Notes on metric and Hilbert spaces
An invitation to functional analysis

Alexandru Ghitza*
School of Mathematics and Statistics
University of Melbourne

Version of Sat 23rd Aug, 2025 at 16:17

*(aghitza@alum.mit.edu)

Contents

1. Introduction

1.1. Infinite-dimensional spaces?

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

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

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

First year linear algebra tells us that every finite-dimensional vector space V has a basis[1]. What happens if V is not finite-dimensional?

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

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

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

thus a linear combination of elements of B.

If we have

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

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

This first example worked out great: the space has bases, and we can actually write down a basis (more precisely, the standard basis) explicitly. We owe our luck to the fact that, even

[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.

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 prototypical finite-dimensional spaces we know, say $\mathbf{R}^n$, and "take the limit as $n \longrightarrow \infty$". This leads us to consider the space $\mathbf{R}^{\mathbf{N}}$ 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 $\mathbf{R}^{\mathbf{N}}$, 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 B.5 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 B.6.) In fact, $\mathbf{R}$ has no countable basis over $\mathbf{Q}$. (See Exercise A.9.) 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.*

For the proof of this theorem, see Exercise B.3; it requires the (in)famous Zorn's Lemma (Lemma B.1).

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 arbitrary 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.

1.2. Plan(-ish)

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 (not necessarily vector spaces) 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**. One natural example is 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 in applications, 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 statements 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 $N \longrightarrow \infty$.

Spaces such as this inner product space V are pretty nice, but in general there will be infinite series of elements of V that "look like" they should converge, however their limit is not in V.

We deal with this by restricting to inner product spaces that are complete (every Cauchy sequence in V converges to an element of V); these are called **Hilbert spaces**, and have nice properties.

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.

1.3. 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 $S \not\subseteq T$ (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

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

Unless otherwise specified, $\mathbf{F}$ denotes an arbitrary field in Chapter 2, and it denotes either $\mathbf{R}$ or $\mathbf{C}$ in Chapter 3.

I am not the right person to ask about foundational questions of logic or set theory: I neither know enough nor 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 other natural examples:

Example 2.1. Let Γ be a 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 $x = y$). □

Other examples are quite exotic, see for instance the p-adic metric in Exercise 1.10.

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

Example 2.2. 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 2.3. 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 $(X_1, d_{X_1}), \ldots, (X_n, d_{X_n})$.

(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 *$\ell^1$ metric*.)

Solution. Straightforward. □

Example 2.4. Same setup as Example 2.3, 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 $(X_1, d_{X_1}), \ldots, (X_n, d_{X_n})$.

(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 2.5. Take $X_1 = X_2 = \mathbf{R}$ with the Euclidean metric and convince yourself that neither d_1 from Example 2.3 nor d_∞ from Example 2.4 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 Exercise 1.10 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 $i \in \{1, \ldots, n\}$ such that $x_i \neq y_i$. Then d is a metric on X; it is called the *Hamming metric*. See Exercise 1.6 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 2.6. 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 1.4 and 1.11.

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 2.7. 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 2.8. 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 2.9. *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 2.10. *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 and continuous functions

Let X be a set. Taking a hint from the previous section, we define a *topology* on X to be 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.

While we're at it, here are a few more useful definitions:

- 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.

The prototypical example (for us) of a topology is the collection of open sets defined on X by some distance function d (see Proposition 2.24 for details). However, topological spaces are a very general concept encompassing much more than metric spaces. We will not place a heavy emphasis 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 2.11.** Let X be an arbitrary set and let $\mathcal{T} = \{\varnothing, X\}$. This is called the *trivial topology* on X.

> **Example 2.12.** 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 2.13. 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.

Comparing topologies

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 $\mathcal{T}_1$. For example, the discrete topology is the finest possible topology on X, and the trivial topology is the coarsest possible topology on X.

In Tutorial Question 2.3 you will find all (4) 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 2.14. 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 (aka one-point subsets).

The topology generated by S is the coarsest topology $\mathcal{T}$ such that $S \subseteq \mathcal{T}$ (see Tutorial Question 2.4).

Continuous functions

The appropriate notion of morphism for topological spaces is that of continuous function: if $f: X \longrightarrow Y$ 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 $f: X \longrightarrow Y$ such that $f^{-1}: Y \longrightarrow X$ 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.)

We can now state a couple of equivalent criteria for comparing topologies:

Proposition 2.15. *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 $f: (X, \mathcal{T}_1) \longrightarrow (X, \mathcal{T}_2)$ given by $f(x) = x$ is continuous.

Proof. See Exercise 1.19. □

New topological spaces from old

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

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

Then $\mathcal{T}|_Y$ is a topology on Y, called the *induced (or subspace) topology*.

Example 2.16. While the induced topology is a simple and natural construction, the result can be surprising unless you keep in mind that **openness is a relative notion** rather than an absolute one.

For instance, take $X = \mathbf{R}$ with its usual Euclidean topology.

- If $Y = (0, 2)$ then $(1, 2) = (1, 3) \cap Y$ is open in Y. Of course $(1, 2)$ is also open in X in this case.
- If $Y = (0, 2]$ then $(1, 2] = (1, 3) \cap Y$ is open in Y, but certainly not open in X.
- If Y is a complicated subset of X, then its open subsets will necessarily look very different to the open subsets of X. You could take Y to be the Cantor ternary set, for instance.

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

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

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

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

□

Proposition 2.18. *Let X_1, X_2 be topological spaces and endow $X_1 \times X_2$ with the product topology. Then the two projection maps*

$$\begin{aligned} \pi_1 \colon X_1 \times X_2 &\longrightarrow X_1, & \pi_1(x_1, x_2) &= x_1 \\ \pi_2 \colon X_1 \times X_2 &\longrightarrow X_2, & \pi_2(x_1, x_2) &= x_2 \end{aligned}$$

are continuous.

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

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

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

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

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

Separation property: Hausdorff

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 other separation properties, weaker or stronger than Hausdorffness.)

Example 2.19. Consider $\mathbf{R}$ with the cofinite topology (see Example 2.13). This is not a Hausdorff topological space.

Solution. We can do this with any two distinct points of $\mathbf{R}$, but let's take $x = 0$ and $y = 1$ for concreteness.

Suppose the space is Hausdorff, and let U be an open neighbourhood of x and V one of y such that $U \cap V = \varnothing$.

