqslant 0}$ by:

$$\widehat{d}\big([(x_n)], [(y_n)]\big) = \lim_{n \longrightarrow \infty} d(x_n, y_n).$$

The limit exists as the sequence $\big(d(x_n, y_n)\big)$ is Cauchy in $\mathbf{R}$ (Proposition 2.56) and $\mathbf{R}$ is complete; moreover $\widehat{d}$ is well-defined, see Exercise 2.42.

It is easy to see that $\widehat{d}$ is a metric on $\widehat{X}$.

We have for all $x, y \in X$:

$$\widehat{d}\big(\iota(x), \iota(y)\big) = \lim_{n \longrightarrow \infty} d(x,y) = d(x,y),$$

so ι is distance-preserving.

To show that $\iota(X)$ is dense in $\widehat{X}$, let $[(x_n)] \in \widehat{X}$ and let $\varepsilon > 0$; we will show that there exists $x \in X$ such that $\widehat{d}\big(\iota(x), [(x_n)]\big) < \varepsilon$. As (x_n) is Cauchy, there exists $N \in \mathbf{N}$ such that $d(x_m, x_n) < \varepsilon/2$ for all $m, n \geqslant N$. Letting $x = x_N$, we have $d(x, x_n) < \varepsilon$ for all $n \geqslant N$, so taking limits:

$$\widehat{d}\big(\iota(x), (x_n)\big) = \lim_{n \longrightarrow \infty} d(x, x_n) \leqslant \frac{\varepsilon}{2} < \epsilon.$$

Let's check that the metric space $\big(\widehat{X}, \widehat{d}\,\big)$ is complete. Suppose (a_n) is a Cauchy sequence in $\widehat{X}$. As $\iota(X)$ is dense in $\widehat{X}$, for each $n \in \mathbf{N}$ there exists $x_n \in X$ such that $\widehat{d}(\iota(x_n), a_n) < \frac{1}{n}$. We get a sequence $(\iota(x_n)) \sim (a_n)$. As (a_n) is Cauchy in $\widehat{X}$, by Proposition 2.57 so is the sequence $(\iota(x_n))$ in $\widehat{X}$, and hence so is the sequence (x_n) in X as $\iota(X)$ is isometric to X. So we have an element $\widehat{x} := [(x_n)] \in \widehat{X}$.

I claim that (a_n) converges to $\widehat{x}$. Let $\varepsilon > 0$. We want to show that there exists $N \in \mathbf{N}$ such that for all $n \geqslant N$ we have

$$\widehat{d}\big(a_n, \widehat{x}\big) = \lim_{m \longrightarrow \infty} d\big(a_n(m), x_m\big) < \varepsilon.$$

Here $a_n \in \widehat{X}$, so it is represented by a Cauchy sequence $\big(a_n(m)\big)$ where the varying quantity is $m \in \mathbf{N}$.

We have by the triangle inequality

$$d\big(a_n(m), x_m\big) \leqslant d\big(a_n(m), x_n\big) + d(x_n, x_m),$$

so taking limits:

$$\lim_{m \longrightarrow \infty} d\big(a_n(m), x_m\big) \leqslant \lim_{m \longrightarrow \infty} d\big(a_n(m), x_n\big) + \lim_{m \longrightarrow \infty} d(x_n, x_m).$$

As (x_n) is Cauchy, there exists $N_1 \in \mathbf{N}$ such that $d(x_n, x_m) < \varepsilon/2$ for all $n, m \geqslant N_1$. Take $N_2 \in \mathbf{N}$ such that $1/N_2 < \varepsilon/2$ and $N = \max\{N_1, N_2\}$, then for all $n \geqslant N$ we have

$$\widehat{d}\big(a_n, \widehat{x}\big) \leqslant \widehat{d}\big(a_n, \iota(x_n)\big) + \lim_{m \longrightarrow \infty} d(x_n, x_m) < \frac{1}{n} + \frac{\varepsilon}{2} < \epsilon. \qquad \square$$

If $f\colon X \longrightarrow Y$ is some kind of function between metric spaces and $\widehat{X}$, $\widehat{Y}$ are completions of X, Y, we may ask whether f can be *extended* to a function of a similar kind $\widehat{f}\colon \widehat{X} \longrightarrow \widehat{Y}$. Since X is not actually a subset of $\widehat{X}$ (and similarly for Y), what we mean here is that we identify X with its isometric copy $\iota_X(X) \subseteq \widehat{X}$, and we identify Y with its isometric copy $\iota_Y(Y) \subseteq \widehat{Y}$. In other words, we say that a function $\widehat{f}\colon \widehat{X} \longrightarrow \widehat{Y}$ is an *extension* of $f\colon X \longrightarrow Y$ if

$$\widehat{f}\big(\iota_X(x)\big) = \iota_Y\big(f(x)\big) \qquad \text{for all } x \in X,$$

or, put more elegantly, if the following diagram commutes:

$$\begin{array}{ccc} X & \xrightarrow{\iota_X} & \widehat{X} \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle \widehat{f}} \\ Y & \xrightarrow{\iota_Y} & \widehat{Y} \end{array}$$

A reasonable first attempt would be to see if any **continuous** function $f\colon X \longrightarrow Y$ extends to a **continuous** function $\widehat{f}\colon \widehat{X} \longrightarrow \widehat{Y}$. It turns out that such a continuous extension may not exist (Exercise 2.43), but when it does, it is unique (this follows from the more general result of Question 4 on Assignment 1).

The following result assures us, however, that any **uniformly continuous** (resp. **distance-preserving**) function $f\colon X \longrightarrow Y$ extends uniquely to a **uniformly continuous** (resp. **distance-preserving**) function $\widehat{f}\colon \widehat{X} \longrightarrow \widehat{Y}$.

Proposition 2.63. *Let Z be a metric space and W a complete metric space. Let $D \subseteq Z$ be a dense subset and $f\colon D \longrightarrow W$ a uniformly continuous function.*

(a) The function f has a unique uniformly continuous extension to Z, that is there exists a unique uniformly continuous function

$$\widehat{f}\colon Z \longrightarrow W \quad \text{such that} \quad \widehat{f}(x) = f(x) \quad \text{for all } x \in D.$$

(b) If, in addition, f is distance-preserving, then so is the extension $\widehat{f}$.

Proof.

(a) The first task is to **construct the function** $\widehat{f}\colon Z \longrightarrow W$. Let $z \in Z$. Since D is dense in Z, there exists a sequence (x_n) in D such that $(x_n) \longrightarrow z$. In particular, (x_n) is Cauchy in D. Since $f\colon D \longrightarrow W$ is uniformly continuous, $\big(f(x_n)\big)$ is Cauchy in W. As W is complete, $\big(f(x_n)\big)$ has a limit $w \in W$.

Define $\widehat{f}(z) = w$.

Is this **well-defined**? We did make one choice in the construction, namely a sequence (x_n) in D that converges to z. Any other valid choice is a sequence (x'_n) in D with the same limit z, so $(x'_n) \sim (x_n)$. As f is continuous, we have $\big(f(x'_n)\big) \sim \big(f(x_n)\big)$, which implies that $\big(f(x'_n)\big) \longrightarrow w \in W$.

Is $\widehat{f}$ an **extension of** f? If $x \in D$ and we work through the above construction, we see that we can take $x_n = x$ for all $n \in \mathbf{N}$, so $f(x_n) = f(x)$ for all $n \in \mathbf{N}$, and finally $\widehat{f}(x) = w = f(x)$. In other words, $\widehat{f}(x) = f(x)$ for $x \in D$, as claimed.

Next we prove **uniform continuity** of $\widehat{f}$. Let $\varepsilon > 0$. Since $f\colon D \longrightarrow W$ is uniformly continuous, there exists $\delta > 0$ such that for all $x, x' \in D$, if $d_Z(x, x') < \delta$, then $d_W(f(x), f(x')) < \varepsilon/2$. Now suppose that $z, z' \in Z$ satisfy $d_Z(z, z') < \delta/3$. Let (x_n) be a sequence as in the definition of $\widehat{f}(z)$ above, and similarly with (x'_n) and $\widehat{f}(z')$. As $(x_n) \longrightarrow z$, there exists $N \in \mathbf{N}$ such that $d_Z(x_n, z) < \delta/3$ for all $n \geqslant N$. Similarly, as $(x'_n) \longrightarrow z'$, there exists $N' \in \mathbf{N}$ such that $d_Z(x'_n, z') < \delta/3$ for all $n \geqslant N'$. Letting $M = \max\{N, N'\}$ we get for all $n \geqslant M$:

$$d_Z(x_n, x'_n) \leqslant d_Z(x_n, z) + d_Z(z, z') + d_Z(z', x'_n) < \delta.$$

Therefore $d_W(f(x_n), f(x'_n)) < \varepsilon/2$ for all $n \geqslant M$.

As $\widehat{f}(z) = \lim f(x_n)$ and $\widehat{f}(z') = \lim f(x'_n)$, we conclude that

$$d_W(\widehat{f}(z), \widehat{f}(z')) \leqslant \frac{\varepsilon}{2} < \varepsilon.$$

The **uniqueness** of $\widehat{f}$ follows from Question 4 on Assignment 1, which says that there is at most one continuous extension.

(b) If f is **distance-preserving**, we use the same line of argument, only simpler. Let $(x_n) \longrightarrow z$, $(x'_n) \longrightarrow z'$ with $x_n, x'_n \in D$. Then

$$\begin{aligned} d_W\big(\widehat{f}(z), \widehat{f}(z')\big) &= d_W\Big(\lim_{n\longrightarrow\infty} \widehat{f}(x_n), \lim_{n\longrightarrow\infty} \widehat{f}(x'_n)\Big) \\ &= \lim_{n\longrightarrow\infty} d_W(f(x_n), f(x'_n)) = \lim_{n\longrightarrow\infty} d_Z(x_n, x'_n) = d_Z(z, z'). \end{aligned}$$ □

This has the following consequence:

Corollary 2.64. *Let X be a metric space.*

(a) Let Y be a metric space and fix completions $\big(\widehat{X}, \iota_X\big)$ of X and $\big(\widehat{Y}, \iota_Y\big)$ of Y. Any uniformly continuous (resp. distance-preserving) function $g\colon X \longrightarrow Y$ has a unique uniformly continuous (resp. distance-preserving) extension $\widehat{g}\colon \widehat{X} \longrightarrow \widehat{Y}$.

(b) Any two completions of X are isometric.

Proof.

(a) Let $D = \iota_X(X) \subseteq \widehat{X}$, and apply Proposition 2.63 to the function $\iota_Y \circ g \circ \iota_X^{-1} \colon D \longrightarrow \widehat{Y}$.

It is worth describing $\widehat{g} \colon \widehat{X} \longrightarrow \widehat{Y}$ more directly: given $\widehat{x} \in \widehat{X}$, let $\iota_X(x_n)$ be a sequence in the dense subset $\iota_X(X)$ that converges to $\widehat{x}$, then set

$$\widehat{g}(\widehat{x}) = \lim_{n \longrightarrow \infty} \iota_Y\big(g(x_n)\big). \tag{2.1}$$

(b) Let $\big(\widehat{X}_1, \widehat{d_1}\big)$ and $\big(\widehat{X}_2, \widehat{d_2}\big)$ be two completions.

We have isometries $\iota_1 \colon X \longrightarrow \iota_1(X) \subseteq \widehat{X}_1$ and $\iota_2 \colon X \longrightarrow \iota_2(X) \subseteq \widehat{X}_2$. Consider the composition $f := \iota_2 \circ \iota_1^{-1} \colon \iota_1(X) \longrightarrow \iota_2(X)$. It is an isometry, in particular it is distance-preserving, so by part (a) it extends uniquely to a distance-preserving function $\widehat{f} \colon \widehat{X}_1 \longrightarrow \widehat{X}_2$.

We check that $\widehat{f}$ is bijective. It is automatically injective since distance-preserving. For surjectivity, let $\widehat{x} \in \widehat{X}_2$ and let (x_n) be a sequence in X such that $\big(\iota_2(x_n)\big) \longrightarrow \widehat{x}$. Let $\widehat{x}_n = \iota_1(x_n)$. Since $\big(\iota_2(x_n)\big)$ converges, it is Cauchy. Since $\iota_2^{-1} \colon \iota_2(X) \longrightarrow X$ is an isometry, (x_n) is Cauchy in X. Since ι_1 is an isometry, $\big(\widehat{x}_n\big)$ is Cauchy in $\widehat{X}_1$. As the latter is complete, $\big(\widehat{x}_n\big) \longrightarrow \widehat{x}' \in \widehat{X}_1$. Therefore

$$\widehat{f}\big(\widehat{x}'\big) = \widehat{f}\Big(\lim_{n \longrightarrow \infty} \widehat{x}_n\Big) = \lim_{n \longrightarrow \infty} \widehat{f}\big(\widehat{x}_n\big) = \lim_{n \longrightarrow \infty} f\big(\iota_1(x_n)\big) = \lim_{n \longrightarrow \infty} \iota_2(x_n) = \widehat{x}. \qquad \square$$

2.10. Banach Fixed Point Theorem

Let (X, d_X) and (Y, d_Y) be metric spaces. A *contraction* is a function $f \colon X \longrightarrow Y$ for which there exists a constant $C \in [0, 1)$ such that

$$d_Y\big(f(x_1), f(x_2)\big) \leqslant C\, d_X(x_1, x_2) \qquad \text{for all } x_1, x_2 \in X.$$

It is easy to see (Exercise 2.44) that contractions are uniformly continuous.

A *fixed point* of a function $f \colon X \longrightarrow Y$ is an element $x \in X$ such that $f(x) = x$.

Proposition 2.65. *Let $f \colon X \longrightarrow X$ be a contraction from a metric space to itself. Then f has at most one fixed point.*

Proof. If x, x' are such that $x = f(x)$ and $x' = f(x')$, then

$$d(x, x') = d\big(f(x), f(x')\big) \leqslant C\, d(x, x').$$

If $x \neq x'$ then $d(x, x') > 0$ and

$$C\, d(x, x') < d(x, x') \quad \text{since} \quad 0 \leqslant C < 1,$$

leading to a contradiction. $\square$

We get a very useful result for complete metric spaces:

Theorem 2.66 (Banach Fixed Point Theorem)**.** *Let (X, d) be a nonempty complete metric space. Let $f \colon X \longrightarrow X$ be a contraction. Then f has a unique fixed point, that is an element $x \in X$ such that $f(x) = x$. Moreover, for any choice of $x_1 \in X$, the sequence (x_n) defined recursively by $x_{n+1} = f(x_n)$ converges to the fixed point x.*

Proof. The uniqueness statement follows from Proposition 2.65.

The proof of existence uses the hint in the last statement. Let $x_1 \in X$ and consider the sequence $(x_n) = \big(f^{\circ n}(x_1)\big)$. For any $m \geqslant 2$ we have

$$d(x_{m+1}, x_m) = d\big(f(x_m), f(x_{m-1})\big) \leqslant C\, d(x_m, x_{m-1}).$$

Applying this repeatedly with decreasing m, we get

$$d(x_{m+1}, x_m) \leqslant C^{m-1}\, d(x_2, x_1).$$

If we now go up from $m+1$ and apply this in conjunction with the triangle inequality, we get for all $n > m$:

$$\begin{aligned} d(x_n, x_m) &\leqslant \big(C^{n-2} + C^{n-3} + \cdots + C^{m-1}\big) d(x_2, x_1) \\ &\leqslant C^{m-1} \frac{1 - C^{n-m}}{1 - C}\, d(x_2, x_1) \\ &\leqslant C^{m-1} \frac{d(x_2, x_1)}{1 - C}. \end{aligned}$$

As $0 \leqslant C < 1$, we know that $C^{m-1} \longrightarrow 0$ as $m \longrightarrow \infty$, so we conclude that the sequence (x_n) is Cauchy. As X is complete, $(x_n) \longrightarrow x \in X$. But we can say more about this limit x, using the continuity of f:

$$f(x) = f\Big(\lim_{n\longrightarrow\infty} x_n\Big) = \lim_{n\longrightarrow\infty} f(x_n) = \lim_{n\longrightarrow\infty} x_{n+1} = x.$$

So x is indeed a fixed point of f. □

Recall the following result from real analysis:

Theorem 2.67 (Mean Value Theorem)**.** *Let $f\colon [a,b] \longrightarrow \mathbf{R}$ be continuous. If f is differentiable on (a,b), then there exists $\xi \in (a,b)$ such that*

$$f'(\xi) = \frac{f(b) - f(a)}{b - a}.$$

This turns out to be very useful in checking that a given function is a contraction:

Example 2.68. Verify that the function $f\colon [1,2] \longrightarrow \mathbf{R}$ defined by

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

has a unique fixed point, and find this point.

Solution. First we show that f is a contraction. We have

$$f'(x) = -\frac{x^2}{4} + 1,$$

and since $1 \leqslant x \leqslant 2$ it is easy to deduce that

$$0 \leqslant f'(x) \leqslant \frac{3}{4},$$

in particular $|f'(x)| \leqslant 3/4$ for all $x \in [1,2]$.

Now let $x_1, x_2 \in [1,2]$. Apply the Mean Value Theorem to f restricted to the interval $[x_1, x_2]$, and deduce that there exists $\xi \in (x_1, x_2) \subseteq [1,2]$ such that

$$|f(x_2) - f(x_1)| = |f'(\xi)|\,|x_2 - x_1| \leqslant \frac{3}{4}\,|x_2 - x_1|,$$

in other words f is a contraction with constant $3/4$.

In order to apply the Banach Fixed Point Theorem we need to know that f is a self-map, that is, that the image of f is contained in $[1,2]$. The global minimum and maximum of f occur either at the boundaries of the interval $[1,2]$, or at some stationary point in the interval. The only zero of $f'(x) = -\frac{x^2}{4} + 1$ in $[1,2]$ is $x = 2$, so we only need to evaluate f at 1 and 2:

$$f(1) = \frac{7}{6} \in [1,2], \qquad f(2) = \frac{19}{12} \in [1,2],$$

so indeed $f\big([1,2]\big) \subseteq [1,2]$.

The Banach Fixed Point Theorem tells us that f has a unique fixed point, which we can find directly by solving

$$x = f(x) = -\frac{1}{12}x^3 + x + \frac{1}{4} \Rightarrow x^3 = 3 \Rightarrow x = \sqrt[3]{3}.$$

Note that this gives us a recursively-defined sequence of rational numbers that converges to $\sqrt[3]{3}$: take $x_1 = 1$ and apply f iteratively, $x_{n+1} = f(x_n)$. □

2.11. Boundedness and compactness in metric spaces

Let (X, d) be a metric space. In this section we will introduce a number of equivalent conditions for a subset $K \subseteq X$ to be compact.

The *diameter* of a nonempty[3] subset $S \subseteq X$ is by definition

$$\operatorname{diam}(S) := \sup\big\{d(x,y) : x, y \in S\big\}.$$

If this is a (finite) real number we say that S is *bounded*. This is equivalent to saying that S is contained in some closed ball with finite radius (see Exercise 2.45). Otherwise we say that S is *unbounded*.

Example 2.69. Let $S \subseteq \mathbf{R}$ be a bounded set. Show that for any $\varepsilon > 0$, there exist $N \in \mathbf{N}$ and open balls $B_1, \dots, B_N$, all of radius ε, such that

$$S \subseteq \bigcup_{n=1}^{N} B_n.$$

Solution. As S is bounded, it is contained in some closed ball, which in $\mathbf{R}$ is some interval $[x, y]$. So it suffices to prove that the conclusion holds for the interval $[x, y]$, which is straightforward: given $\varepsilon > 0$, let $N \in \mathbf{N}$ be such that $N \geqslant \frac{y-x}{\varepsilon}$, then

$$S \subseteq [x,y] \subseteq \bigcup_{n=1}^{N} \big[x + (n-1)\varepsilon, x + n\varepsilon\big] \subseteq \bigcup_{n=1}^{N} \mathbf{B}_\varepsilon\big(x + (2n-1)\varepsilon/2\big). \qquad □$$

[3] Surprisingly, what the diameter of $\varnothing$ should be appears to be a controversial topic. I will steer clear of it.

The property in the last example is called total boundedness: a subset $S \subseteq X$ of a metric space is *totally bounded* if for all $\varepsilon > 0$, there exist $N \in \mathbf{N}$ and $x_1, \ldots, x_N \in X$ such that

$$S \subseteq \bigcup_{n=1}^{N} \mathbf{B}_\varepsilon(x_n).$$

If this makes you think of compact sets, it is not a coincidence: it is easy to see that any compact subset $K \subseteq X$ of a metric space is totally bounded (given $\varepsilon > 0$, cover K with open balls of radius ε centred at each point of K and use compactness).

As you can see in Exercise 2.47, any totally bounded set is bounded; Example 2.69 says that the converse is true if $X = \mathbf{R}$. See Exercise 2.55 for the fact that the product of two totally bounded sets is totally bounded, and Exercise 2.56 for the consequence that in $\mathbf{R}^m$, every bounded set is totally bounded.

Proposition 2.70. *Let $f\colon X \longrightarrow \mathbf{R}$ be a continuous function, where X is a compact metric space. Then the image $f(X)$ is bounded, and the bounds are attained: there exist $x_{min}, x_{max} \in X$ such that*

$$f(x_{min}) \leqslant f(x) \leqslant f(x_{max}) \qquad \text{for all } x \in X.$$

Proof. By Proposition 2.37, $f(X)$ is a compact subset of $\mathbf{R}$. Therefore $f(X)$ is totally bounded, hence bounded. So $f(X)$ has both infimum and supremum, which are boundary points. But $f(X)$ is also closed by Proposition 2.35, therefore it contains its boundary points and hence the infimum and supremum. □

Proposition 2.71. *If $f\colon X \longrightarrow Y$ is a uniformly continuous function between metric spaces and $S \subseteq X$ is totally bounded, then $f(S) \subseteq Y$ is totally bounded.*

Proof. Let $\varepsilon > 0$. As f is uniformly continuous, there exists $\delta > 0$ such that for all $x \in X$ we have

$$f\big(\mathbf{B}_\delta(x)\big) \subseteq \mathbf{B}_\varepsilon\big(f(x)\big).$$

As S is totally bounded, there are open balls $\mathbf{B}_\delta(x_1), \ldots, \mathbf{B}_\delta(x_N)$ such that

$$S \subseteq \bigcup_{j=1}^{N} \mathbf{B}_\delta(x_j),$$

so applying f on both sides we get

$$f(S) \subseteq f\left(\bigcup_{j=1}^{N} \mathbf{B}_\delta(x_j)\right) = \bigcup_{j=1}^{N} f\big(\mathbf{B}_\delta(x_j)\big) \subseteq \bigcup_{j=1}^{N} \mathbf{B}_\varepsilon\big(f(x_j)\big).$$ □

We say that a topological space X is *separable* if it contains a countable dense subset. For instance, $\mathbf{R}^n$ is separable for any $n \in \mathbf{N}$, with $\mathbf{Q}^n$ as countable dense subset.

Proposition 2.72. *Any totally bounded metric space X is separable.*

Proof. For a fixed $n \in \mathbf{Z}_{\geqslant 1}$, cover X with a finite number of open balls of radius $\frac{1}{n}$ and let $D_n \subseteq X$ be the set of centres of these balls. Now let

$$D = \bigcup_{n=1}^{\infty} D_n.$$

This is a countable union of finite sets, hence countable.

Now take $x \in X$ and $\varepsilon > 0$. Let $n \in \mathbf{N}$ be such that $\frac{1}{n} < \varepsilon$. Since X is covered by the open balls of radius $\frac{1}{n}$ centred at elements of D_n, there exists $y \in D_n \subseteq D$ such that $x \in \mathbf{B}_{1/n}(y)$, that is $d(x, y) < \frac{1}{n} < \varepsilon$. So D is dense in X. □

Proposition 2.73. *A subset $S \subseteq X$ of a metric space is totally bounded if and only if every sequence in S has a Cauchy subsequence.*

Proof. Let (s_n) be a sequence in S.

Take a finite cover of S by open balls of radius 1. At least one of these open balls $\mathbf{B}_1(x_1)$ contains infinitely many terms of (s_n); let $\big(s_n^{(1)}\big) = (s_n) \cap \mathbf{B}_1(x_1)$.

Take a finite cover of S by open balls of radius 1/2. As least one of these balls $\mathbf{B}_{1/2}(x_2)$ contains infinitely many terms of $\big(s_n^{(1)}\big)$; let $\big(s_n^{(2)}\big) = \big(s_n^{(1)}\big) \cap \mathbf{B}_{1/2}(x_2)$.

Continuing in this manner, we get a list of successive subsequences $\big(s_n^{(j)}\big) \subseteq \mathbf{B}_{1/j}(x_j)$:

$$
\begin{array}{l|llllllll}
(s_n): & s_1, & s_2, & s_3, & s_4, & \dots & & \dots & \in S \\
\hline
\big(s_n^{(1)}\big): & s_1^{(1)}, & s_2^{(1)}, & s_3^{(1)}, & s_4^{(1)}, & \dots & & \dots & \in \mathbf{B}_1(x_1) \\
\big(s_n^{(2)}\big): & s_1^{(2)}, & s_2^{(2)}, & s_3^{(2)}, & s_4^{(2)}, & \dots & & \dots & \in \mathbf{B}_{1/2}(x_2) \\
\big(s_n^{(3)}\big): & s_1^{(3)}, & s_2^{(3)}, & s_3^{(3)}, & s_4^{(3)}, & \dots & & \dots & \in \mathbf{B}_{1/3}(x_3) \\
\vdots & & & & & & & & \vdots \\
\big(s_n^{(j)}\big): & s_1^{(j)}, & s_2^{(j)}, & s_3^{(j)}, & s_4^{(j)}, & \dots & s_j^{(j)}, & \dots & \in \mathbf{B}_{1/j}(x_j) \\
\vdots & & & & & & & & \vdots
\end{array}
$$

From this list we extract the diagonal, giving rise to a subsequence $\big(s_n^{(n)}\big)$ of (s_n). I claim that $\big(s_n^{(n)}\big)$ is a Cauchy sequence.

