Metric spaces

Lecture notes for MTH224 2025–2026

Paul-Henry Leemann

April 14, 2026

FOREWORD

These notes are based on Alastair Darby's 2024–2025 notes and are a work in progress. They are still plenty of typos (and hopefully few mistakes). If you see some, please let me know by email or directly via learning mall and I will correct them.

Things might be presented in a slightly different way than what was done in class, or in a different order.

Theorem 0.1.

Blue boxes are for theorems.

Definition 0.2.

Green boxes are for definitions.

Remark 0.3.

Orange boxes are for remarks.
Some passages are designed to forewarn the readers against serious errors, where they risk falling; these passages are indicated in the margin with the "dangerous bend" sign.

To go further

In these notes, you will see a few black boxes like this one. These are meant for the interested reader that wants to know more and can be skipped without harm. They are not necessarily be meant to be read in a first reading, but are here to provide more details if you come back to read these notes in the future.
The content that is written inside these "To go further" boxes will **NOT** be in the exam.

Alastair Darby, Gaoming Zhang, Qinen Song, Yang Fu and Hang Yin contributed to reduce the number of typos. These notes would not have been the same without their comments.

CONTENTS

SYMBOLS

$\mathbf{N}$	natural integers excluding 0: $\mathbf{N} = \{1, 2, 3, \dots\}$
$\mathbf{Z}$	relative integers: $\mathbf{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$
$\mathbf{Q}$	rational numbers: $\mathbf{Q} = \left\{ \frac{p}{q} \,\middle|\, p, q \in \mathbf{Z}, q \neq 0 \right\}$
$\mathbf{R}$	real numbers
$\mathbf{C}$	complex numbers: $\mathbf{C} = \{a + b\mathrm{i} \mid a, b \in \mathbf{R}, \mathrm{i}^2 = -1\}$
$\mathbf{F}$	field (for example: $\mathbf{Q}$, $\mathbf{R}$ or $\mathbf{C}$)

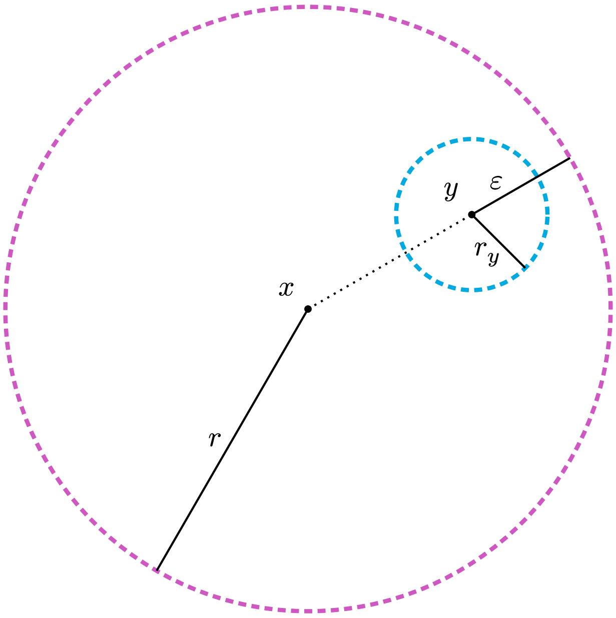
Table 1.: Important sets.

α, A	alpha	ν, N	nu
β, B	beta	$\xi, \varXi$	xi
$\gamma, \varGamma$	gamma	o, O	omicron
$\delta, \varDelta$	delta	$\pi, \varPi$	pi
ε, E	epsilon	ρ, P	rho
ζ, Z	zeta	$\sigma, \varSigma$	sigma
η, H	eta	τ, T	tau
$\theta, \varTheta$	theta	υ, Υ	upsilon
ι, I	iota	$\varphi, \varPhi$	phi
κ, K	kappa	χ, X	chi
$\lambda, \varLambda$	lambda	$\psi, \varPsi$	psi
μ, M	mu	$\omega, \varOmega$	omega

Table 2.: Greek letters.

CHAPTER 1.

CONTINUITY



1.1. Definitions & Examples

As the name of the module hints, we will study metric spaces. These are sets that are enriched with a *distance* that measure how far apart any two points lie. We hence need to define what a distance is. We want it to generalise the natural Euclidean distance in $\mathbf{R}^n$ (which is the usual length of the segment between two points), but also other "natural" notions of distance.

Before introducing an axiomatic definition of a distance, let us go through some real word notions of distance between two points, let us say between cities in China. For two cities A and B, one can measure a "distance" between these two points as the length of the shortest path from A to B, the length of the shortest road from A to B, but also as the minimum price of a train ticket from A to B. All these different notions of distance satisfies some natural commun properties that we formalise in the next definition.

Definition 1.1.1 (Metric Space)**.**

A **metric space** is a pair (X, d) consisting of a non-empty set X and a function, called **distance** or **metric**,

$$d\colon X \times X \to \mathbf{R}$$

that, for all $x, y, z \in X$, satisfies:

(M0) $d(x,y) \geqslant 0$;

(M1) $d(x,x) = 0$;

(M2) $d(x,y) = 0 \implies x = y$; *(positivity)*

(M3) $d(x,y) = d(y,x)$; *(symmetry)*

(M4) $d(x,z) \leqslant d(x,y) + d(y,z)$. *(triangle inequality)*

Axiom (M2) is called *positivity* because it is equivalent to: $x \neq y \implies d(x,y) > 0$. Axiom (M4) is called *triangle inequality* because it says that in a "triangle", the length of any side is smaller than the sum of the lengths of the other sides.

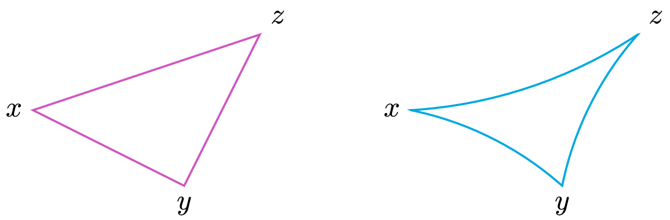


Figure 1.1.: Triangles in the Euclidean plane (on the left) and in the hyperbolic plane (on the right). In both cases the length of a side is smaller than the sum of the lengths of the two other sides, one hence have $d(x,z) \leqslant d(x,y) + d(y,z)$.

Observe that symmetry means that "the shortest time taken to go from A to B" is not a good notion of distance.

We will see in Tutorial 1, Question 1 that the axiom (M0) is redundant: any function $d\colon X \times X \to \mathbf{R}$ satisfying (M1)–(M4) automatically satisfies (M0). However, axiom (M0) is the only superfluous axiom: the axioms (M1)–(M4) are independent. This means that for any $i \in \{1, 2, 3, 4\}$ it is possible to find a set X and a function $d\colon X \times X \to \mathbf{R}$ not that does not satisfy axiom (Mi) but satisfies all the (Mj) for $j \in \{1, 2, 3, 4\} \setminus \{i\}$. Can you find such functions?

Let us start exploring the notion of metric space with a few classical examples you already have encountered in your studies.

Example 1.1.2 (The Real Line)**.** On $\mathbf{R}$ we can define $d_{\mathbf{R}}(x, y) := |x - y|$ and we know from Analysis 1 that this satisfies the axioms (M0)–(M4) of being a metric. So $(\mathbf{R}, d)$ forms a metric space. When we consider $\mathbf{R}$ as a metric space the metric $d_{\mathbf{R}}$ will be understood unless otherwise stated.

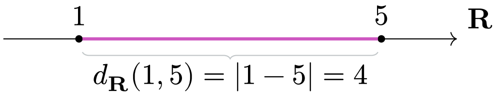


Figure 1.2.: Distance between two points of the real line. The purple line is the unique shortest path between the two points.

Example 1.1.3 (Euclidean n-space)**.** We generalise Example 1.1.2 to n-dimensional **Euclidean space**[1] $\mathbf{R}^n$. For $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ in $\mathbf{R}^n$ we set

$$d_2(x, y) := \left((x_1 - y_1)^2 + \dots + (x_n - y_n)^2\right)^{\frac{1}{2}} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}, \tag{1.1}$$

where we take the positive square root. It is easy to check that the axioms (M0)–(M3) of Definition 1.1.1 are satisfied. We will check (M4), which follows from *Cauchy's*[2] *Inequality*, on Tutorial 1, Question 2.

When we consider $\mathbf{R}^n$ as a metric space the **Euclidean metric** (as defined in equation (1.1)) will be understood unless otherwise stated.

Remark 1.1.4.

For $n = 1$ we have the real line $\mathbf{R}$ where the metric

$$d_2(x, y) = \sqrt{(x - y)^2} = |x - y|$$

is the usual notion of *length* as in Example 1.1.2. For the cases $n = 2$ and 3 we recover the usual 2 and 3-dimensional Euclidean spaces with their normal distance functions.

---

[1] Named after Εὐκλείδης (fl. 300 BC).

[2] Augustin-Louis Cauchy (1789–1857).

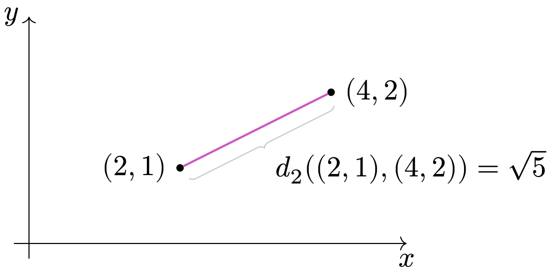


Figure 1.3.: Euclidean distance between two points in the plane. The purple line is the unique shortest path between the two points.

It is possible to have different metrics on the same underlying set as we will see below with $\mathbf{R}^n$.

Example 1.1.5 (Taxicab Metric)**.** On $\mathbf{R}^n$, we define the **taxicab metric** (sometimes called the **Manhattan metric**) by

$$d_1(x,y) := |x_1 - y_1| + \cdots + |x_n - y_n| = \sum_{i=1}^{n} |x_i - y_i|.$$

It is easy to check that (M0)–(M3) hold. For (M4) we write $r_i = x_i - y_i$ and $s_i = y_i - z_i$, where $1 \leqslant i \leqslant n$. So we need to prove that

$$\sum_{i=1}^{n} |r_i + s_i| \leqslant \sum_{i=1}^{n} |r_i| + \sum_{i=1}^{n} |s_i|,$$

which is true since $|r_i + s_i| \leqslant |r_i| + |s_i|$, for all $1 \leqslant i \leqslant n$.

Note that on $\mathbf{R}$ the Euclidean and taxicab metrics are the same.

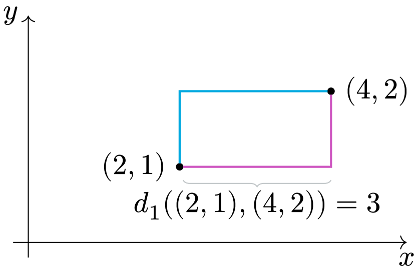


Figure 1.4.: Taxicab distance between two points in the plane. The purple and blue lines are two shortest paths between the two points.

While the taxicab metric is not the usual Euclidean metric, it is not entirely new to you as it comes from a norm on the vector space $\mathbf{R}^n$, see MTH107 Advanced Linear Algebra and Section 1.2 for norms on vector space.

Definition 1.1.6 ($(\mathbf{R}^n, d_p)$).

The metrics d_2 and d_1 on $\mathbf{R}^n$ from Examples 1.1.3 and 1.1.5 can be generalised, which explains the choice of subscripts, as follows. Let $p \geqslant 1$ and define

$$d_p(x,y) := \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{\frac{1}{p}},$$

for $x, y \in \mathbf{R}^n$ and where we take the non-negative real p-th root. The triangle inequality can be proved for d_p in general by using a generalisation of Cauchy's Inequality called *Minkowski's Inequality.*

Definition 1.1.7 ($(\mathbf{R}^n, d_\infty)$).

We can define another metric on $\mathbf{R}^n$:

$$d_\infty(x,y) := \max\{|x_1 - y_1|, \dots, |x_n - y_n|\}.$$

We leave the checking of the axioms as an exercise (Tutorial 1, Question 3).

Let $x = (x_1, x_2)$ and $y = (y_1, y_2)$ be two points in $\mathbf{R}^2$. If $p_1 < p_2 < \dots$ is an unbounded increasing sequence of real numbers, one can show that $d_{p_i}(x,y)$ converges to $\max(|x_1 - y_1|, x_2 - y_2) = d_\infty(x,y)$. The same is true for pair of points in $\mathbf{R}^n$. The metric d_∞ is therefore, in some sense, the limit of the metrics d_p for $p \to \infty$. Hence the name.

We have seen that a given set X might admit more than one metric. The following shows that any non-empty set can be turned into a metric space.

Definition 1.1.8 (Discrete Metric Spaces).

Let X be a non-empty set and define

$$d_0(x,y) := \begin{cases} 1, & x \neq y; \\ 0, & x = y. \end{cases}$$

Clearly the axioms (M0)–(M3) are satisfied. For (M4) observe that $d_0(x,y) + d_0(y,z)$ is either 0, 1, or 2, and is only 0 when $x = y = z$ in which case $d_0(x,z) = 0$ also. The metric d_0 is called the **discrete metric**, and (X, d_0) a **discrete metric space**.

Discrete metric spaces often serve as counterexamples to ideas, suggested by geometric intuition, that may not actually hold in general metric spaces.

Before going further, let us compare the examples of metrics we have seen so far.

Example 1.1.9. Taking points $P = (2,1)$ and $Q = (4,2)$ in $\mathbf{R}^2$ we calculate their distances according to some of the different metrics that we have looked at on $\mathbf{R}^2$:

$$\begin{aligned} d_0(P,Q) &= 1, \quad \text{since } P \neq Q \\ d_1(P,Q) &= |2-4| + |1-2| = 3 \\ d_2(P,Q) &= \sqrt{(2-4)^2 + (1-2)^2} = \sqrt{5} \approx 2.236 \\ d_5(P,Q) &= \sqrt[5]{(2-4)^5 + (1-2)^5} = \sqrt[5]{33} \approx 2.012 \\ d_\infty(P,Q) &= \max\{|2-4|, |1-2|\} = 2 \end{aligned}$$

where we write d_0 for the discrete metric on $\mathbf{R}^2$. This shows that all the above metrics are distinct.

Remark 1.1.10.

Let d be a metric on a set X. It directly follows from the definition that for any positive real number c, scaling the metric by c (that is looking at the function $(x,y) \mapsto c \cdot d(x,y)$) gives a new metric.

The reader can verify that the metrics d_0, d_1, d_2 and d_∞ of the above example are not scalar multiples one of each other. The same is true for the metrics d_p and d_q, $p \neq q$, from Definition 1.1.6.

Examples 1.1.2, 1.1.3 and 1.1.5 and Definitions 1.1.6 and 1.1.7 can easily be adapted to the complex case. In particular, we recover the standard metric on $\mathbf{C}$.

Example 1.1.11. The standard metric on the complex numbers $\mathbf{C}$ is given by

$$d_{\mathbf{C}}(z_1, z_2) := |z_1 - z_2|,$$

where $|z|$ is the **modulus** of the complex number z (that is, $|z| = r$, where $z = re^{i\theta}$). The axioms of Definition (1.1.1) hold by analogy with those for the Euclidean metric on $\mathbf{R}$.

We know (from MTH107 Advanced Linear Algebra) that a vector subspace is a subset that is itself a vector space for the induced operations. A similar definition holds for subgroups (see MTH122 Introduction to Abstract Algebra). We will mimic this for metric subspace.

Definition 1.1.12 (Metric Subspaces).

Let (X,d) be a metric space. A **metric subspace** is a subset $W \subseteq X$ endowed with the restriction of d to W:

$$d|_W : W \times W \to \mathbf{R},$$

i.e. $d|_W(w_1, w_2) := d(w_1, w_2)$.

Observe that we made a slight abuse of notation when writing $d|_W$ to describe the restriction $d|_{W\times W}$ of d to $W \times W \subseteq X \times X$.

As the name hints, a metric subspace $(W, d|_W)$ of a metric space (X, d) is itself a metric space. That is, for any subset $W \subseteq X$ of a metric space (X, d), the function $d|_W$ is automatically a metric on W.

When studying mathematical objects (vector spaces, groups, ...) we are not only interested in those objects, but also in "good" maps between them. We start by defining what it means for two metric spaces to be "identical as metric spaces".

Definition 1.1.13 (Isometry).

An **isometry**[a] between two metric spaces (X, d_X) and (Y, d_Y) is a bijective function $f\colon X \to Y$ such that

$$d_Y(f(x), f(y)) = d_X(x, y), \quad \forall x, y \in X.$$

Two metric spaces are said to be **isometric** if there exists an isometry between them.

---

[a] From the Greek ἴσος μέτρον, meaning equal measure.

Remark 1.1.14.

If $f\colon X \to Y$ is an isometry, then f^{-1} is also an isometry. So, an isometry is a bijective function such that both f and f^{-1} "preserve" the metric space structure. This is similar to the definition of isomorphisms of vector spaces or of group isomorphisms.

It is easily shown that isometry is an equivalence relation (meaning it is reflexive, symmetric and transitive) on metric spaces; see Appendix B for a reminder on equivalence relations. The proof is left as an exercise (Tutorial 3, Question 9). Two metric spaces are isometric if and only if they are indistinguishable as metric spaces. They can however be quite different for some other properties.

Example 1.1.15. Let $f\colon \mathbf{R}^2 \to \mathbf{C}$ be given by $f(x, y) = x + iy$. Then f is an isometry from $\mathbf{R}^2$ with the Euclidean metric to $\mathbf{C}$ with its standard metric (see Examples 1.1.3 and 1.1.11). This means that $\mathbf{R}^2$ with the Euclidean metric and $\mathbf{C}$ with the standard metric are indistinguishable as metric spaces, despite being quite different as rings. For example, in $\mathbf{C}$ every non-zero elements has a multiplicative inverse, but this is not true in $\mathbf{R}^2$.

For algebraic structures (like vector spaces or groups), the homomorphisms are the functions preserving the structure. For metric spaces, asking to preserve the metric is usually too much. Indeed, if $f\colon X \to Y$ is a function such that $d_Y(f(x), f(y)) = d_X(x, y)$ for all $x, y \in X$, then it is automatically injective. In particular, such a function is always an isometry between X and $f(X)$. We will therefore for now only ask our functions to behave well for the notion of "proximity". But see Definition 1.7.1 for another class of interesting functions.

Definition 1.1.16 (Continuous Function)**.**

Given two metric spaces (X, d_X) and (Y, d_Y) and a point $a \in X$ we say that a function $f\colon X \to Y$ is **continuous at** a (with respect to the metrics d_X and d_Y) if

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ such that } d_X(x, a) < \delta \implies d_Y(f(x), f(a)) < \varepsilon.$$

If we need to be precise, we will say that f is (d_X, d_Y)-continuous at $a \in X$.
A function $f\colon X \to Y$ between two metric spaces is **continuous** if it is continuous at a for every $a \in X$.

Remark 1.1.17.

In Analysis 1 we were concerned with functions $f\colon A \to \mathbf{R}$, where $A \subseteq \mathbf{R}$. Note that Definition 1.1.16 coincides with that given in Analysis 1 where we consider $\mathbf{R}$ with the metric $d_{\mathbf{R}}$ defined in Example 1.1.2 and consider $(A, d_{\mathbf{R}}|_A)$ as a metric subspace of $(\mathbf{R}, d_{\mathbf{R}})$.

Remark 1.1.18.

A function $f\colon (X, d_X) \to (Y, d_Y)$ is continuous if

$$\forall x, \forall \varepsilon > 0, \exists \delta > 0, \forall y : d_X(x, a) < \delta \implies d_Y(f(x), f(a)) < \varepsilon.$$

This means that δ might depends not only on ε, but also on x. This happens for example for $f\colon \mathbf{R} \to \mathbf{R}, x \mapsto x^2$.

Example 1.1.19. The function $f\colon \mathbf{R} \to \mathbf{R}, x \mapsto x^2$ is continuous. Indeed, for every $a \in \mathbf{R}$ and $\varepsilon > 0$, there exists $\delta > 0$ such that $2a\delta + \delta^2 \leqslant \varepsilon$. But then if $|x - a| < \delta$ then $|x + a| < 2a + \delta$ and thus $|x^2 - a^2| = |x - a| \cdot |x + a| < 2a\delta + \delta^2 \leqslant \varepsilon$.

Continuous functions possesses some important properties that make them suitable to use as “nice functions” between metric spaces.

Proposition 1.1.20. *Composition of continuous functions is still continuous. More precisely, if* $f\colon (X, d_X) \to (Y, d_Y)$ *and* $g\colon (Y, d_Y) \to (Z, d_Z)$ *are two continuous functions, then the composition* $g \circ f\colon (X, d_X) \to (Z, d_Z)$ *is also continuous.*

Composition of continuous functions has identities. More precisely, the identity function $\mathrm{Id}_X\colon (X, d) \to (X, d)$ *is continuous.*

Composition of continuous functions is associative. That is, if $f\colon (X, d_X) \to (Y, d_Y)$, $g\colon (Y, d_Y) \to (Z, d_Z)$ *and* $h\colon (Z, d_Z) \to (W, d_W)$ *are continuous functions, then* $h \circ (g \circ f) = (h \circ g) \circ f$.

Proof. For the first statement, use the continuity of f at x and of g at $f(x)$. See Tutorial 1, Question 8 for the details.

The second statement is trivial, see also Tutorial 1, Question 7.

The last statement is true for any functions between sets. □

To go further

The above proposition says that metric spaces with continuous functions form a **category**. Category theory is the branch of mathematics that study "objects and nice maps between them" from an abstract viewpoint. Other examples of categories are: vector spaces with linear maps and groups with group homomorphisms.

1.2. Metrics from Norms

In MTH107 Advanced Linear Algebra we studied inner product on vector spaces. We saw that each inner product $\langle\,\cdot\,,\cdot\,\rangle$ gives raise to a norm $\|\cdot\|$ satisfying some nice properties. We will take these properties as the axiomatic definition of a norm.

Definition 1.2.1.

A **norm** on a vector space V (over a field $\mathbf{F} \subseteq \mathbf{C}$) is a function

$$\|\cdot\| \colon V \to \mathbf{R}$$

that, for all $x, y \in V$ and all $\lambda \in \mathbf{F}$, satisfies

(N0) $\|x\| \geqslant 0$;

(N1) $\|0\| = 0$;

(N2) $\|x\| = 0 \implies x = 0$; *(positivity)*

(N3) $\|\lambda x\| = |\lambda| \|x\|$; *(homogeneity)*

(N4) $\|x + y\| \leqslant \|x\| + \|y\|$. *(triangle inequality)*

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

As for the definition of distance, axiom (N0) is superfluous: if $\|\cdot\|$ satisfies (N1) to (N4), it automatically satisfies (N0). Indeed, for every $x \in V$ we have $0 = \|0\| = \|x - x\| \leqslant \|x\| + |-1|\|x\| = 2\|x\|$.

The following elementary examples are inspired by the metrics defined in Section 1.1.

Example 1.2.2 (Norms on $\mathbf{R}^n$ or $\mathbf{C}^n$). The following are norms on $\mathbf{R}^n$ and $\mathbf{C}^n$:

1. Taxicab norm: $\|x\|_1 := \sum_{i=1}^n |x_i|$;
2. Euclidean norm: $\|x\|_2 := \sqrt{(\sum_{i=1}^n |x_i|^2};$
3. L_p norm for $p \geqslant 1$: $\|x\|_p := \sqrt[p]{(\sum_{i=1}^n |x_i|^p};$
4. Max norm: $\|x\|_\infty := \max\{|x_1|, \dots, |x_n|\}$.

Remark 1.2.3.

In the above examples, the only norm that comes from an inner product is $\|\cdot\|_2$. Indeed, none of the other norm satisfies the parallelogram identity $2\|x\|^2 + 2\|y\|^2 = \|x-y\|^2 + \|x+y\|^2$.

The following result explains why we are interested in norms in this module.

Lemma 1.2.4. *If $(V, \|\cdot\|)$ is a normed vector space, then $d(x,y) := \|x-y\|$ is a metric on V.*

Proof. One check that for each $i \in \{1, \dots, 4\}$ the axiom (Ni) implies the corresponding axiom (Mi). Only the triangle inequality needs an argument:

$$\begin{aligned} d(x,z) &= \|x-z\| = \|(x-y)+(y-z)\| \\ &\leqslant \|x-y\| + \|y-z\| = d(x,y) + d(y,z). \end{aligned}$$ ☐

Norms from Example 1.2.2 generate the corresponding metrics from Section 1.1. Observe that the discrete metric on $\mathbf{R}$ does not come from a norm as this would contradict (N3).

1.3. Function Spaces

In this section we will define various metrics on spaces of functions. The set $\mathbf{R}^{[a,b]}$ of all real functions $f\colon [a,b] \to \mathbf{R}$ is in some sense too big to put a nice metric on it. We will therefore restrict ourself to subspaces of $\mathbf{R}^{[a,b]}$.

Recall that for a subset W of $\mathbf{R}$, a function $f\colon W \to \mathbf{R}$ is **bounded** if there exists $K \in \mathbf{R}$ such that $|f(x)| \leqslant K$, for all $x \in W$.

Example 1.3.1 (Function Spaces)**.** Let X be the set of all bounded functions $f\colon W \to \mathbf{R}$. Given two points $f, g \in X$ we define

$$d_{\sup}(f,g) := \sup_{x \in W} |f(x) - g(x)|.$$

This is well-defined (that is, the right-hand side exists) since, as f and g are bounded we can write

$$|f(x)| \leqslant K \quad \text{and} \quad |g(x)| \leqslant L$$

for some $K, L \in \mathbf{R}$ say, then

$$|f(x) - g(x)| \leqslant |f(x)| + |g(x)| \leqslant K + L, \quad \forall x \in W,$$

so the supremum exists. We will now check that $d_{\sup}$ is a metric on X:

“(M0)” Is clear since it is the supremum of a set of non-negative numbers.

“(M1) and (M2)” The supremum $\sup_{x \in W} |f(x) - g(x)|$ is 0 if and only if $f(x) = g(x)$ for all $x \in W$, that is if and only if $f = g$ as functions from W to $\mathbf{R}$.

“(M3)” Since $|f(x) - g(x)| = |g(x) - f(x)|$, for all $x \in W$, it follows that $d_{\sup}(f,g) = d_{\sup}(g,f)$.

“(M4)” Let $f, g, h \in X$. For any $x \in [a, b]$,

$$\begin{aligned} |f(x) - h(x)| &\leqslant |f(x) - g(x)| + |g(x) - h(x)| \\ &\leqslant \sup_{x\in W} |f(x) - g(x)| + \sup_{x\in W} |g(x) - h(x)| \\ &= d_{\sup}(f, g) + d_{\sup}(g, h). \end{aligned}$$

So $d_{\sup}(f, g) + d_{\sup}(g, h)$ is an upper bound for the set $S = \{|f(x) - g(x)| \mid x \in [a, b]\}$. Hence, $\sup S = d_{\sup}(f, h) \leqslant d_{\sup}(f, g) + d_{\sup}(g, h)$.

This metric is usually called the **sup metric** and we usually write $\mathcal{B}(W) := (X, d_{\sup})$ for this metric space.

If $f : [a, b] \to \mathbf{R}$ is continuous, then we know from Analysis 1 that f is bounded. Let Y denote the set of all continuous functions $f : [a, b] \to \mathbf{R}$. Then $\mathscr{C}[a, b] := (Y, d_{\sup})$ forms a metric subspace of $\mathcal{B}[a, b]$.

Example 1.3.2. The polynomial functions $f(x) = x^4$ and $g(x) = 1$ are continuous on $[0, 1]$ and therefore $f, g \in \mathscr{C}[0, 1]$. We find that

$$d_{\sup}(f, g) = d_{\sup}(x^4, 1) = \sup_{x\in[0,1]} |x^4 - 1| = |0^4 - 1| = 1.$$

Therefore, $d_{\sup}(x^4, 1) = 1$ in $\mathscr{C}[0, 1]$ and $\mathcal{B}[0, 1]$.

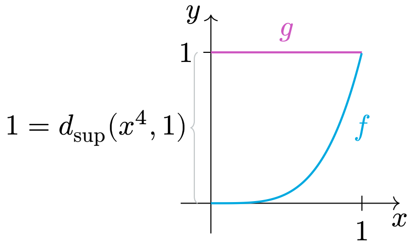


Figure 1.5.: Graphs of the polynomial functions $f(x) = x^4$ and $g(x) = 1$. The $d_{\sup}$ distance between the functions is the supremum of the vertical distance between their graphs.

On the set of all continuous functions $f : [a, b] \to \mathbf{R}$ we can define

$$d_1(f, g) := \int_a^b |f(x) - g(x)| \,\mathrm{d}x. \tag{1.2}$$

Notice that $|f(x) - g(x)| \geqslant 0$, for all $x \in [a, b]$, so

$$\int_a^b |f(x) - g(x)| \,\mathrm{d}x = 0 \iff \forall x \in [a, b] : f(x) = g(x) \tag{1.3}$$

Lemma 1.3.3. *The function d_1 from (1.2) is a metric on the set Y of all continuous functions $f\colon [a,b] \to \mathbf{R}$.*

Proof. The axiom (M0) clearly holds and for (M2) we know by (1.3) that $f = g$ exactly when $d_1(f,g) = 0$. Clearly (M3) holds.

For (M4), take continuous functions $f, g, h\colon [a,b] \to \mathbf{R}$ and $x \in [a,b]$. Then

$$|f(x) - h(x)| \leqslant |f(x) - g(x)| + |g(x) - h(x)|$$

so

$$\int_a^b |f(x) - h(x)| \, \mathrm{d}x \leqslant \int_a^b |f(x) - g(x)| \, \mathrm{d}x + \int_a^b |g(x) - h(x)| \, \mathrm{d}x. \qquad \square$$

This metric is known as the L_1 **metric** and we denote the space of all continuous functions $f\colon [a,b] \to \mathbf{R}$ with the L_1 metric by $\mathcal{L}_1[a,b] := (Y, d_1)$.

Example 1.3.4. The polynomials x^4 and 1 both lie in $\mathcal{L}_1[0,1]$ and now

$$d_1(x^4, 1) = \int_0^1 |x^4 - 1| \, \mathrm{d}x = \left[x - \frac{x^5}{5}\right]_0^1 = \frac{4}{5}.$$

Therefore, $d_1(x^4, 1) = \frac{4}{5}$ in $\mathcal{L}_1[0,1]$.

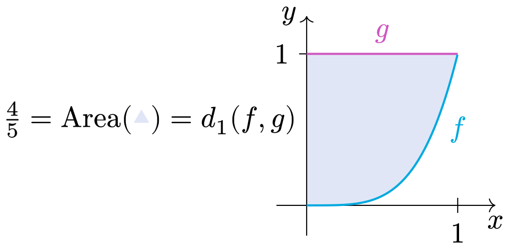


Figure 1.6.: Graphs of the polynomial functions $f(x) = x^4$ and $g(x) = 1$. The d_1 distance between the functions is the area of the surface between their graphs.

There are other metrics on the set of all continuous functions $f\colon [a,b] \to \mathbf{R}$, such as

$$d_p(f,g) := \left(\int_a^b |f(x) - g(x)|^p \, \mathrm{d}x\right)^{\frac{1}{p}}$$

where $p \geqslant 1$, which is known as the L_p **metric**, which results in a metric space $\mathcal{L}_p[a,b] := (Y, d_p)$.

All the metrics from this section come from norms. Let us see that in details for d_{sup} and d_1.

Example 1.3.5. The set $X = \{f\colon [a,b] \to \mathbf{R} \mid f \text{ is bounded}\}$ forms a real vector space by defining

$$(f+g)(x) := f(x) + g(x) \quad \text{and} \quad (\lambda f)(x) := \lambda f(x),$$

for all $f, g \in X$, $x \in [a,b]$ and $\lambda \in \mathbf{R}$.

We can then define a norm $\|\cdot\|_{\sup}$ on X by $\|f\|_{\sup} := \sup\{|f(x)| \mid x \in [a,b]\}$ which leads to the $d_{\sup}$ metric on X and the metric space $\mathscr{B}[a,b]$.

The set $Y = \{f\colon [a,b] \to \mathbf{R} \mid f \text{ is continuous}\}$ is a vector subspace of X and we can define a different norm $\|\cdot\|_1$ on Y defined by $\|f\|_1 := \int_a^b |f(x)|\,\mathrm{d}x$ which leads to the metric d_1 and the metric space $\mathscr{L}_1[a,b]$.

We have seen in Examples 1.3.2 and 1.3.4 that sometimes $d_1(f,g)$ is smaller than $d_{\sup}(f,g)$. This is not always the case as demonstrated in the following example.

Example 1.3.6. The polynomial functions $f(x) = 0$, $g(x) = 1$ and $h(x) = x$ all belong to $\mathscr{C}[0,2]$. Straightforward computations give

$$d_{\sup}(0,1) = 1 \leqslant 2 = d_1(0,1)$$

but

$$d_{\sup}(0,x) = 2 \geqslant 1 = d_1(0,x).$$

1.4. Sequence Spaces

Sequence spaces are discrete versions of the function spaces that we discussed in the previous section. Recall that an (infinite) sequence of real numbers is a function $a\colon \mathbf{N} \to \mathbf{R}$ where we usually write $a_n := a(n)$. The set of all these sequences $\mathbf{R}^{\mathbf{N}}$ forms a real vector space by defining

$$(a+b)_n := a_n + b_n \quad \text{and} \quad (\lambda a)_n := \lambda a_n,$$

for $a, b \in \mathbf{R}^{\mathbf{N}}$ and $\lambda \in \mathbf{R}$.

Definition 1.4.1.

A sequence a is **bounded** if there exists $K \in \mathbf{R}$ such that $|a_n| \leqslant K$, for all $n \in \mathbf{N}$. The space of bounded sequence is denoted by $\mathcal{X}$.

Proposition 1.4.2. *The function* $\|\cdot\|_\infty\colon \mathcal{X} \to \mathbf{R}$ *defined by* $\|a\|_\infty := \sup_{n\in\mathbf{N}} |a_n|$ *is a norm on* $\mathcal{X}$.

Proof. The proof is an easy verification and is left as an exercise (Tutorial 2, Question 5). ☐

Observe that, by Lemma 1.2.4, $\ell^\infty := (\mathcal{X}, d_\infty)$ forms a metric space, where $d_\infty(a,b) := \|a-b\|_\infty$.

Definition 1.4.3.

A sequence a is **absolutely convergent** if $\sum_{n\in\mathbf{N}}|a_n| < \infty$. The space of absolutely convergent sequences is denoted by $\mathcal{Y}$.

Proposition 1.4.4. *The function* $\|\cdot\|_1 \colon \mathcal{Y} \to \mathbf{R}$ *defined by* $\|a\|_1 := \sum_{n\in\mathbf{N}}|a_n|$ *is a norm on* $\mathcal{Y}$.

Proof. The proof is an easy verification and is left as an exercise (Tutorial 2, Question 6). ☐

We call the metric space that this forms $\ell^1 := (\mathcal{Y}, d_1)$, where $d_1(a,b) := \|a-b\|_1$.

Remark 1.4.5.

Although this is beyond the scope of this module we will note that, for all $0 < p < \infty$, we can consider the vector subspace of $\mathbf{R}^{\mathbf{N}}$ consisting of all sequences a where $\sum_{n\in\mathbf{N}}|a_n|^p < \infty$. For $p \geqslant 1$ we can define on this subspace a norm $\|\cdot\|_p$ given by $\|a\|_p := \left(\sum_{n\in N}|a_n|^p\right)^{\frac{1}{p}}$, which gives us the metric space ℓ^p. For $0 < p < 1$ there is no such norm, but we can still define a metric given by $d_p(a,b) := \sum_{n\in N}|a_n - b_n|^p$.

1.5. Open Sets

Having a metric, we can define open and closed subsets, which generalise the notion of open and closed subsets of $\mathbf{R}^n$. We start by defining the simplest open subsets, the open balls. We want them to generalise the simplest open subsets of $\mathbf{R}$: the open intervals.

In order to generalise open intervals, we would like to be able to characterise them in terms of distance. Indeed, the definition $(a,b) = \{x \in \mathbf{R} \mid a < x < b\}$ use the order on $\mathbf{R}$, and thus cannot be generalised to metric spaces. We have $(a,b) = \{x \in \mathbf{R} \mid d(x, a + \frac{b-a}{2}) < \frac{b-a}{2}\}$, is the set of points not too far away from the middle $a + \frac{b-a}{2}$ of the interval (a,b). This can easily be generalised to arbitrary metric space.

