Metric and Hilbert spaces
An invitation to functional analysis

Alexandru Ghitza*
School of Mathematics and Statistics
University of Melbourne

Version of Mon 25^{th} Sept, 2023 at 09:51

*(aghitza@alum.mit.edu)

Contents

1. Introduction

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

The discussion in this section is heavily inspired by the lecture notes [4] by Karen Smith.

Despite the inevitable ups and downs, linear algebra as seen in a first-year subject is very satisfying. There is one fundamental construct (the linear combination, built out of the two operations defining the vector space structure) that gives rise to all the other abstract concepts (linear transformation, subspace, span, linear independence, etc.). And one of these abstract concepts (the basis) allows us to identify even the most ill-conceived of vector spaces with one of the friendly standard spaces $\mathbb{F}^n$, 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 ill-conceived 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 $\mathbb{F}$ be a field and V a vector space over $\mathbb{F}$. As you know, a *linear combination* is a **finite** expression of the form

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

If S is a subset of V, the *span* of S is the subspace of V consisting of all possible linear combinations of elements of S:

$$\operatorname{Span}(S) = \left\{ a_1 v_1 + \cdots + a_n v_n \colon n \in \mathbb{N},\, a_1, \ldots, a_n \in \mathbb{F},\, v_1, \ldots, v_n \in S \right\}.$$

The subset S is *linearly independent* if, for any $n \in \mathbb{N}$, and any $v_1, \ldots, v_n \in S$, the equality

$$a_1 v_1 + \cdots + a_n v_n = 0 \qquad \text{with } a_1, \ldots, a_n \in \mathbb{F}$$

only holds for $a_1 = \cdots = a_n = 0$.

Finally, a subset B of V is a *basis* if $\operatorname{Span}(B) = V$ and B is linearly independent, which one easily shows is equivalent to: 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 $\mathbb{R}[x]$ (called $\mathcal{P}(\mathbb{R})$ in linear algebra) has basis $B = \left\{1, x, x^2, \ldots\right\}$.

[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.

Solution. The fact that B spans and is linearly independent is really just a restatement of the definition of polynomial.

Suppose there exists $n \in \mathbb{N}$, $a_1, \dots, a_n \in \mathbb{R}$ and $k_1, \dots, k_n \in \mathbb{Z}_{\geqslant 0}$ such that

$$a_1 x^{k_1} + \cdots + a_n x^{k_n} = 0.$$

This is an equality of polynomials (with the constant zero polynomial on the right hand side), so by definition it forces the coefficients of same degree to be equal, in other words $a_1 = \cdots = a_n = 0$. So B is linearly independent.

By definition, any polynomial in $\mathbb{R}[x]$ is of the form

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

which is in the span of B. □

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

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

See Exercise 1.4 for more details.

For another example, take $V = \mathbb{R}$ viewed as a vector space over $\mathbb{Q}$. One can show that the set $S = \left\{\sqrt{n} \colon n \in \mathbb{N} \text{ squarefree}\right\}$ is $\mathbb{Q}$-linearly independent in $\mathbb{R}$, but not a basis. The same is true of the set $T = \left\{\pi^n \colon n \in \mathbb{N}\right\}$. (See Exercise 1.5.)

In fact, $\mathbb{R}$ has no countable basis over $\mathbb{Q}$. (See Exercise 1.6.) It's a sign that it may be rather difficult to write down an explicit $\mathbb{Q}$-basis of $\mathbb{R}$.

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

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

The proof of this theorem requires the (in)famous

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

Say what?

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 $\leqslant$.

Here's a good example to keep in mind:

Example 1.4. Fix a set Ω and let X be the set of all subsets of Ω. Then $\subseteq$ is a partial order on X. It is not a total order if Ω has at least two distinct elements.

Solution. The fact that $\subseteq$ is a partial order follows directly from known properties of set inclusion.

If Ω has at least two distinct elements x_1 and x_2, then $\{x_1\}$ and $\{x_2\}$ are not comparable under $\subseteq$, so the latter is not a total order. □

The point of Zorn's Lemma is in dealing with infinite posets, because any nonempty finite poset automatically has a maximal element. (See Exercise 1.7.)

Back to

Proof of Theorem 1.2. If $V = \{0\}$, then $\varnothing$ is vacuously a (in fact, the only) basis of V.

Suppose $V \neq \{0\}$. If $v \in V \setminus \{0\}$, then $\{v\}$ is a linearly independent subset of V. Let X be the set of all linearly independent subsets of V, then X is nonempty. We consider the partial order $\subseteq$ on X given by inclusion of subsets.

Let C be a nonempty chain in X and define

$$U = \bigcup_{S \in C} S,$$

then clearly $S \subseteq U$ for all $S \in C$, so we'll know that U is an upper bound for C as soon as we can show that it is linearly independent (so that $U \in X$).

Suppose there exist $n \in \mathbb{N}$, $a_1, \ldots, a_n \in \mathbb{F}$, and $u_1, \ldots, u_n \in U$ such that

$$a_1 u_1 + \cdots + a_n u_n = 0. \tag{1.1}$$

Let $J = \{1, \ldots, n\}$. For each $j \in J$, there exists $S_j \in C$ such that $u_j \in S_j$. As C is totally ordered, there exists $i \in J$ such that $S_j \subseteq S_i$ for all $j \in J$. But this means that $u_1, \ldots, u_n \in S_i$, so that the linear relation of Equation (1.1) takes place in the linearly independent set S_i. Therefore $a_1 = \cdots = a_n = 0$.

We conclude that X satisfies the conditions of Zorn's Lemma, hence it has a maximal element B. I claim that B spans V, so that it is a basis of V.

We prove this last claim by contradiction: if $v \in V \setminus \operatorname{Span}(B)$, then $B' := B \cup \{v\}$ is linearly independent, hence an element of X. But $B \subseteq B'$ and $B \neq B'$, contradicting the maximality of B. □

I leave it to you (if you are so inclined) to read more about Zorn's Lemma, including the fact that it is equivalent to the Axiom of Choice. (See Section 1.3 for my philosophy regarding the latter.)

For now let's celebrate the fact that we have bases for all vector spaces... but decry the fact that the proof gives us absolutely no handle on what a basis looks like or how to compute one explicitly. This severely reduces the usefulness of the notion of a basis for an infinite-dimensional vector space.

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

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

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

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

The first thing that we will do is axiomatise the essential properties of the Euclidean distance function. We do this on arbitrary sets and obtain the notion of a **metric space**, and see that a surprising amount of results from real analysis carry through to this much more general setting. (Sometimes with different, but typically more conceptual, proofs.)

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

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

$$\langle f, g\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f g \, dx.$$

(The normalising factor is just a matter of convenience, and to some extent of convention.)

The distance function is of course

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

This allows us to bring rigorous meaning to expressions such as

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

In our setting, we have

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

all of them elements of $V = \mathrm{Cts}\big([-\pi,\pi],\mathbb{R}\big)$, and the claim is that $d(f, s_N) \longrightarrow 0$ as $N \longrightarrow \infty$.

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

$$\frac{1}{\sqrt{2}}, \quad \sin(nx) \text{ for } n \in \mathbb{N}, \quad \cos(nx) \text{ for } n \in \mathbb{N},$$

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

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

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

Example 1.5. Consider the sequence

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

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

$$d(f,g) = \sqrt{\frac{1}{\pi}\int_{-\pi}^{\pi} (f-g)^2(x)\,dx}.$$

Show that the function

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

is the pointwise limit of the sequence (f_n) (that is, $f_n(x) \longrightarrow f(x)$ as $n \longrightarrow \infty$), but that $f \notin V$, so V is not complete.

Solution. TODO □

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

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

exists and is finite.

Example 1.6. Check that the function defined in Example 1.5

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

defines an element of $L^2\big([-\pi,\pi]\big)$.

Solution. TODO □

1.2. Exercises

First, some exercises on countability/uncountability. See Section 1.3 for clarification on our use of the term "countable". You may assume without proof that any subset of a countable set is finite or countable.

Exercise 1.1. Let $f\colon X \longrightarrow Y$ be a function, with X a countable set. Then $\operatorname{im}(f)$ is finite or countable.

[*Hint*: Reduce to the case $f\colon \mathbb{N} \longrightarrow Y$ is surjective; construct a right inverse $g\colon Y \longrightarrow \mathbb{N}$, which has to be injective, of f.]

Solution. Without loss of generality, we may assume that f is surjective and we want to show that Y is finite or countable.

Also without loss of generality (by pre-composing f with any bijection $\mathbb{N} \longrightarrow X$), we may assume that $f\colon \mathbb{N} \longrightarrow Y$ is surjective.

As $f\colon \mathbb{N} \longrightarrow Y$ is surjective, there exists a right inverse $g\colon Y \longrightarrow \mathbb{N}$, in other words $f \circ g\colon Y \longrightarrow Y$ is the identity function id_Y: given $y \in Y$, the pre-image $f^{-1}(y) \subseteq \mathbb{N}$ is nonempty, so it has a smallest element n_y; we let $g(y) = n_y$. For any $y \in Y$, we have $f(g(y)) = f(n_y) = y$ as $n_y \in f^{-1}(y)$. So $f \circ g = \mathrm{id}_Y$.

In particular, this forces $g\colon Y \longrightarrow \mathbb{N}$ to be injective, hence realising Y as a subset of the countable set $\mathbb{N}$. We conclude that Y is finite or countable. $\square$

Exercise 1.2. Show that the union S of any countable collection of countable sets is a countable set.

[*Hint*: Construct a surjective function $\mathbb{N} \times \mathbb{N} \longrightarrow S$.]

Solution. Write

$$S = \bigcup_{n \in \mathbb{N}} S_n,$$

with each S_n a countable set. It is clear that S is infinite (as, say, S_1 is, and $S_1 \subseteq S$).

For each $n \in \mathbb{N}$, fix a bijection $\varphi_n\colon \mathbb{N} \longrightarrow S_n$. (As Chengjing rightfully points out to me, this uses the Axiom of Countable Choice.) Define a function $\psi\colon \mathbb{N} \times \mathbb{N} \longrightarrow S$ by:

$$\psi\big((n, m)\big) = \varphi_n(m) \in S_n \subseteq S.$$

This is surjective, and $\mathbb{N} \times \mathbb{N}$ is countable, so S is finite or countable, and we ruled out finite above. $\square$

Exercise 1.3. Modify the proof of Theorem 1.2 to prove that any inner product space V has a maximal orthonormal set. (Note: "maximal" is used here in the sense of the poset of subsets of V under inclusion.)

Solution. TODO $\square$

Exercise 1.4. [tut02] Let $\mathbb{R}^\infty$ be the set of arbitrary sequences $(x_1, x_2, \dots)$ of elements of $\mathbb{R}$.

This is a vector space under the naturally-defined addition of sequences and multiplication by a scalar.

Let $e_j \in \mathbb{R}^\infty$ be the sequence whose j-th entry is 1, and all the others are 0. Describe the subspace $\operatorname{Span}\{e_1, e_2, \dots\}$ of $\mathbb{R}^\infty$. Is the set $\{e_1, e_2, \dots\}$ a basis of $\mathbb{R}^\infty$?

Solution. Let $S = \{e_1, e_2, \dots\}$ and $W = \operatorname{Span}(S)$.

For each $n \in \mathbb{N}$, define

$$W_n = \operatorname{Span}\{e_1, e_2, \dots, e_n\} \subseteq W.$$

I claim that

$$W = \bigcup_{n\in\mathbb{N}} W_n.$$

One inclusion is clear, as $W_n \subseteq W$ for all $n \in \mathbb{N}$.

For the other inclusion, let $w \in W$. Then there exist $m \in \mathbb{N}$, $a_1, \dots, a_m \in \mathbb{R}$ and $k_1, \dots, k_m \in \mathbb{N}$ such that

$$w = a_1 e_{k_1} + \cdots + a_m e_{k_m}.$$

Set $n = \max\{k_1, \dots, k_m\}$, then $w \in W_n$.

Is $W = \mathbb{R}^\infty$? No. Any $w \in W$ appears in a W_n for some $n \in \mathbb{N}$, therefore only the first n entries of w can be nonzero. This means, for instance, that $v = (1, 1, 1, \dots) \notin W$. So S does not span $\mathbb{R}^\infty$. □

Exercise 1.5. [tut02] Let $V = \mathbb{R}$ viewed as a vector space over $\mathbb{Q}$.

Let $\alpha \in \mathbb{R}$. Show that the set $T = \{\alpha^n : n \in \mathbb{N}\}$ is $\mathbb{Q}$-linearly independent if and only if α is transcendental.

(Note: An element $\alpha \in \mathbb{R}$ is called algebraic if there exists a monic polynomial $f \in \mathbb{Q}[x]$ such that $f(\alpha) = 0$. An element $\alpha \in \mathbb{R}$ is called transcendental if it is not algebraic.)

Solution. This is a straightforward rewriting of the definition of algebraic: α is algebraic if and only if it satisfies a polynomial equation with coefficients in $\mathbb{Q}$, which is equivalent to a nontrivial linear relation between the powers of α, which exists if and only if T is linearly dependent. □

Exercise 1.6. [tut02] Let W be a $\mathbb{Q}$-vector space with a countable basis B. Show that W is a countable set.

[*Hint*: Use Exercise 1.2.]

Conclude that $\mathbb{R}$ does not have a countable basis as a vector space over $\mathbb{Q}$.

Solution. Since B is countable we can enumerate it as $B = \{b_n : n \in \mathbb{N}\}$. For each $n \in \mathbb{N}$, let $W_n = \operatorname{Span}\{b_1, \dots, b_n\}$. Then for each $n \in \mathbb{N}$, W_n is isomorphic (as a $\mathbb{Q}$-vector space) to $\mathbb{Q}^n$, hence W_n is countable. I claim that

$$W = \bigcup_{n\in\mathbb{N}} W_n.$$

One inclusion is obvious, as $W_n \subseteq W$ for all $n \in \mathbb{N}$. For the other direction, let $w \in W = \operatorname{Span}(B)$, so there exist $n \in \mathbb{N}$, $a_1, \dots, a_n \in \mathbb{Q}$ and $k_1, \dots, k_n \in \mathbb{N}$ such that

$$w = a_1 b_{k_1} + \cdots + a_n b_{k_n}.$$

Let $k = \max\{k_1, \dots, k_n\}$, then $w \in W_k$.

So W is a countable union of countable sets, hence countable by Exercise 1.2.

The last claim follows directly from the fact that $\mathbb{R}$ is an uncountable set. □

Exercise 1.7. [tut02] Let $(X, \leqslant)$ be a nonempty finite poset. (This just means that X is a nonempty finite set with a partial order $\leqslant$.) Prove that X has a maximal element.

[*Hint*: You could, for instance, use induction on the number of elements of X.]

Solution. We proceed by induction on n, the cardinality of X.

Base case: if $n = 1$ then $X = \{x\}$ for a single element x. Then trivially x is a maximal element of X.

For the induction step, fix $n \in \mathbb{N}$ and suppose that any poset of cardinality n has a maximal element. Let X be a poset of cardinality $n+1$ and choose an arbitrary element $x \in X$. Let $Y = X \setminus \{x\}$, then Y is a poset of cardinality n so by the induction hypothesis has a maximal element m_Y, and clearly $m_Y \neq x$.

We have two possibilities now:

- If $m_Y \leqslant x$, then x is a maximal element of X. Why? Suppose that x is not maximal in X, so that there exists $z \in X$ such that $z \neq x$ and $x \leqslant z$. Since $z \neq x$, we must have $z \in Y$. If $z = m_Y$, then $z \leqslant x$ and $x \leqslant z$ so $z = x$, contradiction. So $z \neq m_Y$, and $m_Y \leqslant x$ and $x \leqslant z$, so $m_Y \leqslant z$, contradicting the maximality of m_Y in Y.

- Otherwise, (if it is not true that $m_Y \leqslant x$), m_Y is a maximal element of X. Why? Suppose there exists $z \in X$ such that $z \neq m_Y$ and $m_Y \leqslant z$. Since $m_Y \leqslant x$ is not true, we have $z \neq x$, so $z \in Y$, contradicting the maximality of m_Y in Y.

In either case we found a maximal element for X. □

Solution. An alternative approach is to proceed by contradiction: suppose $(X, \leqslant)$ is a nonempty finite poset that does not have a maximal element. Use this to construct an unbounded chain of elements of X, contradicting finiteness. □

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 try to avoid using $S \subset T$ (as it is ambiguous), as well as $S \nsubseteq T$ (not ambiguous, but too easily confused with $S \subsetneq T$).

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

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

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

For better or worse, the natural numbers

$$\mathbb{N} = \{1, 2, 3, \dots\}$$

start at 1. The variant starting at 0 is

$$\mathbb{Z}_{\geqslant 0} = \{0, 1, 2, 3, \dots\}.$$

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

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

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

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

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

Acknowledgements

Thanks go to Thomas Black, Isaac Doosey-Shaw, Jack Gardiner, Leigh Greville, Ethan Husband, Peter Karapalidis, Rose-Maree Locsei, Hai Ou, Kashma Pillay, Guozhen Wu, Corey Zelez, and Chengjing Zhang for corrections and suggestions on these notes.

2. Metric spaces

2.1. Metrics

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

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

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 ($\mathbb{R}^n$ 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

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

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 $\mathbb{R}$ with the Euclidean distance. But there are many other examples, some of which are quite exotic:

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

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

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

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

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

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

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

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

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

and

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

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

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

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

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

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

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

and

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

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

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

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

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

Solution.

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

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

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

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

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

For $v_p(x+y)$, without loss of generality assume $v := v_p(x) \leqslant v_p(y) =: u$. Then

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

where m, n, a, b are not divisible by p. Therefore v_p of the quantity in parentheses is non-negative, and we conclude that $v_p(x+y) \geqslant v = \min\{v_p(x), v_p(y)\}$.

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

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

(iv) We have

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

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

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

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

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

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). $\square$

Or we can construct metric spaces as Cartesian products of other metric spaces. There are many ways of doing this, neither 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 $(X_1, d_{X_1}), \dots, (X_n, d_{X_n})$.

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

Solution. Straightforward. $\square$

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 $(X_1, d_{X_1}), \dots, (X_n, d_{X_n})$.

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

Solution. Straightforward; proving the triangle inequality uses

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

Example 2.5. Take $X_1 = X_2 = \mathbb{R}$ 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 $\mathbb{R}^2$.

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

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

Not every metric has to do with lengths and geometry in an obvious way. The p-adic metric in Example 2.1 is an example of something a little different. For another example, let $n \in \mathbb{N}$, $X = \mathbb{F}_2^n$, and let $d(x, y)$ be the number of indices $i \in \{1, \dots, n\}$ such that $x_i \neq y_i$. Then d is a metric on X; it is called the *Hamming metric*. See Exercise 2.17 for more details.

2.2. Open sets and closed sets

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 $r \in \mathbb{R}_{\geqslant 0}$ of a fixed point $c \in X$: the *open ball* of radius r and centre c is

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

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

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

and a corresponding *sphere*

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

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 $\mathbb{Z}$ (endowed with the metric induced from the Euclidean metric on $\mathbb{R}$).

Solution. In addition to the empty set $\varnothing = \mathbb{B}_0(0)$, we have $\{0\} = \mathbb{B}_1(0)$, and for all $n \in \mathbb{N}$ the set

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

For another intuition-busting example, see Exercise 2.19.
And you won't believe how weird p-adic balls are:

Example 2.7. Fix a prime p and consider the metric space $(\mathbb{Q}, d_p)$ where d_p is the p-adic metric from Example 2.1.

(a) Let $p = 3$ and write down 4 elements of $\mathbb{B}_1(2)$ and 4 elements of $\mathbb{B}_{1/9}(3)$.

(b) Back to general prime p now: show that every triangle is isosceles. In other words, given three points in $\mathbb{Q}$, at least two of the three resulting (p-adic) distances are equal.

(c) Show that every point of an open ball is a centre. In other words, take an open ball $\mathbb{B}_r(c)$ with $r \in \mathbb{R}_{\geqslant 0}$ and $c \in \mathbb{Q}$ and suppose $x \in \mathbb{B}_r(c)$; prove that $\mathbb{B}_r(c) = \mathbb{B}_r(x)$.

(d) Show that given any two open balls, either one of them is contained in the other, or they are completely disjoint.

Solution.

(a) We have

$$\left\{2, 5, -7, \frac{4}{5}\right\} \subseteq \mathbb{B}_1(2)$$
$$\left\{3, 30, -24, \frac{39}{4}\right\} \subseteq \mathbb{B}_{1/9}(3).$$

(b) Recall that in the proof of the triangle inequality for the p-adic metric in Example 2.1, the following stronger result was shown:

$$d_p(x,y) \leqslant \max\{d_p(x,t), d_p(t,y)\}.$$

with equality holding if $d_p(x,t) \neq d_p(t,y)$. But this precisely says that if $d_p(x,t) \neq d_p(t,y)$, then $d_p(x,y)$ has to be equal to the largest of $d_p(x,t)$ and $d_p(t,y)$.

(c) First $x \in \mathbb{B}_r(c)$ iff $c \in \mathbb{B}_r(x)$ (this is true for any metric space). So it suffices to show that $x \in \mathbb{B}_r(c)$ implies $\mathbb{B}_r(x) \subseteq \mathbb{B}_r(c)$. Let $y \in \mathbb{B}_r(x)$, then $d_p(y,x) < r$, so that

$$d_p(y,c) \leqslant \max\left\{d_p(y,x), d_p(x,c)\right\} < r,$$

in other words $y \in \mathbb{B}_r(c)$.

(d) Consider two open balls $\mathbb{B}_r(x)$ and $\mathbb{B}_t(y)$. Without loss of generality $r \leqslant t$. Suppose that the balls are not disjoint and let $z \in \mathbb{B}_r(x) \cap \mathbb{B}_t(y)$. By part (c) this implies that $\mathbb{B}_r(z) = \mathbb{B}_r(x)$ and $\mathbb{B}_t(z) = \mathbb{B}_t(y)$, so that

$$\mathbb{B}_r(x) = \mathbb{B}_r(z) \subseteq \mathbb{B}_t(z) = \mathbb{B}_t(y). \qquad \square$$

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 $r \in \mathbb{R}_{>0}$ such that $\mathbb{B}_r(u) \subseteq U$.

If $x \in U$ and $U \subseteq X$ is an open set, we say that U is an *open neighbourhood* of x.

If $A \subseteq X$, we say that $a \in A$ is an *interior point* of A if there exists $r \in \mathbb{R}_{>0}$ such that $\mathbb{B}_r(a) \subseteq A$. Let A° denote the set of all interior points of A. Then $U \subseteq X$ is an open set if and only if $U^\circ = U$.

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

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

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

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

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

Solution. Let $U = \mathbb{B}_r(x)$. 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 $\mathbb{B}_t(u) \subseteq U$. Let $w \in \mathbb{B}_t(u)$, so that $d(w,u) < t$. Then

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

What happens if we combine open sets using set operations?

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

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

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

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

Intersections are a bit more delicate:

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

Proof. Let $n \in \mathbb{N}$ 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 $\mathbb{B}_{r_i}(u) \subseteq U_i$. Let $r = \min\{r_1, \dots, r_n\}$, then $\mathbb{B}_r(u) \subseteq \mathbb{B}_{r_i}(u) \subseteq U_i$ for each $i = 1, \dots, n$. Therefore $\mathbb{B}_r(u) \subseteq U$. □

Wondering about the necessity of the word "finite" in the statement of the proposition? See Exercise 2.20.

Example 2.13. Let $A \subseteq X$. The set A is open if and only if A is the union of a collection of open balls.

Solution. In one direction, if A is a union of a collection of open balls, then A is open by Example 2.10 and Proposition 2.11.

In the other direction, suppose A is open. Let $a \in A$, then there exists an open ball $\mathbb{B}_{r(a)}(a) \subseteq A$. Then

$$A = \bigcup_{a \in A} \mathbb{B}_{r(a)}(a).$$ □

Example 2.14. Let S be a subset of a metric space (X, d) and consider the induced metric on S. Let $A \subseteq S$. Prove that A is an open set in S if and only if there exists an open set U in X such that $A = U \cap S$.

Solution. Here we are working with two different but related metric spaces: (X, d) and (S, d). To avoid confusion, we denote open balls in (X, d) by $\mathbb{B}_r^X(x)$ and open balls in (S, d) by $\mathbb{B}_r^S(s)$. Of course, as $S \subseteq X$, we have $\mathbb{B}_r^S(s) = \mathbb{B}_r^X(s) \cap S$ for any $s \in S$.

Now, suppose A is open in S; by Example 2.13 we have some indexing set I such that

$$A = \bigcup_{i \in I} \mathbb{B}_{r_i}^S(a_i),$$

with $r_i > 0$ and $a_i \in A$ for all $i \in I$. We can then let

$$U = \bigcup_{i \in I} \mathbb{B}_{r_i}^X(a_i),$$

which by Example 2.13 is an open in X. It is clear that $A = U \cap S$ from the discussion in the first paragraph of the solution.

Conversely, suppose $A = U \cap S$ with U open in X. Let $a \in A$, then $a \in U$ so there exists an open (in X) ball $\mathbb{B}_r^X(a)$ such that $\mathbb{B}_r^X(a) \subseteq U$. Consider $\mathbb{B}_r^S(a) = \mathbb{B}_r^X(a) \cap S \subseteq U \cap S = A$. So every point $a \in A$ is contained in an open (in S) ball, hence A is open in S. □

A subset $C \subseteq X$ is a *closed set* 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.

Example 2.15. Show that the union of any finite collection of closed sets is closed. Show that the intersection of any arbitrary collection of closed sets is closed.

Solution. Let $n \in \mathbb{N}$ and let $C_1, \ldots, C_n$ be closed subsets of X. Let

$$C = \bigcup_{i=1}^{n} C_i,$$

then the complement of C is

$$X \setminus C = X \setminus \left(\bigcup_{i=1}^{n} C_i \right) = \bigcap_{i=1}^{n} (X \setminus C_i).$$

For each $i = 1, \ldots, n$, C_i is closed so $X \setminus C_i$ is open, therefore $X \setminus C$ is the intersection of finitely many open sets, hence is itself open by Proposition 2.12. We conclude that C is closed.

For the second statement, let $\{C_i : i \in I\}$ be a collection of closed subsets of X, indexed by a set I. Let

$$C = \bigcap_{i \in I} C_i,$$

then the complement of C is

$$X \setminus C = X \setminus \left(\bigcap_{i \in I} C_i \right) = \bigcup_{i \in I} (X \setminus C_i).$$

For each $i \in I$, C_i is closed so $X \setminus C_i$ is open, therefore $X \setminus C$ is the union of a collection of open sets, hence is itself open by Proposition 2.11. We conclude that C is closed. □

Let (X, d) be a metric space and A a subset of X. Here are a few properties that a point of X might have with respect to the subset A:

(a) Recall that $a \in A$ is an *interior point* of A if there exists $r \in \mathbb{R}_{>0}$ such that $\mathbb{B}_r(a) \subseteq A$.

(b) $x \in X$ is a *boundary point* of A if, for every $r \in \mathbb{R}_{>0}$, we have

$$\mathbb{B}_r(x) \cap A \neq \varnothing \qquad \text{and} \qquad \mathbb{B}_r(x) \cap (X \setminus A) \neq \varnothing.$$

(c) $a \in A$ is an *isolated point* of A if there exists $r \in \mathbb{R}_{>0}$ such that

$$\mathbb{B}_r(a) \cap A = \{a\}.$$

(d) $x \in X$ is an *accumulation point* (or *limit point*) of A if, for every $r \in \mathbb{R}_{>0}$, there exists $a \in \mathbb{B}_r(x) \cap A$ such that $a \neq x$.

This gives rise to:

- The *interior* of A:

$$A^\circ = \{a \in A : a \text{ is an interior point of } A\}.$$

- The *boundary* of A:

$$\partial A = \{x \in X : x \text{ is a boundary point of } A\}.$$

- The *closure* of A:

$$\overline{A} = \{x \in X : x \text{ is a limit point or an isolated point of } A\}.$$

Note that $A \subseteq \overline{A}$: if $a \in A$ is not a limit point, then there exists $r > 0$ such that $\mathbb{B}_r(a) \cap A = \{a\}$, so a is isolated. Hence $a \in \overline{A}$.

Example 2.16. Prove that $A^\circ = A \smallsetminus \partial A$.

Solution. In one direction, it is clear that $A^\circ \subseteq A$. If $a \in A^\circ$ then there exists $\mathbb{B}_r(a) \subseteq A$ of radius $r > 0$, so $\mathbb{B}_r(a) \cap \big(X \smallsetminus A\big) = \varnothing$, hence $a \notin \partial A$.

In the opposite direction, suppose $a \in A \smallsetminus \partial A$. Then there exists $r \in \mathbb{R}_{>0}$ such that $\mathbb{B}_r(a) \cap A = \varnothing$ or $\mathbb{B}_r(a) \cap \big(X \smallsetminus A\big) = \varnothing$. But the former is impossible as $a \in A$, hence we conclude that the latter must hold, implying that $a \in A^\circ$. □

Example 2.17. Prove that $\partial A = \partial(X \smallsetminus A)$.

Solution. Obvious since the statement in the definition is symmetric in A and $X \smallsetminus A$. □

Example 2.18. Prove that $\overline{A} = A^\circ \cup \partial A$.

Solution. Let $x \in \overline{A}$. Suppose x is isolated and let $r > 0$ be such that $\mathbb{B}_r(x) \cap A = \{x\}$. If $x \notin A^\circ$, then $\mathbb{B}_r(x)$ is not contained in A, so that $\mathbb{B}_r(x) \cap \big(X \smallsetminus A\big) \neq \varnothing$, hence x is a boundary point. We conclude that $x \in A^\circ \cup \partial A$.

Suppose now that x is a limit point of A. If $x \notin A^\circ$, for every $r \in \mathbb{R} > 0$, there exists $a \in \mathbb{B}_r(x) \cap A$ such that $a \neq x$, so $\mathbb{B}_r(x) \cap A \neq \varnothing$. On the other hand, $\mathbb{B}_r(x)$ is not a subset of A, so there exists $b \in \mathbb{B}_r(x)$ such that $b \notin A$, therefore $\mathbb{B}_r(x) \cap \big(X \smallsetminus A\big) \neq \varnothing$. We conclude that x is a boundary point of A. In any case, $x \in A^\circ \cup \partial A$.

For the other inclusion, recall that $A \subseteq \overline{A}$. So it remains to deal with the points of $\partial A \cap \big(X \smallsetminus A\big)$. Suppose $x \notin A$ is a boundary point of A. Then for every $r > 0$, $\mathbb{B}_r(x) \cap A \neq \varnothing$, so there exists $a \in \mathbb{B}_r(x) \cap A$, and $a \neq x$ because $x \notin A$. Hence x is a limit point of A, in particular $x \in \overline{A}$. □

Example 2.19. Prove that C is closed if and only if $\overline{C} = C$.

Solution. We already know that $C \subseteq \overline{C}$.

Suppose C is closed, so that $X \smallsetminus C$ is open. Any isolated point of C is by definition in C. If $x \in X \smallsetminus C$ then there exists $r > 0$ such that $\mathbb{B}_r(x) \subseteq \big(X \smallsetminus C\big)$, so that x is not a limit point. Therefore all limit points of C are also in C, hence $\overline{C} = C$.

Conversely, suppose $\overline{C} = C$, then every limit point of C is an element of C. Therefore if $x \in X \smallsetminus C$, x is not a limit point of C. So there exists $r > 0$ such that $\mathbb{B}_r(x) \cap C$ has no points $\neq x$; but x is also not an element of this intersection, which must therefore be empty, so that $\mathbb{B}_r(x) \subseteq \big(X \smallsetminus C\big)$. □

A subset D of X is *dense* if $X = \overline{D}$. This is an extremely useful concept, as the subset D is sometimes easier to work with, for instance:

Example 2.20. The set of rational numbers $\mathbb{Q}$ is dense in $\mathbb{R}$.

Solution. We'll do this in a slightly handwavy way. We want to show that for every $x \in \mathbb{R}$ we have $x \in \overline{\mathbb{Q}}$. Consider the decimal expansion of an arbitrary real number x:

$$x = m.x_1x_2x_3\ldots$$

where $m \in \mathbb{Z}$ and $x_i \in \{0, \ldots, 9\}$. We want to show that for every $\varepsilon > 0$, there exists $q \in \mathbb{Q}$ such that $q \in \mathbb{B}_\varepsilon(x)$. Given such ε, let $n \in \mathbb{N}$ be such that $10^{-n} < \varepsilon$. Set

$$q = m.x_1\ldots x_n = m + \frac{10^{n-1}x_1 + 10^{n-2}x_2 + \cdots + x_n}{10^n},$$

then $|x - q| \leqslant 10^{-n} < \varepsilon$, as claimed. $\square$

Example 2.21. Prove that the set of irrational numbers $\mathbb{R} \smallsetminus \mathbb{Q}$ is dense in $\mathbb{R}$.

Solution. We have seen in Example 2.20 that for any $x \in \mathbb{R}$ and any $\varepsilon > 0$, there exists $q \in \mathbb{Q}$ with $q \in \mathbb{B}_\varepsilon(x)$. But $\mathbb{B}_\varepsilon(x) = (x - \varepsilon, x + \varepsilon)$ is a (nonempty) open interval in $\mathbb{R}$, hence is uncountable, therefore it must also contain some irrational number η (as the rationals are countable).

(To see that any nonempty open interval (a, b) is uncountable, recall that Cantor's diagonal argument shows that $(0, 1)$ is uncountable, then note that $f\colon (a, b) \longrightarrow (0, 1)$ given by $f(x) = (x - a)/(b - a)$ is a bijection.) $\square$

So we have two disjoint sets, each of which is dense in $\mathbb{R}$. The situation is very different if we ask for the sets to be both dense and open, which we do in Exercise 2.12.

A subset N of X is *nowhere dense* if $\left(\overline{N}\right)^\circ = \varnothing$, in other words $\overline{N}$ contains no nonempty open balls of X. An obvious example is $\mathbb{Z}$ as a nowhere dense subset of $\mathbb{R}$.

2.3. Continuous functions

Let (X, d) be a metric space.

A *sequence* in X is a function $\mathbb{N} \longrightarrow X$, 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 $\varepsilon \in \mathbb{R}_{>0}$ there exists $N \in \mathbb{N}$ such that

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

Limits of sequences with terms in a subset A of X belong to the closure of A:

Example 2.22. Let A be a subset of a metric space (X, d). If (x_n) is a sequence in A that converges to $x \in X$, then $x \in \overline{A}$.

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

In particular, a subset $A \subseteq X$ is closed if and only if for every sequence $(a_n) \longrightarrow x \in X$ with $a_n \in A$, we have $x \in A$.

Solution. If $x \in A$ then clearly $x \in \overline{A}$ and we are done.

So suppose (x_n) converges to $x \notin A$. Then for any $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $x_n \in \mathbb{B}_\varepsilon(x)$ for all $n \geqslant N$. In particular, $x_N \in \mathbb{B}_\varepsilon(x) \cap A$, and $x_N \neq x$ as one is in A and the other is not. Therefore x is a limit point of the set A, in particular $x \in \overline{A}$.

For the converse statement: let $x \in \overline{A} = A^\circ \cup \partial A$ (recall Example 2.18). Given $n \in \mathbb{N}$, consider $\mathbb{B}_{1/n}(x) \cap A$. Either $x \in A^\circ$ or $x \in \partial A$, in both cases $\mathbb{B}_{1/n}(x) \cap A \neq \varnothing$, so let x_n be some element in the intersection.

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

Example 2.23. Let $U \subseteq X$ be an open subset of a metric space X and let (x_n) be a sequence in X that converges to $x \in U$. Then there exists $N \in \mathbb{N}$ 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.)

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

Example 2.24. You may encounter "multivariable sequences" such as $\mathbb{N} \times \mathbb{N} \longrightarrow X$, $(m, n) \longmapsto x_{mn}$. You may be tempted, based on your experience with multivariable calculus, to try to deal with this one variable at a time.

However, in general:

$$\lim_{m \longrightarrow \infty} \lim_{n \longrightarrow \infty} x_{mn} \neq \lim_{n \longrightarrow \infty} \lim_{m \longrightarrow \infty} x_{mn}.$$

Convince yourself of this by considering $x_{mn} = \frac{n}{n+m}$.

Solution. We get $\lim_{m \longrightarrow \infty} 1 = 1 \neq 0 = \lim_{n \longrightarrow \infty} 0$. □

Theorem 2.25. *Let X and Y be metric spaces, let $f\colon X \longrightarrow Y$ be a function, let $x \in X$ and $y = f(x) \in Y$. The following are equivalent:*

(a) Given any open neighbourhood $V \subseteq Y$ of y, there exists an open neighbourhood $U \subseteq X$ of x such that $f(U) \subseteq V$.

(b) Given any $\varepsilon \in \mathbb{R}_{>0}$ there exists $\delta \in \mathbb{R}_{>0}$ such that if $x' \in \mathbb{B}_\delta(x)$, then $f(x') \in \mathbb{B}_\varepsilon(y)$.

