

NOTES ON METRIC AND HILBERT SPACES AN INVITATION TO FUNCTIONAL ANALYSIS

Alexandru Ghitza*

School of Mathematics and Statistics
University of Melbourne

Version of Sat 11th Oct, 2025 at 15:46

*(aghitza@alum.mit.edu)

CONTENTS

1. INTRODUCTION	5
1.1. Infinite-dimensional spaces?	5
1.2. Plan(-ish)	6
1.3. Notations and conventions	7
2. METRIC AND TOPOLOGICAL SPACES	9
2.1. Metrics	9
2.2. Open subsets of metric spaces	11
2.3. Topological spaces and continuous functions	12
2.4. Properties of topologies induced by metrics	16
2.5. Connectedness	17
2.6. Compactness	20
2.7. Sequences in metric spaces	23
2.8. Cauchy sequences	25
2.9. Completions	26
2.10. Banach Fixed Point Theorem	30
2.11. (Total) Boundedness	32
2.12. Compactness in metric spaces	34
2.13. Spaces of bounded continuous functions	35
2.14. Function spaces: approximation by polynomials	37
2.15. Intersection of dense open sets	40
3. NORMED AND HILBERT SPACES	43
3.1. Norms	43
3.2. Inner products	47
3.3. Banach and Hilbert spaces	48
3.4. Bounded linear transformations	51
3.5. Convexity and inequalities	54
3.6. Sequence spaces	55
3.7. Dual normed spaces and completeness of sequence spaces	57
3.8. Orthogonality and projections	60
A. APPENDIX: PREREQUISITES	65
A.1. Equivalence relations	65
A.2. Cardinality	66
A.3. Maps between vector spaces	67
A.4. Inner products	68
A.5. Uniform continuity and uniform convergence	69
B. APPENDIX: MISCELLANEOUS	73
B.1. Zorn's Lemma	73
B.2. Bilinear maps	73
B.3. Dual vector space	73
B.4. A diversion: Topological groups	75

1. INTRODUCTION

1.1. INFINITE-DIMENSIONAL SPACES?

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

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

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

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

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

First year linear algebra tells us that every finite-dimensional vector space V has a basis¹. What happens if V is not finite-dimensional?

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

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

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

thus a linear combination of elements of B .

If we have

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

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

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

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

though the space of polynomials is not finite-dimensional, each element of the space is in some sense “finitely generated”.

Something we can try is to start with the prototypical finite-dimensional spaces we know, say $\mathbf{R}^n$, and “take the limit as $n \rightarrow \infty$ ”. This leads us to consider the space $\mathbf{R}^{\mathbf{N}}$ of arbitrary real sequences $(x_1, x_2, \dots)$. We may naively hope that, since $\{e_1, e_2, \dots, e_n\}$ is a basis for $\mathbf{R}^n$, and these standard bases nest nicely as n increases, we end up with $\{e_1, e_2, \dots\}$ being a basis for $\mathbf{R}^{\mathbf{N}}$, but that is not the case because, for instance, the constant sequence $(1, 1, \dots)$ is not in the span of $\{e_1, e_2, \dots\}$. (See [Exercise B.5](#) for more details.)

For another example, take $V = \mathbf{R}$ viewed as a vector space over $\mathbf{Q}$. One can show that the set $S = \{\sqrt{n} : n \in \mathbf{N} \text{ squarefree}\}$ is $\mathbf{Q}$ -linearly independent in $\mathbf{R}$, but not a basis. The same is true of the set $T = \{\pi^n : n \in \mathbf{N}\}$. (See [Exercise B.6](#).) In fact, $\mathbf{R}$ has no countable basis over $\mathbf{Q}$. (See [Exercise A.9](#).) It’s a sign that it may be rather difficult to write down an explicit $\mathbf{Q}$ -basis of $\mathbf{R}$.

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

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

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

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

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

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

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

1.2. PLAN(-ISH)

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

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

Once we have a grasp on the behaviour of general metric spaces and their topology, we consider the special case where the underlying set has a vector space structure. These are

called **normed vector spaces** (in this setting, it is customary to single out the norm of a vector rather than the distance between two vectors; the two are equivalent).

Finally, because of their importance in many applications, we specialise further to **inner product spaces**. One natural example is the space $V = \text{Cts}([- \pi, \pi], \mathbf{R})$ of continuous functions $f: [- \pi, \pi] \rightarrow \mathbf{R}$, endowed with the inner product

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

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

The distance function is of course

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

This allows us to bring rigorous meaning to statements such as

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

In our setting, we have

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

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

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

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

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

1.3. NOTATIONS AND CONVENTIONS

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

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

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

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

For better or worse, the natural numbers

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

start at 0. The variant starting at 1 is

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

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

I am not the right person to ask about foundational questions of logic or set theory: I neither know enough nor care sufficiently about the topic. It's of course okay if you care and (want to) know more about these things. I am happy to spend my mathematical life in ZFC (Zermelo–Fraenkel set theory plus the Axiom of Choice), and these notes are part of my life so they are also hanging out in ZFC. In particular, I am very likely to use the Axiom of Choice without comment (and sometimes without noticing); I may occasionally point it out if someone brings my attention to it.

ACKNOWLEDGEMENTS

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

2. METRIC AND TOPOLOGICAL SPACES

2.1. METRICS

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

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

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

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

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

$$d: X \times X \longrightarrow \mathbf{R}_{\geq 0}$$

such that:

- (a) $d(x, y) = d(y, x)$ for all $x, y \in X$;
- (b) $d(x, y) \leq d(x, t) + d(t, y)$ for all $x, y, t \in X$;
- (c) $d(x, y) = 0$ with $x, y \in X$ if and only if $x = y$.

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

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

Example 2.1. Let Γ be a connected undirected simple graph (finitely many vertices, each pair of which are joined by at most one undirected edge; no loops). Given vertices x and y , we let $d(x, y)$ denote the minimum length of any path joining x and y .

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

Solution.

- (a) Symmetry follows directly from the fact that Γ is undirected.
- (b) Let $x, y, t \in \Gamma$, let p_1 be a shortest path (of length $d(x, t)$) joining x and t , and p_2 a shortest path (of length $d(t, y)$) joining t and y . Concatenating p_1 and p_2 we get a path of length $d(x, t) + d(t, y)$ from x to y , therefore $d(x, y)$ is at most equal to this length.
- (c) Clear (if $x = y$ then the empty path goes from x to y ; conversely, if $d(x, y) = 0$ then there is an empty path joining x to y , forcing $x = y$). $\square$

Other examples are quite exotic, see for instance the p -adic metric in [Exercise 1.10](#).

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

Example 2.2. If (X, d) is a metric space, then for any subset S of X , the restriction of d to S gives a metric on S . (This is called the *induced metric*.)

Solution. Straightforward (follows immediately from the definitions). $\square$

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

Example 2.3. Let (X_1, d_{X_1}) and (X_2, d_{X_2}) denote two metric spaces. Prove that the function d_1 defined by

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

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

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

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

Solution. Straightforward. $\square$

Example 2.4. Same setup as [Example 2.3](#), but with the function

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

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

Solution. Straightforward; proving the triangle inequality uses

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

$\square$

Example 2.5. Take $X_1 = X_2 = \mathbf{R}$ with the Euclidean metric and convince yourself that neither d_1 from [Example 2.3](#) nor d_∞ from [Example 2.4](#) is the Euclidean metric on $\mathbf{R}^2$.

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

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

$\square$

Not every metric has to do with lengths and geometry in an obvious way. The p -adic metric in [Exercise 1.10](#) is an example of something a little different. For another example, let $n \in \mathbf{Z}_{\geq 1}$, $X = \mathbf{F}_2^n$, and let $d(x, y)$ be the number of indices $i \in \{1, \dots, n\}$ such that $x_i \neq y_i$. Then d is a metric on X ; it is called the *Hamming metric*. See [Exercise 1.6](#) for more details.

2.2. OPEN SUBSETS OF METRIC SPACES

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

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

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

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

and a corresponding *sphere*

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

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

Example 2.6. Describe the Euclidean open balls centred at 0 in $\mathbf{Z}$ (endowed with the metric induced from the Euclidean metric on $\mathbf{R}$).

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

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

For more intuition-challenging examples, see [Exercises 1.4](#) and [1.11](#).

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

Example 2.7. Prove that $\emptyset$ and X are open sets.

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

Example 2.8. Prove that any open ball is an open set.

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

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

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

What happens if we combine open sets using set operations?

Proposition 2.9. *Let X be a metric space. The union of an arbitrary collection of open sets is an open set.*

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

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

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

Intersections are a bit more delicate:

Proposition 2.10. *Let X be a metric space. The intersection of a finite collection of open sets is an open set.*

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

Wondering about the necessity of the word “finite” in the statement of the proposition? See [Tutorial Question 2.2](#).

2.3. TOPOLOGICAL SPACES AND CONTINUOUS FUNCTIONS

Let X be a set. Taking a hint from the previous section, we define a *topology* on X to be a subset $\mathcal{T} \subseteq \mathcal{P}(X)$ (in other words, $\mathcal{T}$ is a collection of subsets of X) such that

- (a) $\emptyset \in \mathcal{T}$ and $X \in \mathcal{T}$;
- (b) if $\{U_i : i \in I\}$ is an arbitrary collection of elements of $\mathcal{T}$, then $\bigcup_{i \in I} U_i \in \mathcal{T}$;
- (c) if $\{U_1, \dots, U_n\}$ is a finite collection of elements of $\mathcal{T}$, then $\bigcap_{j=1}^n U_j \in \mathcal{T}$.

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

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

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

The prototypical example (for us) of a topology is the collection of open sets defined on X by some distance function d (see [Proposition 2.24](#) for details). However, topological spaces are a very general concept encompassing much more than metric spaces. We will not place a heavy emphasis on them in this subject, using them mostly to separate those properties of metric spaces that actually depend on the metric from those that depend only on the configuration of open subsets.

Example 2.11. Let X be an arbitrary set and let $\mathcal{T} = \{\emptyset, X\}$. This is called the *trivial topology* on X .

Example 2.12. Let X be an arbitrary set and let $\mathcal{T} = \mathcal{P}(X)$. (Every subset is an open subset.) This is called the *discrete topology* on X .

Example 2.13. Let X be an arbitrary set and let

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

This is called the *cofinite topology* on X .

COMPARING TOPOLOGIES

If $\mathcal{T}_1$ and $\mathcal{T}_2$ are two topologies on the same set X and $\mathcal{T}_1 \subseteq \mathcal{T}_2$ we say that $\mathcal{T}_1$ is *coarser* than $\mathcal{T}_2$ and $\mathcal{T}_2$ is *finer* than $\mathcal{T}_1$. For example, the discrete topology is the finest possible topology on X , and the trivial topology is the coarsest possible topology on X .

In [Tutorial Question 2.3](#) you will find all (4) possible topologies on a set with two elements. (This game quickly becomes complicated as the size of the set increases, for instance a set of three elements has 29 distinct topologies.)

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

Example 2.14. Let X be a set and $S \subseteq \mathcal{P}(X)$. The *topology generated by S* is obtained by letting S' consist of all finite intersections of elements of S , then letting $\mathcal{T}$ consist of all arbitrary unions of elements of S' .

For instance, the discrete topology on X is generated by the set of singletons (aka one-point subsets).

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

CONTINUOUS FUNCTIONS

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

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

Proposition 2.15. Let X be a set and $\mathcal{T}_1, \mathcal{T}_2$ two topologies on X . The following statements are equivalent:

- (a) $\mathcal{T}_2$ is coarser than $\mathcal{T}_1$ (that is, $\mathcal{T}_2 \subseteq \mathcal{T}_1$);
- (b) for any $x \in X$ and any $\mathcal{T}_2$ -open neighbourhood U_x^2 of x , there exists a $\mathcal{T}_1$ -open neighbourhood U_x^1 of x such that $U_x^1 \subseteq U_x^2$;
- (c) the function $f: (X, \mathcal{T}_1) \rightarrow (X, \mathcal{T}_2)$ given by $f(x) = x$ is continuous.

Proof. See [Exercise 1.19](#). □

NEW TOPOLOGICAL SPACES FROM OLD

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

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

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

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

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

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

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

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

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

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

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

Let $R = U_1 \times U_2$, $R' = U'_1 \times U'_2$ be two rectangles. Then

$$\begin{aligned} R \cap R' &= \{(x_1, x_2) \in X_1 \times X_2 : x_1 \in U_1, x_2 \in U_2\} \cap \{(x_1, x_2) \in X_1 \times X_2 : x_1 \in U'_1, x_2 \in U'_2\} \\ &= \{(x_1, x_2) \in X_1 \times X_2 : x_1 \in U_1 \cap U'_1, x_2 \in U_2 \cap U'_2\} \\ &= (U_1 \cap U'_1) \times (U_2 \cap U'_2). \end{aligned}$$

□

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

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

are continuous.

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

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

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

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

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

SEPARATION PROPERTY: HAUSDORFF

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

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

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

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

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

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

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

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

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

INTERIOR, CLOSURE

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

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

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

- (a) the *interior* A° of A to be the union of all open subsets of A , equivalently the largest open subset of A ;
- (b) the *closure* $\overline{A}$ of A to be the intersection of all closed sets that contain A , equivalently the smallest closed set that contains A ;
- (c) the *boundary* ∂A of A to be $\partial A = \overline{A} \cap \overline{X \setminus A}$.

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

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

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

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

Proposition 2.21. *For any subset A of a topological space X we have:*

- (a) $\partial A \cap A^\circ = \emptyset$;
- (b) $\overline{A} = A^\circ \cup \partial A$;
- (c) $A^\circ = A \setminus \partial A$.

Proof. See [Tutorial Question 2.6](#). $\square$

We say that A is *nowhere dense* in X if $(\overline{A})^\circ = \emptyset$. A simple example of this is $\mathbf{Z}$ as a subset of $\mathbf{R}$, see [Exercise 1.31](#).

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

Proposition 2.22. *If A is a subset of a topological space X , then A is dense in X if and only if every nonempty open subset of X intersects A nontrivially.*

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

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

Example 2.23. Consider $\mathbf{R}$ with its usual topology. Both $\mathbf{Q}$ and $\mathbf{R} \setminus \mathbf{Q}$ are dense in $\mathbf{R}$.

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

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

So we have two disjoint sets, each of which is dense in $\mathbf{R}$. The situation is very different if we ask for the sets to be both dense and open, which we do in [Tutorial Question 3.3](#).

2.4. PROPERTIES OF TOPOLOGIES INDUCED BY METRICS

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

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

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

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

Proof. Put together [Example 2.7](#), [Propositions 2.9](#) and [2.10](#), and [Tutorial Question 3.4](#). $\square$

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

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

In a metric space, the concept of continuous function has equivalent formulations that are more familiar from calculus and analysis. For example, the equivalence to the ε - δ definition is detailed in [Tutorial Question 2.8](#).

Example 2.25. Let (X, d) be a metric space and fix a point $t \in X$. Define $f: X \rightarrow \mathbf{R}_{\geq 0}$ by

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

Then f is a continuous function.