Definition 1.5.1 (Open Ball)**.**

Given a metric space (X,d), the **open ball** of radius $r > 0$ around a point $x \in X$ is the set

$$B_r(x) = B_r(x;d) := \{y \in X \mid d(x,y) < r\} \subseteq X.$$

For $n \leqslant 3$ we recover the classical notions of open intervals/disks/balls in the in the Euclidean space $\mathbf{R}^n$. Examples of open balls in $\mathbf{R}^2$ for the metrics d_1, d_2 and d_∞ are given in Tutorial 2, Question 8, while Tutorial 2, Question 10 shows what happens for the British Railway metric from Tutorial 1, Question 5.

Example 1.5.2 (Euclidean Metric)**.** In $\mathbf{R}$, $B_r(x) = (x-r, x+r)$. In $\mathbf{R}^2$, $B_r(x)$ is the interior of a disc of radius r centred around x. In $\mathbf{R}^3$, $B_r(x)$ is the interior of a solid ball centred on x.

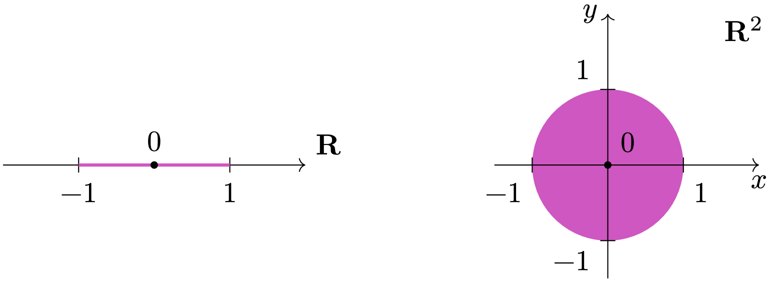


Figure 1.7.: Open balls of radius 1 around the origin for the Euclidean metric on $\mathbf{R}$ (on the left) and on $\mathbf{R}^2$.

As usual, it is enlightening to consider the case of discrete spaces.

Example 1.5.3 (Discrete Metric)**.** In a discrete metric space (X, d),

$$B_r(x) = \begin{cases} \{x\}, & \text{if } r \leqslant 1; \\ X, & \text{if } r > 1. \end{cases}$$

It is important to keep in mind that the definition of an open ball depends heavily on the ambiant space, as demonstrated by the following example.

Example 1.5.4. Let $I = [0,1] \subseteq \mathbf{R}$ with its standard metric $d_{\mathbf{R}}$ on $\mathbf{R}$. Then $B_1(1; d_{\mathbf{R}}) = (0,2)$ whilst $B_1(1; d_{\mathbf{R}}|_I) = (0,1]$, where $(I, d_{\mathbf{R}}|_I) \subseteq (\mathbf{R}, d_{\mathbf{R}})$ is a metric subspace.

More generally, if $(Y, d|_Y) \subseteq (X, d)$ is a subspace, then the open balls of Y are exactly the $B \cap Y$ for B an open ball of X. Indeed, we have

$$B_r(y; d|_Y) = \{z \in Y \mid d|_Y(z,y) < r\} = \{z \in Y \mid d(z,y) < r\} = B_r(y; d) \cap Y.$$

We can now express the continuity of a function using the language of open balls:

Definition 1.5.5 (Continuous Function)**.**

A function $f \colon X \to Y$ between metric spaces is **continuous at** $a \in X$ if

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ such that } f(B_\delta(a)) \subseteq B_\varepsilon(f(a)).$$

This is just a translation of Definition 1.1.16. Indeed, we have $d(x,a) < \delta \iff x \in B_r(a)$ and similarly for $f(x)$ and $f(a)$. While being equivalent to Definition 1.1.16, Definition 1.5.5 it is often more useful in practice.

Using open balls, one can define general open sets. To do so, we first need to define what is the interior of a set. Once again, we want this to generalise the notion of interior points in $\mathbf{R}$.

Definition 1.5.6 (Interior)**.**

Let W be a subset of a metric space (X, d). We say that $w \in X$ is an **interior point** of W if there exists $\varepsilon > 0$ such that $B_\varepsilon(w) \subseteq W$. The **interior of** W is the set W° of all interior points of W.

Observe that $W^\circ \subseteq W$. Interior points obviously generalise the natural notion of interior points of intervals of $\mathbf{R}$.

We can now define the main notion of this section.

Definition 1.5.7 (Open Set)**.**

A subset $U \subseteq (X, d)$ is **open in** X if every $u \in U$ is an interior point, that is if $U^\circ = U$. Equivalently, U is open if

$$\forall u \in U, \exists \varepsilon > 0 \text{ such that } B_\varepsilon(u) \subseteq U.$$

Remark 1.5.8.

Be aware that in the above definition ε depends on u and can take different values for different points in U. A simple example of this fact is given by the open interval $(0, 1)$ in $\mathbf{R}$ for which all $u_n = \frac{1}{n}$ are interior points but there is no given ε that can work for all of them. Indeed, for a given ε, there exists n such that $\frac{1}{n} < \epsilon$ and thus $B_\varepsilon(\frac{1}{n})$ is not contained in $(0, 1)$.

Remark 1.5.9.

The notions of interior points and of openness of U are not absolute, but depend on the ambiant metric space (X, d) containing U. See the below examples for counterexamples.

Example 1.5.10. Let $Y = [0, 2) \cup \{4\} \subseteq \mathbf{R}$ with the induced metric $d|_Y$. Let $W_1 = [0, 1)$ and $W_2 = \{4\}$. Then: $\{0\}$ is an interior point of W_1 in Y, but not an interior point of U_2 in $\mathbf{R}$. Similarly, $\{4\}$ is an interior point of W_2 in Y, but not an interior point of W_2 in $\mathbf{R}$.

More generally, we have $W_1^\circ = [0, 1)$ *in* Y but $W_1^\circ = (0, 1)$ *in* $\mathbf{R}$. We also have $W_2^\circ = \{4\}$ *in* Y but $W_2^\circ = \varnothing$ *in* $\mathbf{R}$.

The only assertions that need some explanation are that $\{0\}$ is an interior point of W_1 in Y and that 4 is an interior point of W_2 in Y. For 0, we have $B_1(0, d|_Y) = \{x \in Y \mid d(x, 0) < 1\} = Y \cap (-1, 1) = [0, 1) \subseteq W_1$. Similarly, for 4 we have $B_1(4, d|_Y) = \{x \in Y \mid d(x, 4) < 1\} = Y \cap (3, 5) = \{4\} \subseteq W_2$.

Example 1.5.11. Let $W := \{(x, 0) \mid x \in \mathbf{R}\} \subseteq \mathbf{R}^2$ be the x-axis in $\mathbf{R}^2$. One easily shows that $(W, d_2|_W)$ is isometric to $\mathbf{R}$ with the standard metric. However, W is not open in $\mathbf{R}$. In fact, $W^\circ = \varnothing$. Indeed, for any $(x, 0)$ in W and any $\varepsilon > 0$, the point $(x, \frac{\varepsilon}{2})$ is in $B_\varepsilon((x, 0))$ but not in W. Therefore, $(x, 0)$ is not in W°.

Now that we have defined both open balls and open sets, we need to verify that the names are not misleading.

Lemma 1.5.12. *Every open ball $B_r(x)$ in a metric space (X, d) is open in X.*

Proof. Suppose we have a point $y \in B_r(x)$ and let $\varepsilon := r - d(x, y)$, which is positive. If $z \in B_\varepsilon(y)$, then $d(y, z) < \varepsilon$ and

$$d(x, z) \leqslant d(x, y) + d(y, z) < d(x, y) + \varepsilon = r,$$

so $z \in B_r(x)$. Therefore, for every $y \in B_r(x)$ there exists $\varepsilon > 0$ such that $B_\varepsilon(y) \subseteq B_r(x)$, and so $B_r(x)$ is open in X. $\square$

See the chapter illustration on page 1 for an illustration of the above proof (The illustration shows the ball $B_{r_y}(y)$ for some $0 < r_y < \varepsilon$).

The converse of Lemma 1.5.12 does not hold.

Remark 1.5.13.

Not every open set is an open ball, for example consider the subset $(0, 1) \times (0, 1)$ in $\mathbf{R}^2$. This is an open set with the Euclidean metric but not an open ball.

However, we have a weak converse to Lemma 1.5.12. Namely:

Lemma 1.5.14. *An open subset U of a metric space is a union of open balls.*

Proof. The proof is left to the reader (see Tutorial 3, Question 2). $\square$

The example of $(0, 1) \times (0, 1)$ in $\mathbf{R}^2$ shows that it is not always possible to write an open set as a *finite* union of open balls.

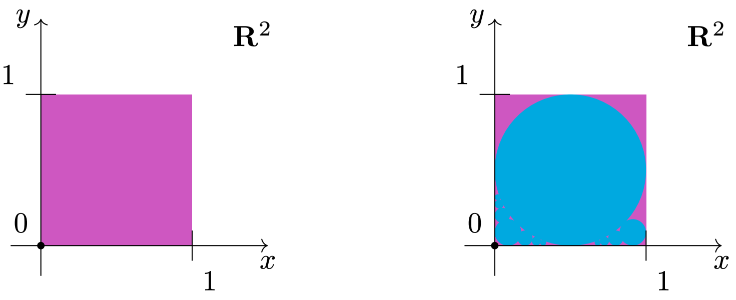


Figure 1.8.: The open set $(0, 1) \times (0, 1)$ is the interior of a square in $\mathbf{R}^2$ and cannot be covered by finitely many balls (interior of a disk).

Below are a few examples of open sets. The last one shows that being open is not an intrinsic property of a subset of X, but depends on the metric we have put on X.

Example 1.5.15. The subset $[a, b] \subseteq \mathbf{R}$ is not open in $(\mathbf{R}, d_\mathbf{R})$, but $[a, b]$ is open in itself, that is, in $([a, b], d_\mathbf{R}|_{[a,b]})$.

Example 1.5.16. In a discrete metric space X, every subset of X is open in X.

Example 1.5.17. The set $\{0\}$ is open in $\mathbf{R}$ with the discrete metric but not with the usual metric.

The operation of taking the interior satisfies some interesting properties that are summarised below.

Proposition 1.5.18. *Given any two subsets $U, V \subseteq (X, d)$ of a metric space we have*

1. $U \subseteq V \implies U^\circ \subseteq V^\circ$;
2. $(U^\circ)^\circ = U^\circ$;
3. U° *is the largest open subset of X that is contained in U.*

Proof. The proof is left to the reader (see Tutorial 3, Question 1). ☐

The collection of open subsets satisfies some nice closure properties.

Proposition 1.5.19. *Let (X, d) be a metric space. Then*

1. *The empty set $\varnothing$ and the whole space X are open in X;*
2. *If $U_1, \dots, U_m$ are open sets, then $\bigcap_{i=1}^m U_i$ is open in X;*
3. *The union of any collection of open sets is open in X.*

Proof. Since the empty has no elements, the condition that all its points are interior points is empty and thus true. The fact that all points of X are interior point is also trivially true.

Let $x \in \bigcap_{i=1}^m U_i$. Then $x \in U_i$, for every $i = 1, \dots, m$, and U_i is open in X which implies that there exists $\varepsilon_i > 0$ such that $B_{\varepsilon_i}(x) \subseteq U_i$. Take $\varepsilon = \min\{\varepsilon_1, \dots, \varepsilon_m\}$. Then $B_\varepsilon(x) \subseteq B_{\varepsilon_i}(x) \subseteq U_i$, for $i = 1, \dots, m$, so $B_\varepsilon(x) \subseteq \bigcap_{i=1}^m U_i$.

Let $x \in \bigcup_{i \in I} U_i$, where I is some indexing set, and each U_i is open in X. Then $x \in U_k$, for some k, and since U_k is open there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq U_k \subseteq \bigcup_{i \in I} U_i$. ☐

Combining the above proposition and Lemma 1.5.14, one obtain a characterisation of open sets.

Corollary 1.5.20. *Let $W \subseteq (X, d)$ be a subset of a metric space. Then W is open if and only if it is a union of open balls.*

Remark 1.5.21.

It is not necessarily true that the intersection of an infinite number of open sets is open. For example, in $\mathbf{R}$ consider $U_i := (-\frac{1}{i}, \frac{1}{i})$, for $i \in \mathbf{N}$. Then each U_i is open in $\mathbf{R}$ but $\bigcap_{i=1}^\infty U_i = \{0\}$ which is not open in $\mathbf{R}$.

Topology[3] is the branch of mathematics that studies the notion of open sets. While the exact definition of a topology is not part of this module[4], we will sometimes talk about "topological properties" meaning all properties following uniquely from Proposition 1.5.19 without needing to use the metric.

To go further

A **topological space** is a couple $(X, \mathscr{T})$ of a set and a collection $\mathscr{T}$, called a **topology**, of subsets of X satisfying Proposition 1.5.19. More precisely, a collection of subsets of X is a topology if it contains $\varnothing$ and X and is closed under finite intersections and arbitrary unions.

Using the characterisation of open sets as union of open balls, one can describe open sets of subspaces.

Proposition 1.5.22. *Let (X, d) be a metric space and $(Y, d|_Y)$ be a subspace. Then*

1. *The open subsets of Y are exactly the $U \cap Y$ for U open in X;*
2. *If Y is open and $U \subseteq Y$ is open in Y, then U is open in X.*

Proof. "1" Let y be any point of $Y \cap U$. There exists $\varepsilon > 0$ such that $B_\varepsilon(y; d) \subseteq U$. Since $B_\varepsilon(y; d|_Y) = B_\varepsilon(y; d) \cap Y$ we conclude that $B_\varepsilon(y; d|_Y)$ is contained in $Y \cap U$. This shows that $Y \cap U$ is open in Y.

For the other direction, let U be an open subset of Y, so U is a union of balls by Lemma 1.5.14: $U = \bigcup_{i \in I} B_{r_i}(y_i; d|_Y)$. Then $\tilde{U} = \bigcup_{i \in I} B_{r_i}(y_i; d)$ is open in X and $\tilde{U} \cap Y = U$.

"2" We have $U = Y \cap \tilde{U}$ for some open subset $\tilde{U}$ of X. If Y is open in X, then so is U as the intersection of two open subsets. $\square$

Using open sets, one obtain a third equivalent characterisation of continuity. This characterisation does not use the metric, only the notion of open sets, hence its name.

Remark 1.5.23.

Recall that for a function $f\colon X \to Y$ (not necessarily injective) and a subset $W \subseteq Y$ that

$$f^{-1}(W) := \{x \in X \mid f(x) \in W\} \subseteq X.$$

Theorem 1.5.24 (Topological Continuity)**.**

Suppose that $f\colon X \to Y$ is a function between metric spaces. Then f is continuous if and only if for every open set U in Y, $f^{-1}(U)$ is open in X.

---

[3] From the Greek τόπος and λογία, meaning place/location and study.

[4] Topology is the subject of MTH323.

Proof. "⇒" Suppose that f is continuous and let U be open in Y. We have to show that $f^{-1}(U)$ is open in X. Let $x \in f^{-1}(U)$. Then $f(x) \in U$ and, since U is open in Y, there exists $\varepsilon > 0$ such that $B_\varepsilon(f(x)) \subseteq U$. Since f is continuous, by Definition 1.5.5, there exists $\delta > 0$ such that $f(B_\delta(x)) \subseteq B_\varepsilon(f(x))$. So $f(B_\delta(x)) \subseteq U$, which implies that $B_\delta(x) \subseteq f^{-1}(U)$ and we have proved that $f^{-1}(U)$ is open in X.

"⇐" Assume the condition on open sets holds. We need to prove that f is continuous at every point $x \in X$. Given $\varepsilon > 0$, then $B_\varepsilon(f(x))$ is open in Y by Lemma 1.5.12. So by the assumption $f^{-1}(B_\varepsilon(f(x)))$ is open in X. Also, $x \in f^{-1}(B_\varepsilon(f(x)))$ so by the definition of open set, there exists $\delta > 0$ such that $B_\delta(x) \subseteq f^{-1}(B_\varepsilon(f(x)))$. Then $f(B_\delta(x)) \subseteq B_\varepsilon(f(x))$ which proves that f is continuous at x according to Definition 1.5.5. ☐

Remark 1.5.25.

Be careful that the topological characterisation of continuity is about the preimage of open sets and *not* about the image of open sets.

A function between two metric spaces is said to be **open** if the image of any open set is open. There exist bijective continuous but non-open functions. In other words, there exists a function f that is both bijective and continuous, but such that the inverse is not continuous. Similarly, there exist bijective open but non-continuous functions. See Examples 1.5.26 and 1.5.27.

Example 1.5.26. Take $X = \mathbf{R}$, $d = d_0$ the discrete metric and $d' = d_2$ the Euclidean metric. Then $\mathrm{Id}_{\mathbf{R}} \colon (\mathbf{R}, d_0) \to (\mathbf{R}, d_2)$ is continuous since its domain has the discrete metric. But $\mathrm{Id}_{\mathbf{R}} \colon (\mathbf{R}, d_2) \to (\mathbf{R}, d_0)$ is not as $\{0\}$ is open in $(\mathbf{R}, d_0)$ whilst its preimage $\mathrm{Id}^{-1}(\{0\}) = \{0\}$ is not open in $(\mathbf{R}, d_2)$.

Example 1.5.27. Let $([0,1), d|_{[0,1)})$ where $d|_{[0,1)}$ is the restriction of the standard metric on $\mathbf{R}$. Let $S^1 = \{z \in \mathbf{C} \mid |z| = 1\}$ be the unit circle with metric $d_{S^1}(z,w) = \theta$, where $\theta \in [0, \pi]$ is the angle zOw (or wOz if zOw is bigger than π). Then the function $f \colon [0,1) \to S^1, x \mapsto \mathrm{e}^{2\pi \mathrm{i} x}$ is a bijection. One can verify that f is continuous, but not open. Indeed, the subset $U = [0, 0.1) \subseteq [0, 0.1)$ is open, but $1 = e^{1\pi \mathrm{i} 0}$ belongs to $f(U) = \{\mathrm{e}^{2\pi \mathrm{i}\theta} \mid \theta \in [0, 0.1)\}$ but is not an interior point of it.

The inverse function $f^{-1} \colon S^{-1} \to [0,1)$ is open but not continuous.

The following result follows from Theorem 1.5.24 and the fact that any subset of a discrete metric space is open.

Lemma 1.5.28. *If (X, d_X) is a discrete metric space, then any function $f \colon X \to Y$ to another metric space (Y, d_Y) is automatically continuous.*

We conclude this section by one last example.

Example 1.5.29. The function $f \colon \mathbf{R} \to \mathbf{R}$ given by $f(x) = \begin{cases} x, & x \neq 0; \\ 1, & x = 0, \end{cases}$ is not continuous as can be shown using the result of Theorem 1.5.24 by observing that $f^{-1}(0, +\infty) = [0, +\infty)$ is not open in $\mathbf{R}$.

1.6. Equivalent Metrics

In Definition 1.1.13 we defined isometries between metric spaces (X, d) and (Y, d'), and we have seen that this is an equivalence relation; see Appendix B for a reminder on equivalence relations. By taking $X = Y$ and the identity function, this also defines an equivalence relation on the set of all metric on a given set X. That is, two metrics d d' on the same set are isometric if $\mathrm{Id}(X, d) \to (X, d')$ is an isometry, so if $d = d'$. In this section we will see two other, coarser, more interesting equivalence relations on this set.

In Remark 1.1.10 we defined what it means for two metrics to be scalar multiples of each other. This is also an equivalence relation, but still too restrictive for what we want to do. We will define another equivalence relation, by "d' is bounded above and below by scalar multiples of d".

Definition 1.6.1 (Lipschitz[5] Equivalence)**.**

Two metrics d and d' on a set X are **Lipschitz equivalent** if there exist $h, k \in (0, \infty)$ such that

$$hd'(x, y) \leqslant d(x, y) \leqslant kd'(x, y), \quad \forall x, y \in X.$$

Lemma 1.6.2. *Lipschitz equivalence is an equivalence relation on the collection of all metrics on a set X.*

Proof. The proof is left as an exercise (Tutorial 3, Question 9). ☐

It is clear that isometric metrics are Lipschitz equivalent (for $h = k = 1$). In other words, Lipschitz equivalence is a coarser equivalence relation than isometry. The converse is not true, as demonstrated by the following example.

Example 1.6.3 (Metrics on $\mathbf{R}^n$)**.** In Examples 1.1.3 and 1.1.5 and Definition 1.1.7 we have the metrics d_1, d_2, d_∞ on $\mathbf{R}^n$. We can check (Tutorial 3, Question 4) that

$$d_\infty(x, y) \leqslant d_2(x, y) \leqslant d_1(x, y) \leqslant nd_\infty(x, y), \quad \text{for all } x, y \in \mathbf{R}^n.$$

This proves that they are all Lipschitz equivalent, despite not being isometric.

Besides isometry and Lipschitz equivalence, we can define a third interesting equivalence relation on the space of metrics on X.

Definition 1.6.4 (Topological Equivalence)**.**

Two metrics d and d' on the set X are **topologically equivalent** if they give rise to the same open sets, that means, a subset $U \subseteq X$ is open with respect to the metric d if and only if it is open with respect to the metric d'.

---

[5] Named after Rudolf Otto Sigismund Lipschitz (1832–1903).

Lemma 1.6.5. *Topological equivalence defines an equivalence relation on the collection of all metrics on a set X.*

Proof. The proof is left as an exercise (Tutorial 3, Question 9). ☐

Topological equivalence is the coarser equivalence relation among our three relations (isometry, Lipschitz equivalence and topological equivalence), as demonstrated below.

Proposition 1.6.6. *Lipschitz equivalent metrics are topologically equivalent.*

Proof. Suppose that d and d' are Lipschitz equivalent metrics on a set X, so we have $h, k \in (0, \infty)$ such that

$$hd'(x,y) \leqslant d(x,y) \leqslant kd'(x,y), \quad \text{for all } x, y \in X.$$

Suppose also that $U \subseteq X$ is open with respect to the metric d. Let $x \in U$. Then there exists $\varepsilon > 0$ such that $B_\varepsilon(x;d) \subseteq U$. Take $\delta := \frac{\varepsilon}{k}$. Then $d(x,y) \leqslant kd'(x,y) < \varepsilon$, for all $y \in B_\delta(x;d')$. So $B_\delta(x;d') \subseteq B_\varepsilon(x;d) \subseteq U$ and U is open with respect to the metric d'.

Using the constant h on can shows that $B_{\frac{r}{h}}(x;d) \subseteq B_r(x;d')$. This implies that any open set for d' is also open for d. The details are left as an exercise. ☐

Remark 1.6.7.

It is not true that if two metrics are topologically equivalent then they are Lipschitz equivalent. See Tutorial 3, Question 7 for a counterexample.

Not all metrics on a given sets are topologically equivalent.

Example 1.6.8. Let Y be the set of continuous functions from $[a,b]$ to $\mathbf{R}$. It can be endowed with the d_{sup} metric (see page 11) as well as with the d_1 metric (see page 12). The metric spaces $\mathscr{C}[a,b] = (Y, d_{\text{sup}})$ and $\mathcal{L}_1[a,b] = (Y, d_1)$ are not topologically equivalent. Let $f_0 \colon [a,b] \to \mathbf{R}$ denote the constant function with value 0. We will show that $B_1(f_0; d_{\text{sup}})$ is not d_1-open. If it were then there would exist $\varepsilon > 0$ such that $B_\varepsilon(f_0; d_1) \subseteq B_1(f_0; d_{\text{sup}})$. But there always exists a continuous function $f \colon [a,b] \to \mathbf{R}$ such that $d_1(f, f_0) < \varepsilon$ but where $d_{\text{sup}}(f, f_0) = 1$ so $f \in B_\varepsilon(f_0; d_1)$ but $f \notin B_1(f_0; d_{\text{sup}})$. As an example of such a function, one can take

$$f(x) = \begin{cases} \frac{a-x}{\delta} + 1 & \text{if } x \in [a, a+\delta] \\ 0 & \text{if } x \in [a+\delta, b] \end{cases}$$

for $\delta < b - a$ small enough. Since $d_1(f, f_0) = \frac{\delta}{2}$, it is enough to take $\delta < \min b - a, 2\varepsilon$.

It follows from Proposition 1.6.6 that the metrics d_1, d_2, d_∞ (and d_p for $p \geqslant 1$) on $\mathbf{R}^n$ are all topologically equivalent. However, the discrete metric d_0 on $\mathbf{R}^n$ has one-point sets as open sets so it cannot be equivalent (either topologically or Lipschitz) to d_1, d_2 or d_∞.

Remark 1.6.9.

Since the metrics d_1, d_2, d_∞ on $\mathbf{R}^n$ are all topologically equivalent, even though the open balls in these metrics are different, the open sets are the same.

When studying vector spaces and groups, you have seen products, which were operations on the base set $X \times Y$ satisfying some natural compatibility conditions with the original operations. Both for vector spaces and for groups, there existes a unique such structure on $X \times Y$ satisfying the desired conditions. We would like to do the same for metric spaces. That is, given (X, d_X) and (Y, d_Y) we want to find a natural metric on the set $X \times Y$. Sadly, for metric spaces the situation is not as well-behaved as for vector spaces (or groups) and we can define many such natural metrics on the product. Below are a few examples that generalise the ideas of different metrics on $\mathbf{R}^2$ (which is a product of 2 copies of $\mathbf{R}$) to products of metric spaces in general.

Definition 1.6.10 (Cartesian Product)**.**

A **Cartesian product**[6] of two metric spaces (X, d_X) and (Y, d_Y) is the set $X \times Y$ with one of the following metrics:

- $d_a((x_1, y_1), (x_2, y_2)) = d_X(x_1, x_2) + d_Y(y_1, y_2)$;
- $d_b((x_1, y_1), (x_2, y_2)) = (d_X(x_1, x_2)^2 + d_Y(y_1, y_2)^2)^{\frac{1}{2}}$;
- $d_c((x_1, y_1), (x_2, y_2)) = \max\{d_X(x_1, x_2), d_Y(y_1, y_2)\}$,

for any $(x_1, y_1), (x_2, y_2) \in X \times Y$. The proofs that d_a, d_b and d_c satisfy the axioms (M0)–(M4) are derived directly from those for $X = Y = \mathbf{R}$ of Examples 1.1.3 and 1.1.5 and Definition 1.1.7 respectively, where $d_a = d_1$, $d_b = d_2$ and $d_c = d_\infty$.

Proposition 1.6.11. *The metrics d_a, d_b and d_c on $X \times Y$ are Lipschitz equivalent, and hence topologically equivalent.*

Proof. It follows directly from Definition 1.6.10 that

$$d_c(\alpha, \beta) \leqslant d_b(\alpha, \beta) \leqslant d_a(\alpha, \beta) \leqslant 2d_c(\alpha, \beta)$$

for every $\alpha, \beta \in X \times Y$. So Definition 1.6.1 applies to each pair in turn. ☐

These results may be extended to n-fold products, and particularly $\mathbf{R}^n$ which we looked at in Example 1.6.3. For example, the analog of the metric d_a on $X \times Y \times Z$ would be $d_a((x_1, y_1, z_1), (x_2, y_2, z_2)) = d_X(x_1, x_2) + d_Y(y_1, y_2) + d_Z(z_1, z_2)$.

---

[6] Named after René Descartes (1596–1650).

Remark 1.6.12.

For any $p \geqslant 1$, it is also possible to define a metric $d_{a_p}((x_1, y_1), (x_2, y_2)) = (d_X(x_1, x_2)^p + d_Y(y_1, y_2)^p)^{\frac{1}{p}}$ on $X \times Y$ similarly to what we did in Definition 1.1.6. This metric is Lipschitz equivalent to d_a (and hence also to d_b and d_c) and hence topologically equivalent to d_a, and also to d_b and d_c. We also call it *a* Cartesian product.

More generally, one say that a metric d on $X \times Y$ is **a Cartesian product** if it is topologically equivalent to one of the metric d_b from Definition 1.6.10. That is, a Cartesian product is a metric on $X \times Y$ that has the same open sets as d_b.

For example, the discrete metric d_0 on $\mathbf{R}^2$ is not a Cartesian product, as it has too many open subsets.

Lemma 1.6.13. *If a metric d is a Cartesian metric on $X \times Y$, then the projections $\pi_X \colon X \times Y \to X$ and $\pi_Y \colon X \times Y \to Y$ are continuous.*

Proof. The proof is left to the reader (Tutorial 3, Question 5). ☐

To go further

It can be shown that if d is a metric on $X \times Y$ such that the projections to (X, d_X) and (Y, d_Y) are continuous, then every d_a-open subset is d-open. In other words, $(X \times Y, d)$ is a Cartesian product of (X, d_X) and (Y, d_Y) if and only if it has the smallest number of open subsets while still making the projections continuous.

1.7. Equivalent Metric Spaces

We now extend the notions of Lipschitz and topological equivalences between metrics on the same set to metric spaces in general.

Definition 1.7.1.

A **Lipschitz equivalence** $f \colon (X, d_X) \to (Y, d_Y)$ between two metric spaces X and Y is a bijection $f \colon X \to Y$ such that there exist $h, k \in (0, \infty)$ such that

$$h d_Y(f(x), f(y)) \leqslant d_X(x, y) \leqslant k d_Y(f(x), f(y))$$

for all $x, y \in X$.

Two metric spaces are **Lipschitz equivalent** fif there exists a Lipschitz equivalence between them.

It is easy to see that an isometry is a special kind of Lipschitz equivalence (by taking $h = k = 1$), and that the identity function $\mathrm{Id}_X \colon (X, d) \to (X, d')$ is a Lipschitz equivalence if and only if d and d' are Lipschitz equivalent metrics on X.

We now define a third equivalence relation on metric spaces, in a similar fashion to what we did for metrics.

Definition 1.7.2 (Homeomorphism)**.**

A **homeomorphism**[7] (or **topological equivalence**) $f\colon (X, d_X) \to (Y, d_Y)$ between two metric spaces X and Y is a bijection $f\colon X \to Y$ such that both f and f^{-1} are continuous.
Two metric spaces are **homeomorphic** if there exists an homeomorphism between them.

Remark 1.7.3.

A continuous bijective function is not necessarily an homeomorphism. See Remark 1.5.25 and Examples 1.5.26 and 1.5.27 for counterexamples.

Remark 1.7.4.

An homeomorphism is a bijective function $f\colon X \to Y$ such both f and f^{-1} are continuous. By Theorem 1.5.24, f is continuous if and only if $f^{-1}(U)$ is open for every open subset $U \subseteq Y$. Since $(f^{-1})^{-1} = f$ we have that f^{-1} is continuous if and only if $f(U)$ is open for every open subset $U \subseteq X$. In other words, a homeomorphism is a bijective function that is both open and continuous.

Like with isometries, Lipschitz equivalences and homeomorphisms define an equivalence relation on the collection of all metric spaces, see Tutorial 3, Question 9. The entire subject of topology focuses on properties of spaces that do not change under homeomorphism.

Example 1.7.5. All open intervals are homeomorphic. More precisely, we have

1. Any open interval $(a, b) \subseteq \mathbf{R}$ is homeomorphic to any other open interval $(c, d) \subseteq \mathbf{R}$ via the homeomorphism $f\colon (a, b) \to (c, d)$ given by

   $$f(x) := \frac{d-c}{b-a}(x-a) + c.$$

   This is an affine function (of the form $f(x) = \alpha x + \beta$ for some $\alpha, \beta \in \mathbf{R}$ with $\alpha \neq 0$) and hence a continuous function. Its inverse is $g(x) := \frac{b-a}{d-c}(x-c) + a$ and is thus also continuous.

2. The function $x \mapsto \frac{x}{(1-|x|)}$ maps the interval $(-1, 1)$ homeomorphically onto $\mathbf{R}$.

3. The function

   $$x \mapsto \begin{cases} x, & \text{if } x \leqslant 0 \\ \frac{1}{1-x} & \text{if } x \geqslant 0 \end{cases}$$

   maps the interval $(-\infty, 1)$ homeomorphically onto $\mathbf{R}$.

---

[7] From the Greek ὅμοιος μορφή: similar shape.

4. The function $x \mapsto x + b - a$ maps $(-\infty, a)$ homeomorphically onto $(-\infty, b)$.
5. The function $x \mapsto -x$ maps the interval $(-\infty, a)$ homeomorphically onto $(-a, +\infty)$.

Altogether, every open interval is homeomorphic to $\mathbf{R}$.

It is evident that isometry is a finer equivalence relation than Lipschitz equivalence. The forthcoming proposition, which is similar to Proposition 1.6.6, shows that homeomorphism is an even coarser equivalence relation.

Proposition 1.7.6. *Lipschitz equivalent metric spaces are homeomorphic.*

Proof. Suppose that $f\colon X \to Y$ is a Lipschitz equivalence. Then the following hold:

$$f(B_{h\varepsilon}(x)) \subseteq B_\varepsilon(f(x)) \quad \text{and} \quad f^{-1}(B_{\frac{\varepsilon}{k}}(y)) \subseteq B_\varepsilon(f^{-1}(y)),$$

see Tutorial 3, Question 10.

Then f is continuous by the first condition and f^{-1} is continuous by the second. □

So we have the following chain of implications:

$$\text{isometry} \quad \Longrightarrow \quad \text{Lipschitz equivalence} \quad \Longrightarrow \quad \text{homeomorphism.}$$

Both implications are strict. The function $x \mapsto 2x$ is a Lipschitz equivalence from $\mathbf{R}$ to itself, which is not an isometry. Tutorial 3, Question 7 gives an example of an homeomorphism which is not a Lipschitz equivalence.

We used the same name, topological equivalence, to describe a relation between two metrics on the same set and to describe a kind the property (homeomorphism) of functions between spaces. This is justified by the forthcoming result.

Proposition 1.7.7. *Two metrics d and d' on a set X are topologically equivalent if and only if the identity function $\mathrm{Id}_X\colon (X, d) \to (X, d')$ is a homeomorphism.*

Proof. "⇒" Let $U \subseteq (X, d')$ be open. Then $\mathrm{Id}_X^{-1}(U) = U$ is open in (X, d) as the metrics are topologically equivalent. Therefore Id_X is continuous. Proving that the inverse is continuous works similarly.

"⇐" Suppose that $U \subseteq X$ is open with respect to the metric d'; then $\mathrm{Id}_X^{-1}(U) = U$ is open with respect to d since Id_X is continuous. Similarly, if $W \subseteq X$ is open with respect to d, then it is open with respect to d' because Id_X^{-1} is continuous. □

Remark 1.7.8.

It is possible to have at the same time $\mathrm{Id}_X\colon (X, d) \to (X, d')$ continuous but $\mathrm{Id}_X\colon (X, d') \to (X, d)$ not continuous!

For example, take $X = \mathbf{R}$, $d = d_0$ the discrete metric and $d' = d_2$ the Euclidean metric. Then $\mathrm{Id}_{\mathbf{R}}\colon (\mathbf{R}, d_0) \to (\mathbf{R}, d_2)$ is continuous since its domain has the discrete metric. But $\mathrm{Id}_{\mathbf{R}}\colon (\mathbf{R}, d_2) \to (\mathbf{R}, d_0)$ is not as $\{0\}$ is open in $(\mathbf{R}, d_0)$ whilst its preimage $\mathrm{Id}^{-1}(\{0\}) = \{0\}$ is not open in $(\mathbf{R}, d_2)$.

1.8. Closed Sets

Closed sets can be thought of as "*dual*" to open sets. This section will thus be similar to Section 1.5 and we will also start with the definition of the simplest closed sets: the closed balls.

Definition 1.8.1 (Closed Ball)**.**

Given a metric space (X, d), the **closed ball** of radius $r > 0$ around a point $x \in X$ is the set

$$\overline{B}_r(x) = \overline{B}_r(x; d) := \{y \in X \mid d(x, y) \leqslant r\} \subseteq X.$$

For $n \in \{1, 2, 3\}$, we recover the classical notions of closed intervals/disks/balls in the Euclidean space $\mathbf{R}^n$.

Similarly to what happens for open balls, if $(Y, d|_Y) \subseteq (X, d)$ is a subspace, then the closed balls of Y are exactly the $\overline{B} \cap Y$ for $\overline{B}$ a closed ball of X.

Example 1.8.2. In a discrete metric space, all open balls are closed balls, but not necessarily for the same radius. Indeed, for any point $x \in X$ we have $B_1(x) = \{x\} = \overline{B}_{\frac{1}{2}}(x)$ while $\overline{B}_1(x) = X$ is not equal to $\{x\}$ if X has at least 2 elements.

We defined open sets via the notion of interior points. The dual notion of interior point is closure points.

Definition 1.8.3 (Closure)**.**

Let W be a subset of a metric space (X, d). We say that $x \in X$ is a **closure point** of W if for every $\varepsilon > 0$, $B_\varepsilon(x) \cap W \neq \varnothing$. The **closure** of W is the set $\overline{W}$ of all closure points.