(c) If (x_n) is a sequence that converges to x in X, then the sequence $\big(f(x_n)\big)$ converges to y in Y.

If the equivalent conditions listed in the theorem hold, we say that f is *continuous at x*.

Proof of Theorem 2.25.

(a) ⇒ (b): Let $\varepsilon \in \mathbb{R}_{>0}$ and set $V := \mathbb{B}_\varepsilon(y)$. This is an open subset of Y, so by (a) we get an open neighbourhood $U \subseteq X$ of x such that $f(U) \subseteq V$. Since U is open, there exists $\delta \in \mathbb{R}_{>0}$ such that $\mathbb{B}_\delta(x) \subseteq U$. But then

$$f\big(\mathbb{B}_\delta(x)\big) \subseteq f(U) \subseteq V = \mathbb{B}_\varepsilon(y),$$

which is precisely what (b) states.

(b) ⇒ (c): Let (x_n) be a sequence converging to x in X.

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

(c) ⇒ (a): Let $V \subseteq Y$ be an open neighbourhood of y.

We prove the existence of U by contradiction: suppose that for every open neighbourhood $U \subseteq X$ of x, there exists $u \in U$ such that $f(u) \notin V$. In particular, this is true for the open balls of radius $1/n$ and centre x, for all $n \in \mathbb{N}$: there exists $x_n \in \mathbb{B}_{1/n}(x)$ such that $f(x_n) \notin V$. We obtain a sequence (x_n) that converges to x in X, hence by part (c), the sequence $\big(f(x_n)\big)$ converges to y in Y. Since $y \in V$ and V is open, this implies (Example 2.23) that $f(x_n) \in V$ for sufficiently large n, contradiction. □

We say that $f: X \longrightarrow Y$ is *continuous* if it is continuous at all $x \in X$.

Example 2.26. The function $f: X \longrightarrow Y$ is continuous if and only if: for any open subset $V \subseteq Y$, the inverse image $f^{-1}(V) \subseteq X$ is an open subset.

Solution. Suppose f is continuous and let $V \subseteq Y$ be an open subset. If $f^{-1}(V)$ is empty, then it is open and we are done. Otherwise take an arbitrary $x \in f^{-1}(V)$ and let $y = f(x) \in V$. As f is continuous at x, there exists an open neighbourhood U of x such that $f(U) \subseteq V$, but this implies that $U \subseteq f^{-1}(V)$. So $f^{-1}(V)$ has the property that every point x has an open neighbourhood contained in $f^{-1}(V)$, therefore $f^{-1}(V)$ is open.

Conversely, suppose that f is such that the inverse image $f^{-1}(V)$ of any open $V \subseteq Y$ is open in X. Let $x \in X$. Let $y = f(x)$ and consider an arbitrary open neighbourhood V of y. Then $f^{-1}(V)$ is open in X, and $x \in f^{-1}(V)$, so certainly there is an open ball centred at x and contained in $f^{-1}(V)$. Therefore f is continuous at x, for all $x \in X$. □

The notion of continuous function is one possible type of morphism that we can consider between metric spaces. The corresponding concept of isomorphism is given by: a continuous function $f: X \longrightarrow Y$ is a *homeomorphism* if f is bijective and $f^{-1}: Y \longrightarrow X$ is continuous.

An important related idea is that of (topological) equivalence of metrics. We have seen that a given set X may have many different metric functions on it. Depending of which features we are focusing on, we may want to identify different metrics. For instance, even though a metric d and its rescaling $\frac{1}{2}d$ are not the same function, it is easy to see that they give rise to the same collection of open sets, closed sets, convergent sequences and their limits, and so on. If we only care about these concepts, rather than the exact distance between points, we may want to treat d and $\frac{1}{2}d$ as equivalent metrics.

To give a precise definition of this notion, we start with a number of logically equivalent ways of comparing two metrics on a set:

Proposition 2.27. *Let X be a set and d_1, d_2 metrics on X. The following are equivalent:*

(a) Every open subset of (X, d_2) is open in (X, d_1).

(b) Every closed subset of (X, d_2) is closed in (X, d_1).

(c) For any $x \in X$, every open ball of (X, d_2) centred at x contains an open ball of (X, d_1) centred at x.

(d) *For any $x \in X$, every sequence that converges to x in (X, d_1) also converges to x in (X, d_2).*

(e) *The function $f\colon (X, d_1) \longrightarrow (X, d_2)$ given by $f(x) = x$ for all $x \in X$ is continuous.*

If any (and therefore all) of the statements in Proposition 2.27 hold, we say that the metric d_1 is *finer* than d_2, and that d_2 is *coarser* than d_1.

We say that d_1 and d_2 are *(topologically) equivalent* if d_1 is both finer and coarser than d_2.

Proof of Proposition 2.27. This is not the most economical way there, but whatever.

(a) ⇔ (b): Let C be closed in (X, d_2), then $X \setminus C$ is open in (X, d_2), so by (a) $X \setminus C$ is open in (X, d_1), hence, C is closed in (X, d_1).

Interchange "closed" and "open" everywhere in the previous sentence to get the other direction.

(a) ⇒ (c): Consider $\mathbb{B}_r^{d_2}(x)$. It is open in (X, d_2), so by (a) it is open in (X, d_1), so there exists $\mathbb{B}_t^{d_1}(x) \subseteq \mathbb{B}_r^{d_2}(x)$, as wanted.

(c) ⇔ (d): Special case of Theorem 2.25, with $f\colon X \longrightarrow X$ the identity function.

(d) ⇔ (e): Special case of Theorem 2.25, with $f\colon X \longrightarrow X$ the identity function.

(a) ⇔ (e): Follows immediately from Example 2.26. □

An important special case concerns the metric space structures on Cartesian products.

If (X, d_X) and (Y, d_Y) are two metric spaces, a metric d on $X \times Y$ is said to be *conserving* if

$$d_\infty\big((x_1, y_1), (x_2, y_2)\big) \leqslant d\big((x_1, y_1), (x_2, y_2)\big) \leqslant d_1\big((x_1, y_1), (x_2, y_2)\big)$$

for all $(x_1, y_1), (x_2, y_2) \in X \times Y$.

Trivially, d_∞ and d_1 are conserving metrics. Slightly less trivial: d_2 is a conserving metric.

Proposition 2.28. *If d is a conserving metric, then the collection of open subsets of $X \times Y$ with respect to d consists precisely of arbitrary unions of sets of the form*

$$U \times V, \quad \text{with } U \text{ open in } X \text{ and } V \text{ open in } Y.$$

In particular, all conserving metrics on $X \times Y$ are equivalent.

Proof. The second statement follows from the first and Proposition 2.27.

For the first statement, we start by showing that $U \times V$ is open with respect to d, if U is open in X and V is open in Y. Consider an arbitrary element $(u, v) \in U \times V$. Since u is open in U, there exists $s > 0$ such that $\mathbb{B}_s(u) \subseteq U$. Similarly, there exists $t > 0$ such that $\mathbb{B}_t(v) \subseteq V$. Let $r = \min\{s, t\} > 0$. I claim that the d-open ball $B \coloneqq \mathbb{B}_r\big((u, v)\big) \subseteq U \times V$. Why? If $(x, y) \in B$ then since d is conserving,

$$\max\big\{d_X(x, u), d_Y(y, v)\big\} = d_\infty\big((x, y), (u, v)\big) \leqslant d\big((x, y), (u, v)\big) < r,$$

so $d_X(x, u) < r \leqslant s$ hence $x \in U$, and $d_Y(y, v) < r \leqslant t$ hence $y \in V$.

We conclude that sets of the form $U \times V$ are open with respect to d. By Proposition 2.11, so are arbitrary unions of such sets.

It remains to prove that any d-open set of $X \times Y$ is of this form. The empty set and $X \times Y$ are clearly of this form, so let $\varnothing \neq W \neq X \times Y$ be open with respect to d. For each $w \in W$, we will exhibit opens $U_w \subseteq X$ and $V_w \subseteq Y$ such that

$$W = \bigcup_{w \in W} U_w \times V_w.$$

Fix $w = (u, v) \in W$. Since W is d-open, there exists $r > 0$ such that $\mathbb{B}_r(w) \subseteq W$. Let U_w be the d_X-open ball $\mathbb{B}_{r/2}(u) \subseteq X$, and let V_w be the d_Y-open ball $\mathbb{B}_{r/2}(v) \subseteq Y$. I claim that $U_w \times V_w \subseteq \mathbb{B}_r(w) \subseteq W$. Why? If $(x, y) \in U_w \times V_w$, since d is conserving,

$$d\big((x, y), (u, v)\big) \leqslant d_X(x, u) + d_Y(y, v) < \frac{r}{2} + \frac{r}{2} = r. \qquad \square$$

There are situations where a stricter notion of morphism between metric spaces is needed. A function $f\colon X \longrightarrow Y$ between metric spaces (X, d_X) and (Y, d_Y) is said to be *distance-preserving* if

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

An *isometry* $f\colon X \longrightarrow Y$ is a bijective distance-preserving function.

Two metric spaces X and Y are said to be *isometric* if there exists an isometry $f\colon X \longrightarrow Y$. This is an equivalence relation on any set of metric spaces.

Example 2.29. Show that the inverse of an isometry $f\colon X \longrightarrow Y$ is distance-preserving.

Solution. Let $y_1, y_2 \in Y$. Set $x_1 = f^{-1}(y_1)$, $x_2 = f^{-1}(y_2)$. Then

$$d_Y(y_1, y_2) = d_Y\big(f(x_1), f(x_2)\big) = d_X(x_1, x_2) = d_X\big(f^{-1}(y_1), f^{-1}(y_2)\big). \qquad \square$$

Example 2.30. Show that any distance-preserving function $f\colon X \longrightarrow Y$ is continuous.
In particular, any isometry is a homeomorphism.

Solution. Let $x \in X$. Given $\varepsilon > 0$, if $x' \in \mathbb{B}_\varepsilon(x)$ then $d_X(x, x') < \varepsilon$, so

$$d_Y\big(f(x), f(x')\big) = d_X(x, x') < \varepsilon,$$

hence $f(x') \in \mathbb{B}_\varepsilon(f(x))$. $\square$

2.4. Completeness

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

Example 2.31. Let (X, d) be a metric space and suppose $(x_n) \longrightarrow x \in X$. Then, given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $d(x_n, x_m) < \varepsilon$ for all $n, m \geqslant N$.

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

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

A sequence (x_n) that satisfies the conclusion of Example 2.31 is said to be *Cauchy*.

A natural question is whether the converse of Example 2.31 holds: does every Cauchy sequence converge?

A metric space (X, d) is said to be *complete* if every Cauchy sequence converges to an element of X.

Example 2.32. If (X,d) is a complete metric space and $S \subseteq X$, then S is complete if and only if S is closed.

Solution. TODO $\square$

Recall that the intersection of two open dense subsets is open and dense (Exercise 2.12), hence the same is true for the intersection of any finite collection of open dense subsets. What happens if we drop the finiteness assumption? In general we cannot expect anything good:

Example 2.33. Note that for every $x \in \mathbb{R}$, $U_x := \mathbb{R} \setminus \{x\}$ is dense and open in $\mathbb{R}$.

Conclude that the intersection of an uncountable collection of dense open subsets need not be dense.

Solution. Clear, as $\bigcap_{x\in\mathbb{R}} U_x = \varnothing$, which is certainly not dense in $\mathbb{R}$. $\square$

But if we stick to complete metric spaces and to countable collections, we are in good shape again:

Theorem 2.34 (Baire Category Theorem)**.** *Let (X,d) be a complete metric space. Then the intersection of any countable collection of dense open subsets of X is dense in X.*

Proof. Let $\{U_n : n \in \mathbb{N}\}$ be a countable collection of dense open subsets of X and let D be their intersection.

We use the criterion in Exercise 2.10. Let W be a nonempty open set. We want to show that $D \cap W \neq \varnothing$.

Since U_1 is dense, $W \cap U_1$ is a nonempty open set. Let $x_1 \in W \cap U_1$ and let $0 < r_1 < 1$ be such that

$$\mathbb{D}_{r_1}(x_1) \subseteq W \cap U_1.$$

Since U_2 is dense, $\mathbb{B}_{r_1}(x_1) \cap U_2$ is a nonempty open set. Let $x_2 \in \mathbb{B}_{r_1}(x_1) \cap U_2$ and let $0 < r_2 < \frac{1}{2}$ be such that

$$\mathbb{D}_{r_2}(x_2) \subseteq \mathbb{B}_{r_1}(x_1) \cap U_2.$$

We continue in this manner; for each $n \geqslant 2$, U_n is dense, so $\mathbb{B}_{r_{n-1}}(x_{n-1}) \cap U_n$ is nonempty and open. Let $x_n \in \mathbb{B}_{r_{n-1}}(x_{n-1}) \cap U_n$ and let $0 < r_n < \frac{1}{n}$ be such that

$$\mathbb{D}_{r_n}(x_n) \subseteq \mathbb{B}_{r_{n-1}}(x_{n-1}) \cap U_n.$$

We obtain a sequence (x_n). It is Cauchy by construction: if $n \geqslant m$ then $x_n \in \mathbb{B}_{r_m}(x_m) \subseteq \mathbb{B}_{1/m}(x_m)$. Since X is complete, $(x_n) \longrightarrow x \in X$.

For each $m \in \mathbb{N}$, $(x_n)_{n \geqslant m}$ is a convergent sequence of elements of the closed set $\mathbb{D}_{r_m}(x_m)$, hence its limit $x \in \mathbb{D}_{r_m}(x_m) \subseteq W \cap U_m$. Therefore $x \in W$ and $x \in U_m$ for all $m \in \mathbb{N}$, in other words $x \in W \cap D$. $\square$

It might be hard to see the point of this result right now, but it is used in multiple ways in functional analysis, so we'll see it come up later.

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

> **Example 2.35.** Let (X, d) be a complete metric space and let $S \subseteq X$. Then the closure $\overline{S}$ (with the metric induced from $\overline{S} \subseteq X$) is a completion of S (with the metric induced from $S \subseteq X$).
>
> *Solution.* Of course, $\overline{S}$ is complete: if (x_n) is a Cauchy sequence in $\overline{S}$, then it is a Cauchy sequence in X, so $(x_n) \longrightarrow x \in X$ since X is complete. But $\overline{S}$ is closed, so $(x_n) \longrightarrow x \in \overline{S}$.
>
> We let $\iota\colon S \longrightarrow \overline{S}$ be the inclusion map: $\iota(s) = s$ for all $s \in S$. It is injective and distance-preserving (as d_S and $d_{\overline{S}}$ are both induced from d_X).
>
> Finally, S is dense in $\overline{S}$: by Example 2.22, for every $x \in \overline{S}$ there exists a sequence (s_n) in S such that $(s_n) \longrightarrow x$. □

Theorem 2.36. *Any metric space (X, d) has a completion.*

We will see later (Example 2.44) that any two completions of (X, d) are isometric.

We give a proof of the Theorem using the fact that $\mathbb{R}$ is complete. (This can be proved by defining $\mathbb{R}$ as the completion of $(\mathbb{Q}, d_2)$ and using arguments similar to the ones given below.)

Proof of Theorem 2.36. Given (X, d), consider the set $\mathcal{C}$ of all Cauchy sequences, and define an equivalence relation on $\mathcal{C}$ by:

$$(x_n) \sim (x'_n) \qquad \text{if given } \varepsilon > 0, \text{ there exists } N \in \mathbb{N} \text{ such that } d(x_n, x'_n) < \varepsilon \text{ for all } n \geqslant N.$$

Put more concisely, $(x_n) \sim (x'_n)$ if $\big(d(x_n, x'_n)\big) \longrightarrow 0 \in \mathbb{R}$.

Let $\widehat{X}$ be the resulting set of equivalence classes $[(x_n)]$. Define $\widehat{d}\colon \widehat{X} \times \widehat{X} \longrightarrow \mathbb{R}_{\geqslant 0}$ 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 $\mathbb{R}$ (Example 2.37) and $\mathbb{R}$ is complete; moreover $\widehat{d}$ is well-defined, see Example 2.39.

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

Consider the map $\iota\colon X \longrightarrow \widehat{X}$ given by

$$\iota(x) = [(x, x, \dots)].$$

If $\iota(x) = \iota(y)$ then $(x, x, \dots) \sim (y, y, \dots)$, but the distance between the n-th elements of these two sequences is the constant $d(x, y)$, which is then forced to be 0, so that $x = y$.

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

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

so ι is distance-preserving.

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

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

Let's check that the metric space $\big(\widehat{X}, \widehat{d}\,\big)$ is complete. Suppose (a_n) is a Cauchy sequence in $\widehat{X}$. As $\iota(X)$ is dense in $\widehat{X}$, for each $n \in \mathbb{N}$ there exists $x_n \in X$ such that $\widehat{d}(\iota(x_n), a_n) < \frac{1}{n}$. As (a_n) is Cauchy in $\widehat{X}$, by Example 2.38 so is the sequence $(\iota(x_n))$ in $\widehat{X}$, and hence so is the sequence (x_n) in X as $\iota(X)$ is isometric to X. So we have an element $\widehat{x} \coloneqq [(x_n)] \in \widehat{X}$.

I claim that (a_n) converges to $\widehat{x}$. Let $\varepsilon > 0$. We want to show that there exists $N \in \mathbb{N}$ 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 $m \in \mathbb{N}$.

We have by the triangle inequality

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

so taking limits:

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

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

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

Example 2.37. 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 $\mathbb{R}$.

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

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

so

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

Similarly:

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

so that

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

We can summarise this as

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

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

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

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

Example 2.38. Let (X, d) be a metric space and let (a_n) be a Cauchy sequence in X. Suppose (b_n) is a sequence in X such that $(b_n) \sim (a_n)$. Prove that (b_n) is Cauchy.

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

$$d(b_n, b_m) \leqslant d(b_n, a_n) + d(a_n, a_m) + d(a_m, b_m) < \epsilon. \qquad \square$$

Example 2.39. In the context of the proof of Theorem 2.36, show that if $(x_n) \sim (x'_n)$ and $(y_n) \sim (y'_n)$, then

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

Solution. This uses the same approach as Example 2.37: we have

$$\left| d(x'_n, y'_n) - d(x_n, y_n) \right| \leqslant d(x'_n, x_n) + d(y'_n, y_n).$$

But by assumption the two distances on the RHS can be made arbitrarily small, so we conclude that $d(x'_n, y'_n)$ and $d(x_n, y_n)$ can be made arbitrarily close, hence they have the same limit.

(This explanation shouldn't keep you from writing a more rigorous proof.) $\square$

Do you remember from Theorem 2.25 how amazing continuous functions are at preserving all sorts of stuff? Well, it turns out that nobody's perfect:

Example 2.40. Give an example of a continuous function $f\colon X \longrightarrow Y$ between metric spaces, and a Cauchy sequence (x_n) in X, such that the sequence $\big(f(x_n)\big)$ is not Cauchy in Y.

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

If you want your functions to preserve Cauchyness, 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 $f\big(\mathbb{B}_\delta(x)\big) \subseteq \mathbb{B}_\varepsilon\big(f(x)\big)$.

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 $x \in X$.)

Example 2.41. Prove that any uniformly continuous function maps Cauchy sequences to Cauchy sequences.

Solution. Let $f\colon X \longrightarrow Y$ be uniformly continuous and let (x_n) be a Cauchy sequence in X. For all $n \in \mathbb{N}$, 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 $N \in \mathbb{N}$ such that $d_X(x_n, x_m) < \delta$ for all $n, m \geqslant N$. Therefore $d_Y\big(y_n, y_m\big) < \varepsilon$ for all $n, m \geqslant N$. $\square$

If $f: X \longrightarrow Y$ is some kind of function between metric spaces and $\widehat{X}$, $\widehat{Y}$ are completions of X, Y, we may ask whether f can be extended to a function of a similar kind $\widehat{f}: \widehat{X} \longrightarrow \widehat{Y}$. 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 $\widehat{f}: \widehat{X} \longrightarrow \widehat{Y}$ is an *extension* of $f: X \longrightarrow Y$ if

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

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

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

A reasonable first attempt would be to see if any **continuous** function $f: X \longrightarrow Y$ extends to a **continuous** function $\widehat{f}: \widehat{X} \longrightarrow \widehat{Y}$. It turns out that such a continuous extension may not exist (Example 2.42), but when it does, it is unique (Example 2.43).

Example 2.42. Let $X = \mathbb{R}_{>0}$, $Y = \mathbb{R}$, $f: X \longrightarrow Y$ given by $f(x) = \frac{1}{x}$. For $\widehat{X} = \mathbb{R}_{\geqslant 0}$ and $\widehat{Y} = Y = \mathbb{R}$, prove that there is no continuous function $\widehat{f}: \widehat{X} \longrightarrow \widehat{Y}$ such that $\widehat{f}\,|_X = f$.

Solution. Suppose that a continuous extension $\widehat{f}: \mathbb{R}_{\geqslant 0} \longrightarrow \mathbb{R}_{\geqslant 0}$ exists. Consider the sequence $(x_n) = \left(\frac{1}{n}\right) \longrightarrow 0 \in \mathbb{R}_{\geqslant 0}$. By continuity of $\widehat{f}$ we must have

$$\widehat{f}(0) = \widehat{f}\left(\lim_{n\longrightarrow\infty} \frac{1}{n}\right) = \lim_{n\longrightarrow\infty} \widehat{f}\left(\frac{1}{n}\right) = \lim_{n\longrightarrow\infty} f\left(\frac{1}{n}\right) = \lim_{n\longrightarrow\infty} n.$$

But the rightmost limit does not exist (in $\mathbb{R}_{\geqslant 0}$), contradiction. □

Example 2.43. Let $f_1, f_2: X \longrightarrow Y$ be two continuous functions between metric spaces. Suppose that there exists a dense subset $D \subseteq X$ such that $f_1|_D = f_2|_D$, in other words that

$$f_1(d) = f_2(d) \qquad \text{for all } d \in D.$$

Then $f_1 = f_2$.

Deduce that any two continuous extensions $g_1, g_2: \widehat{X} \longrightarrow \widehat{Y}$ of a continuous function $g: X \longrightarrow Y$ to completions $\widehat{X}$, $\widehat{Y}$ must be equal.

Solution. Let $x \in X$. As D is dense, we have $x \in \overline{D}$ so there is a sequence $(x_n) \longrightarrow x$ with $x_n \in D$ for all $n \in \mathbb{N}$. But f_1 and f_2 are continuous on X and they agree at each x_n, so

$$f_1(x) = f_1\left(\lim_{n\longrightarrow\infty} x_n\right) = \lim_{n\longrightarrow\infty} f_1(x_n) = \lim_{n\longrightarrow\infty} f_2(x_n) = f_2\left(\lim_{n\longrightarrow\infty} x_n\right) = f_2(x).$$

For the case of completions, let $D = \iota_X(X) \subseteq \widehat{X}$ and use the above. □

It is, however, the case that any **uniformly continuous** (resp. **distance-preserving**) function $f: X \longrightarrow Y$ extends uniquely to a **uniformly continuous** (resp. **distance-preserving**) function $\widehat{f}: \widehat{X} \longrightarrow \widehat{Y}$. See Exercise 2.40.

This has the following consequence:

Example 2.44. Let (X,d) be a metric space. Prove that any two completions of (X,d) are isometric.

Solution. Let $\big(\widehat{X}_1, \widehat{d}_1\big)$ and $\big(\widehat{X}_2, \widehat{d}_2\big)$ be two completions.

We have isometries $\iota_1 \colon X \longrightarrow \iota_1(X) \subseteq \widehat{X}_1$ and $\iota_2 \colon X \longrightarrow \iota_2(X) \subseteq \widehat{X}_2$. Consider the composition $f \coloneqq \iota_2 \circ \iota_1^{-1} \colon \iota_1(X) \longrightarrow \iota_2(X)$. It is an isometry, in particular it is distance-preserving, so by Exercise 2.40 it extends uniquely to a distance-preserving function $\widehat{f} \colon \widehat{X}_1 \longrightarrow \widehat{X}_2$.

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

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

Example 2.45. Let (X, d_X) and (Y, d_Y) be metric spaces. A *contraction* is a function $f \colon X \longrightarrow Y$ for which there exists a constant $C \in (0,1)$ such that

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

Prove that any contraction is uniformly continuous.

Solution. Let $\varepsilon > 0$ and set $\delta = \frac{\varepsilon}{C}$, then for all $x_1, x_2 \in X$ such that $d_X(x_1,x_2) < \delta$, we have

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

Contraction self-maps on a complete metric space have an amazing property that is incredibly useful:

Theorem 2.46 (Banach Fixed Point Theorem)**.** *Let (X,d) be a nonempty complete metric space. Let $f \colon X \longrightarrow X$ be a contraction. Then f has a unique fixed point, that is an element $x \in X$ such that $f(x) = x$. Moreover, for any choice of $x_1 \in X$, the sequence (x_n) defined recursively by $x_{n+1} = f(x_n)$ converges to the fixed point x.*

Proof. The uniqueness claim is easy to show: if x, x' are such that $x = f(x)$ and $x' = f(x')$, then

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

If $x \neq x'$ then $d(x,x') > 0$ and $C\, d(x,x') < d(x,x')$ since $0 < C < 1$, leading to a contradiction.

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

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

Applying this repeatedly with decreasing m, we get

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

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

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

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

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

So x is indeed a fixed point of f. □

Recall the following result from real analysis:

Theorem 2.47 (Mean Value Theorem). *Let $f\colon [a,b] \longrightarrow \mathbb{R}$ be continuous. If f is differentiable on (a,b), then there exists $\xi \in (a,b)$ such that*

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

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

Example 2.48. Verify that the function $f\colon [1,2] \longrightarrow \mathbb{R}$ defined by

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

has a unique fixed point, and find this point.

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

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

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

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

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

Now let $x_1, x_2 \in [1,2]$. Apply Theorem 2.47 to f restricted to the interval $[x_1, x_2]$, and deduce that there exists $\xi \in (x_1, x_2) \subseteq [1,2]$ such that

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

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

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

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

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

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

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

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

2.5. Connected sets

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

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

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

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

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

We say that a metric space (X,d) is *totally disconnected* if the only connected subsets of X are the empty set and the singletons.

Example 2.49. In any metric space X, $\varnothing$ and the singletons $\{x\}$, $x \in X$, are (vacuously) connected.

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

Example 2.50. A metric space (X,d) is disconnected if and only if there exists a non-constant continuous function $g\colon X \longrightarrow \{0,1\}$. (Of course the metric on $\{0,1\}$ is taken to be discrete.)

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

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

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

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

Example 2.51. A subset A of a metric space (X,d) is both closed and open if and only if $\partial A = \varnothing$.

Solution. This follows easily from Examples 2.16 and 2.18, which say that $A^\circ = A \smallsetminus \partial A$ and $\overline{A} = A^\circ \cup \partial A$:

If A is closed and open then $A = \overline{A}$ and $A^\circ = A$, therefore $\partial A \subseteq A^\circ$, but since $A^\circ = A \smallsetminus \partial A$, we get $A^\circ \cap \partial A = \varnothing$. So we conclude that $\partial A = \varnothing$.

Conversely, if $\partial A = \varnothing$ then $A^\circ = A$ and A is open, but also $\overline{A} = A^\circ = A$ and A is closed. □

This leads to another characterisation of disconnectedness:

Example 2.52. A metric space (X,d) is disconnected if and only if it has a nonempty subset $U \subsetneq X$ with $\partial U = \varnothing$.

Solution. Suppose there exists a nonempty subset $U \subsetneq X$ with $\partial U = \varnothing$, and let $V := X \smallsetminus U$. By Example 2.51 U is both closed and open, so its complement V is both closed and open.

In the other direction, suppose X is disconnected and write $X = U \cup V$, $U \cap V = \varnothing$, both U and V open nonempty. Then U is both open and closed, so by Example 2.51 $\partial U = \varnothing$. □

Proposition 2.53. *If $f\colon X \longrightarrow Y$ is a continuous function between metric spaces and X is connected, then $f(X)$ is connected.*

Proof. Suppose $f(X)$ is disconnected, then by Example 2.50 there exists a non-constant continuous function $g\colon f(X) \longrightarrow \{0,1\}$. In particular, $f(X)$ has at least two elements. Then the composition $g \circ f\colon X \longrightarrow f(X) \longrightarrow \{0,1\}$ is a non-constant continuous function, implying that X is discontinuous. □

Example 2.54. The metric space $\mathbb{R}$ is connected.

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

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

There is a similar notion of *infimum*.

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

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

We use the criterion from Example 2.52, so we need to show that every nonempty subset $A \subsetneq \mathbb{R}$ has nonempty boundary.

Let $x \in \mathbb{R} \smallsetminus A$. We have two possibilities:

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

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

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

Example 2.55. The nonempty connected subsets of $\mathbb{R}$ are the intervals.

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

Conversely, suppose I is an interval in $\mathbb{R}$. Then (Exercise 2.52) there exists a surjective continuous function $f\colon \mathbb{R} \longrightarrow I$, hence I is connected because $\mathbb{R}$ is connected. □

Theorem 2.56 (Intermediate Value Theorem)**.** *Let $f\colon X \longrightarrow \mathbb{R}$ be a continuous function, with X a connected metric space. For any $x, y \in X$ and any $r \in \mathbb{R}$ 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 $\mathbb{R}$, hence an interval, from which the conclusion follows. □

2.6. Compactness

Let (X, d) be a metric space.

The *diameter* of a subset $S \subseteq X$ is by definition

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

If this is a real number we say that S is *bounded*. Otherwise we say that S is *unbounded*.

Example 2.57. A subset $S \subseteq X$ is bounded if and only if $S \subseteq \mathbb{D}_r(x)$ for some $r \geqslant 0$ and some $x \in X$.

Solution. If $S \subseteq \mathbb{D}_r(x)$ then $\operatorname{diam}(S) \leqslant \operatorname{diam}\left(\mathbb{D}_r(x)\right) = 2r$ so S is bounded.

Conversely, suppose S is bounded and let $r = \operatorname{diam}(S)$. Let $x \in S$ be any point, then $d(x,y) \leqslant r$ for all $y \in S$, so that $S \subseteq \mathbb{D}_r(x)$. □

Example 2.58. Let (X, d) be a metric space and let A, B be bounded sets. Then $A \cup B$ is bounded.

Solution. Let $a \in A$, $b \in B$, and $r = d(a, b)$. I claim that the diameter of $A \cup B$ is at most $\operatorname{diam}(A) + r + \operatorname{diam}(B)$. If $x, y \in A \cup B$ then

$$d(x,y) \leqslant \begin{cases} \operatorname{diam}(A) & \text{if } x, y \in A \\ \operatorname{diam}(B) & \text{if } x, y \in B \\ d(x,a) + d(a,b) + d(b,y) \leqslant \operatorname{diam}(A) + r + \operatorname{diam}(B) & \text{if } x \in A, y \in B \\ d(y,a) + d(a,b) + d(b,x) \leqslant \operatorname{diam}(A) + r + \operatorname{diam}(B) & \text{if } x \in B, y \in A. \end{cases} \quad \square$$

Example 2.59. Let $S \subseteq \mathbb{R}$ be a bounded set. Show that for any $\varepsilon > 0$, there exist $N \in \mathbb{N}$ and open balls $B_1, \ldots, B_N$, all of radius ε, such that

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

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

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

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

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

Example 2.60. In any metric space (X, d), any totally bounded set S is bounded.

Solution. Take $\varepsilon = 1$ and let $B_1, \ldots, B_N$ be a cover of S by open balls of radius 1. Each B_n is bounded, so by Example 2.58 the finite union $B_1 \cup \cdots \cup B_N$ is bounded, hence so is its subset S. $\square$

We have seen that the converse is true if $X = \mathbb{R}$.

Example 2.61. If $f \colon X \longrightarrow Y$ is a uniformly continuous function between metric spaces and $S \subseteq X$ is totally bounded, then $f(S) \subseteq Y$ is totally bounded.

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

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

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

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

so applying f on both sides we get

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

Before we delve into the next result, let's define a notation that will hopefully simplify things. If (x_n) is a sequence in a metric space (X, d) and $A \subseteq X$ is a subset such that $x_n \in A$ for infinitely many A, we define

$$(x_n) \cap A$$

to be the subsequence (x_{n_j}) with $\{n_j \colon j \in \mathbb{N}\} = \{n \in \mathbb{N} \colon x_n \in A\}$, enumerated in the natural order on $\mathbb{N}$.

For example, if $x_n = \frac{(-1)^n}{n} \in \mathbb{R}$ and $A = [0, 1]$, then

$$(x_n) \cap A = \left(\frac{1}{2j}\right) = (x_{n_j}) \qquad \text{where } n_j = 2j \text{ for } j \in \mathbb{N}.$$

Proposition 2.62. *A subset $S \subseteq X$ is totally bounded if and only if every sequence in S has a Cauchy subsequence.*

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

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

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

Continuing in this manner, we get a subsequence $\big(s_n^{(n)}\big)$ of (s_n). I claim that $\big(s_n^{(n)}\big)$ is a Cauchy sequence.

Given $\varepsilon > 0$, let $N \in \mathbb{N}$ be such that $2/N \leqslant \varepsilon$. For $n \geqslant m \geqslant N$ we have $s_m^{(m)}, s_n^{(n)} \in \big(s_j^{(m)}\big) \subseteq \mathbb{B}_{1/m}(x_m)$, hence

$$d\big(s_m^{(m)}, s_n^{(n)}\big) \leqslant d\big(s_m^{(m)}, x_m\big) + d\big(x_m, s_n^{(n)}\big) < \frac{2}{m} \leqslant \frac{2}{N} \leqslant \varepsilon.$$

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

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

We have been experimenting with various flavours of finiteness. Here is yet another one, which turns out to be more generally useful:

A subset $K \subseteq X$ of a metric space is *compact* if any arbitrary *open cover* of K, that is a collection $\{U_i : i \in I\}$ of open sets $U_i \subseteq X$ such that

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

has a finite *subcover*, that is there exist $n \in \mathbb{N}$ and $i_1, \dots, i_n \in I$ such that

$$K \subseteq \bigcup_{j=1}^{n} U_{i_j}.$$

Example 2.63. Any compact subset $K \subseteq X$ is totally bounded.

Solution. This is straightforward: given $\varepsilon > 0$, consider the open cover

$$K \subseteq \bigcup_{x \in K} \mathbb{B}_\varepsilon(x).$$

As K is compact, this has a finite subcover, proving total boundedness. ☐

Example 2.64. Prove that a subset $K \subseteq X$ is compact if and only if: any open cover of K by open balls

$$K \subseteq \bigcup_{i \in I} \mathbb{B}_{\varepsilon_i}(x_i)$$

there is a finite subcover

$$K \subseteq \bigcup_{k=1}^{n} \mathbb{B}_{\varepsilon_{i_k}}(x_{i_k}), \qquad n \in \mathbb{N}, i_k \in I.$$

Solution. The direction left to right is clear.

For the converse, suppose we are given an open cover

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

By Example 2.13 we know that each U_i is a union of open balls:

$$U_i = \bigcup_{j\in J_i} \mathbb{B}_{\varepsilon_j}(x_j),$$

so that we get an open cover of K by open balls

$$K \subseteq \bigcup_{j\in\bigcup_{i\in I} J_i} \mathbb{B}_{\varepsilon_j}(x_j).$$

This has a finite subcover

$$K \subseteq \bigcup_{k=1}^{n} \mathbb{B}_{\varepsilon_{j_k}}(x_{j_k}),$$

but for each k we have $\mathbb{B}_{\varepsilon_{j_k}}(x_{j_k}) \subseteq U_{i_k}$ for some $i_k \in I$, so that

$$K \subseteq \bigcup_{k=1}^{n} U_{i_k}$$

is a subcover of the original cover. □

Example 2.65. Any compact subset $K \subseteq X$ is closed.

Solution. Let $x \in X \smallsetminus K$.

For any $k \in K$, let $r_k = d(x,k)/2$, then $\mathbb{B}_{r_k}(k) \cap \mathbb{B}_{r_k}(x) = \varnothing$.

We now have an open cover of K:

$$K \subseteq \bigcup_{k\in K} \mathbb{B}_{r_k}(k),$$

which has a finite subcover

$$K \subseteq \bigcup_{n=1}^{N} \mathbb{B}_{r_n}(k_n).$$

Letting $r = \min\{r_n : n = 1,\dots,N\}$, we have $\mathbb{B}_{r_n}(k_n) \cap \mathbb{B}_r(x) = \varnothing$ for all n, hence $K \cap \mathbb{B}_r(x) = \varnothing$ and $X \smallsetminus K$ is open. □

Proposition 2.66. *If $f\colon X \longrightarrow Y$ is a continuous function between metric spaces and $K \subseteq X$ is a compact subset, then $f(K)$ is a compact subset of Y.*

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

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

Then

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

which is an open cover of K as f is continuous. As K is compact, this has a finite subcover

$$K \subseteq \bigcup_{n=1}^{N} f^{-1}(V_{i_n}),$$

therefore
$$f(K) \subseteq \bigcup_{n=1}^{N} V_{i_n}. \qquad \square$$

Proposition 2.67. *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 $f\big(\mathbb{B}_{\delta(x)}(x)\big) \subseteq \mathbb{B}_{\varepsilon/2}\big(f(x)\big)$. We get an open cover of X:
$$X \subseteq \bigcup_{x\in X} \mathbb{B}_{\delta(x)/2}(x),$$
which therefore has a finite subcover
$$X \subseteq \bigcup_{n=1}^{N} \mathbb{B}_{\delta(x_n)/2}(x_n).$$
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 $s \in \mathbb{B}_{\delta(x_n)/2}(x_n)$ for some $n \in \{1, \dots, N\}$. I claim that $t \in \mathbb{B}_{\delta(x_n)/2}(x_n)$ as well:
$$d_X(t, x_n) \leqslant d_X(t,s) + d_X(s, x_n) < \delta + \frac{\delta(x_n)}{2} \leqslant \delta(x_n).$$
Therefore $f(s), f(t) \in \mathbb{B}_{\varepsilon/2}\big(f(x_n)\big)$, hence $d_Y(f(s), f(t)) < \varepsilon$. $\square$

Example 2.68. Any bijective continuous function $f\colon X \longrightarrow Y$ from a compact metric space to a metric space is a homeomorphism.