Solution. Here is a proof that pretends to avoid the ε - δ formalism. By [Exercise 1.22](#) it suffices to consider opens $U \subseteq \mathbf{R}_{\geq 0}$ in a set that generates the topology on $\mathbf{R}_{\geq 0} \subseteq \mathbf{R}$; from real analysis, or a special case of [Tutorial Question 3.4](#), we can take $U = (a, b) \subseteq \mathbf{R}_{\geq 0}$ to be an open interval of finite length. Then

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

which is open in X as it is the intersection of two open sets. (Here we also used [Exercise 1.9](#) to deduce that $\mathbf{D}_a(t)$ is a closed set.) $\square$

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

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

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

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

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

Example 2.27. Any metric space (X, d) is Hausdorff.

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

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

2.5. CONNECTEDNESS

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

$$U \cup V = X \quad \text{and} \quad U \cap V = \emptyset.$$

A nonempty¹ topological space X is *connected* if it is not disconnected.

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

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

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

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

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

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

Proof. In [Tutorial Question 2.6](#) we see that for any subset S of X we have

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

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

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

The relevance of clopen subsets to connectedness is given by

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

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

$$U \cup V = X, \quad U \cap V = \emptyset.$$

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

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

$$U \cup V = X, \quad U \cap V = \emptyset. \quad \square$$

In practice, a very useful criterion for connectedness is

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

- (a) X is connected.
- (b) For every discrete space D , if $f: X \rightarrow D$ is a continuous function, then f is constant.
- (c) For the discrete space $\{0, 1\}$, if $f: X \rightarrow \{0, 1\}$ is a continuous function, then f is constant.

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

(b) $\Rightarrow$ (c): Clear.

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

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

$$U \cup V = X, \quad U \cap V = \emptyset.$$

Define $f: X \rightarrow \{0, 1\}$ by

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

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

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

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

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

We conclude by [Proposition 2.31](#). $\square$

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

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

Take any $x \in U$ and define

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

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

We proceed by contradiction. Suppose S is bounded above. Then $s = \sup S$ exists².

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

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

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

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

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

- $\emptyset$;
- for $a \leq b$: $[a, b]$;
- for $a < b$: (a, b) and $(a, b]$ and $[a, b)$;

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

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

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

A different way of describing intervals is sometimes handy:

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

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

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

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

Corollary 2.35 (Intermediate Value Theorem). *Let $f: X \rightarrow \mathbf{R}$ be a continuous function, with X a connected topological space. For any $x, y \in X$ and any $r \in \mathbf{R}$ such that $f(x) < r < f(y)$, there exists $\xi \in X$ such that $f(\xi) = r$.*

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

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

2.6. COMPACTNESS

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

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

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

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

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

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

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

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

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

Proof. See Exercise 1.52. $\square$

Proposition 2.37. *If X is a Hausdorff topological space and $K \subseteq X$ is a compact subset, then K is closed.*

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

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

which by compactness has some finite subcover

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

Consider

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

which is an open neighbourhood of x . We have $U \cap V = \emptyset$, therefore $V \subseteq X \setminus U \subseteq X \setminus K$ is an open neighbourhood of x contained in $X \setminus K$. By [Exercise 1.14](#), $X \setminus K$ is open. $\square$

Proposition 2.38. *If X is a compact topological space and $K \subseteq X$ is a closed subset, then K is compact.*

Proof. Consider an open cover of K :

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

We can turn this into an open cover of X :

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

As X is compact, there is a finite subcover

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

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

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

hence we have found a finite subcover of the original open cover. $\square$

Proposition 2.39. *If $f: X \rightarrow Y$ is a continuous function between topological spaces and X is compact, then $f(X)$ is compact.*

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

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

Then

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

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

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

therefore

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

A map $f: X \longrightarrow Y$ between topological spaces is

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

Proposition 2.40. *Let $f: X \longrightarrow Y$ be a closed map between topological spaces such that $f^{-1}(y) \subseteq X$ is compact for all $y \in Y$. Then f is proper.*

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

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

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

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

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

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

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

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

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

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

which by the compactness of K has a finite subcover

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

Then

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

which is a finite subcover of the original

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

Theorem 2.41. *Let X_1, X_2 be topological spaces.*

- If X_1 is compact then the map $\pi_2: X_1 \times X_2 \longrightarrow X_2$ is closed and proper.*
- If X_1 and X_2 are compact topological spaces, then their product $X_1 \times X_2$ is compact.*

Proof.

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

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

Setting

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

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

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

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

The fact that π_2 is proper now follows from [Proposition 2.40](#), since for any $x_2 \in X_2$ we have $\pi_2^{-1}(x_2) = X_1 \times \{x_2\}$, which is homeomorphic to X_1 by [Exercise 1.24](#), hence compact.

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

2.7. SEQUENCES IN METRIC SPACES

Let (X, d) be a metric space.

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

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

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

Proposition 2.42. *Let (X, d) be a metric space and let (x_n) be a sequence that converges to $x \in X$.*

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

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

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

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

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

Proof.

- (a) As $x \in U$ and U is open, there exists $\varepsilon > 0$ such that $\mathbf{B}_\varepsilon(x) \subseteq U$. But as $(x_n) \rightarrow x$, there exists $N \in \mathbf{N}$ such that $x_n \in \mathbf{B}_\varepsilon(x) \subseteq U$ for all $n \geq N$.
- (b) Let $U \subseteq X$ be an open neighbourhood of x . By part (a), there exists $N \in \mathbf{N}$ such that $x_n \in U$ for all $n \geq N$. In particular, U intersects A nontrivially. By [Proposition 2.20](#), we conclude that $x \in \overline{A}$.

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

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

(c) Follows immediately from (b). $\square$

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

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

By [Exercise 1.53](#), this is an equivalence relation on the set of sequences in (X, d) .

Proposition 2.43. *Let (x_n) and (y_n) be equivalent sequences in a metric space (X, d) and let $x \in X$. Then (x_n) converges to x if and only if (y_n) converges to x .*

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

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

$$d(y_n, x) \leq d(y_n, x_n) + d(x_n, x) < \varepsilon. \quad \square$$

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

Theorem 2.44. *Let $f: X \longrightarrow Y$ be a function between metric spaces and let $x \in X$. Then f is continuous at x if and only if for all sequences $(x_n) \longrightarrow x$, the sequence $(f(x_n)) \longrightarrow f(x)$.*

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

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

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

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

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

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

Note that an isometry must be injective, as well as continuous.

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

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

2.8. CAUCHY SEQUENCES

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

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

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

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

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

A natural question is whether the converse of [Proposition 2.45](#) holds: does every Cauchy sequence converge? In an arbitrary metric space, the answer is no. We say that a metric space X is *complete* if every Cauchy sequence converges to an element of X .

Example 2.46. (I hope) we know from real analysis that $\mathbf{R}$ is a complete metric space.

However, $\mathbf{Q}$ is not complete, as you can see in [Exercise 1.62](#).

Proposition 2.47. *If X is a complete metric space and $S \subseteq X$, then S is complete if and only if S is closed.*

Proof. Suppose S is complete and let $x \in \overline{S}$. Then there exists a sequence (s_n) in S such that $(s_n) \rightarrow x \in X$; by [Proposition 2.45](#) we know that (s_n) is Cauchy, so by the completeness of S we have $x \in S$. Therefore $\overline{S} = S$.

Conversely, suppose S is closed in X . Let (s_n) be a Cauchy sequence in S , then (s_n) is a Cauchy sequence in X , which is complete, so $(s_n) \rightarrow x \in X$. By [Proposition 2.42](#) we have $x \in \overline{S} = S$ since S is closed. $\square$

Proposition 2.48. *If (x_n) and (y_n) are Cauchy sequences in a metric space (X, d) , then $(d(x_n, y_n))$ is a Cauchy sequence in $\mathbf{R}$.*

Proof. Inequality chase, see [Exercise 1.56](#). $\square$

The equivalence relation on sequences preserves the Cauchy property:

Proposition 2.49. *Let (x_n) and (y_n) be equivalent sequences in a metric space (X, d) . Then (x_n) is Cauchy if and only if (y_n) is Cauchy.*

Proof. Another inequality chase, see [Exercise 1.57](#). $\square$

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

Example 2.50. Take $X = Y = \mathbf{R}_{>0}$ with the induced metric from $\mathbf{R}$, and $f: X \rightarrow Y$ given by $f(x) = \frac{1}{x}$. The function f is continuous on X . Take the sequence (x_n) with $x_n = \frac{1}{n}$ for all $n \in \mathbf{N}$. Then (x_n) is Cauchy, but $(f(x_n)) = (n)$ is most certainly not Cauchy.

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

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

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

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

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

Proposition 2.52. *Any uniformly continuous function maps Cauchy sequences to Cauchy sequences.*

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

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

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

Proposition 2.53. *Let $f: X \rightarrow Y$ 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 $f(\mathbf{B}_{\delta(x)}(x)) \subseteq \mathbf{B}_{\varepsilon/2}(f(x))$. We get an open cover of X :

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

which therefore has a finite subcover

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

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

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

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

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

2.9. COMPLETIONS

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

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

$$\widehat{x} = \lim_{n \rightarrow \infty} \iota(x_n) \quad \text{and} \quad \widehat{y} = \lim_{n \rightarrow \infty} \iota(y_n).$$

Then

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

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

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

we have TODO: make it clear that we are using the continuity of $\widehat{d}$, and have a reference for this

$$\lim_{n \rightarrow \infty} d(x_n, y_n) = \lim_{n \rightarrow \infty} \widehat{d}(\iota(x_n), \iota(y_n)) = \widehat{d}\left(\lim_{n \rightarrow \infty} (\iota(x_n), \iota(y_n))\right) = \widehat{d}(\widehat{x}, \widehat{y}). \quad \square$$

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

We will see later ([Corollary 2.57](#)) that any two completions of (X, d) are isometric.

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

$\widehat{X}$ as a metric space: [Proposition 2.54](#) hints that we should define $\widehat{d}: \widehat{X} \times \widehat{X} \rightarrow \mathbf{R}_{\geq 0}$ by:

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

The limit exists as the sequence $(d(x_n, y_n))$ is Cauchy in $\mathbf{R}$ ([Proposition 2.48](#)) and $\mathbf{R}$ is complete; moreover $\widehat{d}$ is well-defined, see [Exercise 1.65](#). It is easy to see that $\widehat{d}$ is a metric on $\widehat{X}$.

The isometry ι : For any $x \in X$, let $\iota(x) = [(x)]$ be the equivalence class of the constant sequence $(x, x, \dots)$. We have for all $x, y \in X$:

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

so ι is an isometry.

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

$$\widehat{d}(\iota(x), [(x_n)]) = \lim_{n \rightarrow \infty} d(x, x_n) \leq \frac{\varepsilon}{2} < \varepsilon.$$

The metric space $\widehat{X}$ is complete: Suppose (a_n) is a Cauchy sequence in $\widehat{X}$. As $\iota(X)$ is dense in $\widehat{X}$, for each $n \in \mathbf{N}$ there exists $x_n \in X$ such that $\widehat{d}(\iota(x_n), a_n) < \frac{1}{n}$. We get a sequence $(\iota(x_n)) \sim (a_n)$. As (a_n) is Cauchy in $\widehat{X}$, by [Proposition 2.49](#) so is the sequence $(\iota(x_n))$ in $\iota(X) \subseteq \widehat{X}$, and hence so is the sequence (x_n) in X as $\iota(X)$ is isometric to X . So we have an element $\widehat{x} := [(x_n)] \in \widehat{X}$.

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

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

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

For any $n \in \mathbf{N}$, we have by the triangle inequality

$$d(a_n(m), x_m) \leq d(a_n(m), x_n) + d(x_n, x_m),$$

so taking limits:

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

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

$$\widehat{d}(a_n, \widehat{x}) \leq \widehat{d}(a_n, \iota(x_n)) + \lim_{m \rightarrow \infty} d(x_n, x_m) < \frac{1}{n} + \frac{\varepsilon}{2} < \varepsilon. \quad \square$$

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

$$\widehat{g}(\iota_X(x)) = \iota_Y(g(x)) \quad \text{for all } x \in X,$$

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

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

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

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

Proposition 2.56. *Let Z be a metric space and W a complete metric space. Let $D \subseteq Z$ be a dense subset and $f: D \rightarrow W$ a uniformly continuous function.*

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

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

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

Proof.

$$\begin{array}{ccc} D & \subseteq & Z \\ & \searrow f & \downarrow \widehat{f} \\ & & W \end{array}$$

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

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

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

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

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

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

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

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

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

The **uniqueness** of $\widehat{f}$ follows from [Exercise 1.58](#), which says that there is at most one continuous extension.

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

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

This has the following consequence:

Corollary 2.57. *Let X be a metric space.*

- (a) *Let Y be a metric space and fix completions $(\widehat{X}, \iota_X)$ of X and $(\widehat{Y}, \iota_Y)$ of Y . Any uniformly continuous function (resp. isometry) $g: X \rightarrow Y$ has a unique uniformly continuous (resp. isometry) extension $\widehat{g}: \widehat{X} \rightarrow \widehat{Y}$.*
- (b) *Any two completions of X are isometric.*

Proof.

- (a) Let $D = \iota_X(X) \subseteq \widehat{X}$, and apply [Proposition 2.56](#) to the function $f := \iota_Y \circ g \circ \iota_X^{-1}: D \rightarrow \widehat{Y}$:

$$\begin{array}{ccccc}
 X & \xrightarrow[\simeq]{\iota_X} & \iota_X(X) & \subseteq & \widehat{X} \\
 \downarrow g & & & \searrow f & \downarrow \widehat{g} := \widehat{f} \\
 Y & \xrightarrow{\iota_Y} & & & \widehat{Y}
 \end{array}$$

Then $\widehat{g}$ is the function $\widehat{f}$ given by [Proposition 2.56](#).

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

$$(2.1) \quad \widehat{g}(\widehat{x}) = \lim_{n \rightarrow \infty} \iota_Y(g(x_n)).$$

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

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

$$\begin{array}{ccccc}
 & & \iota_1(X) & \subseteq & \widehat{X}_1 \\
 & \nearrow \iota_1 & \downarrow g & & \downarrow \widehat{g} \\
 X & & & & \\
 & \searrow \iota_2 & \downarrow & & \\
 & & \iota_2(X) & \subseteq & \widehat{X}_2
 \end{array}$$

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

$$\widehat{g}(\widehat{x}') = \widehat{g}\left(\lim_{n \rightarrow \infty} \widehat{x}_n\right) = \lim_{n \rightarrow \infty} \widehat{g}(\widehat{x}_n) = \lim_{n \rightarrow \infty} g(\iota_1(x_n)) = \lim_{n \rightarrow \infty} \iota_2(x_n) = \widehat{x}. \quad \square$$

2.10. BANACH FIXED POINT THEOREM

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

$$d_Y(f(x_1), f(x_2)) \leq C d_X(x_1, x_2) \quad \text{for all } x_1, x_2 \in X.$$

It is easy to see ([Tutorial Question 6.1](#)) that contractions are uniformly continuous.

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

Proposition 2.58. *Let $f: X \rightarrow X$ be a contraction from a metric space to itself. Then f has at most one fixed point.*

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

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

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

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

leading to a contradiction. $\square$

We get a very useful result for complete metric spaces:

Theorem 2.59 (Banach Fixed Point Theorem). *Let (X, d) be a nonempty complete metric space. Let $f: X \rightarrow X$ 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 $x_0 \in X$, the sequence (x_n) defined recursively by $x_{n+1} = f(x_n)$ for $n \in \mathbf{N}$ converges to the fixed point x .*

Proof. The uniqueness statement follows from [Proposition 2.58](#).