Observe that we always have $W^\circ \subseteq W \subseteq \overline{W}$, which is coherent with the English meaning of interior and closure.

Remark 1.8.4.

Even if we are defining closure points, the definition is still made using *open* balls and not closed ones!

We can finally define closed sets.

Definition 1.8.5 (Closed Set)**.**

A subset $V \subseteq (X, d)$ is **closed in** X if it contains all of its closure points, that is, $\overline{V} = V$.

Example 1.8.6. The open ball $B_1(0)$ is not closed in the Euclidean plane $\mathbf{R}^2$, because every point x with $|x| = 1$ is a closure point.

In a metric space, singletons are closed.

Lemma 1.8.7. *Any singleton $\{x\} \subseteq X$ in a metric space is closed.*

Proof. Let $y \neq x$ be any other element and let $\delta := d(x, y) > 0$. Then the open ball $B_{\frac{\delta}{2}}(y)$ does not intersect $\{x\}$, so y does not belong to $\overline{\{x\}}$. We conclude that $\overline{\{x\}} = \{x\}$. $\square$

The following proposition formalise the fact that closed and open sets are dual notions.

Proposition 1.8.8. *A set V is closed in X if and only if its complement $U := X \setminus V$ is open.*

Proof. "$\Rightarrow$" Suppose that V is closed. Since $\overline{V} = V$, no point of $u \in U$ can have $B_\varepsilon(u) \cap V \neq \varnothing$, for every $\varepsilon > 0$. So there exists some $\delta > 0$ such that $B_\delta(u) \cap V = \varnothing$. So $B_\delta(u) \subseteq U$.

"$\Leftarrow$" Suppose U is open and that there is an element $u \in \overline{V}$ such that $u \notin V$. Then $u \in U$ so $\exists \varepsilon > 0$ such that $B_\varepsilon(u) \subseteq U$. So $B_\varepsilon(u) \cap V = \varnothing$ which is a contradiction and hence $u \in V$ and V is closed. $\square$

Note that this proposition also says that U is open in X if and only if its complement $X \setminus U$ is closed in X.

Duality between open and closed subsets implies the dual of Proposition 1.5.22

Proposition 1.8.9. *Let (X, d) be a metric space and $(Y, d|_Y)$ be a subspace. Then*

1. *The closed subsets of Y are exactly the $V \cap Y$ for V closed in Y;*
2. *If Y is closed and $V \subseteq Y$ is closed in Y, then V is closed in X.*

Proof. "1" A subset $V \subseteq Y$ is closed in Y if and only if $Y \setminus V$ is open, if and only if there exists an open subset $U \subseteq X$ such that $U \cap Y = Y \setminus V$, if and only if $(X \setminus U) \cap Y = V$.

"2" The subset $U := Y \setminus V$ is open in Y and so equal to $\tilde{U} \cap Y$ for some open subset $\tilde{U}$ of X. We also have $X \setminus Y$ open in X. Therefore $(X \setminus Y) \cup \tilde{U}$ is open in X and $V = X \setminus \big((X \setminus Y) \cup \tilde{U}\big)$ is closed in X. $\square$

Closure and interior are also dual notions.

Proposition 1.8.10.

For a subset $W \subseteq (X, d)$ of a metric space we have

1. $W^\circ = X \setminus (\overline{X \setminus W})$;
2. $\overline{W} = X \setminus (X \setminus W)^\circ$.

Proof. The proof is left to the reader (see Tutorial 4, Question 5). $\square$

To better understand Proposition 1.8.10 it is interesting to start with $W = [0, 1) \subseteq \mathbf{R}$ and to compute all the sets $X \setminus W$, $\overline{X \setminus W}$ and $X \setminus \overline{X \setminus W}$.

To go further

Interaction between interior and closure is not straightforward. On the positive side, for any subset $W \subseteq (X, d)$ of a metric space we have

$$\overline{W^\circ} = \overline{(\overline{W^\circ})^\circ} \quad \text{and} \quad \overline{W}^\circ = \overline{(\overline{W}^\circ)}^\circ .$$

Since $(W^\circ)^\circ = W^\circ$ and $\overline{(\overline{W})} = \overline{W}$, it follows that starting from a subset W and taking interior and closure it is possible to obtain at most 7 different sets. These subsets are: W, W°, $\overline{W^\circ}$, $(\overline{W^\circ})^\circ$, $\overline{W}$, $\overline{W}^\circ$ and $\overline{(\overline{W}^\circ)}$. The set $W = (0, 1) \cup (1, 2) \cup \{3\} \cup (\mathbf{Q} \cap [4, 5]) \subseteq \mathbf{R}$ shows that all the above subsets can indeed be pairwise different.

Similarly to Lemma 1.5.12, we verify that the names closed sets and closed balls are not misleading.

Corollary 1.8.11. *Every closed ball $\overline{B}_r(x)$ is closed in X.*

Proof. The complement $U := X \setminus \overline{B}_r(x)$ of a closed ball is given by $\{y \in X \mid d(x, y) > r\}$. So for any $y \in U$ with $d(x, y) = t$ we have $B_{t-r}(y) \subseteq U$. So U is open in X. ☐

Closure satisfies properties dual to the ones from interior, see Proposition 1.5.18.

Proposition 1.8.12. *Given any two subsets $U, V \subseteq (X, d)$ of a metric space we have*

1. $U \subseteq V \implies \overline{U} \subseteq \overline{V}$;
2. $\overline{(\overline{V})} = \overline{V}$;
3. *$\overline{V}$ is the smallest closed subset of X that contains V.*

Proof. The proof is left to the reader (see Tutorial 4, Question 3). ☐

By combining Corollary 1.8.11 and Proposition 1.8.12 we obtain that in general $\overline{B_r(x)} \subseteq \overline{B}_r(x)$. In $\mathbf{R}^n$ this is an equality: the closure of an open ball is exactly the closed ball of the same radius.

Example 1.8.13. Let x be a point in the metric space $(\mathbf{R}^n, d_p)$ and let $r > 0$. Then $\overline{B_r(x)} = \overline{B}_r(x)$. We have to show that every point of $\overline{B}_r(x)$ is in the closure of $B_r(x)$.

Let y be any point in $\overline{B}_r(x)$ so $d_p(x, y) \leqslant r$. If $d_p(x, y) > r$ the y belongs to $B_r(x) \subseteq \overline{B_r(x)}$. If $d_p(x, y) = r$, for any $\varepsilon > 0$ let $z := \frac{\varepsilon}{2r}x + (1 - \frac{\varepsilon}{2r})y$. We have

$$d_p(z, y) = \|z - y\|_p = \frac{\varepsilon}{2} \quad \text{and} \quad d_p(z, x) = \|x - z\|_p = \Big(1 - \frac{\varepsilon}{2r}\Big)r < r.$$

This shows that z belongs to $B_\varepsilon(y) \cap B_r(x)$ and therefore that y is a closure point of $B_r(x)$. We conclude that $\overline{B_r(x)} \supseteq \overline{B}_r(x)$ as desired.

Remark 1.8.14.

Beware, for general metric spaces it is not always true that $\overline{B_r(x)} = \overline{B}_r(x)$. Can you find a counterexample? *Hint: think of the discrete metric.* See Tutorial 4, Question 6.

The collection of closed subsets satisfies some nice closure properties, dual to the ones satisfied by open sets from Proposition 1.5.19.

Proposition 1.8.15. *Let (X, d) be a metric space. Then*

1. *The empty set $\varnothing$ and the whole space X are closed in X;*
2. *If $V_1, \dots, V_m$ are closed sets, then $\bigcup_{i=1}^{m} V_i$ is closed in X;*
3. *The intersection of any collection of closed sets is closed in X.*

Proof. All these properties follow from the dual properties for open sets and the following facts:

$$X \setminus X = \varnothing, \quad X \setminus \varnothing = X, \quad X \setminus \left(\bigcup_{i=1}^{m} V_i \right) = \bigcap_{i=1}^{m} (X \setminus V_i), \quad X \setminus \left(\bigcap_{i \in I} V_i \right) = \bigcup_{i=1}^{m} (X \setminus V_i). \quad \square$$

There are subsets of a metric space that can be neither open nor closed:

Example 1.8.16. Let x be a point in a metric space (X, d) and let P be a **proper subset** of $\{y \in X \mid d(x, y) = r\}$, that is, P is not the empty set nor the whole set. The set $P_r(x) := B_r(x) \cup P$ is known as a **partially open ball**.

A partially open ball in $\mathbf{R}^2$ is neither open nor closed. The points $y \in P$ are not in the interior of $P_r(x)$ and the points $y \notin P$, where $d(x, y) = r$ are closure points of $P_r(x)$.

They are also subsets of metric spaces that are both open and closed, as $\varnothing$ and X. Such subsets are sometimes called **clopen**. In a discrete metric space, every subset is clopen, whilst the only clopen subsets of $\mathbf{R}^n$ are $\varnothing$ and $\mathbf{R}^n$, see Chapter 2 for more details.

We have seen in Theorem 1.5.24 that continuity can be characterised using open sets. It is thus not a surprise that we can give another characterisation of it using closed sets.

Theorem 1.8.17.

A function between metric spaces is continuous if and only if the preimage of every closed set is closed.

Proof. "⇒" Let $f \colon X \to Y$ be continuous and $V \subseteq Y$ be closed. Then

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

is open, since $Y \setminus V$ is open, which implies that $f^{-1}(V)$ is closed.

"$\Leftarrow$" Suppose that $f^{-1}(V)$ is closed in X for every closed subset $V \subseteq Y$. Suppose that U is open in Y. Then $Y \setminus U$ is closed, which implies that $f^{-1}(Y \setminus U)$ is closed in X and consequently that

$$X \setminus f^{-1}(Y \setminus U) = f^{-1}(U)$$

is open in X. Thus, f is continuous. $\square$

Using closure, one can define what it means for a subset U to be dense in X. Intuitively, this means that every point in X is either in U or arbitrarily close to a point in U, or that points of U are tightly clustered.

Definition 1.8.18.

A subset $W \subseteq (X, d)$ of a metric space is **dense in** X if $\overline{W} = X$.

From what we learned in Analysis 1, both $\mathbf{Q}$ and $\mathbf{R} \setminus \mathbf{Q}$ are dense in $\mathbf{R}$.

The subset $\mathbf{Q} \cap (0,1)$ is not dense in $\mathbf{R}$, but it is still "locally dense" in the sense that $\overline{\mathbf{Q} \cap (0,1)} = (0,1)$. For a strong negation of denseness, the correct notion is of being nowhere dense.

Definition 1.8.19.

A subset $W \subseteq (X, d)$ of a metric space is **nowhere dense in** X if $(\overline{W})^\circ = \varnothing$.

Remark 1.8.20.

When checking if a subset W is nowhere dense, do not forget to first take the closure of W before taking the interior! It is indeed possible that $W^\circ = \varnothing$ but W is not nowhere dense. This happens for example with $\mathbf{Q}$ in $\mathbf{R}$.

Example 1.8.21. The subset $\left\{\frac{1}{n} \mid n \in \mathbf{N}\right\}$ is nowhere dense in $\mathbf{R}$. Indeed, $\overline{\left\{\frac{1}{n} \mid n \in \mathbf{N}\right\}} = \left\{\frac{1}{n} \mid n \in \mathbf{N}\right\} \cup \{0\} \subseteq \mathbf{Q}$ has empty interior.

The following shows that being nowhere dense is, sort of, a dual notion of being dense.

Proposition 1.8.22. *A subset $W \subseteq (X, d)$ is nowhere dense in X if and only if $X \setminus \overline{W}$ is dense in X.*

Proof.

$$\begin{aligned}
(\overline{W})^\circ = \varnothing &\iff \text{no } x \in X \text{ is an interior point of } \overline{W}, \\
&\iff \forall \varepsilon > 0, B_\varepsilon(x) \not\subseteq \overline{W}, \quad \text{for all } x \in X, \\
&\iff \forall \varepsilon > 0, B_\varepsilon(x) \cap (X \setminus \overline{W}) \neq \varnothing, \quad \text{for all } x \in X, \\
&\iff x \in \overline{(X \setminus \overline{W})}, \quad \text{for all } x \in X, \\
&\iff X = \overline{(X \setminus \overline{W})}.
\end{aligned}$$

$\square$

Corollary 1.8.23. *A closed subset $V \subseteq (X, d)$ is nowhere dense in X if and only if $X \setminus V$ is dense in X.*

Example 1.8.24. The above corollary is not true for an arbitrary, not necessarily closed, subset V. For example, $\mathbf{R} \setminus \mathbf{Q}$ is dense in $\mathbf{R}$, but $\mathbf{Q}$ is far of being nowhere dense as $(\overline{\mathbf{Q}})^\circ = \mathbf{R}$.

We will conclude this section by a proposition that justifies the name nowhere dense. Before that we prove a small technical but useful lemma.

Lemma 1.8.25. *Let (X, d) be a metric space. Let $U \subseteq X$ be an open set and $W \subseteq X$ be any subset. Then $U \cap W \neq \varnothing$ if and only if $U \cap \overline{W} \neq \varnothing$.*

Proof. The left-to-right implication is trivial as $W \subseteq \overline{W}$.

For the other implication, let x be any element in $U \cap \overline{W} \neq \varnothing$. On one hand, since x is in the open set U, there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq U$. On the other hand, x being a closure point of W, the intersection $B_\varepsilon(x) \cap W \subseteq U \cap W$ is not empty. □

Proposition 1.8.26. *For a subset $W \subseteq (X, d)$ of a metric space the following are equivalent:*

I. *W is nowhere dense: $(\overline{W})^\circ = \varnothing$;*

II. *$\overline{W}$ contains no non-empty open subset;*

III. *For any non-empty open subset $U \subseteq X$, the subset $W \cap U$ is not dense in U (i.e. $U \nsubseteq \overline{W \cap U}$).*

Proof. "I $\iff$ II" This follows from the fact that V° is the biggest open subset of V.

"II $\implies$ III" We do the proof by contrapositive. Suppose that there exists a non-empty open subset U with $W \cap U$ dense in U. Then $U \subseteq \overline{U \cap W} \subseteq \overline{W}$.

"III $\implies$ II" Let U be an arbitrary non-empty open set. Since U is not contained in $\overline{U \cap W}$, the set $V := U \setminus \overline{U \cap W}$ is a non-empty open subset of U such that $V \cap (\overline{U \cap W}) = \varnothing$. We conclude that $V \cap (U \cap W) = V \cap W$ is empty and, by the previous lemma, that $V \cap \overline{W} = \varnothing$. This implies that V is contained in $X \setminus \overline{W}$, and in particular that U is not contained in $\overline{W}$ (as witnessed by any point of V), which finishes the proof. □

1.9. Boundedness

Having a notion of distance between points is enough to define a notion of "size" of subsets. We start by defining what it means for a subset to not be too big.

Definition 1.9.1 (Bounded Subset)**.**

A subset W of a metric space (X, d) is **bounded** if $W \subseteq \overline{B}_r(x)$ for some closed ball $\overline{B}_r(x)$ in X.

Observe that a subset $W \subseteq X$ is bounded if and only if it is contained in some open ball. Indeed, this follows from the fact that $B_r(X) \subseteq \overline{B}_r < (x) \subseteq B_{r+1}(x)$.

If W is a bounded subset of (X, d), witnessed by $\overline{B}_r(x)$, then for any $w_1, w_2 \in W$,

$$d(w_1, w_2) \leqslant d(w_1, x) + d(x, w_2) \leqslant 2r.$$

Therefore, the following definition makes sense:

Definition 1.9.2 (Diameter)**.**

If W is a non-empty bounded subset of a metric space (X, d), then the **diameter** of W is

$$\operatorname{diam} W := \sup\{d(w_1, w_2) \mid w_1, w_2 \in W\}.$$

Alternatively, we could have first use the formula in Definition 1.9.2 to define the diameter $d(W) \in \mathbf{R}_{\geqslant 0} \cup \{+\infty\}$ of any non-empty subset W of X and then define a bounded subset to be a subset of finite diameter. More precisely, we have

Lemma 1.9.3. *A non-empty subset $W \subseteq X$ is bounded if and only if* $\sup\{d(w_1, w_2) \mid w_1, w_2 \in W\} \in \mathbf{R}_{\geqslant 0}$.

Proof. We already have discussed the right-to-left implication, so let us prove the other implication. Suppose that $\sup\{d(w_1, w_2) \mid w_1, w_2 \in W\} = r$ is a well-defined real number. Then for any point $x \in W$ one have $W \subseteq \overline{B}_r(x)$ and thus W is bounded □

In practice, we will often prove that a subset is bounded by showing that $\{d(w_1, w_2) \mid w_1, w_2 \in W\} \leqslant K$ for some $K \in \mathbf{R}$

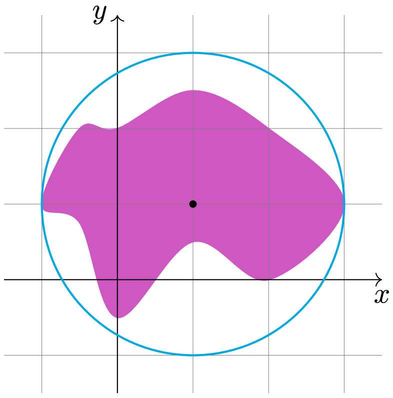


Figure 1.9.: A subset of $\mathbf{R}^2$ of diameter 2.

Remark 1.9.4.

It is in general not true that a bounded subset W is contained in a closed ball of radius $r = \frac{1}{2} \operatorname{diam}(W)$. For example, in a discrete metric space if W has at least two elements then its diameter is 1, but any closed ball containing W has radius at least 1.

Bounded subsets are preserved by Lipschitz equivalences. Namely,

Lemma 1.9.5. *Let $f\colon (X, d_X) \to (Y, d_Y)$ be a Lipschitz equivalence. Then a subset W of X is bounded if and only if $f(W)$ is bounded.*

Proof. There exists $h, k \in (0, \infty)$ such that $hd_Y(f(x), f(y)) \leqslant d_X(x, y) \leqslant kd_Y(f(x), f(y))$ for all $x, y \in X$. It follows that $h \operatorname{diam}_{d_Y} f(W) \leqslant \operatorname{diam}_{d_X} W \leqslant k \operatorname{diam}_{d_Y} f(W)$. ☐

The above statement is not true anymore if we replace Lipschitz equivalence by topological equivalence. In fact, if (X, d) is a metric space, then there exists a topologically equivalent metric d' such that (X, d') is bounded, see Tutorial 3, Question 7.

The collection of bounded subsets satisfies some nice closure properties.

Proposition 1.9.6. *Let X be a metric space. Then the following holds:*

1. *If $Y \subseteq Z$ are two subsets of $X$with Z bounded, then Y is also bounded;*
2. *Any point x in X belong to some bounded subset;*
3. *The union of any finite number of bounded subsets of a X is bounded.*

Proof. "1" The proof is left to the reader (see Tutorial 5, Question 1).

"2" We have $x \in \overline{B}_1(x)$.

"3" We only need to prove this for two subsets as the result follows by induction. Suppose that W_1 and W_2 are bounded subsets of a metric space (X, d) and let $x_1, x_2 \in X$ and $r_1, r_2 \in \mathbf{R}$ be such that $d(w, x_i) \leqslant r_i$, for all $w \in W_i$ $(i = 1, 2)$. Put $x = x_1$ say and $r = \max\{r_1, r_2 + d(x_1, x_2)\}$. Then for any $w \in W_1 \cup W_2$, either $w \in W_1$ and $d(w, x) \leqslant r_1 \leqslant r$, or $w \in W_2$ and $d(w, x) = d(w, x_1) \leqslant d(w, x_2) + d(x_2, x_1) \leqslant r_2 + d(x_2, x_1) \leqslant r$. So $W_1 \cup W_2 \subseteq \overline{B}_r(x)$. ☐

Observe that Proposition 1.9.6 1 implies that an intersection $\bigcap_{i \in I} W_i$ of bounded subsets is still bounded if $I \neq \varnothing$. Combining of 2 and 3 imply that any finite subset is bounded.

An arbitrary union of bounded subsets is not necessary bounded. A counterexample is given by $\bigcup_{n \geqslant 1} [-n, n]$ in $\mathbf{R}$.

To go further

A **bornological space** is a couple $(X, \mathscr{B})$ of a set and a collection $\mathscr{B}$, called a **bornology**, of subsets of X satisfying Proposition 1.9.6. More precisely, a collection $\mathscr{B}$ of subsets of X is a bornology if it is closed by taking subsets and finite unions and if $X = \bigcup_{B \in \mathscr{B}} B$.

Using bounded subsets, one can generalise the classical notion of real bounded functions, see Section 1.3, to functions with values in an arbitrary metric space.

Definition 1.9.7 (Bounded Function)**.**

If $f\colon A \to X$ is a function into a metric space (X, d), then f is **bounded** if the set

$$f(A) := \{f(a) \mid a \in A\} \subseteq X$$

is a bounded subset of (X, d).

Observe that in the definition of a bounded function $f\colon A \to X$ we do not require A to be a metric space.

The boundedness of a function f depends on: the "rule" defining f, the domain of f and the metric on the codomain.

Example 1.9.8. The function $f_1\colon (0,1) \to \mathbf{R}, x \mapsto x^2$ is bounded as $\mathrm{Im}(f_1) = (0,1) \subseteq \mathbf{R}$ is bounded. The function $f_2\colon \mathbf{R} \to \mathbf{R}, x \mapsto x^2$ is unbounded as $\mathrm{Im}(f_2) = \mathbf{R}_{\geqslant 0}$ is unbounded.

The function $g\colon (0,1) \to \mathbf{R}, x \mapsto \log(x)$ is unbounded as $\mathrm{Im}(f) = (-\infty, 0)$ is unbounded (but the domain of g is bounded).

The function $h\colon \mathbf{R} \to \mathbf{R}, x \mapsto 1$ is bounded as $\mathrm{Im}(f_1) = \{1\}$ is bounded (but the domain of g is unbounded).

It follows from Tutorial 3, Question 7 that there exists a metric d' on $\mathbf{R}$ which is topologically equivalent to the standard metric, but such that $(\mathbf{R}, d')$ is bounded. For this metric, any function $f\colon A \to (\mathbf{R}, d')$ is bounded. This is in particular true for the the functions f_1, f_2, g and h described above.

1.10. Boundary

In $\mathbf{R}$ with the standard metric, all four intervals $[0,1]$, $[0,1)$, $(0,1]$ and $(0,1)$ have the same boundary points: 0 and 1. We can generalise the notion of boundary points to subsets of arbitrary metric spaces. We do this by defining a boundary point as a point that is both arbitrarily close to W and arbitrarily close to its complement $X \setminus W$.

Definition 1.10.1.

Suppose $W \subseteq (X, d)$. An element $x \in X$ is a **boundary point** of W if, for every $\varepsilon > 0$,

$$B_\varepsilon(x) \cap W \neq \varnothing \quad \text{and} \quad B_\varepsilon(x) \cap (X \setminus W) \neq \varnothing.$$

The **boundary** ∂W of W is the set of all such boundary points.

Remark 1.10.2.

It follows directly from the definition that $\partial W = \partial(X \setminus W)$.

Interior and closure of a subset satisfies $W^\circ \subseteq W \subseteq \overline{W}$. We can use this to use an alternative definition of the boundary.

Proposition 1.10.3. *The boundary of any subset $W \subseteq (X, d)$ is the set $\overline{W} \setminus W^\circ$.*

Proof. Suppose that $x \in \partial W$. Then every $B_\varepsilon(x)$ meets W, so $x \in \overline{W}$; but $B_\varepsilon(x)$ meets $X \setminus W$, so $x \notin W^\circ$. Hence $\partial W \subseteq \overline{W} \setminus W^\circ$.

Conversely, suppose that $x \in \overline{W} \setminus W^\circ$. Then $x \in \overline{W}$, so every $B_\varepsilon(x)$ meets W; and $x \notin W^\circ$, so $B_\varepsilon(x)$ meets $X \setminus W$. Therefore, $B_\varepsilon(x) \cap W$ and $B_\varepsilon(x) \cap (X \setminus W)$ are both non-empty and $x \in \partial W$. Hence $\overline{W} \setminus W^\circ \subseteq \partial W$, as required. ☐

Boundaries are not any subsets, they are always closed.

Corollary 1.10.4. *Let $W \subseteq (X, d)$. Then ∂W is closed in X.*

Proof. We set $\partial W = \overline{W} \setminus W^\circ = \overline{W} \cap (X \setminus W^\circ)$ is the intersection of two closed subsets. It is therefore closed. ☐

If W is not closed, then it has closure points not in W, so $\partial W \not\subseteq W$. If W° is empty, then $\partial W = \overline{W}$.

The following characterisation of boundary explains the naming.

Lemma 1.10.5. *For any subset $W \subseteq (X, d)$ of a metric space we have $\partial W = \overline{W} \cap \overline{(X \setminus W)}$.*

Proof. The proof is left to the reader (see Tutorial 5, Question 3). ☐

We conclude this section with some examples.

Example 1.10.6. In $\mathbf{R}^n$, the closed ball $\overline{B}_r(0)$ is closed and its interior is the open ball $B_r(0)$. So $\partial \overline{B}_r(0) = \{x \in \mathbf{R}^n \mid |x| = r\}$ is the sphere of radius r. Similarly, $B_r(0)$ is open and its closure is the closed ball $\overline{B}_r(0)$. So $\partial B_r(0) = \{x \in \mathbf{R}^n \mid |x| = r\}$ also!

Example 1.10.7. The subset $\mathbf{Q} \subseteq \mathbf{R}$ is dense but $\mathbf{Q}^\circ$ is empty. Therefore $\partial \mathbf{Q} = \mathbf{R}$.

Example 1.10.8. The punctured plane $S := \mathbf{R}^n \setminus \{0\}$ of $\mathbf{R}^n$ is open, because $\{0\}$ is closed. Moreover, $\overline{S} = \mathbf{R}^n$, because, for every $\varepsilon > 0$, $B_\varepsilon(0) \cap S \neq \varnothing$. Therefore, $\partial S = \mathbf{R}^n \setminus S = \{0\}$.

1.11. Sequences

We know from Analysis 1 that a sequence of real number converges to some limit l if the points in the sequence become arbitrary close to l. Since this definition only use the distance, we can generalise it to any metric space.

Remind that for any set X a **sequence** in X is a function $s \colon \mathbf{N} \to X$. It is usual to write terms of the sequence $s(n)$ as x_n and denote the whole sequence by $(x_n)_{n \in \mathbf{N}}$ or simply by (x_n).

Definition 1.11.1.

A sequence (x_n) in a metric space **converges** to the point $x \in X$ if

$$\forall \varepsilon > 0, \exists N \in \mathbf{N} \text{ such that } n \geqslant N \implies d(x, x_n) < \varepsilon.$$

The point x is known as the **limit** of (x_n).

We can write the convergence of (x_n) to x as $\lim_n x_n = x$, or $x_n \to x$ (as $n \to \infty$).

Remark 1.11.2.

We write $\lim_n x_n = x$ using equality, but not all sequences are convergent so the left hand side might not exist. Can you find a counterexample? *Hint: remember Analysis 1.*

Example 1.11.3. In $\mathbf{R}$, Definition 1.11.1 coincides with the usual definition of $\lim_n x_n = x$ as we learnt in Analysis 1.

If we rewrite the definition of the limit in terms of open balls we have:

$$\lim_n x_n = x \iff \forall \varepsilon > 0, \exists N \in \mathbf{N}, \forall n \geqslant N : x_n \in B_\varepsilon(x).$$

Example 1.11.4. In a discrete metric space (X, d), a ball $B_\varepsilon(x) = \{x\}$ for any $\varepsilon \leqslant 1$. So $x_n \to x$ if and only if the sequence is eventually constant (i.e. there exists $N \in \mathbf{N}$ such that $x_n = x$, for all $n \geqslant N$).

Here is yet another characterisation of the limit.

Lemma 1.11.5. *In a metric space (X, d), $\lim_n x_n = x$ if and only if $d(x, x_n) \to 0$ in $\mathbf{R}$.*

Proof. Simple consequence of Definition 1.11.1. ☐

Example 1.11.6. In $\mathscr{C}[0,1]$ the sequence of continuous functions $f_n := x^n$ does not converge to the constant zero function since $\sup_{x\in[0,1]} |x^n| = 1$. But on the other hand, in $\mathscr{L}_1[0,1]$

$$d_1(x^n, 0) = \int_0^1 |x^n| \, \mathrm{d}x = \frac{1}{(n+1)} \to 0, \quad \text{as } n \to \infty,$$

so (f_n) converges to the constant zero function in $\mathscr{L}_1[0,1]$ but not in $\mathscr{C}[0,1]$. It is even possible to show that (f_n) does not converge in $\mathscr{C}[0,1]$. Indeed, if a limit g would exists, it should satisfy $g(x) = 0$ for $x < 1$ and $g(1) = 1$. But such a function is not continuous, see Example 4.3.2

In Definition 1.11.1 we talked about *the* limit (and not *a* limit) of a sequence. This is justified by the following result.

Proposition 1.11.7. *In any metric space (X, d) the limit of a convergent sequence is unique.*

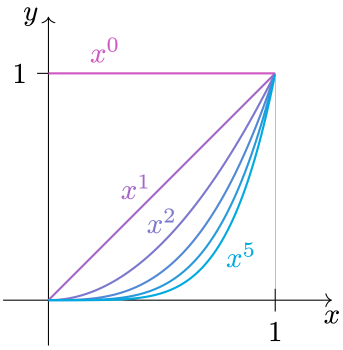


Figure 1.10.: Graphs of the polynomial functions $x \mapsto x^n$ for $n \in \{0, \dots, 5\}$. The area below the curve tends to 0, but the supremum of the function is always 1.

Proof. Suppose that $x_n \to x$ and $x_n \to x'$, where $x \neq x'$. Then $d := d(x, x') > 0$. Choose $\varepsilon = \frac{d}{2}$. So $B_\varepsilon(x) \cap B_\varepsilon(x') = \varnothing$. On the other hand, $x_n \in B_\varepsilon(x)$, for all $n \geqslant N$, and $x_n \in B_\varepsilon(x')$, for all $n \geqslant N'$. So $x_n \in B_\varepsilon(x) \cap B_\varepsilon(x') = \varnothing$, for any $n \geqslant \max\{N, N'\}$, which is absurd. Hence, $x = x'$. ☐

Limit points can be used to characterise the closure of a subset.

Proposition 1.11.8. *Suppose that $W \subseteq (X, d)$ and $w \in X$. Then $w \in \overline{W}$ if and only if there exists a sequence (w_n) in W such that $\lim_n w_n = w$.*

Proof. "⇒" Let $w \in \overline{W}$. Then $B_{\frac{1}{n}}(w) \cap W$ is non-empty for any integer $n \geqslant 1$. Choose a point $w_n \in B_{\frac{1}{n}}(w) \cap W$ for every $n \in \mathbf{N}$. Moreover, for any $\varepsilon > 0$, there exists $n \in \mathbf{N}$, such that $0 < \frac{1}{n} < \varepsilon$; so $w_n \in B_\varepsilon(w)$. Hence, $\lim_n w_n = w$.

"⇐" Suppose that $\lim_n w_n = w$. Then $w_n \in B_\varepsilon(w)$ for all sufficiently large n and any $\varepsilon > 0$. But $w_n \in W$, so $B_\varepsilon(w) \cap W \neq \varnothing$ and so $w \in \overline{W}$. ☐

Using the above result, one can characterise closed sets in terms of converging sequences. We can also characterise open sets in term of converging sequences.

Proposition 1.11.9. *Let $W \subseteq X$ be a subset of a metric space. Then*

1. *W is closed if and only for all $(x_n) \subseteq W$, if the $\lim_n x_n$ exists (in X) then it belongs to W;*
2. *W is open if and only if for every sequence (x_n) of X that converges to a point in W, there exists $K \in \mathbf{N}$ such that $x_k \in W$, for all $k \geqslant K$.*

Proof. "1" This follows from Proposition 1.11.8 and the fact that W is closed if and only if $\overline{W} = W$.

"2" The proof is left to the reader (see Tutorial 5, Question 6). ☐

Example 1.11.10. The subset $W := \{\frac{1}{n} \mid n \in \mathbf{N}\}$ of $\mathbf{R}$ does not contain 0. But the sequence $(w_n = \frac{1}{n})$ lies in W, and has limit 0; so $0 \in \overline{W}$. Moreover, any convergent sequence in W must tend to some $\frac{1}{m}$ or 0. So $\overline{W} = W \cup \{0\}$.

We can use convergence to give yet another characterisation of continuity.

Theorem 1.11.11.

A function $f\colon X \to Y$ between two metric spaces is continuous at $x \in X$ if and only if

$$(x_n) \text{ converges to } x \text{ in } X \implies f(x_n) \text{ converges to } f(x) \text{ in } Y.$$

Proof. "$\Rightarrow$" Suppose that f is continuous at $x \in X$ and that (x_n) is a sequence in X that converges to x. Then, for any $\varepsilon > 0$, there exists $\delta > 0$ such that $f(B_\delta(x)) \subseteq B_\varepsilon(f(x))$ and there also exists a $N \in \mathbf{N}$ such that $n \geqslant N$ implies $x_n \in B_\delta(x)$. Therefore, for $n \geqslant N$, $f(x_n) \in B_\varepsilon(f(x))$, and $f(x_n)$ converges to $f(x)$.

"$\Leftarrow$" Now suppose that $f(x_n)$ converges to $f(x)$ in Y for every (x_n) that converges to x in X. Assume by contradiction that f is not continuous at x. So there exists $\varepsilon > 0$ such that for every $\delta > 0$ we have $f(B_\delta(x)) \nsubseteq B_\varepsilon(f(x))$. Then for each $n \geqslant 1$, there exists (taking $\delta = \frac{1}{n}$) a $x_n \in B_{\frac{1}{n}}(x)$ with $f(x_n) \notin B_\varepsilon(f(x))$. So (x_n) converges to x by construction, but $f(x_n)$ does not converge to $f(x)$. This is a the desired contradiction. ☐

Corollary 1.11.12. *A function $f\colon X \to Y$ is continuous if and only if*

$$(x_n) \text{ converges to } x \text{ in } X \implies f(x_n) \text{ converges to } f(x) \text{ in } Y,$$

for every $x \in X$.

Observe that even if f is continuous, knowing that $(f(x_n))$ is convergent is a priori not enough to conclude about the convergence of (x_n). This is demonstrated in the following example.

Example 1.11.13. The function $e\colon \mathbf{R} \to \mathbf{R}^2$ given by $e(x) = (\cos x, \sin x)$ is continuous [can you prove it?]. We know that the sequence $(x_n = \frac{\pi}{n})$ converges to $0 \in \mathbf{R}$ so $(e(x_n))$ converges to $e(0) = (1,0) \in \mathbf{R}^2$. It is interesting to compare this with the sequence $(w_n = (2n + \frac{1}{n})\pi)$ which does not converge in $\mathbf{R}$ but $(e(w_n) = (\cos(\frac{\pi}{n}), \sin(\frac{\pi}{n})))$ also converges to $(1,0)$ since $e(x_n) = e(w_n)$.

Given a subset W of a metric space, we know from Proposition 1.11.8 that $\overline{W}$ is the set of all limit points of all sequences of elements of W. But any w in W is such a limit point for the trivial reason that $\lim_n w = w$. We thus might want to be more precise and to look at all points of X that can be "approximated" by sequences of elements of W. That is, we are interested at limits of not eventually constant sequences of elements of W.

Definition 1.11.14.

Let $W \subseteq (X,d)$ be a subset of a metric space. A point $x \in X$ is an **accumulation point** of W if, for every $\varepsilon > 0$, $B_\varepsilon(x) \cap W \setminus \{x\} \neq \varnothing$. We denote by $\mathrm{Acc}(W)$ the set of all accumulation points of W.

Note that for any subsets of X we have $A \cap (B \setminus C) = (A \cap B) \setminus C$. So $B_\varepsilon(x) \cap W \setminus \{x\}$ is not ambiguous.

Let us now formally prove that $\operatorname{Acc}(W)$ is the set of limit points of non-eventually constant sequences.

Lemma 1.11.15. *Let $W \subseteq (X, d)$ be a subset of a metric space. Then*

$$\operatorname{Acc}(W) = \{\lim w_n \mid (w_n) \in W \text{ is convergent in } X \text{ and not eventually constant}\}.$$

Proof. "$\subseteq$" Let $x \in \operatorname{Acc}(W)$. We want to show that $x = \lim w_n$ for a non-eventually constant sequence (w_n) of elements of W. Let $\varepsilon_1 = 1$. Then $B_{\varepsilon_1}(x) \cap W \setminus \{x\} \neq \varnothing$, we can hence choose an element w_1 inside. Suppose now that w_n has been constructed. By assumption $w_n \neq x$, so $\varepsilon_{n+1} := \min(d(w_n, x), \frac{1}{n+1}$ is a strictly positive real number. Therefore, there exists w_{n+1} in $B_\varepsilon(x) \cap W \setminus \{x\} \neq \varnothing$. We claim that the sequence (w_n) we just constructed is a non-eventually constant sequence converging to x. If $m < n$ then $d(w_m, x) > d(w_n, x)$ by construction. So the sequence (w_n) is not eventually constant. Moreover, $d(w_n, x) < \varepsilon_n \leqslant \frac{1}{n}$ which converges to 0. Hence the sequence (w_n) converges to x.