Solution. We have to prove that $f^{-1}\colon Y \longrightarrow X$ is continuous. Let $U \subset X$ be open, then $X \setminus U$ is a closed subset of the compact space X, hence $X \setminus U$ is compact, hence its image $f(X \setminus U) = Y \setminus f(U)$ is compact in Y, hence closed in Y, hence $f(U) = \big(f^{-1}\big)^{-1}(U)$ is open in Y. $\square$

Example 2.69. Let S be a subset of a metric space (X, d). Prove that every sequence in S has a subsequence converging in S if and only if every infinite subset of S has an accumulation point in S.

Solution. The statement is clear if S is a finite set (any sequence must take the same value infinitely often, so has a constant subsequence, which is obviously converging in S; the other condition is vacuously true as there are no infinite subsets of S).

Let A be an infinite subset of S. Then there is a sequence (a_n) consisting of distinct elements of A. This has a subsequence $(a_{n_j}) \longrightarrow a \in S$. For any $\varepsilon > 0$, $\mathbb{B}_\varepsilon(a)$ contains all but finitely many elements of (a_{n_j}), and since these are distinct, at least one of them is $\neq a$, hence $a \in S$ is an accumulation point of A.

Conversely, let (s_n) be a sequence in S. Let $A = \{s_n\colon n \in \mathbb{N}\}$. If A is finite, then (s_n) takes the same value infinitely many times, so it has a constant sequence, which converges in S. Otherwise A is infinite so it has an accumulation point $a \in S$. Hence for every $j \in \mathbb{N}$ we have that $\mathbb{B}_{1/j}(a)$ contains some $a_j \in A$ with $a_j \neq a$. Then $(a_j) \longrightarrow a \in S$ is a converging subsequence of (s_n).

(Note: the last part of this is not entirely correct, namely it is not clear that the way in which we chose (a_j) gives a subsequence of (s_n), rather than just a sequence consisting of terms of (s_n). Do you see how to fix this point?

Answer: for $j = 1$, look at $\mathbb{B}_1(a)$; as a is an accumulation point of A, $\mathbb{B}_1(a) \cap A$ is an infinite set. Choose $a_1 \in \mathbb{B}_1(a) \cap A$ and let $n_1 \in \mathbb{N}$ denote the index of its first appearance in (s_n), that is $s_{n_1} = a_1$. Next, for $j = 2$, look at $\mathbb{B}_{1/2}(a) \cap A$, which again must be an infinite set. Choose $a_2 \in \mathbb{B}_{1/2}(a) \cap A$ such that $a_2 \notin \{s_k : k \leqslant n_1\}$ and let $n_2 \in \mathbb{N}$ denote the index of its first appearance in (s_n), that is $s_{n_2} = a_2$. Continue.) ☐

Example 2.70. Suppose (x_n) is a Cauchy sequence in a metric space (X, d). If (x_n) has a subsequence that converges to some $x \in X$, then (x_n) also converges to x.

Solution. Suppose $(x_{n_j}) \longrightarrow x$. Let $\varepsilon > 0$, then there exists $N \in \mathbb{N}$ such that $d(x_m, x_n) < \varepsilon/2$ for all $m, n \geqslant N$, and there exists $J \in \mathbb{N}$ such that $d(x_{n_j}, x) < \varepsilon/2$ for all $j \geqslant J$. Let $N' = \max\{N, n_J\}$, then for all $n \geqslant N'$ we have

$$d(x_n, x) \leqslant d(x_n, x_{n_J}) + d(x_{n_J}, x) < \varepsilon.$$ ☐

Theorem 2.71 (Heine–Borel)**.** *Let (X, d) be a metric space. A subset $K \subseteq X$ is compact if and only if K is complete and totally bounded.*

Proof. Suppose K is compact. We already know from Example 2.63 that K is totally bounded. Suppose it is not complete, and take a Cauchy sequence (x_n) in K that does not converge in K. (In particular, this forces the set $\{x_n : n \in \mathbb{N}\}$ to be infinite, otherwise the sequence would have a constant, hence converging, subsequence, therefore (x_n) would also converge to the same limit, being Cauchy.)

This means that, for every $k \in K$, there exists $\varepsilon > 0$ such that for all $N \in \mathbb{N}$ there exists $n \geqslant N$ with $d(x_n, k) \geqslant \varepsilon$. Since (x_n) is Cauchy, there exists $N' \in \mathbb{N}$ such that $d(x_m, x_n) < \varepsilon/2$ for all $m, n \geqslant N'$. Pick an $n \geqslant N'$ with $d(x_n, k) \geqslant \varepsilon$, then for all $m \geqslant N'$ we have

$$d(x_m, k) + \frac{\varepsilon}{2} > d(x_m, k) + d(x_m, x_n) \geqslant d(x_n, k) \geqslant \varepsilon,$$

hence $d(x_m, k) \geqslant \varepsilon/2$ for all $m \geqslant N'$.

So, for a given $k \in K$, letting $\varepsilon(k) = \min\left\{\{d(x_j, k) : j < N', x_j \neq k\} \cup \{\varepsilon/2\}\right\}$, we have that $\mathbb{B}_{\varepsilon(k)}(k)$ contains no points of the sequence (x_n) unless $x_n = k$.

We get an open cover of K:

$$K \subseteq \bigcup_{k \in K} \mathbb{B}_{\varepsilon(k)}(k),$$

which has a finite subcover

$$K \subseteq \bigcup_{j=1}^{M} \mathbb{B}_{\varepsilon(k_j)}(k_j).$$

But (x_n) is a sequence of K, so it must only take the finitely many values $\{k_1, \ldots, k_M\}$, contradiction.

Conversely, suppose K is complete and totally bounded, but not compact.

From now on, we restrict our attention to K and work with the induced metric from X. In particular, all sets that appear are subsets of K.

Take an open cover of K that does not have a finite subcover:

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

As K is totally bounded, for $\varepsilon = 1$ we have a cover

$$K = \bigcup_{j=1}^{N} \mathbb{B}_1(x_j).$$

It follows that at least one of the open balls $\mathbb{B}_1(x_j)$ cannot be covered by finitely many of the U_i's. After renumbering we may assume that $j = 1$ so the open ball in question is $\mathbb{B}_1(x_1)$.

Now apply the same process with $\mathbb{B}_1(x_1)$ and $\varepsilon = \frac{1}{2}$, to get an open ball $\mathbb{B}_{1/2}(x_2) \subseteq \mathbb{B}_1(x_1)$ and $\mathbb{B}_{1/2}(x_2)$ cannot be covered by finitely many of the U_i's.

Continuing, we get a sequence of subsets

$$\mathbb{B}_1(x_1) \supseteq \mathbb{B}_{1/2}(x_2) \supseteq \mathbb{B}_{1/3}(x_3) \supseteq \dots$$

such that $\mathbb{B}_{1/n}(x_n)$ cannot be covered by finitely many of the U_i's. The resulting sequence (x_n) is Cauchy, hence converges to some $x \in K$ as K is complete.

There exists $i_x \in I$ such that $x \in U_{i_x}$. Since U_{i_x} is open, we can find an open ball $\mathbb{B}_\varepsilon(x) \subseteq U_{i_x}$. Choose $N \in \mathbb{N}$ such that $N > 2/\varepsilon$ and $d(x_N, x) < \varepsilon/2$. Then

$$\mathbb{B}_{1/N}(x_N) \subseteq \mathbb{B}_\varepsilon(x) \subseteq U_{i_x},$$

contradicting the fact that $\mathbb{B}_{1/N}(x_N)$ cannot be covered by finitely many U_i's. ☐

The following compactness criterion is sometimes called *sequential compactness*:

Theorem 2.72 (Bolzano–Weierstraß). *Let (X, d) be a metric space. A subset $K \subseteq X$ is compact if and only if every sequence in K has a subsequence converging in K.*

Proof. If K is compact, then it is totally bounded by Example 2.63, so any sequence in K has a Cauchy subsequence by Proposition 2.62, which converges since K is complete by Theorem 2.71.

In the other direction: suppose (x_n) is a Cauchy sequence in K. Since it has a subsequence that converges to some $x \in K$, (x_n) itself converges to x by Example 2.70. So K is complete.

If (x_n) is an arbitrary sequence in K, it has a subsequence that converges, hence is Cauchy, so by Proposition 2.62, K is totally bounded. ☐

Example 2.73. Let $f\colon X \longrightarrow \mathbb{R}$ be a continuous function, where X is a compact metric space. Then the image $f(X)$ is bounded, and the bounds are attained: there exist $x_{\min}, x_{\max} \in X$ such that

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

Solution. By Proposition 2.66, $f(X)$ is a compact subset of $\mathbb{R}$. Therefore $f(X)$ is totally bounded by Example 2.63, hence bounded by Example 2.59. So $f(X)$ has both infimum and supremum, which are boundary points. But $f(X)$ is also closed by Example 2.65, therefore it contains its boundary points and hence the infimum and supremum. ☐

2.7. Exercises

Exercise 2.1. Let (X, d) be a metric space. Show that

$$|d(x, y) - d(t, y)| \leqslant d(x, t)$$

for all $x, y, t \in X$.

Solution. We need to show that

$$-d(x, t) \leqslant d(x, y) - d(t, y) \leqslant d(x, t).$$

One application of the triangle inequality gives

$$d(x, y) \leqslant d(x, t) + d(t, y) \qquad \Rightarrow \qquad d(x, y) - d(t, y) \leqslant d(x, t).$$

Another application gives

$$d(t, y) \leqslant d(t, x) + d(x, y) \qquad \Rightarrow \qquad -d(x, t) \leqslant d(x, y) - d(t, y). \qquad \square$$

Exercise 2.2. Let (X, d) be a metric space. Show that

$$|d(x, y) - d(s, t)| \leqslant d(x, s) + d(y, t)$$

for all $x, s, y, t \in X$.

Solution. We have

$$\begin{aligned} |d(x, y) - d(s, t)| &= |d(x, y) - d(y, s) + d(y, s) - d(s, t)| \\ &\leqslant |d(x, y) - d(y, s)| + |d(y, s) - d(s, t)| \\ &\leqslant d(x, s) + d(y, t) \end{aligned}$$

after one application of the triangle inequality and two applications of Exercise 2.1. $\square$

Exercise 2.3. If $A \subseteq B$ then $A^\circ \subseteq B^\circ$.

Solution. Let $a \in A^\circ$, then there exists an open ball $\mathbb{B}_r(a) \subseteq A \subseteq B$, so $a \in B^\circ$. $\square$

Exercise 2.4. For any subset $A \subseteq X$, A° is the largest open set contained in A.

Solution. Suppose U is an open set such that $A^\circ \subseteq U \subseteq A$. If $u \in U$ then there exists an open ball $\mathbb{B}_r(u) \subseteq U \subseteq A$. Therefore $u \in A^\circ$, so $U \subseteq A^\circ$, hence $U = A^\circ$.

Let $a \in A^\circ$. There exists an open ball $\mathbb{B}_r(a) \subseteq A$. Therefore $\mathbb{B}_r(a) = \mathbb{B}_r(a)^\circ \subseteq A^\circ$. $\square$

Exercise 2.5. Prove that any closed ball is a closed set.

Solution. This is a variation on Example 2.10 (draw the pictures!), and a generalisation of Example 2.9 (which is the case $r = 0$).

Consider $C = \mathbb{D}_r(x)$ with $x \in X$, $r \in \mathbb{R}_{\geqslant 0}$. Let $y \in X \smallsetminus C$, then $d(x,y) > r$. Set $t = d(x,y) - r$ and consider the open ball $\mathbb{B}_t(y)$.

I claim that $\mathbb{B}_t(y) \subseteq \big(X \smallsetminus C\big)$: if $w \in \mathbb{B}_t(y)$ then $d(w,y) < t$ so

$$d(x,y) \leqslant d(x,w) + d(w,y) \leqslant d(x,w) + t \qquad \Rightarrow \qquad d(x,w) \geqslant d(x,y) - t = r,$$

hence $w \notin C$. □

Exercise 2.6. Prove that C is closed if and only if $\partial C \subseteq C$.

Solution. Suppose that C is closed and let $x \in \partial C$. We proceed by contradiction to prove that $x \in C$: suppose $x \notin C$, then $x \in \big(X \smallsetminus C\big)$, which is open as C is closed. Therefore there exists $r > 0$ such that $\mathbb{B}_r(x) \subseteq \big(X \smallsetminus C\big)$, in particular $\mathbb{B}_r(x) \cap C = \varnothing$, contradicting the fact that x is a boundary point.

For the opposite implication, suppose that $\partial C \subseteq C$. We proceed by contradiction to prove that C is closed: suppose $x \in X \smallsetminus C$ is such that for all $r > 0$, $\mathbb{B}_r(x)$ is not a subset of $X \smallsetminus C$. Therefore $\mathbb{B}_r(x) \cap C \neq \varnothing$, but also $\mathbb{B}_r(x) \cap \big(X \smallsetminus C\big) \neq \varnothing$, as this intersection contains x. We conclude that x is a boundary point of C, so by assumption $x \in C$, contradicting the fact that $x \in X \smallsetminus C$. □

Exercise 2.7. Show that any p-adic open ball is both open and closed.

Solution. Any open ball in any metric space is open (Example 2.10). Let's show that an arbitrary p-adic open ball $\mathbb{B}_r(c)$ is closed.

Let $x \in \partial\mathbb{B}_r(c)$. Let $s \leqslant r$ and consider $\mathbb{B}_s(x)$. Since x is a boundary point, there exists $y \in \mathbb{B}_s(x) \cap \mathbb{B}_r(c)$. In other words, $|x-y|_p < s \leqslant r$ and $|y-c|_p < r$. Therefore

$$|x-c|_p \leqslant \max\big\{|x-y|_p, |y-c|_p\big\} < \max\{s,r\} \leqslant r,$$

where we used the p-adic triangle inequality.

We conclude that $x \in \mathbb{B}_r(c)$, so by Exercise 2.6 $\mathbb{B}_r(c)$ is closed. □

Exercise 2.8. Let A be a finite subset of a metric space (X,d). Show that A has no limit points.

Solution. Suppose x is a limit point of A. Let $r_1 = 1$; there exists $a_1 \in \mathbb{B}_{r_1}(x) \cap A$ such that $a_1 \neq x$. Now let $r_2 = d(x,a_1)/2$; there exists $a_2 \in \mathbb{B}_{r_2}(x) \cap A$ such that $a_2 \neq x$. Note that we also have $a_2 \neq a_1$. Continue in this manner. For each $n \geqslant 3$, let $r_n = d(x, a_{n-1})/2$, then there exists $a_n \in \mathbb{B}_{r_n}(x) \cap A$ such that $a_n \neq x$ and $a_n \neq a_k$ for $k = 1, \dots, n-1$. This constructs an infinite subset $\{a_1, a_2, \dots\} \subseteq A$, contradicting the fact that A is finite. □

Exercise 2.9. Let (X,d) be a metric space.

(a) Prove that $X \smallsetminus \overline{A} = \big(X \smallsetminus A\big)^\circ$. (In other words, $x \in \overline{A}$ if and only if every open neighbourhood of x intersects A.)

Conclude that $\overline{A}$ is a closed set.

(b) If $A \subseteq B$, then $\overline{A} \subseteq \overline{B}$.

(c) Prove that $\overline{A}$ is the smallest closed subset of X that contains A.

Solution.

(a) Suppose that $x \notin \overline{A}$. Then $x \notin A$ and x is not a limit point of A. So there exists $r > 0$ such that $\mathbb{B}_r(x) \cap A$ has no points $\neq x$; but $x \notin A$ is also not an element of this intersection, which must therefore be empty, so that $\mathbb{B}_r(x) \subseteq \big(X \smallsetminus A\big)$. We conclude that $x \in \big(X \smallsetminus A\big)^\circ$.

Conversely, let $x \in \big(X \smallsetminus A\big)^\circ$, then $x \in X \smallsetminus A$ and there exists $r > 0$ such that $\mathbb{B}_r(x) \subseteq \big(X \smallsetminus A\big)$. In particular, $\mathbb{B}_r(x) \cap A = \varnothing$, so x is not a limit point of A. As it is also not an element of A (hence not an isolated point of A), we conclude that $x \notin \overline{A}$.

(b)

$$A \subseteq B \Rightarrow \big(X \smallsetminus B\big) \subseteq \big(X \smallsetminus A\big) \Rightarrow \big(X \smallsetminus B\big)^\circ \subseteq \big(X \smallsetminus A\big)^\circ \Rightarrow \big(X \smallsetminus \overline{B}\big) \subseteq \big(X \smallsetminus \overline{A}\big) \Rightarrow \overline{A} \subseteq \overline{B}.$$

(c) We already know from part (a) that $\overline{A}$ is a closed set.

If $A \subseteq C$ and C is closed, then $\overline{A} \subseteq \overline{C} = C$. ☐

Exercise 2.10. Let (X, d) be a metric space. A subset $D \subseteq X$ is dense in X if and only if $D \cap U \neq \varnothing$ for all nonempty open sets U in X.

Solution. Suppose D is dense and let U be nonempty open. Let $x \in U$. As U is open, there exists $\mathbb{B}_r(x) \subseteq U$ with $r > 0$. If $x \in D$, we are done. Otherwise, $x \in \overline{D} \smallsetminus D$, so it is a limit point of D, so there exists $a \in \mathbb{B}_r(x) \cap D$ such that $a \neq x$, hence $a \in U \cap D$.

Conversely, suppose $D \cap U$ is nonempty for any nonempty open U. Let $x \in X \smallsetminus D$. For every $r > 0$, $U := \mathbb{B}_r(x)$ is open so $D \cap \mathbb{B}_r(x)$ is nonempty, and x is not in this intersection so there must be a point distinct from x in it, hence x is a limit point of D. ☐

Exercise 2.11. Let (X, d) be a metric space, D a dense subset, and x an isolated point of X. Then $x \in D$.

Solution. Since x is isolated, there exists $r > 0$ such that $\mathbb{B}_r(x) = \{x\}$. But D is dense and $\mathbb{B}_r(x)$ is a nonempty open subset of X, hence $D \cap \mathbb{B}_r(x) \neq \varnothing$, which forces $D \cap \mathbb{B}_r(x) = \{x\}$. In particular $x \in D$. ☐

Exercise 2.12. Let (X, d) be a metric space. The intersection of two dense open sets U_1 and U_2 is dense and open.

Solution. Let $U_{12} = U_1 \cap U_2$. We know already that U is open.

To show that U_{12} is dense, we use Exercise 2.10 and show that $U_{12} \cap U \neq \varnothing$ for all nonempty open U:

$$U_{12} \cap U = \big(U_1 \cap U_2\big) \cap U = \mathrm{U}_1 \cap \big(U_2 \cap U\big).$$

Since U_2 is dense and open, $U_2 \cap U$ is nonempty and open. Since U_1 is dense, $U_1 \cap \big(U_2 \cap U\big)$ is nonempty. So $U_{12} \cap U \neq \varnothing$, hence U_{12} is dense. ☐

Exercise 2.13. Any sequence has at most one limit.

Solution. Suppose x and x' are two limits of a sequence (x_n). For any $\varepsilon > 0$, there exist $N, N' \in \mathbb{N}$ such that

$$x_n \in \mathbb{B}_{\varepsilon/2}(x) \qquad \text{for all } n \geqslant N \quad \text{and} \quad x_n \in \mathbb{B}_{\varepsilon/2}(x') \qquad \text{for all } n \geqslant N'.$$

Therefore, for $n = \max\{N, N'\}$ we have $x_n \in \mathbb{B}_{\varepsilon/2}(x) \cap \mathbb{B}_{\varepsilon/2}(x')$, which (via the triangle inequality) implies that $d(x, x') < \varepsilon$.

Since this holds for all $\varepsilon > 0$, we conclude that $d(x, x') = 0$ so that $x = x'$. □

Exercise 2.14 (tut02)**.** Let $n \in \mathbb{N}$, $X = \mathbb{R}^n$ with the dot product $\cdot$, $\|x\| = \sqrt{x \cdot x}$ for $x \in X$, and $d(x, y) = \|x - y\|$ for $x, y \in X$. Then (X, d) is a metric space. (The function d is called the *Euclidean metric* or ℓ^2 *metric* on $\mathbb{R}^n$.)

[*Hint*: The Cauchy–Schwarz inequality can be useful for checking the triangle inequality.]

Solution. We have

(a) $d(x, y) = \|x - y\| = \sqrt{(x - y) \cdot (x - y)} = \sqrt{(-1)^2\,(y - x) \cdot (y - x)} = \|y - x\| = d(y, x)$;

(b) Let $u = x - t$ and $v = t - y$, then we are looking to show that $\|u + v\| \leqslant \|u\| + \|v\|$. But:

$$\begin{aligned}\|u + v\|^2 = (u + v) \cdot (u + v) = \|u\|^2 + 2\,u \cdot v + \|v\|^2 &\leqslant \|u\|^2 + 2\,|u \cdot v| + \|v\|^2 \\ &\leqslant \|u\|^2 + 2\,\|u\|\,\|v\| + \|v\|^2 = \big(\|u\| + \|v\|\big)^2,\end{aligned}$$

where the last inequality sign comes from the Cauchy–Schwarz inequality.

(c) $d(x, y) = 0$ iff $(x - y) \cdot (x - y) = 0$ iff $x - y = 0$ iff $x = y$. □

Exercise 2.15. Fix $n \in \mathbb{N}$. Put a metric space structure on $\mathbb{C}^n$.

Solution. We mimic the construction in Exercise 2.14, replacing the real dot product with the standard Hermitian inner product:

$$\langle x, y \rangle = x_1 \overline{y}_1 + \cdots + x_n \overline{y}_n.$$ □

Exercise 2.16 (tut02)**.** Let X be a nonempty set and define

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

Prove that (X, d) is a metric space. (The function d is called the *discrete metric* on X.)

Solution. It is clear from the definition that $d(y, x) = d(x, y)$ and that $d(x, y) = 0$ iff $x = y$.

For the triangle inequality, take $x, y, t \in X$ and consider the different cases:

$x = y$	$x = t$	$t = y$	$d(x, y)$	$d(x, t) + d(t, y)$
True	True	True	0	$0 + 0 = 0$
True	False	False	0	$1 + 1 = 2$
False	True	False	1	$1 + 0 = 1$
False	False	True	1	$0 + 1 = 1$
False	False	False	1	$1 + 1 = 2$

In all cases we see that $d(x, y) \leqslant d(x, t) + d(t, y)$. $\square$

Exercise 2.17. Let $n \in \mathbb{N}$, $X = \mathbb{F}_2^n$, and let $d(x, y)$ be the number of indices $i \in \{1, \dots, n\}$ such that $x_i \neq y_i$. Prove that (X, d) is a metric space. (The function d is called the *Hamming metric*.)

Solution. Do this from scratch if you want to, but I prefer to deduce it from other examples we have seen.

First look at the case $n = 1$, $X = \mathbb{F}_2$. Then $d(x, y)$ is precisely the discrete metric on $\mathbb{F}_2$ (see Exercise 2.16), in particular it is a metric. I'll denote it $d_{\mathbb{F}_2}$ for a moment to minimise confusion.

Back in the arbitrary $n \in \mathbb{N}$ case, note that $d(x, y)$ defined above can be expressed as

$$d(x, y) = d_{\mathbb{F}_2}(x_1, y_1) + \cdots + d_{\mathbb{F}_2}(x_n, y_n),$$

which is a special case of Example 2.3, therefore also a metric. $\square$

Exercise 2.18 (tut02)**.** Let (X, d) be a metric space and define

$$d'(x, y) = \frac{d(x, y)}{1 + d(x, y)}.$$

Prove that (X, d') is a metric space.

[*Hint*: Before tackling the triangle inequality, show that if $a, b, c \in \mathbb{R}_{\geqslant 0}$ satisfy $c \leqslant a + b$, then $\frac{c}{1+c} \leqslant \frac{a}{1+a} + \frac{b}{1+b}$.]

Solution. It is clear from the definition that $d'(x, y) = d'(y, x)$ and that $d'(x, y) = 0$ iff $d(x, y) = 0$ iff $x = y$.

For the triangle inequality, apply the inequality in the hint with $c = d(x, y)$, $a = d(x, t)$, $b = d(t, y)$. $\square$

Exercise 2.19 (tut02)**.** Draw the unit open balls in the metric spaces $(\mathbb{R}^2, d_1)$ (Example 2.3), $(\mathbb{R}^2, d_2)$ (Exercise 2.14), and $(\mathbb{R}^2, d_\infty)$ (Example 2.4).

Solution. The Manhattan unit open ball is the interior of the square with vertices $(1, 0)$, $(0, -1)$, $(-1, 0)$, and $(0, 1)$.

The Euclidean unit open ball is the interior of the unit circle centred at $(0, 0)$.

The sup metric unit open ball is the interior of the square with vertices $(1, 1)$, $(1, -1)$, $(-1, -1)$, and $(-1, 1)$. $\square$

Exercise 2.20 (tut02)**.** Is the word "finite" necessary in the statement of Proposition 2.12? If no, give a proof of the statement without "finite". If yes, give an example of an infinite collection of open sets whose intersection is not an open set.

Solution. The word "finite" is necessary. For a counterexample to the more general statement, for each $n \in \mathbb{N}$ take $U_n = (-1/n, 1/n)$ as an open set in $\mathbb{R}$ with the Euclidean metric. I claim that

$$U := \bigcap_{n \in \mathbb{N}} U_n = \{0\}.$$

This can be proved by contradiction: suppose $u \in U$, $u \neq 0$. Let $n \in \mathbb{N}$ be such that $n \geqslant \frac{1}{|u|}$. Then $|u| \geqslant \frac{1}{n}$, therefore $u \notin (-1/n, 1/n) = U_n$, contradiction.

Finally, U is not open: for any $r \in \mathbb{R}_{>0}$, $\frac{r}{2} \in \mathbb{B}_r(0)$ but $\frac{r}{2} \notin \{0\} = U$, so $\mathbb{B}_r(0)$ is not a subset of U. □

Exercise 2.21. Let (X, d) be a discrete metric space (see Exercise 2.16).

(a) Prove that $\{x\}$ is an open subset of X for all $x \in X$.

(b) Prove that every subset of X is open.

(c) Prove that every subset of X is closed.

(d) Let A be a subset of X. Find all the (i) interior, (ii) boundary, (iii) isolated, and (iv) limit points of A.

Solution.

(a) Since $d(x, y) = 1$ for any $y \neq x$, we have $\mathbb{B}_1(x) = \{x\}$, which is an open ball, hence an open set.

(b) Any subset is the union of all its one-element subsets, hence is open by the previous part.

(c) The previous part says that the complement of any subset is open.

(d) Since A is open, $A^\circ = A$.

If $x \in X$ is a boundary point of A, then $\mathbb{B}_1(x)$ intersects both A and $X \setminus A$ nontrivially, so it must have at least two elements, but $\mathbb{B}_1(x) = \{x\}$. So $\partial A = \varnothing$.

We have seen in the proof of part (a) that every point of X is isolated, so the set of isolated points of A is A.

If $x \in X$ is a limit point of A, then $\mathbb{B}_1(x) \cap A$ should contain some $a \neq x$, but we have seen that $\mathbb{B}_1(x) = \{x\}$. So the set of limit points of A is $\varnothing$. □

Exercise 2.22 (tut03)**.** Let X and Y be two metric spaces and endow the Cartesian product $X \times Y$ with the Manhattan metric from Example 2.3. Prove that a sequence $\big((x_n, y_n)\big)$ in $X \times Y$ converges to (x, y) if and only if (x_n) converges to x and (y_n) converges to y.

Solution. By definition,

$$d\big((x_n, y_n), (x, y)\big) = d_X(x_n, x) + d_Y(y_n, y).$$

Suppose $(x_n) \longrightarrow x$ and $(y_n) \longrightarrow y$. Let $\varepsilon > 0$, $N_x \in \mathbb{N}$ such that $x_n \in \mathbb{B}_{\varepsilon/2}(x)$ for all $n \geqslant N_x$, and $N_y \in \mathbb{N}$ such that $y_n \in \mathbb{B}_{\varepsilon/2}(y)$ for all $n \geqslant N_y$. Set $N = \max\{N_x, N_y\}$, then

$$d\big((x_n, y_n), (x, y)\big) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \quad \text{for all } n \geqslant N.$$

Conversely, suppose $\big((x_n, y_n)\big) \longrightarrow (x, y)$. Given $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $(x_n, y_n) \in \mathbb{B}_\varepsilon\big((x, y)\big)$ for all $n \geqslant N$, therefore

$$d_X(x_n, x) + d_Y(y_n, y) = d\big((x_n, y_n), (x, y)\big) < \varepsilon.$$

Since both d_X and d_Y are non-negative, we conclude that each summand is strictly bounded by ε for all $n \geqslant N$. $\square$

Exercise 2.23 (tut03)**.** Let (x_n) be a sequence in X, let $\varphi\colon \mathbb{N} \longrightarrow \mathbb{N}$ be an injective function, and consider the sequence $(y_n) = \big(x_{\varphi(n)}\big)$ in X. Prove that if (x_n) converges to x, then so does (y_n).

Does the converse hold?

Solution. Suppose $(x_n) \longrightarrow x$. Given $\varepsilon > 0$, let $N \in \mathbb{N}$ be such that $x_n \in \mathbb{B}_\varepsilon(x)$ for all $n \geqslant N$.

Since $\varphi\colon \mathbb{N} \longrightarrow \mathbb{N}$ is injective, the inverse image $\varphi^{-1}\big(\{1, \dots, N-1\}\big)$ is a finite set, so it has a maximal element M. (If the set is empty, just take $M = 0$.) For all $n \geqslant M+1$, we have $\varphi(n) \geqslant N$, so $y_n = x_{\varphi(n)} \in \mathbb{B}_\varepsilon(x)$.

The converse certainly does not hold. For instance, take $(x_n) = (1, 0, 1, 0, 1, 0, \dots)$ and $\varphi(n) = 2n$, then the sequence $(y_n) = (0, 0, 0, \dots)$ converges to 0 but (x_n) does not converge. $\square$

Exercise 2.24 (tut03)**.**

(a) Let $f\colon X \longrightarrow Y$ be a function between two sets X and Y, and let $S \subseteq Y$. Prove that

$$f^{-1}(S) = X \smallsetminus f^{-1}(Y \smallsetminus S).$$

(b) Let $f\colon X \longrightarrow Y$ be a function between metric spaces. Prove that f is continuous if and only if: for any closed subset $C \subseteq Y$, the inverse image $f^{-1}(C) \subseteq X$ is a closed subset.

Solution.

(a) We have $x \in f^{-1}(S)$ iff $f(x) \in S$ iff $f(x) \notin \big(Y \smallsetminus S\big)$ iff $x \notin f^{-1}(Y \smallsetminus S)$.

(b) Suppose f is continuous and $C \subseteq Y$ is closed. By part (a) we have

$$f^{-1}(C) = X \smallsetminus f^{-1}(Y \smallsetminus C).$$

Then $\big(Y \smallsetminus C\big) \subseteq Y$ is open, so by Example 2.26, $f^{-1}(Y \smallsetminus C) \subseteq X$ is open, therefore $f^{-1}(C)$ is closed.

Conversely, suppose the inverse image of any closed subset is closed. Let $V \subseteq Y$ be open, then by part (a) we have

$$f^{-1}(V) = X \smallsetminus f^{-1}(Y \smallsetminus V).$$

So $\big(Y \smallsetminus V\big) \subseteq Y$ is closed, so $f^{-1}(Y \smallsetminus V) \subseteq X$ is closed, hence $f^{-1}(V)$ is open. By Example 2.26, f is continuous. $\square$

Exercise 2.25 (tut03)**.** Show that if $f\colon X \longrightarrow Y$ is a continuous map between metric spaces and $A \subseteq X$ then $f\big(\overline{A}\big) \subseteq \overline{f(A)}$.

Solution. Let $x \in \overline{A}$, let $y = f(x)$, and suppose that $y \notin \overline{f(A)}$. By Exercise 2.9 part (a), there exists an open neighbourhood $V \subseteq \big(Y \smallsetminus f(A)\big)$ with $y \in V$. As f is continuous, there exists an open neighbourhood $U \subseteq X$ of x with $f(U) \subseteq V$; as V does not intersect $f(A)$, we get that U does not intersect A, contradicting the fact that $x \in \overline{A}$. $\square$

Exercise 2.26 (tut03)**.** Give $\mathbb{N}$ the metric induced from $\mathbb{R}$. Let (X, d) be a metric space and (x_n) a sequence in X. Prove that (x_n) is a continuous function $\mathbb{N} \longrightarrow X$.

Solution. First note that the induced metric on $\mathbb{N} \subseteq \mathbb{R}$ is equivalent to the discrete metric: for any $n \in \mathbb{N}$, we have $\{n\} = (n-1, n+1) \cap \mathbb{N}$, so $\{n\}$ is open in $\mathbb{N}$. Therefore every subset of $\mathbb{N}$ is open, hence every function $\mathbb{N} \longrightarrow X$ is continuous. $\square$

Exercise 2.27 (tut03)**.**

(a) Let $f\colon X \longrightarrow Y$ and $g\colon Y \longrightarrow Z$ be functions, where X, Y, Z are sets, and let $S \subseteq Z$. Then

$$f^{-1}\big(g^{-1}(S)\big) = (g \circ f)^{-1}(S).$$

(b) Let $f\colon X \longrightarrow Y$ and $g\colon Y \longrightarrow Z$ be continuous functions, where X, Y, Z are metric spaces. Prove that $g \circ f\colon X \longrightarrow Z$ is continuous.

Solution.

(a) We have $x \in (g \circ f)^{-1}(S)$ iff $(g \circ f)(x) \in S$ iff $g(f(x)) \in S$ iff $f(x) \in g^{-1}(S)$ iff $x \in f^{-1}\big(g^{-1}(S)\big)$.

(b) Let $W \subseteq Z$ be open. As $g\colon Y \longrightarrow Z$ is continuous, $g^{-1}(W) \subseteq Y$ is open. As $f\colon X \longrightarrow Y$ is continuous, $(g \circ f)^{-1}(W) = f^{-1}\big(g^{-1}(W)\big) \subseteq X$ is open. So $g \circ f$ is continuous. $\square$

Exercise 2.28 (tut03)**.** Let $f\colon X \longrightarrow Y$ be a continuous map between metric spaces and let $S \subseteq Y$ be such that $f(X) \subseteq S$. Endowing S with the metric induced from Y, show that $f\colon X \longrightarrow S$ is continuous.

Solution. Since $f(X) \subseteq S$, we have that $f^{-1}(Y \setminus S) = \varnothing$.

Let $W \subseteq S$ be open in (the induced metric on) S, then there exists $V \subseteq Y$ open in Y such that $W = V \cap S$. Since $f\colon X \longrightarrow Y$ is continuous, we have that $U := f^{-1}(V)$ is open in X. But $f^{-1}(V) = f^{-1}(V \cap S) \cup f^{-1}(V \setminus S)$, and $f^{-1}(V \setminus S) \subseteq f^{-1}(Y \setminus S) = \varnothing$, so $f^{-1}(V) = f^{-1}(V \cap S) = f^{-1}(W)$ is open in X. $\square$

Exercise 2.29 (tut03)**.** Let $g_1\colon X \longrightarrow Y_1$ and $g_2\colon X \longrightarrow Y_2$ be continuous maps, with X, Y_1, Y_2 metric spaces.

Define $f\colon X \longrightarrow Y_1 \times Y_2$ by $f(x) = \big(g_1(x), g_2(x)\big)$. Endow $Y_1 \times Y_2$ with the Manhattan metric.

Show that f is continuous if and only if both g_1 and g_2 are continuous.

