NOTES ON METRIC AND HILBERT SPACES
AN INVITATION TO FUNCTIONAL ANALYSIS

Alexandru Ghitza*
School of Mathematics and Statistics
University of Melbourne

Version of Fri 6th Sept, 2024 at 19:52

---

*(`aghitza@alum.mit.edu`)

CONTENTS

1. INTRODUCTION

1.1. WHAT'S UP WITH INFINITE-DIMENSIONAL VECTOR SPACES?

The discussion in this section is heavily inspired by the lecture notes [1] 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

$$a_1v_1 + \cdots + a_nv_n \qquad \text{where } n \in \mathbf{N}, \quad a_1, \ldots, a_n \in \mathbf{F}, \quad v_1, \ldots, v_n \in V.$$

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 $B = \left\{1, x, x^2, \ldots\right\}$.

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

$$f = a_0 + a_1x + \cdots + a_nx^n,$$

thus a linear combination of elements of B.

If we have

$$f = a_0 + a_1x + \cdots + a_nx^n = b_0 + b_1x + \cdots + b_mx^m,$$

---

[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 $i = 0, \dots, n$. ☐

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 $v := v_p(x) \leqslant v_p(y) =: u$ 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
>
> $$\begin{aligned} f^{-1}(U) &= f^{-1}\big((a,b)\big) \\ &= \{x \in X\colon a < d(x,t) < b\} \\ &= \big\{x \in X\colon a < d(x,t)\big\} \cap \{x \in X\colon d(x,t) < b\} \\ &= \big(X \smallsetminus \mathbf{D}_a(t)\big) \cap \mathbf{B}_b(t), \end{aligned}$$
>
> 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

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

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

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

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

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

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:

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

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 $X \smallsetminus C$ 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 $\partial A = \overline{A} \cap \overline{X \smallsetminus A}$.

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: $x \in X \smallsetminus \overline{A}$ if and only if there exists an open neighbourhood U_x of x such that $U_x \cap A = \varnothing$.

Suppose $x \in X \smallsetminus \overline{A}$. Letting $U_x = X \smallsetminus \overline{A}$, 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, $X \smallsetminus U_x$ is closed and contains A, so it contains the closure $\overline{A}$. Hence $x \in X \smallsetminus \overline{A}$. ☐

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 $\partial A = \overline{A} \cap \overline{(X \smallsetminus A)}$.

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

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 $V := X \smallsetminus U$. 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 $x \in \mathbf{R} \smallsetminus A$. 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 $(M, x] \subseteq \mathbf{R} \smallsetminus A$, therefore $M \in \overline{\mathbf{R} \smallsetminus A}$ but $M \notin \big(\mathbf{R} \smallsetminus A\big)^\circ$, hence $M \in \partial\big(\mathbf{R} \smallsetminus A\big) = \partial A$.
- $S := (x, \infty) \cap A \neq \varnothing$, 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 $x \in \mathbf{R} \smallsetminus S$ 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 $i_1, \dots, i_n \in I$ such that

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

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 $X \setminus K$ is open. Let $x \in X \setminus K$. 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

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

Consider

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

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

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

As X is compact, there is a finite subcover

$$X \subseteq (X \setminus K) \cup U_{i_1} \cup \dots \cup U_{i_n}.$$

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

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

☐

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 $U = X_2 \smallsetminus \pi_2(C)$ and let $u \in U$. Then $u \notin \pi_2(C)$; so for any $x \in X_1$, we have that $(x,u) \in (X_1 \times X_2) \smallsetminus C$. 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 $V \subseteq X_2 \smallsetminus \pi_2(C) = U$, 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 $\delta \in \mathbf{R}_{>0}$, $f\big(\mathbf{B}_\delta(x)\big) \smallsetminus \mathbf{B}_\varepsilon(f(x)) \neq \varnothing$. 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

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

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

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 $d_Y(f(s), f(t)) < \varepsilon$. $\square$

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:

$$\begin{tikzcd}
X \arrow[r, hook, "\iota_X"] \arrow[d, "f"'] & \widehat{X} \arrow[d, "\widehat{f}"] \\
Y \arrow[r, hook, "\iota_Y"] & \widehat{Y}
\end{tikzcd}$$

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

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

(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

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

This has the following consequence:

Corollary 2.64. *Let (X, d) be a metric space.*

(a) Any uniformly continuous (resp. distance-preserving) function $g \colon X \longrightarrow Y$ between arbitrary metric spaces has a unique uniformly continuous (resp. distance-preserving) extension to completions, $\widehat{g} \colon \widehat{X} \longrightarrow \widehat{Y}$.