"$\supseteq$" Let w_n be a non-eventually constant converging sequence of elements of W and let x be its limit. Then, for every $\varepsilon > 0$ there exists N such that $\frac{1}{N} \leqslant \varepsilon$. So for every $n \geqslant N$ we have $d(w_n, x) < \varepsilon$ and w_n belongs to s. Since the sequence is not eventually constant, there exists $n, m \geqslant N$ such that $w_n \neq w_m$. Therefore, at least one of w_n and w_m belongs to $B_{\frac{1}{n}}(x) \cap W \setminus \{x\}$. We conclude that x is an accumulation point of W. ☐

Example 1.11.16. In $\mathbf{R}$ we have $\operatorname{Acc}([0, 1) \cup \{2\}) = [0, 1]$.

Accumulation points are related to the closure of W in the following way.

Proposition 1.11.17. *For any subset $W \subseteq X$ of a metric space we have $\overline{W} = W \cup \operatorname{Acc}(W)$.*

Proof. The proof is left as an exercise (Tutorial 4, Question 7). ☐

The following concept is the dual of accumulation points.

Definition 1.11.18.

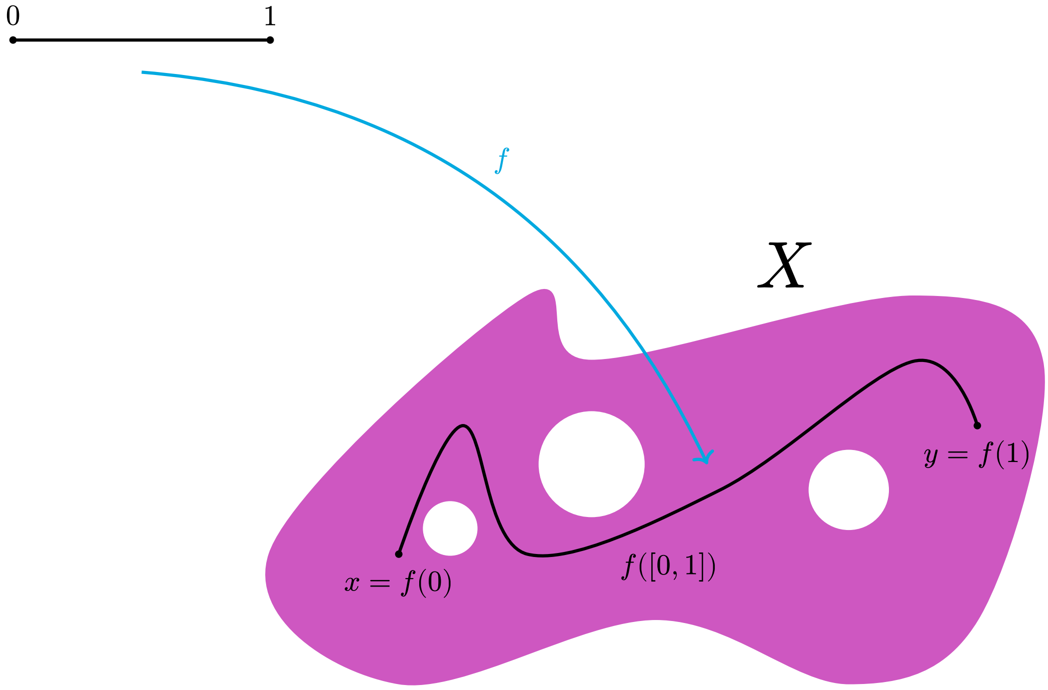
Let $W \subseteq (X, d)$ be a subset of a metric space. A point $w \in W$ is an **isolated point** of W if there exists $\varepsilon > 0$ such that $B_\varepsilon(w) \cap W = \{w\}$. We write $\operatorname{Iso}(W)$ for the set of all isolated points of W.

Proposition 1.11.19. *For any subset $W \subseteq X$ of a metric space we have* $\operatorname{Iso}(W) = W \setminus \operatorname{Acc}(W)$.

Proof. The proof is left as an exercise (Tutorial 4, Question 7). ☐

CHAPTER 2.

CONNECTEDNESS



2.1. Connectedness

A common intuition is that a subset W of $\mathbf{R}^n$ is connected if for any two points x and y in W it is possible to "draw a path" from x to y that stays in W. Whilst this gives an interesting notion, which we will study in Section 2.2, it is sometimes not general enough as there are subsets from $\mathbf{R}^2$ that we would like to be connected but do not satisfy this definition. The intuition for the forthcoming general definition of connexity is the following. First of all, in $\{0,1\}$ with the discrete metric, 0 should not be connected to $\{1\}$, so $\{0,1\}$ is not connected. Then we want the fact of being connected to be preserved by continuous functions. So, a subset W of X has at least two "connected components" if we can colour elements of W in purple an blue such that elements coloured in purple stay away from elements coloured in blue.

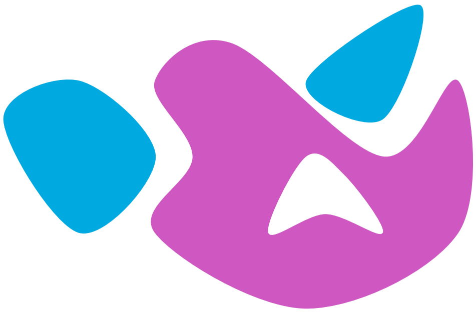


Figure 2.1.: A non-connected subset of $\mathbf{R}^2$, with at least two connected components.

Definition 2.1.1 (Connected)**.**

A non-empty metric space (X,d) is **connected** if there does not exist a surjective continuous function from X onto a discrete two-point space.
A subset $W \subseteq X$ is **connected** if $(W, d|_W)$ is connected.

In Figure 2.1 we can define a map $f\colon W \to \{0,1\}$ sending purple points to 0 and blue points to 1. This map is surjective and continuous if we endow $\{0,1\}$ with the discrete metric.

Definition 2.1.2 (Partition)**.**

A **partition** $A \mid B$ of a metric space X is a pair of non-empty subspaces $A, B \subseteq X$ such that $A \cup B = X$, $A \cap B = \varnothing$, and both A and B are open in X.
We say that $A \mid B$ **partitions** X.

Note that A and B are also both closed in X when $A \mid B$ partitions X.

Theorem 2.1.3.

A non-empty metric space X is connected if and only if it admits no partition.

Proof. "$\Rightarrow$" Let D denote $\{0,1\}$ with the discrete metric. Suppose that $A \mid B$ partitions X. Define $f\colon X \to D$ by

$$f(x) := \begin{cases} 0, & x \in A; \\ 1, & x \in B. \end{cases}$$

Since A and B are non-empty, f is surjective. The only open sets of D are: $\varnothing, \{0\}, \{1\}, D$. Now

$$f^{-1}(\varnothing) = \varnothing, \quad f^{-1}(0) = A, \quad f^{-1}(1) = B, \quad f^{-1}(D) = X.$$

So f is continuous as the preimage of every open set in D is open in X. So by definition, X is not connected.

"$\Leftarrow$" If X is not connected there exits a surjective continuous function $f\colon X \to D$ and it is easy to show that $f^{-1}(0) \mid f^{-1}(1)$ partitions X. ☐

To go further

One can generalise Definitions 2.1.1 and 2.1.2 and Theorem 2.1.3 in the following way. Let (X,d) be a metric space and $W \subseteq$ be a subspace. We say that W has **at least n connected components** if there exists a surjective map $f\colon X \to \{1,\dots,n\}$ which is continuous for the discrete metric on $\{1,\dots,n\}$. We say that W admits a **n-partition** if there exists n non-emtpy open pairswise disjoint subsets U_i such that $W = \bigcup_i U_i$. The **number of connected components** of W is the minimal n, if it exists, such that W has at least n connected component, equivalently such that there exists a n-partition. If no such n exists and W is non-empty, we say that W has infinitely many components. Finely, the number of connected components of $\varnothing$ is 0. For example, the subset of $\mathbf{R}^2$ depicted in Figure 2.1 has 3 connected components.

One can even generalise more and define for any, possibly infinite, cardinal κ the fact to have at least κ-connected components, which is equivalent to admitting a κ-partition.

It is easy to prove that a set is connected if and only if it has at most 1 connected component.

The characterisation of connectedness in terms of the absence of partition is sometimes easier to use as demonstrated in the following example.

Example 2.1.4. The rational numbers $\mathbf{Q}$ are not connected. For if $\alpha \in \mathbf{R} \setminus \mathbf{Q}$ is any irrational number, then $(\mathbf{Q} \cap (-\infty, \alpha)) \mid (\mathbf{Q} \cap (\alpha, +\infty))$ partitions $\mathbf{Q}$.

The following corollary of Theorem 2.1.3 gives us a third characterisation of connectedness.

Theorem 2.1.5.

A non-empty metric space is connected if and only if the only subsets of X which are both open and closed in X are $\varnothing$ and X.

Proof. If $A \mid B$ partitions X, then A (and B) is an open and closed subset of X which is neither $\varnothing$ nor X. Conversely, if there exists an open and closed set A in X other than $\varnothing$ and X, then $A \mid (X \setminus A)$ partitions X. ☐

By an **interval** in $\mathbf{R}$ we mean subsets of the form:

$$[a, b],\ (a, b),\ (a, b],\ [a, b),\ (a, +\infty),\ [a, +\infty),\ (-\infty, b),\ (-\infty, b],\ (-\infty, +\infty) = \mathbf{R},$$

where $a \leqslant b$ are two real numbers. Observe that $a = b$ gives $[a, a] = \{a\}$ in the first interval and $(a, a) = (a, a] = [a, a) = \varnothing$ in intervals 2 to 4.

We will shows that non-empty intervals are exactly the connected subsets of $\mathbf{R}$. To do so, we will use the following characterisation in term of "betweeness".

Lemma 2.1.6. *A subset $A \subseteq \mathbf{R}$ is an interval if and only if it satisfies the following property: if $x, y \in A$ and $z \in \mathbf{R}$ are such that $x < z < y$, then $z \in A$.*

Proof. If A is an interval it clearly has the stated property. Conversely, suppose A has the stated property. Let

$$a = \inf A \quad (\text{or } a = -\infty \text{ if } A \text{ is not bounded below}),$$
$$b = \sup A \quad (\text{or } b = +\infty \text{ if } A \text{ is not bounded above}).$$

We shall prove that $(a, b) \subseteq A \subseteq [a, b]$.

First, if $z \in (a, b)$, then $z > a$, so by definition of a there exists x in A with $x < z$. Similarly, there exists $y \in A$ with $y > z$. So by the hypothesis, $z \in A$. This proves that $(a, b) \subseteq A$, and the inclusion $A \subseteq [a, b]$ follows from the definitions of a and b. ☐

Lemma 2.1.7. *Connected subspaces of $\mathbf{R}$ are intervals.*

Proof. We will prove the contrapositive. Suppose $A \subseteq \mathbf{R}$ is not an interval. Then by Lemma 2.1.6 there exist $x, y \in A$ and $z \in \mathbf{R} \setminus A$ such that $x < z < y$. Then $(A \cap (-\infty, z)) \mid (A \cap (z, +\infty))$ partitions A. Indeed both $A \cap (-\infty, z)$ and $A \cap (z, +\infty)$ are open in A, they are non-empty since they contain x and y respectively, they are clearly disjoint and their union is A since $z \notin A$. ☐

Lemma 2.1.8. *Any non-empty interval $I \subseteq \mathbf{R}$ is connected.*

Proof. Suppose by contradiction that $A \mid B$ partitions I. Let $a \in A$ and $b \in B$ and suppose that $a < b$ (otherwise change the names of A and B). Since $a, b \in I$ and I is an interval $[a, b] \subseteq I$. Let $A' := A \cap [a, b]$ and $B' := B \cap [a, b]$. Since B is closed in I, B' is closed in $[a, b]$. Moreover, since $[a, b]$ is closed in $\mathbf{R}$, B' is closed in $\mathbf{R}$.

Since $B' \neq \varnothing$ and B' is bounded below (by a for example) $b' = \inf B'$ exists by the completeness axiom, and $b' \geqslant a$. Since B' is closed in $\mathbf{R}$, $b' \in B'$. Since $a \in A'$ and

$A' \cap B' = \varnothing$, it follows that $b' > a$. Now let $A'' := A' \cap [a, b']$. Arguing as for B', we get that $a'' = \sup A''$ exists, $a'' \in A''$ and that $a'' < b'$. Now $(a'', b') \cap A = \varnothing$, since otherwise a'' is not an upper bound for A''. Similarly, $(a'', b') \cap B' = \varnothing$, otherwise b' is not a lower bound for B'. But $A' \cup B' = [a, b]$ and $(a'', b') \subseteq [a, b]$. Therefore, we have a contradiction. ☐

As a corollary of the previous two lemmas, we directly obtain:

Proposition 2.1.9. *A subspace of R is connected if and only if it is a non-emtpy interval.*

We now show that connectedness is preserved by continuous maps.

Proposition 2.1.10. *The continuous image of a connected metric space is connected. That is, if $f\colon X \to Y$ is a continuous function between metric spaces where X is connected, then $f(X)$ is connected.*

Proof. Using the above notation, if $A \mid B$ partitions $f(X)$, then $f^{-1}(A) \mid f^{-1}(B)$ partitions X. ☐

As a corollary we have:

Theorem 2.1.11.

Connectedness is a topological property. That is, if X and Y are homeomorphic, then X is connected if and only if Y is connected.

Proposition 2.1.10 is an useful tool to show that some space as connected, as demonstrated below.

Example 2.1.12. The unit circle $S^1 = \{(x, y) \in \mathbf{R}^2 \mid x^2 + y^2 = 1\}$ is connected since $e\colon [0, 2\pi] \to \mathbf{R}^2$, $e(x) = (\cos x, \sin x)$, is continuous and $e([0, 2\pi]) = S$.

Another consequence of Proposition 2.1.10 (and of the description of connected subsets of $\mathbf{R}$ of Lemma 2.1.7) is:

Corollary 2.1.13. *If X is a connected space and $f\colon X \to \mathbf{R}$ is continuous, then $f(X)$ is an interval.*

Using the above, we recover an important results from Analysis:

Corollary 2.1.14 (Intermediate Value Theorem)**.** *If $f\colon X \to \mathbf{R}$ is a continuous function from a connected metric space X and $y \in (\inf f(X), \sup f(X))$, then there exists $x \in X$ such that $f(x) = y$. (If $f(X)$ is not bounded below then we take $\inf f(X) = -\infty$ and similarly if $f(X)$ is not bounded above.)*

Proof. Since X is connected, $f(X)$ is a connected subspace of $\mathbf{R}$ and therefore an interval. It directly follows that $(\inf f(X), \sup f(X)) \subseteq f(X)$, which concludes the proof. ☐

A real function $f\colon \mathbf{R} \to \mathbf{R}$ is often identified with its graph, that is with a subset of $\mathbf{R}^2$. Studying the graph of f is useful to understand better the function f. The same thing can be done for arbitrary function. Let $f\colon X \to Y$ be a function between two sets. The **graph of** f is the set

$$G_f := \{(x, y) \in X \times Y \mid f(x) = y\}.$$

The fact that the graph of a continuous real function $f\colon \mathbf{R} \to \mathbf{R}$ is a curve in $\mathbf{R}^2$ can be generalised to metric spaces.

Proposition 2.1.15. *If $f\colon X \to Y$ is a continuous function between metric spaces, then G_f is homeomorphic to X.*

Proof. Let $i\colon G_f \to X \times Y$ be the inclusion map and let $p_1 : X \times Y \to X$ be the projection onto the first factor. Let $\theta\colon G_f \to X$ and $\psi\colon X \to G_f$ be defined respectively by $\theta = p_1 \circ i$ and $\psi(x) = (x, f(x))$ for all $x \in X$. It is easily checked that θ and ψ are mutually inverse. Continuity of θ follows from continuity of p_1 and i. For the continuity of ψ note that we can write $\psi = (\mathrm{Id} \times f) \circ \Delta$, where $\Delta\colon X \to X \times X$ is the continuous function $\Delta(x) = (x, x)$. Continuity of Δ is evident and continuity of $\mathrm{Id} \times f$ follows from the continuity of Id and f. ☐

Corollary 2.1.16. *If $f\colon X \to Y$ is continuous with X connected, then G_f is a connected subset of $X \times Y$.*

Proposition 2.1.17. *If W is a connected subspace of a metric space X and if $W \subseteq Y \subseteq \overline{W}$, then Y is connected.*

Proof. Let $f\colon Y \to D$ be any continuous function, where D is the discrete space $\{0, 1\}$. Since W is connected and $f|_W$ is continuous, $f(W) = \{0\}$ or $\{1\}$. Suppose, without loss of generality, that $f(W) = \{0\}$ and suppose, to get a contradiction, that there exists $y \in Y$ such that $f(y) = 1$. By continuity, $f^{-1}(1)$ contains a small ball $B_\varepsilon(y) \subseteq Y$. Since Y is in the closure of W there exists a point $x \in B_\varepsilon(y) \cap W$. But then $f(x) = 0$ since x is in W and $f(x) = 1$ since x belongs to $f^{-1}(1)$. This is the desired contradiction. ☐

The similar statement for the interior is not true as demonstrated by the following example.

Example 2.1.18. Let $W = \overline{B}_1(-1) \cup \overline{B}_1(1) \subseteq \mathbf{R}^2$. Then W is connected, but $W^\circ = B_1(-1) \cup B_1(1)$ is not.

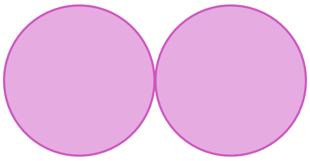


Figure 2.2.: A connected set with non-connected interior.

We conclude this section by constructing an example of a pathological connected subset of $\mathbf{R}^2$.

Example 2.1.19 (Topologist's Sine Curve)**.** Let G be the graph of $\sin\frac{1}{x}$ for $x > 0$ and let $J := \{(0,y) \in \mathbf{R}^2 \mid -1 \leqslant y \leqslant 1\}$. Since $(0,+\infty)$ is connected then, by Corollary 2.1.16,

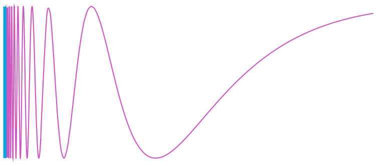


Figure 2.3.: The graph G of $\sin(\frac{1}{x})$ in purple and the segment J in blue.

G is connected. To show that $G \cup J$ is connected we only need to show that $J \subseteq \overline{G}$. Let $p = (0,y) \in J$. We need to show that for any $\varepsilon > 0$, $B_\varepsilon(p) \cap G \neq \varnothing$. Choose $n \in \mathbf{N}$ such that $\frac{1}{2n\pi} < \varepsilon$. Since $\sin(\frac{1}{2}(4n+1)\pi) = 1$ and $\sin(\frac{1}{2}(4n+3)\pi) = -1$, by the Intermediate Value Theorem $\sin\frac{1}{x}$ takes on every value between -1 and 1 in the interval $[\frac{2}{(4n+3)\pi}, \frac{2}{(4n+1)\pi}]$. In particular, $\sin\frac{1}{x_0} = y$ for some x_0 in this interval. The distance between $(0,y)$ and $(x_0, \sin\frac{1}{x_0}) = (x_0, y)$ is $x_0 < \varepsilon$, so $(x_0, \sin\frac{1}{x_0}) \in B_\varepsilon(p) \cap G$ as required. (Actually, one can shows that $\overline{G} = G \sqcup J$).

2.2. Path-connectedness

In this section, we will study a stronger notion of connectedness, which corresponds better to the intuition that two points are connected if one can draw a path from the first one the second.

We will write I for the closed interval $[0,1] \subseteq \mathbf{R}$.

To make formal the notion of "connected by a path", we need to define what is a path.

Definition 2.2.1 (Path)**.**

Given two points x, y in a metric space X, a **path** from x to y is a continuous function $f\colon [0,1] \to X$ such that $f(0) = x$ and $f(1) = y$.
We say that x and y are **joined** by the path f.

See the chapter illustration on page 41 for an illustration of a path between two points.

Observe that if f is a path from x to y, then g defined by $g(t) = f(1-t)$ is a path from y to x. Also, for any point x in a metric space, the constant function $f\colon [0,1] \to \mathbf{R}, t \mapsto f(t) = x$ is a path from x to itself.

Usually, when talking about a path f, we will use the letter "t" for the dependent variable. This is because we think of $f(t)$ as the position at *time* t.

Example 2.2.2. Let $x = (1,1)$ and $y = (4,2)$ in $\mathbf{R}^2$. Then

$$f(t) = (1-t)x + ty \quad \text{and} \quad g(t) = ((t+1)^2, (t+1)))$$

are paths from x to y. The graph of f is the line-segment between x and y.

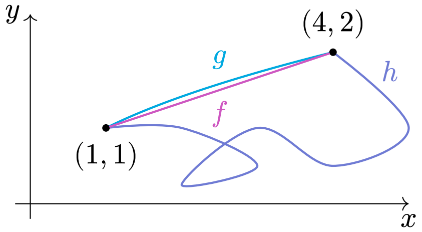


Figure 2.4.: Three paths from $(1,1)$ to $(4,2)$.

If we can go with a path f from x to y and from y to z, it seems natural that we should be able to go from x to z by concatenating the two paths. This basic idea almost works, except that it will gives a continuous function $h\colon [0,2] \to X$, so we need to rescale our two paths.

Lemma 2.2.3. *Suppose $f, g\colon I \to X$ are paths in a metric space X such that $f(1) = g(0)$. Then*

$$h(x) := \begin{cases} f(2x), & x \in [0, \frac{1}{2}]; \\ g(2x-1), & x \in [\frac{1}{2}, 1], \end{cases}$$

is a path in X from $f(0)$ to $g(1)$.

Proof. The proof is left to the reader (see Tutorial 7, Question 4). ☐

It follows from the previous Lemma and the discussion after Definition 2.2.1 that "being connected by a path" is an equivalence relation on the points of a metric space.

Corollary 2.2.4. *The relation $x \sim y$ if there is a path from x to y is an equivalence relation on the points of X.*

Proof. The proof is left to the reader (see Tutorial 7, Question 5). ☐

We now define the main concept of this section.

Definition 2.2.5 (Path-connected)**.**

A non-empty metric space X is **path-connected** if any two points in X can be joined by a path in X.
A subspace $W \subseteq X$ is **path-connected** if $(W, d|_W)$ is a path-connected metric space.

Given a metric space X, a subset W is path-connected if and only if for any x and y in W there exists a path from x to y that stays in W. That is, if and only if there exists a continuous function $f\colon [0,1] \to X$ with $f(0) = 1$, $f(1) = y$ and $f([0,1]) \subseteq W$.

It follows from Corollary 2.2.4 that a subspace W is path-connected if and only if there exists $w \in W$ such that for every x in W there is a path from w to x.

Similarly to what happens for connectedness (Proposition 2.1.10 and Theorem 2.1.11), path-connectedness is preserved by homeomorphism.

Theorem 2.2.6.

The continuous image of a path-connected metric space is path-connected. It follows that path-connectedness is a topological property. That is, if X and Y are homeomorphic, then X is path-connected if and only if Y is path-connected.

Proof. The proof is left to the reader (see Tutorial 7, Question 3). ☐

Let us now explore some examples of path-connected metric spaces. Firstly, intervals in $\mathbf{R}$ are path-connected, but the same is true more generally for balls in $\mathbf{R}^n$.

Lemma 2.2.7. *Balls (open or closed) in a normed vector space are path-connected. In particular, balls in the Euclidean space $\mathbf{R}^n$ are path-connected.*

Proof. Let $(V, \|\cdot\|)$ be a normed vector space and let c be a point in V and $r > 0$ be a real number. Let x and y be two points in $B = B_r(c)$. The function $f\colon V \to \mathbf{R}$ defined by $f(t) := (1-t)x + t(y)$ is continuous, and therefore a path from x to y in V. We need to check that $f(t)$ stays inside $B_r(c)$ for every $t \in [0,1]$.

For any $t \in [0,1]$, we have

$$\begin{aligned} d(c, f(t)) &= \|f(t) - c\| = \|(1-t)x + ty - c\| = \|(1-t)(x-c) + t(y-c)\| \\ &\leqslant |1-t|\|x-c\| + |t|\|y-c\| < (1-t)r + tr = r \end{aligned}$$

So $f(t)$ belongs to B. If $B = \overline{B}_r(c)$ is a closed ball, the proof is the same, but with $<$ replaced by $\leqslant$. ☐

As a corollary, we obtain that $\mathbf{R}^n$ is connected.

Example 2.2.8. Normed vector spaces are path-connected. Indeed, for given x and y, there exists $r = \max\{d(x,0), d(y,0)\}$ such that both x and y belongs to the closed ball $\overline{B}_r(0)$.

If $n \neq 1$, then $\mathbf{R}^n \setminus \{0\}$ is also path-connected. If x and y are not colinear, then one can take the straight line segment between the two. If x and y are colinear, it is easy to modify the straight line segment between them in order to avoid 0, see Figure 2.5.

Finally, $\mathbf{R} \setminus \{0\}$ is not path-connected. Indeed, there is no path from -1 to 1, as such a path should take the value 0 by the intermediate value theorem.

As announced, path-connectedness is a stronger notion than connectedness.

Theorem 2.2.9.

A path-connected metric space X is connected.

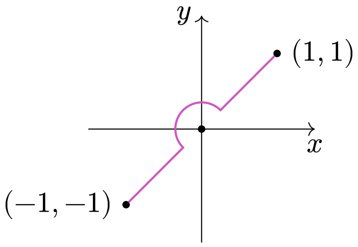


Figure 2.5.: A path from $(-1,-1)$ to $(1,1)$ that avoids 0.

Proof. Suppose by contradiction that X is path-connected but not connected, and let $f\colon X \to \{0,1\}$ be a surjective continuous function where $\{0,1\}$ is endowed with the discrete metric. Let $x, y \in X$ with $f(x) = 0$ and $f(y) = 1$. Let $g\colon I \to X$ be a path from x to y. Then $f \circ g\colon I \to D$ is a continuous and surjective function contradicting the connectedness of I. So X is connected. ☐

Intervals in $\mathbf{R}$ are path-connected (Lemma 2.2.7). It thus follows from Proposition 2.1.9 that a subset of $\mathbf{R}$ is connected if and only if it is path connected. However, in general path-connectedness is a strictly stronger notion than connectedness as demonstrated by the next example.

Example 2.2.10. The topologist's sine curve $G \cup J$ from Example 2.1.19 is connected. We will show that it is not path-connected by proving that any continuous path $f\colon I \to G \cup J$ which begins at $f(0) = (0,0) \in J \subseteq \mathbf{R}^2$ will never leave J and therefore cannot reach any point of G.

Let $i\colon G \cup J \to \mathbf{R}^2$ be the inclusion and $p_j\colon \mathbf{R}^2 \to \mathbf{R}$ be the projection maps for $j = 1, 2$. Then

$$f_1 := p_1 \circ i \circ f\colon I \to \mathbf{R} \quad \text{and} \quad f_2 := p_2 \circ i \circ f\colon I \to \mathbf{R}$$

are both continuous as they are the composition of continuous functions. We also have $f_j(0) = 0$ for $j = 1, 2$.

Let $\alpha := \sup\{x \in [0,1] \mid f_1([0,x]) = 0\}$. We know that $\alpha \in [0,1]$. If $\alpha = 1$, then $f_1(x) = 0$, for all $x \in [0,1]$, which implies that $f(x) \in J$, for all $x \in [0,1]$, which is what we want to prove.

Suppose by contradiction that $\alpha < 1$. Note that $f_1(\alpha) = 0$ by continuity and $f_2(\alpha) \in [-1,1]$. By the continuity of f_2 at α, there exists $\delta > 0$ such that

$$|f_2(x) - f_2(\alpha)| < 1, \quad \text{for all } x \in (\alpha - \delta, \alpha + \delta) \cap [0,1]. \tag{2.1}$$

By definition of α, there exists $x_0 \in (\alpha, \alpha + \delta) \cap [0,1]$ such that $f_1(x_0) > 0$. By the Intermediate Value Theorem, $[0, f_1(x_0)] \subseteq f_1([\alpha, x_0])$.

Suppose that $f_2(\alpha) \geqslant 0$. Then $f_2(\alpha) - 1 \in [-1, 0]$. So there exists $t \in (0, f_1(x_0)]$ such that $\sin(\frac{1}{t}) = f_2(\alpha) - 1$. Also, there exists $\tilde{x} \in [\alpha, x_0] \subseteq (\alpha - \delta, \alpha + \delta) \cap [0,1]$ such that $t = f_1(\tilde{x}) > 0$. So $f_2(\tilde{x}) = \sin(\frac{1}{f_1(\tilde{x})}) = \sin(\frac{1}{t}) = f_2(\alpha) - 1$. This implies that $|f_2(\tilde{x}) - f_2(\alpha)| = 1$ which contradicts Equation (2.1).

The case for $f_2(\alpha) \leqslant 0$ follows similarly by observing that $f_2(\alpha) + 1 \in [0, 1]$.

We have seen that while a subset W of $\mathbf{R}$ is connected if and only if it is path-connected, this is not true anymore for subsets of $\mathbf{R}^2$. This is however still true if we restrict ourselves to open subsets of $\mathbf{R}^n$.

Proposition 2.2.11. *A connected open subset U of a normed vector space is path-connected. In particular, connected open subsets of $\mathbf{R}^n$ are path-connected.*

Proof. Let $u \in U$ and $A \subseteq U$ be the subset of points that can be joined to u by a path in U, and let $B = U \setminus A$. We shall prove that if B were non-empty, then $A \mid B$ would partition U, contradicting its connectedness.

First we show that A is open. Let $x \in A$. Then there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq U$. Given any $y \in B_\varepsilon(x)$, there is a path (straight line segment) g in $B_\varepsilon(x)$ from x to y. But since $x \in A$, there is a path f in U joining u to x. The path h, as given in Lemma 2.2.3, is a path in U from u to y. So $y \in A$ and $B_\varepsilon(x) \subseteq A$ and A is open.

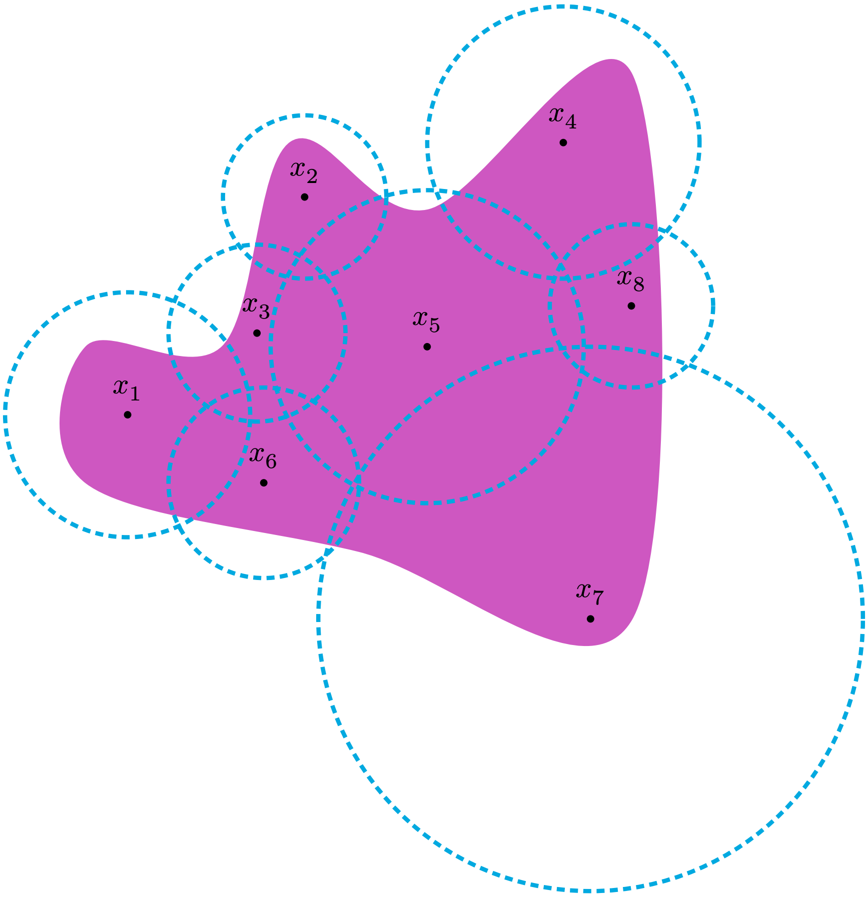
We now show that B is open. So let $b \in B \subseteq U$ and let $\varepsilon >$ be such that $B_\varepsilon(b) \subseteq U$. The ball $B_\varepsilon(b)$ cannot any element a of A, as otherwise we would have a path from b to a and hence from b to u, which is absurd. We conclude that $B_\varepsilon(b) \subseteq U \setminus A$ and thus that B is open.

Finally, $A \cap B = \varnothing$ and $A \cup B = U$ so, by the connectedness of U, $B = \varnothing$. ☐

One can show that the topologist's sine curve from Example 2.1.19 is closed. Proposition 2.2.11 is thus not true for closed subsets.

CHAPTER 3.

COMPACTNESS



3.1. Definition and First Properties

The aim of this chapter is to study compact (sub)spaces, which are a particularly well-behaved kind of (sub)spaces. Compact subspaces generalise finite subspaces and are, in some sense, spaces that are "manageable/approximable by finite data". In Euclidean spaces, compact subspaces are exactly the closed bounded subsets of $\mathbf{R}^n$. However, taking this as a definition of compactness will fail as in a discrete metric space (X, d) every subset is both closed and bounded but not necessarily "well-behaved". We hence need a more precise definition of compactness.

Definition 3.1.1 (Cover)**.**

Let (X, d) be a metric space and let $W \subseteq X$ be a subset. A **cover** of W is a collection $\mathcal{U} := \{U_i \subseteq X \mid i \in I\}$ of subsets such that

$$W \subseteq \bigcup_{i \in I} U_i;$$

a **subcover** of $\mathcal{U}$ is a subcollection $\{U_i \subseteq X \mid i \in J\}$ for some $J \subseteq I$ which also covers W. If I is a finite set, then $\mathcal{U}$ is called a **finite cover**. If every U_i is open in X, then $\mathcal{U}$ is called an **open cover** of W.

Remark 3.1.2.

Remember that the above definition also holds for $W = X$.

See the chapter illustration on the previous page for an illustration of a finite open cover by open balls in $\mathbf{R}^2$.

It is easy to come up with example of covers.

Example 3.1.3. Let $W \subseteq X$ be a subset of a metric space and let $w \in W$. Then the following are covers of W:

$$\mathcal{U}_1 = \{X\}, \quad \mathcal{U}_2 = \{B_1(w) \mid w \in W\} \quad \text{and for any } w_0 \in W : \mathcal{U}_3 = \{B_n(w_0) \mid n \in \mathbf{N}\}.$$

This shows that the notion of (finite) cover is not interesting in itself. What is interesting is to be able to extract a finite subcover from an arbitrary cover. This is exactly the definition of compactness.

Definition 3.1.4 (Compact)**.**

A subset $W \subseteq X$ of a metric space is **compact** if every open cover of W has a finite subcover.

A metric space (X, d) is **compact** if X is a compact subset of itself.

Remark 3.1.5.

Let d and d' be two topologically equivalent metric on a set X. Since d and d' have the same open subsets, a subset $W \subseteq X$ is compact in (X, d) if and only if it is compact in (X, d'). In other words, compactness is a topological property.

Remark 3.1.6.

We defined compactness using open subsets. Using the duality between open and closed subsets, one can obtain a characterisation of compact subsets in terms of closed subsets. It is an interesting exercise to try to write down such a characterisation. See Tutorial 8, Question 1 for the details.

To better understand the definition of compactness, let us see how it can fail. Firstly, W can fail to be compact because it is not closed (see Proposition 3.1.11).

