Yet another book on topology I: point-set topology

Andrew Putman

Department of Mathematics, University of Notre Dame, 255 Hurley Hall, Notre Dame, IN 46556

Email address: andyp@nd.edu

Contents

Preface

This is the first volume of what I anticipate will be a four-volume textbook on topology:

- **Vol. 1**: Point-set topology
- **Vol. 2**: Covering spaces and the fundamental group
- **Vol. 3**: Homology
- **Vol. 4**: Cohomology and Poincaré duality

This preface describes the overall goals for this series of books. For a discussion of the mathematical choices I made in this particular volume and how they differ from other topology books, see the introduction for the instructor. That introduction also explains how this book can be used for different kinds of courses. There is also an introduction for the student that sets the stage for the various topics in this book.

Structure of books

Each volume of this series of books attempts to do three things:

(a) give a linear, well-structured treatment of the subject that is suitable for a first course; and
(b) explain a large number of advanced topics and examples; and
(c) describe (with complete proofs!) how the tools developed in that volume can be applied in other areas of mathematics.

These goals are not compatible with each other.

The solution I have adopted is to divide each volume into two parts. The first part ("Basic topics") represents what I believe should be covered in a first course on the subject. These chapters are intended to be read linearly. Even here there are options, and some chapters and sections are marked with an $*$ to indicate that a first-time reader might want to skip them.

The second part of each volume is a collection of essays on various topics: examples, applications, extensions, topological lore that does not fit neatly into a course, etc. These essays are less self-contained than the basic topics, though they include many proofs and detailed examples. They are not intended to be read linearly, though some refer to other essays for details.[1] They could be used for various kinds of advanced short courses or to enrich a course based on the first part of the volume.

Details and precision

Topology has an overwhelming amount of machinery. Many theorems have multiple versions with subtly different hypotheses and conclusions. As a writer, it is easy to get lost in technical details. The result is a book that is more of an encyclopedia for experts than a textbook.

However, there is also an opposite danger. Students find it frustrating if they have trouble finding precise definitions and statements of theorems. It can also be frustrating to have to flip between multiple sections to figure out what the author is trying to say.

I therefore tried to strike a balance between technical precision and narrative. For instance, I have not hesitated to repeat definitions for things that last appeared many pages ago. I also try to state theorems in their most natural level of generality, always with an eye to what is useful in applications.

I also try to include proofs that are detailed enough that a typical PhD student can read them without suffering unduly. Of course, a balance has to be struck here: too many details can make it

[1] I promise that there are no circular chains of references!

hard for a reader to figure out the main idea. However, I think that overall I include more details than many authors. This is partially responsible for the lengths of these books.

Pictures

I include many pictures, and freely use colors to make those pictures easier to understand. Typically these pictures are included in-line in the middle of a paragraph rather than in a separate stand-alone figure box. One technical issue here is that students with various forms of color blindness might have trouble interpreting the pictures. To help alleviate this, I used the palettes from the article [**3**], which I recommend to anyone who wants to use color but is worried about accessibility.

Audience

Volume 1 (point-set topology) is geared to advanced undergraduates, though it also includes many special topics that are hard to find in the literature and might be of interest to more advanced readers. The subsequent three volumes are aimed at beginning graduate students, roughly at the level of our first-year students at Notre Dame. The level of sophistication does rise as the books progress, largely driven by the increasing difficulty of the material.

Smooth manifolds

A lot of the early literature on topology focused on spaces that were triangulated. This is reflected in the textbook literature, which often uses combinatorial/piecewise-linear tools. However, outside of algebraic topology smooth manifolds are far more important. There are now many excellent textbooks on smooth manifolds, and I think it makes sense to replace simplicial/piecewise-linear techniques with smooth ones whenever possible.

The level of familiarity with smooth manifolds that I assume gradually increases through the books. Volume 1 (point-set topology) does not assume that students have seen manifolds of any kind before. Volume 2 (covering spaces and the fundamental group) mentions smooth manifolds, but only in remarks and essays. Subsequent volumes assume more, roughly at the level of the classic books by Milnor [**2**] or Guillemin–Pollack [**1**]. This is compatible with the sequence of courses taken by PhD students at most institutions I am familiar with.

Abstraction vs. concreteness

The following two quotations paraphrase things told to me by topologists whose work I admire:

(a) "I don't care about abstract nonsense or category theory at all. It's a distraction from what really matters. Fundamentally, I don't feel like I understand something until I've figured out a picture that illustrates what is going on."

(b) "I'm not interested in spaces for their own sake. What I care about are definitions and analogies. Category theory is, for me, the central framework for understanding mathematics as a unified whole."

I find myself in the unusual position of being sympathetic to both of these points of view. I love concrete geometry and abstract machinery, and think the best mathematics involves a mixture of both.

I do think that some students come to topology with an unhealthy attitude towards category theory. For some parts of topology and algebra, it has a marvelous unifying power. I cannot imagine ignoring category theory while teaching those subjects. However, it is not magic and needs to be combined with other tools to give interesting results. Students must learn to be eclectic and not expect any single tool to unify all of mathematics.

In keeping with the above, I try to use categorical language (functors, natural transformations, universal properties, etc) when appropriate, but not to over-emphasize it. Topology has many beautiful things to say about concrete geometric questions, and I hope even readers with a taste for abstraction will come to appreciate how all these tools can be used to prove results that seemingly have little to do with high-powered machinery.

Stylistic goals

I tried hard to obey the following stylistic rules:

(a) Each chapter is short (typically at most 8 pages, not including exercises), does exactly one thing, and is broken up into short sections with descriptive titles.
(b) I avoid overly long unbroken blocks of text. More concretely, I try to avoid having more than 6 lines of text appear without a paragraph break or math display.
(c) I give examples of each new concept, and try to illustrate major results with multiple examples.
(d) I include a large number of exercises in almost every chapter. These are an essential part of the text. Exercises marked with an $*$ are more difficult than the others, and the particularly difficult exercises are marked with $**$.

Acknowledgements

This is where I thank my many teachers, mentors, collaborators, friends, etc. It will be written later.