Given $\varepsilon > 0$, let $N \in \mathbf{N}$ be such that $2/N \leqslant \varepsilon$. For $i \geqslant j \geqslant N$ we have

$$s_j^{(j)}, s_i^{(i)} \in \big(s_n^{(j)}\big) \subseteq \big(s_n^{(N)}\big) \subseteq \mathbf{B}_{1/N}(x_N) \subseteq \mathbf{B}_{\varepsilon/2}(x_N),$$

hence

$$d\big(s_j^{(j)}, s_i^{(i)}\big) \leqslant d\big(s_j^{(j)}, x_N\big) + d\big(x_N, s_i^{(i)}\big) < \varepsilon.$$

In the other direction, let $\varepsilon > 0$. Choose an arbitrary $s_1 \in S$. If $S \subseteq \mathbf{B}_\varepsilon(s_1)$, we are done. Otherwise, there exists $s_2 \in S \setminus \mathbf{B}_\varepsilon(s_1)$. If $S \subseteq \mathbf{B}_\varepsilon(s_1) \cup \mathbf{B}_\varepsilon(s_2)$, we are done. Otherwise, there exists $s_3 \in S \setminus \big(\mathbf{B}_\varepsilon(s_1) \cup \mathbf{B}_\varepsilon(s_2)\big)$.

Suppose that this process does not stop after finitely many steps, then we obtain a sequence (s_n) in S with the property that $d(s_n, s_m) \geqslant \varepsilon$ for all $n, m \in \mathbf{N}$, so that (s_n) has no Cauchy subsequence, contradiction. □

A *Lebesgue number* of an open cover

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

is a real number $\delta > 0$ such that for any subset $A \subseteq K$ with $\operatorname{diam}(A) < \delta$, there exists $i \in I$ such that $A \subseteq U_i$.

It is the case that any open cover of a sequentially compact subset $K \subseteq X$ has a Lebesgue number, see Exercise 2.57.

The following is the main result of the section, an amalgamation of various theorems attributed to Heine–Borel, Bolzano–Weierstrass, and very possibly others.

Theorem 2.74. *Let K be a subset of a metric space X. The following are equivalent:*

(a) K is compact.

(b) K is complete and totally bounded.

(c) K is sequentially compact, *that is every sequence in K has a subsequence that converges to an element of K.*

Proof. **(a)⇒(b):** Suppose K is compact. We have already seen that K is totally bounded. Let $\iota\colon K \longrightarrow \widehat{K}$ be a completion of K. Then $\iota(K)$ is a compact subset of $\widehat{K}$, hence closed by Proposition 2.35. But $\iota(K)$ is also dense in $\widehat{K}$, so $\iota(K) = \widehat{K}$ and K is complete.

(b)⇒(c): Suppose K is complete and totally bounded and let (x_n) be a sequence in K. Since K is totally bounded, (x_n) has a Cauchy subsequence by Proposition 2.73, which converges in K, since K is complete.

(c)⇒(a): Suppose K is sequentially compact and consider an open cover

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

By Exercise 2.57 this cover has a Lebesgue number $\delta > 0$. By Proposition 2.73, K is totally bounded, so it has a finite cover by open balls of radius $\delta/2$:

$$K \subseteq B_1 \cup \cdots \cup B_n.$$

For each $j = 1, \dots, n$ we have $\operatorname{diam}(K \cap B_j) < \delta$ so there exists $i_j \in I$ such that $K \cap B_j \subseteq U_{i_j}$. Overall we get a finite subcover

$$K \subseteq U_{i_1} \cup \cdots \cup U_{i_n}.$$

□

2.12. Spaces of bounded continuous functions

Let X be a set and Y a metric space.

A function $f\colon X \longrightarrow Y$ is *bounded* if there exists $y \in Y$ and $M \in \mathbf{R}$ such that

$$d_Y\big(y, f(x)\big) \leqslant M \qquad \text{for all } x \in X.$$

Equivalently, the direct image $f(X)$ is a bounded subset of Y, see Exercise 2.48.

Let $B(X,Y)$ denote the set of all bounded functions $X \longrightarrow Y$. For $f, g \in B(X,Y)$ define

$$d_\infty(f,g) = \sup_{x\in X}\big\{d_Y\big(f(x), g(x)\big)\big\}.$$

Proposition 2.75. *The function d_∞ is a metric on $B(X,Y)$, called the* uniform metric.

Proof. First we check that d_∞ takes values in $\mathbf{R}_{\geqslant 0}$: if $f, g \in B(X,Y)$, there exist $y_f, y_g \in Y$ and $M_f, M_g \in \mathbf{R}$ such that

$$d_Y\big(y_f, f(x)\big) \leqslant M_f \quad \text{and} \quad d_Y\big(y_g, g(x)\big) \leqslant M_g \qquad \text{for all } x \in X.$$

Letting $M = d_Y(y_f, y_g)$ we see that for all $x \in X$ we have

$$d_Y(f(x), g(x)) \leqslant d_Y(f(x), y_f) + d_Y(y_f, y_g) + d_Y(y_g, g(x)) \leqslant M_f + M + M_g.$$

As $M_f + M + M_g$ is a finite upper bound for the set in the definition of d_∞, we conclude that the supremum is finite as well.

The symmetry of d_∞ follows directly from the symmetry of d_Y.

For the triangle inequality, let $h \in B(X,Y)$ and note that for all $x \in X$ we have

$$d_Y(f(x), g(x)) \leqslant d_Y(f(x), h(x)) + d_Y(h(x), g(x)).$$

By the upper bound property of the supremum we get that for all $x \in X$

$$d_Y(f(x), g(x)) \leqslant d_\infty(f,h) + d_\infty(h,g).$$

By the minimality of the supremum we get

$$d_\infty(f,g) \leqslant d_\infty(f,h) + d_\infty(h,g).$$

For the non-degeneracy of d_∞, note that if $d_\infty(f,g) = 0$ then

$$\sup_{x \in X}\{d_Y(f(x), g(x))\} = 0,$$

so by the non-negativity of d_Y we get that $d_Y(f(x), g(x)) = 0$ for all $x \in X$. Therefore $f(x) = g(x)$ for all $x \in X$, hence $f = g$. □

We say that a sequence (f_n) in $B(X,Y)$ *converges pointwise* to a function $f\colon X \longrightarrow Y$ if, for every $x \in X$, the sequence $\big(f_n(x)\big)$ in Y converges to $f(x) \in Y$:

given $x \in X$ and $\varepsilon > 0$, there exists $N = N(x,\varepsilon) \in \mathbf{N}$ s.t. $d_Y(f_n(x), f(x)) < \varepsilon$ for all $n \geqslant N$.

Example 2.76. The pointwise limit of a sequence of bounded functions need not be bounded.

For instance, take $f_n\colon \mathbf{R}_{\geqslant 0} \longrightarrow \mathbf{R}$ given by

$$f_n(x) = \begin{cases} x & \text{if } x \leqslant n \\ 0 & \text{otherwise.} \end{cases}$$

Then f_n is bounded as $|f_n(x)| \leqslant n$ for all $x \in \mathbf{R}_{\geqslant 0}$, but the pointwise limit is $f(x) = x$, which is not bounded on $\mathbf{R}_{\geqslant 0}$.

We say that a sequence (f_n) in $B(X,Y)$ *converges uniformly* to a function $f\colon X \longrightarrow Y$ if:

given $\varepsilon > 0$, there exists $N = N(\varepsilon) \in \mathbf{N}$ s.t. $d_Y(f_n(x), f(x)) < \varepsilon$ for all $n \geqslant N$ and all $x \in X$.

Proposition 2.77. *Let X be a set and Y a metric space.*

(a) The uniform limit f of a sequence (f_n) of bounded functions $X \longrightarrow Y$ is bounded.

(b) A sequence (f_n) in $B(X,Y)$ converges uniformly to $f \in B(X,Y)$ if and only if $(f_n) \longrightarrow f$ with respect to the uniform metric d_∞ on $B(X,Y)$.

Proof.

(a) Let $\varepsilon = 1$ and consider the corresponding $N \in \mathbf{N}$. Since f_N is bounded, there exist $y \in Y$ and $M \in \mathbf{R}$ such that

$$d_Y(y, f_N(x)) \leqslant M \qquad \text{for all } x \in X.$$

Therefore, for all $x \in X$ we have

$$d_Y(y, f(x)) \leqslant d_Y(y, f_N(x)) + d_Y(f_N(x), f(x)) \leqslant M + 1,$$

which shows that f is bounded.

(b) See Exercise 2.49. □

Proposition 2.78. *Given a set X and a metric space Y, if Y is complete then the metric space $B(X,Y)$ (with the uniform metric d_∞) is complete.*

Proof. Let (f_n) be a Cauchy sequence in $B(X,Y)$. We define $f\colon X \longrightarrow Y$ as follows.

Given $\varepsilon > 0$, there exists $N \in \mathbf{N}$ such that for all $m,n \geqslant N$ we have $d_\infty(f_n, f_m) < \varepsilon/2$, that is

$$d_Y(f_n(x), f_m(x)) < \frac{\varepsilon}{2} < \varepsilon \qquad \text{for all } x \in X.$$

In particular, for any $x \in X$ the sequence $(f_n(x))$ is Cauchy in Y, which is complete, so we can define $f(x)$ to be its limit.

It remains to prove that (f_n) converges to f uniformly. Given $\varepsilon > 0$, take $N \in \mathbf{N}$ exactly as in the previous paragraph and let $n \geqslant N$. Given $x \in X$, let $m(x) \geqslant N$ be such that $d_Y(f_{m(x)}(x), f(x)) < \varepsilon/2$, then

$$d_Y(f_n(x), f(x)) \leqslant d_Y(f_n(x), f_{m(x)}(x)) + d_Y(f_{m(x)}(x), f(x)) < \varepsilon.$$

The conclusion is that $d_Y(f_n(x), f(x)) < \varepsilon$ for all $n \geqslant N$, so $(f_n) \longrightarrow f$.

As we have shown that f is the uniform limit of the sequence of bounded functions (f_n), f is bounded by Proposition 2.77. ☐

Suppose now that both X and Y are metric spaces. Let $C_0(X,Y)$ denote the subset of $B(X,Y)$ consisting of all bounded continuous functions $X \longrightarrow Y$.

Proposition 2.79. *Given metric spaces X and Y, $C_0(X,Y)$ is a closed subset of $B(X,Y)$ with the uniform metric d_∞. In other words, the uniform limit of a sequence of bounded continuous functions is a bounded continuous function.*

Proof. Let $(f_n) \longrightarrow f$ with respect to the uniform norm, where $f_n \in C_0(X,Y)$ for all $n \in \mathbf{N}$. Fix $x_0 \in X$. Given $\varepsilon > 0$, there exists $N \in \mathbf{N}$ such that if $n \geqslant N$ then

$$d_Y(f_n(x), f(x)) < \varepsilon/3 \qquad \text{for all } x \in X.$$

Let $\delta > 0$ be such that

$$d_Y(f_N(x_0), f_N(x)) < \varepsilon/3 \qquad \text{for all } x \in X \text{ such that } d_X(x_0, x) < \delta.$$

We then have that for any $x \in X$ such that $d_X(x_0, x) < \delta$:

$$d_Y(f(x_0), f(x)) \leqslant d_Y(f(x_0), f_N(x_0)) + d_Y(f_N(x_0), f_N(x)) + d_Y(f_N(x), f(x)) < \epsilon.$$ ☐

Example 2.80. The pointwise limit of a sequence of bounded continuous functions need not be continuous.

For $n \in \mathbf{N}$, take $f_n\colon [0,1] \longrightarrow \mathbf{R}$ given by $f_n(x) = x^n$, then the pointwise limit is

$$f\colon [0,1] \longrightarrow \mathbf{R}, \qquad f(x) = \begin{cases} 0 & \text{if } 0 \leqslant x < 1 \\ 1 & \text{if } x = 1, \end{cases}$$

which is clearly not continuous.

2.13. Function spaces: compactness

In this section we specialise to the case where X is a compact metric space and $Y = \mathbf{R}^m$, and consider the space $C_0(X, \mathbf{R}^m)$ of (bounded[4]) continuous functions $X \longrightarrow \mathbf{R}^m$.

[4]Since X is compact, every continuous function is automatically bounded.

Our aim is to give necessary and sufficient conditions for a subset of $C_0(X, \mathbf{R}^m)$ to be compact. These conditions will turn out to be: closed, bounded, and equicontinuous.

We say that a collection F of functions $X \longrightarrow Y$ between metric spaces is *equicontinuous* if given $\varepsilon > 0$ there exists $\delta > 0$ such that for all $f \in F$ and all $x_1, x_2 \in X$ with $d_X(x_1, x_2) < \delta$ we have $d_Y(f(x_1), f(x_2)) < \varepsilon$.

For instance, a singleton $F = \{f\}$ is equicontinuous if and only if f is uniformly continuous.

Example 2.81. The set F of all contractions $X \longrightarrow Y$ is equicontinuous: given $\varepsilon > 0$, let $\delta = \varepsilon$. For any $f \in F$ there exists $C_f \in [0, 1)$ such that

$$d_Y(f(x_1), f(x_2)) \leqslant C_f d_X(x_1, x_2) < d_X(x_1, x_2) < \delta = \varepsilon.$$

Proposition 2.82. *Let X be a totally bounded metric space and Y a complete metric space. Suppose (f_n) is an equicontinuous sequence in $C_0(X, Y)$ such that $\big(f_n(z)\big)$ converges in Y for every z in a dense subset Z of X. Then (f_n) converges uniformly in $C_0(X, Y)$.*

Proof. Since Y is complete, so is $C_0(X, Y)$ by Propositions 2.78 and 2.79. Therefore it suffices to show that the sequence (f_n) is Cauchy in $C_0(X, Y)$.

Let $\varepsilon > 0$. Since (f_n) is equicontinuous, there exists $\delta > 0$ such that for all $n \in \mathbf{N}$ and all $x_1, x_2 \in X$ with $d_X(x_1, x_2) < \delta$ we have $d(f_n(x_1), f_n(x_2)) < \varepsilon/4$.

Let

$$X \subseteq \mathbf{B}_{\delta/2}(x_1) \cup \cdots \cup \mathbf{B}_{\delta/2}(x_k)$$

be a finite open cover of X by open balls of radius $\delta/2$. Since Z is dense in X, for each $i = 1, \ldots, k$ there exists $z_i \in Z \cap \mathbf{B}_{\delta/2}(x_i)$, so that $\mathbf{B}_{\delta/2}(x_i) \subseteq \mathbf{B}_\delta(z_i)$ and

$$X \subseteq \mathbf{B}_\delta(z_1) \cup \cdots \cup \mathbf{B}_\delta(z_k).$$

The sequences $(f_n(z_1)), \ldots, (f_n(z_k))$ are convergent, hence Cauchy, so there exists $N \in \mathbf{N}$ such that for all $n, m \geqslant N$ we have

$$d(f_n(z_i), f_m(z_i)) < \frac{\varepsilon}{4} \qquad \text{for } i = 1, \ldots, k.$$

Given $x \in X$, there exists $i = 1, \ldots, k$ such that $x \in \mathbf{B}_\delta(z_i)$. For all $n, m \geqslant N$ we have

$$d(f_n(x), f_m(x)) \leqslant d(f_n(x), f_n(z_i)) + d(f_n(z_i), f_m(z_i)) + d(f_m(z_i), f_m(x)) < \frac{3\varepsilon}{4}.$$

Therefore

$$d_\infty(f_n, f_m) = \sup_{x \in X}\{d(f_n(x), f_m(x))\} \leqslant \frac{3\varepsilon}{4} < \varepsilon,$$

so the sequence (f_n) is indeed Cauchy. □

Proposition 2.83. *Let X be a metric space and let Z be a countable subset of X. Then every bounded sequence (f_n) in $C_0(X, \mathbf{R}^m)$ has a subsequence (f_{n_k}) such that $(f_{n_k}(z))$ converges in $\mathbf{R}^m$ for every $z \in Z$.*

Proof. Enumerate $Z = \{z_1, z_2, \ldots\}$.

The sequence $(f_n(z_1))$ is bounded in $\mathbf{R}^m$, hence has a convergent subsequence $(f_{n_k^1}(z_1))$.

The sequence $(f_{n_k^1}(z_2))$ is bounded in $\mathbf{R}^m$, hence has a convergent subsequence $(f_{n_k^2}(z_2))$.

We continue in this manner. At the j-th step, we get a subsequence $(f_{n_k^j})$ of $(f_{n_k^{j-1}})$ such that $(f_{n_k^j}(z_i))$ converges for $i = 1, 2, \dots, j$:

$$\begin{array}{lccccccll}
(f_n): & f_1, & f_2, & f_3, & f_4, & \dots & & \text{s.t.} & (f_n) \subseteq C_0(X, \mathbf{R}^m) \\
\hline
(f_{n_k^1}): & f_{n_1^1}, & f_{n_2^1}, & f_{n_3^1}, & f_{n_4^1}, & \dots & & \text{s.t.} & (f_{n_k^1}(z_1)) \text{ converges} \\
(f_{n_k^2}): & f_{n_1^2}, & f_{n_2^2}, & f_{n_3^2}, & f_{n_4^2}, & \dots & & \text{s.t.} & (f_{n_k^2}(z_2)) \text{ converges} \\
(f_{n_k^3}): & f_{n_1^3}, & f_{n_2^3}, & f_{n_3^3}, & f_{n_4^3}, & \dots & & \text{s.t.} & (f_{n_k^3}(z_3)) \text{ converges} \\
\vdots & & & & & & & & \vdots \\
(f_{n_k^j}): & f_{n_1^j}, & f_{n_2^j}, & f_{n_3^j}, & f_{n_4^j}, & \dots & f_{n_j^j} & \text{s.t.} & (f_{n_k^j}(z_j)) \text{ converges} \\
\vdots & & & & & & & & \vdots
\end{array}$$

We turn these nested subsequences into the subsequence desired in the statement by the diagonal argument we used in Proposition 2.73: let f_{n_1} be the first term of the sequence $(f_{n_k^1})$, let f_{n_2} be the second term of the sequence $(f_{n_k^2})$, etc.

Given $j \in \mathbf{N}$, $(f_{n_k}(z_j))$ converges, since after ignoring the first j terms, (f_{n_k}) is a subsequence of $(f_{n_k^j})$. Since this holds for all j, we get that $(f_{n_k}(z))$ converges for every $z \in Z$. □

Theorem 2.84 (Arzelà–Ascoli)**.** *If X is a totally bounded metric space and $K \subseteq C_0(X, \mathbf{R}^m)$ is a bounded, closed, and equicontinuous subset, then K is compact.*

Proof. Let (f_n) be a sequence in K, then (f_n) is bounded and equicontinuous. Since X is totally bounded, it is separable by Proposition 2.72; let Z be a countable dense subset. By Proposition 2.83, (f_n) has a subsequence (f_{n_k}) that converges at every $z \in Z$. By Proposition 2.82, (f_{n_k}) converges in $C_0(X, \mathbf{R}^m)$. Since K is closed, (f_{n_k}) converges to an element of K.

By Theorem 2.74, K is compact. □

If in Theorem 2.84 we require that X be compact (which is how the Arzelà–Ascoli Theorem is usually stated), then the converse also holds: every compact subset $K \subseteq C_0(X, \mathbf{R}^m)$ is bounded, closed, and equicontinuous. See Exercise 2.59.

Another useful class of results involving function spaces describes certain nice dense subsets (for instance, the Weierstrass Approximation Theorem says that polynomials are dense in the space of continuous functions on a closed interval). We will return to this in the following chapter, once we have established some of the language of normed spaces.

3. Normed and Hilbert spaces

After a long detour into the world of sets with a distance function (that is, metric spaces), or more generally with a notion of neighbourhoods of points (that is, topological spaces), we return to the setting of vector spaces and investigate some consequences of endowing these with a notion of distance. This can done in many ways, but we will be interested in pursuing distance functions that are compatible with the vector space structure (just as we tend to study functions between vector spaces that are compatible with the vector space structure, in other words, linear transformations). Such considerations (and a look back at the properties of Euclidean distance in $\mathbf{R}^n$, which we are hoping to emulate and generalise) lead us to the notion of norm defined below, and the associated distance function.

Notation

In this chapter, $\mathbf{F}$ will denote one of the fields $\mathbf{R}$, $\mathbf{C}$, each endowed with its Euclidean metric. The function $\alpha \longmapsto |\alpha|$ is the real or complex absolute value, as appropriate. The function $\alpha \longmapsto \overline{\alpha}$ is the complex conjugation, which restricts to the identity function if $\mathbf{F} = \mathbf{R}$.

Given subsets S, T of a vector space V over $\mathbf{F}$ and $\alpha \in \mathbf{F}$, we write

$$S + T = \{s + t : s \in S, t \in T\},$$
$$\alpha S = \{\alpha s : s \in S\}.$$

3.1. Norms

Let V be a vector space over $\mathbf{F}$.

A *norm* on V is a function

$$\|\cdot\| : V \longrightarrow \mathbf{R}_{\geqslant 0}$$

such that

(a) $\|v + w\| \leqslant \|v\| + \|w\|$ for all $v, w \in V$;

(b) $\|\alpha v\| = |\alpha| \, \|v\|$ for all $v \in V$, $\alpha \in \mathbf{F}$;

(c) $\|v\| = 0$ if and only if $v = 0$.

(If we remove (c), we get what is called a *semi-norm*.)

The pair $(V, \|\cdot\|)$ is called a *normed space*.

Proposition 3.1. *Let $(V, \|\cdot\|)$ be a normed space. Define $d : V \times V \longrightarrow \mathbf{R}_{\geqslant 0}$ by*

$$d(v, w) = \|v - w\|.$$

Then d is a metric on V, and satisfies the following additional properties:

(d) $d(v + u, w + u) = d(v, w)$ for all $u, v, w \in V$;

(e) $d(\alpha v, \alpha w) = |\alpha| \, d(v, w)$ for all $v, w \in V$, $\alpha \in \mathbf{F}$.

So every normed space is a metric space.

Proof.

(a) $d(w,v) = \|w-v\| = \|(-1)(v-w)\| = |-1|\,\|v-w\| = d(v,w)$;

(b) $d(v,u) + d(u,w) = \|v-u\| + \|u-w\| \geqslant \|v-u+u-w\| = \|v-w\| = d(v,w)$;

(c) $d(v,w) = 0$ iff $\|v-w\| = 0$ iff $v-w=0$ iff $v=w$;

(d) $d(v+u, w+u) = \|v+u-w-u\| = \|v-w\| = d(v,w)$;

(e) $d(\alpha v, \alpha w) = \|\alpha v - \alpha w\| = |\alpha|\,\|v-w\| = |\alpha|\,d(v,w)$. □

It is easy to see that the norm $V \longrightarrow \mathbf{R}_{\geqslant 0}$, $v \longmapsto \|v\|$, is a uniformly continuous function with respect to the metric defined by the norm on V, and the Euclidean metric on $\mathbf{R}_{\geqslant 0}$, see Exercise 3.1.

Suppose $\|\cdot\|_1$ and $\|\cdot\|_2$ are norms on a vector space V. We say that they are *equivalent* if there exist $m, M > 0$ such that

$$m\|v\|_1 \leqslant \|v\|_2 \leqslant M\|v\|_1 \qquad \text{for all } v \in V.$$

Equivalent norms on V give rise to equivalent metrics on V (and therefore to the same topology on V), see Exercise 3.2.

If W is a subspace of a normed space $\big(V, \|\cdot\|\big)$, we always endow W with the restriction of $\|\cdot\|$ to W, which is a norm on W.

Proposition 3.2. *Any normed space* $\big(V, \|\cdot\|\big)$ *is a* topological vector space, *that is a vector space such that*

(a) the vector addition $a\colon V \times V \longrightarrow V$, $a(v,w) = v + w$, *is a continuous function;*

(b) the scalar multiplication $s\colon \mathbf{F} \times V \longrightarrow V$, $s(\alpha, v) = \alpha v$, *is a continuous function.*

(Continuity is defined with respect to the product topologies on $V \times V$ *and on* $\mathbf{F} \times V$*.)*

Proof. Since the topology on V is generated by the set of open balls, in both cases it suffices to take an arbitrary open ball $\mathbf{B}_\varepsilon(x)$ and show that its inverse image is open; we do this by taking an arbitrary element of this inverse image and fitting an appropriately small open rectangle around it.

(a) Let $(v_0, w_0) \in a^{-1}(\mathbf{B}_\varepsilon(x))$, then letting $r = \|v_0 + w_0 - x\|$ we have $r < \varepsilon$.

Take $\delta_1, \delta_2 > 0$ such that $\delta_1 + \delta_2 = \varepsilon - r$. (For instance we could let each of them be half of $\varepsilon - r$.)

We check that the open rectangle $\mathbf{B}_{\delta_1}(v_0) \times \mathbf{B}_{\delta_2}(w_0) \subseteq a^{-1}(\mathbf{B}_\varepsilon(x))$: for any (v,w) in the rectangle we have

$$\|v + w - x\| \leqslant \|v - v_0\| + \|w - w_0\| + \|v_0 + w_0 - x\| < \delta_1 + \delta_2 + r = \varepsilon.$$

(b) This is slightly more delicate.

Let $(\alpha_0, v_0) \in s^{-1}(\mathbf{B}_\varepsilon(x))$, then letting $r = \|\alpha_0 v_0 - x\|$ we have $r < \varepsilon$.