The proof of existence uses the hint in the last statement. Let $x_0 \in X$ and consider the sequence $(x_n) = (f^{\circ n}(x_0))$. For any $m \geq 1$ we have

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

Applying this repeatedly with decreasing m , we get

$$d(x_{m+1}, x_m) \leq C^m d(x_1, x_0).$$

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

$$\begin{aligned} d(x_n, x_m) &\leq (C^{n-1} + C^{n-2} + \cdots + C^m) d(x_1, x_0) \\ &\leq C^m \frac{1 - C^{n-m}}{1 - C} d(x_1, x_0) \\ &\leq C^m \frac{d(x_1, x_0)}{1 - C}. \end{aligned}$$

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

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

So x is indeed a fixed point of f . □

Recall the following result from real analysis:

Theorem 2.60 (Mean Value Theorem). *Let $f: [a, b] \rightarrow \mathbf{R}$ be continuous. If f is differentiable on (a, b) , then there exists $\xi \in (a, b)$ such that*

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

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

Example 2.61. Verify that the function $f: [1, 2] \rightarrow \mathbf{R}$ defined by

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

has a unique fixed point, and find this point.

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

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

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

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

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

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

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

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:

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

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

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 $x_0 = 1$ and apply f iteratively, $x_{n+1} = f(x_n)$. $\square$

2.11. (TOTAL) BOUNDEDNESS

Let (X, d) be a metric space.

The *diameter* of a nonempty³ subset $S \subseteq X$ is by definition

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

If this is a (finite) real number we say that S is *bounded*. This is equivalent to saying that S is contained in some closed ball with finite radius (see [Exercise 1.70](#)). Otherwise we say that S is *unbounded*.

Example 2.62. Let $S \subseteq \mathbf{R}$ be a bounded set. Show that for any $\varepsilon > 0$, there exist $N \in \mathbf{N}$ and open balls $B_1, \dots, B_N$, all of radius ε , such that

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

Solution. As S is bounded, it is contained in some closed ball, which in $\mathbf{R}$ is some interval $[x, y]$. So it suffices to prove that the conclusion holds for the interval $[x, y]$, which is

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

straightforward: given $\varepsilon > 0$, let $N \in \mathbf{N}$ be such that $N \geq \frac{y-x}{\varepsilon}$, then

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

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

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

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

As you can see in [Tutorial Question 7.1](#), any totally bounded set is bounded; [Example 2.62](#) says that the converse is true if $X = \mathbf{R}$. See [Exercise 1.75](#) for the fact that the product of two totally bounded sets is totally bounded, and [Exercise 1.76](#) for the consequence that in $\mathbf{R}^m$, every bounded set is totally bounded.

Proposition 2.63. *Let $f: X \rightarrow \mathbf{R}$ be a continuous function, where X is a compact metric space. Then the image $f(X)$ is bounded, and the bounds are attained: there exist $x_{\min}, x_{\max} \in X$ such that*

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

Proof. By [Proposition 2.39](#), $f(X)$ is a compact subset of $\mathbf{R}$. Therefore $f(X)$ is totally bounded, hence bounded. So $f(X)$ has both infimum and supremum, which are boundary points. But $f(X)$ is also closed by [Proposition 2.37](#), therefore it contains its boundary points and hence the infimum and supremum. $\square$

Proposition 2.64. *If $f: X \rightarrow Y$ is a uniformly continuous function between metric spaces and $S \subseteq X$ is totally bounded, then $f(S) \subseteq Y$ is totally bounded.*

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

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

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

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

so applying f on both sides we get

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

Proposition 2.65. *A subset $S \subseteq X$ of a metric space is totally bounded if and only if every sequence in S has a Cauchy subsequence.*

Proof. In one direction, suppose S is totally bounded and let (s_n) be a sequence in S .

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

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

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

$(s_n):$	$s_1,$	$s_2,$	$s_3,$	$s_4,$	$\dots$	$\dots$	$\in S$
$(s_n^{(1)}):$	$s_1^{(1)},$	$s_2^{(1)},$	$s_3^{(1)},$	$s_4^{(1)},$	$\dots$	$\dots$	$\in \mathbf{B}_1(x_1)$
$(s_n^{(2)}):$	$s_1^{(2)},$	$s_2^{(2)},$	$s_3^{(2)},$	$s_4^{(2)},$	$\dots$	$\dots$	$\in \mathbf{B}_{1/2}(x_2)$
$(s_n^{(3)}):$	$s_1^{(3)},$	$s_2^{(3)},$	$s_3^{(3)},$	$s_4^{(3)},$	$\dots$	$\dots$	$\in \mathbf{B}_{1/3}(x_3)$
$\vdots$							$\vdots$
$(s_n^{(j)}):$	$s_1^{(j)},$	$s_2^{(j)},$	$s_3^{(j)},$	$s_4^{(j)},$	$\dots$	$s_j^{(j)}, \dots$	$\in \mathbf{B}_{1/j}(x_j)$
$\vdots$							$\vdots$

From this list we extract the diagonal⁴, giving rise to a subsequence $(s_n^{(n)})$ of (s_n) . I claim that $(s_n^{(n)})$ is a Cauchy sequence.

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

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

hence

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

In the other direction, suppose that every sequence in S has a Cauchy subsequence. We proceed by contradiction: assume that S is not totally bounded. Then there exists $\varepsilon > 0$ so that no finite collection of open balls of radius ε can cover S .

Choose an arbitrary $s_1 \in S$. By assumption, we cannot have $S \subseteq \mathbf{B}_\varepsilon(s_1)$. So there exists $s_2 \in S \setminus \mathbf{B}_\varepsilon(s_1)$. Again by assumption, we cannot have $S \subseteq \mathbf{B}_\varepsilon(s_1) \cup \mathbf{B}_\varepsilon(s_2)$. So there exists $s_3 \in S \setminus (\mathbf{B}_\varepsilon(s_1) \cup \mathbf{B}_\varepsilon(s_2))$.

Continuing in this manner, we obtain a sequence (s_n) in S such that $d(s_n, s_m) \geq \varepsilon$ for all $n, m \in \mathbf{N}$, so that (s_n) has no Cauchy subsequence, contradiction. $\square$

2.12. COMPACTNESS IN METRIC SPACES

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

Theorem 2.66. *Let K be a subset of a metric space X . The following are equivalent:*

- (a) K is compact.
- (b) K is complete and totally bounded.
- (c) K is **sequentially compact**, that is every sequence in K has a subsequence that converges to an element of K .

Before proving this theorem, we need the following notion: given a subset K of X and an open cover

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

⁴It was pointed out to me that the argument should work for other choices than the diagonal in the table. The restrictions are: (a) we need to get a subsequence of the original sequence (to ensure this, we should be moving to the right at each step); (b) we need to end up with a Cauchy sequence (for which we should be moving down at each step).

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

It is the case that any open cover of a sequentially compact subset $K \subseteq X$ has a Lebesgue number, see [Exercise 1.77](#).

Proof of Theorem 2.66. (a) $\Rightarrow$ (b): Suppose K is compact. We have already seen that K is totally bounded. Let $\iota: K \rightarrow \widehat{K}$ be a completion of K . Then $\iota(K)$ is a compact subset of $\widehat{K}$, hence closed by [Proposition 2.37](#). But $\iota(K)$ is also dense in $\widehat{K}$, so $\iota(K) = \widehat{K}$ and K is complete.

(b) $\Rightarrow$ (c): Suppose K is complete and totally bounded and let (x_n) be a sequence in K . Since K is totally bounded, (x_n) has a Cauchy subsequence by [Proposition 2.65](#), which converges in K , since K is complete.

(c) $\Rightarrow$ (a): Suppose K is sequentially compact and consider an open cover

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

By [Exercise 1.77](#) this cover has a Lebesgue number $\delta > 0$. By [Proposition 2.65](#), K is totally bounded, so it has a finite cover by open balls of radius $\delta/2$:

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

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

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

2.13. SPACES OF BOUNDED CONTINUOUS FUNCTIONS

Let X be a set and Y a metric space.

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

$$d_Y(y, f(x)) \leq M \quad \text{for all } x \in X.$$

Equivalently, the direct image $f(X)$ is a bounded subset of Y , see [Exercise 1.74](#).

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

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

Proposition 2.67. *The function d_∞ is a metric on $B(X, Y)$, called the **uniform metric**.*

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

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

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

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

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

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

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

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

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

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

By the minimality of the supremum we get

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

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

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

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

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

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

In this situation, we call the function f the *pointwise limit* of the sequence (f_n) .

Example 2.68. The pointwise limit of a sequence of bounded functions need not be bounded.

For instance, take $f_n: \mathbf{R}_{\geq 0} \rightarrow \mathbf{R}$ given by

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

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

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

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

In this situation, we call the function f the *uniform limit* of the sequence (f_n) .

Proposition 2.69. Let X be a set and Y a metric space.

- (a) The uniform limit f of a sequence (f_n) of bounded functions $X \rightarrow Y$ is bounded.
- (b) A sequence (f_n) in $B(X, Y)$ converges uniformly to $f \in B(X, Y)$ if and only if $(f_n) \rightarrow f$ with respect to the uniform metric d_∞ on $B(X, Y)$.

Proof.

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

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

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

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

which shows that f is bounded.

(b) See [Tutorial Question 7.6](#). □

Proposition 2.70. *Given a set X and a metric space Y , if Y is complete then the metric space $B(X, Y)$ (with the uniform metric d_∞) is complete.*

Proof. Let (f_n) be a Cauchy sequence in $B(X, Y)$. For any $x \in X$, consider the evaluation function $\text{ev}_x: B(X, Y) \rightarrow Y$ given by $\text{ev}_x(h) = h(x)$. It is easy to see that ev_x is uniformly continuous (see [Exercise 1.82](#)). Therefore the sequence $(f_n(x))$ is Cauchy in Y , which is complete, so we define $f(x) \in Y$ to be the limit of this sequence.

We now prove that (f_n) converges to f uniformly. Given $\varepsilon > 0$, take $N \in \mathbf{N}$ such that

$$d_Y(f_n(t), f_m(t)) < \frac{\varepsilon}{2} \quad \text{for all } m, n \geq N, t \in X.$$

Let $n \geq N$. Given $x \in X$, let $m(x) \geq N$ be such that $d_Y(f_{m(x)}(x), f(x)) < \varepsilon/2$, then

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

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

As we have shown that f is the uniform limit of the sequence of bounded functions (f_n) , f is bounded by [Proposition 2.69](#). □

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

Proposition 2.71. *Given metric spaces X and Y , $C_0(X, Y)$ is a closed subset of $B(X, Y)$ with the uniform metric d_∞ . In other words, the uniform limit of a sequence of bounded continuous functions is a bounded continuous function.*

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

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

Let $\delta > 0$ be such that

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

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

$$d_Y(f(x_0), f(x)) \leq d_Y(f(x_0), f_N(x_0)) + d_Y(f_N(x_0), f_N(x)) + d_Y(f_N(x), f(x)) < \varepsilon. \quad \square$$

The **pointwise** limit of a sequence of bounded continuous functions need not be continuous, see [Tutorial Question 7.7](#).

2.14. FUNCTION SPACES: APPROXIMATION BY POLYNOMIALS

There are **a lot** of bounded continuous functions $X \rightarrow \mathbf{R}$ even for relatively simple X , e.g. closed intervals in $\mathbf{R}$. But:

Theorem 2.72 (Weierstrass Approximation Theorem). *Given $a < b$, let $\mathcal{A}$ be the subset of $C_0([a, b], \mathbf{R})$ consisting of polynomial functions. Then $\mathcal{A}$ is dense in $C_0([a, b], \mathbf{R})$.*

We will obtain this as a corollary of a more general result. We need some preliminaries.

Lemma 2.73. *There is a sequence (p_n) in $x\mathbf{R}[x]$ such that $(p_n) \rightarrow |x|$ uniformly on $[-1, 1]$.*

Sketch of proof. Let $p_1(x) = 0$ and

$$p_{n+1}(x) = p_n(x) - \frac{p_n(x)^2 - x^2}{2} = p_n(x) - \frac{(p_n(x) - |x|)(p_n(x) + |x|)}{2} \quad \text{for } n \geq 1.$$

One can use induction to prove that, for all $x \in [-1, 1]$ and all $n \geq 1$:

- (a) $0 \leq p_n(x) \leq |x|$;
- (b) $p_n(x) \leq p_{n+1}(x)$;
- (c) $|x| - p_{n+1}(x) \leq |x| \left(1 - \frac{|x|}{2}\right)^n$.

(See [Exercise 1.85](#).)

A little calculus ([Exercise 1.86](#)) tells us that for any $n \geq 1$

$$|x| \left(1 - \frac{|x|}{2}\right)^n < \frac{2}{n+1} \quad \text{for all } x \in [-1, 1],$$

which then implies that $(p_n) \rightarrow |x|$ uniformly on $[-1, 1]$. □

Corollary 2.74. *For any $a > 0$, there is a sequence (q_n) in $x\mathbf{R}[x]$ such that $(q_n) \rightarrow |x|$ uniformly on $[-a, a]$.*

Proof. See [Exercise 1.87](#). □

For any metric space X , the set $C_0(X, \mathbf{R})$ is an $\mathbf{R}$ -algebra. (The multiplication of functions is done pointwise, just like the addition.)

Proposition 2.75. *Let X be a metric space and let $\mathcal{C}$ be a closed subalgebra of $C_0(X, \mathbf{R})$. Then*

- (a) if $g \in \mathcal{C}$ then $|g| \in \mathcal{C}$;
- (b) if $g_1, g_2 \in \mathcal{C}$ then $\max\{g_1, g_2\}, \min\{g_1, g_2\} \in \mathcal{C}$.

Proof.

- (a) Let $a > 0$ be an upper bound for $|g|$. For any $n \in \mathbf{N}$, we have by [Corollary 2.74](#) a polynomial $q_n(y) \in y\mathbf{R}[y]$ such that

$$|q_n(y) - |y|| < \frac{1}{n} \quad \text{for all } y \in [-a, a].$$

Let $h_n = q_n(g)$, then $h_n \in \mathcal{C}$ since the latter is an algebra; therefore applying the above inequality with $y = g(x)$ we have

$$|h_n(x) - |g(x)|| = |q_n(g(x)) - |g(x)|| < \frac{1}{n} \quad \text{for all } x \in X,$$

in other words

$$\|h_n - |g|\|_{L^\infty} < \frac{1}{n},$$

which shows that $(h_n) \rightarrow |g|$ in $\mathcal{C}$, which is closed, so $|g| \in \mathcal{C}$.

(b) The claim follows directly from the relations

$$\begin{aligned} 2 \max\{g_1, g_2\} &= g_1 + g_2 + |g_1 - g_2| \\ 2 \min\{g_1, g_2\} &= g_1 + g_2 - |g_1 - g_2|. \end{aligned}$$

□

Let $\mathcal{A}$ be a subalgebra of $C_0(X, \mathbf{R})$.

- We say that $\mathcal{A}$ *separates points of X* if for any $x_1, x_2 \in X$ with $x_1 \neq x_2$, there exists $f \in \mathcal{A}$ such that $f(x_1) \neq f(x_2)$.
- We say that $\mathcal{A}$ *is non-vanishing on X* if for any $x \in X$ there exists $f \in \mathcal{A}$ such that $f(x) \neq 0$.