(b) Any two completions of (X, d) are isometric.

Proof.

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

(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 $f \coloneqq \iota_2 \circ \iota_1^{-1}\colon \iota_1(X) \longrightarrow \iota_2(X)$. 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 (x_n) is Cauchy and 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. □

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:

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

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, $x_{n+1} = f(x_n)$. $\square$

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

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

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 $x_1, \dots, x_N \in X$ 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

---

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

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 $\big(s_n^{(j)}\big) \subseteq \mathbf{B}_{1/j}(x_j)$:

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

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 $s_2 \in S \smallsetminus \mathbf{B}_\varepsilon(s_1)$. If $S \subseteq \mathbf{B}_\varepsilon(s_1) \cup \mathbf{B}_\varepsilon(s_2)$, we are done. Otherwise, there exists $s_3 \in S \smallsetminus \big(\mathbf{B}_\varepsilon(s_1) \cup \mathbf{B}_\varepsilon(s_2)\big)$.

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 $j = 1, \ldots, n$ 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 and Y be metric spaces.

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 $f(x) \in Y$:

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

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:

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

Proposition 2.77. *Let X, Y be metric spaces.*

(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 metric spaces X and 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. ☐

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 $d_X(x_0,x) < \delta$:

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

> **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 $X \longrightarrow \mathbf{R}^m$.

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 $C_0(X,Y)$.*

---

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

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 $i = 1,2,\ldots,j$:

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

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

$$\begin{aligned} S + T &= \{s + t\colon\ s \in S, t \in T\}, \\ \alpha S &= \{\alpha s\colon\ s \in S\}. \end{aligned}$$

3.1. NORMS

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

A *norm* on V is a function

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

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

Proposition 3.1. *Let* $\big(V, \|\cdot\|\big)$ *be a normed space. Define* $d\colon V \times V \longrightarrow \mathbf{R}_{\geqslant 0}$ *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 $(V, \|\cdot\|)$, we always endow W with the restriction of $\|\cdot\|$ to W, which is a norm on W.

Proposition 3.2. *Any normed space $(V, \|\cdot\|)$ 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:

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

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) $\lim\limits_{n\longrightarrow\infty}(v_n + w_n) = \lim\limits_{n\longrightarrow\infty} v_n + \lim\limits_{n\longrightarrow\infty} w_n$;

(b) $\lim\limits_{n\longrightarrow\infty}(\alpha\, v_n) = \alpha \lim\limits_{n\longrightarrow\infty} v_n$;

(c) $\lim\limits_{n\longrightarrow\infty}\|v_n\| = \Big\|\lim\limits_{n\longrightarrow\infty} v_n\Big\|$.

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. Each of the following functions give norms on the vector space $V \times W$:*

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

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* $\big(V,\|\cdot\|\big)$*. Then there exists* $m > 0$ *such that*

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

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

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

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

To do this, consider the set

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

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

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

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

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

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

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

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

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

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

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

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

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 $v_1, \dots, v_n$ 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

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

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 $j = 1, \dots, n$, $(\alpha_j^{(m)})$ 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 $j = 1, \dots, n$, there exists $M_j \in \mathbf{N}$ such that $|\alpha_j^{(m)} - \beta_j| < \varepsilon/n$ for all $m \geqslant M_j$. Let $M = \max\{M_j : j = 1, \dots, n\}$, then for all $m \geqslant M$ we have

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

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.2.2 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* $\big(V, \langle\cdot,\cdot\rangle\big)$ *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* $\big(V, \langle\cdot,\cdot\rangle\big)$ *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

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

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 $\big(V,\langle\cdot,\cdot\rangle\big)$ is an inner product space then*

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

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

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

then the function $\langle\cdot,\cdot\rangle$ defined by

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

is an inner product on V with associated norm $\|\cdot\|$.