Before we start in earnest, let's note

$$\begin{aligned}\|\alpha_0 v_0 - \alpha v\| &\leqslant \|\alpha_0 v_0 - \alpha v_0\| + \|\alpha v_0 - \alpha v\| \\ &= |\alpha_0 - \alpha|\,\|v_0\| + |\alpha|\,\|v_0 - v\| \\ &\leqslant |\alpha_0 - \alpha|\,\|v_0\| + |\alpha_0|\,\|v_0 - v\| + |\alpha_0 - \alpha|\,\|v_0 - v\| \\ &= |\alpha_0 - \alpha|\big(\|v_0\| + \|v_0 - v\|\big) + |\alpha_0|\,\|v_0 - v\|.\end{aligned}$$

Set:

$$\begin{aligned} &\text{if } \alpha_0 = 0: &\quad \delta_2 &= 1 &\quad \delta_1 &= \frac{\varepsilon - r}{\|v_0\| + \delta_2} \\ &\text{if } \alpha_0 \neq 0: &\quad \delta_2 &= \frac{\varepsilon - r}{2|\alpha_0|} &\quad \delta_1 &= \frac{\varepsilon - r}{2(\|v_0\| + \delta_2)}.\end{aligned}$$

Suppose $(\alpha, v) \in \mathbf{B}_{\delta_1}(\alpha_0) \times \mathbf{B}_{\delta_2}(v_0)$, then

$$\|\alpha v - x\| \leqslant \|\alpha v - \alpha_0 v_0\| + \|\alpha_0 v_0 - x\| < \delta_1\big(\|v_0\| + \delta_2\big) + |\alpha_0|\delta_2 + r = \varepsilon - r + r = \varepsilon,$$

therefore $\mathbf{B}_{\delta_1}(\alpha_0) \times \mathbf{B}_{\delta_2}(v_0) \subseteq s^{-1}(\mathbf{B}_\varepsilon(x))$ is an open rectangle containing (α_0, v_0). □

Corollary 3.3. *If $\big(V, \|\cdot\|\big)$ is a normed space, (v_n), (w_n) are sequences converging in V, and $\alpha \in \mathbf{F}$ is a scalar, then*

(a) $\lim_{n\to\infty}(v_n + w_n) = \lim_{n\to\infty} v_n + \lim_{n\to\infty} w_n$;

(b) $\lim_{n\to\infty}(\alpha\, v_n) = \alpha \lim_{n\to\infty} v_n$;

(c) $\lim_{n\to\infty}\|v_n\| = \Big\|\lim_{n\to\infty} v_n\Big\|$.

The proof of Proposition 3.2 went directly through the product topology on $V \times V$. You may have wondered about the possibility of defining a norm on the product space and using that instead. That is certainly possible (although it would not have simplified the proof very much):

Proposition 3.4. *Let $\big(V, \|\cdot\|_V\big)$ and $\big(W, \|\cdot\|_W\big)$ be normed vector spaces. The following functions give norms on the vector space $V \times W$:*

$$\begin{aligned} \|\cdot\|_1 \colon V \times W &\longrightarrow \mathbf{R}_{\geqslant 0} &\qquad \|(v,w)\|_1 &= \|v\|_V + \|w\|_W \\ \|\cdot\|_\infty \colon V \times W &\longrightarrow \mathbf{R}_{\geqslant 0} &\qquad \|(v,w)\|_\infty &= \max\{\|v\|_V, \|w\|_W\}.\end{aligned}$$

The norm $\|\cdot\|_1$ gives rise to the Manhattan metric d_1, the norm $\|\cdot\|_\infty$ gives rise to the sup metric d_∞, and any norm $\|\cdot\|$ on $V \times W$ such that

$$\|(v,w)\|_\infty \leqslant \|(v,w)\| \leqslant \|(v,w)\|_1 \qquad \textit{for all } (v,w) \in V \times W$$

gives rise to a conserving metric on $V \times W$. In particular, all these norms give rise to the product topology on $V \times W$.

Proof. We prove that $\|\cdot\|_1$ is a norm and leave $\|\cdot\|_\infty$ as an exercise. The other claims follow immediately from the definition of the metric given by a norm, and by Exercise 2.52.

We have

$$\begin{aligned}\|(v_1, w_1) + (v_2, w_2)\|_1 &= \|v_1 + v_2\|_V + \|w_1 + w_2\|_W \\ &\leqslant \|v_1\|_V + \|v_2\|_V + \|w_1\|_W + \|w_2\|_W \\ &= \|(v_1, w_1)\|_1 + \|(v_2, w_2)\|_1.\end{aligned}$$

Next for all α in the field of scalars $\mathbf{F}$:

$$\|\alpha(v,w)\|_1 = \|\alpha v\|_V + \|\alpha w\|_W = |\alpha|\,\|v\|_V + |\alpha|\,\|w\|_W = |\alpha|\,\|(v,w)\|_1.$$

Finally

$$\begin{aligned}\|(v,w)\|_1 = 0 &\iff \|v\|_V + \|w\|_W = 0\\ &\iff \|v\|_V = 0 \text{ and } \|w\|_W = 0\\ &\iff v = 0, w = 0 \iff (v,w) = (0,0).\end{aligned}$$ □

Proposition 3.5. *Let $\{v_1,\dots,v_n\}$ be a linearly independent subset of a normed space $(V, \|\cdot\|)$. Then there exists $m > 0$ such that*

$$\|\alpha_1 v_1 + \cdots + \alpha_n v_n\| \geqslant m\big(|\alpha_1| + \cdots + |\alpha_n|\big) \qquad \text{for all } \alpha_1,\dots,\alpha_n \in \mathbf{F}.$$

Proof. Let $A = |\alpha_1| + \cdots + |\alpha_n|$.

If $A = 0$, then the inequality is trivially true.

So suppose $A > 0$. Then, dividing by A, we have reduced to proving that there exists $m > 0$ such that

$$\|\beta_1 v_1 + \cdots + \beta_n v_n\| \geqslant m \qquad \text{for all } \beta_1,\dots,\beta_n \in \mathbf{F} \text{ such that } |\beta_1| + \cdots + |\beta_n| = 1.$$

To do this, consider the set

$$K = \{(\beta_1,\dots,\beta_n) \in \mathbf{F}^n : |\beta_1| + \cdots + |\beta_n| = 1\}.$$

It is closed and bounded in $\mathbf{F}^n$ (which is $\mathbf{C}^n$ or $\mathbf{R}^n$), so K is compact.

Now look at the function $F\colon K \longrightarrow \mathbf{R}$ given by

$$F(\beta_1,\dots,\beta_n) = \|\beta_1 v_1 + \cdots + \beta_n v_n\|.$$

This is a composition of continuous functions, hence is itself continuous. Since K is compact, F attains its minimum m on K. A priori we know that $m \geqslant 0$. But if $m = 0$, then for some $\beta_1,\dots,\beta_n \in K$ we have

$$\|\beta_1 v_1 + \cdots + \beta_n v_n\| = 0 \Rightarrow \beta_1 v_1 + \cdots + \beta_n v_n = 0,$$

contradicting the linear independence of the vectors.

Hence $m > 0$ and we are done. □

We are now in a good position to prove that

Theorem 3.6. *Any two norms on a finite-dimensional vector space V are equivalent.*

Proof. Let $v_1,\dots,v_n$ be a basis of V. Consider the norm on V defined by

$$\|\alpha_1 v_1 + \cdots + \alpha_n v_n\|_1 = |\alpha_1| + \cdots + |\alpha_n|.$$

We want to prove that any norm $\|\cdot\|$ on V is equivalent to $\|\cdot\|_1$.

Let $M = \max\{\|v_1\|,\dots,\|v_n\|\}$. Then

$$\|\alpha_1 v_1 + \cdots + \alpha_n v_n\| \leqslant |\alpha_1|\,\|v_1\| + \cdots + |\alpha_n|\,\|v_n\| \leqslant M\big(|\alpha_1| + \cdots + |\alpha_n|\big).$$

From Proposition 3.5 we also have $m > 0$ such that

$$m\big(|\alpha_1| + \cdots + |\alpha_n|\big) \leqslant \|\alpha_1 v_1 + \cdots + \alpha_n v_n\|,$$

We conclude that the norms $\|\cdot\|$ and $\|\cdot\|_1$ are equivalent. □

The following is (a special case of) the topological vector space analogue of Proposition 2.49:

Proposition 3.7. *Let $(V, \|\cdot\|)$ be a normed space and let $W \subseteq V$ be a subspace. Then its closure $\overline{W}$ is also a subspace.*

Proof. Suppose $u, v \in \overline{W}$, then there exist sequences (u_n) and (v_n) in W such that $(u_n) \longrightarrow u$ and $(v_n) \longrightarrow v$. Therefore $u_n + v_n \in W$ for all n, and by Proposition 3.2 we have

$$u + v = \lim(u_n) + \lim(v_n) = \lim(u_n + v_n) \in \overline{W}.$$

Similarly for scalar multiplication. □

If a normed space $(V, \|\cdot\|)$ is complete as a metric space, we say that it is a *Banach space*.

Proposition 3.8. *Any finite-dimensional normed space $(V, \|\cdot\|)$ is Banach.*

Proof. We need to show that V is complete. Let $v_1, \ldots, v_n$ be a basis of V.

By Proposition 3.5 we know that without loss of generality we can take the norm to be given by

$$\|\alpha_1 v_1 + \cdots + \alpha_n v_n\| = |\alpha_1| + \cdots + |\alpha_n| \qquad \text{for all } \alpha_1, \ldots, \alpha_n \in \mathbf{F}.$$

Consider a Cauchy sequence in V, and express each term as a linear combination of the chosen basis:

$$\left(u^{(m)}\right) = \left(\alpha_1^{(m)} v_1 + \cdots + \alpha_n^{(m)} v_n\right).$$

The Cauchyness means that for any $\varepsilon > 0$ there exists $M \in \mathbf{N}$ such that for all $m, k \geqslant M$ we have $\|u^{(m)} - u^{(k)}\| < \varepsilon$, in other words

$$\varepsilon > \|u^{(m)} - u^{(k)}\| = |\alpha_1^{(m)} - \alpha_1^{(k)}| + \cdots + |\alpha_n^{(m)} - \alpha_n^{(k)}|.$$

This means that for each $j = 1, \ldots, n$, $(\alpha_j^{(m)})$ is a Cauchy sequence in $\mathbf{F}$. As $\mathbf{F}$ is complete, $(\alpha_j^{(m)}) \longrightarrow \beta_j \in \mathbf{F}$.

We now let $u = \beta_1 v_1 + \cdots + \beta_n v_n$ and show that $(u^{(m)}) \longrightarrow u \in V$. Let $\varepsilon > 0$. For $j = 1, \ldots, n$, there exists $M_j \in \mathbf{N}$ such that $|\alpha_j^{(m)} - \beta_j| < \varepsilon/n$ for all $m \geqslant M_j$. Let $M = \max\{M_j : j = 1, \ldots, n\}$, then for all $m \geqslant M$ we have

$$\|u^{(m)} - u\| = |\alpha_1^{(m)} - \beta_1| + \cdots + |\alpha_n^{(m)} - \beta_n| < \varepsilon. \qquad \square$$

We are overdue for some infinite-dimensional examples of normed spaces, but we will first take a detour.

3.2. Inner products

We continue to take $\mathbf{F}$ to be either $\mathbf{R}$ or $\mathbf{C}$, and we denote by $\bar{\cdot}$ the complex conjugation (which is just the identity if $\mathbf{F} = \mathbf{R}$).

Let V be a vector space over $\mathbf{F}$. Recall from linear algebra (see Appendix A.2.3 for a summary) that an *inner product* on V is a function

$$\langle \cdot, \cdot \rangle : V \times V \longrightarrow \mathbf{F}$$

that is linear in the first variable, conjugate-linear in the second variable, and positive-definite.

Proposition 3.9. *If* $(V, \langle\cdot,\cdot\rangle)$ *is an inner product space, then the function* $\|\cdot\| : V \longrightarrow \mathbf{R}_{\geqslant 0}$ *defined by*

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

is a norm on V.

Proof. For any $v \in V$, $\alpha \in \mathbf{F}$ we have

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

Note also that

$$\|v\| = 0 \iff \sqrt{\langle v, v\rangle} = 0 \iff \langle v, v\rangle = 0 \iff v = 0.$$

Finally, by the Cauchy–Schwarz Inequality we have

$$\operatorname{Re}\langle v, w\rangle \leqslant \big|\langle v, w\rangle\big| \leqslant \|v\|\,\|w\|.$$

Therefore

$$\begin{aligned}
\|v + w\|^2 &= \langle v + w, v + w\rangle \\
&= \langle v, v\rangle + \langle v, w\rangle + \langle w, v\rangle + \langle w, w\rangle \\
&= \|v\|^2 + 2\operatorname{Re}\langle v, w\rangle + \|w\|^2 \\
&\leqslant \|v\|^2 + 2\|v\|\,\|w\| + \|w\|^2 \\
&= \big(\|v\| + \|w\|\big)^2,
\end{aligned}$$

which means that the triangle inequality holds for $\|\cdot\|$. □

Obviously then:

Corollary 3.10. *Any inner product space is a normed space, and a metric space.*

A *Hilbert space* is a complete inner product space.

Proposition 3.11. *For any* $n \in \mathbf{N}$, $\mathbf{F}^n$ *is a Hilbert space.*

Proof. We know that $\mathbf{F}^n$ is an inner product space, see Example A.5. We also know that finite-dimensional normed spaces are complete, by Proposition 3.8, so $\mathbf{F}^n$ is a Hilbert space. □

An inner product gives rise to a norm. Given a norm, how can we determine whether it comes from an inner product? It turns out that there's a fun way to check:

Proposition 3.12 (Parallelogram Law). *If* $(V, \langle\cdot,\cdot\rangle)$ *is an inner product space, then its norm satisfies*

$$\|v + w\|^2 + \|v - w\|^2 = 2\big(\|v\|^2 + \|w\|^2\big) \qquad \text{for all } v, w \in V.$$

Proof. Recall from the proof of Proposition 3.9 that

$$\|v + w\|^2 = \|v\|^2 + 2\operatorname{Re}\langle v, w\rangle + \|w\|^2.$$

Then

$$\|v - w\|^2 = \|v\|^2 - 2\operatorname{Re}\langle v, w\rangle + \|w\|^2,$$

and adding the two equalities gives the identity in the statement. □

In the proof of the Parallelogram Law we added the two equalities

$$\|v+w\|^2 = \|v\|^2 + 2\operatorname{Re}\langle v,w\rangle + \|w\|^2$$
$$\|v-w\|^2 = \|v\|^2 - 2\operatorname{Re}\langle v,w\rangle + \|w\|^2.$$

Subtracting them instead also gives an interesting fact:

$$4\operatorname{Re}\langle v,w\rangle = \|v+w\|^2 - \|v-w\|^2.$$

When $\mathbf{F} = \mathbf{C}$, can we recover all of the inner product $\langle v,w\rangle$ (as opposed to just the real part)? Yes, because

$$\operatorname{Im}\langle v,w\rangle = \operatorname{Re}\langle v,iw\rangle,$$

which leads us to conclude

Proposition 3.13 (Polarisation Identity). *If $(V,\langle\cdot,\cdot\rangle)$ is an inner product space then*

$$4\langle v,w\rangle = \begin{cases} \|v+w\|^2 - \|v-w\|^2 & \text{if } \mathbf{F}=\mathbf{R} \\ \|v+w\|^2 - \|v-w\|^2 + i\|v+iw\|^2 - i\|v-iw\|^2 & \text{if } \mathbf{F}=\mathbf{C}. \end{cases}$$

Corollary 3.14 (Converse to the Parallelogram Law). *If $(V,\|\cdot\|)$ is a normed space such that*

$$\|v+w\|^2 + \|v-w\|^2 = 2\big(\|v\|^2 + \|w\|^2\big) \qquad \text{for all } v,w \in V,$$

then the function $\langle\cdot,\cdot\rangle$ defined by

$$4\langle v,w\rangle = \begin{cases} \|v+w\|^2 - \|v-w\|^2 & \text{if } \mathbf{F}=\mathbf{R} \\ \|v+w\|^2 - \|v-w\|^2 + i\|v+iw\|^2 - i\|v-iw\|^2 & \text{if } \mathbf{F}=\mathbf{C} \end{cases}$$

is an inner product on V with associated norm $\|\cdot\|$.

Proof. See Question 1 on Assignment 2. □

3.3. Convexity and inequalities

A subset S of a vector space V over $\mathbf{F}$ is *convex* if for all $v,w \in S$ and all $a,b \in \mathbf{R}_{\geqslant 0}$ such that $a+b=1$, we have $av+bw \in S$. (In other words, for any two points in S, the line segment joining the two points is entirely contained in S.)

Example 3.15. Any subspace W of V is convex.

Solution. Suppose $v,w \in W$, $a,b \in \mathbf{R}_{\geqslant 0}$ such that $a+b=1$. Then $av+bw$ is an $\mathbf{F}$-linear combination of elements of W. Since W is a subspace, $av+bw \in W$. □

Example 3.16. Any interval $I \subseteq \mathbf{R}$ is convex.

Solution. Let $I \subseteq \mathbf{R}$ be an interval and let $v,w \in I$, $a,b \in \mathbf{R}_{\geqslant 0}$ such that $a+b=1$.
Without loss of generality, $v \leqslant w$. Then

$$av+bw-v = (a-1)v+bw = b(w-v) \geqslant 0 \Rightarrow v \leqslant av+bw$$

and

$$av+bw-w = av+(b-1)w = a(v-w) \leqslant 0 \Rightarrow av+bw \leqslant w.$$

Therefore $v \leqslant av + bw \leqslant w$, hence $av + bw \in I$ by the definition of an interval. □

If V is a vector space over $\mathbf{F}$ and $S \subseteq V$ is a convex set, we say that a function $f: S \longrightarrow \mathbf{R}$ is *convex* if for all $v, w \in S$ and all $a, b \in \mathbf{R}_{\geqslant 0}$ such that $a + b = 1$, we have

$$f(av + bw) \leqslant af(v) + bf(w).$$

For instance, if $(V, \|\cdot\|)$ is a normed space, then the norm $\|\cdot\|: V \longrightarrow \mathbf{R}_{\geqslant 0}$ is a convex function, see Exercise 3.4.

More interestingly, the notion of convex function is closely related to the concept of concavity in single-variable calculus:

Proposition 3.17. *Let $I \subseteq \mathbf{R}$ be an interval and let $f: I \longrightarrow \mathbf{R}$ be a twice-differentiable function.*

Then f is convex if and only if $f''(x) \geqslant 0$ for all $x \in I$.

Proof. See Exercise 3.5. □

Example 3.18. The functions

$$\begin{aligned} &\text{(i)} \quad f: (0, \infty) \longrightarrow \mathbf{R}, \quad f(x) = x^p, \quad p \geqslant 1 \text{ fixed},\\ &\text{(ii)} \quad \exp: \mathbf{R} \longrightarrow \mathbf{R}, \quad \exp(x) = e^x, \end{aligned}$$

are convex.

Solution.

(i) We have $f''(x) = p(p-1)x^{p-2} \geqslant 0$ for all $x > 0$, as $p \geqslant 1$.

(ii) We have $\exp''(x) = e^x \geqslant 0$ for all $x \in \mathbf{R}$. □

Proposition 3.19. *Let $x, y \in \mathbf{R}_{\geqslant 0}$.*

(a) For any $p \geqslant 1$ and any $a, b \geqslant 0$ such that $a + b = 1$, we have

$$(ax + by)^p \leqslant ax^p + by^p.$$

(b) For any $a, b \geqslant 0$ such that $a + b = 1$, we have

$$x^a y^b \leqslant ax + by.$$

(c) For any $p \geqslant 1$, we have

$$x^p + y^p \leqslant (x + y)^p.$$

Proof.

(a) This is exactly the definition of $x \longmapsto x^p$ being a convex function.

(b) If $x = 0$ or $y = 0$, the inequality is trivial, so we may assume $x, y > 0$. Setting $x = e^s$, $y = e^t$, we are trying to prove that

$$e^{as+bt} \leqslant ae^s + be^t,$$

which is the same as e^x being a convex function.

(c) If $y = 0$, the inequality is obvious, so we may assume $y > 0$. Setting $t = x/y$, we are trying to show that
$$t^p + 1 \leqslant (t+1)^p \qquad \text{for all } t \geqslant 0.$$
Let $f\colon \mathbf{R}_{\geqslant 0} \longrightarrow \mathbf{R}$ be given by $f(t) = t^p+1$, and $g(t)\colon \mathbf{R}_{\geqslant 0} \longrightarrow \mathbf{R}$ be given by $g(t) = (t+1)^p$. We have $f(0) = g(0) = 1$. Also
$$f'(t) = pt^{p-1} \leqslant p(t+1)^{p-1} = g'(t) \qquad \text{for all } t > 0,$$
therefore $f(t) \leqslant g(t)$ for all $t \geqslant 0$, as desired. (There's an appeal to the Mean Value Theorem hiding in here, if you want to write out all the details.) □

Corollary 3.20. *Let $p \geqslant 1$, $q > 0$, $x, y \geqslant 0$, and $a, b \geqslant 0$ such that $a + b = 1$, then:*
$$\begin{aligned}\min\{x,y\} &\leqslant \left(ax^{-q} + by^{-q}\right)^{-1/q} \\ &\leqslant x^a y^b \\ &\leqslant \left(ax^{1/p} + by^{1/p}\right)^p \\ &\leqslant ax + by \\ &\leqslant \left(ax^p + by^p\right)^{1/p} \\ &\leqslant \max\{x,y\}.\end{aligned}$$

Proof. Without loss of generality $x \leqslant y$ so $\min\{x,y\} = x$ and $\max\{x,y\} = y$.

(a) $x \leqslant y$ so $x^{-1} \geqslant y^{-1}$ so $x^{-q} \geqslant y^{-q}$ so $bx^{-q} \geqslant by^{-q}$ so $ax^{-q} + bx^{-q} \geqslant ax^{-q} + by^{-q}$ so
$$\min\{x,y\} = x = (ax^{-q} + bx^{-q})^{-1/q} \leqslant (ax^{-q} + by^{-q})^{-1/q}.$$

(b) Let $X = x^{-q}$, $Y = y^{-q}$, then by Proposition 3.19 part (b) we have
$$\begin{aligned} X^a Y^b \leqslant aX + bY &\Rightarrow x^{-aq} y^{-bq} \leqslant ax^{-q} + by^{-q} \\ &\Rightarrow x^{aq} y^{bq} \geqslant (ax^{-q} + by^{-q})^{-1} \\ &\Rightarrow (ax^{-q} + by^{-q})^{-1/q} \leqslant x^a y^b. \end{aligned}$$

(c) Similar to (b), use Proposition 3.19 part (b) with $X = x^{1/p}$, $Y = y^{1/p}$.

(d) Use Proposition 3.19 part (a) with $X = x^{1/p}$, $Y = y^{1/p}$.

(e) Precisely Proposition 3.19 part (a).

(f) Similar to (a). □

3.4. Sequence spaces

The set of all sequences $\mathbf{F}^{\mathbf{N}} = \{(a_n)\colon a_n \in \mathbf{F} \text{ for all } n \in \mathbf{N}\}$ is of course a vector space over $\mathbf{F}$ with the usual addition and scalar multiplication.

We consider a family of subsets of $\mathbf{F}^{\mathbf{N}}$, parametrised by a real number $p \geqslant 1$:
$$\ell^p = \left\{(a_n) \in \mathbf{F}^{\mathbf{N}}\colon \sum_{n=1}^{\infty} |a_n|^p < \infty\right\},$$

the elements of which are called *p-summable* sequences. (If $p = 1$ we simply call them *summable*, and if $p = 2$, *square-summable*.) We consider the function $\|\cdot\|_{\ell^p} : \ell^p \longrightarrow \mathbf{R}_{\geqslant 0}$ defined by

$$\|(a_n)\|_{\ell^p} = \left(\sum_{n=1}^{\infty} |a_n|^p\right)^{1/p}.$$

There is also an exceptional case $p = \infty$ given by *bounded* sequences

$$\begin{aligned}\ell^\infty &= \left\{(a_n) \in \mathbf{F}^{\mathbf{N}} : \sup(|a_n|) < \infty\right\} \\ &= \left\{(a_n) \in \mathbf{F}^{\mathbf{N}} : \text{there exists } M \text{ such that } |a_n| \leqslant M \text{ for all } n \in \mathbf{N}\right\},\end{aligned}$$

with function $\|\cdot\|_{\ell^\infty} : \ell^\infty \longrightarrow \mathbf{R}_{\geqslant 0}$ given by

$$\|(a_n)\|_{\ell^\infty} = \sup(|a_n|).$$

The upshot is that all these subsets of $\mathbf{F}^{\mathbf{N}}$ are normed spaces, as we now see.

Proposition 3.21 (Minkowski's Inequality)**.** *Let $1 \leqslant p \leqslant \infty$ and let $u = (u_n), v = (v_n) \in \ell^p$. Then*

$$\|u + v\|_{\ell^p} \leqslant \|u\|_{\ell^p} + \|v\|_{\ell^p}.$$

Proof. Fix p and write $\|\cdot\|$ instead of $\|\cdot\|_{\ell^p}$ to simplify notation.

To start with, let $x = (x_n), y = (y_n) \in \ell^p$, and let $a, b \geqslant 0$ be such that $a + b = 1$. Then

$$\sum_{n=1}^{\infty} |ax_n + by_n|^p \leqslant \sum_{n=1}^{\infty} \big(a|x_n| + b|y_n|\big)^p \leqslant a \sum_{n=1}^{\infty} |x_n|^p + b \sum_{n=1}^{\infty} |y_n|^p,$$

where we applied first the triangle inequality for the absolute value, and second the inequality from Proposition 3.19, part (a). Therefore

$$\|ax + by\|^p \leqslant a\|x\|^p + b\|y\|^p.$$

In other words, $\|\cdot\|^p$ is a convex function.

Now we go back to the context of the statement of the proposition. Given $u, v \in \ell^p$, define

$$x = \frac{1}{\|u\|}u, \qquad y = \frac{1}{\|v\|}v, \qquad a = \frac{\|u\|}{\|u\| + \|v\|}, \qquad b = \frac{\|v\|}{\|u\| + \|v\|},$$

then we have

$$\left(\frac{\|u + v\|}{\|u\| + \|v\|}\right)^p = \|ax + by\|^p \leqslant a + b = 1.$$ □

Corollary 3.22. *The set ℓ^p is a vector subspace of $\mathbf{F}^{\mathbf{N}}$, and $\|\cdot\|_{\ell^p}$ is a norm on ℓ^p.*

Proof. It is clear from the definition of ℓ^p that it contains the constant zero sequence $\mathbf{0}$, and that it is closed under scalar multiplication. By Minkowski's Inequality it is also closed under vector addition, so it is a subspace.