Then U and V be non-empty open subsets of $\mathbf{R}$ (in the cofinite topology). Therefore

$$U = \mathbf{R} \setminus \{x_1, \dots, x_n\}$$
$$V = \mathbf{R} \setminus \{y_1, \dots, y_m\},$$

for some $m, n \in \mathbf{N}$. Then

$$U \cap V = \mathbf{R} \setminus \{x_1, \dots, x_n, y_1, \dots, y_m\}.$$

In particular, $U \cap V$ is uncountable, and so most definitely non-empty, contradiction. □

Interior, closure

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

Because of the interplay between open and closed sets, collections of closed sets have properties that are complementary to those of collections of open sets, see Exercise 1.13.

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

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

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

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

Proposition 2.20. *If A is a subset of a topological space X, then $x \in \overline{A}$ if and only if every open neighbourhood of x intersects A nontrivially.*

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

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

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

Proposition 2.21. *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) $A^\circ = A \smallsetminus \partial A$.

Proof. See Tutorial Question 2.6. ☐

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 Exercise 1.31.

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

Proposition 2.22. *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 $A \subseteq \left(X \smallsetminus U\right)$. The latter is a closed set containing A, so by the definition of the closure we have $\overline{A} \subseteq \left(X \smallsetminus U\right) \subsetneq X$, 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 $U := X \smallsetminus \overline{A}$ 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 2.23. Consider $\mathbf{R}$ with its usual topology. Both $\mathbf{Q}$ and $\mathbf{R} \smallsetminus \mathbf{Q}$ 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 Tutorial Question 3.3.

2.4. Properties of topologies induced by metrics

In this section, we revisit some of the concepts introduced above in the special context of metric spaces.

Proposition 2.24. *Every metric space is a topological space.*

In other words, if (X, d) is a metric space and $\mathcal{T}$ is the set of open subsets of X (defined as in Section 2.2), then $\mathcal{T}$ is a topology on X (called the metric topology on X).

The metric topology is generated by the set of open balls in X.

Proof. Put together Example 2.7, Propositions 2.9 and 2.10, and Tutorial Question 3.4. ☐

Looking at things from the other end, a topological space $(X, \mathcal{T})$ is said to be *metrisable* if there exists a metric d on X such that the resulting collection of open sets is precisely $\mathcal{T}$.

Not every topological space is metrisable! (For a (non)-example, see Tutorial Question 2.3.)

In 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 detailed in Tutorial Question 2.8.

Example 2.25. 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 Exercise 1.22 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 Tutorial Question 3.4, 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 1.9 to deduce that $\mathbf{D}_a(t)$ is a closed set.) □

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.

To see an important example of this, let's revisit the product of metric spaces:

Example 2.26. In Exercise 1.4 we considered $X = \mathbf{R}$ and $X \times X = \mathbf{R}^2$ endowed with three different metrics:

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

These three different metrics give rise to the same topology on $\mathbf{R}^2$ (which is the same as the product topology); to see this, use the criterion from Proposition 2.15.

Example 2.27. 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 $U \cap V = \varnothing$. □

2.5. Connectedness

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

$$U \cup V = X \qquad \text{and} \qquad U \cap V = \varnothing.$$

A nonempty[1] topological space X is *connected* if it is not disconnected.

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

The set $\{0,1\} = \{0\} \cup \{1\}$ with the discrete topology is clearly disconnected.

A subset S of a topological space X is *connected* if it is connected when considered as a topological space with the subspace topology. (It is sometimes useful to express this directly in terms of the topology on X, see Exercise 1.40.)

Connectedness is an **intrinsic** property (it does not depend on the ambient space). By contrast, being open is a **non-intrinsic** property: $(0,1)$ is open in $\mathbf{R}$ but not open in $\mathbf{R}^2$.

A *clopen* subset of a topological space is a subset that is both open and closed.

Proposition 2.29. *A subset S of a topological space X is clopen if and only if it has empty boundary: $\partial S = \varnothing$.*

Proof. In Tutorial Question 2.6 we see that for any subset S of X we have

$$\overline{S} = \partial S \sqcup S^\circ \qquad \text{(a disjoint union).}$$

In particular, $\partial S = \overline{S} \setminus S^\circ$. But also $S^\circ \subseteq S \subseteq \overline{S}$.

Therefore S is clopen if and only if $\overline{S} = S = S^\circ$ if and only if $\partial S = \varnothing$. □

The relevance of clopen subsets to connectedness is given by

Proposition 2.30. *A topological space X is disconnected if and only if X has a non-empty proper clopen subset.*

Proof. If X is disconnected, then there exist non-empty open subsets U and V such that

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

Then $U = X \setminus V$ is both open and closed, and it is nonempty and proper.

Conversely, suppose U is a non-empty proper clopen subset of X. Let $V = X \setminus U$, then V is non-empty and open and

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

In practice, a very useful criterion for connectedness is

Proposition 2.31. *Let X be a topological space. The following are equivalent:*

(a) X is connected.

(b) For every discrete space D, if $f\colon X \longrightarrow D$ is a continuous function, then f is constant.

(c) For the discrete space $\{0,1\}$, if $f\colon X \longrightarrow \{0,1\}$ is a continuous function, then f is constant.

Proof. **(a) ⇒ (b):** Suppose there exist a discrete space D and a non-constant continuous function $f\colon X \longrightarrow D$. Let $d \in \operatorname{im}(f)$ and let $U := f^{-1}(d)$. Then U is non-empty (since $d \in \operatorname{im}(f)$), proper (since f is non-constant), clopen (since $\{d\}$ is clopen in the discrete topology on D), hence X is disconnected.

(b) ⇒ (c): Clear.

[1] The empty topological space $\varnothing$ is considered neither connected nor disconnected. (A bit like: 1 is considered neither prime nor composite.)

(c) ⇒ (a): Suppose X is disconnected and let U, V be non-empty open subsets of X such that

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

Define $f\colon X \longrightarrow \{0,1\}$ by

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

This is well-defined since every $x \in X$ is in U or in V, but not in both. It is also continuous since $f^{-1}(\{0\}) = U$ is open and $f^{-1}(\{1\}) = V$ is open. Finally, f is non-constant since both U and V are non-empty. □

Often, proofs of (dis)connectedness are simpler and more elegant when using this criterion. For instance:

Proposition 2.32. *If $f\colon X \longrightarrow Y$ is a continuous surjective function between topological spaces and X is connected, then Y is connected.*

Proof. Let $g\colon Y \longrightarrow \{0,1\}$ be a continuous map. Then $g \circ f\colon X \longrightarrow Y \longrightarrow \{0,1\}$ is continuous, therefore is constant (since X is connected), hence g must be constant. (Otherwise g would be surjective, so $g \circ f$ would be a composition of surjective functions, hence not constant.)