Lemma 2.76. *Let X be a metric space with at least two points. A subalgebra $\mathcal{A}$ of $C_0(X, \mathbf{R})$ separates points of X and is non-vanishing on X if and only if, for every $(x_1, y_1), (x_2, y_2) \in X \times \mathbf{R}$ with $x_1 \neq x_2$, there exists $h \in \mathcal{A}$ such that*

$$h(x_1) = y_1 \quad \text{and} \quad h(x_2) = y_2.$$

Proof. See [Exercise 1.88](#).

□

Theorem 2.77 (Stone–Weierstrass). *Let X be a nonempty compact metric space and let $\mathcal{A}$ be a subalgebra of $C_0(X, \mathbf{R})$. If $\mathcal{A}$ separates points of X and is non-vanishing on X , then it is dense in $C_0(X, \mathbf{R})$ (with respect to the uniform norm).*

Proof. The corner case where X is a singleton is easily dispatched: then $C_0(X, \mathbf{R}) = \mathbf{R}$ and $\mathcal{A} = \mathbf{R}$ (since nonzero).

So we may assume that X has at least two distinct elements (and we can therefore apply [Lemma 2.76](#)).

Let $\mathcal{C}$ denote the closure of $\mathcal{A}$ in $C_0(X, \mathbf{R})$. We will show that $\mathcal{C}$ is dense in $C_0(X, \mathbf{R})$: let $f \in C_0(X, \mathbf{R})$ and let $\varepsilon > 0$.

Fix $x' \in X$.

For every $x \in X$:

- If $x \neq x'$ then by [Lemma 2.76](#), there exists $h_x \in \mathcal{A}$ such that

$$h_x(x) = f(x) \quad \text{and} \quad h_x(x') = f(x').$$

- If $x = x'$, let $t \in X$, $t \neq x'$ and choose $h_{x'} \in \mathcal{A}$ such that

$$h_{x'}(t) = f(t) \quad \text{and} \quad h_{x'}(x') = f(x').$$

Let U_x be an open neighbourhood of x such that

$$h_x(x'') > f(x'') - \varepsilon \quad \text{for all } x'' \in U_x.$$

The sets $\{U_x : x \in X\}$ form an open cover of the compact space X , so there exist $x_1, \dots, x_m$ such that

$$X \subseteq U_{x_1} \cup \dots \cup U_{x_m}.$$

Let $g_{x'} = \max\{h_{x_1}, \dots, h_{x_m}\}$, an element of $\mathcal{C}$ by [Proposition 2.75](#).

Then $g_{x'}(x') = f(x')$ and

$$g_{x'}(x'') > f(x'') - \varepsilon \quad \text{for all } x'' \in X.$$

We have such a function $g_{x'} \in \mathcal{C}$ for each $x' \in X$.

Let $V_{x'}$ be an open neighbourhood of x' such that

$$g_{x'}(x'') < f(x'') + \varepsilon \quad \text{for all } x'' \in V_{x'}.$$

The sets $\{V_{x'} : x' \in X\}$ form an open cover of the compact space X , so there exist $x'_1, \dots, x'_n$ such that

$$X \subseteq V_{x'_1} \cup \dots \cup V_{x'_n}.$$

We let $g = \min\{g_{x'_1}, \dots, g_{x'_n}\}$ so that $g \in \mathcal{C}$ and

$$f(x'') - \varepsilon < g(x'') < f(x'') + \varepsilon \quad \text{for all } x'' \in X,$$

so we conclude that

$$\|f - g\|_{L^\infty} < \varepsilon. \quad \square$$

We now specialise to

Proof of Theorem 2.72. We take $\mathcal{A}$ to be the subset of $C_0([a, b], \mathbf{R})$ consisting of polynomial functions. It is clear that $\mathcal{A}$ is an algebra. Also:

- $\mathcal{A}$ is non-vanishing on $[a, b]$ since the constant polynomial 1 is nowhere zero.
- $\mathcal{A}$ separates points of $[a, b]$: if $x_1 \neq x_2 \in [a, b]$, the polynomial $f(x) = x$ satisfies $f(x_1) = x_1 \neq x_2 = f(x_2)$.

It follows from the Stone–Weierstrass Theorem that $\mathcal{A}$ is dense. $\square$

2.15. INTERSECTION OF DENSE OPEN SETS

Recall that the intersection of two open dense subsets is open and dense ([Tutorial Question 3.3](#)), 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 2.78. Note that for every $x \in \mathbf{R}$, $U_x := \mathbf{R} \setminus \{x\}$ is dense and open in $\mathbf{R}$.

Therefore the intersection of an uncountable collection of dense open subsets need not be dense: this is clear, as $\bigcap_{x \in \mathbf{R}} U_x = \emptyset$, which is certainly not dense in $\mathbf{R}$.

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

Theorem 2.79 (Baire Category Theorem). *Let (X, d) be a complete metric space. Let $\{U_n : n \in \mathbf{N}\}$ be a (countable) collection of dense open subsets of X . Then*

$$D := \bigcap_{n \in \mathbf{N}} U_n \quad \text{is dense in } X.$$

Proof. We use the criterion in [Tutorial Question 3.2](#). Let W be a nonempty open set. We want to show that $D \cap W \neq \emptyset$.

Since U_1 is dense, $W \cap U_1$ is a nonempty open set. Let $x_1 \in W \cap U_1$ and let $0 < 2r_1 < 2$ be such that

$$\mathbf{D}_{r_1}(x_1) \subseteq \mathbf{B}_{2r_1}(x_1) \subseteq W \cap U_1.$$

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

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

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

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

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

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

There are variants of this result that are stated differently. Here is a form that we will use in the context of Banach spaces in the next chapter:

Corollary 2.80. *Let (X, d) be a nonempty complete metric space. If*

$$X = \bigcup_{n \in \mathbf{N}} C_n \quad \text{with each } C_n \text{ a closed subset of } X,$$

then there exists $n \in \mathbf{N}$ such that $C_n^\circ \neq \emptyset$.

Proof. See [Tutorial Question 8.5](#). $\square$

3. NORMED AND HILBERT SPACES

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

NOTATION

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

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

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

3.1. NORMS

Let V be a vector space over $\mathbf{F}$. A *norm* on V is a function

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

such that

- (a) $\|v + w\| \leq \|v\| + \|w\|$ for all $v, w \in V$;
- (b) $\|\alpha v\| = |\alpha| \|v\|$ for all $v \in V$, $\alpha \in \mathbf{F}$;
- (c) $\|v\| = 0$ if and only if $v = 0$.

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

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

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

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

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

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

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

So every normed space is a metric space.

Proof.

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

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

$$(c) \quad d(v, w) = 0 \text{ iff } \|v - w\| = 0 \text{ iff } v - w = 0 \text{ iff } v = w;$$

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

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

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

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

Equivalent norms on V give rise to the same topology on V , see [Exercise 2.2](#).

There are various ways of putting a norm on the product of two normed spaces, but many of these ways are topologically equivalent:

Proposition 3.2. *Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be normed vector spaces. The following functions give norms on the vector space $V \times W$:*

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

Moreover, any norm $\|\cdot\|$ on $V \times W$ such that

$$\|(v, w)\|_\infty \leq \|(v, w)\| \leq \|(v, w)\|_1 \quad \text{for all } (v, w) \in V \times W$$

gives rise to the product topology on $V \times W$.

Proof. We prove that $\|\cdot\|_1$ is a norm and leave $\|\cdot\|_\infty$ as an exercise. The other claim follows immediately from the definition of the metric given by a norm, and by [Exercise 1.38](#).

We have

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

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

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

Finally

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

It is easy to see that the norm $V \longrightarrow \mathbf{R}_{\geq 0}$, $v \longmapsto \|v\|$, is a uniformly continuous function with respect to the metric defined by the norm on V , and the Euclidean metric on $\mathbf{R}_{\geq 0}$, see [Exercise 2.1](#).

Proposition 3.3. *Any normed space $(V, \|\cdot\|)$ is a **topological vector space**, that is a vector space such that*

- (a) *vector addition $a: V \times V \longrightarrow V$, $a(v, w) = v + w$, is a continuous function;*
- (b) *scalar multiplication $s: \mathbf{F} \times V \longrightarrow V$, $s(\alpha, v) = \alpha v$, is a continuous function.*

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

Proof.

- (a) Let $(v_0, w_0) \in V \times V$ be arbitrary. We want to show that a is continuous at (v_0, w_0) . Let $x = a(v_0, w_0) = v_0 + w_0$ and let $\varepsilon > 0$. We will find an open rectangle $R = \mathbf{B}_{\delta_1}(v_0) \times \mathbf{B}_{\delta_2}(w_0) \subseteq V \times V$ such that $a(R) \subseteq \mathbf{B}_\varepsilon(x)$.

Let $\delta_1 = \delta_2 = \varepsilon/2$. If $(v, w) \in R$ then $\|v - v_0\| < \varepsilon/2$ and $\|w - w_0\| < \varepsilon/2$, so that

$$\|a(v, w) - a(v_0, w_0)\| = \|(v + w) - (v_0 + w_0)\| = \|(v - v_0) + (w - w_0)\| \leq \|v - v_0\| + \|w - w_0\| < \varepsilon.$$

- (b) Let $(\alpha_0, v_0) \in \mathbf{F} \times V$ be arbitrary. We want to show that s is continuous at (α_0, v_0) . Let $x = s(\alpha_0, v_0) = \alpha_0 v_0$ and let $\varepsilon > 0$. We will find an open rectangle $R = \mathbf{B}_{\delta_1}(\alpha_0) \times \mathbf{B}_{\delta_2}(v_0) \subseteq \mathbf{F} \times V$ such that $s(R) \subseteq \mathbf{B}_\varepsilon(x)$.

- If $\|v_0\| = 0$, let $\delta_1 = 1$ and $\delta_2 = \varepsilon/(1 + |\alpha_0|)$.
- If $\|v_0\| > 0$, let $\delta_1 = \varepsilon/(2\|v_0\|)$ and $\delta_2 = \varepsilon/(2(\delta_1 + |\alpha_0|))$.

If $(\alpha, v) \in R$ then $|\alpha - \alpha_0| < \delta_1$ and $\|v - v_0\| < \delta_2$. In particular, $\|v\| < \|v_0\| + \delta_2$. We have

$$\begin{aligned} \|s(\alpha, v) - s(\alpha_0, v_0)\| &= \|\alpha v - \alpha_0 v_0\| \\ &= \|(\alpha - \alpha_0)v - \alpha_0(v_0 - v)\| \\ &\leq |\alpha - \alpha_0| \|v\| + |\alpha_0| \|v_0 - v\| \\ &< \delta_1 (\|v_0\| + \delta_2) + |\alpha_0| \delta_2 \\ &= \delta_1 \|v_0\| + \delta_2 (\delta_1 + |\alpha_0|) < \varepsilon. \end{aligned}$$

□

Corollary 3.4. *If $(V, \|\cdot\|)$ is a normed space, (v_n) , (w_n) are sequences converging in V , and $\alpha \in \mathbf{F}$ is a scalar, then*

- (a) $\lim_{n \rightarrow \infty} (v_n + w_n) = \lim_{n \rightarrow \infty} v_n + \lim_{n \rightarrow \infty} w_n$;
- (b) $\lim_{n \rightarrow \infty} (\alpha v_n) = \alpha \lim_{n \rightarrow \infty} v_n$;
- (c) $\lim_{n \rightarrow \infty} \|v_n\| = \left\| \lim_{n \rightarrow \infty} v_n \right\|$.

Corollary 3.5. *Let $(V, \|\cdot\|)$ be a normed space and let $W \subseteq V$ be a subspace. Then its closure $\overline{W}$ is also a subspace.*

Proof. Suppose $u, v \in \overline{W}$, then there exist sequences (u_n) and (v_n) in W such that $(u_n) \longrightarrow u$ and $(v_n) \longrightarrow v$. Therefore $u_n + v_n \in W$ for all n , and by [Corollary 3.4](#) we have

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

Similarly for scalar multiplication.

□

Proposition 3.6. *Let $\{v_1, \dots, v_n\}$ be a linearly independent subset of a normed space $(V, \|\cdot\|)$. Then there exists $m > 0$ such that*

$$\|\alpha_1 v_1 + \dots + \alpha_n v_n\| \geq m(|\alpha_1| + \dots + |\alpha_n|) \quad \text{for all } \alpha_1, \dots, \alpha_n \in \mathbf{F}.$$

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

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

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

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

To do this, consider the set

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

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

Now look at the function $F: K \rightarrow \mathbf{R}$ given by

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

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

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

contradicting the linear independence of the vectors.

Hence $m > 0$ and we are done. □

We are now in a good position to prove that

Theorem 3.7. *Any two norms on a finite-dimensional vector space V are equivalent.*

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

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

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

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

$$\|\alpha_1 v_1 + \dots + \alpha_n v_n\| \leq |\alpha_1| \|v_1\| + \dots + |\alpha_n| \|v_n\| \leq M(|\alpha_1| + \dots + |\alpha_n|).$$

From [Proposition 3.6](#) we also have $m > 0$ such that

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

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

3.2. INNER PRODUCTS

Let V be a vector space over $\mathbf{F}$. Recall from linear algebra (see [Section A.4](#) for a summary) that an *inner product* on V is a function

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

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

Proposition 3.8. *If $(V, \langle \cdot, \cdot \rangle)$ is an inner product space, then the function $\| \cdot \|: V \longrightarrow \mathbf{R}_{\geq 0}$ defined by*

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

is a norm on V .

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

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

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

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

Therefore

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

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

Obviously then:

Corollary 3.9. *Any inner product space is a normed space, and a metric space.*

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

Proposition 3.10 (Parallelogram Law). *If $(V, \langle \cdot, \cdot \rangle)$ is an inner product space, then its norm satisfies*

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

Proof. Recall from the proof of [Proposition 3.8](#) that

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

Then

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

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

Given a norm, you can use the Polarisation Identity (see [Exercise 2.5](#)) to see if it comes from an inner product. There is also a converse to the Parallelogram Law, see [Exercise 2.6](#) for the ridiculously fun proof.

3.3. BANACH AND HILBERT SPACES

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

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

The first examples appear in a familiar context:

Proposition 3.11. *Any finite-dimensional normed space $(V, \|\cdot\|)$ is Banach.*

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

By Proposition 3.6 we know that without loss of generality we can take the norm to be given by

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

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

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

The Cauchy-ness means that for any $\varepsilon > 0$ there exists $M \in \mathbf{N}$ such that for all $m, k \geq M$ 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 $\mathbf{F}$. As $\mathbf{F}$ is complete, $(\alpha_j^{(m)}) \rightarrow \beta_j \in \mathbf{F}$.

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

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

Corollary 3.12. *Any finite-dimensional inner product space (in particular, $\mathbf{F}^n$ for any $n \in \mathbf{N}$) is a Hilbert space.*

For an infinite-dimensional Banach space example, take X to be an infinite set and consider the vector space of bounded functions on X :

$$B(X, \mathbf{F}) = \{f : X \rightarrow \mathbf{F} : \text{there exists } c \text{ such that } |f(x)| \leq c \text{ for all } x \in X\}.$$

Then we have

Proposition 3.13. *The set $B(X, \mathbf{F})$ is a Banach space with respect to the uniform norm given by*

$$\|f\|_{L^\infty} = \sup_{x \in X} |f(x)|.$$

If X is a metric space, then the subset $C_0(X, \mathbf{F})$ of bounded continuous functions $X \rightarrow \mathbf{F}$ is a Banach subspace of $B(X, \mathbf{F})$.