Minkowski's Inequality also gives us the triangle inequality for $\|\cdot\|_{\ell^p}$, as well as the behaviour under scalar multiplication. Finally, if (a_n) is such that there exists $n \in \mathbf{N}$ with $|a_n| > 0$, then $\|(a_n)\|_{\ell^p} \geqslant |a_n| > 0$. So $\|(a_n)\|_{\ell^p} = 0$ if and only if $(a_n) = \mathbf{0}$. □

Here is our first example of a normed space that is not Banach:

Example 3.23. Consider

$$c_{00} = \left\{ (a_n) \in \mathbf{F}^{\mathbf{N}} : \text{there exists } N \in \mathbf{N} \text{ such that } a_n = 0 \text{ for all } n \geqslant N \right\}$$

consisting of sequences in $\mathbf{F}$ with only finitely many nonzero terms.

This is clearly a vector subspace of ℓ^∞, and of course inherits the ℓ^∞ norm from it.

I claim that it is **not** complete, and **not** a closed subspace of ℓ^∞.

Consider the sequence (v_n) in c_{00} given by

$$v_n = \left(1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}, 0, 0, 0, \dots\right).$$

It is Cauchy: given $\varepsilon > 0$, let $N \in \mathbf{N}$ be such that $1/N < \varepsilon$, then for all $n \geqslant m \geqslant N$ we have

$$\|v_n - v_m\|_{\ell^\infty} = \sup\left\{0, \frac{1}{m+1}, \frac{1}{m+2}, \dots, \frac{1}{n}\right\} = \frac{1}{m+1} < \frac{1}{N} < \varepsilon.$$

As a sequence in ℓ^∞, it converges to the following element of ℓ^∞:

$$u = \left(1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}, \dots\right),$$

which is easy to see since

$$\|u - v_n\|_{\ell^\infty} = \frac{1}{n+1} \longrightarrow 0.$$

But u is not in c_{00}, so (v_n) is a Cauchy sequence in c_{00} that does not converge in c_{00}, so c_{00} is not complete. Moreover, (v_n) converges in ℓ^∞, so its limit u is in the closure of c_{00} in ℓ^∞, but not in c_{00} itself.

It is therefore worth thinking about completions of normed spaces, which thankfully are an easy add-on to the topic of completions of metric spaces.

Let $(V, \|\cdot\|)$ be a normed space over $\mathbf{F}$. A *completion* of $(V, \|\cdot\|_V)$ is a Banach space $(\widehat{V}, \|\cdot\|_{\widehat{V}})$ over $\mathbf{F}$ together with an $\mathbf{F}$-linear distance-preserving map

$$\iota \colon V \longrightarrow \widehat{V}$$

such that $\iota(V)$ is a dense normed subspace of $\widehat{V}$.

Proposition 3.24. *Any normed space* $(V, \|\cdot\|_V)$ *has a completion* $(\widehat{V}, \|\cdot\|_{\widehat{V}})$.

Proof. We know from Theorem 2.62 that V has a completion that is a metric space. We have to show that the particular complete metric space $(\widehat{V}, \widehat{d})$ constructed in the proof of Theorem 2.62 is actually a normed space such that $\iota(V)$ is a normed subspace.

This is essentially straightforward, just has a lot of tiny little parts.

Let $\widehat{u} = [(u_n)], \widehat{v} = [(v_n)] \in \widehat{V}$. We define

$$\widehat{u} + \widehat{v} = [(u_n + v_n)].$$

To see why this works, first take a Cauchy sequence (u_n) representing the equivalence class $\widehat{u}$ and a Cauchy sequence (v_n) representing the equivalence class $\widehat{v}$. The sequence $(u_n + v_n)$ is Cauchy in V, as

$$\|(u_n + v_n) - (u_m + v_m)\|_V \leqslant \|u_n - u_m\|_V + \|v_n - v_m\|_V,$$

and (u_n) and (v_n) are Cauchy in V.

Had we chosen other representatives $(u'_n) \sim (u_n)$ and $(v'_n) \sim (v_n)$, we would have ended up with $(u'_n + v'_n)$, which is easily seen to be equivalent to $(u_n + v_n)$, so the equivalence class $[(u_n) + (v_n)]$ is indeed well-defined.

Scalar multiplication and the norm are defined on $\widehat{V}$ as:

$$\alpha\widehat{u} = [(\alpha u_n)], \qquad \|\widehat{u}\|_{\widehat{V}} = \lim\left(\|u_n\|_V\right)$$

and their well-definedness is argued similarly.

Checking the vector space axioms for $\widehat{V}$ is done by using the vector space axioms for V and the continuity of the operations.

Note also that the metric $\widehat{d}$ on $\widehat{V}$ constructed in Theorem 2.62 is the metric associated with the norm $\|\cdot\|_{\widehat{V}}$:

$$\widehat{d}(\widehat{u},\widehat{v}) = \lim d(u_n, v_n) = \lim \|u_n - v_n\|_V = \|\widehat{u} - \widehat{v}\|_{\widehat{V}}.$$ □

3.5. Continuous linear transformations

Let V and W be normed spaces.

A linear transformation $f : V \longrightarrow W$ is said to be *Lipschitz* if there exists $c > 0$ such that

$$\|f(v)\|_W \leqslant c\,\|v\|_V \qquad \text{for all } v \in V.$$

(In the literature you will find that these linear transformations are referred to as "bounded", but I will try very hard to avoid this as it clashes, for non-compact V, with the notion of bounded function we discussed in Section 2.12. The relation between the two notions will be clarified below.)

Proposition 3.25. *A linear transformation $f : V \longrightarrow W$ between normed spaces is continuous if and only if it is Lipschitz if and only if it is uniformly continuous.*

Proof. Suppose f is Lipschitz with constant $c > 0$. Given $\varepsilon > 0$, let $\delta = \varepsilon/c$. If $v_1, v_2 \in V$ are such that $\|v_1 - v_2\|_V < \delta$, then

$$\|f(v_1) - f(v_2)\|_W = \|f(v_1 - v_2)\|_W \leqslant c\|v_1 - v_2\|_V < c\delta = \varepsilon.$$

Therefore f is uniformly continuous, hence continuous.

Suppose f is not Lipschitz. Let $n \in \mathbf{N}$. There exists $v_n \in V$ such that

$$\|f(v_n)\|_W \geqslant n\,\|v_n\|_V.$$

Let $\alpha_n = 1/\|f(v_n)\|_W$ and $u_n = \alpha_n v_n$, then

$$\|u_n\|_V = |\alpha_n|\,\|v_n\|_V = \frac{\|v_n\|_V}{\|f(v_n)\|_W} \leqslant \frac{1}{n},$$

which implies that the sequence $(u_n) \longrightarrow 0 \in V$.

But

$$\|f(u_n)\|_W = |\alpha_n|\,\|f(v_n)\|_W = 1,$$

so the sequence $(f(u_n))$ does not converge to $f(0) = 0$ in W, hence f is not continuous. □

We will write $L(V,W)$ for the set of Lipschitz (aka continuous, aka uniformly continuous) linear transformations between the normed spaces V and W. In the special case $V = W$ we simply write $L(V) = L(V,V)$.

Consider the following function $\|\cdot\|\colon L(V,W) \longrightarrow \mathbf{R}_{\geqslant 0}$:

$$\|f\| = \sup_{v\neq 0} \frac{\|f(v)\|_W}{\|v\|_V}.$$

As $f \in L(V,W)$, there exists $c > 0$ such that

$$\frac{\|f(v)\|_W}{\|v\|_V} \leqslant c \qquad \text{for all } v \neq 0,$$

so that there is a finite supremum $\|f\|$.

We also note the obvious fact that

$$\|f(v)\|_W \leqslant \|f\|\,\|v\|_V \qquad \text{for all } v \in V,$$

and that the linearity of f allows us to rewrite

$$\|f\| = \sup_{\|v\|_V=1} \|f(v)\|_W.$$

Theorem 3.26. *Let V and W be normed spaces.*

(a) The set $L(V,W)$ is a normed space with norm given by

$$\|f\| = \sup_{v\neq 0} \frac{\|f(v)\|_W}{\|v\|_V} = \sup_{\|v\|_V=1} \|f(v)\|_W.$$

(b) Consider the map $N\colon L(V,W) \longrightarrow B\big(V\smallsetminus\{0\},W\big)$ given by $N(f) = F$, where

$$F\colon V\smallsetminus\{0\} \longrightarrow W, \qquad F(v) := \frac{1}{\|v\|_V} f(v).$$

Then N is distance-preserving, and its image is a closed subset of $B\big(V\smallsetminus\{0\},W\big)$.

(c) If W is a Banach space then $L(V,W)$ is also Banach.

Proof.

(a) As $L(V,W)$ is a subset of $\operatorname{Hom}(V,W)$ and the latter is a vector space, we check that $L(V,W)$ is a subspace.

We have

$$\begin{aligned}\|f+g\| &= \sup_{\|v\|_V=1} \|f(v)+g(v)\|_W \\ &\leqslant \sup_{\|v\|_V=1} \big(\|f(v)\|_W + \|g(v)\|_W\big) \\ &\leqslant \sup_{\|v\|_V=1} \|f(v)\|_W + \sup_{\|v\|_V=1} \|g(v)\|_W \\ &= \|f\| + \|g\|,\end{aligned}$$

so that if both f and g are in $L(V,W)$, so is $f+g$.

Similarly:

$$\|\alpha f\| = \sup_{\|v\|_V=1} \|\alpha f(v)\|_W = \sup_{\|v\|_V=1} |\alpha|\,\|f(v)\|_W = |\alpha|\,\|f\|,$$

so that if f is in $L(V,W)$ and $\alpha \in \mathbf{F}$, then αf is in $L(V,W)$.

In addition to showing that $L(V,W)$ is a vector space, these identities also give two of the three norm axioms, leaving to check that $\|f\| = 0$ if and only if $\|f(v)\|_W = 0$ for all $v \in V$ if and only if $f(v) = 0$ for all $v \in V$ if and only if $f = 0$.

(b) Let $f \in L(V, W)$ and let $F = N(f)$. First note that if f is Lipschitz with constant $c > 0$, then $d_W(F(v), 0) \leqslant c$ for all $v \in V \setminus \{0\}$, so F is bounded.

Next we see that if $g \in L(V, W)$ and $G = N(g)$, then

$$\begin{aligned} d_\infty(F, G) &= \sup_{v \in V \setminus \{0\}} \{d_W(F(v), G(v))\} \\ &= \sup_{v \in V \setminus \{0\}} \left\{d_W\left(\frac{1}{\|v\|_V} f(v), \frac{1}{\|v\|_V} g(v)\right)\right\} \\ &= \sup_{v \in V \setminus \{0\}} \left\{\frac{1}{\|v\|_V} \|f(v) - g(v)\|_W\right\} \\ &= \|f - g\|, \end{aligned}$$

so N is indeed distance-preserving.

Let F be in the closure of the image of N and let (F_n) be a sequence with $F_n = N(f_n)$ such that $(F_n) \longrightarrow F$ with respect to the uniform metric.

Define $f\colon V \longrightarrow W$ by setting

$$f(0) = 0 \qquad \text{and} \qquad f(v) = \|v\|_V F(v) \quad \text{for } v \in V \setminus \{0\}.$$

If we can show that $f \in L(V, W)$, then we are done, as clearly $N(f) = F$.

For linearity (ignoring corner cases where some vectors might be zero):

$$\begin{aligned} f(\lambda_1 v_1 + \lambda_2 v_2) &= \|\lambda_1 v_1 + \lambda_2 v_2\|_V F(\lambda_1 v_1 + \lambda_2 v_2) \\ &= \|\lambda_1 v_1 + \lambda_2 v_2\|_V \lim_{n \longrightarrow \infty} F_n(\lambda_1 v_1 + \lambda_2 v_2) \\ &= \lim_{n \longrightarrow \infty} (\|\lambda_1 v_1 + \lambda_2 v_2\|_V F_n(\lambda_1 v_1 + \lambda_2 v_2)) \\ &= \lim_{n \longrightarrow \infty} f_n(\lambda_1 v_1 + \lambda_2 v_2) \\ &= \lim_{n \longrightarrow \infty} (\lambda_1 f_n(v_1) + \lambda_2 f_n(v_2)) \\ &= \lambda_1 \lim_{n \longrightarrow \infty} f_n(v_1) + \lambda_2 \lim_{n \longrightarrow \infty} f_n(v_2) \\ &= \lambda_1 \|v_1\|_V \lim_{n \longrightarrow \infty} F_n(v_1) + \lambda_2 \|v_2\|_V \lim_{n \longrightarrow \infty} F_n(v_2) \\ &= \lambda_1 \|v_1\|_V F_n(v_1) + \lambda_2 \|v_2\|_V F_n(v_2) \\ &= \lambda_1 f(v_1) + \lambda_2 f(v_2). \end{aligned}$$

So f is linear. The fact that f is Lipschitz follows immediately from the fact that $F = N(f)$ is bounded.

(c) This follows from part (b), since W Banach implies that $B\big(V \setminus \{0\}, W\big)$ is complete by Proposition 2.78, so the image of N is complete as it is closed (Proposition 2.55), so $L(V, W)$ is complete since it is isometric to the image of N. □

Let's record an important consequence of Theorem 3.26:

Corollary 3.27. *For any normed space V, the dual space $V^\vee = L(V, \mathbf{F})$ is a Banach space with norm*

$$\|\varphi\| = \sup_{v \neq 0} \frac{|\varphi(v)|}{\|v\|_V}.$$

We'll come back to the topic of dual spaces.

To prove that $L(V, W)$ is a normed space, we had to consider the interplay between the addition of functions and the norms on V and W, and similarly for the operation of multiplying a function by a scalar. There is another operation on functions that has been conspicuously missing from this discussion: composition. We look at this now.

Recall (or see Appendix A.2) that an algebra is a vector space A with a vector multiplication map $A \times A \longrightarrow A$, $(u, v) \longmapsto uv$.

For example, given a vector space V over $\mathbf{F}$, the set of all $\mathbf{F}$-linear transformations $V \longrightarrow V$ is an associative unital $\mathbf{F}$-algebra, where multiplication is given by composition and the unit is $\mathbf{1} = \mathrm{id}_V$.

Proposition 3.28. *If $f\colon U \longrightarrow V$ and $g\colon V \longrightarrow W$ are continuous linear transformations between normed spaces, then $g \circ f\colon U \longrightarrow W$ is continuous and linear, and*

$$\|g \circ f\| \leqslant \|g\| \, \|f\|.$$

In particular, for any normed space V, the normed space $L(V)$ is closed under composition, hence is an associative unital $\mathbf{F}$-algebra.

Proof. We know already that the composition of linear maps is linear, and that the composition of continuous maps is continuous.

As for the norms, for any $u \in U$ we have

$$\|(g \circ f)(u)\|_W = \|g(f(u))\|_W \leqslant \|g\| \, \|f(u)\|_V \leqslant \|g\| \, \|f\| \, \|u\|_U,$$

so that for all $u \neq 0$ we have

$$\frac{\|(g \circ f)(u)\|_W}{\|u\|_U} \leqslant \|g\| \, \|f\|,$$

and we can conclude by taking supremum.

If $U = W = V$ we get the $\mathbf{F}$-algebra $L(V)$ with multiplication given by composition, and with unit element $\mathbf{1} = \mathrm{id}_V$, clearly both linear and continuous. □

Proposition 3.29. *Let V and W be normed spaces and fix completions $(\widehat{V}, \iota_V)$ of V and $(\widehat{W}, \iota_W)$ of W. Then every $f \in L(V, W)$ has a unique extension $\widehat{f} \in L\big(\widehat{V}, \widehat{W}\big)$ and*

$$\|\widehat{f}\| = \|f\|.$$

Proof. We know from Proposition 3.25 that f is uniformly continuous, hence by Proposition 2.63 it extends uniquely to a uniformly continuous function $\widehat{f}\colon \widehat{V} \longrightarrow \widehat{W}$. Moreover, given

$$\widehat{u} = \lim_{n \longrightarrow \infty} \iota_V(u_n),$$

we have

$$\widehat{f}(\widehat{u}) = \widehat{f}\Big(\lim_{n \longrightarrow \infty} \iota_V(u_n)\Big) = \lim_{n \longrightarrow \infty} \iota_W\big(f(u_n)\big).$$

We use this description to prove the linearity of $\widehat{f}$: for $\widehat{u}, \widehat{v} \in \widehat{V}$ and $\alpha, \beta \in \mathbf{F}$ we have

$$\begin{aligned}
\widehat{f}(\alpha\widehat{u} + \beta\widehat{v}) &= \widehat{f}\big(\lim \iota_V(\alpha u_n + \beta v_n)\big) \\
&= \lim \iota_W\big(f(\alpha u_n + \beta v_n)\big) \\
&= \lim \alpha\iota_W\big(f(u_n)\big) + \lim \beta\iota_W\big(f(v_n)\big) \\
&= \alpha\widehat{f}(\widehat{u}) + \alpha\widehat{f}(\widehat{v}).
\end{aligned}$$

Insofar as the norm is concerned, we have

$$\|\widehat{f}(\widehat{v})\|_{\widehat{W}} = \lim \|f(v_n)\|_W \leqslant \|f\| \big(\lim \|v_n\|_V\big) = \|f\| \, \|\widehat{v}\|_{\widehat{V}},$$

which implies that $\|\widehat{f}\| \leqslant \|f\|$.

But there is another relation between these norms, which we obtain by considering the following diagram:

$$\begin{array}{ccccc} V & \xrightarrow[\simeq]{\iota_V} & \iota(V) & \xhookrightarrow{\text{inclusion}} & \widehat{V} \\ {\scriptstyle f}\big\downarrow & & \big\downarrow{\scriptstyle \widetilde{f}} & & \big\downarrow{\scriptstyle \widehat{f}} \\ W & \xrightarrow[\simeq]{\iota_W} & \iota(W) & \xhookrightarrow{\text{inclusion}} & \widehat{W} \end{array}$$

Since ι_V and ι_W are isometries, we have $\|\widetilde{f}\| = \|f\|$. Now $\|\widetilde{f}\|$ and $\|\widehat{f}\|$ are defined by the same formula, but the first is the supremum over the subset $\iota(V)$ of $\widehat{V}$, whereas the second is the supremum over all of $\widehat{V}$. Therefore

$$\|\widehat{f}\| \geqslant \|\widetilde{f}\| = \|f\|. \qquad \square$$

3.6. Series and Schauder bases

A sequence (a_n) in a normed space $\big(V, \|\cdot\|\big)$ defines a *series* in V

$$\sum_{n=1}^{\infty} a_n,$$

which is a shorthand notation for the sequence of *partial sums* (x_m), where

$$x_m = a_1 + \cdots + a_m = \sum_{n=1}^{m} a_n.$$

The series *converges* if there exists $x \in V$ such that $(x_m) \longrightarrow x$, that is

$$\left\| x - \sum_{n=1}^{m} a_n \right\|_V \longrightarrow 0 \qquad \text{as } m \longrightarrow \infty.$$

The limit x is called the *sum* of the series.

The series *converges absolutely* if the series of real numbers

$$\sum_{n=1}^{\infty} \|a_n\|_V$$

converges, that is there exists $r \in \mathbf{R}_{\geqslant 0}$ such that

$$\left(r - \sum_{n=1}^{m} \|a_n\|_V \right) \longrightarrow 0 \qquad \text{as } m \longrightarrow \infty.$$

Proposition 3.30. *Let $\big(V, \|\cdot\|\big)$ be a normed space. V is a Banach space if and only if every absolutely convergent series in V is convergent.*

Proof. In one direction, suppose V is Banach and

$$\sum_{n=1}^{\infty} \|a_n\|_V = r \in \mathbf{R}_{\geqslant 0}.$$

Write

$$x_m = \sum_{n=1}^{m} a_n.$$

Let $\varepsilon > 0$, then there exists $M \in \mathbf{N}$ such that

$$\left| \sum_{n=1}^{m} \|a_n\|_V - r \right| < \frac{\varepsilon}{2} \qquad \text{for all } m \geqslant M.$$

Then for all $m \geqslant k \geqslant M$ we have

$$\begin{aligned} \|x_m - x_k\|_V &= \left\| \sum_{n=k+1}^{m} a_n \right\|_V \leqslant \sum_{n=k+1}^{m} \|a_n\|_V = \sum_{n=1}^{m} \|a_n\|_V - \sum_{n=1}^{k} \|a_n\|_V \\ &= \left| \left(\sum_{n=1}^{m} \|a_n\|_V - r \right) + \left(r - \sum_{n=1}^{k} \|a_n\|_V \right) \right| \leqslant \left| \sum_{n=1}^{m} \|a_n\|_V - r \right| + \left| \sum_{n=1}^{k} \|a_n\|_V - r \right| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

So (x_m) is a Cauchy sequence in V, therefore it converges in V, meaning that the series

$$\sum_{n=1}^{\infty} a_n$$

converges in V.

In the other direction, suppose that every series that converges absolutely also converges in V, and let (a_n) be a Cauchy sequence in V. For each $\varepsilon > 0$ there exists $N(\varepsilon) \in \mathbf{N}$ such that for all $n \geqslant N$ we have $\|a_n - a_N\|_V < \varepsilon$.

Taking $\varepsilon = \frac{1}{2}, \frac{1}{2^2}, \ldots$ we get

$$\begin{gathered} n_1 \geqslant 1 \text{ such that } \left\|a_n - a_{n_1}\right\|_V < \frac{1}{2} \text{ for all } n \geqslant n_1, \\ n_2 > n_1 \text{ such that } \left\|a_n - a_{n_2}\right\|_V < \frac{1}{2^2} \text{ for all } n \geqslant n_2, \\ \vdots \\ n_k > n_{k-1} \text{ such that } \left\|a_n - a_{n_k}\right\|_V < \frac{1}{2^k} \text{ for all } n \geqslant n_k, \\ \vdots \end{gathered}$$

In particular, for all $k \in \mathbf{N}$ we have

$$\left\|a_{n_{k+1}} - a_{n_k}\right\|_V < \frac{1}{2^k},$$

so that

$$\sum_{k=1}^{\infty} \left\|a_{n_{k+1}} - a_{n_k}\right\|_V \leqslant \sum_{k=1}^{\infty} \frac{1}{2^k} = 1,$$

which implies that the series

$$\sum_{k=1}^{\infty} \left(a_{n_{k+1}} - a_{n_k}\right) \qquad \text{absolutely converges,}$$

which by our assumption implies that the series

$$\sum_{k=1}^{\infty} \left(a_{n_{k+1}} - a_{n_k}\right) \qquad \text{converges.}$$

Therefore the sequence of partial sums $\left(a_{n_k} - a_{n_1}\right)$ (observe the telescoping behaviour) converges as $k \longrightarrow \infty$, so the subsequence $\left(a_{n_k}\right)$ of (a_n) converges, which by means that (a_n) converges. (Since any Cauchy sequence with a convergent subsequence is itself convergent, see Question 6 on Assignment 1.) □

Proposition 3.31. *A normed space V is separable (as a metric space, i.e. has a countable dense subset) if and only if*

$$V = \overline{\operatorname{Span}(S)} \qquad \textit{for a countable subset } S \subseteq V.$$

Sketch of proof. (A less sketchy proof of the finite-dimensional case is in Tutorial Question 9.4.)

Suppose $V = \overline{\operatorname{Span}(S)}$ with S countable. Let

$$D = \begin{cases} \operatorname{Span}_{\mathbf{Q}}(S) & \text{if } \mathbf{F} = \mathbf{R} \\ \operatorname{Span}_{\mathbf{Q}[i]}(S) & \text{if } \mathbf{F} = \mathbf{C}. \end{cases}$$

Then D is countable (as S, $\mathbf{Q}$, $\mathbf{Q}[i]$ are countable). But also D is dense in $\operatorname{Span}(S)$ (because $\mathbf{Q}$ is dense in $\mathbf{R}$, and $\mathbf{Q}[i]$ is dense in $\mathbf{C}$), so D is dense in V, hence V is separable.

The converse is much easier: if $V = \overline{D}$ with D countable, then

$$V = \overline{D} \subseteq \overline{\operatorname{Span}(D)} \subseteq V.$$

□

A *Schauder basis* of a normed space V is a sequence $e_1, e_2, \dots$ of **unit** vectors of V such that for every $v \in V$ there exists a **unique** sequence of coefficients $\alpha_1, \alpha_2, \dots \in \mathbf{F}$ with

$$v = \sum_{n=1}^{\infty} \alpha_n e_n,$$

which should be read as meaning that the series on the right hand side converges to $v \in V$.

If V has a Schauder basis, then

$$V = \overline{\operatorname{Span}\{e_1, e_2, \dots\}},$$

so in particular V is separable. Note that not every separable normed space has a Schauder basis.

Example 3.32. For any $1 \leqslant p < \infty$, the sequence space ℓ^p has Schauder basis $\{e_1, e_2, \dots\}$, where

$$e_n = (0, \dots, 0, 1, 0, \dots) \quad \text{with the 1 in the } n\text{-th spot.}$$

In particular, ℓ^p is separable.

Solution. This is an essentially trivial exercise in checking the definition.

Take an arbitrary element $v = (v_n) \in \ell^p$, then

$$\sum_{n=1}^{\infty} |v_n|^p$$

converges with sum $\|v\|^p$.

I claim that the series

$$\sum_{n=1}^{\infty} v_n e_n$$

converges to v with respect to the ℓ^p-norm:

$$\left\| v - \sum_{n=1}^{m} v_n e_n \right\|_{\ell^p}^p = \left\| (0, \dots, 0, v_{m+1}, v_{m+2}, v_{m+3}, \dots) \right\|_{\ell^p}^p = \sum_{n=m+1}^{\infty} |v_n|^p,$$

and the latter converges to 0 as $m \longrightarrow \infty$.

The uniqueness of the sequence of coefficients follows from the fact that

$$(v_1, v_2, \dots) = \sum_{n=1}^{\infty} v_n e_n = v = \sum_{n=1}^{\infty} u_n e_n = (u_1, u_2, \dots)$$

implies $v_n = u_n$ for all $n \in \mathbf{N}$. □

3.7. Dual normed spaces and completeness of sequence spaces

You may want to have a look at Appendices A.2.1 and A.2.2 and read the discussion of bilinear maps and dual spaces. We will only touch on some basic points here, focussing on the new aspects coming from the norm.