Example 3.1.7. The open interval $(0, 1) \subseteq \mathbf{R}$ is not compact since the collection

$$\mathcal{U} := \left\{ U_n = \left(\frac{1}{n}, 1\right) \,\middle|\, n \geqslant 2 \right\}$$

is an open cover of $(0, 1)$ but any finite subcollection $\{U_{n_1}, \dots, U_{n_j}\}$ contains a largest interval, let's say $(\frac{1}{n_k}, 1)$, so it fails to be a subcover since $\frac{1}{2n_k}$ does not lie in U_{n_i} for any $1 \leqslant i \leqslant j$.

Secondly, W can fail to be compact because is not bounded (see Proposition 3.1.11).

Example 3.1.8. The collection $\mathcal{U} := \{U_n = (-n, n) \mid n \in \mathbf{N}\}$ is an open cover of $\mathbf{R}$, but any finite subcollection has a largest open interval and therefore $\mathbf{R}$ is not compact.

To go further

Even if $\mathbf{R}$ is not compact (it is "too big") it can still be considered as a "nice" space as it is **locally compact** ("locally small"). That is, every point $x \in \mathbf{R}$ is contained in a compact subset $x \in K \subseteq \mathbf{R}$. The normed vector space of real polynomials $\mathscr{P}(\mathbf{R})$ is not locally compact.

The next example shows that even if W is both closed and bounded, it is not necessarily compact.

Example 3.1.9. Let (X, d_0) be a discrete metric space. Then all subsets of X are closed and bounded, but a subset of X is compact if and only if it is finite (Tutorial 7, Question 6).

Compact (sub)spaces are small enough, or manageable, in some sense. In particular, finite subsets are compact.

Proposition 3.1.10. *If W is finite, then it is compact.*

Proof. Suppose that $\mathcal{U} = \{U_i \mid i \in I\}$ is an open cover of $W = \{w_1, \dots, w_N\}$. For each $1 \leqslant k \leqslant N$ it is possible to choose an i_k such that $w_k \in U_{i_k}$. Therefore, $W \subseteq \bigcup_{k=1}^{N} U_{i_k}$, and $\{U_{i_k} \mid 1 \leqslant k \leqslant N\}$ is a finite subcover. ☐

It is easy to come with examples of non-compact spaces. Indeed, it is enough to exhibit one cover that does not admit a finite subcover. Showing that a subset is compact is usually more difficult as it requires to show that all covers admit a finite subcover. In particular, showing that $[0,1]$ is compact already requires some work, see Theorem 3.1.14.

The converse of Proposition 3.1.10 is true for discrete metric spaces. Namely, a discrete metric space is compact if and only if it is finite, see Example 3.1.9. It however fails in general, as for example $[0,1]$ is compact but not finite.

Compactness implies boundedness as well as closedness.

Proposition 3.1.11. *If $W \subseteq X$ is compact subspace of a metric space, then it is both bounded and closed*

Proof. We first shows that W is bounded. Choose $w \in W$ and consider the collection $\{B_n(w) \mid n \geqslant 1\}$. This is an open covering of W since $d(w,y)$ is finite for all $y \in W$. So there must exist a finite subcover $\{B_{n_i}(w) \mid 1 \leqslant i \leqslant k\}$. This implies that $W \subseteq B_r(w)$ for $r = \max\{n_1, \dots, n_k\}$. So W is bounded.

We now show that W is closed. Let $x \in X \setminus W$. We will show that $X \setminus W$ is open by proving that $x \in (X \setminus W)^\circ$. Define $d(w) := \frac{d(w,x)}{2}$, for each $w \in W$. Therefore, the open balls $B_{d(w)}(w)$ and $B_{d(w)}(x)$ are disjoint. The collection $\{B_{d(w)}(w) \mid w \in W\}$ is an open cover of W and therefore has a finite subcover which we can write as $\{B_{d(w_i)}(w_i) \mid 1 \leqslant i \leqslant k\}$ for some points $w_1, \dots, w_k \in W$. Let $\varepsilon = \min\{d(w_1), \dots, d(w_k)\}$. Then $B_\varepsilon(x)$ is disjoint from W and so $x \in (X \setminus W)^\circ$ which implies that W is closed in X. See Figure 3.1 for an illustration of the proof. ☐

If A is a compact subset of a metric space and $B \subseteq A$, then B is not necessarily compact. For example, one can take $(0,1) \subseteq [0,1]$ in $\mathbf{R}$. However, compactness passes to closed subsets:

Proposition 3.1.12. *Let W be a compact subset of a metric space (X,d) and let $V \subseteq W$ be a close subset. Then V is compact.*

Proof. Let $\mathcal{U}$ be an open cover of V. Since V is closed, $W \setminus V$ is open. If we add the open subset $W \setminus V$ to $\mathcal{U}$ we get an open cover $\mathcal{W}$ of W. By the compactness of W we get a finite subcover $\tilde{\mathcal{W}}$ of $\mathcal{W}$, which is also a finite cover of V. If $W \setminus V$ is in $\tilde{\mathcal{W}}$, then we can remove it to obtain a finite subcover of $\mathcal{U}$ for V. ☐

As for closed subsets, the collection of compact subsets is closed under finite unions and arbitrary intersections.

Lemma 3.1.13. *If A and B are two compact subsets of a metric space X, then both $A \cap B$ and $A \cup B$ are compact. Moreover, if $I \neq \varnothing$ and all the A_i are compact, then $\bigcap_{i \in I} A_i$ is also compact.*

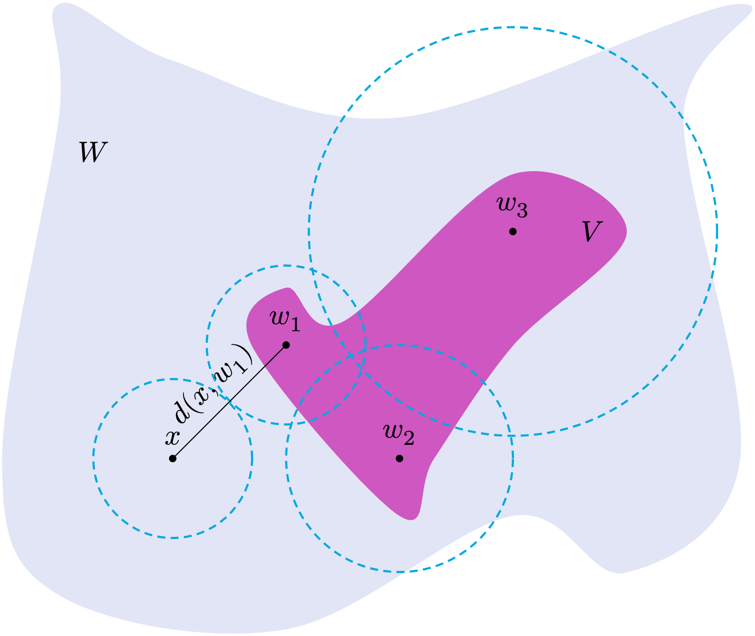


Figure 3.1.: If W is compact and $V \subseteq W$ is closed, then for every $x \in W \setminus V$ one can find ε such that $B_\varepsilon(x) \subseteq W \setminus V$.

Proof. The proof for finite union and intersections if left to the reader (see Tutorial 7, Question 7).

Let us do the case of an arbitrary intersection of compact subspaces $A := \bigcap_{i \in I} A_i$. Since all the A_i are compact, they are also closed and so A is a closed subset of the compact subset A_i (for any $i \in I$). □

The converse of Proposition 3.1.11 does not hold in general, as demonstrated by discrete metric spaces, see Example 3.1.9. However, the converse of Proposition 3.1.11 is true for subsets of $\mathbf{R}^n$. This is the subject of the Heine–Borel theorem (Theorem 3.3.3) which will we prove in the next section. Before proving the full version of the Heine–Borel theorem, we start with a simple case: intervals in $\mathbf{R}$.

Theorem 3.1.14 (Heine–Borel[1] (1-dimensional))**.**

Any closed bounded interval $[a, b]$ *in* $\mathbf{R}$ *is compact.*

Proof. Let $\mathcal{U}$ be an open cover of $[a, b]$, and define

$$G := \{x \in \mathbf{R} \mid x \geqslant a \text{ and } [a, x] \text{ is covered by a finite subcollection of } \mathcal{U}\}.$$

We need to prove that $b \in G$.

---

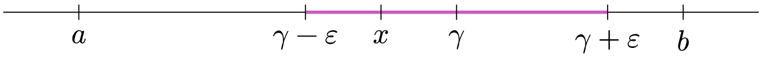
[1] Heinrich Eduard Heine (1821–1881) and Félix Édouard Justin Émile Borel (1871–1956).

First note that if $x \in G$, then $y \in G$ for all $a \leqslant y \leqslant x$.

We see that $a \in G$ since a must belong to some $U \in \mathcal{U}$, since $\mathcal{U}$ covers $[a, b]$. Since U is open, there exists $\delta > 0$ such that $[a, a + \delta) \subseteq U$ which implies that $[a, a + \delta) \subseteq G$ also.

Now we know that $G \neq \varnothing$, either G is bounded above or not. If G is not bounded above, then there exists $c \in G$ with $c > b$ which implies that $b \in G$. If G is bounded above, then its supremum $\gamma := \sup G$ exists. Note that $\gamma > a$. If $\gamma > b$, then there is some $c \in G$, with $c > b$, which implies that $b \in G$.

Now suppose that $\gamma \leqslant b$. There exists some $U_\gamma \in \mathcal{U}$ with $\gamma \in U_\gamma$, and some $\varepsilon > 0$ with $(\gamma - \varepsilon, \gamma + \varepsilon) \subseteq U_\gamma$. Moreover, some $x \in G$ must satisfy $x > a$ and $x > \gamma - \varepsilon$, by the definition of the supremum.



But $x \in G$ implies that $[a, x]$ admits a finite subcover of $\mathcal{U}$, to which we may add U_γ and cover $[a, \gamma + \frac{\varepsilon}{2}]$ by a finite subcover of $\mathcal{U}$. Thus, $\gamma + \frac{\varepsilon}{2}$ lies in G, which contradicts the assumption that $\gamma = \sup G$. ☐

3.2. Continuous Functions on Compact Spaces

In this section we will see that continuous functions with compact domain are particularly well-behaved. We start by showing that the image of a compact space is compact.

Proposition 3.2.1. *If $f\colon X \to Y$ is continuous and X is compact, then the image $f(X) \subseteq Y$ is also compact.*

Proof. Let $\mathcal{U}$ be an open cover for $f(X)$. Since f is continuous, $f^{-1}(U)$ is open in X, for each $U \in \mathcal{U}$. The collection $\{f^{-1}(U) \mid U \in \mathcal{U}\}$ is an open cover for X since

$$x \in X \implies f(x) \in U \text{ for some } U \in \mathcal{U} \implies x \in f^{-1}(U).$$

Since X is compact there exists a finite subcover $\{f^{-1}(U_1), \dots, f^{-1}(U_n)\}$ of X and the collection $\{U_1, \dots, U_n\}$ is a finite subcover of $f(X)$ since

$$X \subseteq \bigcup_{i=1}^{n} f^{-1}(U_i) \implies f(X) \subseteq f\Big(\bigcup_{i=1}^{n} f^{-1}(U_i)\Big) = \bigcup_{i=1}^{n} f(f^{-1}(U_i)) = \bigcup_{i=1}^{n} U_i. \qquad \square$$

Example 3.2.2. The map $e\colon \mathbf{R} \to \mathbf{R}^2$ given by $e(x) = (\cos x, \sin x)$ is continuous and maps the closed interval $[0, 2\pi]$ onto the circle S^1 of radius 1 around the origin. Therefore, S^1 is a compact subset of $\mathbf{R}^2$.

As a direct corollary of Proposition 3.2.1 we obtain:

Corollary 3.2.3. *Compactness is a topological property. That is, if two spaces X and Y are homeomorphic, then X is compact if and only if Y is.*

Since compact subsets are bounded, we also have:

Corollary 3.2.4. *Any continuous function from a compact metric space is bounded.*

Proof. This follows from Propositions 3.2.1 and 3.1.11. ☐

In general, if a function $f\colon X \to \mathbf{R}$ is bounded, this means that $f(X)$ is a bounded subset of $\mathbf{R}$, which implies that $\sup f(X)$ and $\inf f(X)$ exist. These are bounds of f on X but they may not be in $f(X)$ themselves, that is, there does not necessarily exist $m, M \in X$ such that $f(M) = \sup f(X)$ and $f(m) = \inf f(X)$. If they are in X, then we say that f **attains its bounds** on X.

Corollary 3.2.5. *If $f\colon X \to \mathbf{R}$ is continuous and X is compact and non-empty, then f attains its bounds on X.*

Proof. By Corollary 3.2.4 $f(X)$ is bounded which implies that $\sup f(X)$ and $\inf f(X)$ exist and are bounds for $f(X)$. Also, $\sup f(X)$ and $\inf f(X)$ are elements of $\overline{f(X)}$ (Tutorial 8, Question 6). Since $f(X)$ is compact it is closed, which implies that $\sup f(X), \inf f(X) \in f(X) = \overline{f(X)}$. ☐

Theorem 3.2.6.

Suppose that $f\colon X \to Y$ is a continuous bijection where X is compact. Then f is a homeomorphism.

Proof. We need to show that $g := f^{-1}\colon Y \to X$ is continuous. We shall prove that if V is closed in X, then $g^{-1}(V)$ is closed in Y, which will imply that g is continuous.

Let V be closed in X. Since X is compact, V is compact by Proposition 3.1.12. So $f(V)$ is compact by Proposition 3.2.1, which implies that $f(V)$ is closed in Y. But $f(V) = g^{-1}(V)$ so $g^{-1}(V)$ is closed in Y. ☐

Corollary 3.2.7. *If $f\colon X \to Y$ is a continuous injective function where X is compact, then X is homeomorphic to $f(X)$.*

Recall that a function $f\colon [a, b] \to \mathbf{R}$ is strictly monotonic if it is either strictly increasing (if $x < y$, then $f(x) < f(y)$) or strictly decreasing (if $x < y$, then $f(x) > f(y)$).

Example 3.2.8. If $f\colon [a, b] \to \mathbf{R}$ is a continuous strictly monotonic function, then it has a continuous inverse $f^{-1}\colon f([a, b]) \to [a, b]$.

3.3. The Heine–Borel Theorem

The aim of this section is to prove the Heine–Borel Theorem, which generalises Theorem 3.1.14 to closed bounded subsets of $\mathbf{R}^n$.

We start by using results from the previous section to show that compactness is preserved by taking products.

Theorem 3.3.1.

A product $X \times Y$ of metric spaces is compact if and only if both X and Y are compact.

Proof. "$\Rightarrow$" The projection maps $p_X : X \times Y \to X$ and $p_Y : X \times Y \to Y$ are continuous and so X and Y are both compact as the continuous images of $X \times Y$ by Proposition 3.2.1.

"$\Leftarrow$" Suppose X and Y are both compact and let $\mathcal{U}$ be an open cover of $X \times Y$. We shall say that $W \subseteq X$ is *good* if $W \times Y$ is covered by a finite subcollection of $\mathcal{U}$. We want to prove that X is good. In order to do that, we will do two intermediate steps. Firstly, we will prove that every x belongs to some good open subset of U. Then we will show that the union of finitely many good subsets is still good. Finally, we will use these two intermediate results to show that X is good.

Step 1. *For all $x \in X$ there exists an open subset $U(x) \subseteq X$ such that $x \in U(x)$ and $U(x)$ is good.*

Proof. Fix $x \in X$. Then for every $y \in Y$, $(x, y) \in W(y)$, for some $W(y) \in \mathcal{U}$ since $\mathcal{U}$ covers $X \times Y$. We now prove an intermediary lemma for those $W(y)$.

Lemma 3.3.2. *There are open subsets $U(y) \subseteq X$ and $V(y) \subseteq Y$ such that*

$$(x, y) \in U(y) \times V(y) \subseteq W(y).$$

Proof. Since $W(y)$ is open in $X \times Y$, there exists $\varepsilon > 0$ such that $B_\varepsilon((x, y)) \subseteq W(y)$. So far we have not specified exactly the metric we are considering on $X \times Y$ but we know that the metrics d_a, d_b, d_c on $X \times Y$ (from Definition 1.6.10) are all topologically equivalent and so we can choose which one we take without loss of generality. For this proof we will take d_c. Then

$$B_\varepsilon((x, y); d_c) = \{(x', y') \in X \times Y \mid \max\{d_X(x, x'), d_Y(y, y') < r\}\},$$

where d_X is the metric on X and d_Y is the metric on Y. It is then a fairly easy exercise to check that

$$B_\varepsilon((x, y); d_c) = B_\varepsilon(x; d_X) \times B_\varepsilon(y; d_Y).$$

We can then take $U(y) = B_\varepsilon(x; d_X)$ and $V(y) = B_\varepsilon(y; d_Y)$ to complete the proof. ■

The collection $\{V(y) \mid y \in Y\}$ is an open cover of Y and so has a finite subcover $\{V(y_1), V(y_2), \dots, V(y_r)\}$. Let

$$U(x) := U(y_1) \cap U(y_2) \cap \dots \cap U(y_r).$$

Then for each $i \in \{1, \dots, r\}$ we have

$$U(x) \times Y = U(x) \times \bigcup_{i=1}^{r} V(y_i) = \bigcup_{i=1}^{r} (U(x) \times V(y_i)) \subseteq \bigcup_{i=1}^{r} W(y_i),$$

which implies that $U(x)$ is good. Also, $x \in U(x)$ and $U(x)$ is open in X as a finite intersection of open sets. ■

Step 2. *If $W_1, W_2, \dots, W_r \subseteq X$ are good, then $\bigcup_{i=1}^r W_i$ is good.*

Proof. Given any $i \in \{1, \dots, r\}$ there is a finite subcollection say $\mathcal{U}_i$ of $\mathcal{U}$ which covers $W_i \times Y$. Then $\bigcup_{i=1}^r W_i \times Y$ is covered by the finite subcollection

$$\mathcal{U}_1 \cup \mathcal{U}_2 \cup \dots \cup \mathcal{U}_r$$

of $\mathcal{U}$. ■

We now use the above two steps to show that X is good. The collection $\{U(x) \mid x \in X\}$ is an open cover of X and therefore has a finite subcover $\{U(x_1), U(x_2), \dots, U(x_s)\}$. By Step 1, each $U(x_i)$ is good, for $i = 1, \dots, s$, and so $X = \bigcup_{i=1}^s U(x_i)$ is good by Step 2. □

Theorem 3.3.3 (Heine–Borel).

A subspace $A \subseteq \mathbf{R}^n$ is compact if and only if it is closed and bounded.

We can immediately prove this in dimension 1 as if $A \subseteq \mathbf{R}$ is closed and bounded, then $A \subseteq [a, b]$ for some closed interval (since A is bounded), and since we know that $[a, b]$ is compact A is a closed subset of a compact set and therefore compact.

Proof. "$\Rightarrow$" This is Proposition 3.1.11.

"$\Leftarrow$" Let $A \subseteq \mathbf{R}^n$ be bounded and closed. Since A is bounded, $A \subseteq [a, b]^n$ for some $a, b \in \mathbf{R}$ (Tutorial 5, Question 9). Since $[a, b]$ is compact (in $\mathbf{R}$) so is $[a, b]^n$ by induction on n and Theorem 3.3.1. Also, A is closed in $[a, b]^n$. Therefore, A is compact as a closed subspace of a compact space by Proposition 3.1.12. □

Observe that by Remark 3.1.5, $A \subseteq \mathbf{R}^n$ is compact for the Euclidean metric if and only if it is compact for d_p for any $p \in [1, +\infty) \cup \{\infty\}$.

3.4. Equivalence of norms on $\mathbf{R}^n$

This section will be devoted to proving the following theorem:

Theorem 3.4.1.

Any two norms on $\mathbf{R}^n$ give Lipschitz equivalent metrics.

In order to prove this we will show that any norm $\|\cdot\|$ is equivalent to the taxicab norm $\|\cdot\|_1$ on $\mathbf{R}^n$, i.e. there exist $h, k \in (0, +\infty)$ such that

$$h\|x\|_1 \leqslant \|x\| \leqslant k\|x\|_1, \quad \forall x \in \mathbf{R}^n. \tag{3.1}$$

This will prove that any metric coming from a norm on $\mathbf{R}^n$ is Lipschitz equivalent to the taxicab metric on $\mathbf{R}^n$ and therefore, since Lipschitz equivalence is an equivalence relation, that all metrics on $\mathbf{R}^n$ coming from norms are Lipschitz equivalent.

Let $\boldsymbol{e}_i \in \mathbf{R}^n$ be the standard i^{th} basis vector, i.e. its i^{th} coordintate is 1 and the rest are 0.

Let $\|\cdot\|$ be a norm on $\mathbf{R}^n$ and let $k := \max\{\|\boldsymbol{e}_1\|, \dots, \|\boldsymbol{e}_n\|\}$. Then for any $x \in \mathbf{R}^n$,

$$\|x\| = \left\|\sum_{i=1}^{n} x_i \boldsymbol{e}_i\right\| \leqslant \sum_{i=1}^{n} |x_i| \cdot \|\boldsymbol{e}_i\| \leqslant k \sum_{i=1}^{n} |x_i| = k\|x\|_1, \tag{3.2}$$

which gives us the second inequality in Equation (3.1).

Let us recall the following easy consequence of the triangle inequality for norms (see Tutorial 2, Question 7).

Lemma 3.4.2. *For any norm* $\|\cdot\|$ *we have* $\|x \pm y\| \geqslant |\|x\| - \|y\||$.

We will use the above Lemma to show that any norm on $\mathbf{R}^n$ is continuous with respect to $\|\cdot\|_1$.

Lemma 3.4.3. *The function* $\|\cdot\| \colon \mathbf{R}^n \to \mathbf{R}$ *is continuous with respect to the taxicab norms. That is,* $\|\cdot\|$ *is a continuous function from* $(\mathbf{R}^n, d_1)$ *to* $(\mathbf{R}, d_1)$.

Proof. Let $x \in \mathbf{R}^n$ and $\varepsilon > 0$. Then take $\delta = \frac{\varepsilon}{k}$, where k is as in Equation (3.2). If $\|x - y\|_1 < \delta$, then

$$|\|x\| - \|y\|| \leqslant \|x - y\| \leqslant k\|x - y\|_1 < \varepsilon. \qquad \square$$

Now $S := \{x \in \mathbf{R}^n \mid \|x\|_1 = 1\} = \partial B_1(0; d_1)$ is a compact set by the Heine–Borel Theorem as it is closed (being the boundary of a subset) and bounded (being a subset of $\overline{B}_1(0; d_1)$). Therefore, the continuous function $\|\cdot\| \colon \mathbf{R}^n \to \mathbf{R}$, when restricted to S, attains its lower bound h on S and since $0 \notin S$, we have $h > 0$. Hence,

$$\|x\| \geqslant h > 0, \quad \text{for all } x \in S.$$

Lemma 3.4.4. *For every* $x \in \mathbf{R}^n$, $\|x\| \geqslant h\|x\|_1$, *where* $h = \inf\{\|x\| \mid \|x\|_1 = 1\}$.

Proof. If $x = 0$, then the statement holds. Suppose that $x \neq 0$ and set $\lambda := \frac{1}{\|x\|_1} > 0$. Then

$$\|\lambda x\|_1 = \lambda\|x\|_1 = 1 \implies \lambda x \in S.$$

So

$$h \leqslant \|\lambda x\| = \lambda\|x\| \implies \|x\| \geqslant h\|x\|_1. \qquad \square$$

We have therefore proven the first inequality in Equation (3.1) and thus Theorem 3.4.1.

3.5. Uniform Continuity

The following definition strengthen the notion of continuity (Definition 1.1.16).

Definition 3.5.1 (Uniform Continuity)**.**

A function $f\colon (X, d_X) \to (Y, d_Y)$ is **uniformly continuous** if

$$\forall \varepsilon > 0, \exists \delta > 0 \quad \text{such that} \quad d_X(x,y) < \delta \implies d_Y(f(x), f(y)) < \varepsilon.$$

Remark 3.5.2.

In the ordinary definition of continuity, at a point x in the domain, δ depends on x as well as on ε. In the definition of uniform continuity, δ only depends on ε. This is better seen on the fully formal definitions of continuity and uniform continuity:

$$\forall x, \forall \varepsilon > 0, \exists \delta > 0, \forall y : d_X(x, a) < \delta \implies d_Y(f(x), f(a)) < \varepsilon$$

versus

$$\forall \varepsilon > 0, \exists \delta > 0, \forall x, \forall y : d_X(x, a) < \delta \implies d_Y(f(x), f(a)) < \varepsilon.$$

Ordinary continuity is a local property whereas uniform continuity is a global property.

It is clear that uniform continuity implies continuity. The converse does not always hold.

Example 3.5.3. The function $f\colon (0,1) \to \mathbf{R}$, given by $f(x) = \frac{1}{x}$, is continuous but not uniformly continuous since, for $\varepsilon = 1$, and any $\delta > 0$, if we let $x = \min\{\frac{1}{2}, \delta\}$ and $y = \frac{x}{2}$, then we have $|x - y| = \frac{x}{2} < \delta$ but $|f(x) - f(y)| = |\frac{1}{x} - \frac{2}{x}| = \frac{1}{x} \geqslant 1 = \varepsilon$.

The good news it that when the domain of a function is compact, then continuity and uniform continuity are equivalent.

Proposition 3.5.4. *Let $f\colon (X, d_X) \to (Y, d_Y)$ be continuous with X compact. Then f is uniformly continuous.*

Proof. Let $\varepsilon > 0$. Then for each $x \in X$, since f is continuous at x, there exists $\delta(x) > 0$ such that

$$d_X(x,y) < 2\delta(x) \implies d_Y(f(x), f(y)) < \frac{\varepsilon}{2}.$$

The collection $\{B_{\delta(x)}(x) \mid x \in X\}$ is an open cover of X and so, since X is compact, has a finite subcover $\{B_{\delta(x_1)}(x_1), \dots, B_{\delta(x_n)}(x_n)\}$, for some points $x_1, \dots, x_n \in X$. Let $\delta = \min\{\delta(x_1), \dots, \delta(x_n)\}$. Then for $x, y \in X$ with $d_X(x,y) < \delta$ we have

1. there exists $1 \leqslant i \leqslant n$ such that $d_X(x, x_i) < \delta(x_i)$;
2. $d_X(y, x_i) \leqslant d_X(y, x) + d_X(x, x_i) < \delta + \delta(x_i) \leqslant 2\delta(x_i)$.

Now 1 implies that $d_Y(f(x), f(x_i)) < \frac{\varepsilon}{2}$ and 2 implies that $d_Y(f(y), f(x_i)) < \frac{\varepsilon}{2}$ so

$$d_Y(f(x), f(y)) \leqslant d_Y(f(x), f(x_i)) + d_Y(f(x_i), f(y)) < \varepsilon. \qquad \square$$

Of course a function does not need to have a compact domain to be uniformly continuous; any identity function $\mathrm{Id}_X \colon (X,d) \to (X,d)$ is uniformly continuous.

Definition 3.5.5 (Uniformly equivalent metrics)**.**

Two metrics d and d' on a set X are **uniformly equivalent** if both $\mathrm{Id}\colon (X,d) \to (X,d')$ and $\mathrm{Id}\colon (X,d') \to (X,d)$ are uniformly continuous.

So if (X,d) is compact and is topologically equivalent to (X,d'), then d and d' are uniformly equivalent.

Uniform equivalence of metrics is an equivalence relation on the set of metrics on set X. It directly follows from the definition that it is finer than topological equivalence. Uniform equivalence is a stricly finer equivalence relation than topological equivalence as demonstrated in Tutorial 8 Question 4. It is a strictly coarser equivalence relation than Lipschitz equivalence, see Tutorial 8 Questions 2 and 3

Definition 3.5.6 (Uniform Equivalence)**.**

Given metric spaces (X,d_X) and (Y,d_Y), a function $f\colon X \to Y$ is a **uniform equivalence** if f is a bijection where both f and f^{-1} are uniformly continuous. Two metric spaces (X,d_X) and (Y,d_Y) are **uniformly equivalent** if there exists a uniform equivalence $f\colon X \to Y$.

Uniform equivalence is an equivalence relation on the class of metric spaces which is finer than topological equivalence. We hence have the following sequence of strict implications, which is a detailed version of the sequence of page 26.

$$\text{isometry} \implies \text{Lipschitz equivalence} \implies \text{uniform equivalence} \implies \text{homeomorphism.}$$

3.6. Sequential Compactness

In this section, we introduce a variation of the notion of compactness, which for metric spaces will turn to be equivalent to compactness.

Definition 3.6.1 (Sequentially Compact)**.**

A subset $W \subseteq (X,d)$ is **sequentially compact** if every sequence (w_n) in W has a subsequence that converges to a point in W.

Example 3.6.2. The real line $\mathbf{R}$ is not sequentially compact since the sequence $x_n = n$, for all $n \in \mathbf{N}$, does not have a convergent subsequence.

Lemma 3.6.3. *Let (x_n) be a sequence in a metric space X and let $x \in X$. If, for every $\varepsilon > 0$, $B_\varepsilon(x)$ contains x_n for infinitely many values of $n \in \mathbf{N}$, then (x_n) contains a subsequence that converges to $x \in X$.*

Proof. Let n_1 be such that $x_{n_1} \in B_1(x)$ and proceed inductively by choosing integers $n_1 < \cdots < n_r$ such that $x_{n_j} \in B_{\frac{1}{j}}(x)$, for each $1 \leqslant j \leqslant r$. Since $B_{\frac{1}{r+1}}(x)$ contains x_n for infinitely many n, there are infinitely many that satisfy $n_r < n$. Choose n_{r+1} to be one of these. Then $x_{n_r} \in B_{\frac{1}{r}}(x)$ for every $r \in \mathbf{N}$ and $x_{n_r} \to x$, as $r \to \infty$, in X. ☐

Example 3.6.4. In $\mathbf{R}$, the sequence $x_n = \begin{cases} m, & \text{if } n = 2m-1; \\ \frac{1}{m}, & \text{if } n = 2m, \end{cases}$ has a convergent subsequence $x_{2m} \to 0$, and $B_\varepsilon(0)$ contains every x_{2m} for $m > \frac{1}{\varepsilon}$, of which there are infinitely many.

The contrapositive of Lemma 3.6.3 is often useful.

Lemma 3.6.5. *Suppose that (x_n) contains no convergent subsequence. Then, for every $x \in X$, there exists $\varepsilon(x) > 0$, such that $B_{\varepsilon(x)}(x)$ contains x_n for at most finitely many $n \in \mathbf{N}$.*

Example 3.6.6.

1. In $\mathbf{R}$, the sequence $(x_n = n)$ has no convergent subsequences. Given any $x \in \mathbf{R}$, the $\varepsilon(x)$ of Lemma 3.6.5 can be taken to be $\frac{1}{2}$, in which case $B_{\frac{1}{2}}(x)$ contains at most one element of x_n.
2. In the interval $(0, 1] \subseteq \mathbf{R}$, the sequence $(\frac{1}{n})_{n \geqslant 1}$ contains no convergent subsequences. Any $x \in (0, 1]$ determines a unique integer $N(x)$ for which

$$\frac{1}{(N(x)+1)} \leqslant x \leqslant \frac{1}{N(x)}.$$

Choosing $\varepsilon(x) := \frac{1}{(N(x)+1)} - \frac{1}{(N(x)+2)}$ ensures that $B_{\varepsilon(x)}(x)$ contains at most one element of $(\frac{1}{n})$.

In a metric space, sequential compactness implies compactness as demonstrated below.

Proposition 3.6.7. *If a subspace $W \subseteq X$ of a metric space is compact, then it is sequentially compact.*

Proof. Let (w_n) be a sequence in W, and suppose that it has no convergent subsequence. So by Lemma 3.6.5, every $x \in W$ determines $\varepsilon(x) > 0$ such that $B_{\varepsilon(x)}(x)$ contains w_n for at most finitely many $n \in \mathbf{N}$. But the collection $\{B_{\varepsilon(x)}(x) \mid x \in W\}$ is an open cover for W and so therefore has a finite subcover $\{B_{\varepsilon(x_i)}(x_i) \mid 1 \leqslant i \leqslant n\}$ for some set of points $x_1, \dots, x_n \in W$. But each $B_{\varepsilon(x_i)}(x_i)$ can only contain finitely many points of (w_n), which implies that there exists only finitely many n such that W contains w_n, which is a contradiction. ☐

We just proved that compactness implies sequential compactness, and we would like to prove that the converse is true in metric spaces. To do so, we introduce a weakening of the notion of convergence for a sequence. This generalise the Cauchy sequences of real number from Analysis 1.

Definition 3.6.8 (Cauchy Sequence)**.**

A sequence (x_n) in (X, d) is **Cauchy** if

$$\forall \varepsilon > 0, \exists N \in \mathbf{N} \quad \text{such that} \quad m, n \geqslant N \implies d(x_m, x_n) < \varepsilon.$$

Proposition 3.6.9. *Every convergent sequence is Cauchy.*

Proof. Suppose (x_n) is a converging sequence in (X, d), with $\lim_n x_n = x \in X$. Let $\varepsilon > 0$. By convergence, there exists $N \in \mathbf{N}$ such that $d(x_n, x) \leqslant \frac{\varepsilon}{2}$ for every $n \geqslant N$. This implies

$$d(x_m, x_n) \leqslant d(x_m, x) + d(x, x_n) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \quad \forall m, n \geqslant N. \qquad \square$$

Not every Cauchy sequence is necessarily convergent as demonstrated by the following example. There exists however metric spaces, as $\mathbf{R}$ for example, in which every Cauchy sequence is convergent. We will study these spaces in Chapter 4.

Example 3.6.10. Consider the metric space $\mathbf{Q}$ with the subspace metric coming from the standard metric on $\mathbf{R}$. Then the $x_1 = 1, x_2 = 1.4, x_3 = 1.41, x_4 = 1.414, \dots$ (that is x_n is the first n decimal digit of $\sqrt{2}$) is Cauchy but does not converge in $\mathbf{Q}$.

We know introduce the notion of subsets S of X such that every point of X is uniformly close to a point of S. Any such subset naturally gives raise to an open cover (by open balls) of X.

Definition 3.6.11 (ε-net)**.**

Given $\varepsilon > 0$, an **ε-net** for a metric space X is a subset $S \subseteq X$ such that $X \subseteq \bigcup_{x \in S} B_\varepsilon(x)$.

It is clear from the definition that if S is a ε-net, then it is also a ε'-net for any $\varepsilon' \geqslant \varepsilon$.

Example 3.6.12. The subset $\mathbf{Z}^n$ is a 2-net in $\mathbf{R}^n$ (and a ε-net for any $\varepsilon > \frac{\sqrt{2}}{2}$).

Of course, for any $\varepsilon > 0$, we can always find an ε-net by taking $S = X$ but the interesting ones are the finite ones. The existence of a finite ε-net S implies that X is bounded (with $\operatorname{diam}(X) \leqslant \varepsilon + \max\{d(x, y) \mid x, y \in S\}$). It is thus not always possible to find finite ε-net. However, finite ε-nets always exist in sequentially compact spaces.

Proposition 3.6.13. *If (X, d) is sequentially compact, then given any $\varepsilon > 0$, there exists a finite ε-net for X.*

Proof. Suppose that for some $\varepsilon > 0$ there is no finite ε-net. We shall get a contradiction by constructing a sequence (x_n) in X with no convergent subsequence.

Let x_1 be a point in X, and suppose inductively that $x_1, x_2, \dots, x_r$ have been chosen in X such that $d(x_i, x_j) \geqslant \varepsilon$, for all $i, j \leqslant r$ with $i \neq j$. Then $\{x_1, \dots, x_r\}$ is not an ε-net and so there exists a point $x_{r+1} \in X$ such that $d(x_i, x_{r+1}) \geqslant \varepsilon$, for all $i \in \{1, \dots, r\}$. This completes the inductive step of constructing a sequence (x_n) for which $d(x_m, x_n) \geqslant \varepsilon$, for all $m \neq n$. This sequence clearly has no Cauchy subsequence and therefore no convergent subsequence. ☐

As discussed, the existence of a finite ε-net for some $\varepsilon > 0$ implies that the space is bounded. Proposition 3.6.13 implies the existence of ε-net for every $\varepsilon > 0$. Let us name this property.

Definition 3.6.14 (Totally Bounded)**.**

A metric space is **totally bounded** if it has a finite ε-net, for every $\varepsilon > 0$.

By Proposition 3.6.13, every sequentially compact implies totally bounded, and we already argued that totally bounded implies bounded. None of this implication is strict. See Tutorial 8 Question 7 for the first implication and the next example for the second implication.

Example 3.6.15. A discrete metric space (X, d_0) is bounded since its diameter is at most 1 (the diameter is 1 if X has at least two elements, and 0 otherwise). For $1 > \varepsilon > 0$, the open balls $B_\varepsilon(x)$ are singletons: $B_\varepsilon(x) = \{x\}$. We conclude that (X, d_0) has a $\frac{1}{2}$-net if and only if it X is finite.