Bibliography

[1] V. W. Guillemin and A. Pollack, *Differential topology*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1974. (Cited on page 2.)
[2] J. W. Milnor, *Topology from the differentiable viewpoint*, revised reprint of the 1965 original, Princeton Landmarks in Mathematics, Princeton Univ. Press, Princeton, NJ, 1997. (Cited on page 2.)
[3] B. Wong, Points of view: Color blindness, Nat. Methods 8 (2011), 441. (Cited on page 2.)

Introduction for the instructor

Introduction for the student

Part 1

Basic topics

Metric spaces

CHAPTER 1

Review of the topology of subsets of Euclidean space

We start with a review of some basic facts about Euclidean space.

1.1. Preliminary discussion

The most familiar spaces are $\mathbb{R}^d$ and its subspaces. Indeed, we live in a space that looks like $\mathbb{R}^3$ and our drawings typically depict objects in $\mathbb{R}^d$ for $d \leq 3$. For instance:

We can imagine subspaces of $\mathbb{R}^d$ for $d \geq 4$ by analogy with $\mathbb{R}^3$. Subspaces of Euclidean space are the geometric objects studied by mathematicians going back to the ancient Greeks. Calculus and linear algebra were developed in part to provide tools for studying these kinds of spaces.

In this book, we will assume that our readers are familiar with the basic topology of subspaces of Euclidean space, roughly at the level of an introductory course in real analysis. This will allow us to quickly introduce interesting examples of the more abstract geometric objects we define. This chapter is a review of this material, mostly without proofs. We do, however, include proofs of three important results: the Bolzano–Weierstrass theorem, the Heine–Borel theorem, and the extreme value theorem.

Readers may want to skip this chapter if they are comfortable with the following topics:

- Continuous functions on $\mathbb{R}^d$.
- The basic facts about sequences and their limits in $\mathbb{R}^d$, including Cauchy sequences.
- Open and closed sets in $\mathbb{R}^d$, including related concepts like limit points.
- Sequential compactness[1] of subsets of $\mathbb{R}^d$, including the Heine–Borel theorem saying that a subset of $\mathbb{R}^d$ is sequentially compact if and only if it is closed and bounded.
- The extreme value theorem for functions on $\mathbb{R}^d$.

Remark 1.1.1. In Chapter 2, we will abstract some of the features of Euclidean space to define metric spaces. Many of the unproved assertions of this chapter have generalizations to metric spaces, and we will prove them in that broader context in Chapter 2. □

1.2. Euclidean distance

Consider a point $\mathbf{x} \in \mathbb{R}^d$. Its norm $\|\mathbf{x}\|$ is the distance from $\mathbf{0}$ to $\mathbf{x}$. In coordinates, if $\mathbf{x} = (x_1, \ldots, x_d)$ then

$$\|\mathbf{x}\| = \sqrt{x_1^2 + \cdots + x_d^2}.$$

For points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, the distance from $\mathbf{x}$ to $\mathbf{y}$ is $\|\mathbf{x} - \mathbf{y}\|$.

[1] A set $K \subset \mathbb{R}^d$ is sequentially compact if every sequence of points of K has a subsequence converging to a point of K. This is equivalent to requiring that every open cover of K has a finite subcover, though for pedagogical reasons we chose to postpone that more sophisticated notion of compactness until we can introduce it for general topological spaces.

1.3. Continuous functions

Let $X \subset \mathbb{R}^d$ and $Y \subset \mathbb{R}^e$ be subsets and let $f\colon X \to Y$ be a function. For $\mathbf{x}_0 \in X$, we say that f is *continuous at* $\mathbf{x}_0$ if it satisfies the classical ϵ-δ definition of continuity:

- For all $\epsilon > 0$, there exists some $\delta > 0$ such that if $\mathbf{x} \in X$ satisfies $\|\mathbf{x} - \mathbf{x}_0\| < \delta$, then $\|f(\mathbf{x}) - f(\mathbf{x}_0)\| < \epsilon$.

We say that f is *continuous* if it is continuous at all points $\mathbf{x}_0 \in X$. Continuous functions satisfy a variety of expected properties. For instance:

LEMMA 1.3.1. *Let $X \subset \mathbb{R}^d$ and $Y \subset \mathbb{R}^e$ and $Z \subset \mathbb{R}^k$ be subsets and let $f\colon X \to Y$ and $g\colon Y \to Z$ be continuous functions. Then $g \circ f\colon X \to Z$ is continuous.*

LEMMA 1.3.2. *Let $X \subset \mathbb{R}^d$ be a subset and let $f, g\colon X \to \mathbb{R}$ be continuous. Then their sum $f + g$ and product fg are continuous functions from X to $\mathbb{R}$.*

We can write down continuous functions using formulas. Here are two examples:

EXAMPLE 1.3.3. Let $X \subset \mathbb{R}^3$ be arbitrary. We can define a continuous function $f\colon X \to \mathbb{R}$ via the formula $f(x_1, x_2, x_3) = x_1 - 2e^{x_2 - 7x_3}$ for $(x_1, x_2, x_3) \in X$. □

EXAMPLE 1.3.4. For $d \geq 0$, recall that the d-sphere is

$$\mathbb{S}^d = \left\{(x_1, \ldots, x_{d+1}) \in \mathbb{R}^{d+1} \mid x_1^2 + \cdots + x_{d+1}^2 = 1\right\}.$$

We can define a continuous function $f\colon \mathbb{S}^d \times \mathbb{R} \to \mathbb{S}^{d+1}$ via the formula

$$f(x_1, \ldots, x_{d+1}, y) = (\cos(y)x_1, \ldots, \cos(y)x_{d+1},\ \sin(y)) \in \mathbb{S}^{d+1} \quad \text{for all } (x_1, \ldots, x_{d+1}, y) \in \mathbb{S}^d \times \mathbb{R}.$$

Here we are implicitly viewing $\mathbb{S}^d \times \mathbb{R}$ as a subset of $\mathbb{R}^{d+2}$. □