If U, V, W are vector spaces over $\mathbf{F}$, a *bilinear map* $\beta \colon U \times V \longrightarrow W$ is a function such that

$$\beta(au_1 + bu_2, v) = a\beta(u_1, v) + b\beta(u_2, v)$$
$$\beta(u, av_1 + bv_2) = a\beta(u, v_1) + b\beta(u, v_2)$$

for all $u, u_1, u_2 \in U$, $v, v_1, v_2 \in V$, $a, b \in \mathbf{F}$.

For instance, given $n \in \mathbf{N}$, there is a bilinear map $\beta \colon \mathbf{F}^n \times \mathbf{F}^n \longrightarrow \mathbf{F}$ given by

$$\beta(u, v) = \sum_{k=1}^{n} u_k v_k.$$

As described in Appendix A.2.2, this defines a linear map $\mathbf{F}^n \longrightarrow (\mathbf{F}^n)^\vee$, $u \longmapsto u^\vee$, given by $u^\vee(v) = \beta(u, v)$.

We'd like to do the same with (subspaces of) $\mathbf{F}^{\mathbf{N}}$: define a bilinear map $\beta \colon \mathbf{F}^{\mathbf{N}} \times \mathbf{F}^{\mathbf{N}} \longrightarrow \mathbf{F}$ by the formula

$$\beta(u, v) = \sum_{n=1}^{\infty} u_n v_n.$$

Of course this would feel more comfortable if we knew that the series $\sum u_n v_n$ actually converges! And of course that does not happen for arbitrary $u, v \in \mathbf{F}^{\mathbf{N}}$, but we can establish some situations where it does work, as follows.

If $p \geqslant 1$, we say that the real number q satisfying

$$\frac{1}{p} + \frac{1}{q} = 1$$

is the *Hölder conjugate* of p. It is easy to see that $q \geqslant 1$. Note that this includes the degenerate pair $p = 1$, $q = \infty$.

Proposition 3.33 (Hölder's Inequality)**.** *Suppose p and q are Hölder conjugate and let $u = (u_n) \in \ell^p$, $v = (v_n) \in \ell^q$. Then*

$$\sum_{n=1}^{\infty} |u_n v_n| \leqslant \|u\|_{\ell^p} \|v\|_{\ell^q}.$$

Proof. We prove the non-degenerate case $p, q \in \mathbf{R}_{>1}$ and leave the (simpler) degenerate one to Tutorial Question 9.1.

Let $x = (x_n) \in \ell^p$, $y = (y_n) \in \ell^q$. For each $n \in \mathbf{N}$ we have

$$|x_n y_n| = \left(|x_n|^p\right)^{1/p}\left(|y_n|^q\right)^{1/q} \leqslant \frac{1}{p}|x_n|^p + \frac{1}{q}|y_n|^q,$$

by an application of Proposition 3.19 part (b), namely $s^a t^b \leqslant as + bt$ where $a + b = 1$.

Therefore

$$\sum_{n=1}^{\infty}|x_n y_n| \leqslant \sum_{n=1}^{\infty}\left(\frac{1}{p}|x_n|^p + \frac{1}{q}|y_n|^q\right) = \frac{1}{p}\|x\|_{\ell^p}^p + \frac{1}{q}\|y\|_{\ell^q}^q.$$

Now start with $u \in \ell^p$, $v \in \ell^q$ and set

$$x = \frac{1}{\|u\|_{\ell^p}}u, \qquad y = \frac{1}{\|v\|_{\ell^q}}v,$$

so that we obtain

$$\frac{\sum_{n=1}^{\infty}|u_n v_n|}{\|u\|_{\ell^p}\|v\|_{\ell^q}} \leqslant \frac{1}{p} + \frac{1}{q} = 1.$$ □

Before we give the main result of this section, we should extend the notion of continuous linear map to the setting of bilinear maps.

If U, V, W are normed spaces, a bilinear map $\beta\colon U \times V \longrightarrow W$ is *Lipschitz* if there exists $c > 0$ such that

$$\|\beta(u,v)\|_W \leqslant c\|u\|_U\|v\|_V \qquad \text{for all } u \in U, v \in V.$$

Proposition 3.34. *If U, V, W are normed spaces, a bilinear map $\beta\colon U \times V \longrightarrow W$ is Lipschitz if and only if it is continuous.*

Proof. Somewhat tedious, following the example of Proposition 3.25. See Exercise 3.12. □

Beware: in contrast to the linear case, continuous bilinear maps are almost never uniformly continuous. See Exercise 3.13.

Theorem 3.35. *If p, q are Hölder conjugates, then $\beta\colon \ell^p \times \ell^q \longrightarrow \mathbf{F}$ given by*

$$\beta(u,v) = \sum_{n=1}^{\infty} u_n v_n$$

is a continuous bilinear map.

Moreover, if $p, q > 1$, the resulting continuous linear map

$$u \longmapsto u^{\vee} : \ell^p \longrightarrow \left(\ell^q\right)^{\vee}$$

is bijective and distance-preserving, hence an isometry $\ell^p \cong \left(\ell^q\right)^{\vee}$.

Proof. By Hölder's Inequality, the series defining $\beta(u,v)$ converges absolutely. It is then straightforward to check that β is bilinear.

Conveniently, Hölder's Inequality also tells us that β is a Lipschitz bilinear form, hence continuous.

Now we know that $u^{\vee} \in \left(\ell^q\right)^{\vee}$, but we can (and will) say something more precise.

We start by proving the surjectivity of $u \longmapsto u^{\vee}$.

Let $\varphi \in (\ell^q)^\vee$ and let $v \in \ell^q$. Let $\{e_1, e_2, \dots\}$ be the Schauder basis for ℓ^q discussed in Example 3.32. We have

$$\varphi(v) = \varphi\left(\sum_{n=1}^{\infty} v_n e_n\right) = \sum_{n=1}^{\infty} v_n \varphi(e_n) \qquad \text{and } \|e_n\|_{\ell^q} = 1 \text{ for all } n \in \mathbf{N}.$$

Define $u_n = \varphi(e_n)$ and $u = (u_n)$. If we show that $u \in \ell^p$ then we have $\varphi(v) = u^\vee(v)$ and we're done.

For any $m \in \mathbf{N}$, consider (ignore all the u_n's that are zero, as they do not contribute to the sums):

$$x = \sum_{n=1}^{m} \frac{|u_n|^p}{u_n} e_n = \left(\frac{|u_1|^p}{u_1}, \dots, \frac{|u_m|^p}{u_m}, 0, 0, \dots\right),$$

so that

$$\|x\|_{\ell^q} = \left(\sum_{n=1}^{m} \left(|u_n|^{p-1}\right)^q\right)^{1/q} = \left(\sum_{n=1}^{m} |u_n|^p\right)^{1/q}.$$

With this in mind, we have

$$\begin{aligned}\sum_{n=1}^{m} |u_n|^p &= \left|\sum_{n=1}^{m} \frac{|u_n|^p}{u_n} u_n\right| = \left|\sum_{n=1}^{m} \varphi\left(\frac{|u_n|^p}{u_n} e_n\right)\right| \\ &= |\varphi(x)| \leqslant \|\varphi\| \, \|x\|_{\ell^q} = \|\varphi\| \left(\sum_{n=1}^{m} |u_n|^p\right)^{1/q}.\end{aligned}$$

Therefore

$$\left(\sum_{n=1}^{m} |u_n|^p\right)^{1/p} = \left(\sum_{n=1}^{m} |u_n|^p\right)^{1-1/q} \leqslant \|\varphi\|.$$

As this holds for all $m \in \mathbf{N}$, we conclude that the series converges, so $u \in \ell^p$. We also proved that $\|u\|_{\ell^p} \leqslant \|\varphi\| = \|u^\vee\|$.

Let $v \neq 0$. By Hölder's Inequality

$$\frac{|u^\vee(v)|}{\|v\|_{\ell^q}} \leqslant \|u\|_{\ell^p},$$

so taking supremum we get $\|u^\vee\| \leqslant \|u\|_{\ell^p}$.

As we established both inequalities, we conclude that $u \longmapsto u^\vee$ is a surjective distance-preserving map from ℓ^p to $(\ell^q)^\vee$, hence an isometry. ☐

Corollary 3.36. *If $p > 1$ then ℓ^p is a Banach space.*

Proof. Follows as $\ell^p \cong (\ell^q)^\vee$ and all dual normed spaces are Banach. ☐

Proposition 3.37. *The sequence space ℓ^2 of square-summable sequences is a Hilbert space.*

Proof. Consider the function $\langle \cdot, \cdot \rangle : \ell^2 \times \ell^2 \longrightarrow \mathbf{F}$ given by

$$\langle a, b \rangle = \sum_{n=1}^{\infty} a_n \bar{b}_n.$$

We use the Cauchy–Schwarz Inequality (Proposition A.6) to see that this converges. For any $m \in \mathbf{N}$, $(a_1, \dots, a_m), (b_1, \dots, b_m) \in \mathbf{F}^m$ so by Cauchy–Schwarz we have

$$\left|\sum_{n=1}^{m} a_n \bar{b}_n\right| \leqslant \left(\sum_{n=1}^{m} a_n \bar{a}_n\right)^{1/2} \left(\sum_{n=1}^{m} b_n \bar{b}_n\right)^{1/2} = \left(\sum_{n=1}^{m} |a_n|^2\right)^{1/2} \left(\sum_{n=1}^{m} |b_n|^2\right)^{1/2}.$$

Taking limits as $m \longrightarrow \infty$, the right hand side becomes $\|a\|_{\ell^2} \|b\|_{\ell^2}$, which is finite since $a, b \in \ell^2$.

The inner product properties are clear. So is the fact that the norm defined by this inner product is exactly the ℓ^2-norm, so we get a Hilbert space by Corollary 3.36. ☐

The rest of the ℓ^p spaces are **not** Hilbert spaces, see Tutorial Question 9.2.

3.8. Orthogonality and projections

For the next few sections, we will explore some special properties of inner product spaces.

Given a normed space V, a *projection* is a continuous linear map $\pi \in L(V)$ such that $\pi^2 = \pi$.

Proposition 3.38. *Let $\pi \in L(V)$ be a projection.*

(a) The map $\mathrm{id}_V - \pi$ is also a projection.

(b) $\mathrm{im}(\pi) = \ker(\mathrm{id}_V - \pi)$ and $\mathrm{im}(\mathrm{id}_V - \pi) = \ker(\pi)$. In particular, the image of a projection is a closed subspace.

(c) We have

$$V = \mathrm{im}(\pi) \oplus \ker(\pi).$$

Solution. (a) Since both id_V and π are continuous and linear, so is $\mathrm{id}_V - \pi$. Also, we have

$$\left(\mathrm{id}_V - \pi\right) \circ \left(\mathrm{id}_V - \pi\right) = \mathrm{id}_V - \pi - \pi + \pi \circ \pi = \mathrm{id}_V - \pi.$$

(b) If $v \in \mathrm{im}(\pi)$ then $v = \pi(w)$ so that

$$\left(\mathrm{id}_V - \pi\right)(v) = v - \pi(v) = \pi(w) - \pi^2(w) = \pi(w) - \pi(w) = 0,$$

so $v \in \ker\left(\mathrm{id}_V - \pi\right)$.

Conversely, if $v \in \ker\left(\mathrm{id}_V - \pi\right)$ then $v - \pi(v) = 0$ so $v = \pi(v) \in \mathrm{im}(\pi)$.

The other identity follows by applying the first identity to the projection $\mathrm{id}_V - \pi$.

Since the image of π is the kernel of $\mathrm{id}_V - \pi$, it is a closed subspace, as the kernel of any linear continuous map is a closed subspace.

(c) We need to prove that $V = \mathrm{im}(\pi) + \ker(\pi)$ and that $\mathrm{im}(\pi) \cap \ker(\pi) = \{0\}$.

Given $v \in V$, we have

$$v = \pi(v) + \left(\mathrm{id}_V - \pi\right)(v) \in \mathrm{im}(\pi) + \ker(\pi).$$

If

$$w \in \mathrm{im}(\pi) \cap \ker(\pi) = \ker\left(\mathrm{id}_V - \pi\right) \cap \ker(\pi),$$

then

$$w = \pi(w) + \left(\mathrm{id}_V - \pi\right)(w) = 0 + 0 = 0.$$ □

Example 3.39. Take $V = \mathbf{R}^2$ with the Euclidean norm. The matrix

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

satisfies $A^2 = A$, so it defines a projection. It is easy to see that $\mathrm{im}(A)$ is the diagonal $y = x$ in $\mathbf{R}^2$, and $\ker(A)$ is the y-axis.

The complementary projection is given by the matrix

$$I - A = \begin{pmatrix} 0 & 0 \\ -1 & 1 \end{pmatrix},$$

where $\mathrm{im}(I - A)$ is the y-axis and $\ker(I - A)$ is the diagonal $y = x$.

▮

Given a subset S of an inner product space V, we define the *orthogonal complement of S* by

$$S^\perp = \{v \in V : \langle v, s\rangle = 0 \text{ for all } s \in S\}.$$

Proposition 3.40. *For any subset $S \subseteq V$, $S^\perp$ is a closed subspace of V.*

Proof. That $S^\perp$ is a vector subspace of V follows easily from the linearity of $\langle\cdot,\cdot\rangle$ in the first variable.

That $S^\perp$ is closed in V follows from the continuity of $\langle\cdot,\cdot\rangle$ in the first variable. □

If V is an inner product space, an *orthogonal projection* is a projection π such that $\ker(\pi) = \big(\operatorname{im}(\pi)\big)^\perp$, so that we have (by Proposition 3.38)

$$V = \operatorname{im}(\pi) \oplus \big(\operatorname{im}(\pi)\big)^\perp.$$

Recall from Exercise 2.72 that for any subset $Y \subseteq X$ of a metric space, we can define a function $d_Y : X \longrightarrow \mathbf{R}_{\geqslant 0}$ that gives the distance to the set Y:

$$d_Y(x) = \inf_{y\in Y} d(x,y).$$

Theorem 3.41 (Hilbert Projection Theorem, Part I)**.** *Let H be a Hilbert space and Y a closed **convex subset** of H. For any $x \in H$, there exists a unique $y_{min} \in Y$ that realises the distance between x and Y:*

$$d_Y(x) = d\big(x, y_{min}\big) = \big\|x - y_{min}\big\|.$$

In other words, y_{min} is the unique point of Y that is as close as possible to x.

Proof. Let

$$D = d_Y(x) = \inf_{y\in Y} d(x,y).$$

Take a sequence (y_n) in Y such that

$$\big(\|x - y_n\|\big) = \big(d(x,y_n)\big) \longrightarrow D.$$

I claim that the sequence (y_n) is Cauchy.

Let $\varepsilon > 0$. Note that

$$\big(\|x - y_n\|^2\big) \longrightarrow D^2,$$

so there exists $N \in \mathbf{N}$ such that

$$\big|\|x - y_n\|^2 - D^2\big| \leqslant \frac{\varepsilon}{4} \qquad \text{for all } n \geqslant N.$$

Let $m, n \geqslant N$. By the Parallelogram Law:

$$\big\|(y_n - x) + (y_m - x)\big\|^2 + \big\|(y_n - x) - (y_m - x)\big\|^2 = 2\|y_n - x\|^2 + 2\|y_m - x\|^2,$$

so that

$$\begin{aligned}\|y_n - y_m\|^2 &= 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - \big\|(y_n + y_m) - 2x\big\|^2 \\ &= 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4\left\|\frac{y_n + y_m}{2} - x\right\|^2.\end{aligned}$$

At this point we notice that since $y_n, y_m \in Y$ and Y is convex, $(1/2)y_n + (1/2)y_m \in Y$; we can then continue with

$$\begin{aligned} 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4\left\|\frac{y_n + y_m}{2} - x\right\|^2 &\leqslant 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4D^2 \\ &= 2\big(\|y_n - x\|^2 - D^2\big) + 2\big(\|y_m - x\|^2 - D^2\big) \\ &< \varepsilon. \end{aligned}$$

So (y_n) is Cauchy in Y, which is complete (being a closed subset of the Hilbert space H). Therefore (y_n) converges in Y to some point that we will call $y_{\min}$. Since the distance function is continuous, we have

$$d(x, y_{\min}) = \lim_{n \longrightarrow \infty} d(x, y_n) = D = d_Y(x).$$

It remains to prove the uniqueness of $y_{\min}$. Suppose $y' \in Y$ satisfies $d(x, y') = D$. By the Parallelogram Law

$$\big\|(y_{\min} - x) + (y' - x)\big\|^2 + \big\|(y_{\min} - x) - (y' - x)\big\|^2 = 2\|y_{\min} - x\|^2 + 2\|y' - x\|^2,$$

so that

$$\|y_{\min} - y'\|^2 = 2\|y_{\min} - x\|^2 + 2\|y' - x\|^2 - \|(y_{\min} + y') - 2x\|^2 \leqslant 2D^2 + 2D^2 - 4D^2 = 0,$$

which implies $y' = y_{\min}$. □

Theorem 3.42 (Hilbert Projection Theorem, Part II)**.** *Let H be a Hilbert space and W a closed **vector subspace** of H. Let $x \in H$ and let y_{min} be the unique point of W that realises the distance between x and W, as given by the Hilbert Projection Theorem, Part I. For any $y \in W$, we have*

$$y = y_{min} \qquad \text{if and only if} \qquad x - y \in W^{\perp}.$$

The map $\pi\colon H \longrightarrow H$ given by $\pi(x) = y_{min}$ is an orthogonal projection with image W. In particular, we have a decomposition

$$H = W \oplus W^{\perp}.$$

Proof. **First we prove that $x - y_{\mathbf{min}} \in W^{\perp}$.**

Let $w \in W$ be a unit vector, so $\|w\| = 1$. Letting $\alpha = \langle x - y_{\min}, w\rangle$, we want to show that $\alpha = 0$.

Set $v = x - \big(y_{\min} + \alpha w\big)$, then

$$\begin{aligned} \langle v, w\rangle &= \langle x - y_{\min} - \alpha w, w\rangle \\ &= \langle x - y_{\min}, w\rangle - \alpha\langle w, w\rangle \\ &= \alpha - \alpha = 0, \end{aligned}$$

so $v \perp w$. Therefore

$$\big\|x - y_{\min}\big\|^2 = \|v + \alpha w\|^2 = \|v\|^2 + |\alpha|^2\,\|w\|^2 = \|v\|^2 + |\alpha|^2 \geqslant \|v\|^2,$$

in other words

$$\big\|x - y_{\min}\big\| \geqslant \big\|x - (y_{\min} + \alpha w)\big\|.$$

By the minimality and uniqueness of $y_{\min}$, we must have $y_{\min} = y_{\min} + \alpha w$, so $\alpha = 0$.

Next we show that if $y \in W$ and $x - y \in W^{\perp}$ then $y = y_{\mathbf{min}}$.

We have

$$\begin{aligned} x-y\in W^{\perp} &\Rightarrow \langle x-y,w\rangle = 0 && \text{for all } w\in W\\ &\Rightarrow \langle x-y,w-y\rangle = 0 && \text{for all } w\in W\\ &\Rightarrow \|x-w\|^2 = \|x-y\|^2+\|w-y\|^2 && \text{for all } w\in W\\ &\Rightarrow \|x-w\|^2 \geqslant \|x-y\|^2 && \text{for all } w\in W, \end{aligned}$$

implying that $y\in W$ is closest to x; by the uniqueness of $y_{\min}$, we conclude that $y=y_{\min}$.

We now move on to the function π. As we have just seen, for each $x\in H$, $\pi(x)$ is the unique element of W with the property that $x-\pi(x)\in W^{\perp}$.

We check that π is linear.

If $x_1,x_2\in H$, we have $\pi(x_1)+\pi(x_2)\in W$ and

$$(x_1+x_2)-\big(\pi(x_1)+\pi(x_2)\big) = \big(x_1-\pi(x_1)\big)+\big(x_2-\pi(x_2)\big)\in W^{\perp},$$

so $\pi(x_1)+\pi(x_2)=\pi(x_1+x_2)$.

Similarly, if $x\in H$ and $\alpha\in\mathbf{F}$ we have $\alpha\pi(x)\in W$ and

$$\alpha x-\alpha\pi(x)=\alpha\big(x-\pi(x)\big)\in W^{\perp},$$

so $\alpha\pi(x)=\pi(\alpha x)$.

We check that π is Lipschitz, hence continuous.

For any $x\in H$, we have $\pi(x)\in W$ and $x-\pi(x)\in W^{\perp}$, so $\big(x-\pi(x)\big)\perp\pi(x)$ and

$$\|x\|^2=\big\|(x-\pi(x))+\pi(x)\big\|^2=\|x-\pi(x)\|^2+\|\pi(x)\|^2\geqslant\|\pi(x)\|^2,$$

so $\|\pi(x)\|\leqslant\|x\|$.

We check that π is a projection with image W.

Certainly $\operatorname{im}\pi\subseteq W$. If $y\in W$ then $\pi(y)=y$ (closest point to y is y itself), so in fact $\operatorname{im}\pi=W$. Hence for all $x\in H$ we get $\pi^2(x)=\pi\big(\pi(x)\big)=\pi(x)$, so $\pi^2=\pi$.

Finally, we check that π is an orthogonal projection.

We want to show that $W^{\perp}=\ker\pi$. But $x\in W^{\perp}$ if and only if $x-0\in W^{\perp}$ if and only if $\pi(x)=0$ if and only if $x\in\ker\pi$. □

Corollary 3.43. *If H is a Hilbert space and S is a subset of H, then*

$$\big(S^{\perp}\big)^{\perp}=\overline{\operatorname{Span}(S)}.$$

Proof. Let $W=\overline{\operatorname{Span}(S)}$.

By Exercise 3.23, for any inner product space V and subset S we have $S\subseteq\big(S^{\perp}\big)^{\perp}$. Since $\big(S^{\perp}\big)^{\perp}$ is a closed subspace, we have

$$W=\overline{\operatorname{Span}(S)}\subseteq\big(S^{\perp}\big)^{\perp},$$

so it remains to show that $\big(S^{\perp}\big)^{\perp}\subseteq W$.

But Exercise 3.23 also tells us that

$$S^{\perp}=\overline{\operatorname{Span}(S)}^{\perp}=W^{\perp},$$

so what we need to prove is that $\big(W^{\perp}\big)^{\perp}\subseteq W$ for any closed subspace W of a Hilbert space H.

Let $x\in\big(W^{\perp}\big)^{\perp}$. By the Hilbert Projection Theorem Part II, we can decompose

$$H=W\oplus W^{\perp}.$$

So we have $x=y+z$ with $y\in W$ and $z\in W^{\perp}$. Then

$$0=\langle x,z\rangle=\langle y+z,z\rangle=\langle y,z\rangle+\langle z,z\rangle=0+\|z\|^2,$$

implying that $z=0$ and $x=y\in W$. □

3.9. Duality in Hilbert spaces

Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space. Similarly to the case of a bilinear form, for any $w \in V$ the inner product gives rise to a function

$$w^\vee : V \longrightarrow \mathbf{F} \qquad \text{defined by } w^\vee(v) = \langle v, w\rangle$$

that is linear:

$$\begin{aligned} w^\vee(\alpha_1 v_1 + \alpha_2 v_2) &= \langle \alpha_1 v_1 + \alpha_2 v_2, w\rangle \\ &= \alpha_1\langle v_1, w\rangle + \alpha_2\langle v_2, w\rangle \\ &= \alpha_1 w^\vee(v_1) + \alpha_2 w^\vee(v_2) \qquad \text{for all } v_1, v_2 \in V, \alpha_1, \alpha_2 \in \mathbf{F}, \end{aligned}$$

and Lipschitz, hence continuous:

$$|w^\vee(v)| = |\langle v, w\rangle| \leqslant \|v\|\,\|w\| \qquad \text{for all } v \in V,$$

where we used the Cauchy–Schwarz Inequality (and noted that w is fixed hence $\|w\|$ is constant).

We conclude that $w^\vee \in V^\vee = L(V, \mathbf{F})$. Varying w now, we obtain a function $\Phi\colon V \longrightarrow V^\vee$ given by $w \longmapsto w^\vee$.

Proposition 3.44. *The map Φ is conjugate-linear and distance-preserving (hence injective).*

Proof. Let $w_1, w_2 \in V$, then for any $v \in V$ we have

$$\begin{aligned} (\Phi(w_1 + w_2))(v) &= (w_1 + w_2)^\vee(v) \\ &= \langle v, w_1 + w_2\rangle \\ &= \langle v, w_1\rangle + \langle v, w_2\rangle \\ &= w_1^\vee(v) + w_2^\vee(v) \\ &= (\Phi(w_1) + \Phi(w_2))(v). \end{aligned}$$

If $w \in V$ and $\alpha \in \mathbf{F}$, then for any $v \in V$ we have

$$\begin{aligned} (\Phi(\alpha w))(v) &= (\alpha w)^\vee(v) \\ &= \langle v, \alpha w\rangle \\ &= \overline{\alpha}\,\langle v, w\rangle \\ &= \overline{\alpha}\, w^\vee(v) \\ &= (\overline{\alpha}\,\Phi(w))(v). \end{aligned}$$

It remains to check that Φ is norm-preserving (and hence distance-preserving):

$$\|\Phi(w)\|_{V^\vee} = \|w^\vee\|_{V^\vee} = \sup_{v\neq 0} \frac{|w^\vee(v)|}{\|v\|_V} = \sup_{v \neq 0} \frac{|\langle v, w\rangle|}{\|v\|_V} \leqslant \|w\|_V,$$

where we used the Cauchy–Schwarz Inequality. If $w = 0$, then we certainly have equality $\|\Phi(0)\| = 0 = \|0\|$. Otherwise, note that in Cauchy–Schwarz we can take $v = w$ and obtain an equality, so that for all $w \in V$ we have $\|\Phi(w)\| = \|w\|$. □

In the case of a Hilbert space, we can say something very precise about the map Φ:

Theorem 3.45 (Riesz Representation Theorem)**.** *If H is a Hilbert space, then the map $\Phi\colon H \longrightarrow H^\vee$ is surjective, hence a conjugate-linear isometry.*

In other words, for any $\varphi \in H^\vee$ there exists a unique $z \in H$ such that $\varphi(x) = z^\vee(x) = \langle x, z\rangle$ for all $x \in H$.