Solution. The function f is continuous iff for any sequence $(x_n) \longrightarrow x \in X$, we have $\big(f(x_n)\big) \longrightarrow f(x) \in Y_1 \times Y_2$, in other words $\big(g_1(x_n), g_2(x_n)\big) \longrightarrow \big(g_1(x), g_2(x)\big) \in Y_1 \times Y_2$. But by Exercise 2.22, the latter holds iff $\big(g_1(x_n)\big) \longrightarrow g_1(x) \in Y_1$ and $\big(g_2(x_n)\big) \longrightarrow g_2(x) \in Y_2$, which precisely says that both g_1 and g_2 are continuous. $\square$

Exercise 2.30 (tut03)**.** If A and B are subsets of a metric space (X, d), then

$$\overline{A \cup B} = \overline{A} \cup \overline{B}.$$

Solution. By Exercise 2.9 part (b), $A \subseteq A \cup B$ implies $\overline{A} \subseteq \overline{A \cup B}$, and similarly for $\overline{B} \subseteq \overline{A \cup B}$.

For the other inclusion, note that by Example 2.15, $\overline{A} \cup \overline{B}$ is a closed set containing $A \cup B$, so by the minimality of the closure Exercise 2.9, $\overline{A \cup B} \subseteq \overline{A} \cup \overline{B}$. □

Exercise 2.31 (tut03)**.** Let (X, d) be a metric space.

(a) Prove that any subset of a nowhere dense subset of X is nowhere dense in X.

(b) Prove that a subset $N \subseteq X$ is nowhere dense if and only if $X \setminus \overline{N}$ is dense in X.

(c) Prove that the union of any finite collection of nowhere dense subsets of X is nowhere dense in X.

Solution.

(a) Let $N \subseteq X$ be nowhere dense and let $M \subseteq N$. Then $\overline{M} \subseteq \overline{N}$ by Exercise 2.9 part (b), so $\left(\overline{M}\right)^{\circ} \subseteq \left(\overline{N}\right)^{\circ} = \varnothing$ by Exercise 2.3.

(b) Suppose N is nowhere dense and let $U \subseteq X$ be nonempty and open. If $U \cap \left(X \setminus \overline{N}\right) = \varnothing$, then $U \subseteq \overline{N}$, so $U \subseteq \left(\overline{N}\right)^{\circ} = \varnothing$, contradicting the non-emptiness of U. So it must be that U intersects $X \setminus \overline{N}$ nontrivially, hence $X \setminus \overline{N}$ is dense.

Conversely, suppose $X \setminus \overline{N}$ is dense but N is not nowhere dense, that is there exists a nonempty open $U \subseteq \overline{N}$. Then $U \cap \left(X \setminus \overline{N}\right) = \varnothing$, contradicting the denseness of $X \setminus \overline{N}$.

(c) It suffices to prove the case of two nowhere dense sets M and N. Let $L = M \cup N$. Then by Exercise 2.30 we have $\overline{L} = \overline{M} \cup \overline{N}$ so $X \setminus \overline{L} = \left(X \setminus \overline{M}\right) \cap \left(X \setminus \overline{N}\right)$. As $X \setminus \overline{L}$ is the intersection of two dense open subsets, it is dense and open by Exercise 2.12, hence L is nowhere dense. □

Exercise 2.32 (tut03)**.** Let X be a set.

(a) Show that the relation "d_1 is finer than d_2" on metrics on X gives rise to a relation "$[d_1]$ is finer than $[d_2]$" on equivalence classes of metrics on X.

(b) Show that the latter is a partial order on the set of equivalence classes of metrics on X.

(c) In the statement from part (b), can we remove the words "equivalence classes of"?

(d) Show that the partial order from part (b) has a unique maximal element.

Solution.

(a) First we note that the relation "is finer than" on metrics is transitive: if d_1 is finer than d_2 and d_2 is finer than d_3 then d_1 is finer than d_3. (This is clear from any of the equivalent definitions in Proposition 2.27.)

Next we show that the relation "is finer than" on equivalence classes of metrics is well-defined. Let $[d]$ denote the equivalence class of a metric d. We say that a class

$[d_1]$ is finer than a class $[d_2]$ if the metric d_1 is finer than the metric d_2. To check well-definedness of this concept, suppose that d_1' is a metric equivalent to d_1, and d_2' is a metric equivalent to d_2. Is is true that d_1' is finer than d_2'? Well, d_1' is finer than d_1, which is finer than d_2, which is finer than d_2', so the answer is yes, by transitivity.

(b) Given a class $[d]$, it is true that d is a finer metric than d, so $[d]$ is a finer class than $[d]$.

If $[d_1]$ is a finer class than $[d_2]$ and $[d_2]$ is a finer class than $[d_1]$, then d_1 is a finer metric than d_2 and d_2 is a finer metric than d_1, hence d_1 and d_2 are equivalent metrics, so $[d_1] = [d_2]$.

Finally, suppose $[d_1]$ is a finer class than $[d_2]$, which is a finer class than $[d_3]$. Then d_1 is a finer metric than d_2, which is a finer metric than d_3, so by the transitivity we saw in part (a), d_1 is a finer metric than d_3, so $[d_1]$ is a finer class than $[d_3]$.

(c) Not in general, as for metrics, d_1 finer than d_2 and d_2 finer than d_1 does not necessarily imply that $d_1 = d_2$, only that they are equivalent metrics.

(d) The unique maximal element is the equivalence class of the discrete metric on X, as it is clear that the discrete metric is finer than any metric on X. □

Exercise 2.33 (ps01)**.**

(a) Let (X, d) be a metric space with X a finite set. Prove that d is equivalent to the discrete metric on X.

(b) Let X be a set and let d be the discrete metric on X.

Is X (i) complete? (ii) compact? (iii) connected? (iv) bounded?

For each property listed, either give a proof that all discrete metric spaces X have the property, or give a specific counterexample of a discrete metric space X that does not have the property.

Solution.

(a) If X is empty, then d equals to the discrete metric vacuously and of course d is equivalent to the discrete metric. If X is not empty, then we prove that every singleton $\{x\}$ is open with respect to the metric d in two cases:

- If X has only one element, then $X = \{x\}$ is open by Example 2.8.
- If X has more than one element, then there are finitely many pairs $(x, y) \in X \times X$, so we can look at the non-empty set

$$\{d(x, y) : x, y \in X, x \neq y\}.$$

The minimum r of this set is the minimum of finitely many positive numbers, hence $r > 0$. Then $\mathbb{B}_r(x) = \{x\}$ for all $x \in X$, showing that $\{x\}$ is open for all $x \in X$.

(b)

(i) Since d is the discrete metric, $d(x, y) = 1$ iff $x \neq y$, so the only Cauchy sequences are the eventually constant sequences of the form $(x_1, \ldots, x_n, x_n, x_n, \ldots)$, which converges to $x_n \in X$. So yes, X is complete.

(ii) $\mathbb{Z}$ has the open cover
$$\mathbb{Z} = \bigcup_{m \in \mathbb{Z}} \mathbb{B}_1(m),$$
where all the open sets are disjoint, so there is no finite subcover.

In fact, a discrete metric space X is compact iff X is a finite set.

(iii) Let $X = \{x, y\}$ where $x \neq y$. Then $\{x\}$ and $\{y\}$ are open sets and express X as a nontrivial disjoint union of two open sets. So X is disconnected.

In fact, the only connected discrete metric spaces are the empty set and the singletons.

(iv) X is bounded, since $d(x, y) \leqslant 1$ for all $x, y \in X$. $\square$

Exercise 2.34 (ps01)**.** Let X be a set and let d_1, d_2 be two metrics on X.

(a) Suppose that there exist $m, M \in \mathbb{R}_{>0}$ such that

$$m\, d_1(x, y) \leqslant d_2(x, y) \leqslant M\, d_1(x, y) \qquad \text{for all } x, y \in X. \tag{2.1}$$

Show that d_1 and d_2 are equivalent.

(b) Prove that the converse of (a) does not hold.

In other words, find a set X and two equivalent metrics d_1 and d_2 with the property that there **do not** exist positive real numbers m and M such that Equation (2.1) holds.

Solution.

(a) Use Proposition 2.27. Consider an open ball $\mathbb{B}_r^{d_2}(x)$ of (X, d_2). I claim that the open ball $\mathbb{B}_{r/M}^{d_1}(x)$ of (X, d_1) is contained in $\mathbb{B}_r^{d_2}(x)$: if $y \in \mathbb{B}_{r/M}^{d_1}(x)$ then $d_1(x, y) < r/M$, so that
$$d_2(x, y) \leqslant M\, d_1(x, y) < r.$$
So d_1 is finer than d_2.

Now consider an open ball $\mathbb{B}_r^{d_1}(x)$ of (X, d_1). I claim that the open ball $\mathbb{B}_{rm}^{d_2}(x)$ of (X, d_2) is contained in $\mathbb{B}_r^{d_1}(x)$: if $y \in \mathbb{B}_{rm}^{d_2}(x)$ then $d_2(x, y) < rm$, so that
$$d_1(x, y) \leqslant \frac{1}{m}\, d_2(x, y) < r.$$
So d_2 is finer than d_1.

(b) Let $X = \mathbb{Z}$. Let d_1 be the discrete metric on $\mathbb{Z}$. Let d_2 be the induced Euclidean metric from $\mathbb{R}$, that is $d_2(x, y) = |x - y|$ for all $x, y \in \mathbb{Z}$.

First we note that d_1 and d_2 are equivalent metrics. It suffices to show that every singleton $\{x\} \subseteq \mathbb{Z}$ is open with respect to d_2:
$$\mathbb{B}_1^{d_2}(x) = \{y \in \mathbb{Z} \colon |y - x| < 1\} = \{y \in \mathbb{Z} \colon x - 1 < y < x + 1\} = \{x\}.$$

Suppose that d_1 and d_2 satisfy Equation (2.1) for some $m, M > 0$. In particular, if $x \neq y$ we would have
$$m \leqslant |x - y| \leqslant M \qquad \text{for all } x \neq y \in \mathbb{Z},$$
which is blatantly false (take $y = 0$, $x = \lceil M \rceil + 1$). $\square$

Exercise 2.35 (ps01)**.** Let X be a compact metric space and $\{C_i : i \in I\}$ be a collection of closed subsets of X such that

$$\bigcap_{j \in J} C_j \neq \varnothing \qquad \text{for every finite subset } J \subseteq I.$$

Prove that

$$\bigcap_{i \in I} C_i \neq \varnothing.$$

Give an example showing that the conclusion need not hold without the compactness condition.

Solution. Suppose that

$$\bigcap_{i \in I} C_i = \varnothing.$$

Therefore

$$X = \bigcup_{i \in I} U_i, \qquad \text{where } U_i := X \smallsetminus C_i,$$

is an open covering of X. Since X is compact, there exists a finite subset $J \subseteq I$ such that

$$X = \bigcup_{j \in J} U_j,$$

which implies that

$$\bigcup_{j \in J} C_j = \varnothing,$$

contradicting the hypothesis on the collection $\{C_i : i \in I\}$.

Here is a counterexample where X is not compact. Take $X = \mathbb{R}_{>0}$, $I = \mathbb{N}$, and $C_i = (0, 1/i]$ for $i \in I$. Then each C_i is closed in X: since both X and $(1/i, \infty)$ are open in $\mathbb{R}$, we conclude that $X \smallsetminus C_i = X \cap (1/i, \infty)$ is open in X.

Also,

$$\bigcap_{i \in I} C_i = \varnothing,$$

because if $x \in \mathbb{R}_{>0}$ is in C_i for all $i \in I$, then $0 < x \leqslant 1/i$ for all i in I, hence $0 < x \leqslant 0$ by taking limits as $i \longrightarrow \infty$, contradiction.

If $J \subseteq I$ is finite, let $m = \max\{J\}$, then

$$\bigcap_{j \in J} C_j = C_m \neq \varnothing. \qquad \square$$

Exercise 2.36 (ps01)**.** Let (X, d) be a metric space and define $d' : X \times X \longrightarrow \mathbb{R}_{\geqslant 0}$ by

$$d'(x, y) = \min\{d(x, y), 1\}.$$

Prove that d' is a metric on X and that d' is equivalent to d.

Solution. It is clear that $d'(y, x) = d'(x, y)$ and that $d'(x, y) = 0$ if and only if $d(x, y) = 0$ if and only if $x = y$.

For the triangle inequality: $d'(x, y) \leqslant 1$ so if at least one of $d'(x, t)$, $d'(t, y)$ is 1, the triangle inequality holds. So we may assume that $d'(x, t) = d(x, t)$ and $d'(t, y) = d(t, y)$. Then

$$d'(x, y) \leqslant d(x, y) \leqslant d(x, t) + d(t, y) = d'(x, t) + d'(t, y).$$

It remains to prove the equivalence of d and d'. Let $x \in X$ and $s \leqslant 1$.
I claim that $\mathbb{B}_s^d(x) = \mathbb{B}_s^{d'}(x)$. To see this, let $y \in \mathbb{B}_s^d(x)$, then $d(x,y) < s \leqslant 1$, so

$$d'(x,y) = \min\{d(x,y),1)\} = d(x,y) < s.$$

In the other direction, let $y \in \mathbb{B}_s^{d'}(x)$, then

$$\min\{d(x,y),1\} = d'(x,y) < s \leqslant 1,$$

which forces $d(x,y) = d'(x,y) < s$.

We conclude by noting that for any $r > 0$, if we set $s = \min\{r,1\}$ we get $\mathbb{B}_s^d(x) = \mathbb{B}_s^{d'}(x) \subseteq \mathbb{B}_r^{d'}(x)$, and $\mathbb{B}_s^{d'}(x) = \mathbb{B}_s^d(x) \subseteq \mathbb{B}_r^d(x)$. In other words, any d'-open ball contains a d-open ball, and vice-versa. □

Exercise 2.37 (ps01)**.** Let (X,d) be a metric space.

(a) Fix an arbitrary element $y \in X$ and consider the function $f\colon X \longrightarrow \mathbb{R}$ given by $f(x) = d(x,y)$. Prove that f is uniformly continuous.

(b) Give $X \times X$ any conserving metric D coming from d. Prove that $d\colon X \times X \longrightarrow \mathbb{R}$ is uniformly continuous (with respect to D).

(c) Let d' be a metric on X and put on $X \times X$ any conserving metric D' coming from d'. Suppose that $d\colon X \times X \longrightarrow \mathbb{R}$ is continuous with respect to D'. Prove that d' is a finer metric than d.

Solution.

(a) Let $\varepsilon > 0$. Set $\delta = \varepsilon$. If $x, x' \in X$ satisfy $d(x,x') < \delta = \varepsilon$, then

$$|f(x) - f(x')| = |d(x,y) - d(x',y)| \leqslant d(x,x') < \varepsilon.$$

(b) Let $\varepsilon > 0$. Set $\delta = \varepsilon/2$. If $(x_1,x_2),(x_1',x_2') \in X \times X$ satisfy

$$\max\{d(x_1,x_1'),d(x_2,x_2')\} \leqslant D((x_1,x_2),(x_1',x_2')) < \delta = \frac{\varepsilon}{2},$$

(where we used the fact that D is conserving), then

$$|d(x_1,x_2) - d(x_1',x_2')| \leqslant d(x_1,x_1') + d(x_2,x_2') < \varepsilon,$$

where the first inequality is obtained by applying the triangle inequality a couple of times, as in Example 2.37.

(c) We prove that if $(x_n) \longrightarrow x \in X$ with respect to d', then $(x_n) \longrightarrow x$ with respect to x.

Suppose $(x_n) \longrightarrow x \in X$ with respect to d'. Then $\big((x_n,x)\big) \longrightarrow (x,x) \in X \times X$ with respect to D'. But $d\colon X \times X \longrightarrow \mathbb{R}$ is continuous with respect to D', so $(d(x_n,x)) \longrightarrow d(x,x) = 0 \in \mathbb{R}$. Therefore $(x_n) \longrightarrow x \in X$ with respect to d. □

Exercise 2.38 (ps01)**.** Give $\mathbb{Q} \subseteq \mathbb{R}$ the induced metric and consider the sequence (x_n)

defined recursively by

$$x_1 = 1, \qquad x_{n+1} = \frac{x_n}{2} + \frac{1}{x_n} \quad \text{for } n \in \mathbb{N}.$$

(a) Prove that $1 \leqslant x_n \leqslant 2$ for all $n \in \mathbb{N}$ and breathe a sigh of relief that the recursive definition does not accidentally divide by 0.

(b) For $n \in \mathbb{N}$, let $y_n = x_{n+1} - x_n$. Prove that

$$y_{n+1} = -\frac{y_n^2}{2x_{n+1}} \qquad \text{for all } n \in \mathbb{N}.$$

(c) Prove that

$$|y_n| \leqslant \frac{1}{2^n} \qquad \text{for all } n \in \mathbb{N}.$$

(d) Show that (x_n) is Cauchy.

(e) Consider the function $f\colon [1,2] \longrightarrow [1,2]$ given by

$$f(x) = \frac{x}{2} + \frac{1}{x}.$$

Prove that f is a contraction. What is the fixed point of f?

Solution.

(a) Induction on n. Base case $x_1 = 1$ clear.

Fix $n \in \mathbb{N}$ and suppose $1 \leqslant x_n \leqslant 2$. Then

$$\frac{1}{2} \leqslant \frac{x_n}{2} \leqslant 1 \qquad \text{and} \qquad \frac{1}{2} \leqslant \frac{1}{x_n} \leqslant 1,$$

so $1 \leqslant x_{n+1} \leqslant 2$.

(b) Fix $n \in \mathbb{N}$. Noting that $2x_n x_{n+1} = x_n^2 + 2$, we have

$$y_n^2 = \left(x_{n+1} - x_n\right)^2 = x_{n+1}^2 - 2x_{n+1}x_n + x_n^2 = x_{n+1}^2 - 2$$
$$2x_{n+1}y_{n+1} = 2x_{n+1}\left(\frac{1}{x_{n+1}} - \frac{x_{n+1}}{2}\right) = 2 - x_{n+1}^2 = -y_n^2.$$

(c) From part (b) we have

$$|y_{n+1}| = \frac{|y_n|^2}{2x_{n+1}} \qquad \text{for all } n \in \mathbb{N}.$$

We can use this, part (a), and induction by n.

For the base case we have $y_1 = \frac{1}{2}$.

For the induction step, fix $n \in \mathbb{N}$ and suppose $|y_n| \leqslant \frac{1}{2^n}$, then

$$|y_{n+1}| = \frac{|y_n|^2}{2x_{n+1}} \leqslant \frac{|y_n|^2}{2} \leqslant \frac{1}{2^{2n+1}} \leqslant \frac{1}{2^{n+1}}.$$

(d) Let $\varepsilon > 0$ and let $N \in \mathbb{N}$ be such that $2^N > 1/\varepsilon$. If $n \geqslant m \geqslant N$ then

$$\begin{aligned}
|x_n - x_m| &= |y_{n-1} + y_{n-2} + \cdots + y_{m+1}| \\
&\leqslant |y_{n-1}| + \cdots + |y_{m+1}| \\
&\leqslant \frac{1}{2^{n-1}} + \cdots + \frac{1}{2^{m+1}} \\
&= \left(\frac{1}{2^{n-m-2}} + \frac{1}{2^{n-m-3}} + \cdots + 1 \right) \frac{1}{2^{m+1}} \\
&\leqslant \frac{2}{2^{m+1}} \leqslant \frac{1}{2^N} < \varepsilon.
\end{aligned}$$

Here we used the fact that the geometric series with ratio $1/2$ sums up to 2.

(e) Let $x_1, x_2 \in [1,2]$. The function f is continuous on $[x_1, x_2]$ and differentiable on (x_1, x_2), so there exists $\xi \in (x_1, x_2)$ such that

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} = f'(\xi),$$

from which we deduce that

$$|f(x_2) - f(x_1)| = |f'(\xi)|\,|x_2 - x_1|.$$

But since $\xi \in (1,2)$ we have

$$1 < \xi < 2 \Rightarrow \frac{1}{4} < \frac{1}{\xi^2} < 1 \Rightarrow -\frac{1}{2} < f'(\xi) < \frac{1}{2}.$$

We conclude that f is a contraction with constant $1/2$.

Since f is a contraction and $[1,2]$ is complete, we know that f has a unique fixed point, which is precisely the limit of the sequence (x_n) defined above. We can find it explicitly as

$$x = f(x) = \frac{x}{2} + \frac{1}{x} \Rightarrow x^2 = 2,$$

and since $x \in [1,2]$ we get $x = \sqrt{2}$. □

Exercise 2.39 (ps01). Let $\mathbb{S}^1 = \mathbb{S}_1\big((0,0)\big) = \big\{x, y \in \mathbb{R} \colon x^2 + y^2 = 1\big\}$ be the unit circle in $\mathbb{R}^2$.
Consider the function $f \colon [0,1) \longrightarrow \mathbb{S}^1$ given by the parametrisation

$$f(t) = \big(\cos(2\pi t), \sin(2\pi t)\big).$$

Endow $[0,1)$ with the induced metric from $\mathbb{R}$ and $\mathbb{S}^1$ with the induced metric from $\mathbb{R}^2$.
Prove that f is a bijective continuous function, but not a homeomorphism.
(You may use without proof whatever properties of the functions sin and cos you manage to remember from previous subjects.)

Solution.

(a) We know that $t \longmapsto 2\pi t$, $t \longmapsto \cos(t)$ and $t \longmapsto \sin(t)$ are continuous, so by Exercise 2.29 so is f.

(b) Suppose $t_1 \neq t_2 \in [0,1)$ are such that $f(t_1) = f(t_2)$. Then $\cos(2\pi t_1) = \cos(2\pi t_2)$, which implies that $t_2 = 1 - t_1$. In that case $\sin(2\pi t_2) = \sin(2\pi - 2\pi t_1) = \sin(-2\pi t_1) = -\sin(2\pi t_1)$. But we also have $\sin(2\pi t_2) = \sin(2\pi t_1)$, so $\sin(2\pi t_1) = 0$, hence $t_1 = 0$ and $t_2 = 1 - t_1 = 1$, contradicting $t_2 \in [0,1)$.

We conclude that f is injective.

For surjectivity, let $(x,y) \in \mathbb{S}^1$, in other words $x^2 + y^2 = 1$. Define $\theta \in [0, 2\pi)$ by

$$\theta = \begin{cases} \arccos(x) & \text{if } y \geqslant 0 \\ 2\pi - \arccos(x) & \text{if } y < 0. \end{cases}$$

Letting $t = \theta/(2\pi)$, we have $f(t) = (x,y)$.

(c) At this point we know that f is a homeomorphism iff $f^{-1}\colon \mathbb{S}^1 \longrightarrow [0,1)$ is continuous. Note that $\mathbb{S}^1 \subseteq \mathbb{R}^2$ is compact: it is clearly bounded as any two points are at distance at most 2 of each other, so we just need to check that it is a closed subset of $\mathbb{R}^2$.

But $\mathbb{S}^1 = \mathbb{D}_1\big((0,0)\big) \cap C$ is the intersection of two closed sets, where

$$C = \{x, y \in \mathbb{R} \colon x^2 + y^2 \geqslant 1\} = \mathbb{R}^2 \smallsetminus \mathbb{B}_1\big((0,0)\big).$$

Since $\mathbb{S}^1$ is compact, if f^{-1} were continuous then $[0,1) = f^{-1}(\mathbb{S}^1)$ would be compact, hence closed in $\mathbb{R}$. This is a contradiction, because 1 is an accumulation point of $[0,1)$ but does not lie in the set. $\square$

Exercise 2.40 (ps01)**.** Let X be a metric space and Y a complete metric space. Let $D \subseteq X$ be a dense subset and $f\colon D \longrightarrow Y$ a uniformly continuous function.

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

$$\widehat{f}\colon X \longrightarrow Y \quad \text{such that} \quad \widehat{f}(u) = f(u) \quad \text{for all } u \in D.$$

(Make sure you give a complete argument: how do you construct $\widehat{f}$? is it well-defined? does it extend f? why is it uniformly continuous? why is it unique?)

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

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

Solution.

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

Define $\widehat{f}(x) = y$.

But wait, is this actually **well-defined**? We did make one choice in the construction, namely a sequence (u_n) in D that converges to x. Any other valid choice is a sequence

(u'_n) in D with the same limit x, so $(u'_n) \sim (u_n)$. As f is continuous, we have $\big(f(u'_n)\big) \sim \big(f(u_n)\big)$, which implies that $\big(f(u'_n)\big) \longrightarrow y \in Y$.

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

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

$$d_X(u_n, u'_n) \leqslant d_X(u_n, x) + d_X(x, x') + d_X(x', u'_n) < \delta.$$

Therefore $d_Y(f(u_n), f(u'_n)) < \varepsilon/2$ for all $n \geqslant M$.

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

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

The **uniqueness** of $\widehat{f}$ follows from Example 2.43, which says that there is at most one continuous extension.

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

$$\begin{aligned} d_Y\big(\widehat{f}(x), \widehat{f}(x')\big) &= d_Y\left(\lim_{n\longrightarrow\infty} \widehat{f}(u_n), \lim_{n\longrightarrow\infty} \widehat{f}(u'_n)\right) \\ &= \lim_{n\longrightarrow\infty} d_Y(f(u_n), f(u'_n)) = \lim_{n\longrightarrow\infty} d_X(u_n, u'_n) = d_X(x, x'). \end{aligned}$$

(c) For the case of completions, let $D = \iota(X) \subseteq \widehat{X}$, and apply the above to the function $\iota_Y \circ g \circ \iota_X^{-1}\colon D \longrightarrow \widehat{Y}$. □

Exercise 2.41 (tut04)**.** Let (X, d_X) and (Y, d_Y) be metric spaces and let d be a conserving metric on $X \times Y$.

(a) Prove that the sequence $\big((x_n, y_n)\big)$ is Cauchy in $X \times Y$ if and only if (x_n) is Cauchy in X and (y_n) is Cauchy in Y.

(b) Prove that if X and Y are complete then $X \times Y$ is complete. Is the converse true?

Solution.

(a) Suppose $\big((x_n, y_n)\big)$ is a Cauchy sequence in $(X \times Y, d)$. Fix $\varepsilon > 0$. There exists $N \in \mathbb{N}$ such that for all $m, n \geqslant N$ we have

$$d_X(x_m, x_n) \leqslant \max\big\{d_X(x_m, x_n), d_Y(y_m, y_n)\big\} \leqslant d\big((x_m, y_m), (x_n, y_n)\big) < \varepsilon,$$

so (x_n) is Cauchy in X. Similarly, (y_n) is Cauchy in Y.

Conversely, suppose (x_n) is Cauchy in X and (y_n) is Cauchy in Y. Fix $\varepsilon > 0$. Let $N_x \in \mathbb{N}$ be such that for all $m, n \geqslant N_x$ we have $d_X(x_m, x_n) < \varepsilon/2$. Let $N_y \in \mathbb{N}$ be such that for all $m, n \geqslant N_y$ we have $d_Y(y_m, y_n) < \varepsilon/2$. Let $N = \max\{N_x, N_y\}$, then for all $m, n \geqslant N$ we have

$$d\big((x_m, y_m), (x_n, y_n)\big) \leqslant d_X(x_m, x_n) + d_Y(y_m, y_n) < \varepsilon,$$

so $\big((x_n, y_n)\big)$ is Cauchy in $X \times Y$.

(b) Let $\big((x_n, y_n)\big)$ be a Cauchy sequence in $X \times Y$. By part (a), (x_n) is Cauchy in X and (y_n) is Cauchy in Y. Since X and Y are complete, we have $(x_n) \longrightarrow x \in X$ and $(y_n) \longrightarrow y \in Y$. By Exercise 2.22, $\big((x_n, y_n)\big) \longrightarrow (x, y) \in X \times Y$.

The converse also holds: suppose $X \times Y$ is complete. Let (x_n) be a Cauchy sequence in X, and fix some $y \in Y$. Then by (a) we have that $\big((x_n, y)\big)$ is Cauchy in $X \times Y$, so $\big((x_n, y)\big) \longrightarrow (x, y) \in X \times Y$, which by Exercise 2.22 implies that $(x_n) \longrightarrow x \in X$. The same proof gives us that Y is complete. □

Exercise 2.42 (tut04)**.** Any distance-preserving function is uniformly continuous.

Solution. This is immediate from the definitions (can take $\delta = \varepsilon$). □

Exercise 2.43 (tut04)**.** Check (directly from the definition of uniform continuity) that $f\colon \mathbb{R}_{>0} \longrightarrow \mathbb{R}_{>0}$ 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, x' such that $x' \in \mathbb{B}_\delta(x)$ and $f(x') \notin \mathbb{B}_\varepsilon\big(f(x)\big)$.

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

$$|x - x'| = \frac{x}{2} \leqslant \frac{\delta}{2} < \delta,$$

so indeed $x' \in \mathbb{B}_\delta(x)$.

Also

$$|f(x) - f(x')| = \left|\frac{1}{x} - \frac{1}{x'}\right| = \left|\frac{1}{x} - \frac{2}{x}\right| = \frac{1}{x} \geqslant 1 = \varepsilon,$$

so indeed $f(x') \notin \mathbb{B}_\varepsilon\big(f(x)\big)$. □

Exercise 2.44 (tut04)**.** Let $f\colon X \longrightarrow Y$ be a uniformly continuous function between two metric spaces and suppose $(x_n) \sim (x'_n)$ are equivalent sequences in X. Prove that $\big(f(x_n)\big) \sim \big(f(x'_n)\big)$ as sequences in Y.

Does the conclusion hold if f is only assumed to be continuous?

Solution. 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 $d_Y(f(x), f(x')) < \varepsilon$. As $(x_n) \sim (x'_n)$, there exists $N \in \mathbb{N}$ such that $d_X(x_n, x'_n) < \delta$ for all $n \geqslant N$. Hence for all $n \geqslant N$ we have $d_Y(f(x_n), f(x'_n)) < \varepsilon$.

The result does not hold in general for continuous functions; for instance one can take $f\colon \mathbb{R}_{>0} \longrightarrow \mathbb{R}_{>0}$ given by $f(x) = \frac{1}{x}$, and $(1/n) \sim (1/n^2)$ but $(f(1/n)) = (n)$, $(f(1/n^2)) = (n^2)$ and $(n) \not\sim (n^2)$. □

Exercise 2.45 (tut04)**.** Let (X, d_X) and (Y, d_Y) be metric spaces and $f\colon X \longrightarrow Y$ a surjective continuous function. Suppose that X is complete and for all $x_1, x_2 \in X$ we have

$$d_X\big(x_1, x_2\big) \leqslant d_Y\big(f(x_1), f(x_2)\big).$$

(a) Prove that Y is complete.

In particular, distance-preserving maps preserve completeness.

(b) Do you feel strongly that uniformly continuous functions ought to preserve completeness? (After all, they preserve Cauchy sequences, and completeness is defined in terms of Cauchy sequences.)

Prove that $f\colon \mathbb{R} \longrightarrow (-\pi/2, \pi/2)$ given by $f(x) = \arctan(x)$ is uniformly continuous, but. . .

Solution.

(a) Let (y_n) be a Cauchy sequence in Y. For each $n \in \mathbb{N}$, let $x_n \in f^{-1}(y_n)$. I claim that (x_n) is a Cauchy sequence in X. Fix $\varepsilon > 0$. Let $N \in \mathbb{N}$ be such that for all $m, n \geqslant N$ we have $d_Y(y_m, y_n) < \varepsilon$. Then for all $m, n \geqslant N$ we have

$$d_X(x_m, x_n) \leqslant d_Y\big(f(x_m), f(x_n)\big) = d_Y(y_m, y_n) < \varepsilon,$$

so (x_n) is indeed Cauchy in X.

Since X is complete, we have $(x_n) \longrightarrow x \in X$, so that by the continuity of f we conclude that $(y_n) = \big(f(x_n)\big) \longrightarrow f(x) \in Y$.

(b) Given $x_1 < x_2$, apply the Mean Value Theorem to $f(x) = \arctan(x)$ on $[x_1, x_2]$ to get some $\xi \in (x_1, x_2)$ such that

$$|f(x_2) - f(x_1)| = |f'(\xi)|\,|x_2 - x_1| = \frac{1}{1+\xi^2}\,|x_2 - x_1| \leqslant |x_2 - x_1|.$$

So for any $\varepsilon > 0$ we can take $\delta = \varepsilon$ and conclude that f is uniformly continuous.

It is also surjective onto $(-\pi/2, \pi/2)$, but the latter is of course not complete. $\square$

Exercise 2.46 (tut04)**.** Any Cauchy sequence (x_n) is *bounded*, that is there exists $C \geqslant 0$ such that $d(x_n, x_m) \leqslant C$ for all $n, m \in \mathbb{N}$.

Solution. Let $N \in \mathbb{N}$ be such that for all $m, n \geqslant N$ we have $d(x_m, x_n) < 1$.
Let $B = \max\{d(x_m, x_N)\colon 1 \leqslant m < N\}$. Let $C = 2B + 1$, then we have

$$d(x_m, x_n) \leqslant \begin{cases} 1 \leqslant C & \text{if } m, n \geqslant N \\ d(x_m, x_N) + d(x_N, x_n) \leqslant B + 1 \leqslant C & \text{if } m < N, n \geqslant N \\ d(x_m, x_N) + d(x_N, x_n) \leqslant 2B \leqslant C & \text{if } m, n < N. \end{cases} \quad \square$$

Exercise 2.47 (tut04)**.** Suppose A and B are abelian groups. A function $f\colon A \longrightarrow B$ is called *additive* if $f(a+b) = f(a) + f(b)$.

(a) Prove that every additive function $f\colon A \longrightarrow B$ satisfies

$$f(0) = 0 \qquad \text{and} \qquad f(-a) = -f(a).$$

(b) Let V be a $\mathbb{Q}$-vector space. Prove that every additive function $f\colon \mathbb{Q} \longrightarrow V$ is $\mathbb{Q}$-linear.

(c) What can you say (and prove) about **continuous** additive functions $\mathbb{R} \longrightarrow \mathbb{R}$?

(d) Suppose that $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ is additive and continuous at 0. Prove that f is continuous on $\mathbb{R}$, and conclude that f is $\mathbb{R}$-linear.

(e) Let B be a basis for $\mathbb{R}$ as a $\mathbb{Q}$-vector space. (Recall from Exercise 1.5 that B is uncountable.) Use two distinct irrational elements of B to construct a $\mathbb{Q}$-linear transformation $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ that is not $\mathbb{R}$-linear.

If you would (and why wouldn't you?), follow the rabbit:

`https://en.wikipedia.org/wiki/Cauchy%27s_functional_equation`

Solution.

(a) $f(0) = f(0+0) = f(0) + f(0)$ so $f(0) = 0$.
$f(-a) + f(a) = f(-a+a) = f(0) = 0$.

(b) Let $v = f(1) \in V$.
For $n \in \mathbb{N}$ we have

$$f(n) = f(1+1+\cdots+1) = f(1) + \cdots + f(1) = nv.$$

For $m \in \mathbb{N}$ we have

$$v = f(1) = f\left(\frac{1}{m} + \cdots + \frac{1}{m}\right) = mf\left(\frac{1}{m}\right),$$

so $f(1/m) = (1/m)\,v$.
Therefore, for any $n, m \in \mathbb{N}$ we have

$$f\left(\frac{n}{m}\right) = nf\left(\frac{1}{m}\right) = \frac{n}{m}\,v.$$

Combining this with $f(-a) = -f(a)$ and $f(0) = 0$, we conclude that $f(x) = xv = xf(1)$ for all $x \in \mathbb{Q}$.

(c) Let $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ be additive. Let $g\colon \mathbb{Q} \longrightarrow \mathbb{R}$ be the restriction of f to $\mathbb{Q} \subseteq \mathbb{R}$. Let $a = g(1) = f(1)$.
By part (b), $g(q) = q\,g(1) = qa$ for all $q \in \mathbb{Q}$. Let $x \in \mathbb{R}$. As $\mathbb{Q}$ is dense in $\mathbb{R}$, there is some sequence $(q_n) \longrightarrow x$ with $q_n \in \mathbb{Q}$; since f is continuous we have

$$f(x) = f\left(\lim_{n\longrightarrow\infty} q_n\right) = \lim_{n\longrightarrow\infty} f(q_n) = \lim_{n\longrightarrow\infty} g(q_n) = \lim_{n\longrightarrow\infty} (q_n a) = xa = xf(1).$$

Hence f is $\mathbb{R}$-linear.

(d) Let $x \in \mathbb{R}$. Fix $\varepsilon > 0$. Let $\delta > 0$ be such that if $|t| < \delta$, then $|f(t)| < \varepsilon$.

Suppose $x' \in \mathbb{R}$ is such that $|x - x'| < \delta$, then

$$|f(x) - f(x')| = |f(x - x')| < \varepsilon.$$

So f is continuous at every $x \in \mathbb{R}$, so by part (c) f is $\mathbb{R}$-linear.

(e) Let B be a $\mathbb{Q}$-basis for $\mathbb{R}$. Exactly one element of B is a nonzero rational, and without loss of generality we may assume it is 1. Since B is uncountable, it also contains uncountably many irrationals. Let $b, c \in B \cap \big(\mathbb{R} \smallsetminus \mathbb{Q}\big)$. Consider the bijective function $\sigma\colon B \longrightarrow B$ given by

$$\sigma(b) = c, \qquad \sigma(c) = b, \qquad \sigma(x) = x \text{ for all } x \in B \smallsetminus \{b, c\}.$$

Since B is a $\mathbb{Q}$-basis of $\mathbb{R}$, σ extends by $\mathbb{Q}$-linearity to a $\mathbb{Q}$-linear transformation $f\colon \mathbb{R} \longrightarrow \mathbb{R}$. In particular, f is additive.