We conclude by Proposition 2.31. □

Proposition 2.33. *The space* $\mathbf{R}$ *(with the Euclidean topology) is connected.*

Proof. Suppose U is a non-empty clopen subset of $\mathbf{R}$. We claim that $U = \mathbf{R}$.

Take any $x \in U$ and define

$$S := \left\{r \in \mathbf{R}_{>0}\colon (x - r, x + r) \subseteq U\right\}.$$

S is non-empty since U is non-empty and open. If we can show that $S = (0, \infty)$, then $U = \mathbf{R}$ and we are done.

We proceed by contradiction. Suppose S is bounded above. Then $s = \sup S$ exists[2].

We must have $x + s \in U$. Otherwise, since U is closed, $\mathbf{R} \smallsetminus U$ is open, so $x + s \in \mathbf{R} \smallsetminus U$ means that there exists $\delta > 0$ such that $(x + s - \delta, x + s + \delta) \subset \mathbf{R} \smallsetminus U$, so $s - \delta \notin S$, contradicting the fact that $s = \sup S$ (more precisely, that s is the **lowest** upper bound of S).

By a similar argument, $x - s \in U$. Since U is open, there exist $\varepsilon_+, \varepsilon_- > 0$ such that

$$(x + s - \varepsilon_+, x + s + \varepsilon_+) \subseteq U, \qquad (x - s - \varepsilon_-, x - s + \varepsilon_-) \subseteq U.$$

Let $\varepsilon = \min\{\varepsilon_+, \varepsilon_-\}$, then $s + \varepsilon \in S$, contradicting the fact that $s = \sup S$ (more precisely, that s is an upper bound for S). □

We are familiar with the notion of intervals in $\mathbf{R}$; they are precisely the subsets appearing on the following list:

- $\varnothing$;
- for $a \leqslant b$: $[a, b]$;
- for $a < b$: (a, b) and $(a, b]$ and $[a, b)$;

[2] Recall that given a subset $S \subseteq \mathbf{R}$, $s \in \mathbf{R}$ is a *supremum* of S if it is an upper bound for S (that is, $x \leqslant s$ for all $x \in S$), and if $m \in \mathbf{R}$ is any upper bound for S then $s \leqslant m$.

$\mathbf{R}$ has the property that every nonempty bounded above subset has a unique supremum. There is also a dual notion of infimum.

- for any $a \in \mathbf{R}$: $(-\infty, a)$ and $(-\infty, a]$ and (a, ∞) and $[a, \infty)$;
- $\mathbf{R}$.

A different way of describing intervals is sometimes handy:

Given a subset S of $\mathbf{R}$, consider its infimum $\inf S$ and supremum $\sup S$. If S is not bounded below, let's agree temporarily that $\inf S = -\infty$; if S is not bounded above, that $\sup S = \infty$. With this convention, S is an interval if and only if: for any $x \in \mathbf{R}$ with $\inf S < x < \sup S$, we have $x \in S$. (You should convince yourself that this is reasonable.)

Proposition 2.34. *The connected subsets of* $\mathbf{R}$ *are the non-empty intervals.*

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

Conversely, suppose S is an interval in $\mathbf{R}$. Then (Exercise 1.29) there exists a surjective continuous function $f : \mathbf{R} \longrightarrow S$, hence S is connected because $\mathbf{R}$ is connected. □

Corollary 2.35 (Intermediate Value Theorem)**.** *Let* $f : X \longrightarrow \mathbf{R}$ *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. □

We say that a topological space X is *totally disconnected* if X has at least two distinct elements and the only connected subsets of X are the singletons.

2.6. Compactness

Let X be a topological space. An *open cover* of X is a collection $\{U_i : i \in I\}$ of open subsets $U_i \subseteq X$ such that

$$X = \bigcup_{i \in I} U_i.$$

We say that X is a *compact* topological space if every open cover $\{U_i : i \in I\}$ has some finite *subcover*, that is there exist $n \in \mathrm{N}$ and $i_1, \ldots, i_n \in I$ such that

$$X = U_{i_1} \cup \cdots \cup U_{i_n}.$$

More generally, given a subset $K \subseteq X$, we say that K is a *compact subset* if K is a compact space with respect to the induced topology. An equivalent formulation is given by

Proposition 2.36. *A subset* K *of a topological space* X *is compact if and only if, for any collection* $\{U_i : i \in I\}$ *of open subsets of* X *such that*

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

there exist $n \in \mathbf{N}$ *and* $i_1, \ldots, i_n \in I$ *such that*

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

Proof. See Exercise 1.52. □

Proposition 2.37. *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 $k \neq x$ we get by Hausdorffness that 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 obtain an open cover

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

which by compactness has some finite subcover

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

Consider

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

which is an open neighbourhood of x. We have $U \cap V = \varnothing$, therefore $V \subseteq X \setminus U \subseteq X \setminus K$ is an open neighbourhood of x contained in $X \setminus K$. By Exercise 1.14, $X \setminus K$ is open. ☐

Proposition 2.38. *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

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

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

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

hence we have found a finite subcover of the original open cover. ☐

Proposition 2.39. *If $f \colon X \longrightarrow Y$ 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

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

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

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

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;
- *proper* if for any compact subset $K \subseteq Y$, the inverse image $f^{-1}(K) \subseteq X$ is compact.

Proposition 2.40. *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

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

□

Theorem 2.41. *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 $\{V_x^1 : x \in X_1\}$ 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 2.40, 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 1.24, hence compact.

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

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 2.42. *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 2.20, 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 1.53, this is an equivalence relation on the set of sequences in (X, d).

Proposition 2.43. *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

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

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 2.44. *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 an *isometry* 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 an isometry must be injective, as well as continuous.

If a bijective isometry $f\colon X \longrightarrow Y$ exists we say that X and Y are *isometric*. (You should check that the inverse of a bijective isometry is also an isometry.)

Whether continuous functions or isometries 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.