1.4. Sequences

Consider a sequence $(\mathbf{x}_n)_{n\geq 1}$ of points of $\mathbb{R}^d$. We say that the sequence $(\mathbf{x}_n)_{n\geq 1}$ *converges* to $\mathbf{x} \in \mathbb{R}^d$ if for all $\epsilon > 0$, there exists some $N \geq 1$ such that $\|\mathbf{x}_n - \mathbf{x}\| < \epsilon$ for all $n \geq N$. In this case, we write $\lim_{n\to\infty} \mathbf{x}_n = \mathbf{x}$ and say that $(\mathbf{x}_n)_{n\geq 1}$ is a *convergent sequence*. The value $\mathbf{x}$ is also called the *limit* of the sequence. If $(\mathbf{x}_n)_{n\geq 1}$ is a convergent sequence, then its limit is unique. We can characterize continuity using limits as follows:

LEMMA 1.4.1. *Let $X \subset \mathbb{R}^d$ and $Y \subset \mathbb{R}^e$ be subsets and let $f\colon X \to Y$ be a function. Then f is continuous if and only if for all convergent sequences $(\mathbf{x}_n)_{n\geq 1}$ of points of X whose limit lies in X, the sequence $(f(\mathbf{x}_n))_{n\geq 1}$ of points of Y is also convergent and we have*

$$\lim_{n\to\infty} f(\mathbf{x}_n) = f(\lim_{n\to\infty} \mathbf{x}_n).$$

A sequence $(\mathbf{x}_n)_{n\geq 1}$ of points of $\mathbb{R}^d$ is a *Cauchy sequence* if for all $\epsilon > 0$, there exists some $N \geq 1$ such that for all $n, m \geq N$ we have $\|\mathbf{x}_n - \mathbf{x}_m\| < \epsilon$. All convergent sequences of points of $\mathbb{R}^d$ are Cauchy sequences. The following converse to this is one of the cornerstones of real analysis:

THEOREM 1.4.2. *Let $(\mathbf{x}_n)_{n\geq 1}$ be a Cauchy sequence of points of $\mathbb{R}^d$. Then $(\mathbf{x}_n)_{n\geq 1}$ converges to a limit $\mathbf{x} \in \mathbb{R}^d$.*

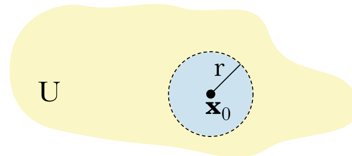
REMARK 1.4.3. This expresses the fact that $\mathbb{R}^d$ is *complete*. The fact that $\mathbb{R} = \mathbb{R}^1$ is complete is one of the most important features of $\mathbb{R}$. Recognizing this was a key insight in the development of the modern foundations for analysis in the 19th century. □

1.5. Open sets

For $\mathbf{x}_0 \in \mathbb{R}^d$ and $r > 0$, define the open ball around $\mathbf{x}_0$ of radius r as follows:

$$B_r(\mathbf{x}_0) = \left\{\mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x} - \mathbf{x}_0\| < r\right\}.$$

A subset $U \subset \mathbb{R}^d$ is *open* if for all $\mathbf{x}_0 \in U$, there exists some $r > 0$ such that $B_r(\mathbf{x}_0) \subset U$:



EXAMPLE 1.5.1. For all $\mathbf{x} \in \mathbb{R}^d$ and $r > 0$, the set $U = B_r(\mathbf{x})$ is open. □

More generally, consider $X \subset \mathbb{R}^d$. A set $U \subset X$ is *open in* X if for all $\mathbf{x}_0 \in U$, there exists some $r > 0$ such that $B_r(\mathbf{x}_0) \cap X \subset U$. Equivalently, U is open in X if $U = X \cap V$ for some open subset $V \subset \mathbb{R}^d$. In particular, a set $U \subset \mathbb{R}^d$ is open if and only if U is open in $\mathbb{R}^d$.

EXAMPLE 1.5.2. Let $X = [0, 1] \subset \mathbb{R}$. The half-open interval $[0, 1)$ is open in X, but is not open as a subset of $\mathbb{R}$. □

EXAMPLE 1.5.3. If $X \subset \mathbb{R}^d$ is open, then a set $U \subset X$ is open in X if and only if U is open as a subset of $\mathbb{R}^d$. □

EXAMPLE 1.5.4. Let $X \subset \mathbb{R}^d$ be a subset and let $f \colon X \to \mathbb{R}$ be a continuous function. Then for $a, b \in \mathbb{R}$ with $a < b$ the sets

$$\begin{aligned} U &= \{\mathbf{x} \in X \mid a < f(\mathbf{x}) < b\}, \\ U' &= \{\mathbf{x} \in X \mid f(\mathbf{x}) < b\}, \\ U'' &= \{\mathbf{x} \in X \mid a < f(\mathbf{x})\} \end{aligned}$$

are all open in X. □

REMARK 1.5.5. Example 1.5.1 is a special case of Example 1.5.4. Indeed, for $\mathbf{x} \in \mathbb{R}^d$ we can define a continuous function $f_{\mathbf{x}} \colon \mathbb{R}^d \to \mathbb{R}$ as follows:

$$f_{\mathbf{x}}(\mathbf{y}) = \|\mathbf{y} - \mathbf{x}\| \quad \text{for } \mathbf{y} \in \mathbb{R}^d.$$

For $r > 0$, we then have $B_r(\mathbf{x}) = \left\{\mathbf{y} \in \mathbb{R}^d \mid f_{\mathbf{x}}(\mathbf{y}) < r\right\}$. □

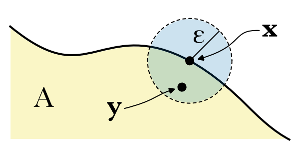
Open sets are closed under arbitrary unions and finite intersections:

LEMMA 1.5.6. *Let $X \subset \mathbb{R}^d$ be a subset. The following hold:*

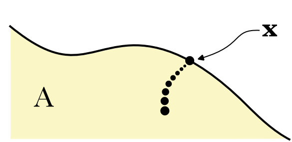
- *The empty set $\emptyset$ and the whole set X are open in X.*
- *If $\{U_i\}_{i \in I}$ is an arbitrary collection of sets that are open in X, then $\cup_{i \in I} U_i$ is open in X.*
- *If $\{U_1, \ldots, U_n\}$ is a finite collection of sets that are open in X, then $\cap_{i=1}^n U_i$ is open in X.*