Proof. That $B(X, \mathbf{F})$ is a vector subspace of the $\mathbf{F}$ -vector space of all functions $X \rightarrow \mathbf{F}$ is straightforward. It is similarly clear that $\|\cdot\|_{L^\infty}$ gives a norm on $B(X, \mathbf{F})$, and that this norm is associated to the uniform distance d_∞ on $B(X, \mathbf{F})$ considered in Section 2.13.

It then follows from Proposition 2.70 that $B(X, \mathbf{F})$ is complete, hence a Banach space. Similarly, the statement about $C_0(X, \mathbf{F})$ follows from Proposition 2.71. $\square$

A sequence (a_n) in a normed space $(V, \|\cdot\|)$ 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 $(x_m) \rightarrow x$, that is

$$\left\| x - \sum_{n=1}^m a_n \right\|_V \rightarrow 0 \quad \text{as } m \rightarrow \infty.$$

The limit x is called the *sum* of the series.

The series *converges absolutely* if the series of real numbers

$$\sum_{n=1}^{\infty} \|a_n\|_V$$

converges, that is there exists $r \in \mathbf{R}_{\geq 0}$ such that

$$\left(r - \sum_{n=1}^m \|a_n\|_V \right) \rightarrow 0 \quad \text{as } m \rightarrow \infty.$$

Proposition 3.14. *Let $(V, \|\cdot\|)$ be a normed space. V is a Banach space if and only if every absolutely convergent series in V is convergent.*

Proof. In one direction, suppose V is Banach and

$$\sum_{n=1}^{\infty} \|a_n\|_V = r \in \mathbf{R}_{\geq 0}.$$

Write

$$x_m = \sum_{n=1}^m a_n.$$

Let $\varepsilon > 0$, then there exists $M \in \mathbf{N}$ such that

$$\left| \sum_{n=1}^m \|a_n\|_V - r \right| < \frac{\varepsilon}{2} \quad \text{for all } m \geq M.$$

Then for all $m \geq k \geq M$ we have

$$\begin{aligned} \|x_m - x_k\|_V &= \left\| \sum_{n=k+1}^m a_n \right\|_V \leq \sum_{n=k+1}^m \|a_n\|_V = \sum_{n=1}^m \|a_n\|_V - \sum_{n=1}^k \|a_n\|_V \\ &= \left| \left(\sum_{n=1}^m \|a_n\|_V - r \right) + \left(r - \sum_{n=1}^k \|a_n\|_V \right) \right| \leq \left| \sum_{n=1}^m \|a_n\|_V - r \right| + \left| \sum_{n=1}^k \|a_n\|_V - r \right| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

So (x_m) is a Cauchy sequence in V , therefore it converges in V , meaning that the series

$$\sum_{n=1}^{\infty} a_n$$

converges in V .

In the other direction, suppose that every series that converges absolutely also converges in V , and let (a_n) be a Cauchy sequence in V . For each $\varepsilon > 0$ there exists $N(\varepsilon) \in \mathbf{N}$ such that for all $n \geq N$ we have $\|a_n - a_N\|_V < \varepsilon$.

Taking $\varepsilon = \frac{1}{2}, \frac{1}{2^2}, \dots$ we get

$$\begin{aligned} n_1 &\geq 1 \text{ such that } \|a_n - a_{n_1}\|_V < \frac{1}{2} \text{ for all } n \geq n_1, \\ n_2 &> n_1 \text{ such that } \|a_n - a_{n_2}\|_V < \frac{1}{2^2} \text{ for all } n \geq n_2, \\ &\vdots \\ n_k &> n_{k-1} \text{ such that } \|a_n - a_{n_k}\|_V < \frac{1}{2^k} \text{ for all } n \geq n_k, \\ &\vdots \end{aligned}$$

In particular, for all $k \in \mathbf{N}$ we have

$$\|a_{n_{k+1}} - a_{n_k}\|_V < \frac{1}{2^k},$$

so that

$$\sum_{k=1}^{\infty} \|a_{n_{k+1}} - a_{n_k}\|_V \leq \sum_{k=1}^{\infty} \frac{1}{2^k} = 1,$$

which implies that the series

$$\sum_{k=1}^{\infty} (a_{n_{k+1}} - a_{n_k}) \quad \text{absolutely converges,}$$

which by our assumption implies that the series

$$\sum_{k=1}^{\infty} (a_{n_{k+1}} - a_{n_k}) \quad \text{converges.}$$

Therefore the sequence of partial sums $(a_{n_k} - a_{n_1})$ (observe the telescoping behaviour) converges as $k \rightarrow \infty$, so the subsequence (a_{n_k}) of (a_n) converges, which by means that (a_n) converges. (Since any Cauchy sequence with a convergent subsequence is itself convergent.) $\square$

Proposition 3.15. *A normed space V is separable (as a metric space, i.e. has a countable dense subset) if and only if*

$$V = \overline{\text{Span}_{\mathbf{F}}(S)} \quad \text{for a countable subset } S \subseteq V.$$

You should start by looking at the special case where V is finite-dimensional, see [Tutorial Question 9.6.](#))

Proof. In one direction, if V is separable, then there is a countable subset D such that $V = \overline{D} \subseteq \overline{\text{Span}_{\mathbf{F}}(D)} \subseteq V$.

For the converse, suppose $V = \overline{\text{Span}_{\mathbf{F}}(S)}$ with S countable. Let

$$\mathbf{K} = \begin{cases} \mathbf{Q} & \text{if } \mathbf{F} = \mathbf{R} \\ \mathbf{Q}[i] & \text{if } \mathbf{F} = \mathbf{C} \end{cases} \quad \text{and let } D = \text{Span}_{\mathbf{K}}(S).$$

Then D is countable (as S and $\mathbf{K}$ are countable).

We claim that D is dense in $\text{Span}_{\mathbf{F}}(S)$. Take an arbitrary element of $\text{Span}_{\mathbf{F}}(S)$:

$$v = \alpha_1 v_1 + \cdots + \alpha_n v_n, \quad v_i \in S, \alpha_i \in \mathbf{F}.$$

Letting $W = \text{Span}_{\mathbf{F}}\{v_1, \dots, v_n\}$, we have a finite-dimensional normed vector space over $\mathbf{F}$, so by [Tutorial Question 9.6](#), given $\varepsilon > 0$ there exists $v' \in \text{Span}_{\mathbf{K}}(S)$ such that $\|v - v'\|_V = \|v - v'\|_W < \varepsilon$.

Since D is dense in $\text{Span}_{\mathbf{F}}(S)$ and $\text{Span}_{\mathbf{F}}(S)$ is dense in V , we conclude that D is dense in V , hence V is separable. $\square$

A *Schauder basis* of a normed space V is a sequence (e_n) of **unit** vectors of V such that for every $v \in V$ there exists a **unique** sequence of coefficients (α_n) in $\mathbf{F}$ with

$$v = \sum_{n=1}^{\infty} \alpha_n e_n,$$

which should be read as meaning that the series on the right hand side converges to $v \in V$.

If V has a Schauder basis, then

$$V = \overline{\text{Span}\{e_1, e_2, \dots\}},$$

so in particular V is separable¹.

3.4. BOUNDED LINEAR TRANSFORMATIONS

Let V and W be normed spaces.

A linear transformation $f: V \rightarrow W$ is said to be *bounded* (or *Lipschitz*²) if there exists $c > 0$ such that

$$\|f(v)\|_W \leq c \|v\|_V \quad \text{for all } v \in V.$$

Proposition 3.16. *A linear transformation $f: V \rightarrow W$ between normed spaces is continuous if and only if it is bounded if and only if it is uniformly continuous.*

Proof. (**bounded** $\Rightarrow$ **uniformly continuous** $\Rightarrow$ **continuous**): Suppose f is bounded with constant $c > 0$. Given $\varepsilon > 0$, let $\delta = \varepsilon/c$. If $v_1, v_2 \in V$ are such that $\|v_1 - v_2\|_V < \delta$, then

$$\|f(v_1) - f(v_2)\|_W = \|f(v_1 - v_2)\|_W \leq c \|v_1 - v_2\|_V < c\delta = \varepsilon.$$

Therefore f is uniformly continuous, hence continuous.

(**continuous** $\Rightarrow$ **bounded**): Suppose f is not bounded. Let $n \in \mathbf{N}$. There exists $v_n \in V$ such that

$$\|f(v_n)\|_W \geq n \|v_n\|_V.$$

Let $\alpha_n = 1/\|f(v_n)\|_W$ and $u_n = \alpha_n v_n$, then

$$\|u_n\|_V = |\alpha_n| \|v_n\|_V = \frac{\|v_n\|_V}{\|f(v_n)\|_W} \leq \frac{1}{n},$$

which implies that the sequence $(u_n) \rightarrow 0 \in V$.

But

$$\|f(u_n)\|_W = |\alpha_n| \|f(v_n)\|_W = 1,$$

so the sequence $(f(u_n))$ does not converge to $f(0) = 0$ in W , hence f is not continuous. $\square$

¹Note that not every separable normed space has a Schauder basis, but this is a nontrivial result.

²We will follow the overwhelmingly popular convention in the literature that such maps be called “bounded”, despite the fact that this clashes, for non-compact V , with the notion of bounded function we discussed in [Section 2.13](#). The two concepts are **related** but **not the same**.

We will write $L(V, W)$ for the set of bounded (aka continuous, aka uniformly continuous) linear transformations between the normed spaces V and W . In the special case $V = W$ we simply write $L(V) = L(V, V)$.

Consider the following function $\|\cdot\|: L(V, W) \rightarrow \mathbf{R}_{\geq 0}$:

$$\|f\| = \sup_{v \neq 0} \frac{\|f(v)\|_W}{\|v\|_V}.$$

As $f \in L(V, W)$, there exists $c > 0$ such that

$$\frac{\|f(v)\|_W}{\|v\|_V} \leq c \quad \text{for all } v \neq 0,$$

so that there is a finite supremum $\|f\|$.

We also note the obvious fact that

$$\|f(v)\|_W \leq \|f\| \|v\|_V \quad \text{for all } v \in V,$$

and that the linearity of f allows us to rewrite

$$\|f\| = \sup_{\|v\|_V=1} \|f(v)\|_W.$$

Theorem 3.17. *Let V and W be normed spaces.*

(a) *The set $L(V, W)$ is a normed space with norm given by*

$$\|f\| = \sup_{v \neq 0} \frac{\|f(v)\|_W}{\|v\|_V} = \sup_{\|v\|_V=1} \|f(v)\|_W.$$

(b) *Consider the map $N: L(V, W) \rightarrow B(V \setminus \{0\}, W)$ given by $N(f) = F$, where*

$$F: V \setminus \{0\} \rightarrow W, \quad F(v) := \frac{1}{\|v\|_V} f(v).$$

Then N is an isometry, and its image is a closed subset of $B(V \setminus \{0\}, W)$.

(c) *If W is a Banach space then $L(V, W)$ is also Banach.*

Proof.

(a) As $L(V, W)$ is a subset of $\text{Hom}(V, W)$ and the latter is a vector space, we check that $L(V, W)$ is a subspace.

We have

$$\begin{aligned} \|f + g\| &= \sup_{\|v\|_V=1} \|f(v) + g(v)\|_W \\ &\leq \sup_{\|v\|_V=1} (\|f(v)\|_W + \|g(v)\|_W) \\ &\leq \sup_{\|v\|_V=1} \|f(v)\|_W + \sup_{\|v\|_V=1} \|g(v)\|_W \\ &= \|f\| + \|g\|, \end{aligned}$$

so that if both f and g are in $L(V, W)$, so is $f + g$.

Similarly:

$$\|\alpha f\| = \sup_{\|v\|_V=1} \|\alpha f(v)\|_W = \sup_{\|v\|_V=1} |\alpha| \|f(v)\|_W = |\alpha| \|f\|,$$

so that if f is in $L(V, W)$ and $\alpha \in \mathbf{F}$, then αf is in $L(V, W)$.

In addition to showing that $L(V, W)$ is a vector space, these identities also give two of the three norm axioms, leaving to check that $\|f\| = 0$ if and only if $\|f(v)\|_W = 0$ for all $v \in V$ if and only if $f(v) = 0$ for all $v \in V$ if and only if $f = 0$.

- (b) Let $f \in L(V, W)$ and let $F = N(f)$. First note that if f is bounded with constant $c > 0$, then $d_W(F(v), 0) \leq c$ for all $v \in V \setminus \{0\}$, so F is bounded.

Next we see that if $g \in L(V, W)$ and $G = N(g)$, then

$$\begin{aligned} d_\infty(F, G) &= \sup_{v \in V \setminus \{0\}} \{d_W(F(v), G(v))\} \\ &= \sup_{v \in V \setminus \{0\}} \left\{ d_W \left(\frac{1}{\|v\|_V} f(v), \frac{1}{\|v\|_V} g(v) \right) \right\} \\ &= \sup_{v \in V \setminus \{0\}} \left\{ \frac{1}{\|v\|_V} \|f(v) - g(v)\|_W \right\} \\ &= \|f - g\|, \end{aligned}$$

so N is indeed an isometry.

Let F be in the closure of the image of N and let (F_n) be a sequence with $F_n = N(f_n)$ such that $(F_n) \rightarrow F$ with respect to the uniform metric.

Define $f: V \rightarrow W$ by setting

$$f(0) = 0 \quad \text{and} \quad f(v) = \|v\|_V F(v) \quad \text{for } v \in V \setminus \{0\}.$$

If we can show that $f \in L(V, W)$, then we are done, as clearly $N(f) = F$.

For linearity (ignoring corner cases where some vectors might be zero):

$$\begin{aligned} f(\lambda_1 v_1 + \lambda_2 v_2) &= \|\lambda_1 v_1 + \lambda_2 v_2\|_V F(\lambda_1 v_1 + \lambda_2 v_2) \\ &= \|\lambda_1 v_1 + \lambda_2 v_2\|_V \lim_{n \rightarrow \infty} F_n(\lambda_1 v_1 + \lambda_2 v_2) \\ &= \lim_{n \rightarrow \infty} (\|\lambda_1 v_1 + \lambda_2 v_2\|_V F_n(\lambda_1 v_1 + \lambda_2 v_2)) \\ &= \lim_{n \rightarrow \infty} f_n(\lambda_1 v_1 + \lambda_2 v_2) \\ &= \lim_{n \rightarrow \infty} (\lambda_1 f_n(v_1) + \lambda_2 f_n(v_2)) \\ &= \lambda_1 \lim_{n \rightarrow \infty} f_n(v_1) + \lambda_2 \lim_{n \rightarrow \infty} f_n(v_2) \\ &= \lambda_1 \|v_1\|_V \lim_{n \rightarrow \infty} F_n(v_1) + \lambda_2 \|v_2\|_V \lim_{n \rightarrow \infty} F_n(v_2) \\ &= \lambda_1 \|v_1\|_V F(v_1) + \lambda_2 \|v_2\|_V F(v_2) \\ &= \lambda_1 f(v_1) + \lambda_2 f(v_2). \end{aligned}$$

So f is linear. The fact that f is bounded follows immediately from the fact that $F = N(f)$ is bounded.