2.8. Cauchy sequences

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

Proposition 2.45. *Let (X,d) be a metric space and suppose $(x_n) \longrightarrow x \in X$. Then, given $\varepsilon > 0$, there exists $N \in \mathbf{N}$ such that $d(x_n, x_m) < \varepsilon$ for all $n, m \geqslant N$.*

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

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

A sequence (x_n) that satisfies the conclusion of Proposition 2.45 is said to be *Cauchy*.

A natural question is whether the converse of Proposition 2.45 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 2.46.** (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 1.62.

Proposition 2.47. *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 2.45 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 2.42 we have $x \in \overline{S} = S$ since S is closed. □

Proposition 2.48. *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 $\mathbf{R}$.*

Proof. Inequality chase, see Exercise 1.56. □

The equivalence relation on sequences preserves the Cauchy property:

Proposition 2.49. *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.*

Proof. Another inequality chase, see Exercise 1.57. □

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

> **Example 2.50.** 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 2.51. Any isometry is uniformly continuous. This is immediate from the definitions (can take $\delta = \varepsilon$).

Proposition 2.52. *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 $n, m \geqslant N$. □

Proposition 2.53. *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:

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

which therefore has a finite subcover

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

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

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

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

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

2.9. Completions

Any metric space can be embedded into a complete metric space. To make this precise, we say that a complete metric space $\big(\widehat{X}, \widehat{d}\,\big)$ is a *completion* of a metric space (X, d) if there exists an isometry $\iota\colon X \longrightarrow \widehat{X}$ 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).)

Proposition 2.54. *Let $(\widehat{X}, \widehat{d})$, $\iota\colon X \longrightarrow \widehat{X}$ be a completion of (X,d). Let $\widehat{x}, \widehat{y} \in \widehat{X}$ and let $(x_n), (y_n)$ be sequences in X such that*

$$\widehat{x} = \lim_{n\longrightarrow\infty} \iota(x_n) \qquad \textit{and} \qquad \widehat{y} = \lim_{n\longrightarrow\infty} \iota(y_n).$$

Then

$$\widehat{d}(\widehat{x}, \widehat{y}) = \lim_{n\longrightarrow\infty} d(x_n, y_n).$$

Proof. Consider $\widehat{X} \times \widehat{X}$ with the sup metric, then since $(\iota(x_n)) \longrightarrow \widehat{x}$ and $(\iota(y_n)) \longrightarrow \widehat{y}$, we have $\big(\iota(x_n), \iota(y_n)\big) \longrightarrow (\widehat{x}, \widehat{y}) \in \widehat{X} \times \widehat{X}$, see Tutorial Question 4.9. Since

$$d(x_n, y_n) = \widehat{d}\big(\iota(x_n), \iota(y_n)\big) \qquad \text{for all } n \in \mathbf{N},$$

we have

$$\lim_{n\longrightarrow\infty} d(x_n, y_n) = \lim_{n\longrightarrow\infty} \widehat{d}\big(\iota(x_n), \iota(y_n)\big) = \widehat{d}\Big(\lim_{n\longrightarrow\infty} \big(\iota(x_n), \iota(y_n)\big)\Big) = \widehat{d}(\widehat{x}, \widehat{y}).$$ □

Theorem 2.55. *Any metric space (X,d) has a completion.*

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

Proof. $\widehat{X}$ **as a set:** Given (X,d), consider the set $\mathcal{C}$ of all Cauchy sequences, equipped with the equivalence relation defined above Proposition 2.43. Let $\widehat{X}$ be the resulting set of equivalence classes $[(x_n)]$.

$\widehat{X}$ **as a metric space:** Proposition 2.54 hints that we should 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 2.48) and $\mathbf{R}$ is complete; moreover $\widehat{d}$ is well-defined, see Exercise 1.65. It is easy to see that $\widehat{d}$ is a metric on $\widehat{X}$.

The isometry ι**:** For any $x \in X$, let $\iota(x) = [(x)]$ be the equivalence class of the constant sequence $(x, x, \ldots)$. 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 an isometry.

$\iota(X)$ **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 $d(x, x_n) < \varepsilon/2$ for all $n \geqslant N$, so taking limits:

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

The metric space $\widehat{X}$ **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 2.49 so is the sequence $(\iota(x_n))$ in $\iota(X) \subseteq \widehat{X}$, 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 $m \in \mathbf{N}$.

For any $n \in \mathbf{N}$, 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:

$$\begin{aligned}
\widehat{d}\big(a_n, \widehat{x}\big) &= \lim_{m \longrightarrow \infty} d\big(a_n(m), x_m\big) \\
&\leqslant \lim_{m \longrightarrow \infty} d\big(a_n(m), x_n\big) + \lim_{m \longrightarrow \infty} d(x_n, x_m) \\
&= \widehat{d}\big(a_n, \iota(x_n)\big) + \lim_{m \longrightarrow \infty} d(x_n, x_m) \qquad \text{for all } n \in \mathbf{N}.
\end{aligned}$$

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

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

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

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

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

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

A reasonable first attempt would be to see if any **continuous** function $f\colon X \longrightarrow Y$ extends to a **continuous** function $\widehat{f}\colon \widehat{X} \longrightarrow \widehat{Y}$. It turns out that such a continuous extension may not exist (Exercise 1.66), but when it does, it is unique (this follows from the more general result of Exercise 1.58.)

We are about to see, however, that any **uniformly continuous** function (resp. **isometry**) $f\colon X \longrightarrow Y$ extends uniquely to a **uniformly continuous** function (resp. **isometry**) $\widehat{f}\colon \widehat{X} \longrightarrow \widehat{Y}$.

3. Normed and Hilbert spaces

A. Appendix: Prerequisites

A.1. Equivalence relations

An *equivalence relation* $\sim$ is a way of identifying elements of a set. More precisely, given a set A and $x, y \in A$, we will write $x \sim y$ to signify that "x is equivalent to y", and we ask for this to satisfy three properties:

- $x \sim x$ for all $x \in A$ (*reflexivity*);
- if $x \sim y$ then $y \sim x$ (*symmetry*);
- if $x \sim y$ and $y \sim z$ then $x \sim z$ (*transitivity*).

The following example should be very familiar:

Example A.1. Fix a natural number n. For $k, m \in \mathbf{Z}$, define $k \sim m$ if $m - k$ is divisible by n. Show that this satisfies the properties of an equivalence relation on $\mathbf{Z}$.

Solution.