Suppose that f is $\mathbb{R}$-linear, then:

$$c = f(b) = bf(1) = b1 = b,$$

contradicting the fact that $b \neq c$. ☐

Exercise 2.48 (tut05)**.** Let A and C be connected subsets of a metric space (X, d). Prove that if $A \cap C \neq \varnothing$, then $A \cup C$ is connected.

Solution. Suppose $A \cup C$ is disconnected, so that $A \cup C = U \cup V$ with U, V nonempty, disjoint, and open in $A \cup C$.

Then $A = \big(A \cap U\big) \cup \big(A \cap V\big)$, with $A \cap U, A \cap V$ disjoint and open in A. As A is connected, $A \cap U$ or $A \cap V$ must be empty. Without loss of generality, say $A \cap U = \varnothing$, so that $A \subseteq V$.

We can apply the same argument to C and get that $C \cap U$ or $C \cap V$ is empty. Since $A \cap C \neq \varnothing$, it must be that $C \cap U = \varnothing$ and $C \subseteq V$.

But then $U \cup V = A \cup C \subseteq V$, implying that $U \subseteq V$, contradicting the fact that $U \cap V = \varnothing$ and $U \neq \varnothing$. ☐

We can generalise this in two different ways:

Exercise 2.49 (tut05)**.** Let (X, d) be a metric space. Suppose $A \subseteq X$ is a connected subset and $\{C_i \colon i \in I\}$ is an arbitrary collection of connected subsets of X such that $A \cap C_i \neq \varnothing$ for all $i \in I$. Then

$$A \cup \bigcup_{i \in I} C_i$$

is a connected subset of X.

[*Hint*: Use the argument from Exercise 2.48.]

Solution. Suppose

$$A \cup \bigcup_{i \in I} C_i = U \cup V,$$

with U, V nonempty disjoint open sets. Then $A = \big(A \cap U\big) \cup \big(A \cap V\big)$, but A is connected so one of these intersections must be empty, say $A \cap U = \varnothing$.

But $U \subseteq A \cup \bigcup_{i \in I} C_i$, so there must be some $i \in I$ such that $U \cap C_i \neq \varnothing$ (otherwise $U = \varnothing$, contradiction). Since $C_i = \big(C_i \cap U\big) \cup \big(C_i \cap V\big)$ and C_i is connected, we must have $C_i \cap V = \varnothing$, therefore $C_i \subseteq U$.

But this forces $A \cap C_i \subseteq A \cap U = \varnothing$, contradicting $A \cap C_i \neq \varnothing$ for all $i \in I$. □

Exercise 2.50 (tut05)**.** Let (X, d) be a metric space. Suppose $\{C_n : n \in \mathbb{N}\}$ is a countable collection of connected subsets of X such that $C_n \cap C_{n+1} \neq \varnothing$ for all $n \in \mathbb{N}$. Then

$$\bigcup_{n \in \mathbb{N}} C_n$$

is a connected subset of X.

[*Hint*: Build the union inductively, and use Exercises 2.48 and 2.49.]

Solution. For any $N \in \mathbb{N}$, let

$$A_N = \bigcup_{n=1}^{N} C_n.$$

We use induction to prove that A_N is connected for all $N \in \mathbb{N}$.

The base case $N = 1$ is clear as $A_1 = C_1$.

For the induction step, fix $N \in \mathbb{N}$ and suppose A_N is connected. Then $A_{N+1} = A_N \cup C_{N+1}$ is connected by Exercise 2.48.

So $\{A_N : N \in \mathbb{N}\}$ is a collection of connected sets, and A_1 is a connected set such that $A_1 \cap A_N \neq \varnothing$ for all $N \in \mathbb{N}$. By Exercise 2.49,

$$A_1 \cup \bigcup_{N \in \mathbb{N}} A_N = \bigcup_{N \in \mathbb{N}} A_N = \bigcup_{n \in \mathbb{N}} C_n$$

is connected. □

Exercise 2.51 (tut05)**.** Let (X, d) be a metric space and define $x \sim x'$ if there exists a connected subset $C \subset X$ such that $x, x' \in C$.

Prove that this is an equivalence relation on the set X, thereby partitioning X into a disjoint union of maximal connected subsets (these are called the *connected components* of X).

[*Hint*: Recall that an equivalence relation has three defining axioms: (a) $x \sim x$ for all $x \in X$; (b) if $x \sim x'$ then $x' \sim x$; (c) if $x \sim x'$ and $x' \sim x''$ then $x \sim x''$.]

Solution.

(a) $x \sim x$: for any $x \in X$, the set $C = \{x\}$ is connected and contains x, so $x \sim x$.

(b) if $x \sim x'$ then $x' \sim x$: clear from the definition, which does not distinguish x and x'.

(c) if $x \sim x'$ and $x' \sim x''$ then $x \sim x''$: since $x \sim x'$ there exists a connected set C_1 such that $x, x' \in C_1$; since $x' \sim x''$ there exists a connected set C_2 such that $x', x'' \in C_2$; by Exercise 2.48, since C_1 and C_2 are connected and $x' \in C_1 \cap C_2$, the union $C_1 \cup C_2$ is connected, and it contains both x and x'', so that $x \sim x''$. □

Exercise 2.52 (tut05)**.** Give explicit continuous surjective functions $f\colon \mathbb{R} \longrightarrow I$, where I is:

(a) $\mathbb{R}$	(b) $(0,\infty)$	(c) $(-\infty,0)$	(d) $(-\infty,0]$	(e) $[-1,1]$
(f) $(0,1]$	(g) $[0,1)$	(h) $(-\pi/2,\pi/2)$	(i) $\{0\}$.	

[*Hint*: Draw some functions you know from calculus and see what their ranges are.]

Solution. These are of course not the only possible answers (well, except for the last one).

(a) $x \longmapsto x$;

(b) $x \longmapsto e^x$;

(c) $x \longmapsto -e^x$;

(d) $x \longmapsto -x^2$;

(e) $x \longmapsto \sin(x)$;

(f) $x \longmapsto \min\{e^x, 1\}$;

(g) $x \longmapsto \max\{-e^x, -1\} + 1$;

(h) $x \longmapsto \arctan(x)$;

(i) $x \longmapsto 0$. □

Exercise 2.53 (tut05)**.** Let (X,d) be a metric space.

If A and B are bounded sets with $A \cap B \neq \varnothing$, then

$$\operatorname{diam}(A \cup B) \leqslant \operatorname{diam}(A) + \operatorname{diam}(B).$$

Solution. It suffices to show that for any $x, y \in A \cup B$ we have

$$d(x,y) \leqslant \operatorname{diam}(A) + \operatorname{diam}(B).$$

If $x, y \in A$, this is obvious as $d(x,y) \leqslant \operatorname{diam}(A)$. Similarly if $x, y \in B$.

It remains to see what happens if $x \in A$ and $y \in B$. Let $t \in A \cap B$. We have

$$d(x,y) \leqslant d(x,t) + d(t,y) \leqslant \operatorname{diam}(A) + \operatorname{diam}(B),$$

as desired. □

Exercise 2.54 (tut05)**.** Let C be a closed subset of a compact subset K of a metric space (X,d). Prove that C is compact.

[*Hint*: $K \subseteq X = C \cup \big(X \smallsetminus C\big)$.]

Solution. Consider an arbitrary open cover of C:

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

Then we have

$$K \subseteq X = C \cup \big(X \smallsetminus C\big) \subseteq \left(\bigcup_{i \in I} U_i\right) \cup \big(X \smallsetminus C\big),$$

which is an open cover of K. As K is compact, there is a finite subcover, so that

$$K \subseteq \left(\bigcup_{n=1}^{N} U_{i_n} \right) \cup \big(X \setminus C \big), \qquad i_n \in I,$$

hence

$$C \subseteq \bigcup_{n=1}^{N} U_{i_n}. \qquad \square$$

Exercise 2.55 (tut05)**.** Let K and L be compact subsets of a metric space (X, d). Prove that $K \cup L$ is compact.

Solution. Consider an arbitrary open cover of $K \cup L$:

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

This is also an open cover of K, so there is a finite subcover that still covers K:

$$K \subseteq \bigcup_{n=1}^{N} U_{i_n}, \qquad i_n \in I.$$

Similarly, we get a finite subcover that covers L:

$$L \subseteq \bigcup_{m=1}^{M} U_{j_m}, \qquad j_m \in I.$$

Letting $S = \{i_1, \dots, i_N\} \cup \{j_1, \dots, j_M\}$, we get a finite subcover that covers $K \cup L$:

$$K \cup L \subseteq \bigcup_{s \in S} U_s. \qquad \square$$

Exercise 2.56 (tut06)**.** Let (X, d_X) and (Y, d_Y) be metric spaces, and let d be any conserving metric on $X \times Y$.

(a) Prove that if X and Y are compact, then $X \times Y$ is compact.

[*Hint*: If you're not sure where to start, try sequential compactness.]

(b) Does the converse hold?

Solution.

(a) Suppose (x_n, y_n) is a sequence in $X \times Y$. Then (x_n) is a sequence in X, and since X is compact, it follows that (x_n) has some converging subsequence $\left(x_{n_k}\right) \longrightarrow x \in X$. Now consider the sequence $\left(y_{n_k}\right)$ in Y. Since Y is compact, it follows that $\left(y_{n_k}\right)$ has some converging subsequence $\left(y_{n_{k_j}}\right) \longrightarrow y \in Y$. Then $\left(x_{n_{k_j}}\right)$ is a subsequence of the converging sequence $\left(x_{n_k}\right) \longrightarrow x \in X$, so that it is itself converging to $x \in X$. We conclude that $\left(x_{n_{k_j}}, y_{n_{k_j}}\right) \longrightarrow (x, y) \in X \times Y$ and is a subsequence of the original sequence (x_n, y_n).

(b) The converse does hold, since the projection maps $\pi_1 \colon X \times Y \longrightarrow X$, $\pi_1(x, y) = x$, and $\pi_2 \colon X \times Y \longrightarrow Y$, $\pi_2(x, y) = y$, are continuous and surjective. $\square$

Exercise 2.57 (tut06)**.** Let C be a nonempty compact subset of a metric space (X, d). Prove that there exist points $a, b \in C$ such that

$$d(a, b) = \sup\left\{d(x, y) : x, y \in C\right\}.$$

In other words, the diameter of C is realised as the distance between two points of C.

Solution. As you know from Assignment 1 Question 5 (Exercise 2.37), the distance function $d : X \times X \longrightarrow \mathbb{R}$ is continuous. By Exercise 2.56, $C \times C$ is compact, so by Example 2.73 there exists $(a_{\max}, b_{\max}) \in C \times C$ such that

$$d(a, b) \leqslant d(a_{\max}, b_{\max}) \qquad \text{for all } (a, b) \in C \times C.$$

Therefore $a_{\max}, b_{\max} \in C$ realise the diameter of C. □

3. Normed vector spaces

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

Notation

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

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

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

Example 3.1. If S and T are subspaces of V, then so are $S + T$ and αS.

Solution. TODO □

3.1. Norms

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

A *norm* on V is a function

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

such that

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

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

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

(If we replace (c) by the weaker "$\|v\| \geqslant 0$ for all $v \in V$", we get what is called a *semi-norm*.)

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

Example 3.2. Let $\big(V, \|\cdot\|\big)$ be a normed space. Define $d \colon V \times V \longrightarrow \mathbb{R}$ by

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

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

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

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

So every normed space is a metric space.

Solution.

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

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

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

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

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

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

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

Example 3.3. If $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent norms on V, then the corresponding metrics d_1 and d_2 (see Example 3.2) are equivalent.

Solution. By Proposition 2.27 we know that d_1 is finer than d_2 if and only if for every $v \in V$, every sequence that converges to v in (V, d_1) also converges to v in (V, d_2).

So let (v_n) be a sequence that converges to v in (V, d_1), that is $\big(d_1(v_n, v)\big) \longrightarrow 0$, so $\big(\|v_n - v\|_1\big) \longrightarrow 0$, hence $\big(m\|v_n - v\|_1\big) \longrightarrow 0$ and $\big(M\|v_n - v\|_1\big) \longrightarrow 0$. Since by assumption

$$m\|v_n - v\|_1 \leqslant \|v_n - v\|_2 \leqslant M\|v_n - v\|_1,$$

this implies by the Sandwich Theorem that $\big(\|v_n - v\|_2\big) \longrightarrow 0$, in other words that $(v_n) \longrightarrow v$ in (V, d_2).

The fact that d_2 is finer than d_1 follows because

$$\frac{1}{M}\|v\|_2 \leqslant \|v\|_1 \leqslant \frac{1}{m}\|v\|_2 \qquad \text{for all } v \in V,$$

so we can interchange the roles of d_1 and d_2 in the previous argument. ☐

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

Example 3.4. Let $\big(V, \|\cdot\|_V\big)$ and $\big(W, \|\cdot\|_W\big)$ be normed vector spaces over $\mathbb{F}$. We endow the vector space $V \times W$ with the function

$$\|\cdot\| : V \times W \longrightarrow \mathbb{R}, \qquad \|(v,w)\| = \|v\|_V + \|w\|_W.$$

Then $\|\cdot\|$ is a norm on $V \times W$.

Solution. We have

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

Next

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

Finally

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

Proposition 3.5. *If $(V,\|\cdot\|)$ is a normed space, then*

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

(b) the scalar multiplication $\mathbb{F}\times V\longrightarrow V$, $(\alpha,v)\longmapsto \alpha v$, is a continuous function;

(c) the norm $V\longrightarrow\mathbb{R}$, $v\longmapsto\|v\|$, is a uniformly continuous function.

Proof.

(a) For $(v_1,w_1),(v_2,w_2)\in V\times V$ we have

$$\begin{aligned}d_V\big(v_1+w_1,v_2+w_2\big) &= \|(v_1+w_1)-(v_2+w_2)\|\\ &= \|(v_1-v_2)+(w_1-w_2)\|\\ &\leqslant \|v_1-v_2\|+\|w_1-w_2\|\\ &= \|(v_1,w_1)-(v_2,w_2)\|\\ &= d_{V\times V}\big((v_1,w_1),(v_2,w_2)\big),\end{aligned}$$

from which the continuity of the addition follows easily.

(Note also that we actually get uniform continuity, at least with respect to the particular norm we are using on $V\times V$.)

(b) This is slightly delicate.

For $v,w\in V$ and $\alpha,\beta\in\mathbb{F}$ we have

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

So given $\varepsilon>0$, let

$$\delta=\min\left\{\frac{\varepsilon}{3\|v\|},\frac{\varepsilon}{3},\frac{\varepsilon}{3|\alpha|},1\right\}.$$

Then if we are told that

$$d_{\mathbb{F}\times V}\big((\alpha,v),(\beta,w)\big)=|\alpha-\beta|+\|v-w\|<\delta,$$

we can follow the inequalities above and conclude that

$$\|\alpha v - \beta w\| < \varepsilon.$$

(c) Given $\varepsilon > 0$, let $\delta = \varepsilon$. I claim that if $d_V(v,w) < \varepsilon$ then $d_{\mathbb{R}}\big(\|v\|, \|w\|\big) = \big|\|v\| - \|w\|\big| < \varepsilon$. To prove this, note that

$$\|v\| = \|v - w + w\| \leqslant \|v - w\| + \|w\| \Rightarrow \|v\| - \|w\| \leqslant \|v - w\|$$
$$\|w\| = \|v + w - v\| \leqslant \|v\| + \|w - v\| \Rightarrow -\|v - w\| \leqslant \|v\| - \|w\|,$$

so that

$$\big|\|v\| - \|w\|\big| \leqslant \|v - w\|,$$

and the rest follows. □

Example 3.6. If $\big(V, \|\cdot\|\big)$ is a normed space, (v_n), (w_n) are sequences converging in V, and $\alpha \in \mathbb{F}$ is a scalar, then

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

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

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

Solution. Direct consequences of the continuity proved in Proposition 3.5. □

Example 3.7. Let $\{v_1, \ldots, v_n\}$ be a linearly independent subset of a normed space $\big(V, \|\cdot\|\big)$. Then there exists $m > 0$ such that

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

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

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

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

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

To do this, consider the set

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

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

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

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

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

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

contradicting the linear independence of the vectors.

Hence $m > 0$ and we are done. □

We are now in a good position to prove that

Proposition 3.8. *Any two norms on a finite-dimensional vector space V are equivalent.*

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

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

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

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

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

From Example 3.7 we also have $m > 0$ such that

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

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

Example 3.9 (Spaces of sequences)**.** The set of all sequences $\mathbb{F}^{\mathbb{N}} = \{(a_n) : n \in \mathbb{N}, a_n \in \mathbb{F}\}$ is of course a vector space over $\mathbb{F}$ with the usual addition and scalar multiplication.

Various subsets of this are normed spaces:

(a) ℓ^∞ given by the *bounded* sequences

$$\begin{aligned}\ell^\infty &= \left\{(a_n) \in \mathbb{F}^{\mathbb{N}} : \sup(|a_n|) < \infty\right\} \\ &= \left\{(a_n) \in \mathbb{F}^{\mathbb{N}} : \text{there exists } M \text{ such that } |a_n| \leqslant M \text{ for all } n \in \mathbb{N}\right\},\end{aligned}$$

with norm given by

$$\|(a_n)\|_{\ell^\infty} = \sup(|a_n|).$$

This is clearly a vector subspace of $\mathbb{F}^{\mathbb{N}}$.

(b) for any $p \geqslant 1$, the subset ℓ^p of all *p-summable* sequences

$$\ell^p = \left\{(a_n) \in \mathbb{F}^{\mathbb{N}} : \sum_{n=1}^{\infty} |a_n|^p < \infty\right\},$$

with norm given by

$$\|(a_n)\|_{\ell^p} = \left(\sum_{n=1}^{\infty} |a_n|^p\right)^{1/p}.$$

The triangle inequality takes a bit of work to establish, see Proposition 3.10 below.

Proposition 3.10 (Minkowski's Inequality)**.** *Let $1 \leqslant p \leqslant \infty$ and let $u = (u_n), v = (v_n) \in \ell^p$. Then*

$$\|u + v\|_{\ell^p} \leqslant \|u\|_{\ell^p} + \|v\|_{\ell^p}.$$

In particular, ℓ^p is closed under vector addition (and hence a vector subspace of $\mathbb{F}^{\mathbb{N}}$), and $\|\cdot\|_{\ell^p}$ is a norm.

Proof. Fix p and write $\|\cdot\|$ instead of $\|\cdot\|_{\ell^p}$ to simplify notation.

To start with, let $x = (x_n), y = (y_n) \in \ell^p$, and let $a, b \geqslant 0$ be such that $a + b = 1$. Then

$$\sum_{n=1}^{\infty} |ax_n + by_n|^p \leqslant \sum_{n=1}^{\infty} \big(a|x_n| + b|y_n|\big)^p \leqslant a \sum_{n=1}^{\infty} |x_n|^p + b \sum_{n=1}^{\infty} |y_n|^p,$$

where we applied first the triangle inequality for the absolute value, and second the inequality from Exercise 3.4, part (b). Therefore

$$\|ax+by\|^p \leqslant a\|x\|^p + b\|y\|^p.$$

In other words, $\|\cdot\|$ is a convex function.

Now we go back to the context of the statement of the proposition. Given $u, v \in \ell^p$, define

$$x = \frac{1}{\|u\|}u, \qquad y = \frac{1}{\|v\|}v, \qquad a = \frac{\|u\|}{\|u\|+\|v\|}, \qquad b = \frac{\|v\|}{\|u\|+\|v\|},$$

then we have

$$\left(\frac{\|u+v\|}{\|u\|+\|v\|}\right)^p = \|ax+by\|^p \leqslant a+b = 1. \qquad \square$$

3.2. Metric properties of normed spaces

Example 3.11. Let $(V, \|\cdot\|)$ be a normed space and let $W \subseteq V$ be a subspace. Then its closure $\overline{W}$ is also a subspace.

Solution. Suppose $u, v \in \overline{W}$, then there exist sequences (u_n) and (v_n) in W such that $(u_n) \longrightarrow u$ and $(v_n) \longrightarrow v$. Therefore $u_n + v_n \in W$ for all n, and by Proposition 3.5 we have

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

Similarly for scalar multiplication. $\square$

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

Example 3.12. Any finite-dimensional normed space $(V, \|\cdot\|)$ is Banach.

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

By Proposition 3.8 we know that without loss of generality we can take the norm to be given by

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

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

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

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

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

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

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

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

Example 3.13. Consider

$$V = \{(a_n) \in \ell^\infty : \text{there exists } N \in \mathbb{N} \text{ such that } a_n = 0 \text{ for all } n \geqslant N\}$$

consisting of all finite sequences with terms in $\mathbb{F}$.

This is clearly a vector subspace of ℓ^∞, and of course inherits the ℓ^∞ norm from it.

Is it complete with respect to this norm?

Solution. Consider the sequence (v_n) in V given by

$$v_n = \left(1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}, 0, 0, 0, \dots\right).$$

It is Cauchy: given $\varepsilon > 0$, let $N \in \mathbb{N}$ be such that $1/N < \varepsilon$, then for all $n \geqslant m \geqslant N$ we have

$$\|v_n - v_m\|_{\ell^\infty} = \sup\left\{0, \frac{1}{m+1}, \frac{1}{m+2}, \dots, \frac{1}{n}\right\} = \frac{1}{m+1} < \frac{1}{N} < \varepsilon.$$

As a sequence in ℓ^∞, it converges to the following element of ℓ^∞:

$$u = \left(1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}, \dots\right),$$

which is easy to see since

$$\|u - v_n\|_{\ell^\infty} = \frac{1}{n+1} \longrightarrow 0.$$

But u is not in V, so V is not complete. $\square$

Let $(V, \|\cdot\|)$ be a normed space over $\mathbb{F}$. A *completion* of $(V, \|\cdot\|)$ is a Banach space $(\widehat{V}, \widehat{\|\cdot\|})$ over $\mathbb{F}$ together with an $\mathbb{F}$-linear distance-preserving map

$$\iota\colon V \longrightarrow \widehat{V}$$

such that $\iota(V)$ is a dense normed subspace of $\widehat{V}$.

Proposition 3.14. *Any normed space* $(V, \|\cdot\|)$ *has a completion* $(\widehat{V}, \widehat{\|\cdot\|})$.

Proof. We know from Theorem 2.36 that V has a completion that is a metric space. We have to show that the particular complete metric space $(\widehat{V}, \widehat{d})$ constructed in the proof of Theorem 2.36 is actually a normed space such that $\iota(V)$ is a normed subspace.

This is essentially straightforward, just has a lot of tiny little parts.

Let $u = [(u_n)], v = [(v_n)] \in \widehat{V}$. We define

$$u + v = [(u_n + v_n)].$$

To see why this works, first take a Cauchy sequence (u_n) representing the equivalence class u and a Cauchy sequence (v_n) representing the equivalence class v. The sequence $(u_n + v_n)$ is Cauchy in V, as

$$\|(u_n + v_n) - (u_m + v_m)\| \leqslant \|u_n - u_m\| + \|v_n - v_m\|,$$

and (u_n) and (v_n) are Cauchy in V.

Had we chosen other representatives $(u_n') \sim (u_n)$ and $(v_n') \sim (v_n)$, we would have ended up with $(u_n' + v_n')$, which is easily seen to be equivalent to $(u_n + v_n)$, so the equivalence class $[(u_n) + (v_n)]$ is indeed well-defined.

Scalar multiplication and the norm are defined on $\widehat{V}$ as:

$$\alpha u = [(\alpha u_n)], \qquad \widehat{\|u\|} = \lim\big(\|u_n\|\big)$$

and their well-definedness is argued similarly.

Checking the vector space axioms for $\widehat{V}$ is done by using the vector space axioms for V and the continuity of the operations.

Note also that the metric $\widehat{d}$ on $\widehat{V}$ constructed in Theorem 2.36 is the metric associated with the norm $\widehat{\|\cdot\|}$:

$$\widehat{d}(u,v) = \lim d(u_n, v_n) = \lim \|u_n - v_n\| = \widehat{\|u - v\|}. \qquad \square$$

3.3. Continuous linear transformations

Let V and W be normed spaces.

A linear transformation $f\colon V \longrightarrow W$ is said to be *bounded* if there exists $c > 0$ such that

$$\|f(v)\|_W \leqslant c\,\|v\|_V \qquad \text{for all } v \in V.$$

Proposition 3.15. *A linear transformation $f\colon V \longrightarrow W$ between normed spaces is continuous if and only if it is bounded if and only if it is uniformly continuous.*

Proof. Suppose f is bounded. Given $\varepsilon > 0$, let $\delta = \varepsilon/c$. If $v_1, v_2 \in V$ are such that $\|v_1 - v_2\|_V < \delta$, then

$$\|f(v_1) - f(v_2)\|_W = \|f(v_1 - v_2)\|_W \leqslant c\|v_1 - v_2\|_V < c\delta = \varepsilon.$$

Therefore f is uniformly continuous, hence continuous.

Suppose f is not bounded. Let $n \in \mathbb{N}$. There exists $v_n \in V$ such that

$$\frac{\|f(v_n)\|_W}{\|v_n\|_V} \geqslant n.$$

Let $\alpha_n = 1/\|f(v_n)\|_W$ and $u_n = \alpha_n v_n$, then

$$1 = \|f(u_n)\| \geqslant n\|u_n\|.$$

We have therefore a sequence (u_n) with $\|u_n\| = 1/n \longrightarrow 0$, so $(u_n) \longrightarrow 0$. But $\|f(u_n)\| = 1$ for all $n \in \mathbb{N}$, hence does not converge to 0, so f is not continuous. $\square$

We will write $B(V,W)$ for the set of bounded (aka continuous, aka uniformly continuous) linear transformations between the normed spaces V and W.

Consider the following function $\|\cdot\|\colon B(V,W) \longrightarrow \mathbb{R}_{\geqslant 0}$:

$$\|f\| = \sup_{v \neq 0} \frac{\|f(v)\|_W}{\|v\|_V}.$$

As $f \in B(V,W)$, there exists $c > 0$ such that

$$\frac{\|f(v)\|_W}{\|v\|_V} \leqslant c \qquad \text{for all } v \neq 0,$$

so that there is a finite supremum $\|f\|$.

We also note the obvious fact that

$$\|f(v)\| \leqslant \|f\|\,\|v\| \qquad \text{for all } v \in V,$$

and that the linearity of f allows us to rewrite.

$$\|f\| = \sup_{\|v\|=1} \|f(v)\|_W.$$

Theorem 3.16. *Let V and W be normed spaces.*

(a) The set $B(V,W)$ is a normed space with norm given by

$$\|f\| = \sup_{v\neq 0} \frac{\|f(v)\|_W}{\|v\|_V} = \sup_{\|v\|=1} \|f(v)\|_W.$$

(b) If W is a Banach space then $B(V,W)$ is also Banach.

Proof.

(a) As $B(V,W)$ is a subset of $\mathrm{Hom}(V,W)$ and the latter is a vector space (Example A.2), we check that $B(V,W)$ is a subspace.

We have

$$\|f+g\| = \sup_{\|v\|=1} \|f(v)+g(v)\| \leqslant \sup_{\|v\|=1} \big(\|f(v)\| + \|g(v)\|\big) \leqslant \sup_{\|v\|=1} \|f(v)\| + \sup_{\|v\|=1} \|g(v)\| = \|f\| + \|g\|,$$

so that if both f and g are in $B(V,W)$, so is $f+g$.

Similarly:

$$\|\alpha f\| = \sup_{\|v\|=1} \|\alpha f(v)\| = \sup_{\|v\|=1} |\alpha|\,\|f(v)\| = |\alpha|\,\|f\|,$$

so that if f is in $B(V,W)$ and $\alpha \in \mathbb{F}$, then αf is in $B(V,W)$.

In addition to showing that $B(V,W)$ is a vector space, these identities also give two of the three norm axioms, leaving to check that $\|f\| = 0$ if and only if $\|f(v)\| = 0$ for all $v \in V$ if and only if $f = 0$.

(b) Let (f_n) be a Cauchy sequence in $B(V,W)$.

We define $f\colon V \longrightarrow W$ as follows. Set $f(0) = 0$. Fix $v \in V$, $v \neq 0$. Given $\varepsilon > 0$, let $N \in \mathbb{N}$ be such that

$$\|f_n - f_m\| < \frac{\varepsilon}{\|v\|} \qquad \text{for all } n,m \geqslant N.$$

Then for all $n,m \geqslant N$ we have

$$\|f_n(v) - f_m(v)\| \leqslant \|f_n - f_m\|\,\|v\| < \varepsilon,$$

so the sequence $\big(f_n(v)\big)$ is Cauchy in W, which is complete. Let $f(v) = \lim\big(f_n(v)\big) \in W$.

Because of the continuity of addition and scalar multiplication in normed spaces, f is linear:

$$\begin{aligned} f(u+v) &= \lim\big(f_n(u+v)\big) = \lim\big(f_n(u)+f_n(v)\big) = \lim\big(f_n(u)\big) + \lim\big(f_n(v)\big) = f(u)+f(v)\\ f(\alpha v) &= \lim\big(f_n(\alpha v)\big) = \lim\big(\alpha f_n(v)\big) = \alpha \lim\big(f_n(v)\big) = \alpha f(v). \end{aligned}$$

Let $\varepsilon > 0$ and let $v \in V$. Since $\big(f_m(v)\big) \longrightarrow f(v)$, there exists $M_1 \in \mathbb{N}$ such that if $m \geqslant M_1$ then $\|f_m(v) - f(v)\| < \varepsilon\|v\|/2$. Since $\big(f_n\big)$ is Cauchy, there exists $M_2 \in \mathbb{N}$ such that if $n,m \geqslant M_2$ then $\|f_m - f_n\| < \varepsilon/2$. So for $n,m \geqslant \max\{M_1,M_2\}$ we have

$$\begin{aligned} \|(f_n - f)(v)\| &= \|(f_n - f_m)(v) + (f_m(v) - f(v))\|\\ &\leqslant \|(f_n - f_m)(v)\| + \|f_m(v) - f(v)\|\\ &\leqslant \|f_n - f_m\|\,\|v\| + \|f_m(v) - f(v)\|\\ &< \varepsilon\|v\|. \end{aligned}$$

The first thing we can deduce from this is that $f_n - f$ is continuous at v, hence f is continuous at v, so that $f \in B(V, W)$.

The second thing we can deduce is that

$$\frac{\|(f_n - f)(v)\|}{\|v\|} < \varepsilon,$$

hence $\|f_n - f\| < \varepsilon$, so that $(f_n) \longrightarrow f$. □

Let's record an important consequence of Theorem 3.16:

Corollary 3.17. *For any normed space V, the dual space $V^\vee = B(V, \mathbb{F})$ is a Banach space with norm*

$$\|\varphi\| = \sup_{v \neq 0} \frac{|\varphi(v)|}{\|v\|_V}.$$

We'll come back to the topic of dual spaces.

To prove that $B(V, W)$ is a normed space, we had to consider the interplay between the addition of functions and the norms on V and W, and similarly for the operation of multiplying a function by a scalar. There is another operation on functions that has been conspicuously missing from this discussion: composition. We look at this now.

Recall (?) that given a field K, a *K-algebra* is a vector space A over K 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 $\alpha, \beta \in K$ 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 example, given a vector space V over K, the set of all K-linear transformations $V \longrightarrow V$ is an associative unital K-algebra, where multiplication is given by composition and the unit is $\mathbf{1} = \mathrm{id}_V$.

Proposition 3.18. *If $f : U \longrightarrow V$ and $g : V \longrightarrow W$ are continuous linear transformations between normed spaces, then $g \circ f : U \longrightarrow W$ is continuous and linear, and*

$$\|g \circ f\| \leqslant \|g\| \, \|f\|.$$

In particular, for any normed space V, the normed space $B(V, V)$ is closed under composition, hence is an associative unital $\mathbb{F}$-algebra.

Proof. We know already that the composition of linear maps is linear, and that the composition of continuous maps is continuous.

As for the norms, for any $u \in U$ we have

$$\|(g \circ f)(u)\| = \|g(f(u))\| \leqslant \|g\| \, \|f(u)\| \leqslant \|g\| \, \|f\| \, \|u\|,$$

so that for all $u \neq 0$ we have
$$\frac{\|(g \circ f)(u)\|}{\|u\|} \leqslant \|g\| \, \|f\|,$$
and we can conclude by taking supremum.

If $U = W = V$ we get the $\mathbb{F}$-algebra $B(V,V)$ with multiplication given by composition, and with unit element $\mathbf{1} = \mathrm{id}_V$, clearly both linear and continuous. □

Proposition 3.19. *Let V and W be normed spaces and let $f \in B(V,W)$. Then f has a unique extension $\widehat{f} \in B\big(\widehat{V},\widehat{W}\big)$ and*
$$\|\widehat{f}\| = \|f\|.$$

Proof. We know from Proposition 3.15 that f is uniformly continuous, hence by Exercise 2.40 it extends uniquely to a uniformly continuous function $\widehat{f} \colon \widehat{V} \longrightarrow \widehat{W}$.

For $u = [(u_n)], v = [(v_n)] \in \widehat{V}$, we have from the proof of Proposition 3.14 that
$$\begin{aligned}
\widehat{f}(u+v) &= \widehat{f}\big(\lim(u_n + v_n)\big) \\
&= \lim\big(f(u_n+v_n)\big) \\
&= \lim\big(f(u_n)\big) + \lim\big(f(v_n)\big) \\
&= \widehat{f}(u) + \widehat{f}(v),
\end{aligned}$$
and similarly for scalar multiplication.

Insofar as the norm is concerned, we have
$$\|\widehat{f}(v)\| = \lim \|f(v_n)\| \leqslant \|f\|\big(\lim \|v_n\|\big) = \|f\| \, \|v\|,$$
which implies that $\|\widehat{f}\| \leqslant \|f\|$.

But there is another relation between these norms, which we obtain by considering the following diagram:

$$\begin{tikzcd}
V \arrow[r, "\iota_V", "\simeq"'] \arrow[d, "f"'] & \iota(V) \arrow[r, hook, "\text{inclusion}"] \arrow[d, "\widetilde{f}"] & \widehat{V} \arrow[d, "\widehat{f}"] \\
W \arrow[r, "\iota_W", "\simeq"'] & \iota(W) \arrow[r, hook, "\text{inclusion}"] & \widehat{W}
\end{tikzcd}$$

Since ι_V and ι_W are isometries, we have $\|\widetilde{f}\| = \|f\|$. Now $\|\widetilde{f}\|$ and $\|\widehat{f}\|$ are defined by the same formula, but the first is the supremum over the subset $\iota(V)$ of $\widehat{V}$, whereas the second is the supremum over all of $\widehat{V}$. Therefore
$$\|\widehat{f}\| \geqslant \|\widetilde{f}\| = \|f\|.$$
□

3.4. Series

A sequence (a_n) in a normed space $\big(V, \|\cdot\|\big)$ defines a *series* in V
$$\sum_{n=1}^{\infty} a_n,$$
which is a shorthand notation for the sequence of *partial sums* (x_m), where
$$x_m = a_1 + \cdots + a_m = \sum_{n=1}^{m} a_n.$$

The series *converges* if there exists $x \in V$ such that $(x_m) \longrightarrow x$, that is

$$\left\| x - \sum_{n=1}^{m} a_n \right\|_V \longrightarrow 0 \qquad \text{as } m \longrightarrow \infty.$$

The limit x is called the *sum* of the series.

The series *converges absolutely* if the series of real numbers

$$\sum_{n=1}^{\infty} \|a_n\|_V$$

converges.

Proposition 3.20. *Let $\big(V, \|\cdot\|\big)$ be a normed space. V is a Banach space if and only if every absolutely convergent series in V is convergent.*

Proof. In one direction, suppose V is Banach and

$$\sum_{n=1}^{\infty} \|a_n\|_V = r \in \mathbb{R}_{\geqslant 0}.$$

Write

$$x_m = \sum_{n=1}^{m} a_n.$$

Let $\varepsilon > 0$, then there exists $M \in \mathbb{N}$ such that

$$\left| \sum_{n=1}^{m} \|a_n\|_V - r \right| < \frac{\varepsilon}{2} \qquad \text{for all } m \geqslant M.$$

Then for all $m \geqslant k \geqslant M$ we have

$$\|x_m - x_k\| = \left\| \sum_{n=k+1}^{m} a_n \right\| \leqslant \sum_{n=k+1}^{m} \|a_n\| = \left(\sum_{n=1}^{m} \|a_n\| - r \right) + \left(r - \sum_{n=1}^{k} \|a_n\| \right) < \varepsilon.$$

So (x_m) is a Cauchy sequence in V, therefore it converges in V, meaning that the series

$$\sum_{n=1}^{\infty} a_n$$

converges in V.

In the other direction, suppose that every series that converges absolutely also converges in V, and let (a_n) be a Cauchy sequence in V.