Proof. See TODO. ☐

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 $f\colon S \longrightarrow \mathbf{R}$ 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 $\|\cdot\|\colon V \longrightarrow \mathbf{R}_{\geqslant 0}$ 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 $f\colon I \longrightarrow \mathbf{R}$ 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

$$\begin{array}{lll} \text{(i)} & f\colon (0,\infty) \longrightarrow \mathbf{R}, & f(x) = x^p, \quad p \geqslant 1 \text{ fixed}, \\ \text{(ii)} & \exp\colon \mathbf{R} \longrightarrow \mathbf{R}, & \exp(x) = e^x, \end{array}$$

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
$$\min\{x,y\} = x = \left(ax^{-q} + bx^{-q}\right)^{-1/q} \leqslant \left(ax^{-q} + by^{-q}\right)^{-1/q}.$$

(b) Let $X = x^{-q}$, $Y = y^{-q}$, then by Proposition 3.19 part (b) we have
$$\begin{aligned}X^a Y^b \leqslant aX + bY &\Rightarrow x^{-aq} y^{-bq} \leqslant ax^{-q} + by^{-q} \\ &\Rightarrow x^{aq} y^{bq} \geqslant \left(ax^{-q} + by^{-q}\right)^{-1} \\ &\Rightarrow \left(ax^{-q} + by^{-q}\right)^{-1/q} \leqslant x^a y^b.\end{aligned}$$

(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.

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 $X_1 = A \smallsetminus g(B)$. 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 $a \in A \smallsetminus X$, then $a \notin X_1 = A \smallsetminus g(B)$, 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

$$h(a) = \begin{cases} f(a) & \text{if } a \in X \\ g^{-1}(a) & \text{if } a \in A \smallsetminus X \end{cases}$$

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 $g(b) \notin X_1 = A \smallsetminus g(B)$. 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 $a_2 \in A \setminus X$ (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.
- $a_1, a_2 \in A \setminus X$, 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

$$\operatorname{Hom}(V, W) = \{f\colon V \longrightarrow W : f \text{ is a linear transformation}\}.$$

This is a vector space over $\mathbf{F}$, with zero vector given by the constant function $\mathbf{0}\colon V \longrightarrow W$, $\mathbf{0}(v) = 0_W$ 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 $\mathbf{F}$, $\operatorname{End}(V) := \operatorname{Hom}(V, V)$ 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 $g\colon B \longrightarrow W$ there exists a unique linear map $f\colon V \longrightarrow W$ 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. 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*

$$v_i^\vee(a_1v_1 + \cdots + a_nv_n) = a_i \qquad \text{for } i = 1, \dots, n.$$

Then $v_i^\vee \in V^\vee$ for $i = 1, \dots, n$ and the set $B^\vee = \left\{v_1^\vee, \dots, v_n^\vee\right\}$ 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:

$$\begin{aligned} v &= a_1v_1 + \cdots + a_nv_n \\ w &= b_1v_1 + \cdots + b_nv_n, \end{aligned}$$

then

$$v_i^\vee(v + w) = v_i^\vee(a_1v_1 + \cdots + a_nv_n + b_1v_1 + \cdots + b_nv_n) = a_i + b_i = v_i^\vee(v) + v_i^\vee(w).$$

Similarly, if $\lambda \in \mathbf{F}$ we have

$$v_i^\vee(\lambda v) = v_i^\vee(\lambda a_1v_1 + \cdots + \lambda a_nv_n) = \lambda a_i = \lambda v_i^\vee(v).$$