ε-nets give especially open covers by balls of radius ε. We now introduce a notion for an arbitrary open cover $\mathcal{U}$ to contains "big" open sets. This will allow us to extract a subcover of $\mathcal{U}$ that comes from ε-net.

Definition 3.6.16 (Lebesgue Number)**.**

Given an open cover $\mathcal{U}$ of a metric space X, a fixed real number $\varepsilon > 0$ is called a **Lebesgue number** for $\mathcal{U}$ if, for any $x \in X$, there exists $U(x) \in \mathcal{U}$ such that $B_\varepsilon(x) \subseteq U(x)$.

If ε is a Lebesgue number for $\mathcal{U}$, then so is any ε' such that $0 < \varepsilon' \leqslant \varepsilon$. There may not be a Lebesgue number for a given open cover $\mathcal{U}$. For example let $X = (0, 1)$, and $\mathcal{U} = \{(\frac{1}{n}, 1) \mid n \geqslant 2\}$. Then for any $\varepsilon > 0$, take $n > \frac{1}{\varepsilon}$ and $x = \frac{1}{n}$. There is no $(\frac{1}{m}, 1)$ such that $B_\varepsilon(x) \subseteq (\frac{1}{m}, 1)$, since $B_\varepsilon(x) = B_\varepsilon(\frac{1}{n}) = (0, \frac{1}{n} + \varepsilon)$. So there does not exist a Lebesgue number for this open cover.

Lemma 3.6.17. *There exists a Lebesgue number for any open cover of a sequentially compact metric space X.*

Proof. Suppose that $\mathcal{U}$ is an open cover for which there does not exist a Lebesgue number. Then for any $n \in \mathbf{N}$, there exists some point $x_n \in X$ such that $B_{\frac{1}{n}}(x_n) \subseteq U$ is false for every $U \in \mathcal{U}$. By sequential compactness (x_n) has a subsequence (x_{n_r}) that converges to a point $x \in X$. Since $\mathcal{U}$ covers X, there exists $U_x \in \mathcal{U}$ such that $x \in U_x$. Since U_x is open, $B_{\frac{2}{m}}(x) \subseteq U_x$, for some $m \in \mathbf{N}$. Now, $B_{\frac{1}{m}}(x)$ contains x_{n_r} for all $r \geqslant R$, for some

$R \in \mathbf{N}$. Fix $r \geqslant R$ so that $n_r \geqslant m$, and write $s = n_r$. Then $B_{\frac{1}{s}}(x_s) \subseteq B_{\frac{2}{m}}(x)$ since

$$d(x_s, y) < \frac{1}{s} \implies d(x, y) \leqslant d(x, x_s) + d(x_s, y) < \frac{1}{m} + \frac{1}{s} \leqslant \frac{2}{m}.$$

Therefore $B_{\frac{1}{s}}(x_s) \subseteq U_x$ contradicting the choice of x_s. So there exists a Lebesgue number for $\mathcal{U}$. □

Even if any open cover $\mathcal{U}$ of a sequentially compact metric space has a Lebesgue number $\varepsilon(\mathcal{U})$, this does not necessarily mean there exists a common Lebesgue number ε working for all open covers.

We can finally prove the converse of Proposition 3.6.7.

Proposition 3.6.18. *Any sequentially compact metric space is compact.*

Proof. Let X be a sequentially compact metric space and let $\mathcal{U}$ be an open cover of X. By Lemma 3.6.17 there exists a Lebesgue number ε for $\mathcal{U}$. By Proposition 3.6.13 there exists a finite ε-net $\{x_1, \dots, x_n\}$, where ε is the same Lebesgue number. Then $B_\varepsilon(x_i) \subseteq U_i$, for some $U_i \in \mathcal{U}$ by the definition of Lebesgue number. Since

$$X \subseteq \bigcup_{i=1}^{n} B_\varepsilon(x_i) \subseteq \bigcup_{i=1}^{n} U_i,$$

we have a finite subcover $\{U_1, \dots, U_n\}$ of $\mathcal{U}$ for X. □

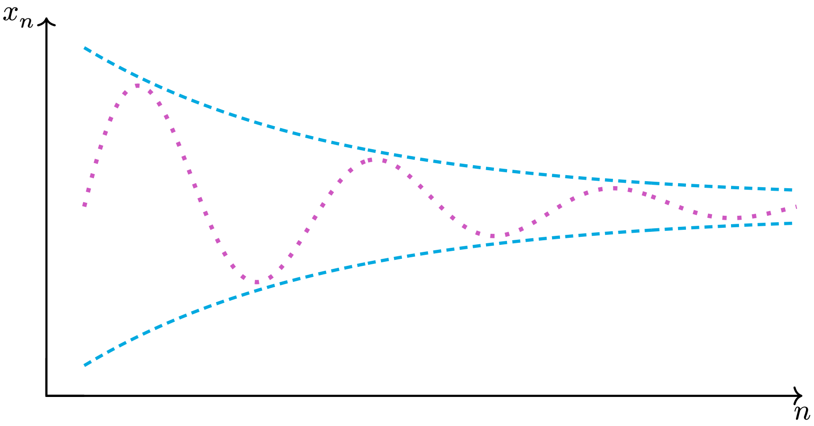
To summarise this Section, we have proved:

Theorem 3.6.19.

A metric space is compact if and only if it is sequentially compact.

CHAPTER 4.

COMPLETENESS



4.1. Definitions

Recall (Definition 3.6.8) that a sequence (x_n) in a metric space (X, d) is **Cauchy** if

$$\forall \varepsilon > 0, \exists N \in \mathbf{N} \quad \text{such that} \quad m, n \geqslant N \implies d(x_m, x_n) < \varepsilon.$$

Every convergent sequence is Cauchy, Proposition 3.6.9, but the converse is not true in general, Example 3.6.10. However, some metric spaces have the property that a sequence is convergent if and only if it is Cauchy.

Definition 4.1.1.

A metric space X is **complete** if every Cauchy sequence in X converges to a point in X.

The space $\mathbf{R}$ is complete by work done in Analysis 1, whilst $\mathbf{Q}$ is not complete by Example 3.6.10.

Example 4.1.2. $(0, 1)$ is not complete: the sequence $(\frac{1}{n})_{n \geqslant 2}$ is Cauchy in $(0, 1)$ but does not converge in $(0, 1)$.

Remark 4.1.3.

Since $\mathbf{R}$ is complete but $(0, 1)$ is not, we see that completeness is not a topological property. However, it is preserved under uniform equivalence (and therefore Lipschitz equivalence), see Tutorial 10, Question 7.

Observe that if d and d' are topologically equivalent metrics on the same set X, then a sequence (x_n) is convergent for d if and only if it is convergent for d'. So if (X, d') is complete but (X, d) is not this is not because (X, d) does not have enough converging sequences, but instead because (X, d') has too many Cauchy sequences. See Question 4.

It is not true that a subspace of a complete metric space is complete; think of $(0, 1) \subseteq \mathbf{R}$. However, this holds for closed subspaces as demonstrated below.

Proposition 4.1.4. *In a complete metric spaces, a subspace is complete if and only if it is closed. More precisely, we have*

1. *A complete subspace W of a metric space X is closed in X.*
2. *A closed subspace W of a complete metric space X is complete.*

Proof. "1" Suppose that W is complete and that $x \in \overline{W}$. So there is a sequence (w_n) in W such that $\lim_n w_n = x$. Since (w_n) is convergent it is Cauchy, so by the completeness of W, (w_n) converges to a point in W. By the uniqueness of limits of sequences in metric spaces, this point must be x and so $x \in W$.

"2" Suppose that X is complete and that W is closed in X, i.e. $W = \overline{W}$. Let (w_n) be a Cauchy sequence in W. By the completeness of X, (w_n) converges to a point $x \in X$ and so $x \in \overline{W} = W$. So W is complete. ☐

In general, convergence of a subsequence does not necessarily implies the convergence of the full sequence. This is for example the case of the sequence (x_n) defined by $x_{2k} = 0$ and $x_{2k+1} = k$. Such a situation does not occur for Cauchy sequences.

Lemma 4.1.5. *If a Cauchy sequence (x_n) in a metric space (X, d) has a subsequence (x_{n_r}) that converges to x, then (x_n) converges to x.*

Proof. Let $\varepsilon > 0$. Since (x_n) is Cauchy,

$$\exists N \in \mathbf{N} \quad \text{such that} \quad m, n \geqslant N \implies d(x_m, x_n) < \frac{\varepsilon}{2}.$$

Since (x_{n_r}) converges to x,

$$\exists R \in \mathbf{N} \quad \text{such that} \quad r \geqslant R \implies d(x_{n_r}, x) < \frac{\varepsilon}{2}.$$

For any $n \geqslant N$, choose $r \geqslant R$ such that $n_r \geqslant N$, and then

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

So (x_n) converges to x. ☐

The following result allows us to make a link between the previous chapter and this one.

Theorem 4.1.6.

Any compact metric space is complete.

Proof. Suppose that X is compact and that (x_n) is a Cauchy sequence in X. By compactness, (x_n) has a convergent subsequence (x_{n_r}), converging to a point $x \in X$. Then (x_n) converges to x by Lemma 4.1.5. ☐

Combining the above with results from Chapter 3, we obtain the following which generalise the Heine–Borel Theorem (Theorem 3.3.3).

Theorem 4.1.7.

A metric space is compact if and only if it is both complete and totally bounded.

Proof. We already know that a compact space is totally bounded (Proposition 3.6.13 and Theorem 3.6.19) and complete (Theorem 4.1.6). It remains to prove the converse.

Let X be a totally bounded space. We claim that any sequence in X admits a subsequence which is Cauchy. As such a subsequence would be convergent, this is enough to conclude. So let (x_n) be a sequence in a totally bounded space. Let D_1 be a finite 1-net of X. That is, $X = \bigcup_{x \in D_n} B_{\frac{1}{n}}(x)$. Since D_1 is finite, by the pigeonhole principle there exists $y_1 \in D_1$ such that $B_1(y_1)$ contains infinitely many of the x_n. This gives us a subsequence $(x_n)_{n \in A_1}$ for some infinite subset A_1 of $\mathbf{N}$. Now, let D_2 be a $\frac{1}{2}$-net for $B_1(y_1)$. Since D_2 is finite, there exists $y_1 \in D_{\frac{1}{2}}$ such that $B_{\frac{1}{2}}(y_2) \subseteq B_1(y_1)$ contains infinitely many of the x_n for $n \in A_1$. We hence obtain an infinite subset $A_2 \subseteq A_1 \subseteq \mathbf{N}$ and a corresponding subsequence $(x_n)_{n \in A_2}$. We can repeat this process indefinitely.

Now, we choose an increasing sequence of integers (n_k) such that for every k we have $n_k \in A_k$. Equivalently, we have a subsequence (x_{n_k}) such that $x_{n_k} \in B_{\frac{1}{k}} y_k$ for every k. Given $\varepsilon > 0$, there exists N such that $\frac{1}{2N} \leqslant \varepsilon$. For every $k, l \geqslant N$, both x_{n_k} and x_{n_l} belong to $B_{\frac{1}{2N}}(y_{2N})$ and so $d(x_{n_k}, x_{n_l}) < \frac{1}{N} \leqslant \varepsilon$. ☐

As for closed subsets, the collection of complete subspaces is closed under finite unions and arbitrary intersections.

Lemma 4.1.8. *If A and B are two complete subsets of a metric space X, then both $A \cap B$ and $A \cup B$ are compact. Moreover, if $I \neq \varnothing$ and all the A_i are complete, then $\bigcap_{i \in I} A_i$ is also complete.*

Proof. The proof for finite union and intersections if left to the reader (see Question 1).

Let us do the case of an arbitrary intersection of complete subspaces $A := \bigcap_{i \in I} A_i$. Since all the A_i are complete, they are also closed in X. It follows that their intersection is closed, and hence complete. ☐

The above result fails for infinite unions. Indeed, all $[\frac{1}{n}, n]$ are complete subspaces of $\mathbf{R}$, but their union is $(0, +\infty)$, which is not complete.

Theorem 4.1.6 can be use to show that some spaces are complete even if they are not compact.

Example 4.1.9 ($\mathbf{R}^n$ is complete)**.** We will prove this with the Euclidean metric d_2 on $\mathbf{R}^n$. Suppose that $({}^r x)_{r \geqslant 1}$ is a Cauchy sequence in $\mathbf{R}^n$ where

$$ {}^r x = ({}^r x_1, {}^r x_2, \dots, {}^r x_n) \in \mathbf{R}^n. $$

For each $i \in \{1, \dots, n\}$, $({}^r x_i)$ is a Cauchy sequence in $\mathbf{R}$, since $|{}^r x_i - {}^s x_i| \leqslant d_2({}^r x, {}^s x)$. So $({}^r x_i)$ converges to x_i say. We shall prove that $({}^r x)$ converges to $x = (x_1, \dots, x_n) \in \mathbf{R}^n$. Given $\varepsilon > 0$, for each i there exists $N_i \in \mathbf{N}$ such that $r \geqslant N_i$ implies that $|{}^r x_i - x_i| < \frac{\varepsilon}{\sqrt{n}}$. Then for all $r \geqslant \max\{N_1, \dots, N_n\}$,

$$ d_2({}^r x, x) \leqslant \sqrt{\sum_{i=1}^{n} ({}^r x_i - x_i)^2} < \varepsilon. $$

So $({}^r x)$ converges to $x \in \mathbf{R}^n$.

Remark 4.1.10.

It turns out that in the above example we have proved more than announced. Indeed, the same proof directly gives that if X and Y are complete metric space, then $(X \times Y, d_b)$ is a complete metric space. Since d_a, d_b and d_c (and all the d_{a_p} for $p \geqslant 1$) metrics are Lipschitz equivalent, the same is true for any of them.

Remark 4.1.11.

It is not true that when X and Y are complete then so is $(X \times Y, d)$ for any product metric. For example,

$$d'((x_1,x_2),(y_1,y_2)) = \sqrt{\left(\frac{x_1}{1+|x_1|} - \frac{y_1}{1+|y_1|}\right)^2 + \left(\frac{x_2}{1+|x_2|} - \frac{y_2}{1+|y_2|}\right)^2}$$

is a non-complete metric on $\mathbf{R}^2$ which is topologically equivalent to d_2. (This metric d' makes $\mathbf{R}^2$ isometric to $(-1,1)^2$ via the map $(x_1,x_2) \mapsto (\frac{x_1}{1+|x_1|}, \frac{x_1}{1+|x_2|})$.)

We now proceed to prove a characterisation of completeness in terms of nested sequences of closed subsets

Theorem 4.1.12 (Cantor's[1] Intersection Theorem)**.**

A metric space X is complete if and only if for any sequence (V_n) of non-empty closed subsets of X, with $V_{n+1} \subseteq V_n$, for all $n \in \mathbf{N}$ and such that $\operatorname{diam} V_n \to 0$, as $n \to \infty$, we have

$$\bigcap_{n\in\mathbf{N}} V_n \neq \varnothing.$$

Remark 4.1.13.

If $\operatorname{diam} V_n \to 0$, as $n \to \infty$, then $|\bigcap_{n\in\mathbf{N}} V_n| \in \{0,1\}$.

Proof. "$\Rightarrow$" Let $x_n \in V_n$. Then (x_n) is a Cauchy sequence since given $\varepsilon > 0$, there exists $N \in \mathbf{N}$ such that $\operatorname{diam} V_N < \varepsilon$, then for any $m,n \geqslant N$, $x_m, x_n \in V_N$, so $d(x_m,x_n) < \varepsilon$. By the completeness of X, (x_n) converges to some $x \in X$. Since $x_n \in V_m$, for all $n \geqslant m$, $x \in \overline{V_m} = V_m$. This is true for any m, so

$$\bigcap_{n\in\mathbf{N}} V_n = \{x\}.$$

"$\Leftarrow$" Suppose X is not complete and let (x_n) be a non-convergent Cauchy sequence in X. Let $V_n = \{x_r \mid r \geqslant n\}$. Then V_n is closed since, if it were not, there would be a sequence (v_j) in V_n that converges to a point $x \in \overline{V_n} \setminus V_n$. Since $v_j = x_r$, for some $r \geqslant n$, (v_j) is a subsequence (x_{n_k}) of (x_n), which would imply that (x_n) converges to x as (x_n) is Cauchy, see Lemma 4.1.5. Clearly, $V_{n+1} \subseteq V_n$, for all $n \in \mathbf{N}$. Also, given $\varepsilon > 0$, there exists $N \in \mathbf{N}$ such that $m,n \geqslant N \Rightarrow d(x_m,x_n) < \varepsilon$, so $\operatorname{diam} V_N \leqslant \varepsilon$ and $\operatorname{diam} V_n \to 0$, as $n \to \infty$. Suppose that $x \in V_n$, for all $n \in \mathbf{N}$. Then $\lim x_n = x$ since, for any $n \in \mathbf{N}$, $d(x,x_n) \leqslant \operatorname{diam} V_n$. This contradicts our assumption and so, $\bigcap_{n\in\mathbf{N}} V_n = \varnothing$. ☐

---

[1] Georg Ferdinand Ludwig Philipp Cantor (1845–1918).

4.2. Contractions

On page 7 we observed that distance preserving functions were automatically injective and hence concluded that this class of functions was too restrictive. We thus studied other class of functions: (uniformly) continuous functions, Lipschitz equivalences, ...If we slightly relax the condition of preserving the distance, we obtain a new interesting class of functions.

Definition 4.2.1 (Contraction)**.**

A function $f\colon X \to X$, where (X, d) is a metric space, is a **contraction** (or **short map**) if there exists a constant $K < 1$ such that

$$d(f(x), f(y)) \leqslant K d(x, y), \quad \forall x, y \in X.$$

Remark 4.2.2.

Obviously, $K \geqslant 0$ and $K = 0$ if and only if f is a constant function.

Observe that the statement that f is a contraction depends heavily on the choice of metric on X: it is possible that two metrics d and d' are Lipschitz equivalent on X and that f is a d-contraction but not a d'-contraction, see Question 2 for a concrete example.

It directly follows from the definition that any metric preserving function is a contraction. The converse is not true as demonstrated by the map $f\colon \mathbf{R} \to \mathbf{R}, x \mapsto \frac{x}{/}2$. Our next result shows that being a contraction is a stronger property than being uniformly continuous.

Lemma 4.2.3. *Any contraction is uniformly continuous.*

Proof. Using the above notation, let $\varepsilon > 0$ and $\delta = \frac{\varepsilon}{K}$. For any $x, y \in X$ satisfying $d(x, y) < \delta$, we have

$$d(f(x), f(y)) \leqslant K d(x, y) < \varepsilon,$$

as required. ☐

To go further

In fact, any contraction is a Lipschitz *function*, where a function $f\colon X \to Y$ is Lipschitz if there exists K such that $d(f(x), f(y)) \leqslant K d(x, y)$ for all $x, y \in X$. So a Lipschitz equivalence is a bijective function f such that both f and f^{-1} are Lipschitz. Any contraction is Lipschitz and any Lipschitz function is uniformly continuous. These two implications are strict.

Being a contraction is a strong condition that implies consequences that are not necessarily true for uniformly continuous. Before stating the first consequence, let us formally define what is a fixed point of a self-map.

Definition 4.2.4.

Given a set S, a **self-map** of S is a function $f\colon S \to S$. A **fixed point** of a self-map f of S is a point $p \in S$ such that $f(p) = p$.

We can now state the main theorem of this section.

Theorem 4.2.5 (Banach's[2] Contraction Mapping Theorem)**.**

If $f\colon X \to X$ is a contraction of a complete non-empty metric space X, then f has a unique fixed point $p \in X$.

Proof. We start by proving the existence of a fixed point. Choose $x_1 \in X$ arbitrarily and let $x_{n+1} := f(x_n)$, for $n \geqslant 1$. We shall prove that (x_n) is a Cauchy sequence. First note that $d(x_r, x_{r+1}) \leqslant K^{r-1} d(x_1, x_2)$, for all $r \geqslant 1$. This follows easily by induction from the contraction condition. Now for $m > n$, by repeated use of the triangle inequality,

$$d(x_m, x_n) \leqslant d(x_m, x_{m-1}) + d(x_{m-1}, x_{m-2}) + \cdots + d(x_{n+1}, x_n).$$

So

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

where we used $0 \leqslant K < 1$ for the last inequality. Using once again $0 \leqslant K < 1$, we also have $K^{n-1} \to 0$ as $n \to \infty$. So (x_n) is a Cauchy sequence and since X is complete it converges to some $p \in X$. By Lemma 4.2.3 f is continuous and so $f(x_n) \to f(p)$, as $n \to \infty$. But $f(x_n) = x_{n+1} \to p$ as $n \to \infty$. So $f(p) = p$.

We now proceed to show the uniqueness of the fixed point. If $f(p) = p$ and $f(q) = q$, then

$$d(p, q) = d(f(p), f(q)) \leqslant K d(p, q).$$

Since $K < 1$, this is a contradiction unless $p = q$. So the fixed point is unique. ☐

Remark 4.2.6.

A remarkable aspect of the proof of the Contraction Mapping Theorem is that, not only does it prove the existence and uniqueness of the fixed point, it tells you how to find it: choose a random element $x_1 \in X$ and set $x_{n+1} = f(x_n)$. Then the unique fixed point p is the limit of the sequence (x_n), i.e. $p = \lim_n x_n$.

---

[2] Stefan Banach (1892–1945).

Remark 4.2.7.

The completeness condition in the Contraction Mapping Theorem is necessary as demonstrated by the function $f\colon (0,1) \to (0,1)$, $x \mapsto \frac{x}{2}$. This function is a contraction with $K = \frac{1}{2}$ but it has no fixed point on $(0,1)$ as $0 \notin (0,1)$.
The contraction condition in the Contraction Mapping Theorem is also necessary as demonstrated by $f\colon [1,\infty) \to [1,\infty)$, $f(x) = x + \frac{1}{x}$ which has no fixed points (since $f(x) = x \Rightarrow \frac{1}{x} = 0$) despite $[1,+\infty)$ being complete. Observe that the map f is nearly a contraction in the sense that $d(f(x), f(y)) < d(x,y)$, for all $x \neq y$. However, f is not a contraction as $\sup_{x\neq y} \frac{d(f(x),f(y))}{d(x,y)} = 1$.

As we observed after Definition 4.2.1, it is possible to construct examples of two Lipschitz equivalent metrics on a set X such that f is a d-contraction but not a d'-contraction. It follows from Lipschitz equivalence that (X,d) is complete if and only if (X,d') is complete, however the theorem only apply to (X,d) and not to (X,d').

To go further

Banach's Contraction Mapping Theorem is one of the numerous useful fixed point theorems in mathematics. Another one is the Brouwer fixed point theorem:

> *Every continuous function from a closed ball of a Euclidean space into itself has a fixed point.*

These fixed point theorems usually came with a procedure on how to compute an actual fixed point. They thus have usually many applications, both in mathematics and for concrete real world problems.

An important consequence of the Contraction Mapping Theorem is that we can use it to compute numerical approximations.

Example 4.2.8 (Approximation Procedure)**.** Consider the function

$$g(x) = \frac{1}{2}\left(x + \frac{2}{x}\right).$$

This is a self-map of the interval $[1,2]$ because

$$1 \leqslant x \leqslant 2 \quad \Longrightarrow \quad \frac{1}{2} \leqslant \frac{x}{2} \leqslant 1 \quad \text{and} \quad \frac{1}{2} \leqslant \frac{1}{x} \leqslant 1,$$

so $1 \leqslant g(x) \leqslant 2$.

If $x, y \in [1,2]$ then

$$|g(x) - g(y)| = \left|\frac{x-y}{2} + \frac{y-x}{xy}\right| = |x-y| \cdot \left|\frac{1}{2} - \frac{1}{xy}\right| \leqslant \frac{1}{2}|x-y|.$$

Therefore, $|g(x) - g(y)| \leqslant \frac{|x-y|}{2}$, and g is a contraction with $K = \frac{1}{2}$. Moreover, $[1,2]$ is complete as it is a closed subspace of the complete space $\mathbf{R}$. So by the Contraction Mapping Theorem g has a unique fixed point $p \in [1,2]$.

The fixed point must satisfy

$$p = g(p) = \frac{(p + \frac{2}{p})}{2}.$$

Therefore, $p^2 = 2$, and $p = +\sqrt{2}$.

Remember that we can pick a starting point $x_1 \in [1, 2]$ at random and the sequence $(x_1, g(x_1), g^2(x_1), \dots)$ converges to the fixed point $\sqrt{2}$ for *every* $x_1 \in [1, 2]$. If we choose $x_1 = 1$ we get

$$\left(1, \frac{3}{2}, \frac{17}{12}, \frac{577}{408}, \dots\right) = (1, 1.5, 1.416666666, 1.414215686, \dots)$$

and if we choose $x_1 = \frac{7}{5}$, then

$$\left(\frac{7}{5}, \frac{99}{70}, \frac{19601}{13860}, \dots\right) = (1.4, 1.414285714, 1.414213564, \dots)$$

to 9 decimal places. This is not bad as the correct answer is $1.414213562\dots$.

Remark 4.2.9.

This procedure can be generalised to compute the decimal expansion of $\sqrt{a}$ for $a \geqslant 1$ by considering

$$g(x) = \frac{1}{2}\left(x + \frac{a}{x}\right)$$

on $[1, a]$. If $0 \leqslant a \leqslant 1$, then one need to consider g as above, but defined on $[a, 1]$.

As the above remark shows, sometimes in order to use the Contraction Mapping Theorem for a given f the difficult part is not to show that f is contracting, but it is to find X such that $f(X) \subseteq X$.

To go further

The Contraction Mapping Theorem admits many partial converses. For example here is one due to Czesław Bessaga:

> *Let $f\colon X \to X$ be a self-map of an abstract set such that each iterate $f^n = f \circ \dots \circ f$ has a unique fixed point. Let $K \in (0, 1)$, then there exists a complete metric on X such that f is a contraction, and K is the contraction constant.*

Observe that if f is a contraction, then so is each of its iterate f^n. In particular, the Contraction Mapping Theorem asserts that each f^n has a unique fixed point. This shows that the condition in Bessaga's result is natural.

4.3. Function Spaces

In Section 1.3 we have seen that there exist a natural metric, namely the $d_{\sup}$ metric, on the set of bounded functions from $[a, b]$ to $\mathbf{R}$. This gaves us the metric space $\mathcal{B}[a, b]$. The restriction of d_{sup} to the subset of continuous functions gives the subspace $\mathscr{C}[a,b]$. The aim of this section is to show that both $\mathcal{B}[a,b]$ and $\mathscr{C}[a,b]$ are complete, see Corollaries 4.3.10 and 4.3.12.

For (not necessarily bounded) functions, we have the following naive notion of convergence.

Definition 4.3.1 (Pointwise Convergence)**.**

Let $A \subseteq \mathbf{R}$. A sequence $(f_n\colon A \to \mathbf{R})$ of real valued functions **converges** to $f\colon A \to \mathbf{R}$ **pointwise** on A if $\lim_n f_n(x) = f(x)$, for all $x \in A$.
We write $f_n \to f$ pointwise on A and call f the **pointwise limit**.

The sequence (f_n) pointwise converge to f if and only if:

$$\forall x \in A, \forall \varepsilon > 0, \exists N, \forall n \geqslant N : |f_n(x) - f(x)| < \varepsilon.$$

Observe that in the above formula N depends a priori on x (and of course on ε).

Let us revisit the power functions from Example 1.11.6.

Example 4.3.2. Define $f_n(x) = x^n$ on $A = [0,1]$, see Figure 1.10 on page 38. Then

$$\lim f_n(x) = f(x) := \begin{cases} 0, & x \in [0,1); \\ 1, & x = 1. \end{cases}$$

Note that whilst each f_n is continuous, f is not.

The fact that, as in the above example, continuous functions might pointwise converge to non-continuous functions is a bit annoying and motivate use to define a stronger notion of convergence for sequences of functions, that will hopefully be better behaved.

Definition 4.3.3.

A sequence $(f_n\colon S \to \mathbf{R})$ **converges** to $f\colon S \to \mathbf{R}$ **uniformly** on S if

$$\forall \varepsilon > 0, \exists N \in \mathbf{N}, \forall n \geqslant N, \forall x \in S : |f_n(x) - f(x)| < \varepsilon.$$

We write $f_n \to f$ uniformly on S and call f the **uniform limit**.

In the definition, we make no assumption on S that can be any set. In practice, we will mostly use for $S = [a,b] \subseteq \mathbf{R}$.

Obviously, if $f_n \to f$ uniformly, then $f_n \to f$ pointwise on S, but the opposite is not true as demonstrated by Remark 4.3.4. Uniform convergence is stronger in that N is independent of x, while for pointwise convergence we can use a different N for each x. It follows from the above that, if the uniform limit of (f_n) exists, then so does the pointwise limit and they are equal.

Remark 4.3.4.

The sequence from Example 4.3.2 is an example of a sequence that converges pointwise but does not converge uniformly. Indeed, let $\varepsilon = \frac{1}{2}$ and for any $N \in \mathbf{N}$ choose $x \in (0,1)$ such that $x^N = \frac{1}{2}$, which exists by the Intermediate Value Theorem. Then

$$|f_N(x) - f(x)| = |x^N - 0| = x^N = \frac{1}{2} = \varepsilon.$$

Warning: the convergence of $f_n(x) = x^n$ is not uniform on $[0,1)$ either – we can use the same proof. However, for any $a \in (0,1)$, the sequence does converge uniformly on $[0,a]$.

A constant sequence of functions (f_n) with $f_n = f_m$ for all m, n is obviously uniformly convergent.

Example 4.3.5. Let (y_n) be a sequence of real numbers and let S be any non-empty set of $\mathbf{R}$. For $n \in \mathbf{N}$, define $f_n \colon S \to \mathbf{R}$ to be the constant function $f(x) = y_n$. Then the sequence (y_n) converges (to $y \in \mathbf{R}$) if and only if the sequence of functions (f_n) converges uniformly (to the constant function $f(x) = y$).

Uniform convergence is not a new concept, but an old one in disguise as demonstrated by the following result.

Proposition 4.3.6. *Let* (f_n) *be a sequence in* $\mathcal{B}[a,b]$ *and* $f \in \mathcal{B}[a,b]$. *Then*

$$\lim_n f_n = f \in \mathcal{B}[a,b] \iff f_n \longrightarrow f \text{ uniformly on } [a,b]$$

Proof. Let us remind that in $\mathcal{B}[a,b]$ we have $\lim f_n = f$ if and only if $d_{\text{sup}}(f_n, f) \to 0$ as $n \to \infty$.

"$\Rightarrow$" To prove that $f_n \to f$ uniformly on $[a,b]$, let $\varepsilon > 0$. Then there exists $N \in \mathbf{N}$ such that $0 \leqslant d_{\text{sup}}(f_n, f) < \varepsilon$. Since $|f_n(x) - f(x)| \leqslant d_{\text{sup}}(f_n, f)$, for all $x \in [a,b]$, by definition, we get that $|f_n(x) - f(x)| < \varepsilon$, for all $n \geqslant N$ and all $x \in [a,b]$.

"$\Leftarrow$" Suppose that $f_n \to f$ uniformly on $[a,b]$ and let $\varepsilon > 0$. Then there exists $N \in \mathbf{N}$ such that

$$n \geqslant N \implies |f_n(x) - f(x)| < \frac{\varepsilon}{2}, \quad \text{for all } x \in [a,b].$$

Then

$$n \geqslant N \implies 0 \leqslant d_{\text{sup}}(f_n, f) = \sup_{x \in [a,b]} |f_n(x) - f(x)| \leqslant \frac{\varepsilon}{2} < \varepsilon. \qquad \square$$

We have defined uniform convergence of a sequence of functions by asking the N in the definition to work for all x. We can similarly define uniform Cauchy sequences.

Definition 4.3.7 (Uniformly Cauchy).

A sequence $(f_n\colon S \to \mathbf{R})$ of functions is **uniformly Cauchy** on S if for all $\varepsilon > 0$, there exists $N \in \mathbf{N}$ such that

$$\forall \varepsilon > 0, \exists N, \forall m, n \geqslant N, \forall x \in S : |f_m(x) - f_n(x)| < \varepsilon.$$

The following is the Cauchy version of Proposition 4.3.6.

Lemma 4.3.8. *Let (f_n) be a sequence in $\mathcal{B}[a,b]$. Then (f_n) is a Cauchy sequence for d_{sup} if and only if it is uniformly Cauchy on $[a,b]$.*

Proof. "⇒" If (f_n) is a Cauchy sequence, then for $\varepsilon > 0$, there exists $N \in \mathbf{N}$ such that

$$m, n \geqslant N \implies |f_m(x) - f_n(x)| \leqslant d_{\text{sup}}(f_m, f_n) < \varepsilon, \quad \forall x \in [a,b].$$

So (f_n) is uniformly Cauchy.

"⇐" Suppose that (f_n) is uniformly Cauchy on $[a,b]$ and let $\varepsilon > 0$. Then here exists $N \in \mathbf{N}$ such that

$$m, n \geqslant N \implies |f_m(x) - f_n(x)| < \frac{\varepsilon}{2}, \quad \forall x \in [a,b].$$

We conclude that

$$m, n \geqslant N \implies d_{\text{sup}}(f_m, f_n) = \sup_{x \in [a,b]} |f_m(x) - f_n(x)| \leqslant \frac{\varepsilon}{2} < \varepsilon, \quad \forall x \in [a,b]$$

and thus (f_n) is a Cauchy sequence for d_{sup}. ☐

For a sequence of, not necessarily bounded, functions, it is equivalent to converge uniformly and to be uniformly Cauchy:

Theorem 4.3.9 (Cauchy's Criteria for Uniform Convergence).

Let $(f_n\colon S \to \mathbf{R})$ be a sequence of real-valued functions. Then (f_n) converges uniformly on S if and only if (f_n) is uniformly Cauchy on S.

Proof. "⇒" Suppose that (f_n) converges uniformly and let $\varepsilon > 0$. Then there exists $N \in \mathbf{N}$ such that $n \geqslant N$ implies $|f_n(x) - f(x)| < \frac{\varepsilon}{2}$, for all $x \in S$. So for $m, n \geqslant N$,

$$|f_m(x) - f_n(x)| \leqslant |f_m(x) - f(x)| + |f(x) - f_n(x)| < \varepsilon, \quad \forall x \in S,$$

showing that the sequence (f_n) is uniformly Cauchy.

"⇐" Suppose that (f_n) is uniformly Cauchy on S and let $\varepsilon > 0$. Then for any $x \in S$, $(f_n(x))$ is a Cauchy sequence in $\mathbf{R}$ and therefore converges to a real number $f(x)$. That is, there exists N_x such that $m \geqslant N_x$ implies that $|f_m(x) - f(x)| < \frac{\varepsilon}{2}$. Now choose $N \in \mathbf{N}$ such that $m, n \geqslant N$ implies that $|f_m(x) - f_n(x)| < \frac{\varepsilon}{2}$, for all $x \in S$. By combining the above two facts, we obtain for any $x \in S$ and any $m \geqslant \max\{N, N_x\}$:

$$n \geqslant N \implies |f_n(x) - f(x)| \leqslant |f_n(x) - f_m(x)| + |f_m(x) - f(x)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

So $f_n \to f$ uniformly on S. ☐

To go further

It is possible to define convergence uniform and uniform Cauchyness for sequence of functions $(f_n\colon S \to X)$ where X is an arbitrary metric space. If X is complete, then Theorem 4.3.9 (and its proof) remains true.

Combining the previous results we obtain the completeness of the space of bounded functions.

Corollary 4.3.10. *The metric space $\mathcal{B}[a,b]$ is complete.*

Proof. Let (f_n) be a Cauchy sequence in $\mathcal{B}[a,b]$. Then (f_n) is uniformly Cauchy on $[a,b]$ by Lemma 4.3.8. So (f_n) converges uniformly on $[a,b]$, by Cauchy's Criteria for Uniform Convergence and hence is convergent in $\mathcal{B}[a,b]$ by Proposition 4.3.6. ☐

We now prove that the phenomenon described in Example 4.3.2 (a sequence of continuous functions converging to a non-continuous function) cannot happens for uniform convergence.

Theorem 4.3.11.

If $f_n\colon [a,b] \to \mathbf{R}$ is continuous at $c \in [a,b]$, for each $n \in \mathbf{N}$, and $f_n \to f$ uniformly on $[a,b]$, then f is continuous at c.