1.6. Limit points and closed sets

Consider $A \subset \mathbb{R}^d$. A point $\mathbf{x} \in \mathbb{R}^d$ is a *limit point* of A if for all $\epsilon > 0$, there exists some $\mathbf{y} \in A$ with $\mathbf{y} \neq \mathbf{x}$ such that $\|\mathbf{y} - \mathbf{x}\| < \epsilon$:



Equivalently, there exists a sequence $(\mathbf{y}_n)_{n \geq 1}$ of points of A with $\mathbf{y}_n \neq \mathbf{x}$ for all $n \geq 1$ such that $\lim_{n \to \infty} \mathbf{y}_n = \mathbf{x}$:



A set $C \subset \mathbb{R}^d$ is *closed* if all limit points of C are contained in C. Equivalently, for all convergent sequences $(\mathbf{x}_n)_{n \geq 1}$ of points of C, we have $\lim_{n \to \infty} \mathbf{x}_n \in C$.

EXAMPLE 1.6.1. For $\mathbf{x} \in \mathbb{R}^d$ and $r \geq 0$, define the *closed ball* of radius r around $\mathbf{x}$ to be[2]

$$B_r^{c\ell}(\mathbf{x}) = \left\{ \mathbf{y} \in \mathbb{R}^d \mid \|\mathbf{y} - \mathbf{x}\| \leq r \right\}.$$

For all $\mathbf{x} \in \mathbb{R}^d$ and $r \geq 0$, the set $C = B_r^{c\ell}(\mathbf{x})$ is closed. □

More generally, consider $X \subset \mathbb{R}^d$. A set $C \subset X$ is *closed in* X if for all limit points c of C with $c \in X$, we have $c \in C$. Equivalently, C is closed in X if $C = X \cap D$ for some closed subset $D \subset \mathbb{R}^d$. In particular, a set $D \subset \mathbb{R}^d$ is closed if and only if D is closed in $\mathbb{R}^d$.

EXAMPLE 1.6.2. Let $X = (0, 1) \subset \mathbb{R}$. The half-open interval $(0, 1/2]$ is closed in X, but is not closed as a subset of $\mathbb{R}$. □

EXAMPLE 1.6.3. If $X \subset \mathbb{R}^d$ is closed, then a set $C \subset X$ is closed in X if and only if C is closed as a subset of $\mathbb{R}^d$. □

EXAMPLE 1.6.4. Let $X \subset \mathbb{R}^d$ be a subset and let $f\colon X \to \mathbb{R}$ be a continuous function. Then for $a, b \in \mathbb{R}$ with $a < b$ the sets

$$\begin{aligned} C &= \{\mathbf{x} \in X \mid a \leq f(\mathbf{x}) \leq b\}, \\ C' &= \{\mathbf{x} \in X \mid f(\mathbf{x}) \leq b\}, \\ C'' &= \{\mathbf{x} \in X \mid a \leq f(\mathbf{x})\} \end{aligned}$$

are all closed in X. In particular, for all $\mathbf{y} \in \mathbb{R}$ the set $f^{-1}(\mathbf{y}) = \{\mathbf{x} \in X \mid f(\mathbf{x}) = \mathbf{y}\}$ is closed in X. □

Closed sets are closed under arbitrary intersections and finite unions:

LEMMA 1.6.5. *Let $X \subset \mathbb{R}^d$ be a subset. The following hold:*

- *The empty set $\emptyset$ and the whole set X are closed in X.*
- *If $\{C_i\}_{i \in I}$ is an arbitrary collection of sets that are closed in X, then $\cap_{i \in I} C_i$ is closed in X.*
- *If $\{C_1, \ldots, C_n\}$ is a finite collection of sets that are closed in X, then $\cup_{i=1}^n C_i$ is closed in X.*

1.7. Relating open and closed sets

Open and closed sets are related as follows:

LEMMA 1.7.1. *Let $X \subset \mathbb{R}^d$ be a subset. A set $U \subset X$ is open in X if and only if $X \setminus U$ is closed in X.*

This might lead an inexperienced reader to think that openness and closedness are opposite conditions. However, this is not true even for open/closed subsets of $\mathbb{R}^d$ itself:

- A typical set is neither open nor closed. For instance, the half-open interval $[0, 1)$ in $\mathbb{R}$ is neither open nor closed.
- The empty set $\emptyset$ and the whole space $\mathbb{R}^d$ are both open and closed. More generally, for $X \subset \mathbb{R}^d$ the empty set $\emptyset$ and the whole set X are both open and closed in X.

1.8. Interior, closure, and boundary of a set

Let $Z \subset \mathbb{R}^d$ be a subset. Associated to Z are the following:

- The *closure* of Z, denoted $\overline{Z}$, is the union of Z and all of its limit points. The set $\overline{Z}$ is closed, and is the smallest closed set containing Z in the sense that if $C \subset \mathbb{R}^d$ is closed and $Z \subset C$, then $\overline{Z} \subset C$.
- The *interior* of Z, denoted $\mathrm{Int}(Z)$, is the set of all points $\mathbf{x}_0 \in Z$ such that there exists some $r > 0$ with $B_r(\mathbf{x}_0) \subset Z$. The set $\mathrm{Int}(Z)$ is open, and is the largest open set contained in Z in the sense that if $U \subset Z$ is open then $U \subset \mathrm{Int}(Z)$.

[2] It is also common to write this as $\overline{B}_r(\mathbf{x})$, but this notation could potentially be confusing. Indeed, later we will define the closure $\overline{A}$ of a set A, and the notation $\overline{B}_r(\mathbf{x})$ suggests that $\overline{B}_r(\mathbf{x})$ is the closure of $B_r(\mathbf{x})$. For $r > 0$ we have $B_r^{c\ell}(\mathbf{x}) = \overline{B_r(\mathbf{x})}$ for $\mathbf{x}$ a point in Euclidean space, but this is false in more general settings.