- (c) This follows from part (b), since W being Banach implies that $B(V \setminus \{0\}, W)$ is complete by [Proposition 2.70](#), so the image of N is complete as it is closed ([Proposition 2.47](#)), so $L(V, W)$ is complete since it is isometric to the image of N . $\square$

Let's record an important consequence of [Theorem 3.17](#):

Corollary 3.18. *For any normed space V , the dual space $V^\vee = L(V, \mathbf{F})$ is a Banach space with norm*

$$\|\varphi\| = \sup_{v \neq 0} \frac{|\varphi(v)|}{\|v\|_V}.$$

We'll come back to the topic of dual spaces.

To prove that $L(V, W)$ is a normed space, we had to consider the interplay between the addition of functions and the norms on V and W , and similarly for the operation of multiplying a function by a scalar. There is another operation on functions that has been conspicuously missing from this discussion: composition. We look at this now.

Recall (or see [Section A.3](#)) that an algebra is a vector space A with a vector multiplication map $A \times A \rightarrow A$, $(u, v) \mapsto uv$.

For example, given a vector space V over $\mathbf{F}$, the set of all $\mathbf{F}$ -linear transformations $V \rightarrow V$ is an associative unital $\mathbf{F}$ -algebra, where multiplication is given by composition and the unit is $\mathbf{1} = \text{id}_V$.

Proposition 3.19. *If $f: U \rightarrow V$ and $g: V \rightarrow W$ are continuous linear transformations between normed spaces, then $g \circ f: U \rightarrow W$ is continuous and linear, and*

$$\|g \circ f\| \leq \|g\| \|f\|.$$

In particular, for any normed space V , the normed space $L(V)$ is closed under composition, hence is an associative unital $\mathbf{F}$ -algebra.

Proof. We know already that the composition of linear maps is linear, and that the composition of continuous maps is continuous.

As for the norms, for any $u \in U$ we have

$$\|(g \circ f)(u)\|_W = \|g(f(u))\|_W \leq \|g\| \|f(u)\|_V \leq \|g\| \|f\| \|u\|_U,$$

so that for all $u \neq 0$ we have

$$\frac{\|(g \circ f)(u)\|_W}{\|u\|_U} \leq \|g\| \|f\|,$$

and we can conclude by taking supremum.

If $U = W = V$ we get the $\mathbf{F}$ -algebra $L(V)$ with multiplication given by composition, and with unit element $\mathbf{1} = \text{id}_V$, clearly both linear and continuous. $\square$

3.5. CONVEXITY AND INEQUALITIES

A subset S of a vector space V over $\mathbf{F}$ is *convex* if for all $v, w \in S$ and all $a, b \in \mathbf{R}_{\geq 0}$ such that $a + b = 1$, we have $av + bw \in S$. (In other words, for any two points in S , the line segment joining the two points is entirely contained in S .)

Example 3.20. Any subspace W of V is convex.

Solution. Suppose $v, w \in W$, $a, b \in \mathbf{R}_{\geq 0}$ such that $a + b = 1$. Then $av + bw$ is an $\mathbf{F}$ -linear combination of elements of W . Since W is a subspace, $av + bw \in W$. $\square$

Another simple example is: any interval $I \subseteq \mathbf{R}$ is convex, see [Exercise 2.20](#).

If V is a vector space over $\mathbf{F}$ and $S \subseteq V$ is a convex set, we say that a function $f: S \rightarrow \mathbf{R}$ is *convex* if for all $v, w \in S$ and all $a, b \in \mathbf{R}_{\geq 0}$ such that $a + b = 1$, we have

$$f(av + bw) \leq af(v) + bf(w).$$

For instance, if $(V, \|\cdot\|)$ is a normed space, then the norm $\|\cdot\|: V \rightarrow \mathbf{R}_{\geq 0}$ is a convex function, see [Exercise 2.21](#).

More interestingly, the notion of convex function is closely related to the concept of concavity in single-variable calculus:

Proposition 3.21. *Let $I \subseteq \mathbf{R}$ be an interval and let $f: I \rightarrow \mathbf{R}$ be a twice-differentiable function.*

Then f is convex if and only if $f''(x) \geq 0$ for all $x \in I$.

Proof. See [Exercise 2.22](#). □

Lemma 3.22. *Let $x, y \geq 0$, $p \geq 1$, $a, b \geq 0$ such that $a + b = 1$. Then*

$$(ax + by)^p \leq ax^p + by^p.$$

Proof. Given that $p \geq 1$, the function $f: (0, \infty) \rightarrow \mathbf{R}$ given by $f(x) = x^p$ is convex by [Proposition 3.21](#), since $f''(x) = p(p-1)x^{p-2} \geq 0$ for all $x > 0$.

The claimed inequality follows from the convexity property. □

3.6. SEQUENCE SPACES

The set of all sequences $\mathbf{F}^{\mathbf{N}} = \{(a_n): a_n \in \mathbf{F} \text{ for all } n \in \mathbf{N}\}$ is of course a vector space over $\mathbf{F}$ with the usual addition and scalar multiplication.

We consider a family of subsets of $\mathbf{F}^{\mathbf{N}}$, parametrised by a real number $p \geq 1$:

$$\ell^p = \left\{ (a_n) \in \mathbf{F}^{\mathbf{N}} : \sum_{n=1}^{\infty} |a_n|^p < \infty \right\},$$

the elements of which are called *p-summable* sequences. (If $p = 1$ we simply call them *summable*, and if $p = 2$, *square-summable*.) We consider the function $\|\cdot\|_{\ell^p}: \ell^p \rightarrow \mathbf{R}_{\geq 0}$ defined by

$$\|(a_n)\|_{\ell^p} = \left(\sum_{n=1}^{\infty} |a_n|^p \right)^{1/p}.$$

There is also an exceptional case $p = \infty$ given by *bounded* sequences

$$\begin{aligned} \ell^{\infty} &= \{(a_n) \in \mathbf{F}^{\mathbf{N}} : \sup(|a_n|) < \infty\} \\ &= \{(a_n) \in \mathbf{F}^{\mathbf{N}} : \text{there exists } M \text{ such that } |a_n| \leq M \text{ for all } n \in \mathbf{N}\}, \end{aligned}$$

with function $\|\cdot\|_{\ell^{\infty}}: \ell^{\infty} \rightarrow \mathbf{R}_{\geq 0}$ given by

$$\|(a_n)\|_{\ell^{\infty}} = \sup(|a_n|).$$

The upshot is that all these subsets of $\mathbf{F}^{\mathbf{N}}$ are normed spaces, as we now see.

Lemma 3.23. *Suppose $p \geq 1$. The function $f: \ell^p \rightarrow \mathbf{R}_{\geq 0}$ given by*

$$f(x) = \|x\|_{\ell^p}^p = \sum_{n=1}^{\infty} |x_n|^p$$

is convex.

Proof. Let $x = (x_n), y = (y_n) \in \ell^p$, and let $a, b \geq 0$ be such that $a + b = 1$. Then

$$\sum_{n=1}^{\infty} |ax_n + by_n|^p \leq \sum_{n=1}^{\infty} (a|x_n| + b|y_n|)^p \leq a \sum_{n=1}^{\infty} |x_n|^p + b \sum_{n=1}^{\infty} |y_n|^p,$$

where we applied first the triangle inequality for the absolute value, and second the inequality from [Lemma 3.22](#). Therefore

$$f(ax + by) = \|ax + by\|^p \leq a\|x\|^p + b\|y\|^p = af(x) + bf(y).$$

In order to finish the proof that f is a convex function, we also need to show that ℓ^p is a convex subset of the vector space $\mathbf{F}^{\mathbf{N}}$. But this is built into the inequality we proved above, which shows that if $x, y \in \ell^p$ and $a, b \geq 0$ are such that $a + b = 1$, then $ax + by \in \ell^p$. $\square$

Proposition 3.24 (Minkowski's Inequality). *Let $p \geq 1$ and let $u = (u_n), v = (v_n) \in \ell^p$. Then*

$$\|u + v\|_{\ell^p} \leq \|u\|_{\ell^p} + \|v\|_{\ell^p}.$$

Proof. Fix p and write $\|\cdot\|$ instead of $\|\cdot\|_{\ell^p}$ to simplify notation.

Given $u, v \in \ell^p$, define

$$x = \frac{1}{\|u\|} u, \quad y = \frac{1}{\|v\|} v, \quad a = \frac{\|u\|}{\|u\| + \|v\|}, \quad b = \frac{\|v\|}{\|u\| + \|v\|},$$

then we have by [Lemma 3.23](#)

$$\left(\frac{\|u + v\|}{\|u\| + \|v\|} \right)^p = \|ax + by\|^p \leq a + b = 1. \quad \square$$

Note that Minkowski's Inequality also holds for $p = \infty$, see [Tutorial Question 10.3](#).

Corollary 3.25. *For $1 \leq p \leq \infty$, the set ℓ^p is a vector subspace of $\mathbf{F}^{\mathbf{N}}$, and $\|\cdot\|_{\ell^p}$ is a norm on ℓ^p .*

Proof. It is clear from the definition of ℓ^p that it contains the constant zero sequence $\mathbf{0}$, and that it is closed under scalar multiplication. By Minkowski's Inequality it is also closed under vector addition, so it is a subspace.

Minkowski's Inequality also gives us the triangle inequality for $\|\cdot\|_{\ell^p}$, as well as the behaviour under scalar multiplication. Finally, if (a_n) is such that there exists $n \in \mathbf{N}$ with $|a_n| > 0$, then $\|(a_n)\|_{\ell^p} \geq |a_n| > 0$. So $\|(a_n)\|_{\ell^p} = 0$ if and only if $(a_n) = \mathbf{0}$. $\square$

The normed spaces ℓ^p have some desirable properties; notably they are Banach, as we will soon see.

For now let us establish:

Proposition 3.26. *For any $1 \leq p < \infty$, the sequence space ℓ^p has Schauder basis $\{e_1, e_2, \dots\}$, where*

$$e_n = (0, \dots, 0, 1, 0, \dots) \quad \text{with the 1 in the } n\text{-th spot.}$$

In particular, ℓ^p is separable.

Proof. This is an essentially trivial exercise in checking the definition.

Take an arbitrary element $v = (v_n) \in \ell^p$, then

$$\sum_{n=1}^{\infty} |v_n|^p$$

converges with sum $\|v\|^p$.

I claim that the series

$$\sum_{n=1}^{\infty} v_n e_n$$

converges to v with respect to the ℓ^p -norm:

$$\left\| v - \sum_{n=1}^m v_n e_n \right\|_{\ell^p}^p = \left\| (0, \dots, 0, v_{m+1}, v_{m+2}, v_{m+3}, \dots) \right\|_{\ell^p}^p = \sum_{n=m+1}^{\infty} |v_n|^p,$$

and the latter converges to 0 as $m \rightarrow \infty$.

The uniqueness of the sequence of coefficients follows from the fact that

$$(v_1, v_2, \dots) = \sum_{n=1}^{\infty} v_n e_n = v = \sum_{n=1}^{\infty} u_n e_n = (u_1, u_2, \dots)$$

implies $v_n = u_n$ for all $n \in \mathbf{N}$. □

Note however that ℓ^∞ is **not** separable, see [Exercise 2.33](#).

3.7. DUAL NORMED SPACES AND COMPLETENESS OF SEQUENCE SPACES

You may want to have a look at [Sections B.2](#) and [B.3](#) and read the discussion of bilinear maps and dual spaces. We will only touch on some basic points here, focussing on the new aspects coming from the norm.

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

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

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

For instance, given $n \in \mathbf{N}$, there is a bilinear map $\beta: \mathbf{F}^n \times \mathbf{F}^n \rightarrow \mathbf{F}$ given by

$$\beta(u, v) = \sum_{k=1}^n u_k v_k.$$

As described in [Section B.3](#), this defines a linear map $\mathbf{F}^n \rightarrow (\mathbf{F}^n)^\vee$, $u \mapsto u^\vee$, given by $u^\vee(v) = \beta(u, v)$.

We'd like to do the same with (subspaces of) $\mathbf{F}^\mathbf{N}$: define a bilinear map $\beta: \mathbf{F}^\mathbf{N} \times \mathbf{F}^\mathbf{N} \rightarrow \mathbf{F}$ by the formula

$$\beta(u, v) = \sum_{n=1}^{\infty} u_n v_n.$$

Of course this would feel more comfortable if we knew that the series $\sum u_n v_n$ actually converges! And of course that does not happen for arbitrary $u, v \in \mathbf{F}^\mathbf{N}$, but we can establish some situations where it does work, as follows.

If $p \geq 1$, 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 $q \geq 1$. Note that this includes the degenerate pair $p = 1$, $q = \infty$.

Proposition 3.27 (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*

$$\sum_{n=1}^{\infty} |u_n v_n| \leq \|u\|_{\ell^p} \|v\|_{\ell^q}.$$

Proof. We prove the non-degenerate case $p, q \in \mathbf{R}_{>1}$ and leave the (simpler) degenerate one to [Tutorial Question 10.4](#).

Let $x = (x_n) \in \ell^p$, $y = (y_n) \in \ell^q$. For each $n \in \mathbf{N}$ we have

$$|x_n y_n| = (|x_n|^p)^{1/p} (|y_n|^q)^{1/q} \leq \frac{1}{p} |x_n|^p + \frac{1}{q} |y_n|^q,$$

by an application of [Exercise 2.23](#), namely $s^a t^b \leq as + bt$ where $a + b = 1$.

Therefore

$$\sum_{n=1}^{\infty} |x_n y_n| \leq \sum_{n=1}^{\infty} \left(\frac{1}{p} |x_n|^p + \frac{1}{q} |y_n|^q \right) = \frac{1}{p} \|x\|_{\ell^p}^p + \frac{1}{q} \|y\|_{\ell^q}^q.$$

Now start with $u \in \ell^p$, $v \in \ell^q$ and set

$$x = \frac{1}{\|u\|_{\ell^p}} u, \quad y = \frac{1}{\|v\|_{\ell^q}} v,$$

so that we obtain

$$\frac{\sum_{n=1}^{\infty} |u_n v_n|}{\|u\|_{\ell^p} \|v\|_{\ell^q}} \leq \frac{1}{p} + \frac{1}{q} = 1. \quad \square$$

Before we give the main result of this section, we should extend the notion of continuous linear map to the setting of bilinear maps.

If U, V, W are normed spaces, a bilinear map $\beta: U \times V \rightarrow W$ is *bounded* if there exists $c > 0$ such that

$$\|\beta(u, v)\|_W \leq c \|u\|_U \|v\|_V \quad \text{for all } u \in U, v \in V.$$

Proposition 3.28. *If U, V, W are normed spaces, a bilinear map $\beta: U \times V \rightarrow W$ is bounded if and only if it is continuous.*

Proof. Somewhat tedious, following the example of [Proposition 3.16](#). See [Exercise 2.36](#). $\square$

Beware: in contrast to the linear case, continuous bilinear maps are almost never uniformly continuous. See [Exercise 2.37](#).

Theorem 3.29. *If p, q are Hölder conjugates, then $\beta: \ell^p \times \ell^q \rightarrow \mathbf{F}$ given by*

$$\beta(u, v) = \sum_{n=1}^{\infty} u_n v_n$$

is a continuous bilinear map.