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

$$\lambda_1v_1^\vee + \cdots + \lambda_nv_n^\vee = 0.$$

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:

$$v = a_1v_1 + \cdots + a_nv_n.$$

Then, since φ is a linear transformation, we have

$$\begin{aligned} \varphi(v) &= a_1\varphi(v_1) + \cdots + a_n\varphi(v_n) \\ &= \lambda_1v_1^\vee(v) + \cdots + \lambda_nv_n^\vee(v), \end{aligned}$$

where we let $\lambda_1 = \varphi(v_1), \dots, \lambda_n = \varphi(v_n)$. This shows that φ is in the span of the set $B^\vee$. □

If V and W are vector spaces over $\mathbf{F}$, then a function $\beta\colon V \times W \longrightarrow \mathbf{F}$ is said to be a *bilinear map* if

(a) $\beta(av_1 + bv_2, w) = a\beta(v_1, w) + b\beta(v_2, w)$ for all $v_1, v_2 \in V$, $w \in W$, $a, b \in \mathbf{F}$;

(b) $\beta(v, aw_1 + bw_2) = a\beta(v, w_1) + b\beta(v, w_2)$ for all $v \in V$, $w_1, w_2 \in W$, $a, b \in \mathbf{F}$.

It is called a *bilinear form* if $W = V$.

Note that β induces linear maps

$$\begin{aligned} \beta_W\colon W &\longrightarrow V^\vee, & w &\longmapsto \bigl(w^\vee\colon v \longmapsto \beta(v,w)\bigr) \\ \beta_V\colon V &\longrightarrow W^\vee, & v &\longmapsto \bigl(v^\vee\colon w \longmapsto \beta(v,w)\bigr). \end{aligned}$$

For instance, we can take $W = V^\vee$ and consider $\beta\colon V \times V^\vee \longrightarrow \mathbf{F}$ given 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 $\beta_V\colon V \longrightarrow \bigl(V^\vee\bigr)^\vee$ given by

$$\beta_V(v)(\varphi) = \beta(v,\varphi) = \varphi(v).$$

Proposition A.3. *If V is finite-dimensional, then $\beta_V\colon V \longrightarrow \bigl(V^\vee\bigr)^\vee$ is invertible.*

Proof. Let $B = \{v_1,\dots,v_n\}$ be a basis for V and let $B^\vee = \{v_1^\vee,\dots,v_n^\vee\}$ 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

$$\begin{aligned} u &= a_1 v_1 + \dots + a_n v_n \\ v &= b_1 v_1 + \dots + b_n v_n \end{aligned}$$

then, for $i = 1,\dots,n$, 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 $\bigl(V^\vee\bigr)^\vee$, 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

$$v = T\bigl(v_1^\vee\bigr)v_1 + \dots + T\bigl(v_n^\vee\bigr)v_n.$$

I claim that $\beta_V(v) = T$. For any $\varphi \in V^\vee$ we have

$$\begin{aligned} \beta_V(v)(\varphi) = \varphi(v) &= T\bigl(v_1^\vee\bigr)\varphi(v_1) + \dots + T\bigl(v_n^\vee\bigr)\varphi(v_n) \\ &= T\bigl(\varphi(v_1)v_1^\vee + \dots + \varphi(v_n)v_n^\vee\bigr) \\ &= T(\varphi), \end{aligned}$$

where we expressed φ in terms of the dual basis $v_1^\vee,\dots,v_n^\vee$ 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

$$T^\vee(\varphi_1 + \varphi_2)(v) = (\varphi_1 + \varphi_2)\bigl(T(v)\bigr) = \varphi_1\bigl(T(v)\bigr) + \varphi_2\bigl(T(v)\bigr) = T^\vee(\varphi_1)(v) + T^\vee(\varphi_2)(v).$$

Similarly, if $\varphi \in W^\vee$ and $\lambda \in \mathbf{F}$, then for any $v \in V$ we have

$$T^\vee(\lambda\varphi)(v) = (\lambda\varphi)\bigl(T(v)\bigr) = \lambda\varphi\bigl(T(v)\bigr) = \lambda T^\vee(\varphi)(v).$$ ☐

A.2.2. Inner products

We take **F** to be either **R** or **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 **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 **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 **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 $|\langle u, v \rangle|^2 \leqslant \|u\|^2 \, \|v\|^2$. $\square$

BIBLIOGRAPHY

[1] 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