- We have $\mathrm{Int}(Z) \subset Z \subset \overline{Z}$. The *boundary* of Z, denoted ∂Z, is $\partial Z = \overline{Z} \setminus \mathrm{Int}(Z)$. The set ∂Z is closed. An alternate way of describing ∂Z is that it is the set consisting of all $\mathbf{z} \in \mathbb{R}^d$ such that for all $r > 0$ the ball $B_r(\mathbf{z})$ contains points of both Z and $\mathbb{R}^d \setminus Z$.

EXAMPLE 1.8.1. Let $Z = [0, 1) \subset \mathbb{R}$. We have $\overline{Z} = [0, 1]$ and $\mathrm{Int}(Z) = (0, 1)$ and $\partial Z = \{0, 1\}$. $\square$

REMARK 1.8.2. More generally, if $Z \subset X \subset \mathbb{R}^d$, then we can talk about the closure, interior, and boundary of Z in X. To avoid multiplying technicalities we will not discuss this. $\square$

1.9. The Bolzano–Weierstrass theorem

Consider a sequence $(\mathbf{x}_n)_{n\geq 1}$ of points of $\mathbb{R}^d$. We introduce the following two definitions:

- The sequence $(\mathbf{x}_n)_{n\geq 1}$ is *bounded* if there exists some $r \geq 0$ such that $\|\mathbf{x}_n\| \leq r$ for all $n \geq 1$.
- A *subsequence* of $(\mathbf{x}_n)_{n\geq 1}$ is a sequence $(\mathbf{y}_m)_{m\geq 1}$ such that there exists an increasing sequence of positive integers

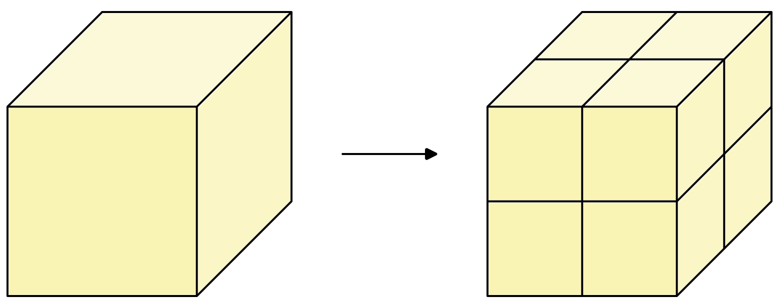
$$1 \leq n_1 < n_2 < \cdots$$

such that $\mathbf{y}_m = \mathbf{x}_{n_m}$ for all $m \geq 1$. For instance, $(\mathbf{y}_m)_{m\geq 1}$ might be the sequence $(\mathbf{x}_{2m})_{m\geq 1}$.

With these definitions, we have the following fundamental result:

THEOREM 1.9.1 (Bolzano–Weierstrass theorem). *Let $(\mathbf{x}_n)_{n\geq 1}$ be a bounded sequence of points of $\mathbb{R}^d$. Then $(\mathbf{x}_n)_{n\geq 1}$ contains a convergent subsequence.*

PROOF. Since $(\mathbf{x}_n)_{n\geq 1}$ is bounded, we can find some sufficiently large $\lambda > 0$ such that each $\mathbf{x}_n$ is contained in the cube $C_1 = [-\lambda/2, \lambda/2]^d$. Set $n_1 = 1$, so $\mathbf{x}_{n_1} \in C_1$. Note that each side of C_1 has length λ. As in the following figure, divide C_1 into 2^d subcubes with side length $\lambda/2$:



At least one of these subcubes must contain infinitely many terms of the sequence $(\mathbf{x}_n)_{n\geq 1}$. Choose such a subcube and call it C_2. Let n_2 be such that $\mathbf{x}_{n_2} \in C_2$ and $n_2 > n_1$. This process can then be repeated: C_2 can be divided into 2^d subcubes with side length $\lambda/2^2$, and among these there must exist a subcube C_3 containing infinitely many terms of $(\mathbf{x}_n)_{n\geq 1}$. Let n_3 be such that $\mathbf{x}_{n_3} \in C_3$ and $n_3 > n_2$. We then divide C_3 into 2^d subcubes with side length $\lambda/2^3$, etc.

This procedure gives a nested sequence

$$C_1 \supset C_2 \supset C_3 \supset \cdots$$

of cubes and an increasing sequence of integers

$$n_1 < n_2 < n_3 < \cdots$$

such that the following holds for all $k \geq 1$:

- the cube C_k has side length $\lambda/2^{k-1}$; and
- $\mathbf{x}_{n_k} \in C_k$. In fact, since the C_k are nested we have $\mathbf{x}_{n_{k'}} \in C_k$ for all $k' \geq k$.

Since the diameters of the C_k are shrinking to 0 and $\mathbf{x}_{n_{k'}} \in C_k$ for all $k' \geq k$, it follows that the subsequence

$$\mathbf{x}_{n_1}, \mathbf{x}_{n_2}, \mathbf{x}_{n_3}, \ldots$$

of $(\mathbf{x}_n)_{n\geq 1}$ is a Cauchy sequence. In particular, it converges (see Theorem 1.4.2), as desired. $\square$

1.10. Sequentially compact sets and the Heine–Borel theorem

The following is perhaps the most important definition in this chapter:

DEFINITION 1.10.1. A set $K \subset \mathbb{R}^d$ is *sequentially compact* if every sequence $(\mathbf{x}_n)_{n\geq 1}$ of points of K has a subsequence that converges to a point of K. □

EXAMPLE 1.10.2. A finite set $K \subset \mathbb{R}^d$ is sequentially compact. □

EXAMPLE 1.10.3. For $d \geq 1$, the whole space $\mathbb{R}^d$ is not sequentially compact. Indeed, for $n \geq 1$ set $\mathbf{x}_n = (n, 0, \ldots, 0) \in \mathbb{R}^d$. Every subsequence of $(\mathbf{x}_n)_{n\geq 1}$ diverges to infinity,[3] so $(\mathbf{x}_n)_{n\geq 1}$ has no convergent subsequence. □

REMARK 1.10.4. We will later define what it means for a general topological space to be compact. The definition involves "open covers", and is rather different from the definition of sequential compactness we gave above. For sufficiently nice spaces (including subsets of $\mathbb{R}^d$, and more generally the metric spaces we will define in the next chapter), compactness is equivalent to sequential compactness. In complete generality, however, neither implies the other. □

It is not easy to directly verify that a set is sequentially compact using the above definition; however, there is a simple characterization of sequentially compact sets. A subset $X \subset \mathbb{R}^d$ is *bounded* if there exists some $r \geq 0$ such that $\|\mathbf{x}\| \leq r$ for all $\mathbf{x} \in X$. We then have the following:

THEOREM 1.10.5 (Heine–Borel theorem). *A set $K \subset \mathbb{R}^d$ is sequentially compact if and only if it is closed and bounded.*

PROOF. We divide the proof into two parts.