- Given $k \in \mathbf{Z}$, $k - k = 0$ is divisible by n.
- If $k \sim m$, then $m - k = na$ for some $a \in \mathbf{Z}$, therefore $k - m = -na$, so $m \sim k$.
- If $k \sim m$ and $m \sim \ell$ then $m - k = na$ and $\ell - m = nb$ for some $a, b \in \mathbf{Z}$. Therefore $\ell - k = n(a + b)$ so $k \sim \ell$.

□

Suppose we are given an equivalence relation on a set A. For any element $x \in A$, we define the *equivalence class* of x as:

$$[x] = \{y \in A : x \sim y\}.$$

Proposition A.2. *For any elements $x, z \in A$, their equivalence classes are either identical or disjoint, in other words:*

$$\text{either } [x] = [z] \quad \text{or } [x] \cap [z] = \varnothing.$$

Proof. Let $x, z \in A$. There are two possibilities:

- $x \sim z$: given $y \in [x]$, we have $x \sim y$, so $y \sim x$, so $y \sim z$, so $y \in [z]$. This tells us that $[x] \subseteq [z]$, and the other inclusion follows the same way from $z \sim x$. Therefore $[x] = [z]$.
- $x \not\sim z$: suppose $[x] \cap [z]$ is not empty, and pick some element y in there. Then $y \in [x]$ so $y \sim x$, and $y \in [z]$ so $y \sim z$, implying that $x \sim z$, contradiction. Therefore $[x] \cap [z] = \varnothing$.

□

Example A.3. How many distinct equivalence classes are there for the equivalence relation on $\mathbf{Z}$ defined in Example A.1?

Solution. Given $m \in \mathbf{Z}$, let $0 \leqslant r \leqslant n - 1$ be the remainder of the division of m by n: $m = qn + r$. Then $m - r$ is divisible by n, hence $m \sim r$. From the previous part, we know that there are at most n equivalence classes, one for each possible value of r. To show that we have exactly n buddy groups, we need to prove that $[r_1] \neq [r_2]$ for any $r_1 \neq r_2$ with $0 \leqslant r_1, r_2 \leqslant n-1$. We do this by contradiction: if $[r_1] = [r_2]$ then $r_1 \sim r_2$, so $r_2 - r_1$ is a multiple of n. But $-(n-1) \leqslant r_2 - r_1 \leqslant (n-1)$, and the only multiple of n in that interval is 0, in other words $r_2 = r_1$, contradiction. □

Suppose we are given an equivalence relation on a set A, and consider the set of equivalence classes

$$(A/\sim) := \{[x] : x \in A\}.$$

$A/\sim$ is read as "A mod tilde", and it is referred to as the *quotient* of A by the relation $\sim$. There is a canonical surjective function (the *quotient map*)

$$\pi : A \longrightarrow A/\sim \qquad \text{given by } \pi(x) = [x].$$

Example A.4 (Row equivalence of matrices)**.** Fix natural numbers m, n and consider the set $M_{m\times n}$ of all $m \times n$ matrices with real entries. Given matrices $X, Y \in M_{m\times n}$, define $X \sim Y$ if and only if there is a finite sequence of elementary row operations that starts at X and ends at Y.

Show that this is an equivalence relation.

Solution.

- Let $X \in M_{m\times n}$. The identity elementary row operation takes X to X, so $X \sim X$.
- Let $X, Y \in M_{m\times n}$ and suppose $X \sim Y$, so there is a sequence $\rho_1 \circ \cdots \circ \rho_k$ of elementary row operations that starts at X and ends at Y. Then each ρ_j is invertible and ρ_j^{-1} is an elementary row operation, so the sequence $\rho_k^{-1} \circ \cdots \circ \rho_1^{-1}$ starts at Y and ends at X, so $Y \sim X$.
- Let $X, Y, Z \in M_{m\times n}$ and suppose that $X \sim Y$ and $Y \sim Z$. Composing the two sequences of elementary row operations, we get a finite sequence that starts at X and ends at Z, so $X \sim Z$. □

Example A.5. Describe explicitly the quotient of $M_{2\times 3}$ by the equivalence relation from Example A.4, by listing a representative element for each equivalence class.

Describe as precisely as you can all the matrices that belong to the equivalence class of the matrix $X = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}$.

Show that the function $f : \left(M_{2\times 3}/\sim\right) \longrightarrow \mathbf{N}$ given by $f([X]) = \operatorname{rank}(X)$ is well-defined.

A.2. Cardinality

We say that two sets S and T have *the same cardinality* if there exists a bijective function $f : S \longrightarrow T$.

This defines an equivalence relation (on the set of subsets of any fixed set Ω, see Exercise A.6). In this subject, the natural numbers

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

start at 0.

A set S is *finite* if it is empty or there exists $n \in \mathbf{N}$ such that S has the same cardinality as $\{0, \ldots, n\}$. A set S is *infinite* if it is not finite.

We will use the term *countable* to mean what is more precisely called *countably infinite*, that is, a set that has the same cardinality as $\mathbf{N}$. A set S is *uncountable* if it is infinite and not countable.

Proposition A.6. *Any subset S of a countable set T is either finite or countable.*

The first uncountable set that most people encounter is the set $\mathbf{R}$ of real numbers. It is easy to see that any interval of length > 0 in $\mathbf{R}$ must also be uncountable.

It can be difficult to find a bijective function between two sets (assuming that one exists). The following result makes it easier to show that two sets have the same cardinality. (The proof is nontrivial, and uses the Axiom of Choice.)

Theorem A.7 (Schröder–Bernstein)**.** *If A and B are sets and $f: A \longrightarrow B$ and $g: B \longrightarrow A$ are injective functions, then A and B have the same cardinality.*

A.3. Maps between vector spaces

Unless specified otherwise, we use $\mathbf{F}$ to denote an arbitrary field.

For vector spaces V, W over $\mathbf{F}$, we write

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

This is a vector space over $\mathbf{F}$, with zero vector given by the constant function $\mathbf{0}: V \longrightarrow W$, $\mathbf{0}(v) = 0_W$ for all $v \in V$, and with vector addition and scalar multiplication defined pointwise:

$$(f_1 + f_2)(v) = f_1(v) + f_2(v) \qquad \text{and} \qquad (\lambda f)(v) = \lambda f(v).$$

An ***F**-algebra* is a vector space A over $\mathbf{F}$ together with a multiplication map $A \times A \longrightarrow A$, $(u, v) \longmapsto uv$, satisfying

- $(u + v)w = uw + vw$ for all $u, v, w \in A$;
- $u(v + w) = uv + uw$ for all $u, v, w \in A$;
- $\big(\alpha u\big)\big(\beta v\big) = \big(\alpha\beta\big)\big(uv\big)$ for all $\alpha, \beta \in \mathbf{F}$ and all $u, v \in A$.