Proof. Given $\varepsilon > 0$, uniform convergence implies that there exists $N \in \mathbf{N}$ such that

$$n \geqslant N \implies |f_n(x) - f(x)| < \frac{\varepsilon}{3}, \quad \forall x \in [a,b].$$

By continuity of f_N at c there exists $\delta > 0$ such that

$$x \in (c-\delta, c+\delta)) \cap [a,b] \implies |f_N(x) - f_N(c)| < \frac{\varepsilon}{3}.$$

So for any such x,

$$|f(x) - f(c)| \leqslant |f(x) - f_N(x)| + |f_N(x) - f_N(c)| + |f_N(c) - f(c)| < \varepsilon.$$ ☐

We conclude this section by showing the completeness of the space of continuous functions.

Corollary 4.3.12. *The metric space $\mathscr{C}[a,b]$ is complete.*

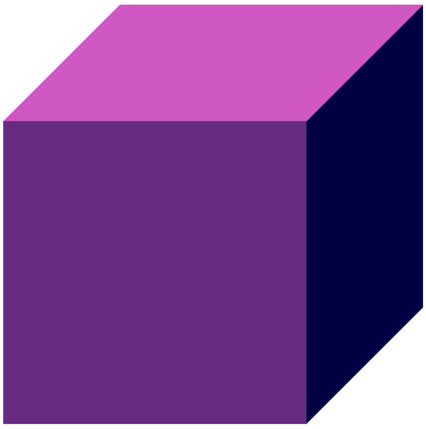
Proof. Let $\mathscr{C}_c$ be the subspace of functions in $\mathcal{B}[a,b]$ that are continuous at c. The previous theorem tells us that $\mathscr{C}_c$ is closed in the complete space $\mathcal{B}[a,b]$. Hence, $\mathscr{C}_c$ is complete. Observe that

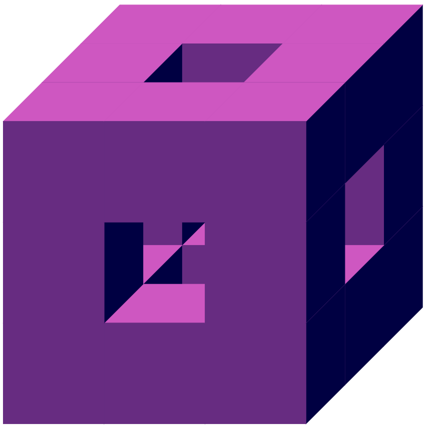
$$\mathscr{C}[a,b] = \bigcap_{c\in[a,b]} \mathscr{C}_c,$$

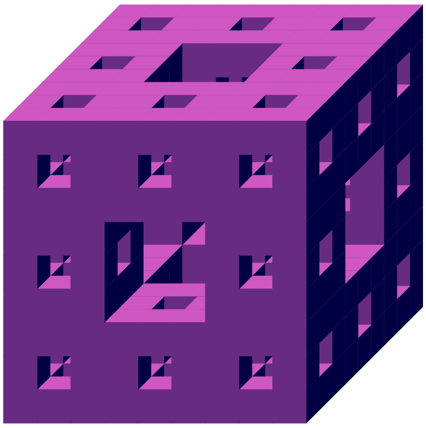
which implies that $\mathscr{C}[a,b]$ is also closed in $\mathcal{B}[a,b]$ and hence complete. ☐

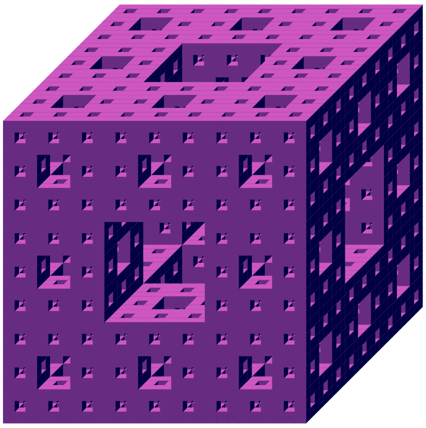
CHAPTER 5.

APPLICATIONS









In this chapter we will see three applications of the methods developed in this module. The three sections of this chapter are about different applications and are independent one of each other.

5.1. Initial Value Problem

An **initial value problem** (**IVP**) is an ordinary differential equation of the form

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(t, y(t)),$$

with $f\colon U \subseteq \mathbf{R}^2 \to \mathbf{R}$, where U is an open set of $\mathbf{R}^2$, together with a point $(t_0, y_0) \in U$ which is called the **initial condition**.

A **solution** to an IVP is a function y that is a solution to the differential equation and satisfies $y(t_0) = y_0$.

We know how to find solutions to simple IVPs: for example, suppose

$$\frac{\mathrm{d}y}{\mathrm{d}t} = 2y \implies \int \frac{\mathrm{d}y}{y} = \int 2\,\mathrm{d}t \implies |y| = \mathrm{e}^{C_1}\mathrm{e}^{2t}.$$

Let $C = \pm\mathrm{e}^{C_1}$ so $y = C\mathrm{e}^{2t}$. When we have an initial condition we can find C, for example, if $y(0) = 19$ we get $C = 19$ and the solution is $y(t) = 19\mathrm{e}^{2t}$.

More complex IVPs are a priori not easy to solve. Hopefully, if f is nice enough, then we can use the theory of metric space to guarantee a unique solution (on a small interval).

Theorem 5.1.1 (Picard, Lindelof, Lipschitz and Cauchy[1]).

Given an IVT, suppose that

$$f\colon U \subseteq \mathbf{R}^2 \to \mathbf{R}$$

is continuous and Lipschitz continuous in the second coordinate, that is, there exists $L > 0$ *such that*

$$|f(t, y_1(t)) - f(t, y_2(t))| \leqslant L|y_1(t) - y_2(t)|, \quad \forall t.$$

Then there is an $\varepsilon > 0$*, such that, there exists a unique solution* $y(t)$ *to the IVP on the interval* $[t_0 - \varepsilon, t_0 + \varepsilon]$.

Before giving a formal proof of Theorem 5.1.1, let us introduce the idea behind it. The idea of the proof is to integrate both sides of the differential equation:

$$y(t) - y(t_0) = \int_{t_0}^{t} f(s, y(s))\,\mathrm{d}s,$$

---

[1] Charles Émile Picard (1856–1941), Ernst Leonard Lindelöf (1870–1946), Rudolf Otto Sigismund Lipschitz (1832–1903) and Augustin Louis Cauchy (1789–1857).

and set $\varphi_0(t) := y_0$ to be the constant function. Then we perform what is called **Picard iteration** by setting:

$$\varphi_{k+1}(t) := y_0 + \int_{t_0}^{t} f(s, \varphi_k(s))\,\mathrm{d}s.$$

We see that φ_k satisfies the initial condition, for all $k \in \mathbf{N}$. The sequence (φ_n) will be Cauchy in $\mathscr{C}[t_0 - \varepsilon, t_0 + \varepsilon]$, which is complete and therefore (φ_n) converges to φ which satisfies

$$\varphi(t) = y_0 + \int_{t_0}^{t} f(s, \varphi(s))\,\mathrm{d}s \implies \frac{\mathrm{d}\varphi}{\mathrm{d}t} = f(t, \varphi(t)) \quad \text{and} \quad \varphi(t_0) = y_0,$$

as the fixed point of a contraction. So setting $y = \varphi(t)$ we have a solution to the IVP.

Example 5.1.2. Let $y(t) = \tan(t)$ and $t_0 = 0$. Then

$$\frac{\mathrm{d}y}{\mathrm{d}t} = 1 + y^2, \quad y_0 = 0 \quad \text{and} \quad \varphi_0(t) = 0.$$

Suppose now that we are given the IVT 5.1.2 and that we forgot that $\tan(t)$ was a solution. Let us try to solve it using Picard iteration:

$$\varphi_{k+1}(t) = \int_0^t (1 + (\varphi_k(s))^2)\,\mathrm{d}s$$

and $\varphi_n(t) \to y(t)$, as $n \to \infty$. Calculating the first few Picard iterates:

$$\varphi_1(t) = \int_0^t (1 + 0^2)\,\mathrm{d}s = t; \quad \varphi_2(t) = \int_0^t (1 + s^2)\,\mathrm{d}s = t + \frac{t^3}{3};$$

$$\varphi_3(t) = \int_0^t \left(1 + \left(s + \frac{s^3}{3}\right)^2\right)\mathrm{d}s = t + \frac{t^3}{3} + \frac{2t^5}{15} + \frac{t^7}{63}.$$

This is calculating the series expansion of $\tan(t)$.

We now proceed to the formal proof of Theorem 5.1.1.

Proof of Theorem 5.1.1. Let $a, b > 0$ be such that

$$C_{a,b} =\subseteq U.$$

So $C_{a,b}$ is a compact subspace on which f is defined. Let

$$M = M(f, a, b) := \sup_{(t,y) \in C_{a,b}} |f(t, y)|.$$

Let $I := [t_0 - a, t_0 + a]$ and $B := [y_0 - b, y_0 + b]$, so $C_{a,b} = I \times B$. Consider the metric space $\mathscr{C} := \mathscr{C}(I, B)$ of continuous functions from I to B with the d_{sup} metric. This is a subspace of the complete metric space $\mathscr{C}[t_0 - a, t_0 + a]$ and it is closed. Indeed, for any sequence (f_n) in $\mathscr{C}$ that converges to a point in $f \in \mathscr{C}[t_0 - a, t_0 + a]$ and for any $\varepsilon > 0$, there exists $N \in \mathbf{N}$ such that

$$n \geqslant N \implies |f(t) - f_n(t)| \leqslant d_{sup}(f, f_n) < \varepsilon, \quad \forall t \in I.$$

So for any $\varepsilon > 0$, there is an $N \in \mathbf{N}$ such that

$$n \geqslant N \implies y_0 - b - \varepsilon \leqslant f_n(t) - \varepsilon \leqslant f(t) \leqslant f_n(t) + \varepsilon \leqslant y_o + b + \varepsilon, \quad \forall t \in I$$

and so $f(t) \in B$, for all $t \in I$. Therefore, $\mathscr{C}$ is closed in $\mathscr{C}[t_0 - a, t_0 + a]$ and thus is complete.

Now we define a self-map $\mathit{\Gamma}\colon \mathscr{C} \to \mathscr{C}$ by

$$\mathit{\Gamma}\varphi(t) := y_0 + \int_{t_0}^{t} f(s, \varphi(s))\,\mathrm{d}s.$$

First, we need to show that $\mathit{\Gamma}$ is actually a self-map, that is, that $\mathit{\Gamma}\varphi \in \mathscr{C}$, for all $\varphi \in \mathscr{C}$. If $\varphi \in \mathscr{C}$, then $|\varphi(t) - y_0| \leqslant b$, for all $t \in I$. And so,

$$|\mathit{\Gamma}\varphi(t) - y_0| = \left|\int_{t_0}^{t} f(s, \varphi(s))\,\mathrm{d}s\right| \leqslant \left|\int_{t_0}^{t'} f(s, \varphi(s))\,\mathrm{d}s\right| \leqslant M|t' - t_0| \leqslant Ma,$$

where t' is some number in I where the maximum is achieved. So f is a self-map if $a \leqslant \frac{b}{M}$. The quantity $M = M(f, a, b)$ does depend on a. But if $0 < a' < a$ then $M(f, a', b) \leqslant M(f, a, b$ as the supremum is taken on a smaller subset. This implies that if $a' \leqslant \frac{b}{M(f,a,b)}$ then $a' \leqslant \frac{b}{M(f,a',b)}$.

Now we will prove that $\mathit{\Gamma}$ is a contraction. Let $t \in I$ be such that

$$d_{\text{sup}}(\mathit{\Gamma}\varphi_1, \mathit{\Gamma}\varphi_2) = |\mathit{\Gamma}\varphi_1(t) - \mathit{\Gamma}\varphi_2(t)|.$$

Then by the definition of $\mathit{\Gamma}$:

$$\begin{aligned} |(\mathit{\Gamma}\varphi_1 - \mathit{\Gamma}\varphi_2)(t)| &= \left|\int_{t_0}^{t} f(s, \varphi_1(s)) - f(s, \varphi_2(s))\,\mathrm{d}s\right| \\ &\leqslant \int_{t_0}^{t} |f(s, \varphi_1(s)) - f(s, \varphi_2(s))|\,\mathrm{d}s \\ &\leqslant L\int_{t_0}^{t} |\varphi_1(s) - \varphi_2(s)|\,\mathrm{d}s \leqslant La \cdot d_{\text{sup}}(\varphi_1, \varphi_2), \end{aligned}$$

since f is Lipschitz continuous in y. So $\mathit{\Gamma}$ is a contraction if a is chosen so that $a < \frac{1}{L}$.

So the Picard's operator $\mathit{\Gamma}$ is a contraction on the complete metric space $\mathscr{C}$ and there hence exists a unique function $\varphi \in \mathscr{C}$ such that $\mathit{\Gamma}\varphi = \varphi$. This function is the unique solution of the IVP valid on $[t_0 - a, t_0 + a]$ where a satisfies the condition that $a < \min\{\frac{b}{M}, \frac{1}{L}\}$. ☐

5.2. Fractals

This section is a first glimpse at fractal objects, that is at subsets of $\mathbf{R}^n$ presenting some interesting self-similar properties.

5.2.1. The Cantor Set

We begin our investigation of fractal objects, by looking at possibly the simplest of them: the Cantor set, which is a subset K of $\mathbf{R}$ obtained by removing more and more chunks from $[0,1]$.

Definition 5.2.1.

The **Cantor Set** $K \subseteq \mathbf{R}$ is defined by the following inductive procedure. Let $K_0 = [0,1]$ and define $K_1 \subseteq K_0$ by deleting its open middle third; so $K_1 = [0,\frac{1}{3}] \cup [\frac{2}{3},1]$ is the union of two closed intervals, each of length $\frac{1}{3}$. Let K_n be the union of 2^n closed intervals, each of length $\frac{1}{3^n}$, and define K_{n+1} by deleting the open middle third of each; so K_{n+1} is the union of 2^{n+1} closed intervals, each of length $\frac{1}{3^{n+1}}$. Then

$$K := K_0 \cap K_1 \cap \cdots \cap K_n \cap \cdots = \bigcap_{n\in\mathbf{N}} K_n.$$

Figure 5.1.: Construction of the Cantor Set.

Since the endpoints 0 and 1 are never removed from $[0,1]$, we conclude that K is non-empty. Moreover, we will see later that each K_n is compact and so is K. We want a way to describe all points in K. In order to do so, we will study the ternary expansion of real numbers.

First, let us recall that if $a \in [0,1]$ is a real number, we can compute a **decimal expansion** by

$$a = 0.a_1a_2\dots a_n \dots := \sum_{n=1}^{\infty} \frac{a_n}{10^n}, \quad a_n \in \{0,\dots,9\}.$$

e.g. $\frac{1}{2} = 0.5 = 0.4999\dots$ and $1 = 0.999\dots$. This expansion is not unique, as demonstrated above. However, each $a \in [0,1)$ admits a unique decimal expansion such that a_i is not eventually 9. In other words, each $a \in [0,1)$ admits a unique decimal expansion such that $\forall N \exists n \geqslant n : a_n \neq 9$.

We already encountered (for example in Tutorial 5, Question 5) the **binary expansion** of a $a \in [0,1]$:

$$a = 0.a_1 a_2 \dots a_n \dots := \sum_{n=1}^{\infty} \frac{a_n}{2^n}, \quad a_n \in \{0,1\}.$$

e.g. $\frac{1}{2} = 0.1 = 0.0111\dots$ and $1 = 0.111\dots$. Similarly to the decimal expansion case, each $a \in [0,1)$ admits a unique binary expansion such that a_i is not eventually 1.

To study the Cantor set, we will need to look at the **ternary expansion** of a real number $a \in [0,1]$:

$$a = 0.a_1 a_2 \dots a_n \dots := \sum_{n=1}^{\infty} \frac{a_n}{3^n}, \quad a_n \in \{0,1,2\}.$$

Ternary expansions are unique except that $0.a_1 \dots a_{n-1} a_n 222\dots = 0.a_1 \dots a_{n-1} a_n' 000\dots$ whenever $a_n \neq 2$ and $a_n' = a_n + 1$. So $0 = 0.000\dots$ and $1 = 0.222\dots$.

We also know that 0, 1, $\frac{1}{3}$ and $\frac{2}{3}$ are all in K while $\frac{1}{2}$ is not. If we look at their ternary expansion, we have

$$\frac{1}{3} = 0.0222\dots = 0.1000\dots \quad \text{and} \quad \frac{2}{3} = 0.2000\dots = 0.1222\dots$$

whilst $\frac{1}{2} = 0.111\dots$. This phenomenon is not random:

Theorem 5.2.2.

The Cantor Set K consists of all numbers $a \in [0,1]$ which have a ternary expansion containing only 0s and 2s.

Proof. The middle third of K_0 consists of exactly a whose ternary expansions have $a_1 = 1$. So $a \in K_1$ if and only if it has a ternary expansion with $a_1 \neq 1$. This includes $\frac{1}{3}$ and $\frac{2}{3}$.

The middle third of $[0, \frac{1}{3}] \subseteq K_1$ consists exactly of those a which have a ternary expansion with $a_1 = 0$ and $a_2 = 1$. Similarly, the middle third of $[\frac{2}{3}, 1]$ consists of those a which have a ternary expansion with $a_1 = 2$ and $a_2 = 1$. So $a \in K_2$ if and only if it has an expansion satisfying $a_1, a_2 \neq 1$.

Proceeding inductively, the middle third of each of the 2^n closed intervals that make up K_n consists of those a whose ternary expansions have $a_1 = 0$ or $2, \dots, a_n = 0$ or 2 and $a_{n+1} = 1$. So $a \in K_{n+1}$ if and only if it has an expansion satisfying $a_m \neq 1$ for all $m \leqslant n+1$. This holds for all $n \geqslant 1$ and the result follows. ☐

It might seem that the only numbers left in K are the endpoints of the closed intervals, i.e. numbers of the form $\frac{k}{3^n}$, but actually $\frac{1}{4} \in K$ since its ternary expansion is $\frac{1}{4} = 0.020202\dots$. However, one can show that points of the form $\frac{k}{3^n}$ are dense in K.

It follows from the above description of K, that is infinite and in fact is uncountable.

Corollary 5.2.3. *The Cantor Set is uncountable.*

Proof. The interval $[0,1]$ is uncountable (because it contains $(0,1)$ which is in bijection with $\mathbf{R}$). We can construct an explicit injection $f\colon [0,1] \to K$ in the following way. For any $a \in [0,1)$ write $a = 0.a_1a_2\dots$ for the unique *binary* expansion of a such that a_i is not eventually 1 . Now, let $f(a)$ be the unique element of $[0,1)$ whose *ternary* expansion is $b = 0.b_1b_2\dots$ where

$$b_i = \begin{cases} 0 & \text{if } a_i = 0 \\ 2 & \text{if } a_i = 1. \end{cases}$$

Finally, define $f(1) := 1$. One easily check that f is well-defined and an injection from $[0,1]$ to K. We conclude that the cardinality of K is at least the cardinality of $\mathbf{R}$. Since $K \subseteq \mathbf{R}$, we conclude that K and $\mathbf{R}$ have the same cardinality, and thus K is uncountable. ☐

In the above proof, we have given an explicit injection $f\colon [0,1] \to K$, alas this was not a bijection. For example, $0.22\dots$ has no preimage. However, since $K \subseteq [0,1]$, we can use the Cantor–Bernstein–Schroeder theorem to construct a bijection from $[0,1] \to K$. Alternatively, it is possible to slightly modify the definition of f to directly obtain a bijection between $[0,1]$ and K. Let's compute the sums of the lengths of the intervals that are removed from $[0,1]$ to create K:

$$\frac{1}{3} + \frac{2}{3^2} + \frac{2^2}{3^3} + \dots + \frac{2^{n-1}}{3^n} + \dots = \frac{1}{3} \cdot \frac{1}{1 - \frac{2}{3}} = 1.$$

So the Cantor Set has length 0 but still the same number of elements as $[0,1]$!

We now investigate a bit more the topological properties of the Cantor Set.

Lemma 5.2.4. *The Cantor Set is closed.*

Proof. The closed interval $K_0 = [0,1]$ is closed and K_1 is obtained by removing an open interval, this implies that K_1 is closed. Assume K_n is closed. Then K_{n+1} is obtained by removing a finite number of open intervals and so K_{n+1} is also closed. So $K = \bigcap_{n \in \mathbf{N}} K_n$ is an infinite intersection of closed sets and is therefore closed. ☐

Corollary 5.2.5. *The Cantor Set is compact and complete.*

Proof. Clearly K is bounded, and since it is also closed, by the Heine–Borel Theorem, it is compact. Also, every compact metric space is complete. ☐

Proposition 5.2.6. *The Cantor Set in nowhere dense and $\partial K = K$.*

Proof. Since $\partial K = \overline{K} \setminus K^\circ$ and $\overline{K} = K$ we just need to show that $K^\circ = \varnothing$. Let $a \in K$ have ternary expansion $0.a_1a_2\dots$ and let $\varepsilon > 0$. Choose $N \in \mathbf{N}$ such that any real number of the form $b = 0.0\dots0b_Nb_{N+1}\dots$ (ternary expansion) must satisfy $0 < b < \varepsilon$. Then $0.a_1a_2\dots a_{N-1}111\dots$ lies in $B_\varepsilon(a) = (a-\varepsilon, a+\varepsilon)$ but not in K. Therefore, $K^\circ = \varnothing$.

Now $(\overline{K})^\circ = \varnothing$, so K is nowhere dense. ☐

Example 5.2.7. It is interesting to compare the Cantor Set K with $\mathbf{Q} \cap [0,1] \subseteq \mathbf{R}$ which is countable, not closed and therefore not compact and $\partial(\mathbf{Q} \cap [0,1]) = [0,1] \neq \mathbf{Q} \cap [0,1]$.

Observe that the Cantor Set is a *self-similar fractal* in the sense that $K \cap [0, \frac{1}{3^n}]$ (and every other similar interval) is in bijection with K; as can be seen:

$$\begin{aligned} K \cap \left[0, \frac{1}{3^n}\right] &\longrightarrow K \\ 0.0 \ldots 0a_{n+1}a_{n+2} \ldots &\longmapsto 0.a_{n+1}a_{n+2} \ldots \end{aligned}$$

Higher dimension analogs of the Cantor Set can be constructed. For example, the chapter illustration on page 81 shows the first steps in the construction of the **Menger**[2] **Sponge** (or **Sierpiński cube**[3]), which is a 3-dimensional analog of the Cantor set.

To go further

The Cantor Set might look like just recreational mathematics, but the truth is far from that. Indeed, the Cantor Set is a ubiquitous object that plays important roles in many fields of mathematics. It for example appears naturally when studying topology, Brownian motion, infinite rooted trees, … To cite just one realisation of the Cantor set: it is the set of vertices of the infinite dimensional cube $[0, 1]^{\mathbf{N}}$.

5.2.2. The Hausdorff Metric

In order to construct other fractals, we introduce a metric on the space of subsets of a metric space.

Let (X, d) be a metric space with a subset $B \subseteq X$ and a point $x \in X$. The **distance between x and B** is defined to be

$$d(x, B) := \inf\{d(x, b) \mid b \in B\}.$$

It is easy to see that if $x \in B$, then $d(x, B) = 0$, but the converse is not true in general. We can also show that if B is compact and non-empty, then there exists $b^* \in B$ such that $d(x, B) = d(x, b^*)$. See Tutorial 12, Question 4 for the details.

If $A, B \subseteq X$ are non-empty compact subspaces, then we define the **distance from A to B** by

$$d(A, B) := \sup\{d(a, B) \mid a \in A\}.$$

Remark 5.2.8.

In general, the "distance from A to B" is not a distance! Indeed, it does not need to be symmetric: it is not true that $d(A, B)$ and $d(B, A)$ are equal in general. As subspaces of $\mathbf{R}$ we see that for $A = [0, 3]$ and $B = [0, 1] \cup [2, 3]$ that

$$d(A, B) = \sup_{a \in A} d(a, B) = \sup_{a \in A}\{\min\{a - 1, 2 - a\} \mid a \notin B\} = \frac{1}{2}.$$

But $d(B, A) = 0$ as $B \subseteq A$.

---

[2] Karl Menger (1902–1985).

[3] Wacław Franciszek Sierpiński (1882–1969).

Luckily, there is an easy trick that allows to overcome the lack of symmetry of d.

Definition 5.2.9.

Let $\mathcal{H}$ be the set of all non-empty compact subsets of X and define

$$d_H(A, B) := \max\{d(A, B), d(B, A)\}.$$

This time, we obtain a real distance on the set $\mathcal{H}$. This is known as the **Hausdorff space**[4] associated to (X, d), which measures how far two compact subspaces of (X, d) are from each other.

Proposition 5.2.10. *The pair* $(\mathcal{H}, d_H)$ *forms a metric space.*

Proof. For (M1) and (M2), we see that $d_H(A, B) = 0$ if and only if $A \subseteq B$ and $B \subseteq A$. So $d_H(A, B) = 0$ if and only if $A = B$. The axiom (M3) is automatically satisfied as the definition is symmetric. For the triangle inequality (M4) it is enough to show that $d(A, B) \leqslant d(A, C) + d(C, B)$ because interchanging the role A and B gives the desired result. For any $a \in A$ we have

$$\begin{aligned} d(a, B) &= \inf_{b\in B} d(a, b) \\ &\leqslant \inf_{b\in B} d(a, c) + d(c, b), && \forall c \in C \\ &= d(a, c) + d(c, B), && \forall c \in C \\ &\leqslant d(a, c) + d(C, B), && \forall c \in C. \end{aligned}$$

Minimising over C gives

$$d(a, B) \leqslant d(a, C) + d(C, B), \quad \forall a \in A.$$

Maximising over A on the right side gives

$$d(a, B) \leqslant d(A, C) + d(C, B), \quad \forall a \in A.$$

Then maximising over A on the left side gives what we are looking for. ☐

The Hausdorff distance between two non-empty compact subsets can alternatively be computed using *neighbourhood.*

Proposition 5.2.11. *If* $A \subseteq X$ *and* $\varepsilon > 0$, *the* ε***-neighbourhood*** *of* A *is given by*

$$A_\varepsilon := \{x \in X \mid d(x, A) < \varepsilon\}.$$

Then $d_H(A, B) < \varepsilon$ *if and only if* $A \subseteq B_\varepsilon$ *and* $B \subseteq A_\varepsilon$. *It follows that* $d_H(A, B) = \inf\{\varepsilon \geqslant 0 \mid A \subseteq B_\varepsilon \text{ and } B \subseteq A_\varepsilon\}$.

Proof. The proof is left to the reader (see Tutorial 12, Question 5). ☐

The space $(\mathcal{H}, d_H)$ inherits some interesting properties from (X, d).

---

[4] Felix Hausdorff (1868–1942).

Theorem 5.2.12.

Let (X, d) be a metric space and let $(\mathcal{H}, d_H)$ be the associated Hausdorff metric space. Then

1. *If (X, d) is complete, then $(\mathcal{H}, d_H)$ is complete;*
2. *If (X, d) is compact, then $(\mathcal{H}, d_H)$ is compact;*
3. *If d and d' are topologically equivalent metrics on X, then d_H and d'_H are topologically equivalent metrics on $\mathcal{H}$.*

Sketch of the Proof. "1" Suppose (A_n) is a Cauchy sequence in $(\mathcal{H}, d_H)$. Define A to be the set of all $x \in X$ with the property that there exists a sequence of points $x_n \in A_n$ with $x_n \to x$. It is elementary but fastidious to prove that A is compact and non-empty and is the limit of (A_n) in the Hausdorff metric.

"2" We will use that a space is compact if and only if it is complete and totally bounded, Theorem 4.1.7. We already know that if X is compact, it is complete and so $\mathcal{H}$ is complete. Since X is totally bounded, for any $\varepsilon > 0$ there exists $S \subseteq X$ such that $X \subseteq \bigcup_{x\in S} B_\varepsilon(x; d)$. Define

$$\mathcal{S} := \{\overline{B}_\varepsilon(x; d) \mid x \in S\}.$$

Since X is compact, all the closed balls $\overline{B}_\varepsilon(x; d)$ are compact and thus $\mathcal{S} \subseteq X$. One can show, using Proposition 5.2.11, that $\mathcal{H} \subseteq \bigcup_{K\in\mathcal{S}} B_\varepsilon(x; d_H)$, that is that $\mathcal{S}$ is a finite ε-net for $\mathcal{H}$. This implies that $\mathcal{H}$ is totally bounded and hence compact.

"3" Firstly, observe that the set $\mathcal{H}$ does not depend on the choice of d or d'. Secondly, using that d and d' are topologically equivalent one can show that $A_{\varepsilon;d}$ is not only open for d, but also for d'. It follows that $B_r(A; d_H)$ is open for d'. □

The elements of $\mathcal{H}$ are not mere points, but compact subsets of X. So if A and B are two elements of $\mathcal{H}$, their union $A \cup B$ is still an element of $\mathcal{H}$. It is natural to ask how the distance d_H behaves when taking the union.

Lemma 5.2.13. *Let (X, d) be a metric space with associated Hausdorff space $(\mathcal{H}, d_H)$ and let A_1, A_2, B_1, B_2 be 4 elements of $\mathcal{H}$. Then:*

$$d_H(A_1 \cup A_2, B_1 \cup B_2) \leqslant \max\{d_H(A_1, B_1), d_H(A_2, B_2)\}.$$

Proof. It is easy to see that

$$\begin{aligned} d(A_1 \cup A_2, B) = \sup_{a\in A_1\cup A_2} d(a, B) &= \max\{\sup_{a\in A_1} d(a, B), \sup_{a\in A_2} d(a, B)\} \\ &= \max\{d(A_1, B), d(A_2, B)\}. \end{aligned}$$

for any $B \in \mathcal{H}$, which implies

$$d(A_1 \cup A_2, B_1 \cup B_2) = \max\{d(A_1, B_1 \cup B_2), d(A_2, B_1 \cup B_2)\}. \tag{5.1}$$

We also have

$$\begin{aligned} d(A, B_1 \cup B_2) &= \sup_{a \in A} d(a, B_1 \cup B_2) = \sup_{a \in A} \inf_{b \in B_1 \cup B_2} d(a, b) \\ &= \sup_{a \in A} \min\{d(a, B_1), d(a, B_2)\} \leqslant \min\{d(A, B_1), d(A, B_2)\}, \end{aligned}$$

which implies

$$d(A, B_1 \cup B_2) \leqslant d(A, B_i), \quad \text{for } i = 1, 2. \tag{5.2}$$

Substituting Equation (5.2) into Equation (5.1) gives

$$d(A_1 \cup A_2, B_1 \cup B_2) \leqslant \max\{d(A_1, B_1), d(A_2, B_2)\}.$$

So

$$\begin{aligned} d_H(A_1 \cup A_2, B_1 \cup B_2) &= \max\{d(A_1 \cup A_2, B_1 \cup B_2), d(B_1 \cup B_2, A_1 \cup A_2)\} \\ &\leqslant \max\{d(A_1, B_1), d(A_2, B_2), d(B_1, A_1), d(B_2, A_2)\} \\ &= \max\{d_H(A_1, B_1), d_H(A_2, B_2)\}. \end{aligned}$$ ☐

The metric d_H also behave nicely with respect of self-maps of X. Before making this more precise, let us introduce a new notion that generalises contractions.

Definition 5.2.14.

Let $f\colon X \to X$ be a self-map of X. Its **Lipschitz constant** is

$$\operatorname{Lip} f := \sup_{x \neq y} \frac{d(f(x), f(y))}{d(x, y)} \tag{5.3}$$

if the supremum exists. Otherwise, we set $\operatorname{Lip} f = +\infty$.

The name Lipschitz constant comes from the fact that $\operatorname{Lip} f \neq +\infty$ if and only if f is Lipschitz. More precisely, we have:

Lemma 5.2.15. *If* $\operatorname{Lip} f = K \in \mathbf{R}$, *then* $K \geqslant 0$, *and*

$$d(f(x), f(y)) \leqslant K d(x, y), \quad \text{for all } x, y \in X,$$

and, moreover, $\operatorname{Lip} f$ *is the least such* K*.*

Proof. If $\operatorname{Lip} f \in \mathbf{R}$, then, since it is the supremum of a set of non-negative numbers, $K \geqslant 0$. Also, for $x \neq y$,

$$\frac{d(f(x), f(y))}{d(x, y)} \leqslant \operatorname{Lip} f = K,$$

which gives the inequality as the case $x = y$ is trivial.

Moreover, $\operatorname{Lip} f$ is the least such K otherwise it would contradict it being the supremum as defined in Equation (5.3). ☐

It follows from the above result that $\operatorname{Lip} f < 1$ if and only if f is a contraction.

The Lipschitz constant is an useful to study how the distances in Hausdorff space changes under the application of a self-map:

Lemma 5.2.16. *Let (X, d) be a metric space with associated Hausdorff space $(\mathcal{H}, d_H)$. Suppose that $f\colon X \to X$ is a self-map. Then:*

$$d_H(f(A_1), f(A_2)) \leqslant \operatorname{Lip} f \cdot d_H(A_1, A_2)$$

for any $A_1, A_2 \in \mathcal{H}$.

Proof. We have

$$\begin{aligned} d(f(A_1), f(A_2)) &= \sup_{a \in A_1} d(f(a), f(A_2)) = \sup_{a \in A_1} \inf_{a' \in A_2} d(f(a), f(a')) \\ &\leqslant \sup_{a \in A_1} \inf_{a' \in A_2} K d(a, a') = K d(A_1, A_2), \end{aligned}$$

where $K = \operatorname{Lip} f$. Similarly, $d(f(A_2), f(A_1)) \leqslant K d(A_2, A_1)$. This implies the result. □

5.2.3. Iterated Function Systems

We will now use the Hausdorff space $\mathcal{H}$ to build fractal objects from contractions.

Definition 5.2.17.

An **iterated function system** (**IFS**) is a finite set of contractions on a complete metric space (X, d), that is,

$$\{f_i\colon X \to X \mid i = 1, 2, \dots, N\},$$

where $N \in \mathbf{N}$ and each f_i is a contraction on the complete metric space X.

We are interested in subsets $A \subseteq X$ that are **invariant** under an IFS, that is,

$$A = \bigcup_{i=1}^{N} f_i(A) = f_1(A) \cup f_2(A) \cup \dots \cup f_N(A).$$

Given a compact subset $K \subseteq X$ we define

$$f(K) := \bigcup_{i=1}^{N} f_i(K) = f_1(K) \cup f_2(K) \cup \dots \cup f_N(K). \tag{5.4}$$

This induces a self-map $f\colon \mathcal{H} \to \mathcal{H}$ as all contractions are continuous, the continuous image of compact spaces are compact, and finite unions of compact spaces are compact. So a compact subset K is invariant for an IFS $\{f_i\colon X \to X \mid 1 \leqslant i \leqslant N\}$ if and only if it is a fixed point of the corresponding self-map $f\colon \mathcal{H} \to \mathcal{H}$. The Banach Contraction Mapping Theorem (Theorem 4.2.5) ensures that every contracting map has a unique fixed point, so it seems reasonable to try to use it to produce invariant subsets for an IFS.

Theorem 5.2.18 (Hutchinson,[5] 1981)**.**

Given an iterated function system on a complete metric space X, there exists a unique non-empty compact invariant subset.

Proof. Let $\{f_i \colon X \to X \,|\, i = 1, 2, \dots, N\}$ be an IFS on X and let $K_i < 1$ be the contraction constant for f_i. Since X is complete we know by Theorem 5.2.12 that its associated Hausdorff space $(\mathcal{H}, d_H)$ is also complete. We will show that the map $f \colon \mathcal{H} \to \mathcal{H}$ given in Equation (5.4) is a contraction.

Let $K = \max\{K_i \mid i = 1, 2, \dots, N\} < 1$. Then, for $A, B \in \mathcal{H}$,

$$\begin{aligned} d_H(f(A), f(B)) &= d_H\left(\bigcup_{i=1}^{N} f_i(A), \bigcup_{i=1}^{N} f_i(B)\right) \\ &\leqslant \max\{d_H(f_i(A), f_i(B)) \mid i = 1, 2, \dots, N\} && \text{by Lemma 5.2.13} \\ &\leqslant \max\{K_i d_H(A, B) \mid i = 1, 2, \dots, N\} && \text{by Lemma 5.2.16} \\ &\leqslant K d_H(A, B). \end{aligned}$$

So f is a contraction on the complete metric space $(\mathcal{H}, d_H)$ and so there is a unique point $C \in \mathcal{H}$ such that $f(C) = C$, that is, there is a unique non-empty compact invariant subspace C of X. ☐

Obviously, if there is only one contraction f_1 in our IFS, with fixed point p, then f has fixed point the compact subset $S = \{p\}$. As an example, let $f \colon \mathbf{R} \to \mathbf{R}$ be the contraction given by $f(x) = \frac{x}{2}$, which has 0 as its unique fixed point. Hutchinson's Theorem tells us that for *any* compact subset $S_0 \subseteq \mathbf{R}$, if we define $S_{k+1} := f(S_k)$, that the sequence (S_n) converges to $\{0\}$, as $n \to \infty$. Taking $S_0 = \{1\}$ gives us $S_n = \{\frac{1}{2^n}\}$. Whilst taking $S_0 = [-1, 1]$ gives us $S_n = [-\frac{1}{2^n}, \frac{1}{2^n}]$.