Moreover, if $p, q > 1$, the resulting continuous linear map

$$u \mapsto u^\vee: \ell^p \rightarrow (\ell^q)^\vee$$

is a bijective isometry $\ell^p \cong (\ell^q)^\vee$.

Proof. By Hölder's Inequality, the series defining $\beta(u, v)$ converges absolutely. It is then straightforward to check that β is bilinear.

Conveniently, Hölder's Inequality also tells us that β is a bounded bilinear form, hence continuous.

Now we know that $u^\vee \in (\ell^q)^\vee$, but we can (and will) say something more precise.

We start by proving the surjectivity of $u \mapsto u^\vee$.

Let $\varphi \in (\ell^q)^\vee$ and let $v \in \ell^q$. Let $\{e_1, e_2, \dots\}$ be the Schauder basis for ℓ^q discussed in Proposition 3.26. We have

$$\varphi(v) = \varphi\left(\sum_{n=1}^{\infty} v_n e_n\right) = \sum_{n=1}^{\infty} v_n \varphi(e_n) \quad \text{and } \|e_n\|_{\ell^q} = 1 \text{ for all } n \in \mathbf{N}.$$

Define $u_n = \varphi(e_n)$ and $u = (u_n)$. If we show that $u \in \ell^p$ then we have $\varphi(v) = u^\vee(v)$ and we're done.

For any $m \in \mathbf{N}$, consider (ignore all the u_n 's that are zero, as they do not contribute to the sums):

$$x = \sum_{n=1}^m \frac{|u_n|^p}{u_n} e_n = \left(\frac{|u_1|^p}{u_1}, \dots, \frac{|u_m|^p}{u_m}, 0, 0, \dots \right),$$

so that

$$\|x\|_{\ell^q} = \left(\sum_{n=1}^m (|u_n|^{p-1})^q \right)^{1/q} = \left(\sum_{n=1}^m |u_n|^p \right)^{1/q}.$$

With this in mind, we have

$$\begin{aligned} \sum_{n=1}^m |u_n|^p &= \left| \sum_{n=1}^m \frac{|u_n|^p}{u_n} u_n \right| = \left| \sum_{n=1}^m \varphi\left(\frac{|u_n|^p}{u_n} e_n\right) \right| \\ &= |\varphi(x)| \leq \|\varphi\| \|x\|_{\ell^q} = \|\varphi\| \left(\sum_{n=1}^m |u_n|^p \right)^{1/q}. \end{aligned}$$

Therefore

$$\left(\sum_{n=1}^m |u_n|^p \right)^{1/p} = \left(\sum_{n=1}^m |u_n|^p \right)^{1-1/q} \leq \|\varphi\|.$$

As this holds for all $m \in \mathbf{N}$, we conclude that the series converges, so $u \in \ell^p$. We also proved that $\|u\|_{\ell^p} \leq \|\varphi\| = \|u^\vee\|$.

Let $v \neq 0$. By Hölder's Inequality

$$\frac{|u^\vee(v)|}{\|v\|_{\ell^q}} \leq \|u\|_{\ell^p},$$

so taking supremum we get $\|u^\vee\| \leq \|u\|_{\ell^p}$.

As we established both inequalities, we conclude that $u \mapsto u^\vee$ is a surjective isometry (hence also injective) from ℓ^p to $(\ell^q)^\vee$. $\square$

Corollary 3.30. *If $p > 1$ then ℓ^p is a Banach space.*

Proof. Follows as $\ell^p \cong (\ell^q)^\vee$ and all dual normed spaces are Banach. $\square$

Proposition 3.31. *The sequence space ℓ^2 of square-summable sequences is a Hilbert space.*

Proof. Consider the function $\langle \cdot, \cdot \rangle: \ell^2 \times \ell^2 \rightarrow \mathbf{F}$ given by

$$\langle a, b \rangle = \sum_{n=1}^{\infty} a_n \bar{b}_n.$$

We use the Cauchy–Schwarz Inequality ([Proposition A.9](#)) to see that this converges. For any $m \in \mathbf{N}$, $(a_1, \dots, a_m), (b_1, \dots, b_m) \in \mathbf{F}^m$ so by Cauchy–Schwarz we have

$$\left| \sum_{n=1}^m a_n \bar{b}_n \right| \leq \left(\sum_{n=1}^m a_n \bar{a}_n \right)^{1/2} \left(\sum_{n=1}^m b_n \bar{b}_n \right)^{1/2} = \left(\sum_{n=1}^m |a_n|^2 \right)^{1/2} \left(\sum_{n=1}^m |b_n|^2 \right)^{1/2}.$$

Taking limits as $m \rightarrow \infty$, 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 3.30](#). $\square$

The rest of the ℓ^p spaces are **not** Hilbert spaces, see [Exercise 2.30](#).

3.8. ORTHOGONALITY AND PROJECTIONS

For the next few sections, we will explore some special properties of inner product spaces.

Given a normed space V , a **projection** is a continuous linear map $\pi \in L(V)$ such that $\pi^2 = \pi$.

Proposition 3.32. *Let $\pi \in L(V)$ be a projection.*

(a) *The map $\text{id}_V - \pi$ is also a projection.*

(b) *$\text{im}(\pi) = \ker(\text{id}_V - \pi)$ and $\text{im}(\text{id}_V - \pi) = \ker(\pi)$. In particular, the image of a projection is a closed subspace.*

(c) *We have*

$$V = \text{im}(\pi) \oplus \ker(\pi).$$

Solution. (a) Since both id_V and π are continuous and linear, so is $\text{id}_V - \pi$. Also, we have

$$(\text{id}_V - \pi) \circ (\text{id}_V - \pi) = \text{id}_V - \pi - \pi + \pi \circ \pi = \text{id}_V - \pi.$$

(b) If $v \in \text{im}(\pi)$ then $v = \pi(w)$ so that

$$(\text{id}_V - \pi)(v) = v - \pi(v) = \pi(w) - \pi^2(w) = \pi(w) - \pi(w) = 0,$$

so $v \in \ker(\text{id}_V - \pi)$.

Conversely, if $v \in \ker(\text{id}_V - \pi)$ then $v - \pi(v) = 0$ so $v = \pi(v) \in \text{im}(\pi)$.

The other identity follows by applying the first identity to the projection $\text{id}_V - \pi$.

Since the image of π is the kernel of $\text{id}_V - \pi$, it is a closed subspace, as the kernel of any linear continuous map is a closed subspace.

(c) We need to prove that $V = \text{im}(\pi) + \ker(\pi)$ and that $\text{im}(\pi) \cap \ker(\pi) = \{0\}$.

Given $v \in V$, we have

$$v = \pi(v) + (\text{id}_V - \pi)(v) \in \text{im}(\pi) + \ker(\pi).$$

If

$$w \in \text{im}(\pi) \cap \ker(\pi) = \ker(\text{id}_V - \pi) \cap \ker(\pi),$$

then

$$w = \pi(w) + (\text{id}_V - \pi)(w) = 0 + 0 = 0.$$

$\square$

Example 3.33. Take $V = \mathbf{R}^2$ with the Euclidean norm. The matrix

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

satisfies $A^2 = A$, so it defines a projection. It is easy to see that $\text{im}(A)$ is the diagonal $y = x$ in $\mathbf{R}^2$, and $\ker(A)$ is the y -axis.

The complementary projection is given by the matrix

$$I - A = \begin{pmatrix} 0 & 0 \\ -1 & 1 \end{pmatrix},$$

where $\text{im}(I - A)$ is the y -axis and $\ker(I - A)$ is the diagonal $y = x$.

Given a subset S of an inner product space V , we define the *orthogonal complement of S* by

$$S^\perp = \{v \in V : \langle v, s \rangle = 0 \text{ for all } s \in S\}.$$

Proposition 3.34. *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. $\square$

If V is an inner product space, an *orthogonal projection* is a projection π such that $\ker(\pi) = (\text{im}(\pi))^\perp$, so that we have (by Proposition 3.32)

$$V = \text{im}(\pi) \oplus (\text{im}(\pi))^\perp.$$

Recall from Tutorial Question 5.6 that for any subset $Y \subseteq X$ of a metric space, we can define a function $d_Y : X \rightarrow \mathbf{R}_{\geq 0}$ that gives the distance to the set Y :

$$d_Y(x) = \inf_{y \in Y} d(x, y).$$

Theorem 3.35 (Hilbert Projection Theorem, Part I). *Let H be a Hilbert space and Y a closed **convex subset** of H . For any $x \in H$, there exists a unique $y_{\min} \in Y$ that realises the distance between x and Y :*

$$d_Y(x) = d(x, y_{\min}) = \|x - y_{\min}\|.$$

In other words, $y_{\min}$ is the unique point of Y that is as close as possible to x .

Proof. Let

$$D = d_Y(x) = \inf_{y \in Y} d(x, y).$$

Take a sequence (y_n) in Y such that

$$(\|x - y_n\|) = (d(x, y_n)) \rightarrow D.$$

I claim that the sequence (y_n) is Cauchy.

Let $\varepsilon > 0$. Note that

$$(\|x - y_n\|^2) \rightarrow D^2,$$

so there exists $N \in \mathbf{N}$ such that

$$\left| \|x - y_n\|^2 - D^2 \right| \leq \frac{\varepsilon}{4} \quad \text{for all } n \geq N.$$

Let $m, n \geq N$. By the Parallelogram Law:

$$\|(y_n - x) + (y_m - x)\|^2 + \|(y_n - x) - (y_m - x)\|^2 = 2\|y_n - x\|^2 + 2\|y_m - x\|^2,$$

so that

$$\begin{aligned} \|y_n - y_m\|^2 &= 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - \|(y_n + y_m) - 2x\|^2 \\ &= 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4\left\| \frac{y_n + y_m}{2} - x \right\|^2. \end{aligned}$$

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

$$\begin{aligned} 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4\left\| \frac{y_n + y_m}{2} - x \right\|^2 &\leq 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4D^2 \\ &= 2(\|y_n - x\|^2 - D^2) + 2(\|y_m - x\|^2 - D^2) \\ &< \varepsilon. \end{aligned}$$

So (y_n) is Cauchy in Y , which is complete (being a closed subset of the Hilbert space H). Therefore (y_n) converges in Y to some point that we will call $y_{\min}$. Since the distance function is continuous, we have

$$d(x, y_{\min}) = \lim_{n \rightarrow \infty} d(x, y_n) = D = d_Y(x).$$

It remains to prove the uniqueness of $y_{\min}$. Suppose $y' \in Y$ satisfies $d(x, y') = D$. By the Parallelogram Law

$$\|(y_{\min} - x) + (y' - x)\|^2 + \|(y_{\min} - x) - (y' - x)\|^2 = 2\|y_{\min} - x\|^2 + 2\|y' - x\|^2,$$

so that

$$\|y_{\min} - y'\|^2 = 2\|y_{\min} - x\|^2 + 2\|y' - x\|^2 - \|(y_{\min} + y') - 2x\|^2 \leq 2D^2 + 2D^2 - 4D^2 = 0,$$

which implies $y' = y_{\min}$. □

Theorem 3.36 (Hilbert Projection Theorem, Part II). *Let H be a Hilbert space and W a closed **vector subspace** of H . Let $x \in H$ and let $y_{\min}$ be the unique point of W that realises the distance between x and W , as given by the Hilbert Projection Theorem, Part I. For any $y \in W$, we have*

$$y = y_{\min} \quad \text{if and only if} \quad x - y \in W^\perp.$$

The map $\pi: H \rightarrow H$ given by $\pi(x) = y_{\min}$ is an orthogonal projection with image W . In particular, we have a decomposition

$$H = W \oplus W^\perp.$$

Proof. First we prove that $x - y_{\min} \in W^\perp$.

Let $w \in W$ be a unit vector, so $\|w\| = 1$. Letting $\alpha = \langle x - y_{\min}, w \rangle$, we want to show that $\alpha = 0$.

Set $v = x - (y_{\min} + \alpha w)$, 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

$$\|x - y_{\min}\|^2 = \|v + \alpha w\|^2 = \|v\|^2 + |\alpha|^2 \|w\|^2 = \|v\|^2 + |\alpha|^2 \geq \|v\|^2,$$

in other words

$$\|x - y_{\min}\| \geq \|x - (y_{\min} + \alpha w)\|.$$

By the minimality and uniqueness of $y_{\min}$, we must have $y_{\min} = y_{\min} + \alpha w$, so $\alpha = 0$.

Next we show that if $y \in W$ and $x - y \in W^\perp$ then $y = y_{\min}$.

We have

$$\begin{aligned}x - y \in W^\perp &\Rightarrow \langle x - y, w \rangle = 0 && \text{for all } w \in W \\ &\Rightarrow \langle x - y, w - y \rangle = 0 && \text{for all } w \in W \\ &\Rightarrow \|x - w\|^2 = \|x - y\|^2 + \|w - y\|^2 && \text{for all } w \in W \\ &\Rightarrow \|x - w\|^2 \geq \|x - y\|^2 && \text{for all } w \in W,\end{aligned}$$

implying that $y \in W$ is closest to x ; by the uniqueness of $y_{\min}$, we conclude that $y = y_{\min}$.

We now move on to the function π . As we have just seen, for each $x \in H$, $\pi(x)$ is the unique element of W with the property that $x - \pi(x) \in W^\perp$.

We check that π is linear.

If $x_1, x_2 \in H$, we have $\pi(x_1) + \pi(x_2) \in W$ and

$$(x_1 + x_2) - (\pi(x_1) + \pi(x_2)) = (x_1 - \pi(x_1)) + (x_2 - \pi(x_2)) \in W^\perp,$$

so $\pi(x_1) + \pi(x_2) = \pi(x_1 + x_2)$.

Similarly, if $x \in H$ and $\alpha \in \mathbf{F}$ we have $\alpha\pi(x) \in W$ and

$$\alpha x - \alpha\pi(x) = \alpha(x - \pi(x)) \in W^\perp,$$

so $\alpha\pi(x) = \pi(\alpha x)$.

We check that π is bounded, hence continuous.

For any $x \in H$, we have $\pi(x) \in W$ and $x - \pi(x) \in W^\perp$, so $(x - \pi(x)) \perp \pi(x)$ and

$$\|x\|^2 = \|(x - \pi(x)) + \pi(x)\|^2 = \|x - \pi(x)\|^2 + \|\pi(x)\|^2 \geq \|\pi(x)\|^2,$$

so $\|\pi(x)\| \leq \|x\|$.

We check that π is a projection with image W .

Certainly $\text{im } \pi \subseteq W$. If $y \in W$ then $\pi(y) = y$ (closest point to y is y itself), so in fact $\text{im } \pi = W$. Hence for all $x \in H$ we get $\pi^2(x) = \pi(\pi(x)) = \pi(x)$, so $\pi^2 = \pi$.

Finally, we check that π is an orthogonal projection.

We want to show that $W^\perp = \ker \pi$. But $x \in W^\perp$ if and only if $x - 0 \in W^\perp$ if and only if $\pi(x) = 0$ if and only if $x \in \ker \pi$. $\square$

Corollary 3.37. *Let H be a Hilbert space.*

(a) *If W is a closed linear subspace of H , then $(W^\perp)^\perp = W$.*

(b) If V is a linear subspace of H , then $(V^\perp)^\perp = \overline{V}$.

(c) If S is a subset of H , then $(S^\perp)^\perp = \overline{\text{Span}(S)}$.

Proof.