There exist

$$\begin{gathered}
n_1 \geqslant 1 \text{ such that } \left\|a_n - a_{n_1}\right\| < \frac{1}{2} \text{ for all } n \geqslant n_1, \\
n_2 > n_1 \text{ such that } \left\|a_n - a_{n_2}\right\| < \frac{1}{2^2} \text{ for all } n \geqslant n_2, \\
\vdots \\
n_k > n_{k-1} \text{ such that } \left\|a_n - a_{n_k}\right\| < \frac{1}{2^k} \text{ for all } n \geqslant n_k, \\
\vdots
\end{gathered}$$

In particular, for all $k \in \mathbb{N}$ we have

$$\left\|a_{n_{k+1}} - a_{n_k}\right\| < \frac{1}{2^k},$$

so that

$$\sum_{k=1}^{\infty} \left\| a_{n_{k+1}} - a_{n_k} \right\| \leqslant \sum_{k=1}^{\infty} \frac{1}{2^k} = 1,$$

which implies that the series

$$\sum_{k=1}^{\infty} \left(a_{n_{k+1}} - a_{n_k} \right) \qquad \text{absolutely converges,}$$

which by our assumption implies that the series

$$\sum_{k=1}^{\infty} \left(a_{n_{k+1}} - a_{n_k} \right) \qquad \text{converges.}$$

Therefore the sequence of partial sums $\left(a_{n_k} - a_{n_1}\right)$ (observe the telescoping behaviour) converges as $k \longrightarrow \infty$, so the subsequence $\left(a_{n_k}\right)$ of (a_n) converges, which by Example 2.70 means that (a_n) converges. □

A sequence $e_1, e_2, \ldots$ of unit vectors of V is a *Schauder basis* of V if for every $v \in V$ there exists a unique sequence of coefficients $\alpha_1, \alpha_2, \cdots \in \mathbb{F}$ such that

$$v = \sum_{n=1}^{\infty} \alpha_n e_n,$$

which should be read as meaning that the series on the right hand side converges in V and its sum is v.

If V has a Schauder basis, then

$$V = \overline{\operatorname{Span}\{e_1, e_2, \ldots\}}.$$

In particular, V is separable (we say that a metric space is *separable* if it has a countable dense subset; for example $\mathbb{R}^n$ is separable). Note that not every separable normed space has a Schauder basis.

Example 3.21. For any $p \geqslant 1$, the sequence space ℓ^p has Schauder basis $\{e_1, e_2, \ldots\}$, where

$$e_n = (0, \ldots, 0, 1, 0, \ldots) \quad \text{with the 1 in the } n\text{-th spot.}$$

In particular, ℓ^p is separable.

Solution. This is an essentially trivial exercise in checking the definition.

Take an arbitrary element $v = (v_n) \in \ell^p$, then

$$\sum_{n=1}^{\infty} |v_n|^p$$

converges with sum $\|v\|^p$.

I claim that the series

$$\sum_{n=1}^{\infty} v_n e_n$$

converges to v with respect to the ℓ^p-norm:

$$\left\| v - \sum_{n=1}^{m} v_n e_n \right\|_{\ell^p}^p = \left\| (0, \ldots, 0, v_{m+1}, v_{m+2}, v_{m+3}, \ldots) \right\|_{\ell^p}^p = \sum_{n=m+1}^{\infty} |v_n|^p,$$

and the latter converges to 0 as $m \longrightarrow \infty$.

The uniqueness of the sequence of coefficients follows from the fact that

$$(v_1, v_2, \ldots) = \sum_{n=1}^{\infty} v_n e_n = v = \sum_{n=1}^{\infty} u_n e_n = (u_1, u_2, \ldots)$$

implies $v_n = u_n$ for all $n \in \mathbb{N}$. □

(You may want to have a look at Appendix A.2.1 and read the definition and discussion of bilinear maps first.)

For $n \in \mathbb{N}$, there is a bilinear map $\beta\colon \mathbb{F}^n \times \mathbb{F}^n \longrightarrow \mathbb{F}$ given by

$$\beta(u,v) = \sum_{k=1}^{n} u_k v_k.$$

As described in Appendix A.2.1, this defines a linear map $\mathbb{F}^n \longrightarrow \big(\mathbb{F}^n\big)^\vee$, $u \longmapsto u^\vee$, given by $u^\vee(v) = \beta(u,v)$.

We'd like to do the same with (subspaces of) $\mathbb{F}^{\mathbb{N}}$: define a bilinear map $\beta\colon \mathbb{F}^{\mathbb{N}} \times \mathbb{F}^{\mathbb{N}} \longrightarrow \mathbb{F}$ by the formula

$$\beta(u,v) = \sum_{n=1}^{\infty} u_n v_n.$$

Of course this would feel more comfortable if we knew that the series $\sum u_n v_n$ actually converges! And of course that does not happen for arbitrary $u, v \in \mathbb{F}^{\mathbb{N}}$, but we can establish some situations where it does work, as follows.

If $p \geqslant 1$, we say that the real number q satisfying

$$\frac{1}{p} + \frac{1}{q} = 1$$

is the *Hölder conjugate* of p. It is easy to see that $q \geqslant 1$. Note that this includes the degenerate pair $p = 1$, $q = \infty$.

Proposition 3.22 (Hölder's Inequality)**.** *Suppose p and q are Hölder conjugate and let $u = (u_n) \in \ell^p$, $v = (v_n) \in \ell^q$. Then*

$$\sum_{n=1}^{\infty} |u_n v_n| \leqslant \|u\|_{\ell^p} \|v\|_{\ell^q}.$$

Proof. We prove the non-degenerate case $p, q \in \mathbb{R}_{>1}$ and leave the (simpler) degenerate one to Exercise 3.15.

Let $x = (x_n) \in \ell^p$, $y = (y_n) \in \ell^q$. For each $n \in \mathbb{N}$ we have

$$|x_n y_n| = \big(|x_n|^p\big)^{1/p}\big(|y_n|^q\big)^{1/q} \leqslant \frac{1}{p}|x_n|^p + \frac{1}{q}|y_n|^q,$$

by an application of Exercise 3.4 part (c), namely $s^a t^b \leqslant as + bt$ where $a + b = 1$.

Therefore

$$\sum_{n=1}^{\infty} |x_n y_n| \leqslant \sum_{n=1}^{\infty}\left(\frac{1}{p}|x_n|^p + \frac{1}{q}|y_n|^q\right) = \frac{1}{p}\|x\|_{\ell^p}^p + \frac{1}{q}\|y\|_{\ell^q}^q.$$

Now start with $u \in \ell^p$, $v \in \ell^q$ and set

$$x = \frac{1}{\|u\|_{\ell^p}}\,u, \qquad y = \frac{1}{\|v\|_{\ell^q}}\,v,$$

so that we obtain

$$\frac{\displaystyle\sum_{n=1}^{\infty} |u_n v_n|}{\|u\|_{\ell^p}\|v\|_{\ell^q}} \leqslant \frac{1}{p} + \frac{1}{q} = 1. \qquad \square$$

Proposition 3.23. *If p, q are Hölder conjugates, then $\beta\colon \ell^p \times \ell^q \longrightarrow \mathbb{F}$ given by*

$$\beta(u,v) = \sum_{n=1}^{\infty} u_n v_n$$

is a continuous bilinear map.

Moreover, if $p, q > 1$, the resulting continuous linear map

$$u \longmapsto u^\vee : \ell^p \longrightarrow \left(\ell^q\right)^\vee$$

is bijective and distance-preserving, hence an isometry $\ell^p \cong \left(\ell^q\right)^\vee$.

Proof. By Hölder's Inequality, the series defining $\beta(u,v)$ converges absolutely. It is then straightforward to check that β is bilinear.

We now show that β is continuous at any $(u',v') \in \ell^p \times \ell^q$. Let $\varepsilon > 0$ and suppose that $(u,v) \in \ell^p \times \ell^q$ satisfies

$$\|(u,v) - (u',v')\| = \|u-u'\|_{\ell^p} + \|v - v'\|_{\ell_q} < \min\left\{\sqrt{\frac{\varepsilon}{3}}, \frac{\varepsilon}{3\|u'\|_{\ell^p}}, \frac{\varepsilon}{3\|v'\|_{\ell^q}}\right\}.$$

Note that

$$\begin{aligned}\beta(u-u', v-v') &= \big(\beta(u,v) - \beta(u',v')\big) - \big(\beta(u,v') - \beta(u',v')\big) - \big(\beta(u',v) - \beta(u',v')\big)\\ &= \big(\beta(u,v) - \beta(u',v')\big) - \beta(u-u',v') - \beta(u',v-v'),\end{aligned}$$

so that

$$\begin{aligned}\big|\beta(u,v) - \beta(u',v')\big| &\leqslant \big|\beta(u-u',v-v')\big| + \big|\beta(u-u',v')\big| + \big|\beta(u',v-v')\big|\\ &= \left|\sum_{n=1}^{\infty}(u_n - u_n')(v_n - v_n')\right| + \left|\sum_{n=1}^{\infty}(u_n-u_n')v_n'\right| + \left|\sum_{n=1}^{\infty} u_n'(v_n - v_n')\right|\\ &\leqslant \sum_{n=1}^{\infty}|u_n - u_n'|\,|v_n - v_n'| + \sum_{n=1}^{\infty}|u_n - u_n'|\,|v_n'| + \sum_{n=1}^{\infty}|u_n'|\,|v_n - v_n'|\\ &\leqslant \|u-u'\|_{\ell^p}\|v-v'\|_{\ell^q} + \|u-u'\|_{\ell^p}\|v'\|_{\ell^q} + \|u'\|_{\ell^p}\|v-v'\|_{\ell^q}\\ &< \varepsilon,\end{aligned}$$

where we first used the triangle inequality a few times and then Hölder's Inequality.

Now we know that $u^\vee \in \left(\ell^q\right)^\vee$, but we can (and will) say something more precise.

Let $v \neq 0$. By Hölder's Inequality

$$\frac{|u^\vee(v)|}{\|v\|_{\ell^q}} \leqslant \|u\|_{\ell^p},$$

so taking supremum we get $\|u^\vee\| \leqslant \|u\|_{\ell^p}$.

I claim that this upper bound is actually attained, that is there exists $v \in \ell^q$ such that

$$\frac{|u^\vee(v)|}{\|v\|_{\ell^q}} = \|u\|_{\ell^p}.$$

Let $v = (v_n)$, where

$$v_n = \begin{cases}\frac{|u_n|^p}{u_n} & \text{if } u_n \neq 0\\ 0 & \text{if } u_n = 0.\end{cases}$$

Then

$$
\begin{aligned}
u^\vee(v) &= \sum_{n=1}^\infty u_n v_n = \sum_{n=1}^\infty |u_n|^p = \|u\|_{\ell^p}^p \\
\|v\|_{\ell^q}^q &= \sum_{n=1}^\infty |v_n|^q = \sum_{n=1}^\infty |u_n|^{(p-1)q} = \sum_{n=1}^\infty |u_n|^p = \|u\|_{\ell^p}^p \\
\frac{|u^\vee(v)|}{\|v\|_{\ell^q}} &= \frac{\|u\|_{\ell_p}^p}{\|u\|_{\ell_p}^{p/q}} = \|u\|_{\ell_p}.
\end{aligned}
$$

We conclude that $\|u^\vee\| = \|u\|_{\ell^p}$. This means that $u \longmapsto u^\vee$ is a distance-preserving map from ℓ^p to $\left(\ell^q\right)^\vee$, hence injective.

It remains to prove surjectivity. Let $\varphi \in \left(\ell^q\right)^\vee$ and let $v \in \ell^q$. Let $\{e_1, e_2, \dots\}$ be the Schauder basis for ℓ^q discussed in Example 3.21. We have

$$
\varphi(v) = \varphi\left(\sum_{n=1}^\infty v_n e_n\right) = \sum_{n=1}^\infty v_n \varphi(e_n) \qquad \text{and } \|e_n\|_{\ell^q} = 1 \text{ for all } n \in \mathbb{N}.
$$

Define $u_n = \varphi(e_n)$ and $u = (u_n)$. If we show that $u \in \ell^p$ then we have $\varphi(v) = u^\vee(v)$ and we're done.

For any $m \in \mathbb{N}$, consider (ignore all the u_n's that are zero, as they do not contribute to the sums):

$$
x = \sum_{n=1}^m \frac{|u_n|^p}{u_n} e_n = \left(\frac{|u_1|^p}{u_1}, \dots, \frac{|u_m|^p}{u_m}, 0, 0, \dots\right),
$$

so that

$$
\|x\|_{\ell^q} = \left(\sum_{n=1}^m \left(|u_n|^{p-1}\right)^q\right)^{1/q} = \left(\sum_{n=1}^m |u_n|^p\right)^{1/q}.
$$

With this in mind, we have

$$
\begin{aligned}
\sum_{n=1}^m |u_n|^p &= \left|\sum_{n=1}^m \frac{|u_n|^p}{u_n} u_n\right| = \left|\sum_{n=1}^m \varphi\left(\frac{|u_n|^p}{u_n} e_n\right)\right| \\
&= \left|\varphi(x)\right| \leqslant \|\varphi\| \, \|x\|_{\ell^q} = \|\varphi\| \left(\sum_{n=1}^m |u_n|^p\right)^{1/q}.
\end{aligned}
$$

Therefore

$$
\left(\sum_{n=1}^m |u_n|^p\right)^{1/p} = \left(\sum_{n=1}^m |u_n|^p\right)^{1-1/q} \leqslant \|\varphi\|.
$$

As this holds for all $m \in \mathbb{N}$, we conclude that the series converges, so $u \in \ell^p$. □

Corollary 3.24. *If $p > 1$ then ℓ^p is a Banach space.*

Proof. Follows as $\ell^p \cong \left(\ell^q\right)^\vee$ and all dual normed spaces are Banach. □

3.5. Exercises

Exercise 3.1 (tut06)**.** A subset S of a vector space V over $\mathbb{F}$ is said to be *convex* if for all $v, w \in S$ and all $a, b \in \mathbb{R}_{\geqslant 0}$ such that $a + b = 1$, we have

$$av + bw \in S.$$

(In other words, for any two points in S, the line segment joining the two points is entirely contained in S.)

Show that:

(a) Any subspace W of V is convex.

(b) The intersection of an arbitrary collection of convex sets is convex.

(c) Any interval $I \subseteq \mathbb{R}$ is convex.

Solution.

(a) Suppose $v, w \in W$, $a, b \in \mathbb{R}_{\geqslant 0}$ such that $a + b = 1$. In particular, $a, b \in \mathbb{F}$ so $av + bw$ is an $\mathbb{F}$-linear combination of elements of W. Since W is a subspace, $av + bw \in W$.

(b) Suppose I is an arbitrary set and S_i is a convex subset of V for all $i \in I$. Let

$$S = \bigcap_{i \in I} S_i$$

and let $v, w \in S$, $a, b \in \mathbb{R}_{\geqslant 0}$ such that $a + b = 1$. Then for all $i \in I$ we have $v, w \in S_i$, so that $av + bw \in S_i$ since S_i is convex. Therefore $av + bw \in S$.

(c) Let $I \subseteq \mathbb{R}$ be an interval and let $v, w \in I$, $a, b \in \mathbb{R}_{\geqslant 0}$ such that $a + b = 1$.

Without loss of generality, $v \leqslant w$. Then

$$av + bw - v = (a-1)v + bw = b(w - v) \geqslant 0 \Rightarrow v \leqslant av + bw$$

and

$$av + bw - w = av + (b-1)w = a(v - w) \leqslant 0 \Rightarrow av + bw \leqslant w.$$

Therefore $v \leqslant av + bw \leqslant w$, hence $av + bw \in I$ by the definition of an interval. □

Exercise 3.2 (tut06)**.** If V is a vector space over $\mathbb{F}$ and $S \subseteq V$ is a convex set, we say that a function $f\colon S \longrightarrow \mathbb{R}$ is *convex* if for all $v, w \in S$ and all $a, b \in \mathbb{R}_{\geqslant 0}$ such that $a + b = 1$, we have

$$f(av + bw) \leqslant af(v) + bf(w).$$

Prove that, if $\big(V, \|\cdot\|\big)$ is a normed space, then $f\colon V \longrightarrow \mathbb{R}$ given by $f(v) = \|v\|$ is a convex function.

Solution. Suppose $v, w \in S$ and $a, b \in \mathbb{R}_{\geqslant 0}$ such that $a + b = 1$. Then

$$f(av + bw) = \|av + bw\| \leqslant \|av\| + \|bw\| = |a|\,\|v\| + |b|\,\|w\| = a\|v\| + b\|w\| = af(v) + bf(w). \quad \square$$

Exercise 3.3 (tut06)**.** Let $I \subseteq \mathbb{R}$ be an interval and let $f\colon I \longrightarrow \mathbb{R}$ be a twice-differentiable function.

The aim of this Exercise is to check the familiar calculus fact: f is convex if and only if $f''(x) \geqslant 0$ for all $x \in I$.

It was heavily inspired by Alexander Nagel's Wisconsin notes [3]:

https://people.math.wisc.edu/~ajnagel/convexity.pdf

(a) For any $s, t \in I$ with $s < t$, define the linear function $L_{s,t}\colon [s,t] \longrightarrow \mathbb{R}$ by

$$L_{s,t}(x) = f(s) + \left(\frac{x-s}{t-s}\right)\big(f(t) - f(s)\big).$$

Convince yourself that this is the equation of the secant line joining $\big(s, f(s)\big)$ to $\big(t, f(t)\big)$.

Prove that f is convex on I if any only if

$$f(x) \leqslant L_{s,t}(x) \qquad \text{for all } s,t \in I \text{ such that } s < t \text{ and all } s \leqslant x \leqslant t.$$

(b) Check that for all $s, t \in I$ such that $s < t$ we have

$$L_{s,t}(x) - f(x) = \frac{x-s}{t-s}\big(f(t) - f(x)\big) - \frac{t-x}{t-s}\big(f(x) - f(s)\big).$$

(c) Use the Mean Value Theorem for f twice to prove that there exist ξ, ζ with $x < \xi < t$ and $s < \zeta < x$ such that

$$L_{s,t}(x) - f(x) = \frac{(t-x)(x-s)}{t-s}\big(f'(\xi) - f'(\zeta)\big).$$

(d) Use the Mean Value Theorem once more to conclude that if $f''(x) \geqslant 0$ for all $x \in I$, then f is convex on I.

(e) Now we prove the converse. From this point on, assume that $f\colon I \longrightarrow \mathbb{R}$ is twice-differentiable and convex, and let $s, t \in I^\circ$.

1. Show that if $s < x < t$ then

$$\frac{f(x) - f(s)}{x - s} \leqslant \frac{f(t) - f(x)}{t - x}.$$

2. Conclude that if $s < x_1 < x_2 < t$ then

$$\frac{f(x_1) - f(s)}{x_1 - s} \leqslant \frac{f(t) - f(x_2)}{t - x_2}.$$

3. Conclude that if $s < t$ then $f'(s) \leqslant f'(t)$, and finally that $f''(x) \geqslant 0$ on I.

Solution. Parts (b)–(d) are pretty thoroughly discussed in the above reference if you need more guidance, so I'll just do parts (a) and (e).

(a) In the definition of convex function, take $v = s$, $w = t$, $a = (t-x)/(t-s)$, $b = (x-s)/(t-s)$, so that $av + bw = x$. Then we know that

$$f(x) \leqslant \frac{t-x}{t-s} f(s) + \frac{x-s}{t-s} f(t) = f(s) + \frac{x-s}{t-s}\big(f(t) - f(s)\big) = L_{s,t}(x).$$

The other direction is straightforward.

(e) 1. From part (a) we have

$$\frac{f(x) - f(s)}{x-s} \leqslant \frac{f(t) - f(s)}{t-s}.$$

Cross-multiplying, we end up with

$$x\big(f(t) - f(s)\big) - s\big(f(t) - f(x)\big) - t\big(f(x) - f(s)\big) \geqslant 0,$$

which is also equivalent to the inequality we are trying to prove.

2. Apply the previous part twice, first with $s < x_1 < x_2$ and then with $x_1 < x_2 < t$, to get

$$\frac{f(x_1) - f(s)}{x_1 - s} \leqslant \frac{f(x_2) - f(x_1)}{x_2 - x_1} \leqslant \frac{f(t) - f(x_2)}{t - x_2}.$$

3. Following from the previous part, we have

$$f'(s) = \lim_{x_1 \searrow s} \frac{f(x_1) - f(s)}{x_1 - s} \leqslant \lim_{x_2 \nearrow t} \frac{f(t) - f(x_2)}{t - x_2} = f'(t).$$

This implies that f' is an increasing function on I°, therefore $f''(x) \geqslant 0$ on I°. □

Exercise 3.4 (tut06)**.** (a) Prove that the functions

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

are convex.

[*Hint*: Use Exercise 3.3.]

(b) Conclude that for any $p \geqslant 1$, any $x, y \geqslant 0$ and any $a, b \geqslant 0$ such that $a + b = 1$, we have

$$(ax + by)^p \leqslant ax^p + by^p.$$

(c) Conclude that for any $x, y \geqslant 0$ and any $a, b \geqslant 0$ such that $a + b = 1$, we have

$$x^a y^b \leqslant ax + by.$$

[*Hint*: Set $x = e^s$, $y = e^t$.]

(d) Show that for any $p \geqslant 1$ and any $x, y \geqslant 0$, we have

$$x^p + y^p \leqslant (x+y)^p.$$

[*Hint*: Let $t = x/y$ and compare derivatives to show that $t^p + 1 \leqslant (t+1)^p$.]

Solution.

(a) (i) We have $f''(x) = p(p-1)x^{p-2} \geqslant 0$ for all $x > 0$, as $p \geqslant 1$.
(ii) We have $\exp''(x) = e^x \geqslant 0$ for all $x \in \mathbb{R}$.

(b) This is exactly the definition of $x \longmapsto x^p$ being a convex function.

(c) If $x = 0$ or $y = 0$, the inequality is trivial, so we may assume $x, y > 0$. Setting $x = e^s$, $y = e^t$, we are trying to prove that

$$e^{as+bt} \leqslant ae^s + be^t,$$

which is the same as e^x being a convex function.

(d) If $y = 0$, the inequality is obvious, so we may assume $y > 0$. Setting $t = x/y$, we are trying to show that

$$t^p + 1 \leqslant (t+1)^p \qquad \text{for all } t \geqslant 0.$$

Let $f\colon \mathbb{R}_{\geqslant 0} \longrightarrow \mathbb{R}$ be given by $f(t) = t^p+1$, and $g(t)\colon \mathbb{R}_{\geqslant 0} \longrightarrow \mathbb{R}$ be given by $g(t) = (t+1)^p$. We have $f(0) = g(0) = 1$. Also

$$f'(t) = pt^{p-1} \leqslant p(t+1)^{p-1} = g'(t) \qquad \text{for all } t > 0,$$

therefore $f(t) \leqslant g(t)$ for all $t \geqslant 0$, as desired. (There's an appeal to the Mean Value Theorem hiding in here, if you want to write out all the details.) $\square$

Exercise 3.5 (tut06)**.** Let $p \geqslant 1$, $q > 0$, $x, y \geqslant 0$, and $a, b \geqslant 0$ such that $a + b = 1$.
Prove that

$$\begin{aligned}\min\{x,y\} &\leqslant \left(ax^{-q} + by^{-q}\right)^{-1/q}\\ &\leqslant x^a y^b\\ &\leqslant \left(ax^{1/p} + by^{1/p}\right)^p\\ &\leqslant ax + by\\ &\leqslant \left(ax^p + by^p\right)^{1/p}\\ &\leqslant \max\{x,y\}.\end{aligned}$$

Solution. Without loss of generality $x \leqslant y$ so $\min\{x,y\} = x$ and $\max\{x,y\} = y$.

(a) $x \leqslant y$ so $x^{-1} \geqslant y^{-1}$ so $x^{-q} \geqslant y^{-q}$ so $bx^{-q} \geqslant by^{-q}$ so $ax^{-q} + bx^{-q} \geqslant ax^{-q} + by^{-q}$ so

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

(b) Let $X = x^{-q}$, $Y = y^{-q}$, then by Exercise 3.4 part (c) we have

$$\begin{aligned}X^aY^b \leqslant aX + bY &\Rightarrow x^{-aq}y^{-bq} \leqslant ax^{-q} + by^{-q}\\ &\Rightarrow x^{aq}y^{bq} \geqslant \left(ax^{-q} + by^{-q}\right)^{-1}\\ &\Rightarrow \left(ax^{-q} + by^{-q}\right)^{-1/q} \leqslant x^a y^b.\end{aligned}$$

(c) Similar to (b), use Exercise 3.4 part (c) with $X = x^{1/p}$, $Y = y^{1/p}$.

(d) Use Exercise 3.4 part (b) with $X = x^{1/p}$, $Y = y^{1/p}$.

(e) Precisely Exercise 3.4 part (b).

(f) Similar to (a). $\square$

Exercise 3.6 (tut07). Let $(V, \|\cdot\|)$ be a normed space and let $S \subseteq V$ be a subset. Then $\overline{\mathrm{Span}(S)}$ is the smallest closed subspace of V that contains S.

Solution. We know that $\mathrm{Span}(S)$ is a subspace of V, and by Example 3.11 that $\overline{\mathrm{Span}(S)}$ is a closed subspace of V.

Let $W \subseteq V$ be some closed subspace of V that contains S. Then $\mathrm{Span}(S) \subseteq W$, and so $\overline{\mathrm{Span}(S)} \subseteq \overline{W} = W$, whence the minimality property. □

Exercise 3.7 (tut07). Let $(V, \|\cdot\|)$ be a normed space and take $r, s > 0$, $u, v \in V$, $\alpha \in \mathbb{F}^\times$. Show that

(a) $\mathbb{B}_r(u+v) = \mathbb{B}_r(u) + \{v\}$;

(b) $\alpha\,\mathbb{B}_1(0) = \mathbb{B}_{|\alpha|}(0)$;

(c) $\mathbb{B}_r(v) = r\mathbb{B}_1(0) + \{v\}$;

(d) $r\mathbb{B}_1(0) + s\mathbb{B}_1(0) = (r+s)\mathbb{B}_1(0)$;

(e) $\mathbb{B}_r(u) + \mathbb{B}_s(v) = \mathbb{B}_{r+s}(u+v)$;

(f) $\mathbb{B}_1(0)$ is a convex subset of V;

(g) any open ball in V is convex.

Solution.

(a)

$$\begin{aligned} w \in \mathbb{B}_r(u+v) &\iff \|(u+v) - w\| < r \\ &\iff \|u - (w - v)\| < r \\ &\iff w - v \in \mathbb{B}_r(u) \\ &\iff w \in \mathbb{B}_r(u) + \{v\}. \end{aligned}$$

(b)

$$\begin{aligned} w \in \alpha\mathbb{B}_1(0) &\iff \frac{1}{\alpha} w \in \mathbb{B}_1(0) \\ &\iff \left\|\frac{1}{\alpha} w\right\| < 1 \\ &\iff \|w\| < |\alpha| \\ &\iff w \in \mathbb{B}_{|\alpha|}(0). \end{aligned}$$

(c) From (a) and (b):

$$\mathbb{B}_r(v) = \mathbb{B}_r(0) + \{v\} = r\mathbb{B}_1(0) + \{v\}.$$

(d) If $\|u\| < r$ and $\|v\| < s$ then $\|u+v\| < r+s$, so $r\mathbb{B}_1(0) + s\mathbb{B}_1(0) \subseteq (r+s)\mathbb{B}_1(0)$.

Conversely, if $\|w\| < r+s$, then

$$w = \frac{r}{r+s} w + \frac{s}{r+s} w \in r\mathbb{B}_1(0) + s\mathbb{B}_1(0).$$

(e) From (c) and (d):

$$\mathbb{B}_r(u) + \mathbb{B}_s(v) = r\mathbb{B}_1(0) + s\mathbb{B}_1(0) + \{u\} + \{v\} = (r+s)\mathbb{B}_1(0) + \{u+v\} = \mathbb{B}_{r+s}(u+v).$$

(f) If $u, v \in \mathbb{B}_1(0)$ and $0 \leqslant a \leqslant 1$, then by (d)

$$au + (1-a)v \in a\mathbb{B}_1(0) + (1-a)\mathbb{B}_1(0) = (a+1-a)\mathbb{B}_1(0) = \mathbb{B}_1(0).$$

(g) $\mathbb{B}_r(u) = r\mathbb{B}_1(0) + \{u\}$ is the translate of a convex set, hence is itself convex. □

Exercise 3.8 (tut07). Let $(V, \|\cdot\|)$ be a normed space and let S, T be subsets of V and $\alpha \in \mathbb{F}$. Prove that

(a) If S and T are bounded, so are $S+T$ and αS.

(b) If S and T are totally bounded, so are $S+T$ and αS.

(c) If S and T are compact, so are $S+T$ and αS.

Solution.

(a) A subset S of V is bounded if and only if $S \subseteq \mathbb{B}_s(0) = s\mathbb{B}_1(0)$ for some $s \geqslant 0$. So $S \subseteq s\mathbb{B}_1(0)$ and $T \subseteq t\mathbb{B}_1(0)$, hence $S + T \subseteq s\mathbb{B}_1(0) + t\mathbb{B}_1(0) = (s+t)\mathbb{B}_1(0)$.

Similarly $\alpha S \subseteq s\alpha\mathbb{B}_1(0) = s\mathbb{B}_{|\alpha|}(0) = (s|\alpha|)\mathbb{B}_1(0)$.

(b) Let $\varepsilon > 0$. Since S and T are totally bounded, they can each be covered by finitely many open balls of radius $\varepsilon/2$:

$$S \subseteq \bigcup_{n=1}^{N} \mathbb{B}_{\varepsilon/2}(s_n)$$
$$T \subseteq \bigcup_{m=1}^{M} \mathbb{B}_{\varepsilon/2}(t_m),$$

but then

$$S + T \subseteq \bigcup_{n=1}^{N} \mathbb{B}_{\varepsilon/2}(s_n) + \bigcup_{m=1}^{M} \mathbb{B}_{\varepsilon/2}(t_m) = \bigcup_{n=1}^{N}\bigcup_{m=1}^{M} \big(\mathbb{B}_{\varepsilon/2}(s_n) + \mathbb{B}_{\varepsilon/2}(t_m)\big) = \bigcup_{n=1}^{N}\bigcup_{m=1}^{M} \mathbb{B}_{\varepsilon}(s_n + t_m).$$

For αS, note that S can be covered by finitely many open balls of radius $\varepsilon/|\alpha|$:

$$S \subseteq \bigcup_{n=1}^{N} \mathbb{B}_{\varepsilon/|\alpha|}(s_n),$$

so that

$$\alpha S \subseteq \bigcup_{n=1}^{N} \alpha\mathbb{B}_{\varepsilon/|\alpha|}(s_n) = \bigcup_{n=1}^{N} \mathbb{B}_{\varepsilon}(s_n).$$

(c) Consider the addition map $a\colon V \times V \longrightarrow V$, $a(v,w) = v + w$. We know that it is continuous, so its restriction

$$a|_{S\times T}\colon S \times T \longrightarrow V, \qquad a(s,t) = s + t$$

is also continuous, and its image is $S+T$. Since S and T are compact, so is $S \times T$, and so is $S + T = a(S \times T)$.

The same argument with scalar multiplication gives compactness of αS. □

Exercise 3.9 (tut07)**.** Let $f \in B(V, W)$.

(a) If U is a subspace of V, then its image $f(U)$ is a subspace of W.

(b) If U is a closed subspace of W, then its preimage $f^{-1}(U)$ is a closed subspace of V.

(c) If S is a convex subset of V, then its image $f(S)$ is a convex subset of W.

(d) If S is a convex subset of W, then its preimage $f^{-1}(S)$ is a convex subset of V.

Solution.

(a) Clear since f is linear so it takes vector subspaces to vector subspaces.

(b) Clear since f is linear so the inverse image of a subspace is a subspace; and f is continuous so the inverse image of a closed set is a closed set.

(c) Let $f(s), f(t) \in f(S)$ and let $a, b \geqslant 0$ such that $a + b = 1$. We have

$$af(s) + bf(t) = f(as + bt) \in f(S),$$

where we used the convexity of S to conclude that $as + bt \in S$.

(d) Let $u, v \in f^{-1}(S)$ and let $a, b \geqslant 0$ such that $a + b = 1$. Then

$$f(au + bv) = af(u) + bf(v) \in S,$$

where we used the convexity of S. We conclude that $au + bv \in f^{-1}(S)$. □

Exercise 3.10 (tut07)**.** Prove that the following subset is a closed subspace of ℓ^1:

$$S = \left\{ (a_n) \in \ell^1 : \sum_{n=1}^{\infty} a_n = 0 \right\}.$$

Solution. Consider the function $f : \ell^1 \longrightarrow \mathbb{F}$ given by

$$f\big((a_n)\big) = \sum_{n=1}^{\infty} a_n.$$

First note that this is a reasonable definition, because the infinite series on the right hand side converges in $\mathbb{F}$:

$$\left| \sum_{n=1}^{N} a_n \right| \leqslant \sum_{n=1}^{N} |a_n|,$$

and the latter converges as $N \longrightarrow \infty$ since $(a_n) \in \ell^1$.

The function f is linear. It is also bounded, because as we have just seen:

$$\big| f\big((a_n)\big) \big| = \left| \sum_{n=1}^{\infty} a_n \right| \leqslant \sum_{n=1}^{\infty} |a_n| = \|(a_n)\|_{\ell^1}.$$

Hence $f \in B(\ell^1, \mathbb{F}) = (\ell^1)^\vee$ and its kernel is S, so S is a closed subspace of ℓ^1. □

Exercise 3.11 (tut07)**.** Suppose $1 \leqslant p \leqslant q$. Prove that

$$\ell^p \subseteq \ell^q.$$

Show that if $p < q$ then the inclusion is strict: $\ell^p \subsetneq \ell^q$.

Solution. We prove that

$$\|x\|_{\ell^q} \leqslant \|x\|_{\ell^p} \qquad \text{for all } x \in \ell^p.$$

If $\|x\|_{\ell^p} = 0$ then $x = 0$ so $\|x\|_{\ell^q} = 0$ and the inequality obviously holds. So suppose $x \neq 0$, then by dividing through by $\|x\|_{\ell^p}$ we can reduce to proving that

$$\|x\|_{\ell^q} \leqslant 1 \qquad \text{for all } x \text{ such that } \|x\|_{\ell^p} = 1.$$

But if $\|x\|_{\ell^p} = 1$ then

$$\sum_{n=1}^{\infty} |x_n|^p = 1,$$

which means that for all $n \in \mathbb{N}$ we have $|x_n|^p \leqslant 1$, so $|x_n| \leqslant 1$. However, $p \leqslant q$ and $|x_n| \leqslant 1$ implies that $|x_n|^q \leqslant |x_n|^p$ for all $n \in \mathbb{N}$, so that

$$\|x\|_{\ell^q}^q = \sum_{n=1}^{\infty} |x_n|^q \leqslant \sum_{n=1}^{\infty} |x_n|^p = 1.$$

If $p < q$ then $\alpha := q/p > 1$. For each $n \in \mathbb{N}$, let

$$x_n = \frac{1}{n^{1/p}},$$

so that

$$|x_n|^p = \frac{1}{n}, \qquad |x_n|^q = \frac{1}{n^\alpha}.$$

We have

$$\|(x_n)\|_{\ell^p} = \sum_{n=1}^{\infty} \frac{1}{n} = \infty, \qquad \|(x_n)\|_{\ell^q} = \sum_{n=1}^{\infty} \frac{1}{n^\alpha} < \infty,$$

so $(x_n) \in \ell^q \setminus \ell^p$. ☐

Exercise 3.12 (tut08)**.** If a series in a normed space $(V, \|\cdot\|)$

$$\sum_{n=1}^{\infty} a_n$$

converges, and converges absolutely, then

$$\left\| \sum_{n=1}^{\infty} a_n \right\| \leqslant \sum_{n=1}^{\infty} \|a_n\|.$$

Solution. This follows from the usual triangle inequality.

For any $m \in \mathbb{N}$, we have

$$\|a_1 + \cdots + a_m\| \leqslant \|a_1\| + \cdots + \|a_m\|.$$

Taking limits as $m \longrightarrow \infty$ we get

$$\left\| \sum_{n=1}^{\infty} a_n \right\| = \left\| \lim_{m \longrightarrow \infty} \sum_{n=1}^{m} a_n \right\| = \lim_{m \longrightarrow \infty} \left\| \sum_{n=1}^{m} a_n \right\| \leqslant \lim_{m \longrightarrow \infty} \sum_{n=1}^{m} \|a_n\| = \sum_{n=1}^{\infty} \|a_n\|.$$ ☐

Exercise 3.13 (tut08)**.** Give an example of a series that converges but does not converge absolutely.

Solution. In $\mathbb{R}$, consider
$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n}.$$
□

Exercise 3.14 (tut08)**.** If $f \in B(V, W)$ with V, W normed spaces, and the series
$$\sum_{n=1}^{\infty} \alpha_n v_n, \qquad \alpha_n \in \mathbb{F}, v_n \in V,$$
converges in V, then the series
$$\sum_{n=1}^{\infty} \alpha_n f(v_n)$$
converges in W to the limit
$$f\left(\sum_{n=1}^{\infty} \alpha_n v_n\right).$$

Solution. Let
$$x_m = \sum_{n=1}^{m} \alpha_n v_n, \qquad x = \sum_{n=1}^{\infty} \alpha_n v_n.$$
We know that $(x_m) \longrightarrow x$ in V.