The algebra A is *associative* if

$$(uv)w = u(vw) \qquad \text{for all } u, v, w \in A.$$

The algebra A is *unital* if there exists an element $\mathbf{1} \in A$ with the property that

$$\mathbf{1}v = v\mathbf{1} = v \qquad \text{for all } v \in A.$$

For any vector space V over $\mathbf{F}$, $\mathrm{End}(V) := \mathrm{Hom}(V, V)$ is an associative unital $\mathbf{F}$-algebra, see Exercise A.10.

An important property of a basis for a vector space is the ability to define a function on that basis and then extend it to a unique linear map. More precisely, let V and W be vector spaces over $\mathbf{F}$. Fix a basis B of V. For any function $g: B \longrightarrow W$ there exists a unique linear map $f: V \longrightarrow W$ such that $g = f|_B$, constructed in the following manner:

Given $v \in V$, there is a unique expression of the form

$$v = a_1 v_1 + \cdots + a_n v_n, \qquad n \in \mathbf{N}, a_j \in \mathbf{F}, v_j \in B.$$

Therefore the only option is to set

$$f(v) = a_1 g(v_1) + \cdots + a_n g(v_n).$$

It is easy to see that f is linear.

We say that f is obtained from g by *extending by linearity*.

A.4. Inner products

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

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

An *inner product* on V is a function

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

such that

(a) $\langle w, v \rangle = \overline{\langle v, w \rangle}$ for all $v, w \in V$;

(b) $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ for all $u, v, w \in V$;

(c) $\langle \alpha v, w \rangle = \alpha \langle v, w \rangle$ for all $v, w \in V$, all $\alpha \in \mathbf{F}$;

(d) $\langle v, v \rangle \geqslant 0$ for all $v \in V$ and $\langle v, v \rangle = 0$ iff $v = 0$.

Properties (a), (b), and (c) say that $\langle \cdot, \cdot \rangle$ is linear in the first variable, but *conjugate-linear* in the second:

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

(Such a function $V \times V \longrightarrow \mathbf{F}$ is called a *sesquilinear form*.)

Property (d) says that $\langle \cdot, \cdot \rangle$ is *positive-definite*.

An *inner product space* is a pair $\big(V, \langle \cdot, \cdot \rangle\big)$, where V is a vector space over $\mathbf{F}$ and $\langle \cdot, \cdot \rangle$ is an inner product on V.

Example A.8. The prototypical inner product on $\mathbf{C}^n$ is

$$\langle u, v \rangle = \sum_{k=1}^{n} u_k \overline{v}_k = \overline{v}^{\mathsf{T}} u,$$

which on $\mathbf{R}^n$ becomes

$$\langle u, v \rangle = \sum_{k=1}^{n} u_k v_k = v^{\mathsf{T}} u.$$

All other inner products on $\mathbf{C}^n$ are of the form

$$\langle u, v \rangle = \overline{v}^{\mathsf{T}} A u,$$

where A is an $n \times n$ *positive-definite Hermitian matrix*, that is

$$\overline{A}^{\mathsf{T}} = A \qquad \text{and all the eigenvalues of } A \text{ are real and positive.}$$

Over $\mathbf{R}$, A is a positive-definite[1] symmetric matrix.

Define

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

Proposition A.9 (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, that is $u = \lambda v$ for some $\lambda \in \mathbf{F}$.

[1] There is a slightly weaker notion of *positive-semidefinite matrix* A, where we ask for the eigenvalues to be real and non-negative. Since we are then allowing 0 to be an eigenvalue, such a matrix may not define an inner product, because there could be nonzero vectors with length zero.

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 $|\langle u, v\rangle|^2 \leqslant \|u\|^2\,\|v\|^2$.

Equality holds if and only if $0 = \langle u - tv, u - tv\rangle$ if and only if $u - tv = \mathbf{0}$ if and only if $u = tv$. □

Let V be a finite-dimensional inner product space. A linear map $f\colon V \longrightarrow V$ is *self-adjoint* if

$$\langle f(u), v\rangle = \langle u, f(v)\rangle \qquad \text{for all } u, v \in V.$$

A set of vectors $S \subseteq V$ is said to be *orthonormal* if

$$\langle u, v\rangle = \begin{cases} 1 & \text{if } u = v \\ 0 & \text{if } u \neq v \end{cases}$$

for all $u, v \in S$.

Theorem A.10 (Spectral Theorem, finite-dimensional case)**.** *Let $f\colon V \longrightarrow V$ be a self-adjoint linear map on a finite-dimensional inner product space V over $\mathbf{F}$. There exists an orthonormal basis of V made of eigenvectors for f.*

In practice, a linear map $f\colon V \longrightarrow V$ is often given by a matrix (representation) M.

- If $\mathbf{F} = \mathbf{R}$, f is self-adjoint if and only if M is real *symmetric* ($M^{\mathsf{T}} = M$), and then the Spectral Theorem implies that M is *orthogonally diagonalisable*: there exists a diagonal matrix D with real entries and a real *orthogonal* matrix Q (that is, $QQ^{\mathsf{T}} = I$) such that $Q^{\mathsf{T}}MQ = D$.

- If $\mathbf{F} = \mathbf{C}$, f is self-adjoint if and only if M is *Hermitian* ($\overline{M}^{\mathsf{T}} = M$), and then the Spectral Theorem implies that M is *unitarily (real-)diagonalisable*: there exists a diagonal matrix D with real entries and a *unitary* matrix U (that is, $U\overline{U}^{\mathsf{T}} = I$) such that $\overline{U}^{\mathsf{T}}MU = D$.

In both cases, D stores the eigenvalues of M and Q or U store normalised eigenvectors of M (so all of these are obtained by computing first the eigenvalues, then bases for the eigenspaces, then orthonormalising the bases using Gram–Schmidt).

A.5. Uniform continuity and uniform convergence

Let $f\colon X \longrightarrow \mathbf{R}$ be a function, with domain $X \subseteq \mathbf{R}$.

The typical first definition of continuity amounts to: f is *continuous on X* if and only if

$$\text{for every } x \in X \text{ and every } \varepsilon > 0, \text{ there exists } \delta > 0 \text{ such that:}$$
$$\text{for all } y \in X, \text{ if } |x - y| < \delta, \text{ then } |f(x) - f(y)| < \varepsilon.$$

The order of appearance of the variables matters! In particular, since δ appears after both x and ε, it may well depend on both of these.