Proof. Let $\varphi \in H^\vee$. The uniqueness of z follows from the injectivity of Φ, proved in Proposition 3.44. So we just need to prove the existence of z. If $\varphi = 0$, then we can take $z = 0$ and be done.

So suppose $\varphi \neq 0$. Therefore $\ker\varphi \neq H$; since H is a Hilbert space and $\ker\varphi$ is a closed subspace of H, we have by the Hilbert Projection Theorem Part II (Theorem 3.42) that

$$H = (\ker\varphi) \oplus (\ker\varphi)^\perp,$$

so $\ker\varphi \neq H$ implies that $(\ker\varphi)^\perp \neq 0$.

Take a unit vector $u \in (\ker\varphi)^\perp$ and let $z = \overline{\varphi(u)}\,u$. For all $x \in H$ we have

$$\begin{aligned}
\big\langle \varphi(x)u - \varphi(u)x, u\big\rangle &= \varphi(x) - \varphi(u)\langle x, u\rangle \\
&= \varphi(x) - \big\langle x, \overline{\varphi(u)}\,u\big\rangle \\
&= \varphi(x) - \langle x, z\rangle \\
&= \varphi(x) - z^\vee(x).
\end{aligned}$$

However, for any $x, u \in H$ we have

$$\varphi\big(\varphi(x)u - \varphi(u)x\big) = 0 \qquad \text{hence} \qquad \varphi(x)u - \varphi(u)x \in \ker\varphi,$$

so in the previous calculation, having chosen $u \in (\ker\varphi)^\perp$, we have $\big\langle \varphi(x)u - \varphi(u)x, u\big\rangle = 0$. Therefore $\varphi(x) - z^\vee(x) = 0$. ☐

Example 3.46. Consider $\psi\colon \ell^2 \longrightarrow \mathbf{F}$ given by

$$\psi\big(x_1, x_2, \dots\big) = \sum_{n=1}^\infty \frac{x_n}{2^n}.$$

The Riesz Representation Theorem says that there exists a unique $z \in \ell^2$ such that $\psi(x) = \langle x, z\rangle$ for all $x \in \ell^2$.

By inspection, if we take $z = (1/2^n)$ then $\psi(x) = \langle x, z\rangle$ for all $x \in \ell^2$.

It remains to check that $z \in \ell^2$:

$$\|z\|_{\ell^2}^2 = \sum_{n=1}^\infty \frac{1}{2^{2n}} = \frac{1}{1 - \frac{1}{4}} - 1 = \frac{1}{3}.$$

We also know that $z \longmapsto z^\vee$ is distance-preserving, so

$$\|\psi\| = \|z^\vee\| = \|z\| = \frac{1}{\sqrt{3}}.$$

3.10. Adjoint maps

We often encounter expressions of the kind $\langle f(x), y\rangle$, where f is a continuous linear map. A very useful trick consists of moving f from the first to the second variable in the inner product, at the cost of perhaps altering f in some way, as we are about to see.

Let X, Y be Hilbert spaces and let $\Phi_X : X \longrightarrow X^\vee$, $\Phi_Y : Y \longrightarrow Y^\vee$ be the corresponding conjugate-linear isometries. Suppose $f : X \longrightarrow Y$ is a continuous linear map. This induces a linear map $f^\vee : Y^\vee \longrightarrow X^\vee$ by setting $f^\vee(\varphi) = \varphi \circ f$ for all $\varphi \in Y^\vee$, see Proposition A.4 (whose statement asks for finite-dimensionality, but whose proof does not require it).

We can illustrate the situation via a diagram:

$$\begin{array}{ccc} Y^\vee & \xrightarrow{\quad f^\vee \quad} & X^\vee \\ {\scriptstyle \Phi_Y}\uparrow & & \uparrow{\scriptstyle \Phi_X} \\ Y & \overset{f^*}{\cdots\cdots\cdots\rightarrow} & X \end{array}$$

We complete this by defining the bottom arrow $f^* : Y \longrightarrow X$ in the unique way that makes the diagram commute:

$$f^* := \Phi_X^{-1} \circ f^\vee \circ \Phi_Y, \qquad \text{in other words} \qquad \Phi_X \circ f^* = f^\vee \circ \Phi_Y.$$

Proposition 3.47. *Let X, Y be Hilbert spaces. Given $f \in L(X, Y)$, the function $f^* : Y \longrightarrow X$ defined above is the unique element of $L(Y, X)$ that satisfies*

$$\big\langle f(x), y \big\rangle_Y = \big\langle x, f^*(y) \big\rangle_X \qquad \textit{for all } x \in X, y \in Y.$$

It is called the adjoint *of f.*

Proof. From the definition $f^* = \Phi_X^{-1} \circ f^\vee \circ \Phi_Y$ we see that f^* is continuous and additive. Its linearity is a consequence of the conjugate-linearity of Φ_X and Φ_Y cancelling each other out:

$$\begin{aligned} f^*(\alpha y) &= \Phi_X^{-1}\Big(f^\vee\big(\Phi_Y(\alpha y)\big)\Big) \\ &= \Phi_X^{-1}\Big(f^\vee\big(\overline{\alpha}\,\Phi_Y(y)\big)\Big) \\ &= \Phi_X^{-1}\Big(\overline{\alpha}\,f^\vee\big(\Phi_Y(y)\big)\Big) \\ &= \alpha\,\Phi_X^{-1}\Big(f^\vee\big(\Phi_Y(y)\big)\Big) \\ &= \alpha f^*(y). \end{aligned}$$

For all $x \in X$, $y \in Y$ we have

$$\begin{aligned} \big\langle f(x), y \big\rangle_Y &= \Phi_Y(y)\big(f(x)\big) \\ &= \big(\Phi_Y(y) \circ f\big)(x) \\ &= f^\vee\big(\Phi_Y(y)\big)(x) \\ &= \big(f^\vee \circ \Phi_Y\big)(y)(x) \\ &= \big(\Phi_X \circ f^*\big)(y)(x) \\ &= \Phi_X\big(f^*(y)\big)(x) \\ &= \big\langle x, f^*(y) \big\rangle_X. \end{aligned}$$

For the uniqueness statement, suppose $g \in L(Y, X)$ satisfies

$$\big\langle f(x), y \big\rangle_Y = \big\langle x, g(y) \big\rangle_X \qquad \text{for all } x \in X, y \in Y.$$

Then for any $y \in Y$ we have

$$\big\langle x, g(y) \big\rangle_X = \big\langle x, f^*(y) \big\rangle_X \qquad \text{for all } x \in X,$$

which implies that $g(y) = f^*(y)$. We conclude that $g = f^*$. □

Proposition 3.48. *Given Hilbert spaces X and Y, the map $L(X,Y) \longrightarrow L(Y,X)$ given by $f \longmapsto f^*$ is a conjugate-linear distance-preserving involution such that*

(a) $(f \circ g)^* = g^* \circ f^*$;

(b) $\mathrm{id}_X^* = \mathrm{id}_X$;

(c) if f is invertible, then so is f^* and $\left(f^*\right)^{-1} = \left(f^{-1}\right)^*$;

(d) $\|f^* \circ f\| = \|f\|^2$;

(e) $\ker\left(f^*\right) = \left(\operatorname{im} f\right)^\perp$ *and* $\overline{\operatorname{im}\left(f^*\right)} = \left(\ker f\right)^\perp$;

(f) if $f\colon X \longrightarrow X$ and W is a closed subspace of X then W is f-invariant if and only if $W^\perp$ is $\left(f^*\right)$-invariant.

Proof. See Exercises 3.27 to 3.32 and Tutorial Question 11.4. □

Example 3.49. Consider a (continuous, automatically) linear map $f\colon \mathbf{C}^m \longrightarrow \mathbf{C}^n$ with standard matrix representation $A \in M_{n\times m}(\mathbf{C})$. Then the adjoint $f^*\colon \mathbf{C}^n \longrightarrow \mathbf{C}^m$ has standard matrix representation $\overline{A}^t$, the conjugate transpose of A.

See Tutorial Question 11.7 for the real vector space case; the argument is the same.

The notion of adjoint leads to certain special types of maps on Hilbert spaces. Let H be a Hilbert space and let $f \in L(H)$. We say that

(a) f is *self-adjoint* if $f^* = f$;

(b) f is *normal* if $f \circ f^* = f^* \circ f$.

Obviously every self-adjoint map is normal.

Example 3.50. If $H = \mathbf{C}^n$ for $n \in \mathbf{N}$, then a linear map $f\colon H \longrightarrow H$ is self-adjoint if and only if its standard matrix representation A is a *Hermitian matrix*, that is $A^* = A$.

For another example of a self-adjoint map, see Exercise 3.33.

3.11. Orthonormal bases

Let V be an inner product space. An *orthonormal system* is a subset $S \subseteq V$ consisting of unit vectors that are pairwise orthogonal, in other words for all $x, y \in S$ we have

$$\langle x, y\rangle = \delta_{x,y} = \begin{cases} 0 & \text{if } x \neq y \\ 1 & \text{if } x = y. \end{cases}$$

An *orthonormal basis*[1] of V is an orthonormal system $B \subseteq V$ such that $\overline{\operatorname{Span}(B)} = V$.

[1]Note that if V is infinite-dimensional, an orthonormal basis of V is not actually a basis of V in the sense of finite-dimensional linear algebra.

Example 3.51. Recall that $B = \{e_1, e_2, \dots\}$ is a Schauder basis for the sequence space ℓ^2, that is $\ell^2 = \overline{\operatorname{Span}(B)}$. But B is also an orthonormal system:

$$\langle e_k, e_n \rangle = \delta_{k,n}.$$

So B is an orthonormal basis of ℓ^2.

If fact, every **Hilbert space** has an orthonormal basis, see Exercise 3.34.

Theorem 3.52. *Let V be an inner product space and let $\{e_n : n \in \mathbf{N}\}$ be a countable orthonormal system.*

(a) (Bessel's Inequality) For all $x \in V$ we have

$$\sum_{n=1}^{\infty} \left|\langle x, e_n \rangle\right|^2 \leqslant \|x\|^2.$$

(b) (Parseval's Identity) If

$$x = \sum_{n=1}^{\infty} \alpha_n e_n \qquad \text{and} \qquad y = \sum_{n=1}^{\infty} \beta_n e_n,$$

then

$$\langle x, y \rangle = \sum_{n=1}^{\infty} \alpha_n \overline{\beta}_n$$
$$\alpha_n = \langle x, e_n \rangle$$
$$\sum_{n=1}^{\infty} \left|\langle x, e_n \rangle\right|^2 = \|x\|^2.$$

Proof.

(a) Let $x \in V$. Let $\alpha_n = \langle x, e_n \rangle$ for all $n \in \mathbf{N}$.

For $m \in \mathbf{N}$ let

$$s_m = \sum_{n=1}^{m} \alpha_n e_n.$$

Then

$$\langle x, s_m \rangle = \sum_{n=1}^{m} \langle x, \alpha_n e_n \rangle = \sum_{n=1}^{m} \overline{\alpha}_n \langle x, e_n \rangle = \sum_{n=1}^{m} \overline{\alpha}_n \alpha_n = \sum_{n=1}^{m} |\alpha_n|^2$$

and

$$\|s_m\|^2 = \sum_{n=1}^{m} \sum_{k=1}^{m} \langle \alpha_n e_n, \alpha_k e_k \rangle = \sum_{n=1}^{m} \sum_{k=1}^{m} \alpha_n \overline{\alpha}_k \langle e_n, e_k \rangle = \sum_{n=1}^{m} \alpha_n \overline{\alpha}_n = \sum_{n=1}^{m} |\alpha_n|^2,$$

so that

$$\begin{aligned} 0 &\leqslant \|x - s_m\|^2 \\ &= \|x\|^2 - \langle x, s_m \rangle - \langle s_m, x \rangle + \|s_m\|^2 \\ &= \|x\|^2 - \sum_{n=1}^{m} |\alpha_n|^2 - \sum_{n=1}^{m} |\alpha_n|^2 + \sum_{n=1}^{m} |\alpha_n|^2 \\ &= \|x\|^2 - \sum_{n=1}^{m} |\alpha_n|^2, \end{aligned}$$

which implies that

$$\sum_{n=1}^{m} |\alpha_n|^2 \leqslant \|x\|^2.$$

This holds for all $m \in \mathbf{N}$, so the left hand side forms an increasing sequence in m that is bounded above, hence it has a limit and the limit satisfies the same inequality.

(b) This is straightforward: using the continuity of the inner product, we have

$$\begin{aligned}
\langle x, y \rangle &= \left\langle \sum_{n=1}^{\infty} \alpha_n e_n, y \right\rangle \\
&= \sum_{n=1}^{\infty} \alpha_n \langle e_n, y \rangle \\
&= \sum_{n=1}^{\infty} \alpha_n \left\langle e_n, \sum_{m=1}^{\infty} \beta_m e_m \right\rangle \\
&= \sum_{n=1}^{\infty} \alpha_n \sum_{m=1}^{\infty} \overline{\beta}_m \langle e_n, e_m \rangle \\
&= \sum_{n=1}^{\infty} \alpha_n \overline{\beta}_n.
\end{aligned}$$

A simplified version of this calculation gives

$$\langle x, e_n \rangle = \alpha_n \quad \text{and} \quad \langle y, e_n \rangle = \beta_n \quad \text{for all } n \in \mathbf{N},$$

and the statement about the norm of x follows immediately from the above. □

Corollary 3.53. *Let H be a Hilbert space.*

(a) Let $\{e_n \colon n \in \mathbf{N}\}$ be a countable orthonormal system in H. Then

$$\sum_{n=1}^{\infty} \alpha_n e_n \quad \textit{converges in } H \textit{ if and only if} \quad (\alpha_n) \in \ell^2.$$

(b) Any countable orthonormal basis of H is a Schauder basis.

Proof.

(a) For $N \in \mathbf{N}$, let

$$S_N = \sum_{n=1}^{N} \alpha_n e_n, \qquad T_N = \sum_{n=1}^{N} |\alpha_n|^2.$$

For all $M \leqslant N$ we have

$$\left\| S_N - S_M \right\|^2 = \left\| \alpha_{M+1} e_{M+1} + \cdots + \alpha_N e_N \right\|^2 = |\alpha_{M+1}|^2 + \cdots + |\alpha_N|^2 = |T_N - T_M|,$$

where we used the orthonormality of $\{e_1, e_2, \ldots\}$.

Therefore the partial sums (S_N) form a Cauchy sequence in H if and only if the partial sums (T_N) form a Cauchy sequence in $\mathbf{R}$.

The statement now follows from the fact that both H and $\mathbf{R}$ are complete.

(b) Given $x \in H$, let $\alpha_n = \langle x, e_n \rangle$ and consider the series

$$\sum_{n=1}^{\infty} \alpha_n e_n.$$

By Bessel's Inequality, the sequence (α_n) is in ℓ^2, so by part (a) we know that the series written above converges to some element $y \in H$.

For any $k \in \mathbf{N}$ we have

$$\langle y - x, e_k \rangle = \sum_{n=1}^{\infty} \langle x, e_n \rangle \langle e_n, e_k \rangle - \langle x, e_k \rangle = \langle x, e_k \rangle - \langle x, e_k \rangle = 0.$$

Therefore

$$y - x \in \{e_1, e_2, \dots\}^{\perp} = 0$$

since $\{e_1, e_2, \dots\}$ is an orthonormal basis of H. □

Proposition 3.54 (Gram–Schmidt Orthogonalisation)**.** *Let V be an inner product space and $A = \{v_n : n \in \mathbf{N}\}$ a countable subset of V. Then there exists an orthonormal system S (finite or countable) such that*

$$\overline{\operatorname{Span}(S)} = \overline{\operatorname{Span}(A)}.$$

Proof. Without loss of generality $v_1 \neq 0$ (otherwise we can remove it from A without changing $\overline{\operatorname{Span}(A)}$).

Let

$$u_1 = v_1, \qquad e_1 = \frac{1}{\|u_1\|} u_1.$$

Proceed iteratively as follows: given $n \geqslant 2$, if $v_n \in \overline{\operatorname{Span}\{v_1, \dots, v_{n-1}\}}$ then remove v_n from A (this does not change $\overline{\operatorname{Span}(A)}$) and move on to the next element of A. Otherwise let

$$u_n = v_n - \sum_{k=1}^{n-1} \langle v_n, e_k \rangle e_k, \qquad e_n = \frac{1}{\|u_n\|} u_n.$$

At each step we have

$$\operatorname{Span}\big(\{e_1, \dots, e_n\}\big) = \operatorname{Span}\big(\{v_1, \dots, v_n\}\big).$$

So letting $S = \{e_1, e_2, \dots\}$ we have

$$\overline{\operatorname{Span}(S)} = \overline{\operatorname{Span}(A)}.$$

It is easy to see that S is an orthonormal system. □

Proposition 3.55. *If H is a separable Hilbert space, then H is linearly isometric to ℓ^2 or to $\mathbf{F}^n$ for some $n \in \mathbf{N}$.*

Proof. Let $A = \{v_1, v_2, \dots\}$ be a dense countable subset of H. Apply Gram–Schmidt to A to produce an orthonormal basis S for H. Since A is countable, S is either finite or countable.

In the finite case, write $S = \{s_1, \dots, s_n\}$ and define a function $f : H \longrightarrow \mathbf{F}^n$ by setting

$$f(s_j) = e_j \qquad \text{for } j = 1, \dots, n$$

and extending by linearity. Here $\{e_1, \dots, e_n\}$ denotes the standard basis of $\mathbf{F}^n$.

It is clear that f is a bijection, and an isometry since we are mapping an orthonormal basis to an orthonormal basis.

In the countable case, write $S = \{s_1, s_2, \dots\}$ and define a function $f \colon H \longrightarrow \ell^2$ by setting

$$f(x) = \big(\langle x, s_n\rangle\big).$$

(Note that $f(s_n) = e_n$, where $\{e_1, e_2, \dots\}$ denotes the standard Schauder basis of ℓ^2.)

The fact that $f(x) \in \ell^2$ as claimed follows from Bessel's Inequality.

Parseval's Identity

$$\|x\|^2 = \sum_{n=1}^{\infty} |\langle x, s_n\rangle|^2 = \|f(x)\|_{\ell^2}^2$$

implies that f is norm-preserving, hence also injective.

Finally, f is surjective: given $(\alpha_n) \in \ell^2$, we know from Corollary 3.53 that there is some $x \in H$ such that

$$x = \sum_{n=1}^{\infty} \alpha_n s_n$$

and

$$f(x) = \left(\left\langle \sum_{m=1}^{\infty} \alpha_m s_m, s_n \right\rangle\right) = \big(\alpha_n\big). \qquad \square$$

3.12. Function spaces: the uniform norm

We spent some time in the previous chapter studying the vector space of sequences $\mathbf{F}^{\mathbf{N}}$ and some subspaces ℓ^p $(p \geqslant 1)$ and ℓ^∞ of it that are endowed with norms with respect to which they are Banach spaces.

Of course, a sequence is just a function $\mathbf{N} \longrightarrow \mathbf{F}$. We could be bold and replace $\mathbf{N}$ with an arbitrary set X, and consider the set of functions $X \longrightarrow \mathbf{F}$. This set is an $\mathbf{F}$-vector space, but for a general X and arbitrary functions, putting a norm (let alone an inner product) on this vector space seems hopeless.

However, if we restrict our attention to bounded functions:

$$B(X, \mathbf{F}) = \big\{f \colon X \longrightarrow \mathbf{F} \colon \text{ there exists } c \text{ such that } |f(x)| \leqslant c \text{ for all } x \in X\big\},$$

we have

Proposition 3.56. *The set $B(X, \mathbf{F})$ is a Banach space with respect to the* uniform norm *given by*

$$\|f\|_{L^\infty} = \sup_{x \in X} |f(x)|.$$

If X is a metric space, then the subset $C_0(X, \mathbf{F})$ of bounded continuous functions $X \longrightarrow \mathbf{F}$ is a Banach subspace of $B(X, \mathbf{F})$.

Proof. That $B(X, \mathbf{F})$ is a vector subspace of the $\mathbf{F}$-vector space of all functions $X \longrightarrow \mathbf{F}$ is straightforward. It is similarly clear that $\|\cdot\|_{L^\infty}$ gives a norm on $B(X, \mathbf{F})$, and that this norm is associated to the uniform distance d_∞ on $B(X, \mathbf{F})$ considered in Section 2.12.

It then follows from Proposition 2.78 that $B(X, \mathbf{F})$ is complete, hence a Banach space. Similarly, the statement about $C_0(X, \mathbf{F})$ follows from Proposition 2.79. $\square$

There are **a lot** of bounded continuous functions $X \longrightarrow \mathbf{F}$ even for relatively simple X, e.g. closed intervals in $\mathbf{R}$. But:

Theorem 3.57 (Weierstrass Approximation Theorem)**.** *Given $a < b$, let $\mathcal{A}$ be the subset of $C_0([a, b], \mathbf{R})$ consisting of polynomial functions. Then $\mathcal{A}$ is dense in $C_0([a, b], \mathbf{R})$.*

We will obtain this as a corollary of a more general result. We need some preliminaries.

Lemma 3.58. *There is a sequence* (p_n) *in* $x\mathbf{R}[x]$ *such that* $(p_n) \longrightarrow |x|$ *uniformly on* $[-1,1]$.

Sketch of proof. Let $p_1(x) = 0$ and

$$p_{n+1}(x) = p_n(x) - \frac{p_n(x)^2 - x^2}{2} = p_n(x) - \frac{\big(p_n(x) - |x|\big)\big(p_n(x) + |x|\big)}{2} \qquad \text{for } n \geqslant 1.$$

One can use induction to prove that, for all $x \in [-1,1]$ and all $n \geqslant 1$:

(a) $0 \leqslant p_n(x) \leqslant |x|$;

(b) $p_n(x) \leqslant p_{n+1}(x)$;

(c) $|x| - p_{n+1}(x) \leqslant |x| \left(1 - \frac{|x|}{2}\right)^n$.

(See Exercise 3.35.)

A little calculus (Exercise 3.36) tells us that for any $n \geqslant 1$

$$|x| \left(1 - \frac{|x|}{2}\right)^n < \frac{2}{n+1} \qquad \text{for all } x \in [-1,1],$$

which then implies that $(p_n) \longrightarrow |x|$ uniformly on $[-1,1]$. □

Corollary 3.59. *For any* $a > 0$, *there is a sequence* (q_n) *in* $x\mathbf{R}[x]$ *such that* $(q_n) \longrightarrow |x|$ *uniformly on* $[-a,a]$.

Proof. See Exercise 3.37. □

For any metric space X, the set $C_0(X,\mathbf{R})$ is an $\mathbf{R}$-algebra. (The multiplication of functions is done pointwise, just like the addition.) Let $\mathcal{A}$ be a subalgebra of $C_0(X,\mathbf{R})$. We say that $\mathcal{A}$ *interpolates pairs of points on* X if for any $(x_1,y_1),(x_2,y_2) \in X \times \mathbf{R}$ with $x_1 \neq x_2$, there exists $h \in \mathcal{A}$ such that

$$h(x_1) = y_1 \qquad \text{and} \qquad h(x_2) = y_2.$$

Proposition 3.60. *Let* X *be a metric space and let* $\mathcal{C}$ *be a closed subalgebra of* $C_0(X,\mathbf{R})$. *Then*

(a) if $g \in \mathcal{C}$ *then* $|g| \in \mathcal{C}$;

(b) if $g_1, g_2 \in \mathcal{C}$ *then* $\max\{g_1,g_2\}, \min\{g_1,g_2\} \in \mathcal{C}$.

Proof.

(a) Let $a > 0$ be an upper bound for $|g|$. For any $n \in \mathbf{N}$, we have by Corollary 3.59 a polynomial $q_n(y) \in y\mathbf{R}[y]$ such that

$$\big|q_n(y) - |y|\big| < \frac{1}{n} \qquad \text{for all } y \in [-a,a].$$

Let $h_n = q_n(g)$, then $h_n \in \mathcal{C}$ since the latter is an algebra; therefore applying the above inequality with $y = g(x)$ we have

$$\big|h_n(x) - |g(x)|\big| = \big|q_n\big(g(x)\big) - |g(x)|\big| < \frac{1}{n} \qquad \text{for all } x \in X,$$

in other words

$$\big\|h_n - |g|\big\|_{L^\infty} < \frac{1}{n},$$

which shows that $(h_n) \longrightarrow |g|$ in $\mathcal{C}$, which is closed, so $|g| \in \mathcal{C}$.

(b) The claim follows directly from the relations

$$2\max\{g_1, g_2\} = g_1 + g_2 + |g_1 - g_2|$$
$$2\min\{g_1, g_2\} = g_1 + g_2 - |g_1 - g_2|. \qquad \square$$

Theorem 3.61 (Stone–Weierstrass)**.** *Let X be a compact metric space and let $\mathcal{A}$ be a nonzero subalgebra of $C_0(X, \mathbf{R})$. If $\mathcal{A}$ interpolates pairs of points on X, then it is dense in $C_0(X, \mathbf{R})$ (with respect to the uniform norm).*

Proof. The corner case where X is a singleton is easily dispatched: then $C_0(X, \mathbf{R}) = \mathbf{R}$ and $\mathcal{A} = \mathbf{R}$ (since nonzero). So we may assume that X has at least two distinct elements, so that the interpolation property is non-vacuous.

Let $\mathcal{C}$ denote the closure of $\mathcal{A}$ in $C_0(X, \mathbf{R})$. We will show that $\mathcal{C}$ is dense in $C_0(X, \mathbf{R})$: let $f \in C_0(X, \mathbf{R})$ and let $\varepsilon > 0$.

Fix $x' \in X$.