(a) By [Exercise 2.42](#) we have $W \subseteq (W^\perp)^\perp$.

Now let $x \in (W^\perp)^\perp$. By the Hilbert Projection Theorem Part II, we can decompose

$$H = W \oplus W^\perp.$$

So we have $x = y + z$ with $y \in W$ and $z \in W^\perp$. Then

$$0 = \langle x, z \rangle = \langle y + z, z \rangle = \langle y, z \rangle + \langle z, z \rangle = 0 + \|z\|^2,$$

implying that $z = 0$ and $x = y \in W$.

(b) See [Tutorial Question 11.1](#).

(c) See [Tutorial Question 11.1](#). □

A. APPENDIX: PREREQUISITES

A.1. EQUIVALENCE RELATIONS

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

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

The following example should be very familiar:

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

Solution.

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

□

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

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

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

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

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

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

□

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

Solution. Given $m \in \mathbf{Z}$, let $0 \leq r \leq n - 1$ be the remainder of the division of m by n : $m = qn + r$. Then $m - r$ is divisible by n , hence $m \sim r$. From the previous part, we know

that there are at most n equivalence classes, one for each possible value of r . To show that we have exactly n buddy groups, we need to prove that $[r_1] \neq [r_2]$ for any $r_1 \neq r_2$ with $0 \leq r_1, r_2 \leq n-1$. We do this by contradiction: if $[r_1] = [r_2]$ then $r_1 \sim r_2$, so $r_2 - r_1$ is a multiple of n . But $-(n-1) \leq r_2 - r_1 \leq (n-1)$, and the only multiple of n in that interval is 0, in other words $r_2 = r_1$, contradiction. $\square$

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

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

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

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

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

Show that this is an equivalence relation.

Solution. • Let $X \in M_{m \times n}$. The identity elementary row operation takes X to X , so $X \sim X$.

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

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

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

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

A.2. CARDINALITY

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

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

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

start at 0.

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

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

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

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

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

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

A.3. MAPS BETWEEN VECTOR SPACES

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

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

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

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

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

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

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

The algebra A is *associative* if

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

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

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

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

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

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

$$v = a_1v_1 + \dots + a_nv_n, \quad n \in \mathbf{N}, a_j \in \mathbf{F}, v_j \in B.$$

Therefore the only option is to set

$$f(v) = a_1g(v_1) + \dots + a_ng(v_n).$$

It is easy to see that f is linear.

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

A.4. INNER PRODUCTS

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

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

An *inner product* on V is a function

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

such that

- (a) $\langle w, v \rangle = \overline{\langle v, w \rangle}$ for all $v, w \in V$;
- (b) $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ for all $u, v, w \in V$;
- (c) $\langle \alpha v, w \rangle = \alpha \langle v, w \rangle$ for all $v, w \in V$, all $\alpha \in \mathbf{F}$;
- (d) $\langle v, v \rangle \geq 0$ for all $v \in V$ and $\langle v, v \rangle = 0$ iff $v = 0$.

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

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

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

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

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

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

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

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

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

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

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

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

$$\bar{A}^T = A \quad \text{and all the eigenvalues of } A \text{ are real and positive.}$$

Over $\mathbf{R}$, A is a positive-definite¹ symmetric matrix.

Define

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

Proposition A.9 (Cauchy–Schwarz Inequality). *Take u, v in an inner product space V . Then*

$$|\langle u, v \rangle| \leq \|u\| \|v\|,$$

where equality holds if and only if u and v are parallel, that is $u = \lambda v$ for some $\lambda \in \mathbf{F}$.

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

Proof. If $u = \mathbf{0}$ or $v = \mathbf{0}$, we have the equality $0 = 0$. Otherwise, for any $t \in \mathbf{F}$ we have

$$\begin{aligned} 0 &\leq \langle u - tv, u - tv \rangle = \langle u, u \rangle - 2 \operatorname{Re}(\bar{t} \langle u, v \rangle) + t \bar{t} \langle v, v \rangle \\ &= \|u\|^2 - 2 \operatorname{Re}(\bar{t} \langle u, v \rangle) + |t|^2 \|v\|^2. \end{aligned}$$

In particular, we can take $t = \frac{\langle u, v \rangle}{\|v\|^2}$:

$$0 \leq \|u\|^2 - 2 \operatorname{Re}\left(\frac{|\langle u, v \rangle|^2}{\|v\|^2}\right) + \frac{|\langle u, v \rangle|^2}{\|v\|^2} = \|u\|^2 - \frac{|\langle u, v \rangle|^2}{\|v\|^2},$$

so $|\langle u, v \rangle|^2 \leq \|u\|^2 \|v\|^2$.

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

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

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

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

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

for all $u, v \in S$.

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

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

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

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

A.5. UNIFORM CONTINUITY AND UNIFORM CONVERGENCE

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

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

for every $x \in X$ and every $\varepsilon > 0$, there exists $\delta > 0$ such that:

$$\text{for all } y \in X, \text{ if } |x - y| < \delta, \text{ then } |f(x) - f(y)| < \varepsilon.$$

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

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

for every $\varepsilon > 0$, there exists $\delta > 0$ such that:

$$\text{for all } x, y \in X, \text{ if } |x - y| < \delta, \text{ then } |f(x) - f(y)| < \varepsilon.$$

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

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

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

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

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

Also

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

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

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

for every $x \in X$ and every $\varepsilon > 0$, there exists $N \in \mathbf{N}$ such that:

$$\text{if } n \geq N, \text{ then } |f_n(x) - f(x)| < \varepsilon.$$

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

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

for every $\varepsilon > 0$, there exists $N \in \mathbf{N}$ such that:

$$\text{for every } x \in X, \text{ if } n \geq N, \text{ then } |f_n(x) - f(x)| < \varepsilon.$$

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

Example A.12. For $n \geq 1$, consider $f_n: \mathbf{R} \rightarrow \mathbf{R}$ given by

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

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

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

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

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

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

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

$$f: \mathbf{R} \rightarrow \mathbf{R}, \quad f(x) = x.$$

Solution. We have

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

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

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

B. APPENDIX: MISCELLANEOUS

B.1. ZORN'S LEMMA

A **partially ordered set** (*poset* for short) is a set X together with a *partial order* $\leq$, that is a relation satisfying

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

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

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

If $S \subseteq X$ is a subset of a poset, then an *upper bound* for S is an element $u \in X$ such that $s \leq u$ 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 $m \leq x$. In other words, for any $x \in X$, either $x = m$, or $x \leq m$, or x and m are not comparable with respect to the partial order $\leq$.

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

Lemma B.1 (Zorn's Lemma). *Let X be a nonempty poset such that every nonempty chain C in X has an upper bound in X . Then X has a maximal element.*

B.2. BILINEAR MAPS

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

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

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

Note that such β induces maps

$$\begin{aligned}\beta_U: U &\longrightarrow \text{Hom}(V, W), & u &\longmapsto (v \longmapsto \beta(u, v)) \\ \beta_V: V &\longrightarrow \text{Hom}(U, W), & v &\longmapsto (u \longmapsto \beta(u, v)).\end{aligned}$$

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

B.3. DUAL VECTOR SPACE

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

$$V^\vee = \text{Hom}(V, \mathbf{F}).$$

This is a vector space over $\mathbf{F}$, called the *dual vector space* to V . Its elements are sometimes called *(linear) functionals* and denoted with Greek letters such as φ .

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

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

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

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

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

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

then

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

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

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

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

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

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

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

$$\lambda_i = 0.$$

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

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

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

Then, since φ is a linear transformation, we have

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

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

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

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

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

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

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

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

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

Proof. Let $B = \{v_1, \dots, v_n\}$ be a basis for V and let $B^\vee = \{v_1^\vee, \dots, v_n^\vee\}$ be the dual basis for $V^\vee$ as in [Proposition B.2](#).

To show that β_V is injective, suppose $u, v \in V$ are such that $\beta_V(u) = \beta_V(v)$, in other words

$$\varphi(u) = \varphi(v) \quad \text{for all } \varphi \in V^\vee.$$

Write

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

then, for $i = 1, \dots, n$, we have

$$a_i = v_i^\vee(u) = v_i^\vee(v) = b_i$$

Therefore $u = v$.

We now prove that β_V is surjective. (Note that we could get away with simply saying that [Proposition B.2](#) tells us that V and $V^\vee$, and therefore also $(V^\vee)^\vee$, have the same dimension n ; so β_V , being injective, is also surjective.)

Let $T: V^\vee \rightarrow \mathbf{F}$ be a linear transformation. Define $v \in V$ by

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

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

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

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

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

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

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

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

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

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

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

$$T^\vee(\lambda\varphi)(v) = (\lambda\varphi)(T(v)) = \lambda\varphi(T(v)) = \lambda T^\vee(\varphi)(v). \quad \square$$

B.4. A DIVERSION: TOPOLOGICAL GROUPS

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

$$G \times G \rightarrow G, \quad (g, h) \mapsto gh$$

and the inverse map

$$G \rightarrow G, \quad g \mapsto g^{-1}$$

are both continuous.

Obviously, this makes the inverse map into a homeomorphism.

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

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

Example B.6. Consider $\mathbf{R}$ with the Euclidean topology, under the addition operation on $\mathbf{R}$.

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

Example B.7 (The circle group). Let

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

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

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

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

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

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

Proposition B.9. Let G be a topological group and $g \in G$. The *left translation* map $L_g : G \rightarrow G$ given by $L_g(x) = gx$ is a homeomorphism. So is the *right translation* map R_g .

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

Corollary B.10. Any topological group G is a *homogeneous topological space*, that is: for every $x, y \in G$ there exists a homeomorphism $f : G \rightarrow G$ such that $f(x) = y$.

Proof. Let $f = L_{yx^{-1}}$. $\square$

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

Example B.11. We know that the inverse map $G \rightarrow G$, $g \mapsto g^{-1}$ is continuous (in fact, a homeomorphism). But it is a group homomorphism (and hence a topological group homomorphism) if and only if G is abelian.

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

Example B.12. The map $\exp : \mathbf{R} \rightarrow \mathbf{R}^\times$ is a topological group homomorphism, where $\mathbf{R}$ has the Euclidean topology and the addition operation, and $\mathbf{R}^\times$ has the subspace

topology and the multiplication operation.

Example B.13. The determinant map $\det: \text{GL}_n(\mathbf{R}) \rightarrow \mathbf{R}^\times$ is a topological group homomorphism.

Proposition B.14. *Let G be a topological group and H a subgroup. Then the closure $\overline{H}$ is a subgroup of G . Moreover, if H is normal, then so is $\overline{H}$.*

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

In the rest of the proof, we will repeatedly use [Proposition 2.20](#): if $A \subseteq X$, then $x \in \overline{A}$ if and only if every open neighbourhood of x intersects A nontrivially.

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

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

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

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

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

BIBLIOGRAPHY

INDEX

- absolute value
 - complex, 7
 - real, 7
- algebra, 67
 - associative, 67
 - unital, 67
- Banach space, 48
- basis, 5
 - Schauder, 51
- bilinear map, 57, 73
- boundary, 15
- bounded bilinear map, 58
- cardinality, 66
- Cauchy sequence, 25
- Cauchy–Schwarz Inequality, 68
- chain, 73
- closed ball, 11
- closed set, 12, 15
- closure, 15
- complete metric space, 25
- completion, 26
- conjugate-linear, 68
- conjugation, 76
- continuous, 13
- contraction, 30
- convergence
 - for a sequence, 23
- convex function, 55
- convex subset, 54
- dense, 16
- diameter, 32
- distance, 9
- dual basis, 74
- dual linear transformation, 75
- dual vector space, 73
- equivalence class, 65
- equivalence of sequences, 24
- equivalence relation, 65
- extend by linearity, 67
- extension, 28
- fixed point, 31
- function
 - bounded, 35
 - convex, 55
- functional, 73
- Hölder conjugate, 57
- Hausdorff
 - topological space, 15
- Hilbert space, 48
- homeomorphic, 13
- homeomorphism, 13
- inequality
 - Cauchy–Schwarz, 68
 - Hölder, 58
 - Minkowski, 56
- inner product, 47, 68
- interior, 15
- isometry, 24
- Lebesgue number, 35
- lemma
 - Zorn, 73
- limit
 - of a sequence, 23
- linear combination, 5
- linear transformation
 - bounded, 51
 - Lipschitz, 51
- Lipschitz, 51
- map
 - closed, 22
 - proper, 22
- matrix
 - Hermitian, 68
 - positive-definite, 68
 - positive-semidefinite, 68
- maximal element, 73
- metric, 9
 - ℓ^1 , 10
 - ℓ^∞ , 10
 - coarse, 17

- equivalent, 17
- fine, 17
- Hamming, 10
- induced, 10
- Manhattan, 10
- sup norm, 10
- taxicab, 10
- uniform, 35
- uniform norm, 10
- metric space, 9
 - complete, 25
- metric spaces
 - isometric, 24
- neighbourhood, 12
 - open, 12
- norm, 43
 - L^∞ , 48
 - equivalence, 44
 - uniform, 48
- normed space, 43
- nowhere dense, 16
- open ball, 11
- open cover, 20
- open rectangle, 14
- open set, 11, 12
- orthogonal complement, 61
- orthogonal projection, 61
- parallelogram law, 47
- partial sum, 49
- pointwise convergence, 36
- pointwise limit, 36
- poset, 73
- positive-definite form, 68
- projection, 60
 - orthogonal, 61
- quotient by an equivalence relation, 66
- reflexivity, 65
- Schauder basis, 51
- separates points, 39
- separation property, 15
- sequence, 23
 - p -summable, 55
 - bounded, 55
 - Cauchy, 25
 - equivalent, 24
 - square-summable, 55
 - summable, 55
- sequential compactness, 34
- series, 49
 - absolutely convergent, 49
 - convergent, 49
- sesquilinear form, 68
- set
 - bounded, 32
 - closed, 15
 - countable, 67
 - countably infinite, 67
 - finite, 67
 - infinite, 67
 - open, 11
 - partially ordered, 73
 - totally bounded, 33
 - totally ordered, 73
 - unbounded, 32
 - uncountable, 67
- space
 - Banach, 48
 - compact, 20
 - connected, 18
 - disconnected, 17
 - Hilbert, 48
 - homogeneous, 76
 - inner product, 68
 - metric, 9
 - normed, 43
 - of sequences, 55
 - totally disconnected, 20
- sphere, 11
- subcover, 20
- subset
 - compact, 20
 - connected, 18
- supremum, 19
- symmetry, 65
- theorem
 - Baire Category, 40
 - Banach Fixed Point, 31
 - Heine–Borel, 34
 - Intermediate Value, 20
 - Mean Value, 31
 - Schröder–Bernstein, 67
 - Weierstrass Approximation, 37
- topological group, 75
- topological space, 12
 - metrisable, 16
- topological vector space, 45

topology, [12](#)

coarse, [13](#)

cofinite, [13](#)

discrete, [12](#)

fine, [13](#)

generated by a set, [13](#)

induced, [14](#)

metric, [16](#)

metrisable, [16](#)

product, [14](#)

subspace, [14](#)

trivial, [12](#)

total order, [73](#)

transitivity, [65](#)

translation

left, [76](#)

right, [76](#)

uniform convergence, [36](#)

uniform limit, [36](#)

uniform norm, [48](#)

uniformly continuous, [25](#)

upper bound, [73](#)

Zorn's Lemma, [73](#)