For various purposes, a stronger notion of continuity is needed. We say that f is *uniformly continuous on X* if and only if

$$\text{for every } \varepsilon > 0, \text{ there exists } \delta > 0 \text{ such that:} \\ \text{for all } x, y \in X, \text{ if } |x - y| < \delta, \text{ then } |f(x) - f(y)| < \varepsilon.$$

In this version, δ only depends on ε (and thus its choice is **uniform over** $x \in X$, hence the name).

Example A.11. The function $f\colon \mathbf{R}_{>0} \longrightarrow \mathbf{R}_{>0}$ given by $f(x) = \frac{1}{x}$ is not uniformly continuous.

Solution. First make sure that you negate the condition in the definition correctly: there exists $\varepsilon > 0$ such that for all $\delta > 0$ there exist x, y such that $|x-y| < \delta$ and $|f(x)-f(y)| \geqslant \varepsilon$.

And now, to work: let $\varepsilon = 1$. Take an arbitrary $\delta > 0$. Set $x = \min\{\delta, 1\}$. I claim that $y := x/2$ satisfies the desired condition. Let's check:

$$|x - y| = \frac{x}{2} \leqslant \frac{\delta}{2} < \delta.$$

Also

$$|f(x) - f(y)| = \left|\frac{1}{x} - \frac{1}{y}\right| = \left|\frac{1}{x} - \frac{2}{x}\right| = \frac{1}{x} \geqslant 1 = \varepsilon. \qquad \square$$

One source of uniformly continuous functions is given by the fact that if X is a closed, bounded subset of $\mathbf{R}$, then any continuous function $f\colon X \longrightarrow \mathbf{R}$ is uniformly continuous on X. We will prove a more general result in the context of metric spaces.

There is a similar pair of the type (more general notion, stronger notion) in the context of sequences of functions. Suppose we have, for each $n \in \mathbf{N}$, a function $f_n\colon X \longrightarrow \mathbf{R}$ with domain $X \subseteq \mathbf{R}$. Suppose we also have a "target" function $f\colon X \longrightarrow \mathbf{R}$. We say that the sequence (f_n) *converges pointwise to f on X* if and only if

$$\text{for every } x \in X \text{ and every } \varepsilon > 0, \text{ there exists } N \in \mathbf{N} \text{ such that:} \\ \text{if } n \geqslant N, \text{ then } |f_n(x) - f(x)| < \varepsilon.$$

Note that N may well depend on both ε and x.

On the other hand, we say that the sequence (f_n) *converges uniformly to f on X* if and only if

$$\text{for every } \varepsilon > 0, \text{ there exists } N \in \mathbf{N} \text{ such that:} \\ \text{for every } x \in X, \text{ if } n \geqslant N, \text{ then } |f_n(x) - f(x)| < \varepsilon.$$

In this case N depends only on ε (and thus is **uniform over** $x \in X$).

Example A.12. For $n \geqslant 1$, consider $f_n\colon \mathbf{R} \longrightarrow \mathbf{R}$ given by

$$f_n(x) = \frac{1}{n(1 + x^2)}.$$

The sequence (f_n) converges uniformly on $\mathbf{R}$ to the constant function zero.

Solution. The key point is to note that $1 + x^2 \geqslant 1$ for all $x \in \mathbf{R}$, hence $0 \leqslant 1/(1 + x^2) \leqslant 1$ for all $x \in \mathbf{R}$. Therefore, given $\varepsilon > 0$, we let $N \in \mathbf{N}$ satisfy $N > 1/\varepsilon$ (independent of x) and get, for all $n \geqslant N$:

$$|f_n(x) - 0| = \left|\frac{1}{n(1 + x^2)}\right| \leqslant \frac{1}{n} \leqslant \frac{1}{N} < \varepsilon.$$

□

Example A.13. For $n \geqslant 1$, consider $f_n : \mathbf{R} \longrightarrow \mathbf{R}$ given by

$$f_n(x) = \frac{x^2 + nx}{n}.$$

The sequence (f_n) converges pointwise, but not uniformly on $\mathbf{R}$ to the function

$$f : \mathbf{R} \longrightarrow \mathbf{R}, \qquad f(x) = x.$$

Solution. We have

$$|f_n(x) - f(x)| = \left|\frac{x^2 + nx}{n} - x\right| = \frac{x^2}{n}.$$

For a fixed $x \in \mathbf{R}$, we can take $N > x^2/\varepsilon$ to get pointwise convergence at x. But to do so uniformly over $x \in \mathbf{R}$ we would need N to satisfy $N > x^2/\varepsilon$ for all $x \in \mathbf{R}$, which is clearly impossible. □

We will discuss uniform convergence at length and in greater generality. Its main attraction over pointwise convergence is that a uniform limit function retains many nice properties of the functions in the sequence (continuity, boundedness, and so on).

B. Appendix: Miscellaneous

B.1. Zorn's Lemma

A **p**artially **o**rdered **set** (*poset* for short) is a set X together with a *partial order* $\leqslant$, that is a relation satisfying

- $x \leqslant x$ for all $x \in X$;
- if $x \leqslant y$ and $y \leqslant x$ then $x = y$;
- if $x \leqslant y$ and $y \leqslant z$ then $x \leqslant z$.

A poset X such that for any $x, y \in X$ we have $x \leqslant y$ or $y \leqslant x$ is called a *totally ordered set*, and $\leqslant$ is called a *total order*.

A *chain* in a poset $(X, \leqslant)$ is a subset $C \subseteq X$ that is totally ordered with respect to $\leqslant$.

If $S \subseteq X$ is a subset of a poset, then an *upper bound* for S is an element $u \in X$ such that $s \leqslant u$ for all $s \in S$.

A *maximal element* of a poset X is an element m of X such that there does not exist any $x \in X$ such that $x \neq m$ and $m \leqslant x$. In other words, for any $x \in X$, either $x = m$, or $x \leqslant m$, or x and m are not comparable with respect to the partial order $\leqslant$.