For every $x \in X$:

- If $x \neq x'$ then by the interpolation property of $\mathcal{A}$, there exists $h_x \in \mathcal{A}$ that interpolates $(x, f(x))$ and $(x', f(x'))$, that is
$$h_x(x) = f(x) \qquad \text{and} \qquad h_x(x') = f(x').$$
- If $x = x'$, let $t \in X$, $t \neq x'$ and choose $h_{x'} \in \mathcal{A}$ such that
$$h_t(t) = f(t) \qquad \text{and} \qquad h_{x'}(x') = f(x').$$

Let U_x be an open neighbourhood of x such that

$$h_x(x'') > f(x'') - \varepsilon \qquad \text{for all } x'' \in U_x.$$

The sets $\{U_x : x \in X\}$ form an open cover of the compact space X, so there exist $x_1, \dots, x_m$ such that

$$X \subseteq U_{x_1} \cup \cdots \cup U_{x_m}.$$

Let $g_{x'} = \max\{h_{x_1}, \dots, h_{x_m}\}$, an element of $\mathcal{C}$ by Proposition 3.60.

Then $g_{x'}(x') = f(x')$ and

$$g_{x'}(x'') > f(x'') - \varepsilon \qquad \text{for all } x'' \in X.$$

We have such a function $g_{x'} \in \mathcal{C}$ for each $x' \in X$.

Let $V_{x'}$ be an open neighbourhood of x' such that

$$g_{x'}(x'') < f(x'') + \varepsilon \qquad \text{for all } x'' \in V_{x'}.$$

The sets $\{V_{x'} : x' \in X\}$ form an open cover of the compact space X, so there exist $x'_1, \dots, x'_n$ such that

$$X \subseteq V_{x'_1} \cup \cdots \cup V_{x'_n}.$$

We let $g = \min\{g_{x'_1}, \dots, g_{x'_n}\}$ so that $g \in \mathcal{C}$ and

$$f(x'') - \varepsilon < g(x'') < f(x'') + \varepsilon \qquad \text{for all } x'' \in X,$$

so we conclude that

$$\|f - g\|_{L^\infty} < \varepsilon. \qquad \square$$

We can now specialise to

Proof of Theorem 3.57. We take $\mathcal{A}$ to be the subset of $C_0([a, b], \mathbf{R})$ consisting of polynomial functions. It is clear that $\mathcal{A}$ is an algebra. Also, $\mathcal{A}$ interpolates pairs of points on X since its subset consisting of linear polynomials already has this property. It follows from the Stone–Weierstrass Theorem that $\mathcal{A}$ is dense. $\square$

3.13. (*) Function spaces: the L^p-norms

We can consider other norms on $C_0(X, \mathbf{F})$ for suitable X. To keep things simple, we restrict to $X = [a, b]$ for real numbers $a < b$.

For $p \geqslant 1$ and $f \in C_0([a, b], \mathbf{F})$, let

$$\|f\|_{L^p} = \left(\int_a^b |f(x)|^p \, dx \right)^{1/p} \in \mathbf{R}_{\geqslant 0}.$$

The proof that this is a norm is similar to the one for ℓ^p, with the appropriate version of Hölder's Inequality substituted in.

Example 3.62. Let $f\colon [-\pi, \pi] \longrightarrow \mathbf{R}$ be given by $f(x) = \sin(x)$. Then

$$\|f\|_{L^\infty} = 1, \quad \|f\|_{L^1} = 4, \quad \|f\|_{L^2}^2 = \pi, \quad \|f\|_{L^3}^3 = \frac{8}{3}, \ldots$$

Just for fun: show that for all $n \in \mathbf{N}$

$$\|f\|_{L^{2n-1}}^{2n-1} = 2^{2n} \frac{\big((n-1)!\big)^2}{(2n-1)!}$$
$$\|f\|_{L^{2n}}^{2n} = \frac{(2n)!}{2^{2n-1}(n!)^2} \, \pi.$$

One issue is that, in contrast with Proposition 3.56, the space of continuous functions is not complete with respect to the L^p-norms.

Example 3.63. Consider $V = C_0([-1, 1], \mathbf{R})$ endowed with the L^1-norm and define for all $n \in \mathbf{N}$:

$$f_n(x) = \begin{cases} 0 & \text{if } -1 \leqslant x < 0 \\ nx & \text{if } 0 \leqslant x \leqslant \frac{1}{n} \\ 1 & \text{if } \frac{1}{n} < x \leqslant 1. \end{cases}$$

It is clear that $f_n \in V$ for all n. Moreover (f_n) is a Cauchy sequence in V with respect to the L^1-norm: given $\varepsilon > 0$ take $N \in \mathbf{N}$ such that $\frac{1}{N} < \varepsilon$, then for all $n, m \geqslant N$ we have

$$\|f_n - f_m\|_{L^1} = \int_{-1}^1 |f_n(x) - f_m(x)| \, dx = \frac{1}{2} \left| \frac{1}{n} - \frac{1}{m} \right| \leqslant \frac{1}{2} \left(\frac{1}{n} + \frac{1}{m} \right) \leqslant \frac{1}{N} < \varepsilon.$$

Suppose V is complete, so $(f_n) \longrightarrow f$ in the L^1-norm with f continuous, then

$$\int_{-1}^0 |f(x)| \, dx + \int_0^{1/n} |f_n(x) - f(x)| \, dx + \int_{1/n}^1 |1 - f(x)| \, dx$$
$$= \int_{-1}^1 |f_n(x) - f(x)| \, dx \longrightarrow 0 \quad \text{as } n \longrightarrow \infty,$$

so that each of the three nonnegative summands must converge to 0 as $n \longrightarrow \infty$. This implies that (given the fact that f is continuous):

$$\int_{-1}^0 |f(x)| \, dx = 0 \quad \Rightarrow \quad f(x) = 0 \text{ for } -1 \leqslant x < 0$$

and
$$\int_0^1 |1 - f(x)|\, dx = 0 \quad \Rightarrow \quad f(x) = 1 \text{ for } 0 < x \leqslant 1.$$
Here is the contradiction, since f is manifestly not continuous at 0.
So $\big(C_0([-1,1],\mathbf{R}), \|\cdot\|_{L^1}\big)$ is not complete.

This state of affairs leaves us no choice but to take the completion of the normed space $\big(C_0([a,b],\mathbf{F}), \|\cdot\|_{L^p}\big)$, which results in a Banach space denoted $L^p([a,b],\mathbf{F})$. An element of $L^p([a,b],\mathbf{F})$ is therefore an equivalence class of Cauchy sequences of continuous functions $[a,b] \longrightarrow \mathbf{F}$ with respect to the L^p-norm.

There is another approach to defining the spaces L^p, via Lebesgue integration (see the next section for a very minimal introduction to this). This can be used to prove a number of results that are reminiscent of the sequence spaces ℓ^p:

- as sets, we have inclusions
$$C_0([a,b],\mathbf{F}) \subseteq \cdots \subseteq L^p([a,b],\mathbf{F}) \subseteq \cdots \subseteq L^1([a,b],\mathbf{F});$$
- all are Banach spaces, that is complete normed spaces;
- polynomials are dense in $L^p([a,b],\mathbf{F})$ for all $p \geqslant 1$;
- $L^p([a,b],\mathbf{F})$ for $p \geqslant 1$ and $C_0([a,b],\mathbf{F})$ have Schauder bases, hence are separable;
- $\big(L^p([a,b],\mathbf{F})\big)^\vee \cong L^q([a,b],\mathbf{F})$ if $\frac{1}{p} + \frac{1}{q} = 1$ and $p > 1$;
- only $L^2([a,b],\mathbf{F})$ is a Hilbert space, with inner product
$$\langle f, g\rangle = \int_a^b f(x)\,\overline{g(x)}\, dx.$$

Let $\mathbf{C}^{\mathbf{Z}} = \{\mathbf{Z} \longrightarrow \mathbf{C}\}$ be the vector space of doubly-infinite sequences
$$(a_n) = \big(\ldots, a_{-2}, a_{-1}, a_0, a_1, a_2, \ldots\big).$$
The definition of the subspace $\ell^\infty(\mathbf{Z})$ is clear, whereas for $1 \leqslant p$ we have
$$\ell^p(\mathbf{Z}) = \left\{(a_n) \in \mathbf{C}^{\mathbf{Z}} : \sum_{n=-\infty}^{\infty} |a_n|^p < \infty\right\},$$
where we say that
$$\sum_{n=-\infty}^{\infty} b_n \quad \text{converges to } \beta$$
if
$$\sum_{n=-\infty}^{-1} b_n \quad \text{converges to } \beta_-$$
$$\sum_{n=0}^{\infty} b_n \quad \text{converges to } \beta_+$$
$$\beta = \beta_- + \beta_+.$$
There is a linear map $\mathcal{F}\colon C_0([0,1],\mathbf{C}) \longrightarrow \mathbf{C}^{\mathbf{Z}}$ given by
$$\mathcal{F}(f) = \widehat{f} = \big(\widehat{f}_n\big), \quad \text{where} \quad \widehat{f}_n = \int_0^1 e^{-2\pi i n x} f(x)\, dx \quad \text{for } n \in \mathbf{Z}.$$

The complex numbers $\widehat{f}_n$ are called the *Fourier coefficients* of f.

Note that

$$\|\mathcal{F}(f)\|_{\ell^\infty(\mathbf{Z})} = \|\widehat{f}\|_{\ell^\infty(\mathbf{Z})} = \sup_{n\in\mathbf{Z}} |\widehat{f}_n| = \sup_{n\in\mathbf{Z}} \left| \int_0^1 e^{-2\pi inx} f(x)\,dx \right| \leqslant \int_0^1 |f(x)|\,dx = \|f\|_{L^1},$$

so that $\widehat{f} \in \ell^\infty(\mathbf{Z})$ and we can view $\mathcal{F}$ as a linear function $C_0([0,1],\mathbf{C}) \longrightarrow \ell^\infty(\mathbf{Z})$. As such, the inequality $\|\mathcal{F}(f)\| \leqslant \|f\|$ we checked above shows that $\mathcal{F}$ is Lipschitz, hence (uniformly) continuous. With some more work, one can show that $\operatorname{im}(\mathcal{F}) \subseteq c_0(\mathbf{Z})$ (this result is called the Riemann–Lebesgue Lemma), where

$$c_0(\mathbf{Z}) = \left\{ (a_n) \in \mathbf{C}^{\mathbf{Z}} \colon (a_n) \longrightarrow 0 \text{ as } n \longrightarrow \infty \text{ and } (a_n) \longrightarrow 0 \text{ as } n \longrightarrow -\infty \right\}.$$

The (uniformly) continuous linear map $\mathcal{F}$ extends uniquely to a (uniformly) continuous linear map between completions

$$\mathcal{F} \colon L^1([0,1],\mathbf{C}) \longrightarrow c_0(\mathbf{Z}).$$

One can show that $\mathcal{F}$ is injective and

$$\operatorname{im}\left(\mathcal{F}|_{L^p([0,1],\mathbf{C})}\right) \subseteq \ell^q(\mathbf{Z}) \qquad \text{if } 1 \leqslant p \leqslant 2 \text{ and } \frac{1}{p} + \frac{1}{q} = 1.$$

In particular, when $p = 2$ we have

$$\mathcal{F} \colon L^2([0,1],\mathbf{C}) \longrightarrow \ell^2(\mathbf{Z}).$$

The set $\left\{ e^{2\pi inx} \colon n \in \mathbf{Z} \right\}$ is an orthonormal basis for $L^2([0,1],\mathbf{C})$; given $f \in L^2([0,1],\mathbf{C})$, the resulting unique expression

$$f(x) = \sum_{n=-\infty}^{\infty} \widehat{f}_n e^{2\pi inx}$$

is called the *Fourier expansion* of f. Note that the equality is misleading: it means convergence with respect to the L^2 norm; it is true that there is pointwise convergence a.e. but this is a hard result (proved by Carleson in 1966).

For a different example of an orthonormal basis in a separable Hilbert space, consider $H = L^2([-1,1],\mathbf{R})$. We saw above that polynomials are dense in this Hilbert space, so certainly $1, x, x^2, \ldots$ is a countable set whose span is dense in H. But it is not an orthonormal basis:

$$\langle 1, x^2 \rangle = \frac{2}{3} \neq 0.$$

However, we can apply Gram–Schmidt to $\{1, x, x^2, \ldots\}$ with respect to the L^2 norm and get

$$\frac{1}{\sqrt{2}}, \sqrt{\frac{3}{2}}\,x, \frac{3}{2}\sqrt{\frac{5}{2}}\left(x^2 - \frac{1}{3}\right), \ldots$$

The elements of this orthonormal basis are called *normalised Legendre polynomials*.

3.14. (*) The Lebesgue integral from scratch

This will be as quick and unsatisfying as a movie trailer. We follow parts of [1, Chapters 1 and 2], to which we refer the reader for more details.

All the functions we consider in this section are $\mathbf{R} \longrightarrow \mathbf{R}$. Let $\mathcal{I}$ denote the set of all closed bounded intervals in $\mathbf{R}$:

$$\mathcal{I} = \{ I = [a,b] \subseteq \mathbf{R} \}.$$

Given a subset $S \subseteq \mathbf{R}$, the *characteristic function* of S is $\chi_S \colon \mathbf{R} \longrightarrow \mathbf{R}$ given by

$$\chi_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{otherwise.} \end{cases}$$

We define the *integral* of the characteristic function $\chi_{[a,b]}$ to be

$$\int_{\mathbf{R}} \chi_{[a,b]} = b - a.$$

Let Λ denote the set of *step functions*:

$$\Lambda = \operatorname{Span}\left\{\chi_I \colon I \in \mathcal{I}\right\}.$$

Then Λ is a vector space, and it is closed under taking min, max, and absolute values. We define a linear function $\Lambda \longrightarrow \mathbf{R}$ by setting

$$\int_{\mathbf{R}} \varphi = \int_{\mathbf{R}} \left(a_1\chi_{I_1} + \cdots + a_n\chi_{I_n}\right) \coloneqq a_1 \int_{\mathbf{R}} \chi_{I_1} + \cdots + a_n \int_{\mathbf{R}} \chi_{I_n}.$$

A subset $E \subseteq \mathbf{R}$ is said to be *of measure zero (or a null set, or negligeable)* if for any $\varepsilon > 0$, there exists a cover of E by open intervals

$$E \subseteq \bigcup_{n=1}^{\infty} U_n \qquad \text{such that} \qquad \sum_{n=1}^{\infty} \operatorname{diam}(U_n) \leqslant \varepsilon.$$

We say that a certain property related to elements $x \in \mathbf{R}$ holds *almost everywhere (a.e.)* if there exists a set E of measure zero such that the property holds for all $x \in \mathbf{R} \smallsetminus E$.

A function $f \colon \mathbf{R} \longrightarrow \mathbf{R}$ is *Lebesgue integrable* if there exist step functions $\{\varphi_n \colon n \in \mathbf{N}\}$ such that

(a) $\displaystyle\sum_{n=1}^{\infty} \varphi_n(x) = f(x)$ a.e.;

(b) $\displaystyle\sum_{n=1}^{\infty} \int_{\mathbf{R}} |\varphi_n| < \infty$.

In this case, we define the *Lebesgue integral of f* to be

$$\int_{\mathbf{R}} f \coloneqq \sum_{n=1}^{\infty} \int_{\mathbf{R}} \varphi_n.$$

If $f \colon \mathbf{R} \longrightarrow \mathbf{R}$ is a function that is continuous on a closed bounded interval $[a, b]$ and is zero on $\mathbf{R} \smallsetminus [a, b]$, then f is Lebesgue integrable and its Lebesgue integral is equal to its Riemann integral:

$$\int_{\mathbf{R}} f = \int_a^b f(x)\, dx.$$

For $1 \leqslant p$ we define

$$\mathcal{L}^p([a,b], \mathbf{R}) = \left\{f \colon [a,b] \longrightarrow \mathbf{R} \colon \int_a^b |f(x)|^p < \infty\right\}.$$

This is not a normed space because $\|f\|_{L^p} = 0$ for any function that is zero almost everywhere on $[a, b]$. We can define an equivalence relation on $\mathcal{L}^p([a,b], \mathbf{R})$ by setting $f \sim g$ if $f - g$ is zero a.e. on $[a, b]$, and we let

$$L^p([a,b], \mathbf{R}) = \mathcal{L}^p([a,b], \mathbf{R}) / \sim$$

be the set of equivalence classes.

There is a variant of this where $\mathbf{R}$ gets replaced by $\mathbf{C}$.

3.15. (*) Some spectral theory

In this section we let H be a Hilbert space over $\mathbf{C}$.

The following result is proved in Group Theory and Linear Algebra:

Theorem 3.64. *Let $f\colon V \longrightarrow V$ be a self-adjoint linear map on a finite-dimensional complex inner product space V. There exists an orthonormal basis of V made of eigenvectors for f.*

This has some generalisations to the infinite-dimensional setting, for instance:

Theorem 3.65 (Spectral Theorem). *Let $f\colon H \longrightarrow H$ be a self-adjoint compact linear map on a separable complex Hilbert space H. Then there exists an orthonormal basis of H of the form*

$$\{u_n : 1 \leqslant n \leqslant \operatorname{rank}(f)\} \sqcup \{z_m : 1 \leqslant m \leqslant \operatorname{nullity}(f)\},$$

where each u_n is an eigenvector of f with nonzero eigenvalue and each z_m is an eigenvector of f with eigenvalue zero, and

$$0 \leqslant \operatorname{rank}(f) := \dim \operatorname{im}(f) \leqslant \infty, \quad 0 \leqslant \operatorname{nullity}(f) := \dim \ker(f) \leqslant \infty.$$

Moreover, if we order the (finite or countable) set of nonzero eigenvalues of f in such a way that $|\lambda_n|$ is non-increasing, then $(\lambda_n) \longrightarrow 0$ as $n \longrightarrow \infty$.

But what is a compact linear map? A natural starting point is to consider maps $f \in L(H)$ that have finite-dimensional image in H. We say that such f is a *finite rank map*.

Example 3.66. Fix $m \in \mathbf{N}$ and consider $f_m\colon \ell^2 \longrightarrow \ell^2$ given by

$$f_m\big((a_n)\big) = \left(\frac{a_1}{1}, \frac{a_2}{2}, \ldots, \frac{a_m}{m}, 0, 0, \ldots\right).$$

Then $\operatorname{rank}(f_m) = m$, as $\operatorname{im}(f_m) = \operatorname{Span}\{e_1, \ldots, e_m\}$.

We let $R(H)$ denote the set of all finite rank maps. It has some interesting properties:

- $R(H)$ is a subspace of $L(H)$, see Exercise 3.48;
- if $f \in R(H)$ and $g_1, g_2 \in L(H)$ then $g_2 \circ f \circ g_1 \in R(H)$, see Exercise 3.49;
- if $f \in R(H)$ then $f^* \in R(H)$, see Exercise 3.50.

However, in general $R(H)$ is not a **closed** subspace of $L(H)$:

Example 3.67. Continuing with the setup of Example 3.66, note that the sequence of finite rank maps (f_m) converges to $f\colon \ell^2 \longrightarrow \ell^2$ given by

$$f\big((a_n)\big) = \left(\frac{a_1}{1}, \frac{a_2}{2}, \ldots\right).$$

But f certainly does not have finite rank, so $R(\ell^2)$ is not closed.

We let $K(H) = \overline{R(H)}$, a closed subspace of $L(H)$. Elements of $K(H)$ are called *compact maps*.

Proposition 3.68. *A map $f \in L(H)$ is compact if and only if $\overline{f\big(\mathbf{D}_1(0)\big)}$ is compact.*

Example 3.69. The identity map id_{ℓ^2} is not compact, since

$$\overline{\mathrm{id}_{\ell^2}\big(\mathbf{D}_1(0)\big)} = \overline{\mathbf{D}_1(0)} = \mathbf{D}_1(0),$$

which contains the standard vector e_n for all $n \in \mathbf{N}$, thus giving a sequence (e_n) that does not have any convergent subsequences (because the distance between e_n and e_m is $\delta_{nm}\sqrt{2}$).

Given $f \in L(H)$, we define the *spectrum* of f to be the set

$$\sigma(f) = \big\{\lambda \in \mathbf{C} \colon\ f - \lambda\,\mathrm{id}_H \in L(H) \text{ is not invertible}\big\}.$$

The complement of the spectrum is called the *resolvent set* of f:

$$\rho(f) = \big\{\lambda \in \mathbf{C} \colon\ f - \lambda\,\mathrm{id}_H \in L(H) \text{ is invertible}\big\}.$$

Some things are similar to what we know from the finite-dimensional case:

Proposition 3.70. *If λ is an eigenvalue of f then $\lambda \in \sigma(f)$.*

Proof. There exists a nonzero element $x \in H$ such that

$$\begin{aligned} f(x) = \lambda x \quad &\iff \quad \big(f - \lambda\,\mathrm{id}_H\big)(x) = 0 \\ &\iff \quad \ker\big(f - \lambda\,\mathrm{id}_H\big) \neq 0 \\ &\iff \quad (f - \lambda\,\mathrm{id}_H) \colon H \longrightarrow H \text{ is not injective} \\ &\implies \quad (f - \lambda\,\mathrm{id}_H) \colon H \longrightarrow H \text{ is not invertible.} \end{aligned}$$ □

If H is finite-dimensional, then the last arrow in the proof is also an equivalence (by the rank-nullity theorem), so $\sigma(f)$ is precisely the set of eigenvalues of f.

Other things are very different in the infinite-dimensional case, for instance there are operators on H that have no complex eigenvalues, like the right shift operator on ℓ^2, see Exercise 3.51.

There is a nice relation between the spectra of adjoint maps:

Proposition 3.71. *If $f \in L(H)$ for a complex Hilbert space H, then*

$$\sigma(f^*) = \big\{\overline{\lambda} \colon\ \lambda \in \sigma(f)\big\}.$$

Proof. Recall from Tutorial Question 11.4 that $g \in L(H)$ is invertible if and only if g^* is invertible.

So $\lambda \in \sigma(f)$ iff $\big(\lambda\,\mathrm{id}_H - f\big)$ is not invertible iff $\big(\lambda\,\mathrm{id}_H - f\big)^*$ is not invertible iff $\big(\overline{\lambda}\,\mathrm{id}_H - f^*\big)$ is not invertible iff $\overline{\lambda} \in \sigma(f^*)$. □

The following result is a useful generalisation of the geometric series identity:

$$\big(1 - x\big)\big(1 + x + x^2 + \dots\big) = 1 \qquad \text{if } |x| < 1.$$

Proposition 3.72. *If $f \in L(H)$ satisfies $\|f\| < 1$ then $\mathrm{id}_H - f$ is invertible.*

Proof. Consider the series in $L(H)$:

$$\sum_{n=0}^{\infty} f^n.$$

We have $\|f^n\| \leqslant \|f\|^n$ for all $n \in \mathbf{N}$, and the series of real numbers $\sum_{n=1}^{\infty} \|f\|^n$ converges since $\|f\| < 1$, so the series $\sum_{n=0}^{\infty} f^n$ is absolutely convergent in $L(H)$, which is a Banach space, hence converges in $L(H)$ to some element g. We have

$$f \circ g = f \circ \sum_{n=0}^{\infty} f^n = \sum_{n=1}^{\infty} f^n = g - \mathrm{id}_H,$$

so that $g \circ (\mathrm{id}_H - f) = \mathrm{id}_H$. A similar calculation gives $(\mathrm{id}_H - f) \circ g = \mathrm{id}_H$, so $\mathrm{id}_H - f$ is invertible with inverse g. □

Corollary 3.73. *For any $f \in L(H)$ we have $\sigma(f) \subseteq \mathbf{D}_{\|f\|}(0)$.*

Proof. Suppose $\lambda \in \mathbf{C}$ satisfies $\lambda \notin \mathbf{D}_{\|f\|}(0)$, so $|\lambda| > \|f\|$. Then $\left\|\lambda^{-1} f\right\| < 1$, so $\mathrm{id}_H - \lambda^{-1} f$ is invertible; let g be its inverse, then

$$\big(f - \lambda \, \mathrm{id}_H\big)\big(-\lambda^{-1} g\big) = (\mathrm{id}_H - \lambda^{-1} f) g = \mathrm{id}_H,$$

and similarly for the composition in the opposite order, therefore $f - \lambda \, \mathrm{id}_H$ is invertible so $\lambda \notin \sigma(f)$. □

In fact (see Exercise 3.53), $\sigma(f)$ is a compact set.

Under the additional assumption that f is self-adjoint, we can say more:

Proposition 3.74. *If $f \in L(H)$ is a self-adjoint map on a complex Hilbert space H then $\sigma(f) \subseteq \mathbf{R}$, so that*

$$\sigma(f) \subseteq \big[-\|f\|, \|f\|\big].$$

Proof. By Exercise 3.54, for any given $\lambda = a + ib \in \mathbf{C}$ we have

$$\big\|\big(f - (a+ib)\,\mathrm{id}_H\big)(x)\big\| \geqslant |b| \, \|x\| \qquad \text{for all } x \in H. \tag{3.1}$$

We show that if $b \neq 0$ then $f - \lambda \, \mathrm{id}_H = f - (a+ib)\,\mathrm{id}_H$ is invertible.

First of all, Equation (3.1) implies that $f - \lambda \, \mathrm{id}_H$ is injective.

Second, it also implies that $\mathrm{im}\big(f - \lambda \, \mathrm{id}_H\big)$ is closed in H, see Exercise 3.39.

Finally, we can apply Equation (3.1) with $\big(f - \lambda \, \mathrm{id}_H\big)^* = f - (a - ib)\,\mathrm{id}_H$ and see that this map is also injective, in other words by Exercise 3.31

$$\mathrm{im}\big(f - \lambda \, \mathrm{id}_H\big)^{\perp} = \ker\big((f - \lambda \, \mathrm{id}_H)^*\big) = 0.$$

So $\mathrm{im}\big(f - \lambda \, \mathrm{id}_H\big)$ is dense in H; since it is also closed in H, it must equal H, so $f - \lambda \, \mathrm{id}_H$ is invertible. □

More is true in fact: at least one of the interval endpoints $\pm\|f\|$ is an element of $\sigma(f)$.