Since $f \in B(V, W)$ is continuous, we have that $\big(f(x_m)\big) \longrightarrow f(x)$ in W.

But f is also linear, so
$$f(x_m) = \sum_{n=1}^{m} \alpha_n f(v_n).$$
Hence
$$\left(\sum_{n=1}^{m} \alpha_n f(v_n)\right) \longrightarrow f(x),$$
so that the series
$$\sum_{n=1}^{\infty} \alpha_n f(v_n) \quad \text{converges to } f(x).$$
□

Exercise 3.15 (tut08)**.** Prove that if $u = \big(u_n\big) \in \ell^\infty$ and $v = \big(v_n\big) \in \ell^1$, then
$$\sum_{n=1}^{\infty} |u_n v_n| \leqslant \|u\|_{\ell^\infty} \|v\|_{\ell^1}.$$

Solution. Very straightforward.

By the definition of ℓ^∞ and the ℓ^∞-norm, we have $|u_n| \leqslant \|u\|_{\ell^\infty}$ for all $n \in \mathbb{N}$. Therefore for any $m \in \mathbb{N}$ we have
$$\sum_{n=1}^{m} |u_n v_n| \leqslant \|u\|_\infty \sum_{n=1}^{m} |v_n|,$$
but the latter series converges because $v \in \ell^1$, to $\|v\|_{\ell^1}$ and we get
$$\sum_{n=1}^{\infty} |u_n v_n| \leqslant \|u\|_{\ell^\infty} \|v\|_{\ell^1}.$$
□

Exercise 3.16 (tut08)**.** Consider the subset $c_0 \subseteq \mathbb{F}^{\mathbb{N}}$ of all sequences with limit 0:

$$c_0 = \{(a_n) \in \mathbb{F}^{\mathbb{N}} : (a_n) \longrightarrow 0\}.$$

Prove that c_0 is a closed subspace of ℓ^∞.
Conclude that c_0 is a Banach space.

Solution. It's pretty clear that c_0 is a subspace of $\mathbb{F}^{\mathbb{N}}$, and hence of ℓ^∞. To show that c_0 is closed in ℓ^∞, let $(x_n) \longrightarrow x \in \ell^\infty$ with $x_n \in c_0$ for all $n \in \mathbb{N}$. We want to prove that $x \in c_0$.
Write $x_n = (a_{nm}) = (a_{n1}, a_{n2}, a_{n3}, \dots)$ and $x = (a_m) = (a_1, a_2, a_3, \dots)$. Let $\varepsilon > 0$. There exists $N \in \mathbb{N}$ such that for all $n \geqslant N$ we have

$$\sup_m |a_m - a_{nm}| = \|x - x_n\|_{\ell^\infty} < \frac{\varepsilon}{2}.$$

Consider the sequence $x_N = (a_{Nm}) \in c_0$. It converges to 0, so that there exists $M \in \mathbb{N}$ such that for any $m \geqslant M$ we have

$$|a_{Nm}| < \frac{\varepsilon}{2}.$$

Therefore, for $m \geqslant M$, we get

$$|a_m| = |a_m - a_{Nm} + a_{Nm}| \leqslant |a_m - a_{Nm}| + |a_{Nm}| < \varepsilon.$$

Hence $x = (a_m) \longrightarrow 0$.
Since c_0 is closed and ℓ^∞ is Banach, c_0 is Banach. □

Exercise 3.17 (tut08)**.** Prove that the space c_0 of sequences with limit 0 is separable, by finding a Schauder basis for c_0.
[*Hint*: You needn't look too hard.]

Solution. I claim that c_0 has the same Schauder basis at the one given in Example 3.21 for ℓ^p: $\{e_1, e_2, \dots\}$ where $e_n = (0, \dots, 0, 1, 0 \dots)$ with the 1 in the n-th spot.
Take $v = (v_n) \in c_0$, then $(v_n) \longrightarrow 0$. I claim that the series

$$\sum_{n=1}^{\infty} v_n e_n$$

converges to v with respect to the norm on c_0, which is the ℓ^∞-norm:

$$\left\| v - \sum_{n=1}^{m} v_n e_n \right\|_{\ell^\infty} = \|(0, \dots, 0, v_{m+1}, v_{m+2}, v_{m+3}, \dots)\|_{\ell^\infty} = \sup_{n \geqslant m+1} |v_n|,$$

and the latter converges to 0 as $m \longrightarrow \infty$, since $(v_n) \longrightarrow 0$. The uniqueness of the coefficients follows in precisely the same way as for Example 3.21. □

Exercise 3.18 (tut08)**.** Consider the space ℓ^∞ of bounded sequences.

(a) Let $S \subseteq \ell^\infty$ be the subset of sequences (a_n) such that $a_n \in \{0, 1\}$ for all $n \in \mathbb{N}$. Prove that S is an uncountable set.

[*Hint*: Mimic Cantor's diagonal argument.]

(b) Use S to construct an uncountable set T of disjoint open balls in ℓ^∞.

(c) Conclude that ℓ^∞ is not separable.

Solution.

(a) Suppose S is countable and enumerate its elements:

$$\begin{aligned} a_1 &= \big(a_{11}, a_{12}, a_{13}, \dots\big) \\ a_2 &= \big(a_{21}, a_{22}, a_{23}, \dots\big) \\ a_3 &= \big(a_{31}, a_{32}, a_{33}, \dots\big) \\ &\vdots \end{aligned}$$

Go down the diagonal of this infinite grid of 0's and 1's, and define $b_n = 1 - a_{nn}$ for all $n \in \mathbb{N}$. Then $b = (b_n) \in S$, but $b \neq a_m$ for any $m \in \mathbb{N}$, contradiction.

(b) If $a = (a_n), b = (b_n) \in S$ with $a \neq b$ then

$$\|a - b\| = \sup_n |a_n - b_n| = 1,$$

so $\mathbb{B}_{1/2}(a) \cap \mathbb{B}_{1/2}(b) = \varnothing$.

Therefore we can take

$$T = \big\{\mathbb{B}_{1/2}(s) \colon s \in S\big\}.$$

(c) Any dense subset D of ℓ^∞ must contain at least one point (in fact, must be dense) in each open ball in the set T. Since T is uncountable, D must also be uncountable, so ℓ^∞ is not separable. □

Exercise 3.19 (tut09)**.** Consider the map $\pi_1 \colon \mathbb{F}^{\mathbb{N}} \longrightarrow \mathbb{F}$ given by

$$\pi_1\big((a_n)\big) = a_1.$$

(a) Show that π_1 is linear.

(b) Prove that the restriction of π_1 to ℓ^∞ or to ℓ^p for $p \geqslant 1$ is continuous and surjective.

Solution.

(a) Straightforward.

(b) We have for $a \in \ell^\infty$:

$$|\pi_1(a)| = |a_1| \leqslant \sup_{n \geqslant 1}\{|a_n|\} = \|a\|_{\ell^\infty},$$

so π_1 is bounded.

Similarly for $a \in \ell^p$:

$$|\pi_1(a)| = |a_1| = \big(|a_1|^p\big)^{1/p} \leqslant \left(\sum_{n \geqslant 1} |a_n|^p\right)^{1/p} = \|a\|_{\ell^p}.$$

For the surjectivity we note that for any $a \in \mathbb{F}$ we have $\pi_1\big((a, 0, 0, \dots)\big) = a$ and $(a, 0, 0 \dots) \in \ell^1 \subseteq \ell^p$ for all $p \geqslant 1$ and for $p = \infty$. □

Exercise 3.20 (tut09)**.** Consider the left shift map $L\colon \mathbb{F}^{\mathbb{N}} \longrightarrow \mathbb{F}^{\mathbb{N}}$ given by $L\big((a_n)\big) = (a_{n+1})$, that is
$$L(a_1, a_2, a_3, \dots) = (a_2, a_3, \dots).$$

(a) Prove that L is a surjective linear map. What is the kernel of L?

(b) Prove that for all $p \geqslant 1$ and for $p = \infty$, the restriction of L to ℓ^p is a surjective continuous map onto ℓ^p.

(c) Define the right shift map $R\colon \mathbb{F}^{\mathbb{N}} \longrightarrow \mathbb{F}^{\mathbb{N}}$ and prove that it is an injective linear map, the restriction of which is distance-preserving for any ℓ^p with $p \geqslant 1$ and $p = \infty$.

(d) Check that $L \circ R = \mathrm{id}_{\mathbb{F}^{\mathbb{N}}} \neq R \circ L$.

Solution.

(a) It is clear that L is surjective. Linearity is pretty straightforward, and it's also clear that $\ker(L) = \mathrm{Span}\{e_1\}$.

(b) We have
$$\|L(a_1, a_2, a_3, \dots)\|_{\ell^p} = \left(\sum_{n=2}^{\infty} |a_n|^p\right)^{1/p} \leqslant \left(\sum_{n=1}^{\infty} |a_n|^p\right)^{1/p} = \big\|(a_1, a_2, \dots)\big\|_{\ell^p},$$
so L is bounded, and $L\big((a_n)\big) \in \ell^p$ if $(a_n) \in \ell^p$.

For the surjectivity note that if $b = (b_1, b_2, \dots) \in \ell^p$, then
$$b = L(a) \qquad \text{for } a = (0, b_1, b_2, \dots)$$
and $\|a\|_{\ell^p} = \|b\|_{\ell^p}$, so $a \in \ell^p$.

The case of ℓ^∞ is done in a similar way.

(c) To get a linear map we need to set
$$R(a_1, a_2, a_3, \dots) = (0, a_1, a_2, a_3, \dots).$$
Both injectivity and linearity are straightforward.

We have, for $p \geqslant 1$ or $p = \infty$:
$$\|R(a_1, a_2, \dots)\|_{\ell^p} = \|(0, a_1, a_2, \dots)\|_{\ell^p} = \|(a_1, a_2, \dots)\|_{\ell^p},$$
so R is distance-preserving and $R(a) \in \ell^p$ if $a \in \ell^p$.

(d) Clear. For any $a = (a_n) \in \mathbb{F}^{\mathbb{N}}$ we have
$$\begin{aligned} L(R(a)) &= L(R(a_1, a_2, \dots)) = L(0, a_1, a_2, \dots) = (a_1, a_2, \dots) = a, \\ R(L(a)) &= R(L(a_1, a_2, \dots)) = R(a_2, a_3, \dots) = (0, a_2, a_3, \dots) \neq a \text{ unless } a_1 = 0. \end{aligned}$$ □

Exercise 3.21 (tut09)**.** Consider the subset c of $\mathbb{F}^{\mathbb{N}}$ consisting of all convergent sequences (with any limit).

(a) Convince yourself that c is a vector subspace of ℓ^∞.

(b) Prove that $\lim\colon c \longrightarrow \mathbb{F}$ given by

$$(a_n) \longmapsto \lim_{n\longrightarrow\infty}(a_n)$$

is a continuous surjective linear map.

(c) Prove that the formula

$$J\big((a_n)\big) = R\big((a_n)\big) - \Big(\lim_{n\longrightarrow\infty} a_n\Big)(1,1,\dots)$$

defines a linear homeomorphism $J\colon c \longrightarrow c_0$. (Here R denotes the right shift map.)

(d) Show that c is separable and find a Schauder basis for c.

Solution. (a) We know that convergent sequences are bounded, so $c \subseteq \ell^\infty$. We also know that the sum of two convergent sequences is convergent, and that a scalar multiple of a convergent sequence is convergent, and that the constant sequence $(0,0,\dots)$ is convergent.

(b) We know that lim is linear, as a consequence of the continuity of addition and of scalar multiplication.

It is certainly surjective, as given any $a \in \mathbb{F}$ the constant sequence $(a,a,\dots)$ converges to a.

Finally, if $a = (a_n) \in c$ then (a_n) is a bounded sequence and

$$\Big|\lim_{n\longrightarrow\infty} a_n\Big| \leqslant \sup_{n\in\mathbb{N}} |a_n| = \|a\|_{\ell^\infty},$$

so lim is a bounded linear map.

(c) It is clear that J is linear and continuous, as R and lim are linear and continuous.

We exhibit an explicit inverse of J: let $K\colon c_0 \longrightarrow c$ be given by

$$K\big((b_n)\big) = L\big((b_n)\big) - b_1(1,1,\dots).$$

Note that K is linear and continuous, as L and $(b_n) \longmapsto b_1$ are linear and continuous.

We check that K and J and inverses. If $b \in c_0$ and $a \in c$ then:

$$\begin{aligned}
J\big(K(b)\big) &= J\big(L(b)\big) - b_1 J(1,1,\dots)\\
&= R\big(L(b)\big) - 0(1,1,\dots) - b_1\big(R(1,1,\dots) - (1,1,\dots)\big)\\
&= (0,b_2,b_3,\dots) - b_1(-1,0,0,\dots)\\
&= b,\\
K\big(J(a)\big) &= K\big(R(a)\big) - \big(\lim a_n\big)K(1,1,\dots)\\
&= L\big(R(a)\big) - \big(\lim a_n\big)\big(L(1,1,\dots) - (1,1,\dots)\big)\\
&= a.
\end{aligned}$$

(d) We know that $\{e_1, e_2, e_3, \dots\}$ is a Schauder basis for c_0, so we apply $K\colon c_0 \longrightarrow c$ to this to get:

$$\begin{aligned}
K(e_1) &= L(e_1) - (1,1,\dots) = -(1,1,\dots) \\
K(e_2) &= L(e_2) - 0(1,1,\dots) = e_1 \\
K(e_3) &= L(e_3) - 0(1,1,\dots) = e_2 \\
&\vdots \\
K(e_n) &= L(e_n) - 0(1,1,\dots) = e_{n-1} \qquad \text{for } n \geqslant 2 \\
&\vdots
\end{aligned}$$

We suspect then that $\{(1,1,\dots), e_1, e_2, e_3, \dots\}$ is a Schauder basis for c.

This is of course true whenever we have a linear homeomorphism $f\colon V \longrightarrow W$ between normed spaces: If $\{b_1, b_2, \dots\}$ is a Schauder basis for V, then $\{f(b_1), f(b_2), \dots\}$ is a Schauder basis for W.

Let $w \in W$ and let $v = f^{-1}(w) \in V$. Write

$$v = \sum_{j\in\mathbb{N}} \alpha_j b_j,$$

then

$$w = f(v) = \sum_{j\in\mathbb{N}} \alpha_j f(b_j).$$

Uniqueness follows from the uniqueness of the expansion for v. □

Exercise 3.22 (tut09)**.** For any $n \in \mathbb{N}$, give a linear distance-preserving map $\mathbb{F}^n \longrightarrow \ell^2$. (Take the Euclidean norm on $\mathbb{F}^n$.)

Solution. Consider $f\colon \mathbb{F}^n \longrightarrow \ell^2$ given by

$$f(a) = f(a_1, a_2, \dots, a_n) = (a_1, a_2, \dots, a_n, 0, 0, \dots).$$

We have

$$\left\|(a_1, a_2, \dots, a_n, 0, 0, \dots)\right\|_{\ell^2} = \left(\sum_{k=1}^{n} |a_k|^2\right)^{1/2} = \|(a_1, a_2, \dots, a_n)\|_{\mathbb{F}^n},$$

so $f(a) \in \ell^2$, and f is distance-preserving.

Linearity is straightforward. □

4. Hilbert spaces

We discussed distance functions on sets in the context of metric spaces, then we specialised to the case where the set is a vector space V over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ and the distance comes from a norm on V. In this chapter we specialise further to the case where the norm (and hence the distance) comes from an inner product on V. In addition to the norm of a vector, this provides us with a notion of angle between vectors.

4.1. Inner products and norms

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

Let V be a vector space over $\mathbb{F}$. Recall from linear algebra (see Appendix A.2.2 for a summary) that an inner product on V is a function

$$\langle \cdot, \cdot \rangle \colon V \times V \longrightarrow \mathbb{F}$$

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

Proposition 4.1. *If $\big(V, \langle \cdot, \cdot \rangle\big)$ is an inner product space, then the function $\| \cdot \| \colon V \longrightarrow \mathbb{R}_{\geqslant 0}$ defined by*

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

is a norm on V.

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

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

Note also that

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

Finally, by the Cauchy–Schwarz Inequality we have

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

Therefore

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

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

Obviously then:

Corollary 4.2. *Any inner product space is a normed space, and a metric space.*

Let's get to the eponymous definition of this chapter: A *Hilbert space* is a complete inner product space.

Example 4.3. For any $n \in \mathbb{N}$, $\mathbb{F}^n$ is a Hilbert space.

Solution. We know that $\mathbb{F}^n$ is an inner product space, see Example A.10. We also know that finite-dimensional normed spaces are complete, by Example 3.12, so $\mathbb{F}^n$ is a Hilbert space. □

Example 4.4. The sequence space ℓ^2 of square-summable sequences is a Hilbert space.

Solution. Consider the function $\langle \cdot, \cdot \rangle \colon \ell^2 \times \ell^2 \longrightarrow \mathbb{F}$ given by

$$\langle a, b \rangle = \sum_{n=1}^{\infty} a_n \overline{b}_n.$$

We use the Cauchy–Schwarz Inequality (Proposition A.11) to see that this converges. For any $m \in \mathbb{N}$, $(a_1, \dots, a_m), (b_1, \dots, b_m) \in \mathbb{F}^m$ so by Cauchy–Schwarz we have

$$\left| \sum_{n=1}^{m} a_n \overline{b}_n \right| \leqslant \left(\sum_{n=1}^{m} a_n \overline{a}_n \right)^{1/2} \left(\sum_{n=1}^{m} b_n \overline{b}_n \right)^{1/2} = \left(\sum_{n=1}^{m} |a_n|^2 \right)^{1/2} \left(\sum_{n=1}^{m} |b_n|^2 \right)^{1/2}.$$

Taking limits as $m \longrightarrow \infty$, the right hand side becomes $\|a\|_{\ell^2} \|b\|_{\ell^2}$, which is finite since $a, b \in \ell^2$.

The inner product properties are clear. So is the fact that the norm defined by this inner product is exactly the ℓ^2-norm, so we get a Hilbert space by Corollary 3.24. □

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

Proposition 4.5 (Parallelogram Law)**.** *If $\big(V, \langle \cdot, \cdot \rangle\big)$ is an inner product space, then its norm satisfies*

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

Proof. Recall from the proof of Proposition 4.1 that

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

Then

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

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

Example 4.6. Prove that the sequence spaces ℓ^∞ and ℓ^p for $p \neq 2$ are **not** Hilbert spaces.

Solution. Consider

$$\begin{aligned} a &= (1, 1, 0, 0, 0, \dots) \\ b &= (-1, 1, 0, 0, 0, \dots) \\ a + b &= (0, 2, 0, 0, 0, \dots) \\ a - b &= (2, 0, 0, 0, 0, \dots). \end{aligned}$$

Then

$$\begin{aligned}\|a\|_{\ell^p} &= \|b\|_{\ell^p} = 2^{1/p}\\ \|a+b\|_{\ell^p} &= \|a-b\|_{\ell^p} = 2\\ \|a\|_{\ell^\infty} &= \|b\|_{\ell^\infty} = 1\\ \|a+b\|_{\ell^\infty} &= \|a-b\|_{\ell^\infty} = 2,\end{aligned}$$

which shows that ℓ^∞ does not satisfy the Parallelogram Law, and ℓ^p satisfies it if and only if $p = 2$. □

In the proof of the Parallelogram Law (Proposition 4.5) we added the two equalities

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

Subtracting them instead also gives an interesting fact:

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

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

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

which leads us to conclude

Proposition 4.7 (Polarisation Identity). *If $(V,\langle\cdot,\cdot\rangle)$ is an inner product space then*

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

Corollary 4.8 (Converse to the Parallelogram Law). *If $(V,\|\cdot\|)$ is a normed space such that*

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

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

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

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

Proof. In terms of

$$4[v,w] := \|v+w\|^2 - \|v-w\|^2,$$

we have

$$\langle v,w\rangle = \begin{cases}[v,w] & \text{if } \mathbb{F}=\mathbb{R}\\ [v,w] + i[v,iw] & \text{if } \mathbb{F}=\mathbb{C}.\end{cases}$$

We describe in detail the case $\mathbb{F}=\mathbb{C}$, as it is slightly more complicated than $\mathbb{F}=\mathbb{R}$.

The conditions for $\langle\cdot,\cdot\rangle$ to be an inner product are equivalent to the following properties of the auxiliary function $[\cdot,\cdot]$:

(a) $[w,v] = [v,w]$ and $[w,iv] = -[v,iw]$ for all $v,w\in V$.

(b) $[u+v,w] = [u,w] + [v,w]$ for all $u,v,w \in V$.

(c) $[\alpha v,w] = \alpha[v,w]$ for all $v,w \in V$, $\alpha \in \mathbb{R}$.

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

The first part of (a) is obvious from the definition. For the second part of (a), we have

$$\begin{aligned} 4[w,iv] &= \|w+iv\|^2 - \|w-iv\|^2 \\ &= \|i(v-iw)\|^2 - \|(-i)(v+iw)\|^2 \\ &= |i|^2\|v-iw\|^2 - |-i|^2\|v+iw\|^2 \\ &= -4[v,iw]. \end{aligned}$$

The key to parts (b) and (c) is the following calculation for all $u,v,w \in V$:

$$\begin{aligned} 4[2u,w] + 4[2v,w] &= \|2u+w\|^2 - \|2u-w\|^2 + \|2v+w\|^2 - \|2v-w\|^2 \\ &= \Big(\big\|(u+v+w)+(u-v)\big\|^2 + \big\|(u+v+w)-(u-v)\big\|^2\Big) \\ &\quad - \Big(\big\|(u+v-w)+(u-v)\big\|^2 + \big\|(u+v-w)-(u-v)\big\|^2\Big) \\ &= 2\big(\|u+v+w\|^2 + \|u-v\|^2\big) - 2\big(\|u+v-w\|^2 + \|u-v\|^2\big) \\ &= 8[u+v,w]. \end{aligned}$$

We conclude that

$$[2u,w] + [2v,w] = 2[u+v,w] \qquad \text{for all } u,v,w \in V. \tag{4.1}$$

In particular, setting $u = 0$ we have $[2u,w] = 0$ (from the definition of $[\cdot,\cdot]$) so that

$$[2v,w] = 2[v,w] \qquad \text{for all } v,w \in V.$$

Using this on the LHS of Equation (4.1) we get part (b):

$$[u,w] + [v,w] = [u+v,w] \qquad \text{for all } u,v,w \in V.$$

We already have part (c) for $\alpha = 0,1,2$. Clearly repeated application of part (b) gives us

$$[nv,w] = n[v,w] \qquad \text{for all } n \in \mathbb{N}.$$

For (-1) we have

$$4[-v,w] = \|-v+w\|^2 - \|-v-w\|^2 = \|v-w\|^2 - \|v+w\|^2 = -4[v,w],$$

hence

$$[nv,w] = n[v,w] \qquad \text{for all } n \in \mathbb{Z}.$$

For any $q \in \mathbb{Q}$, write $q = m/n$ with $\gcd(m,n) = 1$:

$$n[qv,w] = [nqv,w] = [mv,w] = m[v,w],$$

therefore

$$[qv,w] = q[v,w] \qquad \text{for all } q \in \mathbb{Q}.$$

Finally, for any $\alpha \in \mathbb{R}$ choose a rational sequence $(q_n) \longrightarrow \alpha$:

$$\begin{aligned}
[\alpha v, w] &= \left[\left(\lim_{n\longrightarrow\infty} q_n\right) v, w\right] \\
&= \left[\lim_{n\longrightarrow\infty} (q_n v), w\right] \\
&= \lim_{n\longrightarrow\infty} [q_n v, w] \\
&= \lim_{n\longrightarrow\infty} (q_n [v, w]) \\
&= \left(\lim_{n\longrightarrow\infty} q_n\right) [v, w] \\
&= \alpha [v, w].
\end{aligned}$$

Somewhere in the middle we used the fact that $[\cdot,\cdot]$ is continuous in the first variable (which follows easily from the definition of $[\cdot,\cdot]$ and th fact that the norm is continuous).

Part (d) is straightforward, as

$$4[v, v] = 4\|v\|^2.$$ □

4.2. Orthogonality

Given a subset S of an inner product space V, we define

$$S^{\perp} = \left\{ v \in V : \langle v, s \rangle = 0 \text{ for all } s \in S \right\}.$$

Proposition 4.9. *For any subset $S \subseteq V$, $S^{\perp}$ is a closed subspace of V.*

Proof. That $S^{\perp}$ is a vector subspace of V follows easily from the linearity of $\langle \cdot,\cdot \rangle$ in the first variable.

That $S^{\perp}$ is closed in V follows from the continuity of $\langle \cdot,\cdot \rangle$ in the first variable. □

Given a normed space V, a *projection* is a continuous linear map $\varphi \in B(V, V)$ such that $\varphi^2 = \varphi$.

Proposition 4.10. *Let $\varphi \in B(V, V)$ be a projection.*

(a) The map $\mathrm{id}_V - \varphi$ is also a projection.

(b) $\mathrm{im}(\varphi) = \ker(\mathrm{id}_V - \varphi)$ and $\mathrm{im}(\mathrm{id}_V - \varphi) = \ker(\varphi)$. In particular, the image of a projection is a closed subspace.

(c) We have

$$V = \mathrm{im}(\varphi) \oplus \ker(\varphi).$$

Solution. (a) Since both id_V and φ are continuous and linear, so is $\mathrm{id}_V - \varphi$. Also, we have

$$\left(\mathrm{id}_V - \varphi\right) \circ \left(\mathrm{id}_V - \varphi\right) = \mathrm{id}_V - \varphi - \varphi + \varphi \circ \varphi = \mathrm{id}_V - \varphi.$$

(b) If $v \in \mathrm{im}(\varphi)$ then $v = \varphi(w)$ so that

$$\left(\mathrm{id}_V - \varphi\right)(v) = v - \varphi(v) = \varphi(w) - \varphi^2(w) = \varphi(w) - \varphi(w) = 0,$$

so $v \in \ker\left(\mathrm{id}_V - \varphi\right)$.

Conversely, if $v \in \ker\left(\mathrm{id}_V - \varphi\right)$ then $v - \varphi(v) = 0$ so $v = \varphi(v) \in \mathrm{im}(\varphi)$.

The other identity follows by applying the first identity to the projection $\mathrm{id}_V - \varphi$.

Since the image of φ is the kernel of $\mathrm{id}_V - \varphi$, it is a closed subspace, as the kernel of any linear continuous map is a closed subspace.

(c) We need to prove that $V = \operatorname{im}(\varphi) + \ker(\varphi)$ and that $\operatorname{im}(\varphi) \cap \ker(\varphi) = \{0\}$.

Given $v \in V$, we have

$$v = \varphi(v) + \big(\operatorname{id}_V - \varphi\big)(v) \in \operatorname{im}(\varphi) + \ker(\varphi).$$

If

$$w \in \operatorname{im}(\varphi) \cap \ker(\varphi) = \ker\big(\operatorname{id}_V - \varphi\big) \cap \ker(\varphi),$$

then

$$w = \varphi(w) + \big(\operatorname{id}_V - \varphi\big)(w) = 0 + 0 = 0.$$

□

Example 4.11. Take $V = \mathbb{R}^2$ with the Euclidean norm. The matrix

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

satisfies $A^2 = A$, so it defines a projection. It is easy to see that $\operatorname{im}(A)$ is the diagonal $y = x$ in $\mathbb{R}^2$, and $\ker(A)$ is the y-axis.

The complementary projection is given by the matrix

$$I - A = \begin{pmatrix} 0 & 0 \\ -1 & 1 \end{pmatrix},$$

where $\operatorname{im}(I - A)$ is the y-axis and $\ker(I - A)$ is the diagonal $y = x$.

If V is an inner product space, an *orthogonal projection* is a projection φ such that $\ker(\varphi) = \big(\operatorname{im}(\varphi)\big)^{\perp}$.

If $\big(X, d\big)$ is a metric space, and $Y \subseteq X$ is an arbitrary subset, we can define a function $d_Y : X \longrightarrow \mathbb{R}_{\geqslant 0}$ that gives the distance to the set Y:

$$d_Y(x) = \inf_{y \in Y} d(x, y).$$

Theorem 4.12 (Hilbert Projection Theorem)**.** *Let H be a Hilbert space.*

(a) Let Y be a convex closed ***subset*** *of H. For any $x \in H$, there exists a unique $y_{min} \in Y$ that realises the distance between x and Y:*

$$d_Y(x) = d\big(x, y_{min}\big) = \big\|x - y_{min}\big\|.$$

In other words, y_{min} is the unique point of Y that is as close as possible to x.

(b) Suppose now that W is a closed ***vector subspace*** *of H. For any $x \in H$ and any $y \in W$, we have*

$$y = y_{min} \qquad \text{if and only if} \qquad x - y \in W^{\perp}.$$

The map $\varphi \colon H \longrightarrow H$ given by $\varphi(x) = y_{min}$ is an orthogonal projection with image W.

Proof.

(a) Let

$$D = d_Y(x) = \inf_{y \in Y} d(x, y).$$

Take a sequence (y_n) in Y such that

$$\big(\|x - y_n\|\big) = \big(d(x, y_n)\big) \longrightarrow D.$$

I claim that the sequence (y_n) is Cauchy.

Let $\varepsilon > 0$. Note that

$$\big(\|x - y_n\|^2\big) \longrightarrow D^2,$$

so there exists $N \in \mathbb{N}$ such that

$$\big|\|x - y_n\|^2 - D^2\big| \leqslant \frac{\varepsilon}{4} \qquad \text{for all } n \geqslant N.$$

Let $m, n \geqslant N$. By the Parallelogram Law:

$$\big\|(y_n - x) + (y_m - x)\big\|^2 + \big\|(y_n - x) - (y_m - x)\big\|^2 = 2\|y_n - x\|^2 + 2\|y_m - x\|^2,$$

so that

$$\begin{aligned}\|y_n - y_m\|^2 &= 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - \big\|(y_n + y_m) - 2x\big\|^2 \\ &= 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4\left\|\frac{y_n + y_m}{2} - x\right\|^2.\end{aligned}$$

At this point we notice that since $y_n, y_m \in Y$ and Y is convex, $(1/2)y_n + (1/2)y_m \in Y$; we can then continue with

$$\begin{aligned}2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4\left\|\frac{y_n + y_m}{2} - x\right\|^2 &\leqslant 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4D^2 \\ &= 2\big(\|y_n - x\|^2 - D^2\big) + 2\big(\|y_m - x\|^2 - D^2\big) \\ &< \varepsilon.\end{aligned}$$

So (y_n) is Cauchy in Y, which is complete (because a closed subset of the Hilbert space H). Therefore (y_n) converges in Y to some point that we will call $y_{\min}$. Since the distance function is continuous, we have

$$d(x, y_{\min}) = \lim_{n \longrightarrow \infty} d(x, y_n) = D = d_Y(x).$$

It remains to prove the uniqueness of $y_{\min}$. Suppose $y' \in Y$ satisfies $d(x, y') = D$. By the Parallelogram Law

$$\big\|(y_{\min} - x) + (y' - x)\big\|^2 + \big\|(y_{\min} - x) - (y' - x)\big\|^2 = 2\|y_{\min} - x\|^2 + 2\|y' - x\|^2,$$

so that

$$\|y_{\min} - y'\|^2 = 2\|y_{\min} - x\|^2 + 2\|y' - x\|^2 - \|(y_{\min} + y') - 2x\|^2 \leqslant 2D^2 + 2D^2 - 4D^2 = 0,$$

which implies $y' = y_{\min}$.

(b) A subspace of H is convex, so the conclusion of part (a) applies to W.

First we prove that $x - y_{\min} \in W^{\perp}$.

Let $w \in W$ be a unit vector, so $\|w\| = 1$. Define

$$\alpha = \langle x - y_{\min}, w\rangle \qquad \text{and} \qquad v = x - \big(y_{\min} + \alpha w\big).$$

Then

$$\begin{aligned}\langle v, w\rangle &= \langle x - y_{\min} - \alpha w, w\rangle \\ &= \langle x - y_{\min}, w\rangle - \alpha\langle w, w\rangle \\ &= \alpha - \alpha = 0,\end{aligned}$$

so $v \perp w$. Therefore

$$\left\|x - y_{\min}\right\|^2 = \|v + \alpha w\|^2 = \|v\|^2 + |\alpha|^2\,\|w\|^2 = \|v\|^2 + |\alpha|^2 \geqslant \|v\|^2,$$

in other words

$$\left\|x - y_{\min}\right| \geqslant \|x - (y_{\min} + \alpha w)\|.$$

By the minimality property of $y_{\min}$, this inequality must actually be an equality, therefore $\alpha = 0$.

So $\langle x - y_{\min}, w\rangle = 0$ for all unit vectors $w \in W$, which implies that $\langle x - y_{\min}, w\rangle = 0$ for all $w \in W$, so $x - y_{\min} \in W^\perp$.

Next we show that if $y \in W$ and $x - y \in W^\perp$ then $y = y_{\min}$.

We have

$$\begin{aligned}x - y \in W^\perp &\Rightarrow \langle x - y, w\rangle = 0 && \text{for all } w \in W \\ &\Rightarrow \langle x - y, w - y\rangle = 0 && \text{for all } w \in W \\ &\Rightarrow \|x - w\|^2 = \|x - y\|^2 + \|w - y\|^2 && \text{for all } w \in W \\ &\Rightarrow \|x - w\|^2 \geqslant \|x - y\|^2 && \text{for all } w \in W,\end{aligned}$$

implying that $y \in W$ is closest to x; by the uniqueness statement of part (a), we conclude that $y = y_{\min}$.

We now move on to the function φ. By its definition, for each $x \in H$, $\varphi(x)$ is the unique element of W with the property that $x - \varphi(x) \in W^\perp$.

We check that φ is linear.

If $x_1, x_2 \in H$, we have $\varphi(x_1) + \varphi(x_2) \in W$ and

$$\big(x_1 + x_2\big) - \big(\varphi(x_1) + \varphi(x_2)\big) = \big(x_1 - \varphi(x_1)\big) + \big(x_2 - \varphi(x_2)\big) \in W^\perp,$$

so $\varphi(x_1) + \varphi(x_2) = \varphi(x_1 + x_2)$.

Similarly, if $x \in H$ and $\alpha \in \mathbb{F}$ we have $\alpha\varphi(x) \in W$ and

$$\alpha x - \alpha\varphi(x) = \alpha\big(x - \varphi(x)\big) \in W^\perp,$$

so $\alpha\varphi(x) = \varphi(\alpha x)$.

We check that φ is continuous.

For any $x \in H$, we have $\varphi(x) \in W$ and $x - \varphi(x) \in W^\perp$, so $\big(x - \varphi(x)\big) \perp \varphi(x)$ and

$$\|x\|^2 = \left\|(x - \varphi(x)) + \varphi(x)\right\|^2 = \|x - \varphi(x)\|^2 + \|\varphi(x)\|^2 \geqslant \|\varphi(x)\|^2,$$

so $\|\varphi(x)\| \leqslant \|x\|$.

We check that φ is a projection with image W.

Certainly $\operatorname{im}\varphi \subseteq W$. If $y \in W$ then $\varphi(y) = y$ (closest point to y is y itself), so in fact $\operatorname{im}\varphi = W$. Hence for all $x \in H$ we get $\varphi^2(x) = \varphi\big(\varphi(x)\big) = \varphi(x)$, so $\varphi^2 = \varphi$.

Finally, we check that φ is an orthogonal projection.

We want to show that $W^\perp = \ker\varphi$. But $x \in W^\perp$ if and only if $x - 0 \in W^\perp$ if and only if $\varphi(x) = 0$ if and only if $x \in \ker\varphi$. □

Corollary 4.13. *If H is a Hilbert space and S is a subset of H, then*

$$\left(S^{\perp}\right)^{\perp} = \overline{\operatorname{Span}(S)}.$$

Proof. We've seen in Exercise 4.3 that for any inner product space V and subset S we have

$$S \subseteq \left(S^{\perp}\right)^{\perp} \qquad \text{and} \qquad S^{\perp} = \overline{\operatorname{Span}(S)}^{\perp}.$$

Let $W = \overline{\operatorname{Span}(S)}$, then what we are trying to prove is that $\left(W^{\perp}\right)^{\perp} = W$ when H is a Hilbert space.

Let $x \in \left(W^{\perp}\right)^{\perp}$. By the Hilbert Projection Theorem, we can decompose

$$H = W \oplus W^{\perp}.$$

So we have $x = y + z$ with $y \in W$ and $z \in W^{\perp}$. Then

$$0 = \langle x, z\rangle = \langle y + z, z\rangle = \langle y, z\rangle + \langle z, z\rangle = 0 + \|z\|^2,$$

implying that $z = 0$ and $x = y \in W$. $\square$

4.3. Exercises

Exercise 4.1 (tut09)**.** Let $\big(V, \langle\cdot,\cdot\rangle\big)$ be an inner product space. Prove that the inner product is a continuous function.