The following result is used to deduce the existence of maximal elements in infinite posets:

Lemma B.1 (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.*

B.2. Bilinear maps

If U, V, W are vector spaces over $\mathbf{F}$, a *bilinear map* $\beta\colon U \times V \longrightarrow W$ is a function such that

$$\begin{aligned} \beta(au_1 + bu_2, v) &= a\beta(u_1, v) + b\beta(u_2, v) \\ \beta(u, av_1 + bv_2) &= a\beta(u, v_1) + b\beta(u, v_2) \end{aligned}$$

for all $u, u_1, u_2 \in U$, $v, v_1, v_2 \in V$, $a, b \in \mathbf{F}$.

Note that such β induces maps

$$\begin{aligned} \beta_U\colon U &\longrightarrow \operatorname{Hom}(V, W), \qquad u \longmapsto \big(v \longmapsto \beta(u, v)\big) \\ \beta_V\colon V &\longrightarrow \operatorname{Hom}(U, W), \qquad v \longmapsto \big(u \longmapsto \beta(u, v)\big). \end{aligned}$$

It is easy to check that these maps are themselves linear.

B.3. 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 B.2. *Suppose $B = \{v_1, \dots, v_n\}$ is a basis for V. Define $v_1^\vee, \dots, v_n^\vee \in \mathrm{Fun}(V, \mathbf{F})$ by*

$$v_i^\vee(a_1 v_1 + \cdots + a_n v_n) = a_i \qquad \text{for } i = 1, \dots, n.$$

Then $v_i^\vee \in V^\vee$ for $i = 1, \dots, n$ and the set $B^\vee = \{v_1^\vee, \dots, v_n^\vee\}$ is a basis for $V^\vee$. (It is called the dual basis to B.)

Proof. We check that $v_i^\vee$ is a linear transformation.

Given $v, w \in V$, we express them in the basis B:

$$\begin{aligned} v &= a_1 v_1 + \cdots + a_n v_n \\ w &= b_1 v_1 + \cdots + b_n v_n, \end{aligned}$$

then

$$v_i^\vee(v + w) = v_i^\vee(a_1 v_1 + \cdots + a_n v_n + b_1 v_1 + \cdots + b_n v_n) = a_i + b_i = v_i^\vee(v) + v_i^\vee(w).$$

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

$$v_i^\vee(\lambda v) = v_i^\vee(\lambda a_1 v_1 + \cdots + \lambda a_n v_n) = \lambda a_i = \lambda v_i^\vee(v).$$

So $v_i^\vee \in V^\vee$ for any $i = 1, \dots, n$.

Next we show that the set $B^\vee$ is linearly independent. Suppose we have

$$\lambda_1 v_1^\vee + \cdots + \lambda_n v_n^\vee = 0.$$

In particular, we can apply this to the basis vector $v_i \in B$ for any $i = 1, \dots, n$ and get

$$\lambda_i = 0.$$

So all the coefficients in the above linear relation must be zero, therefore $B^\vee$ is linearly independent.

Finally, we show that the set $B^\vee$ spans $V^\vee$. Let $\varphi \in V^\vee$; let $v \in V$ and express v in the basis B:

$$v = a_1 v_1 + \cdots + a_n v_n.$$

Then, since φ is a linear transformation, we have

$$\begin{aligned} \varphi(v) &= a_1 \varphi(v_1) + \cdots + a_n \varphi(v_n) \\ &= \lambda_1 v_1^\vee(v) + \cdots + \lambda_n v_n^\vee(v), \end{aligned}$$

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

Note that a bilinear map $\beta\colon V \times W \longrightarrow \mathbf{F}$ induces linear maps

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

For instance, we can take $W = V^\vee$ and define $\beta\colon V \times V^\vee \longrightarrow \mathbf{F}$ by

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

The corresponding linear maps are $\beta_{V^\vee} = \mathrm{id}_{V^\vee}\colon V^\vee \longrightarrow V^\vee$, and $\beta_V\colon V \longrightarrow \big(V^\vee\big)^\vee$ given by

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

Proposition B.3. *If V is finite-dimensional, then $\beta_V\colon V \longrightarrow \big(V^\vee\big)^\vee$ 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 B.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

$$u = a_1 v_1 + \cdots + a_n v_n$$
$$v = b_1 v_1 + \cdots + b_n v_n$$

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 B.2 tells us that V and $V^\vee$, and therefore also $(V^\vee)^\vee$, 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

$$v = T(v_1^\vee)v_1 + \cdots + T(v_n^\vee)v_n.$$

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

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

where we expressed φ in terms of the dual basis $v_1^\vee, \dots, v_n^\vee$ from Proposition B.2. □

Proposition B.4. *Consider a linear transformation $T\colon V \longrightarrow W$, where W is another finite-dimensional vector space over $\mathbf{F}$. Define $T^\vee\colon W^\vee \longrightarrow V^\vee$ by*

$$T^\vee(\varphi) = \varphi \circ T.$$

Then $T^\vee$ is a linear transformation, called the dual linear transformation to T.

Proof. It is clear that $\varphi \circ T\colon V \longrightarrow \mathbf{F}$ is linear, being the composition of two linear transformations.

To show that $T^\vee\colon W^\vee \longrightarrow V^\vee$ is linear, take $\varphi_1, \varphi_2 \in W^\vee$. For any $v \in V$ we have

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

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

$$T^\vee(\lambda\varphi)(v) = (\lambda\varphi)\big(T(v)\big) = \lambda\varphi\big(T(v)\big) = \lambda T^\vee(\varphi)(v).$$ □

B.4. 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 B.5. Any group G endowed with the discrete topology (or with the trivial topology) is a topological group.

Example B.6. 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 B.7 (The circle group)**.** Let

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

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

Example B.8 (The general linear groups)**.** Let $n \in \mathbf{Z}_{\geqslant 1}$ and

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

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 B.9. *Let G be a topological group and $g \in G$. The left translation map $L_g : G \longrightarrow G$ 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 B.10. *Any topological group G is a homogeneous topological space, that is: for every $x, y \in G$ there exists a homeomorphism $f : G \longrightarrow G$ such that $f(x) = y$.*

Proof. Let $f = L_{yx^{-1}}$. □

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

Example B.11. 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 $c_g : G \longrightarrow G$, $c_g(x) = g^{-1}xg$ 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 B.12. The map $\exp : \mathbf{R} \longrightarrow \mathbf{R}^\times$ 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 B.13. The determinant map $\det\colon \mathrm{GL}_n(\mathbf{R}) \longrightarrow \mathbf{R}^\times$ is a topological group homomorphism.

Proposition B.14. *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 2.20: 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.

Bibliography

Index