CLAIM. *Assume that K is closed and bounded. Then K is sequentially compact.*

Let $(\mathbf{x}_n)_{n\geq 1}$ be a sequence of points of K. We must prove that $(\mathbf{x}_n)_{n\geq 1}$ has a subsequence that converges to a point of K. Since K is bounded, the Bolzano–Weierstrass theorem (Theorem 1.9.1) implies that $(\mathbf{x}_n)_{n\geq 1}$ has a convergent subsequence $(\mathbf{y}_m)_{m\geq 1}$. Since K is closed, we have $\lim_{m\to\infty} \mathbf{y}_m \in K$. The claim follows.

CLAIM. *Assume that $K \subset \mathbb{R}^d$ is sequentially compact. Then K is closed and bounded.*

To see that K is closed, consider a convergent sequence $(\mathbf{x}_n)_{n\geq 1}$ of points of K. We must prove that $\lim_{n\to\infty} \mathbf{x}_n \in K$. Since K is sequentially compact, there is a subsequence $(\mathbf{y}_m)_{m\geq 1}$ of $(\mathbf{x}_n)_{n\geq 1}$ that converges to a point of K. Passing to a subsequence does not change the limit of a sequence (see Exercise 1.3), so

$$\lim_{n\to\infty} \mathbf{x}_n = \lim_{m\to\infty} \mathbf{y}_m \in K,$$

as desired.

To see that K is bounded, assume for the sake of contradiction that it is not bounded. We can then find a sequence $(\mathbf{z}_n)_{n\geq 1}$ of points of K such that $\|\mathbf{z}_n\| \geq n$ for all $n \geq 1$. Since K is sequentially compact, there is a subsequence $(\mathbf{w}_m)_{m\geq 1}$ of $(\mathbf{z}_n)_{n\geq 1}$ that converges to a point of K. However, since $\|\mathbf{z}_n\| \geq n$ for all $n \geq 1$ it follows that $\lim_{m\to\infty} \|\mathbf{w}_m\| = +\infty$, contradicting the fact that $(\mathbf{w}_m)_{m\geq 1}$ is a convergent sequence. We conclude that K is in fact bounded. □

1.11. Continuous functions on sequentially compact sets

Let $X \subset \mathbb{R}^d$ be a nonempty set and $f \colon X \to \mathbb{R}$ be a function. Recall that f is *bounded* if there exist $a, b \in \mathbb{R}$ with $a \leq b$ such that $a \leq f(\mathbf{x}) \leq b$ for all $\mathbf{x} \in X$. Assume that f is bounded. The *infimum* of f, denoted $\inf(f)$, is the infimum of the image

$$f(X) = \{f(\mathbf{x}) \mid \mathbf{x} \in X\} \subset \mathbb{R}$$

of f, i.e., its greatest lower bound. We say that the infimum of f is *attained* if there exists some $\mathbf{x} \in X$ such that $f(x) = \inf(f)$, in which case we write $\min(f) = \inf(f)$. Similarly, the *supremum* of f, denoted $\sup(f)$, is the supremum of $f(X)$, i.e., its least upper bound. We say that the

[3] Recall that a sequence $(\mathbf{y}_m)_{m\geq 1}$ is said to diverge to infinity if for all $C \geq 0$, there exists some $M \geq 1$ such that $\|\mathbf{y}_m\| \geq C$ for all $m \geq M$.

supremum of f is *attained* if there exists some $\mathbf{x} \in X$ such that $f(\mathbf{x}) = \sup(f)$, in which case we write $\max(f) = \sup(f)$.

Neither the infimum nor the supremum of a bounded continuous function need to be attained:

EXAMPLE 1.11.1. Let $f\colon (0,1) \to \mathbb{R}$ be the function $f(t) = t$ for all $t \in (0,1)$. This is a bounded continuous function with $\inf(f) = 0$ and $\sup(f) = 1$. Neither the infimum nor the supremum is attained. □

The following is a key feature of continuous functions on sequentially compact sets:

THEOREM 1.11.2 (Extreme value theorem). *Let $K \subset \mathbb{R}^d$ be a nonempty sequentially compact set and let $f\colon K \to \mathbb{R}$ be a continuous function. Then:*

(i) The function f is bounded.
(ii) The infimum and supremum of f are attained.

PROOF. We prove the two parts separately.

CLAIM (i). *The function f is bounded.*

Assume for the sake of contradiction that there does not exist $b \in \mathbb{R}$ such that $f(\mathbf{x}) \leq b$ for all $\mathbf{x} \in K$. This implies that there exists a sequence $(\mathbf{x}_n)_{n\geq 1}$ of points of K such that $f(\mathbf{x}_n) \geq n$ for all $n \geq 1$. Since K is sequentially compact, there is a convergent subsequence $(\mathbf{y}_m)_{m\geq 1}$ of $(\mathbf{x}_n)_{n\geq 1}$. Set $\mathbf{y} = \lim_{m\to\infty} \mathbf{y}_m \in K$. Since f is continuous, we have

$$\lim_{m\to\infty} f(\mathbf{y}_m) = f(\lim_{m\to\infty} \mathbf{y}_m) = f(\mathbf{y}). \tag{1.11.1}$$

However, since $(\mathbf{y}_m)_{m\geq 1}$ is a subsequence of $(\mathbf{x}_n)_{n\geq 1}$ and $f(\mathbf{x}_n) \geq n$ for all $n \geq 1$, the terms $f(\mathbf{y}_m)$ in the limit on the left-hand side of (1.11.1) tend to $+\infty$, a contradiction. This implies that there does exist some $b \in \mathbb{R}$ such that $f(\mathbf{x}) \leq b$ for all $\mathbf{x} \in K$. A similar argument[4] shows that there exists some $a \in \mathbb{R}$ such that $a \leq f(\mathbf{x})$ for all $\mathbf{x} \in K$. We conclude that f is bounded, as desired.