Example 3.75. Consider the compact map $f \colon \ell^2 \longrightarrow \ell^2$ from Example 3.67:

$$f(a_1, a_2, \dots) = \left(\frac{a_1}{1}, \frac{a_2}{2}, \dots\right).$$

Then for each $n \in \mathbf{N}$, e_n is an eigenvector of f with eigenvalue $1/n$, therefore $1/n \in \sigma(f)$. Since $\sigma(f)$ is closed, it must also contain the limit point 0 of these eigenvalues, that is

$$S := \left\{\frac{1}{n} : n \in \mathbf{N}\right\} \cup \{0\} \subseteq \sigma(f).$$

In fact, we will see now that $S = \sigma(f)$. Suppose λ is nonzero and $\lambda \neq 1/n$ for any $n \in \mathbf{N}$.

Then there exists $c > 0$ such that

$$\left|\lambda - \frac{1}{n}\right| > c \qquad \text{for all } n \in \mathbf{N}.$$

Then for any $x = (x_n) \in \ell^2$ we have

$$\left\|\left(\lambda \operatorname{id}_{\ell^2} - f\right)(x)\right\|^2 = \sum_{n=1}^{\infty} \left|\lambda - \frac{1}{n}\right|^2 |x_n|^2 > c^2 \sum_{n=1}^{\infty} |x_n|^2 = c^2 \|x\|^2.$$

So

$$\left\|\left(\lambda \operatorname{id}_{\ell^2} - f\right)(x)\right\| > c\,\|x\| \qquad \text{for all } x \in \ell^2.$$

This tells us several things:

- $\lambda \operatorname{id}_{\ell^2} - f$ is injective;
- by Exercise 3.31:

  $$\overline{\operatorname{im}\left(\lambda \operatorname{id}_{\ell^2} - f\right)^*} = \left(\ker\left(\lambda \operatorname{id}_{\ell^2} - f\right)\right)^{\perp} = H,$$

  so $\operatorname{im}\left(\lambda \operatorname{id}_{\ell^2} - f\right) = \operatorname{im}\left(\lambda \operatorname{id}_{\ell^2} - f\right)^*$ is dense in ℓ^2;
- by Exercise 3.39, $\operatorname{im}\left(\lambda \operatorname{id}_{\ell^2} - f\right)$ is closed in ℓ^2, so with the previous point we conclude that $\lambda \operatorname{id}_{\ell^2} - f$ is surjective.

Hence $\lambda \notin \sigma(f)$.

Example 3.76. Fix a bijection $\varphi\colon \mathbf{N} \longrightarrow [0,1] \cap \mathbf{Q}$. Define $g\colon \ell^2 \longrightarrow \ell^2$ by

$$g(a_1, a_2, \dots) = \left(\varphi(1)a_1, \varphi(2)a_2, \dots\right).$$

For each $n \in \mathbf{N}$, e_n is an eigenvector of g with eigenvalue $\varphi(n)$, therefore

$$[0,1] = \overline{\left\{\varphi(n)\colon n \in \mathbf{N}\right\}} \subseteq \sigma(g).$$

This is actually an equality, which can be proved in a manner similar to Example 3.75.

Example 3.77. If $L, R\colon \ell^2 \longrightarrow \ell^2$ denote the left shift, respectively right shift maps, then

$$\sigma(L) = \sigma(R) = \mathbf{D}_1(0).$$

Solution. First note that L and R are adjoint maps:

$$\langle R(x), y\rangle = \sum_{n=1}^{\infty} x_n \overline{y}_{n+1} = \langle x, L(y)\rangle.$$

It suffices to prove that any $\lambda \in \mathbf{C}$ with $|\lambda| < 1$ is an eigenvalue of L, which we do below. If that is the case then certainly

$$\mathbf{D}_1(0) = \overline{\mathbf{B}_1(0)} \subseteq \sigma(L).$$

As $\|L\| = 1$, Corollary 3.73 tells us that $\sigma(L) = \mathbf{D}_1(0)$. From this we conclude by Proposition 3.71 that $\sigma(R) = \mathbf{D}_1(0)$.

It remains to prove the claim about eigenvalues of L. As $L(e_1) = 0$, we see that $\lambda = 0$ is an eigenvalue of L.

Now given $0 < |\lambda| < 1$, let $x_n = \lambda^{n-1}$ for all $n \in \mathbf{N}$. Then $x = (x_n) \neq 0$ and

$$\|x\|_{\ell^2}^2 = \sum_{n=1}^{\infty} |x_n|^2 = \sum_{n=0}^{\infty} |\lambda|^{2n} = \frac{1}{1 - \|\lambda\|^2},$$

so $x \in \ell^2$. Finally, $L\big((x_n)\big) = \lambda(x_n)$ so λ is an eigenvalue of L. □

A. Appendix

At the moment, this is just a disorganised pile of miscellanea.

A.1. Set theory

Theorem A.1 (Schröder–Bernstein). *If A and B are sets and $f\colon A \longrightarrow B$ and $g\colon B \longrightarrow A$ are injective functions, then A and B have the same cardinality (that is, there exists some bijective function $h\colon A \longrightarrow B$).*

Proof. If $g(B) = A$ then g is bijective so we can take $h = g^{-1}$.

Otherwise, let $X_1 = A \setminus g(B)$. Define $X_2 = g\big(f(X_1)\big)$, and more generally

$$X_n = g\big(f(X_{n-1})\big), \qquad \text{for } n \geqslant 2.$$

Let

$$X = \bigcup_{n \in \mathbf{N}} X_n.$$

This is a subset of A with the property that

$$g\big(f(X)\big) = \bigcup_{n \in \mathbf{N}} g\big(f(X_n)\big) = \bigcup_{n \in \mathbf{N}} X_{n+1}. \tag{A.1}$$

If $a \in A \setminus X$, then $a \notin X_1 = A \setminus g(B)$, therefore $a \in g(B)$. As g is injective, there is a unique $b \in B$ such that $a = g(b)$, in other words, $g^{-1}(a) = \{b\}$.

This means that

$$h(a) = \begin{cases} f(a) & \text{if } a \in X \\ g^{-1}(a) & \text{if } a \in A \setminus X \end{cases}$$

gives a well-defined function $h\colon A \longrightarrow B$.

Let's check that h is surjective. If $b \in f(X)$, then $b = f(a) = h(a)$ for some $a \in X$ and we are done. If $b \notin f(X)$, then as g is injective, $g(b) \notin g\big(f(X)\big)$. By Equation (A.1), we have

$$g(b) \notin \bigcup_{n \in \mathbf{N}} X_{n+1}.$$

We also have $g(b) \in g(B)$ so $g(b) \notin X_1 = A \setminus g(B)$. Therefore

$$g(b) \notin X = X_1 \cup \bigcup_{n \in \mathbf{N}} X_{n+1},$$

so setting $a = g(b)$ we have

$$h(a) = h(g(b)) = g^{-1}\big(g(b)\big) = b.$$

Finally, we check that h is injective. Suppose $h(a_1) = h(a_2)$. There are three cases to consider:

- $a_1 \in X$ and $a_2 \in A \smallsetminus X$ (or vice-versa). This cannot actually occur: if $h(a_1) = h(a_2)$ then $f(a_1) = g^{-1}(a_2)$, so that
$$a_2 = g\big(g^{-1}(a_2)\big) = g\big(f(a_1)\big) \in g\big(f(X)\big) \subseteq X,$$
contradiction.
- $a_1, a_2 \in X$, then $f(a_1) = f(a_2)$ so $a_1 = a_2$ by the injectivity of f.
- $a_1, a_2 \in A \smallsetminus X$, then $g^{-1}(a_1) = g^{-1}(a_2)$ so $a_1 = a_2$ by applying g. □

A.2. Linear algebra

Unless specified otherwise, we use $\mathbf{F}$ to denote an arbitrary field.

For vector spaces V, W over $\mathbf{F}$, we write

$$\mathrm{Hom}(V, W) = \{f : V \longrightarrow W : f \text{ is a linear transformation}\}.$$

This is a vector space over $\mathbf{F}$, with zero vector given by the constant function $\mathbf{0} : V \longrightarrow W$, $\mathbf{0}(v) = 0_W$ for all $v \in V$, and with vector addition and scalar multiplication defined pointwise:

$$(f_1 + f_2)(v) = f_1(v) + f_2(v) \qquad \text{and} \qquad (\lambda f)(v) = \lambda f(v).$$

An $\mathbf{F}$-*algebra* is a vector space A over $\mathbf{F}$ together with a multiplication map $A \times A \longrightarrow A$, $(u, v) \longmapsto uv$, satisfying

- $(u + v)w = uw + vw$ for all $u, v, w \in A$;
- $u(v + w) = uv + uw$ for all $u, v, w \in A$;
- $\big(\alpha u\big)\big(\beta v\big) = \big(\alpha\beta\big)\big(uv\big)$ for all $\alpha, \beta \in \mathbf{F}$ and all $u, v \in A$.

The algebra A is *associative* if

$$(uv)w = u(vw) \qquad \text{for all } u, v, w \in A.$$

The algebra A is *unital* if there exists an element $\mathbf{1} \in A$ with the property that

$$\mathbf{1}v = v\mathbf{1} = v \qquad \text{for all } v \in A.$$

For any vector space V over $\mathbf{F}$, $\mathrm{End}(V) := \mathrm{Hom}(V, V)$ is an associative unital $\mathbf{F}$-algebra, see Exercise A.1.

An important property of a basis for a vector space is the ability to define a function on that basis and then extend it to a unique linear map. More precisely, let V and W be vector spaces over $\mathbf{F}$. Fix a basis B of V. For any function $g : B \longrightarrow W$ there exists a unique linear map $f : V \longrightarrow W$ such that $g = f|_B$, constructed in the following manner:

Given $v \in V$, there is a unique expression of the form

$$v = a_1 v_1 + \cdots + a_n v_n, \qquad n \in \mathbf{N}, a_j \in \mathbf{F}, v_j \in B.$$

Therefore the only option is to set

$$f(v) = a_1 g(v_1) + \cdots + a_n g(v_n).$$

It is easy to see that f is linear.

We say that f is obtained from g by *extending by linearity*.

A.2.1. Bilinear maps

If U, V, W are vector spaces over $\mathbf{F}$, a *bilinear map* $\beta\colon U \times V \longrightarrow W$ is a function such that

$$\begin{aligned} \beta(au_1 + bu_2, v) &= a\beta(u_1, v) + b\beta(u_2, v) \\ \beta(u, av_1 + bv_2) &= a\beta(u, v_1) + b\beta(u, v_2) \end{aligned}$$

for all $u, u_1, u_2 \in U$, $v, v_1, v_2 \in V$, $a, b \in \mathbf{F}$.

Note that such β induces maps

$$\begin{aligned} \beta_U\colon U &\longrightarrow \operatorname{Hom}(V, W), & u &\longmapsto \big(v \longmapsto \beta(u, v)\big) \\ \beta_V\colon V &\longrightarrow \operatorname{Hom}(U, W), & v &\longmapsto \big(u \longmapsto \beta(u, v)\big). \end{aligned}$$

It is easy to check that these maps are themselves linear.

A.2.2. Dual vector space

Let V be a finite dimensional vector space over $\mathbf{F}$. Define

$$V^\vee = \operatorname{Hom}(V, \mathbf{F}).$$

This is a vector space over $\mathbf{F}$, called the *dual vector space* to V. Its elements are sometimes called *(linear) functionals* and denoted with Greek letters such as φ.

Proposition A.2. *Suppose $B = \{v_1, \dots, v_n\}$ is a basis for V. Define $v_1^\vee, \dots, v_n^\vee \in \operatorname{Fun}(V, \mathbf{F})$ by*

$$v_i^\vee(a_1 v_1 + \dots + a_n v_n) = a_i \qquad \text{for } i = 1, \dots, n.$$

Then $v_i^\vee \in V^\vee$ for $i = 1, \dots, n$ and the set $B^\vee = \{v_1^\vee, \dots, v_n^\vee\}$ is a basis for $V^\vee$. (It is called the dual basis *to B.)*

Proof. We check that $v_i^\vee$ is a linear transformation.

Given $v, w \in V$, we express them in the basis B:

$$\begin{aligned} v &= a_1 v_1 + \dots + a_n v_n \\ w &= b_1 v_1 + \dots + b_n v_n, \end{aligned}$$

then

$$v_i^\vee(v + w) = v_i^\vee(a_1 v_1 + \dots + a_n v_n + b_1 v_1 + \dots + b_n v_n) = a_i + b_i = v_i^\vee(v) + v_i^\vee(w).$$

Similarly, if $\lambda \in \mathbf{F}$ we have

$$v_i^\vee(\lambda v) = v_i^\vee(\lambda a_1 v_1 + \dots + \lambda a_n v_n) = \lambda a_i = \lambda v_i^\vee(v).$$

So $v_i^\vee \in V^\vee$ for any $i = 1, \dots, n$.

Next we show that the set $B^\vee$ is linearly independent. Suppose we have

$$\lambda_1 v_1^\vee + \dots + \lambda_n v_n^\vee = 0.$$

In particular, we can apply this to the basis vector $v_i \in B$ for any $i = 1, \dots, n$ and get

$$\lambda_i = 0.$$

So all the coefficients in the above linear relation must be zero, therefore $B^\vee$ is linearly independent.

Finally, we show that the set $B^\vee$ spans $V^\vee$. Let $\varphi \in V^\vee$; let $v \in V$ and express v in the basis B:

$$v = a_1 v_1 + \cdots + a_n v_n.$$

Then, since φ is a linear transformation, we have

$$\begin{aligned}\varphi(v) &= a_1\varphi(v_1) + \cdots + a_n\varphi(v_n)\\ &= \lambda_1 v_1^\vee(v) + \cdots + \lambda_n v_n^\vee(v),\end{aligned}$$

where we let $\lambda_1 = \varphi(v_1), \ldots, \lambda_n = \varphi(v_n)$. This shows that φ is in the span of the set $B^\vee$. □

Note that a bilinear map $\beta\colon V \times W \longrightarrow \mathbf{F}$ induces linear maps

$$\begin{aligned}\beta_W\colon W \longrightarrow V^\vee, &\qquad w \longmapsto \big(w^\vee\colon v \longmapsto \beta(v,w)\big)\\ \beta_V\colon V \longrightarrow W^\vee, &\qquad v \longmapsto \big(v^\vee\colon w \longmapsto \beta(v,w)\big).\end{aligned}$$

For instance, we can take $W = V^\vee$ and define $\beta\colon V \times V^\vee \longrightarrow \mathbf{F}$ by

$$\beta(v,\varphi) = \varphi(v).$$

The corresponding linear maps are $\beta_{V^\vee} = \mathrm{id}_{V^\vee}\colon V^\vee \longrightarrow V^\vee$, and $\beta_V\colon V \longrightarrow \big(V^\vee\big)^\vee$ given by

$$\beta_V(v)(\varphi) = \beta(v,\varphi) = \varphi(v).$$

Proposition A.3. *If V is finite-dimensional, then $\beta_V\colon V \longrightarrow \big(V^\vee\big)^\vee$ is invertible.*

Proof. Let $B = \{v_1, \ldots, v_n\}$ be a basis for V and let $B^\vee = \{v_1^\vee, \ldots, v_n^\vee\}$ be the dual basis for $V^\vee$ as in Proposition A.2.

To show that β_V is injective, suppose $u, v \in V$ are such that $\beta_V(u) = \beta_V(v)$, in other words

$$\varphi(u) = \varphi(v) \qquad \text{for all } \varphi \in V^\vee.$$

Write

$$\begin{aligned}u &= a_1 v_1 + \cdots + a_n v_n\\ v &= b_1 v_1 + \cdots + b_n v_n\end{aligned}$$

then, for $i = 1, \ldots, n$, we have

$$a_i = v_i^\vee(u) = v_i^\vee(v) = b_i$$

Therefore $u = v$.

We now prove that β_V is surjective. (Note that we could get away with simply saying that Proposition A.2 tells us that V and $V^\vee$, and therefore also $\big(V^\vee\big)^\vee$, have the same dimension n; so β_V, being injective, is also surjective.)

Let $T\colon V^\vee \longrightarrow \mathbf{F}$ be a linear transformation. Define $v \in V$ by

$$v = T\big(v_1^\vee\big)v_1 + \cdots + T\big(v_n^\vee\big)v_n.$$

I claim that $\beta_V(v) = T$. For any $\varphi \in V^\vee$ we have

$$\begin{aligned}\beta_V(v)(\varphi) = \varphi(v) &= T\big(v_1^\vee\big)\varphi(v_1) + \cdots + T\big(v_n^\vee\big)\varphi(v_n)\\ &= T\big(\varphi(v_1)v_1^\vee + \cdots + \varphi(v_n)v_n^\vee\big)\\ &= T(\varphi),\end{aligned}$$

where we expressed φ in terms of the dual basis $v_1^\vee, \ldots, v_n^\vee$ from Proposition A.2. □

Proposition A.4. *Consider a linear transformation $T\colon V \longrightarrow W$, where W is another finite-dimensional vector space over $\mathbf{F}$. Define $T^\vee\colon W^\vee \longrightarrow V^\vee$ by*

$$T^\vee(\varphi) = \varphi \circ T.$$

Then $T^\vee$ is a linear transformation, called the dual linear transformation *to T.*

Proof. It is clear that $\varphi \circ T\colon V \longrightarrow \mathbf{F}$ is linear, being the composition of two linear transformations.

To show that $T^\vee\colon W^\vee \longrightarrow V^\vee$ is linear, take $\varphi_1, \varphi_2 \in W^\vee$. For any $v \in V$ we have

$$T^\vee(\varphi_1 + \varphi_2)(v) = (\varphi_1 + \varphi_2)\big(T(v)\big) = \varphi_1\big(T(v)\big) + \varphi_2\big(T(v)\big) = T^\vee(\varphi_1)(v) + T^\vee(\varphi_2)(v).$$

Similarly, if $\varphi \in W^\vee$ and $\lambda \in \mathbf{F}$, then for any $v \in V$ we have

$$T^\vee(\lambda\varphi)(v) = (\lambda\varphi)\big(T(v)\big) = \lambda\varphi\big(T(v)\big) = \lambda T^\vee(\varphi)(v). \qquad \square$$

A.2.3. Inner products

We take $\mathbf{F}$ to be either $\mathbf{R}$ or $\mathbf{C}$, and we denote by $\bar{\cdot}$ the complex conjugation (which is just the identity if $\mathbf{F} = \mathbf{R}$).

Let V be a vector space over $\mathbf{F}$.

An *inner product* on V is a function

$$\langle \cdot, \cdot \rangle\colon V \times V \longrightarrow \mathbf{F}$$

such that

(a) $\langle w, v\rangle = \overline{\langle v, w\rangle}$ for all $v, w \in V$;

(b) $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ for all $u, v, w \in V$;

(c) $\langle \alpha v, w\rangle = \alpha \langle v, w\rangle$ for all $v, w \in V$, all $\alpha \in \mathbf{F}$;

(d) $\langle v, v\rangle \geqslant 0$ for all $v \in V$ and $\langle v, v\rangle = 0$ iff $v = 0$.

Properties (a), (b), and (c) say that $\langle \cdot, \cdot \rangle$ is linear in the first variable, but *conjugate-linear* in the second:

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

(Such a function $V \times V \longrightarrow \mathbf{F}$ is called a *sesquilinear form.*)

Property (d) says that $\langle \cdot, \cdot \rangle$ is *positive-definite.*

An *inner product space* is a pair $\big(V, \langle \cdot, \cdot \rangle\big)$, where V is a vector space over $\mathbf{F}$ and $\langle \cdot, \cdot \rangle$ is an inner product on V.

Example A.5. The prototypical inner product on $\mathbf{C}^n$ is

$$\langle u, v\rangle = \sum_{k=1}^{n} u_k \overline{v}_k = \overline{v}^T u,$$

which on $\mathbf{R}^n$ becomes

$$\langle u, v\rangle = \sum_{k=1}^{n} u_k v_k = v^T u.$$

All other inner products on $\mathbf{C}^n$ are of the form

$$\langle u, v\rangle = \overline{v}^T A u,$$

where A is an $n \times n$ *positive-definite Hermitian matrix*, that is

$$\overline{A}^T = A \qquad \text{and all the eigenvalues of } A \text{ are real and positive.}$$

Over $\mathbf{R}$, A is a positive-definite symmetric matrix.

Define

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

Proposition A.6 (Cauchy–Schwarz Inequality)**.** *Take u, v in an inner product space V. Then*

$$|\langle u, v\rangle| \leqslant \|u\| \, \|v\|,$$

where equality holds if and only if u and v are parallel.

Proof. If $u = \mathbf{0}$ or $v = \mathbf{0}$, we have the equality $0 = 0$. Otherwise, for any $t \in \mathbf{F}$ we have

$$\begin{aligned} 0 \leqslant \langle u - tv, u - tv\rangle &= \langle u, u\rangle - 2\operatorname{Re}\left(\bar{t}\langle u, v\rangle\right) + t\bar{t}\langle v, v\rangle \\ &= \|u\|^2 - 2\operatorname{Re}\left(\bar{t}\langle u, v\rangle\right) + |t|^2\|v\|^2. \end{aligned}$$

In particular, we can take $t = \frac{\langle u,v\rangle}{\|v\|^2}$:

$$0 \leqslant \|u\|^2 - 2\operatorname{Re}\left(\frac{|\langle u, v\rangle|^2}{\|v\|^2}\right) + \frac{|\langle u, v\rangle|^2}{\|v\|^2} = \|u\|^2 - \frac{|\langle u, v\rangle|^2}{\|v\|^2},$$

so $|\langle u, v\rangle|^2 \leqslant \|u\|^2 \, \|v\|^2$. □

Bibliography

[1] Edgar Asplund and Lutz Bungart. *A first course in integration*. Holt, Rinehart and Winston, New York-Toronto-London, 1966.

[2] Karen E. Smith. Bases for infinite dimensional vector spaces. https://dept.math.lsa.umich.edu/~kesmith/infinite.pdf, University of Michigan lecture notes, 1999. [Online; accessed 21 July 2023].

Index

absolute value
- p-adic, 12
- complex, 9
- real, 9

adjoint map, 72
algebra, 90
- associative, 90
- unital, 90

Banach space, 49
basis, 5
- Schauder, 62

Bessel's Inequality, 74
bilinear map, 63, 91
boundary, 19

Cauchy sequence, 29
Cauchy–Schwarz Inequality, 94
closed ball, 14
closed set, 16, 19
closure, 19
complete metric space, 29
completion, 31
conjugate-linear, 93
conjugation, 26
continuous, 17
contraction, 34
convergence
- for a sequence, 27

convex function, 52
convex subset, 51

dense, 20
diameter, 36
distance, 11
distance-preserving, 28
dual basis, 91
dual linear transformation, 93
dual vector space, 91

equicontinuous, 42
equivalence of sequences, 28
extend by linearity, 90
extension, 32

fixed point, 34
Fourier coefficients, 82
Fourier expansion, 82
function
- bounded, 39
- convex, 52

functional, 91

Gram–Schmidt Orthogonalisation, 76

Hölder conjugate, 63
Hausdorff
- topological space, 19

Hermitian matrix, 73
Hilbert space, 50
homeomorphic, 17
homeomorphism, 17

identity
- polarisation, 51

inequality
- Cauchy–Schwarz, 94
- Hölder, 63
- Minkowski, 54

infimum, 22
inner product, 49, 93
interior, 19
interpolate pair of points, 78
interval, 22
isometry, 28

Lebesgue number, 38
Legendre polynomials, 82
lemma
- Zorn, 6

limit
- of a sequence, 27

linear combination, 5
linear transformation
- bounded, 56
- Lipschitz, 56

Lipschitz, 56
Lipschitz bilinear map, 64

map

- closed, 24
- compact, 84
- finite rank, 84
- proper, 24

matrix
- Hermitian, 94
- positive-definite, 94

metric, 11
- ℓ^1, 14
- ℓ^∞, 14
- p-adic, 12
- coarse, 17
- equivalent, 17
- fine, 17
- Hamming, 14
- induced, 13
- Manhattan, 14
- sup norm, 14
- taxicab, 14
- uniform, 39
- uniform norm, 14

metric space, 11
- complete, 29

metric spaces
- isometric, 28

metrisable, 16

neighbourhood, 18
- open, 18

norm, 45
- L^∞, 77
- equivalence, 46
- uniform, 77

normal map, 73
normed space, 45
- completion, 55

nowhere dense, 20

open ball, 14
open cover, 23
open rectangle, 18
open set, 15, 16
orthogonal complement, 67
orthogonal projection, 67
orthonormal basis, 73
orthonormal system, 73

parallelogram law, 50
Parseval's Identity, 74
partial sum, 60
pointwise convergence, 40
polarisation identity, 51
positive-definite form, 93
projection, 66
- orthogonal, 67

resolvent, 85

Schauder basis, 62
self-adjoint map, 73
semi-norm, 45
separation property, 19
sequence, 27
- p-summable, 54
- bounded, 54
- Cauchy, 29
- equivalent, 28
- square-summable, 54
- summable, 54

sequential compactness, 38
series, 60
- absolutely convergent, 60
- convergent, 60

sesquilinear form, 93
set
- bounded, 36
- closed, 19
- compact, 23
- countable, 9
- countably infinite, 9
- open, 15
- totally bounded, 37
- unbounded, 36

space
- Banach, 49
- connected, 21
- disconnected, 20
- Hilbert, 50
- homogeneous, 26
- inner product, 93
- metric, 11
- normed, 45
- of sequences, 53
- separable, 37
- totally disconnected, 21

spectrum, 85
sphere, 15
subcover, 23
supremum, 22

theorem
- Arzelà–Ascoli, 43

Banach Fixed Point, 34
Heine–Borel, 38
Intermediate Value, 22
Mean Value, 35
Riesz Representation, 71
Schröder–Bernstein, 89
Weierstrass Approximation, 77
topological group, 25
topological space, 16
topological vector space, 46
topology, 16
coarse, 17
cofinite, 16
discrete, 16
fine, 17
generated by a set, 17
induced, 18
metric, 16
product, 18
subspace, 18
trivial, 16
translation
left, 26
right, 26

uniform convergence, 40
uniform norm, 77
uniformly continuous, 30

valuation
p-adic, 11

Zorn's Lemma, 6