Solution. One way is to use the Polarisation Identity and the fact that the norm is continuous.

But we can also proceed more directly: suppose $(x_n, y_n) \longrightarrow (x, y)$, then $(x_n) \longrightarrow x$ and $(y_n) \longrightarrow y$. As (y_n) converges, it is bounded, so there exists $C \geqslant 0$ such that $\|y_n\| \leqslant C$ for all $n \in \mathbb{N}$.

Given $\varepsilon > 0$, let $N \in \mathbb{N}$ be such that

$$\|x_n - x\| < \frac{\varepsilon}{2C} \qquad \text{and} \qquad \|y_n - y\| < \frac{\varepsilon}{2\|x\|} \text{ for all } n \geqslant N.$$

Then

$$\begin{aligned}
\big|\langle x_n, y_n\rangle - \langle x, y\rangle\big| &= \big|\langle x_n, y_n\rangle - \langle x, y_n\rangle + \langle x, y_n\rangle - \langle x, y\rangle\big| \\
&= \big|\langle x_n - x, y_n\rangle + \langle x, y_n - y\rangle\big| \\
&\leqslant \big|\langle x_n - x, y_n\rangle\big| + \big|\langle x, y_n - y\rangle\big| \\
&\leqslant \|x_n - x\|\,\|y_n\| + \|x\|\,\|y_n - y\| \\
&\leqslant C\|x_n - x\| + \|x\|\,\|y_n - y\| \\
&< \varepsilon.
\end{aligned}$$

We conclude that $\big(\langle x_n, y_n\rangle\big) \longrightarrow \langle x, y\rangle$. □

Exercise 4.2 (tut09)**.** Let $\big(V, \langle\cdot,\cdot\rangle\big)$ be an inner product space. For any $v \in V$ we have

$$\|v\| = \sup_{\|w\|=1} \big|\langle v, w\rangle\big|.$$

The supremum is in fact achieved by a well-chosen w.

Solution. If $v = 0$ then the equality is obvious.

So assume now that $v \neq 0$. By Cauchy–Schwarz we have for all $w \in V$:

$$\big|\langle v, w\rangle\big| \leqslant \|v\|\,\|w\|.$$

Therefore for all $w \in V$ with $\|w\| = 1$ we have

$$\big|\langle v, w\rangle\big| \leqslant \|v\|,$$

so that

$$\sup_{\|w\|=1} \big|\langle v, w\rangle\big| \leqslant \|v\|.$$

To get equality, take $w = \frac{1}{\|v\|}\, v$ and see that the LHS is indeed $\|v\|$. □

Exercise 4.3 (tut09)**.** Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space and let R, S be subsets of V.

(a) Prove that $S \cap S^{\perp} = 0$.

(b) Prove that if $R \subseteq S$ then $S^{\perp} \subseteq R^{\perp}$.

(c) Prove that $S \subseteq (S^{\perp})^{\perp}$.

(d) Prove that $S^{\perp} = \overline{\operatorname{Span}(S)}^{\perp}$.

Solution.

(a) If $x \in S^{\perp} \cap S$ then $\langle x, s\rangle = 0$ for all $s \in S$, in particular $\langle x, x\rangle = 0$ so $x = 0$.

(b) Suppose $R \subseteq S$ and $x \in S^{\perp}$. For any $r \in R$ we have $r \in S$ so $\langle x, r\rangle = 0$, hence $x \in R^{\perp}$.

(c) Let $s \in S$. For any $x \in S^{\perp}$, we have

$$\langle s, x\rangle = \overline{\langle x, s\rangle} = 0,$$

so $s \in (S^{\perp})^{\perp}$.

(d) Since $S \subseteq \operatorname{Span}(S) \subseteq \overline{\operatorname{Span}(S)}$, by part (b) we get

$$\overline{\operatorname{Span}(S)}^{\perp} \subseteq S^{\perp}.$$

In the other direction, suppose $x \in S^{\perp}$. For any $v \in \operatorname{Span}(S)$ we have

$$\langle x, v\rangle = \langle x, \alpha_1 s_1 + \cdots + \alpha_n s_n\rangle = \overline{\alpha}_1\langle x, s_1\rangle + \cdots + \overline{\alpha}_n\langle x, s_n\rangle = 0.$$

Now if $(v_n) \longrightarrow w \in \overline{\operatorname{Span}(S)}$ with $v_n \in \operatorname{Span}(S)$, we have

$$\langle x, w\rangle = \langle x, \lim v_n\rangle = \lim\langle x, v_n\rangle = \lim 0 = 0. \qquad \square$$

Exercise 4.4 (tut09)**.** Let (X, d) be a metric space and let $S \subseteq X$. Prove that $d_S(x) = 0$ if and only if $x \in \overline{S}$.

Solution. Suppose $0 = d_S(x) = \inf_{s\in S} d(s, x)$, then there exists a sequence (s_n) with $s_n \in S$ and $d(s_n, x) \longrightarrow d_S(x) = 0$, so $(s_n) \longrightarrow x$, so $x \in \overline{S}$.

Conversely, if $x \in \overline{S}$ then there exists a sequence $(s_n) \longrightarrow x$, so

$$d_S(x) = \inf_{s\in S} d_S(x) \leqslant \inf_{n\in\mathbb{N}} d(s_n, x) = 0. \qquad \square$$

A. Appendix

At the moment, this is just a disorganised pile of miscellanea.

A.1. Set theory

Theorem A.1 (Schröder–Bernstein)**.** *If A and B are sets and $f\colon A \longrightarrow B$ and $g\colon B \longrightarrow A$ are injective functions, then A and B have the same cardinality (that is, there exists some bijective function $h\colon A \longrightarrow B$).*

Proof. If $g(B) = A$ then g is bijective so we can take $h = g^{-1}$.

Otherwise, let $X_1 = A \smallsetminus g(B)$. Define $X_2 = g\big(f(X_1)\big)$, and more generally

$$X_n = g\big(f(X_{n-1})\big), \qquad \text{for } n \geqslant 2.$$

Let

$$X = \bigcup_{n\in\mathbb{N}} X_n.$$

This is a subset of A with the property that

$$g\big(f(X)\big) = \bigcup_{n\in\mathbb{N}} g\big(f(X_n)\big) = \bigcup_{n\in\mathbb{N}} X_{n+1}. \tag{A.1}$$

If $a \in A \smallsetminus X$, then $a \notin X_1 = A \smallsetminus g(B)$, therefore $a \in g(B)$. As g is injective, there is a unique $b \in B$ such that $a = g(b)$, in other words, $g^{-1}(a) = \{b\}$.

This means that

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

gives a well-defined function $h\colon A \longrightarrow B$.

Let's check that h is surjective. If $b \in f(X)$, then $b = f(a) = h(a)$ for some $a \in X$ and we are done. If $b \notin f(X)$, then as g is injective, $g(b) \notin g\big(f(X)\big)$. By Equation (A.1), we have

$$g(b) \notin \bigcup_{n\in\mathbb{N}} X_{n+1}.$$

We also have $g(b) \in g(B)$ so $g(b) \notin X_1 = A \smallsetminus g(B)$. Therefore

$$g(b) \notin X = X_1 \cup \bigcup_{n\in\mathbb{N}} X_{n+1},$$

so setting $a = g(b)$ we have

$$h(a) = h(g(b)) = g^{-1}\big(g(b)\big) = b.$$

Finally, we check that h is injective. Suppose $h(a_1) = h(a_2)$. There are three cases to consider:

- $a_1 \in X$ and $a_2 \in A \setminus X$ (or vice-versa). This cannot actually occur: if $h(a_1) = h(a_2)$ then $f(a_1) = g^{-1}(a_2)$, so that
$$a_2 = g\big(g^{-1}(a_2)\big) = g\big(f(a_1)\big) \in g\big(f(X)\big) \subseteq X,$$
contradiction.
- $a_1, a_2 \in X$, then $f(a_1) = f(a_2)$ so $a_1 = a_2$ by the injectivity of f.
- $a_1, a_2 \in A \setminus X$, then $g^{-1}(a_1) = g^{-1}(a_2)$ so $a_1 = a_2$ by applying g. □

A.2. Linear algebra

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

For vector spaces V, W over $\mathbb{F}$, we write
$$\mathrm{Hom}(V, W) = \{f\colon V \longrightarrow W : f \text{ is a linear transformation}\}.$$

Example A.2. Prove that $\mathrm{Hom}(V, W)$ is a vector space over $\mathbb{F}$.

[*Hint*: You may use without proof the fact that for any set X and any vector space W over $\mathbb{F}$, the set $\mathrm{Fun}(X, W) := \{f\colon X \longrightarrow W\}$ is a vector space over $\mathbb{F}$ with the obvious vector space operations.]

Solution. We apply the Subspace Theorem.

- The zero vector of $\mathrm{Fun}(V, W)$ is the constant function $\mathbf{0}\colon V \longrightarrow W$ given by $\mathbf{0}(v) = 0 \in W$ for all $v \in V$. We check that this is a linear transformation:
$$\begin{aligned}\mathbf{0}(v_1 + v_2) &= 0 = 0 + 0 = \mathbf{0}(v_1) + \mathbf{0}(v_2)\\ \mathbf{0}(\lambda v) &= 0 = \lambda\, 0 = \lambda \mathbf{0}(v)\end{aligned}$$
- Suppose $f_1, f_2 \in \mathrm{Hom}(V, W)$, then both are linear transformations. Their sum in $\mathrm{Fun}(V, W)$ is the function $(f_1 + f_2)\colon V \longrightarrow W$ given by $(f_1 + f_2)(v) = f_1(v) + f_2(v)$. We check that this is linear:
$$\begin{aligned}(f_1 + f_2)(v_1 + v_2) &= f_1(v_1 + v_2) + f_2(v_1 + v_2)\\ &= f_1(v_1) + f_1(v_2) + f_2(v_1) + f_2(v_2)\\ &= (f_1 + f_2)(v_1) + (f_1 + f_2)(v_2)\\ (f_1 + f_2)(\lambda v) &= f_1(\lambda v) + f_2(\lambda v)\\ &= \lambda f_1(v) + \lambda f_2(v)\\ &= \lambda(f_1 + f_2)(v).\end{aligned}$$
So $(f_1 + f_2) \in \mathrm{Hom}(V, W)$.
- Suppose $f \in \mathrm{Hom}(V, W)$ and $\lambda \in \mathbb{F}$. We get the function $(\lambda f)\colon V \longrightarrow W$ given by $(\lambda f)(v) = \lambda f(v)$. We check that this is linear:
$$\begin{aligned}(\lambda f)(v_1 + v_2) &= \lambda f(v_1 + v_2) = \lambda f(v_1) + \lambda f(v_2) = (\lambda f)(v_1) + (\lambda f)(v_2)\\ (\lambda f)(\mu v) &= \lambda f(\mu v) = \lambda \mu f(v) = \mu(\lambda f)(v).\end{aligned}$$
So $(\lambda f) \in \mathrm{Hom}(V, W)$. □

TODO: define $\mathbb{F}$-algebra.

Example A.3. Let V be a vector space over $\mathbb{F}$. Prove that $\mathrm{End}(V) := \mathrm{Hom}(V, V)$ is an associative unital $\mathbb{F}$-algebra.

Solution. TODO □

A.2.1. Dual vector space

Let V be a finite dimensional vector space over $\mathbb{F}$. Define

$$V^\vee = \mathrm{Hom}(V, \mathbb{F}).$$

By Example A.2 we know that this is a vector space over $\mathbb{F}$. It is called the *dual vector space* to V. Its elements are sometimes called *(linear) functionals* and denoted with Greek letters such as φ.

Example A.4. Suppose $B = \{v_1, \ldots, v_n\}$ is a basis for V. Define $v_1^\vee, \ldots, v_n^\vee \in \mathrm{Fun}(V, \mathbb{F})$ by

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

Show that $v_i^\vee \in V^\vee$ for $i = 1, \ldots, n$ and that the set $B^\vee = \left\{v_1^\vee, \ldots, v_n^\vee\right\}$ is a basis for $V^\vee$. It is called the *dual basis* to B.

Solution. 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 $\lambda \in \mathbb{F}$ 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 $i = 1, \ldots, n$.

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 $i = 1, \ldots, n$ and get

$$\lambda_i = 0.$$

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

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

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

Then, since φ is a linear transformation, we have

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

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

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

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

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

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

Note that β induces linear maps

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

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

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

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

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

Example A.5. Prove that if V is finite-dimensional, then $\beta_V\colon V \longrightarrow \big(V^\vee\big)^\vee$ is invertible.

Solution. 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 Example A.4.

To show that β_V is injective, suppose $u, v \in V$ are such that $\beta_V(u) = \beta_V(v)$, in other words

$$\varphi(u) = \varphi(v) \qquad \text{for all } \varphi \in V^\vee.$$

Write

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

then, for $i = 1, \dots, n$, 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 Example A.4 tells us that V and $V^\vee$, and therefore also $\big(V^\vee\big)^\vee$, have the same dimension n; so β_V, being injective, is also surjective.)

Let $T\colon V^\vee \longrightarrow \mathbb{F}$ be a linear transformation. Define $v \in V$ by

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

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

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

where we expressed φ in terms of the dual basis $v_1^\vee, \dots, v_n^\vee$ from Example A.4. $\square$

Example A.6. Consider a linear transformation $T\colon V \longrightarrow W$, where W is another finite-dimensional vector space over $\mathbb{F}$. Define $T^\vee\colon W^\vee \longrightarrow V^\vee$ by

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

Prove that $T^\vee$ is a linear transformation. It is called the *dual linear transformation* to T.

Solution. It is clear that $\varphi \circ T\colon V \longrightarrow \mathbb{F}$ 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 $\lambda \in \mathbb{F}$, then for any $v \in V$ we have

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

Example A.7. In the setup of Example A.6, suppose $W = V$ so that $T\colon V \longrightarrow V$ and $T^\vee\colon V^\vee \longrightarrow V^\vee$.

Let M be the matrix representation of T with respect to an ordered basis B of V, and let $M^\vee$ be the matrix representation of $T^\vee$ with respect to the dual basis $B^\vee$.

Express $M^\vee$ in terms of M.

Solution. As in Example A.4, we have $B = \big(v_1, \dots, v_n\big)$ and $B^\vee = \big(v_1^\vee, \dots, v_n^\vee\big)$. Write $\big(a_{ij}\big)$ for the entries of the matrix M. For future reference, the i-th row of M is

$$\begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{bmatrix}.$$

By the definition of matrix representations, we have

$$\begin{aligned} T(v_1) &= a_{11}v_1 + a_{21}v_2 + \dots + a_{n1}v_n \\ T(v_2) &= a_{12}v_1 + a_{22}v_2 + \dots + a_{n2}v_n \\ &\vdots \\ T(v_n) &= a_{1n}v_1 + a_{2n}v_2 + \dots + a_{nn}v_n. \end{aligned}$$

The i-th column of $M^\vee$ is given by the $B^\vee$-coordinates of the vector $T^\vee(v_i^\vee) = v_i^\vee \circ T$. To determine these, we apply $v_i^\vee \circ T$ to the basis vectors $v_1, \dots, v_n$:

$$T^\vee\big(v_i^\vee\big)(v_j) = \big(v_i^\vee \circ T\big)(v_j) = v_i^\vee\big(T(v_j)\big) = v_i^\vee\big(a_{1j}v_1 + a_{2j}v_2 + \dots + a_{nj}v_n\big) = a_{ij}.$$

This means that

$$T^\vee\big(v_i^\vee\big) = a_{i1}v_1^\vee + a_{i2}v_2^\vee + \dots + a_{in}v_n^\vee$$

and the i-th column of $M^\vee$ is

$$\begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{bmatrix},$$

precisely the i-th row of M.

We conclude that $M^\vee = M^T$, the transpose of the matrix M. □

Example A.8. Let $v_1, \dots, v_n \in V$. Define $\Gamma \colon V^\vee \longrightarrow \mathbb{F}^n$ by

$$\Gamma(\varphi) = \begin{bmatrix} \varphi(v_1) \\ \vdots \\ \varphi(v_n) \end{bmatrix}.$$

(a) Prove that Γ is a linear transformation.

(b) Prove that Γ is injective if and only if $\{v_1, \dots, v_n\}$ spans V.

(c) Prove that Γ is surjective if and only if $\{v_1, \dots, v_n\}$ is linearly independent.

Solution.

(a) Given $\varphi_1, \varphi_2 \in V^\vee$, we have

$$\begin{aligned} \Gamma(\varphi_1 + \varphi_2) &= ((\varphi_1 + \varphi_2)(v_1), \dots, (\varphi_1 + \varphi_2)(v_n)) \\ &= (\varphi_1(v_1), \dots, \varphi_1(v_n)) + (\varphi_2(v_1), \dots, \varphi_2(v_n)) \\ &= \Gamma(\varphi_1) + \Gamma(\varphi_2). \end{aligned}$$

Given $\varphi \in V^\vee$ and $\lambda \in \mathbb{F}$, we have

$$\begin{aligned} \Gamma(\lambda\varphi) &= ((\lambda\varphi)(v_1), \dots, (\lambda\varphi)(v_n)) \\ &= (\lambda\varphi(v_1), \dots, \lambda\varphi(v_n)) \\ &= \lambda\Gamma(\varphi). \end{aligned}$$

(b) Suppose Γ is injective. Let $W = \operatorname{Span}\{v_1, \dots, v_n\}$. We want to prove that $W = V$.

Suppose $W \neq V$. Let $C = \{w_1, \dots, w_k\}$ be a basis of W and extend it to a basis $B = \{w_1, \dots, w_k, w_{k+1}, \dots, w_m\}$ of V.

Let $B^\vee$ be the dual basis to B and consider its last element $v_m^\vee$ given by

$$v_m^\vee(a_1 w_1 + \cdots + a_m w_m) = a_m.$$

Then $v_m^\vee \neq 0$ (since $v_m^\vee(w_m) = 1$, for instance) but $v_m^\vee(w) = 0$ for all $w \in W$. In particular, $v_m^\vee(v_1) = \cdots = v_m^\vee(v_n) = 0$, so $\Gamma(v_m^\vee) = 0$, contradicting the injectivity of Γ.

We conclude that $W = V$, in other words $\{v_1, \dots, v_n\}$ spans V.

Conversely, suppose $\{v_1, \dots, v_n\}$ spans V. If $\varphi_1, \varphi_2 \in V^\vee$ are such that $\Gamma(\varphi_1) = \Gamma(\varphi_2)$, then $\Gamma(\varphi_1 - \varphi_2) = 0$, so setting $\varphi = \varphi_1 - \varphi_2$, we want to show that $\varphi = 0$, the constant zero function.

If $\varphi \neq 0$, then there exists $v \in V - \{0\}$ such that $\varphi(v) \neq 0$. Since $\{v_1, \dots, v_n\}$ spans V, then we can write v as

$$v = b_1 v_1 + \cdots + b_n v_n.$$

But $\Gamma(\varphi) = 0$, so

$$0 \neq \varphi(v) = b_1\varphi(v_1) + \cdots + b_n\varphi(v_n) = 0,$$

which is a contradiction. So we must have $\varphi = 0$, that is $\varphi_1 = \varphi_2$. We conclude that Γ is injective.

(c) Suppose $\Gamma\colon V^\vee \longrightarrow \mathbb{F}^n$ is surjective. Let

$$a_1v_1 + \cdots + a_nv_n = 0$$

be a linear relation.

Let $i \in \{1,\ldots,n\}$. Since Γ is surjective, given the standard basis vector $e_i \in \mathbb{F}^n$ (1 in the i-th entry), there exists $\varphi_i \in V^\vee$ such that $\Gamma(\varphi_i) = e_i$. If we apply φ_i on both sides of the linear relation, we get

$$a_i = 0.$$

Since this holds for all i, the relation is trivial.

Conversely, suppose $\{v_1,\ldots,v_n\}$ is linearly independent. This set can be enlarged to a basis $B = \{v_1,\ldots,v_n,v_{n+1},\ldots,v_m\}$ of V, with dual basis $v_1^\vee,\ldots,v_m^\vee$.

Now take an arbitrary vector in $\mathbb{F}^n$:

$$w = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}.$$

Let

$$\varphi = a_1v_1^\vee + \cdots + a_nv_n^\vee,$$

then

$$\Gamma(\varphi) = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = w.$$

We conclude that Γ is surjective. □

Here's a concrete example of a naturally-occurring linear functional:

Example A.9. Let $V = \mathbb{F}[x]$ be the vector space of polynomials in one variable with coefficients in $\mathbb{F}$. Given a scalar $\alpha \in \mathbb{F}$, consider the function $\mathrm{ev}_\alpha\colon V \longrightarrow \mathbb{F}$ given by evaluation at α:

$$\mathrm{ev}_\alpha(f) = f(\alpha).$$

Prove that $\mathrm{ev}_\alpha \in V^\vee$.

Solution. We have to prove that $\mathrm{ev}_\alpha\colon V \longrightarrow \mathbb{F}$ is linear.

If $f_1, f_2 \in \mathbb{F}[x]$, then

$$\mathrm{ev}_\alpha(f_1 + f_2) = (f_1 + f_2)(\alpha) = f_1(\alpha) + f_2(\alpha) = \mathrm{ev}_\alpha(f_1) + \mathrm{ev}_\alpha(f_2).$$

If $f \in \mathbb{F}[x]$ and $\lambda \in \mathbb{F}$, then

$$\mathrm{ev}_\alpha(\lambda f) = (\lambda f)(\alpha) = \lambda f(\alpha) = \lambda\, \mathrm{ev}_\alpha(f).$$ □

A.2.2. Inner products

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

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

An *inner product* on V is a function

$$\langle\cdot,\cdot\rangle\colon V\times V\longrightarrow\mathbb{F}$$

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 $\alpha\in\mathbb{F}$;

(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. (Such a function $V\times V\longrightarrow\mathbb{F}$ 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 $\mathbb{F}$ and $\langle\cdot,\cdot\rangle$ is an inner product on V.

Example A.10. The prototypical inner product on $\mathbb{C}^n$ is

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

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

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

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

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

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

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

Over $\mathbb{R}$, A is a positive-definite symmetric matrix.

Define

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

Proposition A.11 (Cauchy–Schwarz Inequality)**.** *Let u, v be vectors in an inner product space V. Then*

$$|\langle u,v\rangle|\leqslant\|u\|\,\|v\|,$$

where equality holds if and only if u and v are parallel.

Proof. If $u=\mathbf{0}$ or $v=\mathbf{0}$, we have the equality $0=0$. Otherwise, for any $t\in\mathbb{F}$ we have

$$\begin{aligned}0\leqslant\langle u-tv,u-tv\rangle&=\langle u,u\rangle-2\operatorname{Re}\big(\overline{t}\langle u,v\rangle\big)+t\overline{t}\langle v,v\rangle\\&=\|u\|^2-2\operatorname{Re}\big(\overline{t}\langle u,v\rangle\big)+|t|^2\|v\|^2.\end{aligned}$$

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

A.3. Optional: A tour of the p-adics

Let p be a prime number.

We have introduced in Example 2.1 the p-adic absolute value $|\cdot|_p \colon \mathbb{Q} \longrightarrow \mathbb{Q}_{\geqslant 0}$ and the corresponding p-adic metric d_p on $\mathbb{Q}$. We have also seen in Example 2.7 and Exercise 2.7 that the resulting geometry on $\left(\mathbb{Q}, d_p\right)$ is very strange: every triangle is isosceles, every point in an open ball is a centre, every open ball is also closed.

The aim of this section is to showcase some more properties of the p-adics. For much more detail than we can possibly include here, an excellent starting point is Fernando Gouvêa's book [2].

Let's start by formalising what we want an *absolute value* on a field K to be: a function

$$|\cdot| \colon K \longrightarrow \mathbb{R}_{\geqslant 0}$$

that satisfies

(a) (non-degeneracy) $|x| = 0$ if and only if $x = 0$;

(b) (multiplicativity) $|xy| = |x|\,|y|$ for all $x, y \in K$;

(c) (triangle inequality) $|x+y| \leqslant |x| + |y|$ for all $x, y \in K$.

We say that an absolute value is *non-archimedean* if it satisfies the following strengthening of the triangle inequality:

$$|x+y| \leqslant \max\left\{|x|, |y|\right\} \qquad \text{for all } x, y \in K.$$

Otherwise, we say that $|\cdot|$ is *archimedean*.

Example A.12. The real absolute value $|\cdot| \colon \mathbb{R} \longrightarrow \mathbb{R}_{\geqslant 0}$ and the complex absolute value $|\cdot| \colon \mathbb{C} \longrightarrow \mathbb{R}_{\geqslant 0}$ are archimedean absolute values.

The p-adic absolue value $|\cdot| \colon \mathbb{Q} \longrightarrow \mathbb{R}_{\geqslant 0}$ is a non-archimedean absolute value.

Example A.13. The cheapest way to get an absolute value on any field K is to set

$$|x| = \begin{cases} 1 & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

This is called the *trivial absolute value* on K. The corresponding metric on K is the discrete metric. We will typically exclude it from our considerations.

Example A.14. For any field K, there is a unique ring homomorphism $j \colon \mathbb{Z} \longrightarrow K$ determined by $j(1) = 1$. An absolute value $|\cdot|$ on K is non-archimedean if and only if $j(\mathbb{Z})$ is a bounded subset of K with respect to the metric defined by $|\cdot|$.

We say that two absolute values on K are *equivalent* if their corresponding metrics on K are equivalent.

Theorem A.15 (Ostrowski)**.** *Let $|\cdot|_?$ be a non-trivial absolute value on $\mathbb{Q}$. Then $|\cdot|_?$ is equivalent to the real absolute value $|\cdot|$ on $\mathbb{Q}$, or to the p-adic absolute value $|\cdot|_p$ for some prime number p.*

Example A.16. If $p \neq q$ are two distinct primes, then $|\cdot|_p$ and $|\cdot|_q$ on $\mathbb{Q}$ are not equivalent. If p is a prime number, then $|\cdot|_p$ and the real absolute value $|\cdot|$ on $\mathbb{Q}$ are not equivalent.

Solution. Consider the sequence (x_n) in $\mathbb{Q}$ given by $x_n = p^n$. We have:

$$|p|_p = \frac{1}{p}, \qquad |p|_q = 1, \qquad |p| = p,$$

so that

$$|x_n|_p = \frac{1}{p^n}, \qquad |x_n|_q = 1, \qquad |x_n| = p^n,$$

hence the sequence (x_n)

- converges to 0 with respect to $|\cdot|_p$;
- converges to 1 with respect to $|\cdot|_q$;
- diverges with respect to $|\cdot|$.

We conclude that these absolute values are not equivalent. □

There's a magical way in which all the different absolute values on $\mathbb{Q}$ fit together:

Proposition A.17 (Product Formula)**.** *For any $x \in \mathbb{Q}$, $x \neq 0$ we have*

$$|x| \prod_{p \text{ prime}} |x|_p = 1.$$

Proof. Writing $x = a/b$ in lowest terms, we notice that it suffices to prove the product formula for a positive integer a. If $a = 1$ the formula is clearly true.

If $a > 1$ it follows easily from the unique prime factorisation of integers: write

$$a = p_1^{e_1} \dots p_n^{e_n},$$

then

$$|a|_p = \begin{cases} p_j^{-e_j} & \text{if } p = p_j, j = 1, \dots, n \\ 1 & \text{otherwise.} \end{cases}$$

Hence

$$|a| \prod_{p \text{ prime}} |a|_p = p_1^{e_1} \dots p_n^{e_n} \prod_{j=1}^{n} p_j^{-e_j} = 1.$$ □

Let's go back to fixing a prime p. Then the p-adic absolute value and the corresponding metric d_p make $(\mathbb{Q}, d_p)$ into a metric space. As such, there is a completion denoted $(\mathbb{Q}_p, \widehat{d_p})$ and an embedding $\iota\colon \mathbb{Q} \longrightarrow \mathbb{Q}_p$ such that $\iota(\mathbb{Q})$ is a dense subset of $\mathbb{Q}_p$. *A priori*, $\mathbb{Q}_p$ is just a set, but one can proceed in a manner similar to Proposition 3.14 and make it into a field, in such a way that $\iota\colon \mathbb{Q} \longrightarrow \mathbb{Q}_p$ is a field homomorphism. Moreover, we can define an absolute value on $\mathbb{Q}_p$ by

$$\big|[(q_n)]\big|_p = \lim_{n \longrightarrow \infty} \big(|q_n|_p\big).$$

The elements of $\mathbb{Q}_p$ are called *p-adic numbers*. Just as in everyday life we are not working with real numbers as equivalence classes of Cauchy sequences of rationals, but rather as decimal expansions, p-adic numbers are more easily manipulated in the form

$$x = \sum_{n \geqslant -m} b_n p^n = b_{-m} p^{-m} + \dots + b_{-1} p^{-1} + b_0 + b_1 p + b_2 p^2 + b_3 p^3 + O(p^4),$$

where $m \in \mathbb{Z}$, and the *p-adic digits* of x satisfy $0 \leqslant b_n \leqslant p-1$ for all $n \geqslant -m$, $b_{-m} \neq 0$. The *valuation* of x is $v_p(x) = -m$, and $|x|_p = p^m$.

This looks a bit like a formal Laurent series in the "variable" p. The elements that look like a formal power series (in other words, with $m \leqslant 0$) are called *p-adic integers*.

To recover a Cauchy sequence of rationals from one of these formal Laurent series, take successive truncations of the series. Given

$$x = \sum_{n \geqslant -m} b_n p^n,$$

let

$$\begin{aligned} x_1 &= b_{-m}p^{-m} \\ x_2 &= b_{-m}p^{-m} + b_{-m+1}p^{-m+1} \\ x_3 &= b_{-m}p^{-m} + b_{-m+1}p^{-m+1} + b_{-m+2}p^{-m+2} \end{aligned}$$

and so on. It is easy to see that the resulting rational sequence (x_n) is Cauchy with respect to the p-adic absolute value.

Example A.18. In $\mathbb{Q}_7$ we have

$$\begin{aligned} \frac{444}{49} &= 7^{-2} \cdot 444 = 7^{-2}\big(3 + 2 \cdot 7^2 + 1 \cdot 7^3\big) = 3 \cdot 7^{-2} + 2 + 1 \cdot 7 \\ -1 &= 6 + 6 \cdot 7 + 6 \cdot 7^2 + 6 \cdot 7^3 + 6 \cdot 7^4 + 6 \cdot 7^5 + 6 \cdot 7^6 + 6 \cdot 7^7 + O(7^8) \\ \sqrt{2} &= 3 + 1 \cdot 7 + 2 \cdot 7^2 + 6 \cdot 7^3 + 1 \cdot 7^4 + 2 \cdot 7^5 + 1 \cdot 7^6 + 2 \cdot 7^7 + O(7^8) \\ \sqrt{3} &\text{ does not exist.} \end{aligned}$$

Here the last line should be taken to assert that the equation $x^2 = 3$ has no solutions in $\mathbb{Q}_7$, while the second to last line should be taken to assert that the right hand side of the equality is a 7-adic integer x with the property that $x^2 = 2$. There is another 7-adic integer with the same property, namely:

$$4 + 5 \cdot 7 + 4 \cdot 7^2 + 0 \cdot 7^3 + 5 \cdot 7^4 + 4 \cdot 7^5 + 5 \cdot 7^6 + 4 \cdot 7^7 + O(7^8).$$

This also brings us to the question of performing arithmetic operations on p-adic numbers in this form, which is done in the same way that one treats formal Laurent series, with the exception that one has to carry p-adic digits when they overflow the bounds $0 \leqslant b_n \leqslant p-1$.

For instance, the two alleged 7-adic square roots of 2 really ought to add up to 0, right?

$$\begin{aligned} &\big(3 + 1 \cdot 7 + 2 \cdot 7^2 + 6 \cdot 7^3 + 1 \cdot 7^4 + 2 \cdot 7^5 + 1 \cdot 7^6 + 2 \cdot 7^7 + O(7^8)\big) \\ &\qquad + \big(4 + 5 \cdot 7 + 4 \cdot 7^2 + 0 \cdot 7^3 + 5 \cdot 7^4 + 4 \cdot 7^5 + 5 \cdot 7^6 + 4 \cdot 7^7 + O(7^8)\big) = O(7^8). \end{aligned}$$

And each of them should square to 2:

$$\big(3 + 1 \cdot 7 + 2 \cdot 7^2 + 6 \cdot 7^3 + 1 \cdot 7^4 + 2 \cdot 7^5 + 1 \cdot 7^6 + 2 \cdot 7^7 + O(7^8)\big)^2 = 2 + O(7^8).$$

How on Earth does one prove these claims about solving polynomial equations in $\mathbb{Q}_p$? Here the p-adic world offers an elegant tool that has only imperfect reflections in the reals:

Theorem A.19 (Hensel's Lemma)**.** *Let*

$$f(x) = a_n x^n + \cdots + a_1 x + a_0 \in \mathbb{Z}_p[x]$$

be a polynomial with p-adic integer coefficients.

If there exists $b_0 \in \{0, 1, \ldots, p-1\}$ such that

$$f(b_0) \equiv 0 \pmod p \qquad \textit{but} \qquad f'(b_0) \not\equiv 0 \pmod p,$$

then there exists a unique $y \in \mathbb{Z}_p$ such that

$$f(y) = 0 \qquad \textit{and} \qquad y = b_0 + O(p).$$

The proof of Hensel's Lemma is constructive: the solution $y \in \mathbb{Z}_p$ is built iteratively, one p-adic digit at a time.

I'll demonstrate how it works in the example that lead us here:

Example A.20. Solve $x^2 = 2$ in $\mathbb{Q}_7$.

Solution. Let $f(x) = x^2 - 2$, then $f'(x) = 2x$. There are two values of b_0 with the required property: $b_0 = 3, 4$. Check:

$$\begin{aligned} f(3) - 2 &= 3^2 - 2 = 9 - 2 = 7, & f'(3) &= 6 \\ f(4) - 2 &= 4^2 - 2 = 16 - 2 = 14, & f'(4) &= 8. \end{aligned}$$

Let's pick one of them: $b_0 = 3$. Setting $y_1 = b_0 = 3$, we have

$$f(y_1) = 3^2 - 2 = 7 \equiv 0 \pmod 7.$$

We have $f(y_1) = 7 \cdot 1$; let $x_1 = 1$ and

$$b_1 \equiv -\frac{x_1}{f'(b_0)} \equiv -\frac{1}{6} \equiv -6 \equiv 1 \pmod 7.$$

Let $y_2 = b_0 + b_1 \cdot 7 = 3 + 1 \cdot 7$, then

$$f(y_2) = y_2^2 - 2 = 98 = 2 \cdot 7^2 \equiv 0 \pmod{7^2}.$$

We have $f(y_2) = 7^2 \cdot 2$; let $x_2 = 2$ and

$$b_2 \equiv -\frac{x_2}{f'(b_0)} \equiv 2 \pmod 7.$$

Let $y_3 = b_0 + b_1 \cdot 7 + b_2 \cdot 7^2 = 3 + 1 \cdot 7 + 2 \cdot 7^2$, then

$$f(y_3) = y_3^2 - 2 = 11662 = 34 \cdot 7^3 \equiv 0 \pmod{7^3}.$$

And so on. □

If this whole discussion of recursively constructing a unique solution reminds you of the Banach Fixed Point Theorem (Theorem 2.46), you'll be happy to hear that there is a proof of Hensel's Lemma that uses the Banach Fixed Point Theorem, see [1, Section 6].

Several of you have asked whether our study of normed spaces over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ can be done over other fields as well. Although we won't pursue the topic in this subject, one can definitely look at normed spaces over $\mathbb{Q}_p$: just take the definition of norm in Section 3.1 and replace any appearance of the usual absolute value by the p-adic absolute value $|\cdot|_p$. Many of the results we proved for real and complex normed spaces are also valid in this p-adic setting, including the fact that all norms on a finite-dimensional vector space are equivalent, and that finite-dimensional normed spaces are complete. For a proof, see [2, Theorem 5.2.1]; some of the arguments are the same as the ones we used, but others do not carry over to $\mathbb{Q}_p$ and need replacement. In contrast, there is (as far as I know) no reasonable theory of p-adic Hilbert spaces. Nonetheless, there is a well-developed theory of p-adic functional analysis, which is heavily used in certain parts of number theory and representation theory.

Bibliography

[1] Keith Conrad. Hensel's lemma. https://kconrad.math.uconn.edu/blurbs/gradnumthy/hensel.pdf, University of Connecticut lecture notes. [Online; accessed 1 September 2023].

[2] Fernando Q. Gouvêa. *p-adic numbers. An introduction.* Universitext. Springer, Cham, third edition, 2020.

[3] Alexander Nagel. Convexity and the sign of the second derivative. https://people.math.wisc.edu/~ajnagel/convexity.pdf, University of Wisconsin lecture notes, 2010. [Online; accessed 25 August 2023].

[4] Karen E. Smith. Bases for infinite dimensional vector spaces. https://dept.math.lsa.umich.edu/~kesmith/infinite.pdf, University of Michigan lecture notes, 1999. [Online; accessed 21 July 2023].

Index