CLAIM (ii). *The infimum and supremum of f are attained.*

We will prove that the supremum of f is attained. The proof that the infimum of f is attained is similar. By definition, we can find a sequence $(\mathbf{x}_n)_{n\geq 1}$ of points of K such that $\lim_{n\to\infty} f(\mathbf{x}_n) = \sup(f)$. Since K is sequentially compact, we can replace $(\mathbf{x}_n)_{n\geq 1}$ by a subsequence and assume that it is convergent. Letting $\mathbf{x} = \lim_{n\to\infty} \mathbf{x}_n \in K$, since f is continuous we have

$$f(\mathbf{x}) = f(\lim_{n\to\infty} \mathbf{x}_n) = \lim_{n\to\infty} f(\mathbf{x}_n) = \sup(f),$$

as desired. □

This result has the following useful consequence:

LEMMA 1.11.3. *Let $K \subset \mathbb{R}$ be a nonempty sequentially compact set. There then exist $a, b \in K$ with $a \leq b$ such that $a = \inf(K)$ and $b = \sup(K)$.*

PROOF. Apply Theorem 1.11.2 to the continuous inclusion map $\iota\colon K \to \mathbb{R}$. □

1.12. Exercises

EXERCISE 1.1. Compute the limit points, closure, interior, and boundary of the following:

(a) The open unit ball in $\mathbb{R}^d$ for $d \geq 1$.
(b) The closed unit ball in $\mathbb{R}^d$ for $d \geq 1$.
(c) The set $Z = \{1/n \mid n \geq 1\}$ in $\mathbb{R}$.
(d) The set $W = \{0\} \cup \{1/n \mid n \geq 1\}$ in $\mathbb{R}$.
(e) The subspace $\mathbb{Q}$ of $\mathbb{R}$. □

EXERCISE 1.2. Do the following:

(a) For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, prove the triangle inequality $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$. Hint: use the Cauchy–Schwarz inequality.

[4] Alternatively, one could just apply the above argument to $-f$.

(b) For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, prove the reverse triangle inequality $|\|\mathbf{x}\| - \|\mathbf{y}\|| \leq \|x - y\|$.
(c) Fix some $\mathbf{x}_0 \in \mathbb{R}^d$. Define $f\colon \mathbb{R}^d \to \mathbb{R}$ via the formula

$$f(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}_0\| \quad \text{for all } \mathbf{x} \in \mathbb{R}^d.$$

Prove that f is continuous.
(d) Use (c) to prove that open balls are open. Hint: see Example 1.5.4. □

Exercise 1.3. Let $(\mathbf{x}_n)_{n\geq 1}$ be a sequence in $\mathbb{R}^d$ that converges to a point $\mathbf{x} \in \mathbb{R}^d$. Do the following:
(a) Prove that $(\mathbf{x}_n)_{n\geq 1}$ is Cauchy.
(b) Prove that $(\mathbf{x}_n)_{n\geq 1}$ is bounded.
(c) Prove that every subsequence of $(\mathbf{x}_n)_{n\geq 1}$ also converges to $\mathbf{x}$. □

Exercise 1.4. Do the following:
(a) Find open sets $U_1, U_2, U_3, \ldots$ in $\mathbb{R}$ such that $\cap_{n=1}^{\infty} U_n$ is not open.
(b) Find closed sets $C_1, C_2, C_3, \ldots$ in $\mathbb{R}$ such that $\cup_{n=1}^{\infty} C_n$ is not closed. □

Exercise 1.5. Let $(\mathbf{x}_n)_{n\geq 1}$ be a sequence in $\mathbb{R}^d$. For $n \geq 1$, write $\mathbf{x}_n = (x(n)_1, \ldots, x(n)_d)$ with $x(n)_1, \ldots, x(n)_d \in \mathbb{R}$. Do the following:
(a) Prove that $(\mathbf{x}_n)_{n\geq 1}$ is a Cauchy sequence if and only if for each $1 \leq j \leq d$ the sequence $(x(n)_j)_{n\geq 1}$ of real numbers is a Cauchy sequence.
(b) For some $\mathbf{x} = (x_1, \ldots, x_d)$, prove that $\lim_{n\to\infty} \mathbf{x}_n = \mathbf{x}$ if and only if $\lim_{n\to\infty} x(n)_j = x_j$ for all $1 \leq j \leq d$. □

Exercise 1.6. Do the following:
(a) Prove Lemma 1.3.1 directly from the ϵ-δ definition of continuity. Recall that this lemma says the following. Let $X \subset \mathbb{R}^d$ and $Y \subset \mathbb{R}^e$ and $Z \subset \mathbb{R}^k$ be subsets and let $f\colon X \to Y$ and $g\colon Y \to Z$ be continuous functions. Then $g \circ f\colon X \to Z$ is continuous.
(b) Prove Lemma 1.3.2 directly from the ϵ-δ definition of continuity. Recall that this lemma says the following. Let $X \subset \mathbb{R}^d$ be a subset and let $f, g\colon X \to \mathbb{R}$ be continuous. Then their sum $f + g$ and product fg are continuous functions from X to $\mathbb{R}$. □

Exercise 1.7. Let $Z \subset \mathbb{R}^d$. Do the following:
(a) Prove that $\mathrm{Int}(Z) = \mathbb{R}^d \setminus \overline{\mathbb{R}^d \setminus Z}$.
(b) Prove that $\partial Z = \overline{Z} \cap \overline{\mathbb{R}^d \setminus Z}$.
(c) Prove that $\overline{Z} = Z \cup \partial Z$.
(d) Prove that Z is open if and only if $Z \cap \partial Z = \emptyset$.
(e) Prove that Z is closed if and only if $\partial Z \subset Z$. □