5.2.4. The Chaos Game

It becomes much more interesting when an IFS consists of several contractions. Our unique invariant subsets often take the form of interesting *fractals*!

Our first such example will consist of two contractions on $\mathbf{R}$:

$$\left\{f_1(x) = \frac{x}{3}, f_2(x) = 1 + \frac{(x-1)}{3}\right\}.$$

Then our unique compact invariant set, which we will denote by C, is the famous Cantor Set! We can generate this set by starting with a compact subset $S_0 \subseteq \mathbf{R}$ and then define $S_{n+1} = f(S_n)$ as before.

Exercise 5.2.19. Calculate S_i, for $i = 0, 1, \dots, 3$, when $S_0 = [0, 1]$ and when $S_0 = \{0, 1\}$.

---

[5] John Edward Hutchinson (1946).

The *Chaos Game* comes about when we want to plot an IFS using a computer. This is usually done in $\mathbf{R}^2$. We begin by choosing a point $x_0 \in S \subseteq \mathbf{R}^2$ (we can take it as one of the fixed points of the contractions in our IFS). Then we *randomly* choose one of our contractions f_i from our IFS and apply this to x_0 to obtain $x_1 = f_i(x_0)$. We program the computer to do this for a large number of iterations (e.g. 1,000,000) randomly choosing a contraction from our IFS every time. Plotting all of these points in $\mathbf{R}^2$ will converge to the invariant set S of our IFS *every* time, even though the choices of which contraction to use are random!

Another famous IFS is known as the **Sierpiński Triangle** which is a fractal with the overall shape of an equilateral triangle; see the picture on the title page of these notes for an illustration (the picture on the title pace is actually (half of) the complement of the Sierpiński triangle: the purple part in the picture is what is removed from the triangle).

Exercise 5.2.20. Do some research and write a brief paragraph about the Sierpiński Triangle. What three contractions are used to generate it? Go to the website `https://larryriddle.agnesscott.org/ifs/siertri/siertri.htm` and play the “The Sierpinski Gasket ‘Chaos Game’ Challenge”.

One of the most fascinating things about the fractals coming from iterated function systems is that they can often be used to model objects in nature. A famous example of this is **Barnsley Fern** which is generated by the following four contractions of $\mathbf{R}^2$:

$$f_1(x) = \begin{pmatrix} 0 & 0 \\ 0 & 0.16 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad f_2(x) = \begin{pmatrix} 0.85 & 0.04 \\ -0.04 & 0.85 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1.6 \end{pmatrix},$$

$$f_3(x) = \begin{pmatrix} 0.2 & -0.26 \\ 0.23 & 0.22 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1.6 \end{pmatrix}, \quad f_4(x) = \begin{pmatrix} -0.15 & 0.28 \\ 0.26 & 0.24 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0.44 \end{pmatrix}.$$

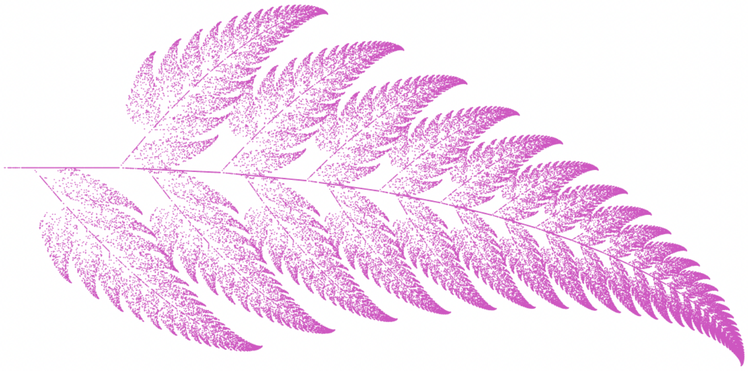


Figure 5.2.: A Barnsley Fern with $100,000$ iterations.

When we play the Chaos Game with the Barnsley Fern we randomly choose one of the four contractions of $\mathbf{R}^2$ with a certain probability p. For the standard Barnsley Fern

$$p(f_1) = 0.01, \quad p(f_2) = 0.85, \quad p(f_3) = 0.07, \quad p(f_4) = 0.07.$$

Exercise 5.2.21. Go to the website `https://www.chradams.co.uk/fern/maker.html` to see the standard Barnsley Fern and play around with this "fern generator". Ask it to plot a small number of points (e.g. 100) and then increase this by a factor of 10 each time to see the fern appear before your eyes. Play around with some of the values defining the contractions and their associated probabilities to generate your own unique fern!

5.3. Data Science and Clustering Algorithms

Data science, also sometimes called *big data*, is a field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from extremely large amounts of data. It is a form of statistics using data analysis, machine learning etc. as well as metric spaces!

A common technique for data analysis is *cluster analysis* which groups a set of objects together in such a way that objects in the same group (called a *cluster*) are more similar (in some sense) to each other than to those in other clusters. To do this we can use the notion of distance/metric: so that nearby objects are more similar than objects that are far away.

In this section, we will present one application of metric spaces to data-science: the **k-means clustering algorithm**.

We partition n observations into k clusters in which each observation belongs to a cluster with the nearest *mean*. We are assuming that n is very very large and that k is much more manageable. The idea is to minimise the distance between points in one cluster and a point which is the centre (mean) of another cluster. An algorithm to do this is:

1. Randomly select k points as the means;
2. Associate each point to its closest mean to create k clusters;
3. The *total cost* is the sum of the distances of points in a cluster to their means;
4. For each mean m and for each non-mean p:
   a) swap m and p, associate each point to its closest mean and recompute the total cost;
   b) if the total cost increases, then undo the swap, otherwise keep it.

There has been recent work (for example: *A Mathematical Theory for Clustering in Metric Spaces* by Chang, Liao, Chen and Liou, 2016) exploring the mathematical theory of clustering in metric spaces. You will be able to find the pdf file for the relevant readable paper on the Learning Mall. We will now try to explain the main results and you will be able to find the relevant proofs in the paper.

Definition 5.3.1 (Relative Distance)**.**

Let (X, d) be a finite metric space with n elements. For points $x, y \in X$, the **relative distance** from x to y is given by

$$\mathrm{RD}(x \mid y) := \frac{1}{n} \sum_{z \in X} (d(x, y) - d(x, z)).$$

The **relative distance** from a random point to y is

$$\mathrm{RD}(y) := \frac{1}{n} \sum_{z \in X} \mathrm{RD}(z \mid y).$$

Remark 5.3.2.

Note that $\mathrm{RD}(x \mid y)$ is not symmetric and may be negative.

Let $S_1, S_2 \subseteq X$ be two subsets of a finite metric space X. Then the **relative distance** from S_1 to S_2 is

$$\mathrm{RD}(S_1 \mid S_2) := \frac{1}{|S_1| \cdot |S_2|} \sum_{x \in S_1, y \in S_2} \mathrm{RD}(x \mid y).$$

Definition 5.3.3 (Cohesion)**.**

The **cohesion** between two points $x, y \in X$ of a finite metric space is given by

$$\gamma(x, y) := \mathrm{RD}(y) - \mathrm{RD}(x \mid y).$$

The points x and y are **cohesive** if $\gamma(x, y) \geqslant 0$ and are **incohesive** if $\gamma(x, y) \leqslant 0$.

Cohesion satisfies the following properties:

Proposition 5.3.4. *For all $x, y \in X$ in a finite metric space:*

1. $\gamma(x, y) = \gamma(y, x)$;
2. $\gamma(x, x) \geqslant 0$;
3. $\gamma(x, x) \geqslant \gamma(x, y)$;
4. $\sum_{y \in X} \gamma(x, y) = 0$.

Again, for two subsets $S_1, S_2 \subseteq X$ of a finite metric space, we define

$$\gamma(S_1, S_2) := \sum_{x \in S_1, y \in S_2} \gamma(x, y),$$

with S_1 and S_2 **cohesive** if $\gamma(S_1, S_2) \geqslant 0$ and **incohesive** if $\gamma(S_1, S_1) \leqslant 0$.

Definition 5.3.5 (Cluster)**.**

Let X be a finite metric space. A subset $S \neq \varnothing$, $S \subseteq X$, is a **cluster** if $\gamma(S, S) \geqslant 0$, that is if S is cohesive to itself.

Being a cluster admits many characterisations:

Theorem 5.3.6.

Let S be a non-empty subset of X with $S \neq X$. Then the following are equivalent:

I. *S is a cluster;*

II. *$X \setminus S$ is a cluster;*

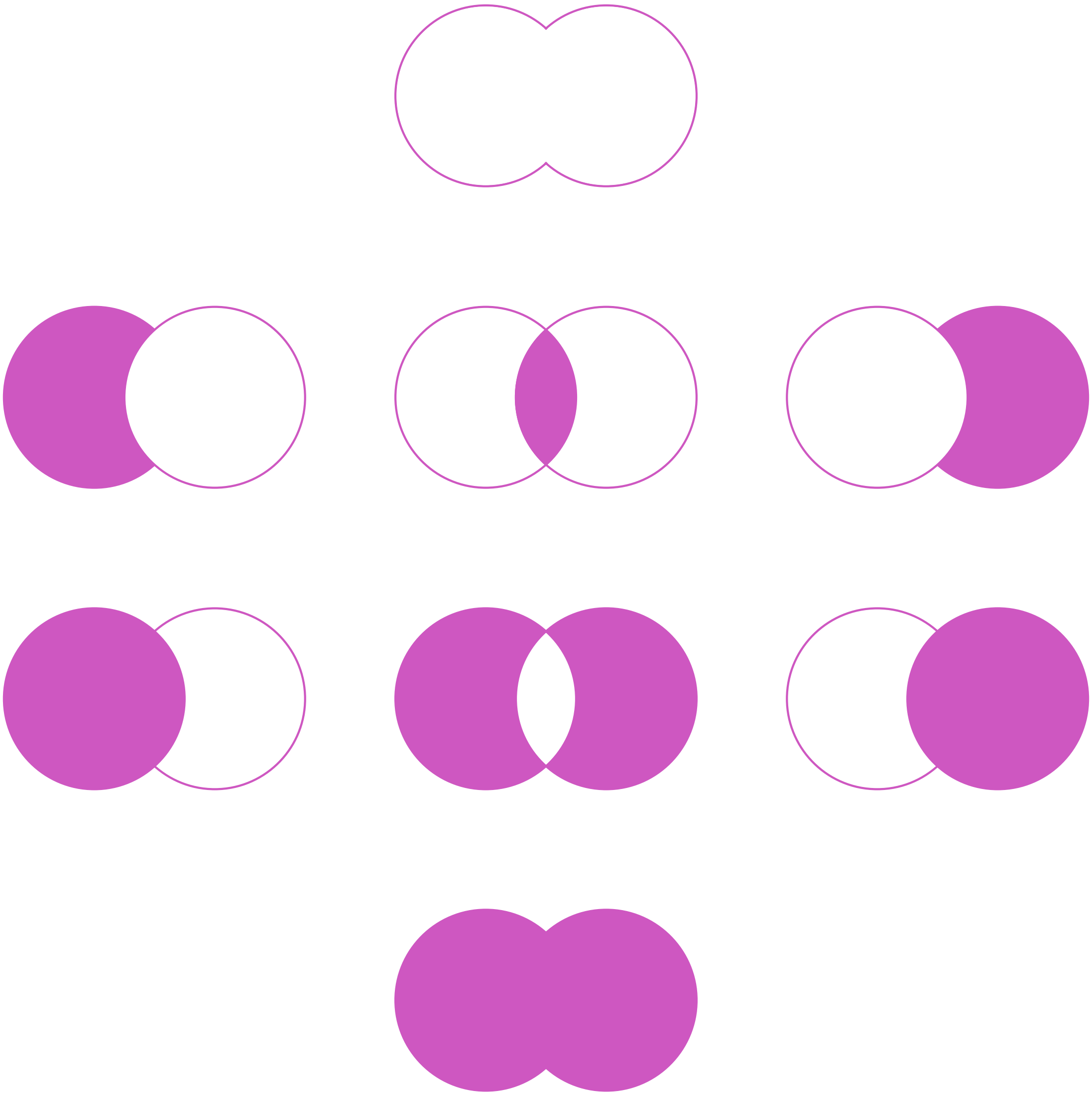
III. $\gamma(S, X \setminus S) \leqslant 0$*;*

IV. $\gamma(S, S) \geqslant \gamma(S, X \setminus S)$*.*

An algorithm for creating clusters is given by first putting each element into its own cluster, i.e. start with $S_i = \{x_i\}$ for $i = 1, \ldots, n$. Now compute $\gamma(S_i, S_j) = \gamma(x_i, x_j)$, for all i, j. If there exists i, j such that $\gamma(S_i, S_j) > 0$ then merge S_i and S_j into a new set $S_k = S_i \cup S_j$. Keep doing this process until it finishes.

APPENDIX A.

SETS AND FUNCTIONS



A.1. Sets and Operations on them

This appendix is devoted to some reminders about sets and operations on them. In this section, we will present many properties. These properties are intended as references and reminders and not as a list to learn by heart. All properties involving finitely many subsets can be recovered from Venn diagrams. In practice, using Venn diagrams is often the fastest and easiest way to find the wanted identity.

We will not present proofs here.

Definition A.1.1.

For A and B two subsets of the same set S, we define their **union**, **intersection** and **complement** by

$$\begin{aligned} A \cup B &:= \{x \in S \mid x \in A \text{ or } x \in B\}, \\ A \cap B &:= \{x \in S \mid x \in A \text{ and } x \in B\}, \\ A \setminus B &:= \{x \in S \mid x \in A \text{ and } x \notin B\}. \end{aligned}$$

If (A_i) is a arbitrary collection of subsets (with $I \neq \varnothing$), then one also define the corresponding arbitrary intersection and union:

$$\begin{aligned} \bigcup_{i\in I} A_i &:= \{x \in S \mid \exists i \in I : x \in A_i\}, \\ \bigcap_{i\in I} A_i &:= \{x \in S \mid \forall i \in I : x \in A_i\}. \end{aligned}$$

Unions and intersections are commutative and associative operations. They are also distributive one over the other.

Proposition A.1.2. *For A, B and C subsets of the same set S the following hold:*

1. *Commutative property:*

$$A \cup B = B \cup A \quad \text{and} \quad A \cap B = B \cap A;$$

2. *Associative property:*

$$(A \cup B) \cup C = A \cup (B \cup C) \quad \text{and} \quad (A \cap B) \cap C = A \cap (B \cap C);$$

3. *Distributive property:*

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \quad \text{and} \quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

If $(A_i)_{i\in I}$ and $(B_j)_{j\in J}$ are arbitrary family of subsets of S, we also have

3*. *Infinite distributive property:*

$$\Big(\bigcap_{i\in I} A_i\Big) \cup B = \bigcap_{i\in I}(A_i \cup B) \quad \text{and} \quad \Big(\bigcup_{i\in I} A_i\Big) \cap B = \bigcup_{i\in I}(A_i \cap B),$$

and more generally

$$\Big(\bigcap_{i\in I} A_i\Big) \cup \Big(\bigcap_{j\in J} B_j\Big) = \bigcap_{\substack{i\in I\\ j\in J}} (A_i \cup B_j) \quad \text{and} \quad \Big(\bigcup_{i\in I} A_i\Big) \cap \Big(\bigcup_{j\in J} B_j\Big) = \bigcup_{\substack{i\in I\\ j\in J}} (A_i \cap B_j)$$

Be careful that we have

$$\Big(\bigcap_{i\in I} A_i\Big) \cup \Big(\bigcap_{i\in I} B_i\Big) = \bigcap_{\substack{i\in I\\ j\in I}} (A_i \cup B_j) \neq \bigcap_{i\in I} (A_i \cup B_i).$$

Finally, we always have

$$\bigcup_{i\in I}\Big(\bigcap_{j\in J} A_{i,j}\Big) \subseteq \bigcap_{j\in J}\Big(\bigcup_{i\in I} A_{i,j}\Big),$$

but the equality does not necessarily hold in general.

We have two additional important properties, the second one involving complements.

Proposition A.1.3. *For $A \subseteq S$ the following hold:*

4. *Identity:*
$$A \cup \varnothing = A \quad \text{and} \quad A \cap S = A;$$

5. *Complement:*
$$A \cup (S \setminus A) = S \quad \text{and} \quad A \cap (S \setminus A) = \varnothing.$$

Before going further, let us show out to use Venn diagrams to verify identities.

Example A.1.4. Imagine we want to verify the double complement law $C \setminus (B \setminus A) = (A \cap V) \cup (C \setminus B)$ (see item 11 on page 101). We start by drawing to Venn diagrams, one for the left-hand side and one for the right-hand side of the equality, see Figure A.1.

On the left Venn diagram, we first identify $(B \setminus A)$ (in purple) by removing A from B. We then remove this from C to obtain $C \setminus (B \setminus A)$ (blue).

For the right Venn diagram, on one hand we look at the intersection of A and B (in brown) and on the other hand we remove B from C to obtain $C \setminus B$ (in blue). We then take the union of these two subsets to obtain $(A \cap V) \cup (C \setminus B)$ (everything that is in brown and/or in blue).

It is then straightforward to verify the equality between the left and the right hand sides of the double complement law.

The following follows from properties 1–5.

Proposition A.1.5. *For A, B and C subsets of the same set S the following hold:*

6. *Idempotent laws:*
$$A \cup A = A \quad \text{and} \quad A \cap A = A;$$

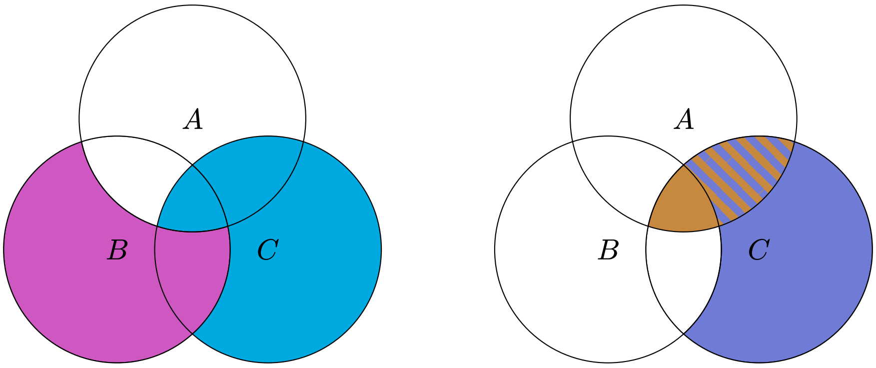


Figure A.1.: Venn diagrams for the double complement law. On the left, $B \setminus A$ in purple and $C \setminus (B \setminus A)$ in blue. On the right, $A \cap C$ in brown and $C \setminus B$ in blue.

7. *Domination laws:*

$$A \cup S = S \quad \text{and} \quad A \cap \varnothing = \varnothing;$$

8. *Absorption laws:*

$$A \cup (A \cap B) = A \quad \text{and} \quad A \cap (A \cup B) = A;$$

9. *De Morgan's laws:*

$$C \setminus (A \cup B) = (C \setminus A) \cap (C \setminus B) \quad \text{and} \quad C \setminus (A \cap B) = (C \setminus A) \cup (C \setminus B);$$

10. *Complement laws:*

$$A \setminus \varnothing = A, \quad \varnothing \setminus A = \varnothing \quad \text{and} \quad A \setminus A = \varnothing;$$

11. *Double complement laws:*

$$C \setminus (B \setminus A) = (A \cap C) \cup (C \setminus B) \quad \text{and} \quad (B \setminus A) \setminus C = B \setminus (A \cup C),$$

which implies

$$A \setminus (A \setminus B) = A \cap B, \quad S \setminus (S \setminus A) = A \quad \text{and} \quad S \setminus (B \setminus A) = A \cup (S \setminus B),$$

12. *Uniqueness of the complement:*

$$(A \cap B = \varnothing \;\&\; A \cup B = S) \implies B = S \setminus A;$$

13. *Misc identites:*

$$\begin{aligned} B \setminus A &= (S \setminus A) \cap B = (S \setminus A) \setminus (S \setminus B), \\ (B \setminus A) \cap C &= (B \cap C) \setminus (A \cap C) = (B \cap C) \setminus A = B \cap (C \setminus A), \\ (B \setminus A) \cup C &= (B \cup C) \setminus (A \setminus C). \end{aligned}$$

If $(A_i)_{i \in Y}$ is an arbitrary family of subsets of S, we also have

9.*

$$S \setminus \Big(\bigcup_{i \in I} A_i\Big) = \bigcap_{i \in I} (S \setminus A_i) \quad \text{and} \quad S \setminus \Big(\bigcap_{i \in I} A_i\Big) = \bigcup_{i \in I} (S \setminus A_i).$$

A.2. The Lattice of Subsets

The relation $A \subseteq B$ defined by $\forall x : x \in A \implies x \in B$ is an order relation on the subsets of S. In other words, $\subseteq$ is

1. Reflexive: $A \subseteq A$;
2. Antisymmetric: If both $A \subseteq B$ and $B \subseteq A$, then $A = B$;
3. Transitive: If both $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.

Not only $\subseteq$ is an order relation, it also interact nicely with intersections and unions.

Proposition A.2.1. *The set $\mathcal{P}(S)$ of all subsets of S together with $\subseteq$, $\cap$ and $\cup$ is a lattice. This means that for all A, B and C subsets of S we have:*

1. *Existence of a least element and a greatest element:*
$$\varnothing \subseteq A \subseteq S;$$
2. *Existence of joins:*
$$A \subseteq A \cup B \quad \text{and} \quad (A \subseteq C \;\&\; B \subseteq C) \implies A \cup B \subseteq C;$$
3. *Existence of meets:*
$$A \cap B \subseteq A \quad \text{and} \quad (C \subseteq A \;\&\; C \subseteq B) \implies C \subseteq A \cap B.$$

Inclusion can be alternatively characterised using intersection, using union or using complement.

Proposition A.2.2. *For A and B two subsets of a same set S, the following are equivalent:*

I. $A \subseteq B$;

II. $A \cap B = A$;

III. $A \cup B = B$;

IV. $A \setminus B = \varnothing$;

V. $S \setminus B \subseteq S \setminus A$.

Once again, the equivalences of Proposition A.2.2 can easily be verified using Venn diagrams.

A.3. Functions between Sets

For a function $f\colon S \to T$ and subsets $A \subseteq S$ and $C \subseteq T$ we write $f(A) = \{f(a) \mid a \in A\}$ and $f^{-1}(C) = \{s \in S \mid f(s) \in C\}$ for the image of A and the preimage of C. It follows from the definition that for every $A \subseteq S$ and $C \subseteq T$ we have

$$A \subseteq f^{-1}(f(A)) \quad \text{and} \quad C = f(f^{-1}(C)).$$

The function $f\colon \mathbf{R} \to \mathbf{R}, x \mapsto x^2$ and the subset $A = \{1\}$ shows that it is possible to have $A \subsetneq f^{-1}(f(A))$.

Proposition A.3.1. *Let $f\colon S \to T$ be a function between two sets. Let A, B be subsets of S and C, D be subsets of T. Then:*

1. *Images preserve unions:*

$$\begin{aligned} f(A \cup B) &= f(A) \cup f(B), \\ f(A \cap B) &\subseteq f(A) \cap f(B), && \text{(sic!)} \\ f(A \setminus B) &\supseteq f(A) \setminus f(B); && \text{(sic!)} \end{aligned}$$

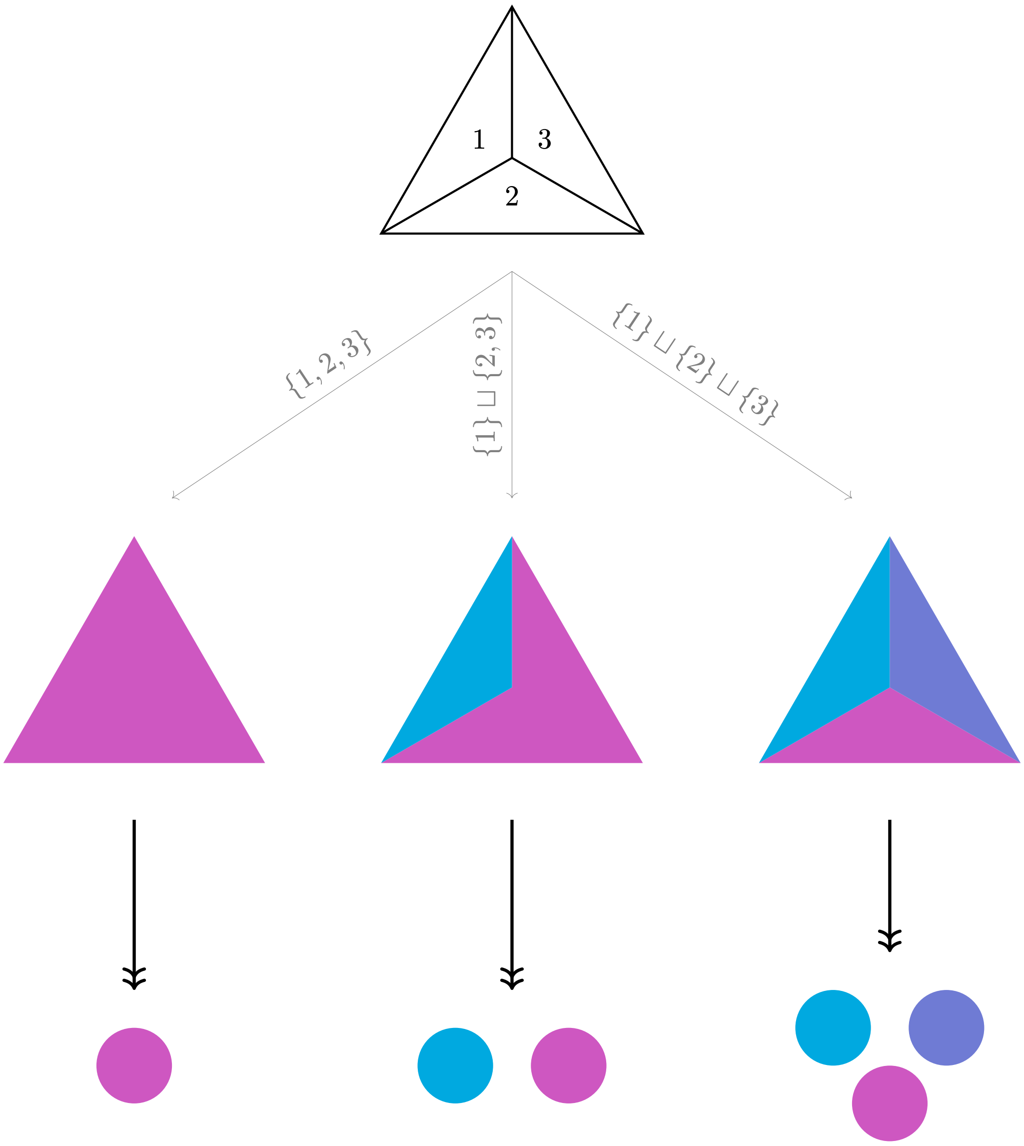
2. *Preimages preserve unions, intersections and complements:*

$$\begin{aligned} f^{-1}(C \cup D) &= f^{-1}(C) \cup f^{-1}(D), \\ f^{-1}(C \cap D) &= f^{-1}(C) \cap f^{-1}(D), \\ f^{-1}(C \setminus D) &= f^{-1}(C) \setminus f^{-1}(D). \end{aligned}$$

The function $f\colon \mathbf{R} \to \mathbf{R}, x \mapsto x^2$ and the subsets $A = \{1\}$ and $B = \{-1\}$ shows that it is possible to have $f(A \cap B) \subsetneq f(A) \cap f(B)$. The same function but with $A = \{-1, 1\}$ and $B = \{-1\}$ shows that it is possible to have $f(A \setminus B) \supsetneq f(A) \setminus f(B)$.

APPENDIX B.

EQUIVALENCE RELATIONS



B.1. Definitions and Examples

This appendix is devoted to some reminders about equivalence relation. We start with the definition and a few (in)formal examples.

Definition B.1.1.

An **equivalence relation** $\sim$ on a set X is a binary relation satisfies the following properties for all $x, y, z \in X$:

(1) $x \sim x$; *(reflexivity)*

(2) $x \sim y \implies y \sim x$; *(symmetry)*

(3) $(x \sim y \;\&\; y \sim z) \implies x \sim z$. *(transitivity)*

There are two equivalence relations we can always put on a set. The first one is the equality: $x \sim y$ if $x = y$. The second one is the trivial relation: $x \sim y$ for all $x, y \in X$. We are of course in general interested in less extreme examples of equivalence relations.

Example B.1.2. We give some examples of equivalence relations where you should check that the above three conditions hold.

1. Consider the set of all people on Earth. The relation

   “Person A $\sim$ Person B if Person A is related to Person B”

   gives an equivalence relation. So does “Person A has the same birthday as Person B”. This is not true for the statement “Person A is a friend of Person B” as transitivity may not hold.

2. For the set of integers $\mathbf{Z}$, the relation “$x \sim y$ if $x + y$ is even” is an equivalence relation. However, the relation “$x \sim y$ if $x + y$ is odd” is not as is it fails to be reflexive.

3. Consider the set $Q := \{(m, n) \in \mathbf{Z}^2 \mid n \neq 0\}$. We can define an equivalence relation on Q in the following way:

   $$(m, n) \sim (p, q) \text{ if } mq = pn.$$

   For example, $(1, 2) \sim (-3, -6)$.

4. Let H be a subgroup of a group G. We can define an equivalence relation on G by setting $g_1 \sim g_2$ if there exists $h \in H$ such that $g_1 h = g_2$ (equivalently, if $g_2^{-1} g_1 \in H$).

If we make explicit what it means to be a binary relation, we obtain the following formal definition of an equivalence relation on X:

Definition B.1.3.

Let X be a set. An **equivalence relation** on X is a subset $R \subseteq X \times X$ such that, for all $x, y, z \in X$,

(1) $(x, x) \in R$; *(reflexivity)*

(2) $(x, y) \in R \implies (y, x) \in R$; *(symmetry)*

(3) $((x, y) \in R \;\&\; (y, z) \in R) \implies (x, z) \in R$. *(transitivity)*

We then write $x \sim y$ if $(x, y) \in R$.

Given a set X with an equivalence relation $\sim$, one can quotient X by $\sim$ by identifying together elements of X that are equivalent. Let us make this formal.

Definition B.1.4.

Suppose we have an equivalence relation $\sim$ on a set X and let $x \in X$. We define the **equivalence class** of x to be:

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

We write $X/_\sim := \{[x] \mid x \in X\}$ for the set of all equivalence classes, which we call the **quotient set** of X by $\sim$. The function $p\colon X \to X/_\sim$, given by $p(x) = [x]$, is called the **projection** of $\sim$.

Note that we always have $x \in [x]$.

Suppose $y \in [x]$. Then we know that $x \sim y$. It follows, using transitivity, that $[x] = [y]$. So we have found that

$$x \sim y \iff [x] = [y].$$

Also, if $y \notin [x]$, then $x \nsim y$, and we find, by transitivity again, that $[x] \cap [y] = \varnothing$, that is,

$$x \nsim y \iff [x] \cap [y] = \varnothing.$$

We thus just have proved:

Lemma B.1.5. *Two equivalence classes in the quotient set are either equal or disjoint.*

Let us see what are the equivalence classes of the examples we previously studied.

Example B.1.6. We consider Example B.1.2.

1. The equivalence class [Person A] contains all the people that Person A is related to and the quotient set is the set of all families.
2. Here we see that that the quotient set has just two elements: the set of all odd integers and the set of all even integers.

3. Here we can construct the rationals as the quotient set, that is, $\mathbf{Q} := Q/_\sim$. We commonly write $\frac{m}{n} = [(m,n)]$ for the equivalence classes. In particular, $\frac{1}{2} = \frac{-3}{-6}$.

4. The equivalence classes are the cosets, that is, $[g] = gH$.

B.2. Important Results

Suppose that $\sim$ is an equivalence relation on X. Then suppose that we are trying to construct a function with the quotient set of $\sim$ as the domain, something like, $f\colon X/_\sim \to Y$. We often need to be careful and check that the f is "well-defined". This means that, if $x_1 \sim x_2$, then $f([x_1]) = f([x_2])$. If this were not true, then f would not be a function.

Example B.2.1. For example, $f\colon Q/\sim \to \mathbf{Z}, [(n,m)] \mapsto n - m$ is not well-defined. Indeed, $[(2,1)] = [(6,3)]$ but $f([(2,1)]) = 2-1 = 1 \neq 3 = 6-3 = f([(6,3)])$ so f is not a function.

Equivalence relation (more precisely, their quotient set) are deeply related to the image of functions as demonstrated below.

Theorem B.2.2 (Bijection Theorem)**.**

Let $f\colon X \to Y$ be a function and define an equivalence relation on X by setting

$$x_1 \sim x_2 \iff f(x_1) = f(x_2).$$

Then the function

$$\begin{aligned}\tilde{f}\colon X/_\sim &\to f(X)\\ x &\mapsto \tilde{f}([x]) := f(x)\end{aligned}$$

is a bijection.

Proof. It is easy to see that $\sim$ is an equivalence relation on X. Next we need to check that $\tilde{f}$ is well-defined. Suppose that $x_1 \sim x_2$. Then, by definition, $f(x_1) = f(x_2)$ and so $\tilde{f}([x_1]) = f(x_1) = f(x_2) = \tilde{f}([x_2])$. Therefore, $\tilde{f}$ is well-defined.

Now we check that $\tilde{f}$ is injective. Suppose that $[x_1] \neq [x_2] \in X/_\sim$. Then $x_1 \nsim x_2$ and we have that $f(x_1) \neq f(x_2)$. Therefore, $\tilde{f}([x_1]) \neq \tilde{f}([x_2])$ and $\tilde{f}$ is injective.

Now we check that $\tilde{f}$ is surjective. Let $y \in f(X)$. Then there exists some $x \in X$ such that $y = f(x)$. Therefore, $\tilde{f}([x]) = f(x) = y$ and $\tilde{f}$ is surjective. We have now proved that $\tilde{f}$ is a bijection. ☐

The Bijection Theorem might remind you of the first isomorphism theorem for groups. This is not a coincidence. Indeed, if $f\colon G \to H$ is a group homomorphism, then $G/\sim = G/\ker(f)$. The details are left to the reader.

A **partition** on a set X is a collection of non-empty disjoint subsets of X whose union is the whole of X. For example, if $X = \{1, 2, 3\}$, then one partition could be $A_1 = \{1\}, A_2 = \{2\}, A_3 = \{3\}$, whilst another could be $B_1 = \{1, 2\}, B_2 = \{3\}$.

Partitions are simply another way to describe equivalence relations. More precisely:

Theorem B.2.3 (Fundamental Theorem of Equivalence Relations)**.**

An equivalence relation on a set X determines a partition on X. Conversely, any partition of X determines an equivalence relation on X.
These are bijective correspondences.

Proof. Suppose we have an equivalence relation $\sim$ on X. Then the equivalence classes are all non-empty and disjoint subsets of X and that every element of X belongs to an equivalence class. Therefore, the quotient set gives a partition of X.

Now suppose that $\{A_i \subseteq X \mid i \in I\}$ is a partition of X, for some indexing set I. We define an equivalence relation on X by setting $x \sim y$ if and only if x and y belong to the same subset A_i in the partition. It can clearly be seen that this is an equivalence relation.

We still need to show that these correspondences are bijective. Let $\sim$ be an equivalence relation on X and let $(A_i)_{i\in I}$ be the corresponding partition (so the A_i are the equivalence classes of $\sim$) given in the first part of the proof. Let $\sim'$ be the equivalence relation obtained from the A_i in the second part of the proof. It is straightforward that $\sim' = \sim$. If we first start with a partition $(A_i)_{i\in I}$, take the corresponding equivalence relation $\sim$ and then the partition $(B_j)_{j\in J}$ given by the equivalence classes of $\sim$, then there is a bijection $I \cong J$ and $(A_i)_{i\in I} = (B_i)_{i\in I}$ (up to reordering the B_i). $\square$

The Chapter's picture on page 104 illustrates the different partitions of a $\{1, 2, 3\}$ (there are 3 of them up to bijections), as well as the quotient sets they induce.

INDEX