Exercise 1.8. Let $K \subset \mathbb{R}^d$ be sequentially compact and let $Z \subset K$ be an infinite set. Prove that Z has at least one limit point in $\mathbb{R}^d$ and that each limit point of Z lies in K. □

Exercise 1.9*. Let $C \subset \mathbb{R}^d$ be a nonempty closed subset and let $\mathbf{x} \in \mathbb{R}^d$. Do the following:
(a) Prove that there exists some $\mathbf{c}_0 \in C$ such that $\|\mathbf{x} - \mathbf{c}_0\| = \inf\{\|\mathbf{x} - \mathbf{c}\| \mid \mathbf{c} \in C\}$.
(b) Let $K \subset \mathbb{R}^d$ be nonempty and sequentially compact. Assume that $K \cap C = \emptyset$. Prove that $\inf\{\|\mathbf{k} - \mathbf{c}\| \mid \mathbf{k} \in K, \mathbf{c} \in C\} > 0$.
(c) Give a counterexample to show that (b) can fail if K is assumed to be merely closed instead of sequentially compact. □

CHAPTER 2

Metric spaces: definition and basic properties

This chapter introduces metric spaces and discusses their basic properties.

2.1. Metric spaces

As we will see throughout this book, there are geometric entities that share features with classical spaces like the subsets of Euclidean space discussed in Chapter 1. Modern formalizations of the notion of "space" give a precise language for extending our geometric imagination to these less easily visualizable kinds of spaces.

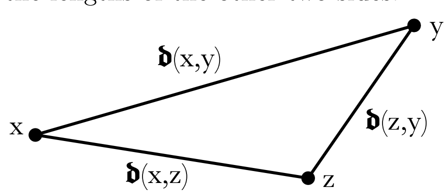
What features should we look for in an abstract notion of a space? One basic feature of spaces is that they have "points" of some kind, and some of those points are "closer" to one another than other points. One way to formalize this is to abstract the idea of distance. For points $x, y \in \mathbb{R}^d$, the distance from x to y is $\|x - y\|$. In a metric space, the distance $\|x - y\|$ is replaced by a number $\mathfrak{d}(x, y)$ satisfying three axioms.

Definition 2.1.1. A *metric space* is a pair $(M, \mathfrak{d})$, where M is a set and $\mathfrak{d}\colon M \times M \to \mathbb{R}$ is a function called the *metric* satisfying the following axioms:

- **Positivity**. For all $x, y \in M$, we have $\mathfrak{d}(x, y) \geq 0$ with equality if and only if $x = y$.
- **Symmetry**. For all $x, y \in M$, we have $\mathfrak{d}(x, y) = \mathfrak{d}(y, x)$.
- **Triangle inequality**. For all $x, y, z \in M$, we have $\mathfrak{d}(x, y) \leq \mathfrak{d}(x, z) + \mathfrak{d}(z, y)$. □

Sometimes we will not mention $\mathfrak{d}$ and just say that M is a metric space.

The triangle inequality has this name because it expresses the fact that the length of a side of a triangle is at most the sum of the lengths of the other two sides:



Another way of thinking about this is that there are no "shortcuts": if you travel from x to z to y, then you must have moved at least as far as you would have if you went directly from x to y.

Example 2.1.2. Define a metric $\mathfrak{d}\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ via the formula

$$\mathfrak{d}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| \quad \text{for } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d.$$

The pair $(\mathbb{R}^d, \mathfrak{d})$ is a metric space. The triangle inequality follows from the triangle inequality for the Euclidean norm: for $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^d$, we have

$$\mathfrak{d}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \|(\mathbf{x} - \mathbf{z}) - (\mathbf{y} - \mathbf{z})\| \leq \|\mathbf{x} - \mathbf{z}\| + \|\mathbf{y} - \mathbf{z}\| = \mathfrak{d}(\mathbf{x}, \mathbf{z}) + \mathfrak{d}(\mathbf{y}, \mathbf{z}).$$

We call this the *Euclidean metric* on $\mathbb{R}^d$. □

Example 2.1.3. Let M be a set. The *discrete metric* on M is the function $\mathfrak{d}\colon M \times M \to \mathbb{R}$ defined by

$$\mathfrak{d}(x, y) = \begin{cases} 1 & \text{if } x \neq y, \\ 0 & \text{if } x = y \end{cases} \quad \text{for } x, y \in M.$$

This is a metric (see Exercise 2.1). □

Example 2.1.4. A set can typically be endowed with many different metrics. For instance, $\mathbb{R}^d$ can also be endowed with the following metrics (see Exercise 2.2):

- The ℓ^1-metric:

$$\mathfrak{d}_1(\mathbf{x}, \mathbf{y}) = \sum_{n=1}^{d} |x_n - y_n| \quad \text{for } \mathbf{x} = (x_1, \ldots, x_d) \text{ and } \mathbf{y} = (y_1, \ldots, y_d).$$

- The ℓ^∞-metric:

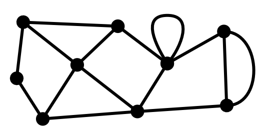
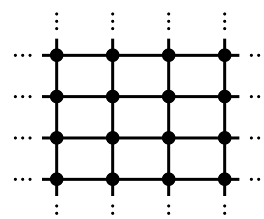
$$\mathfrak{d}_\infty(\mathbf{x}, \mathbf{y}) = \max\{|x_1 - y_1|, \ldots, |x_d - y_d|\} \quad \text{for } \mathbf{x} = (x_1, \ldots, x_d) \text{ and } \mathbf{y} = (y_1, \ldots, y_d).$$

We remark that the Euclidean metric is also often called the ℓ^2-metric. In Chapter 4, we will explain one sense in which these all give the "same" metric space: using language we will introduce in that chapter, they are all bi-Lipschitz equivalent to each other. See Exercise 4.16*. □

Example 2.1.5. If $(M, \mathfrak{d})$ is a metric space and $M' \subset M$ is a subset, then $\mathfrak{d}' = \mathfrak{d}|_{M' \times M'}$ is a metric on M' making $(M', \mathfrak{d}')$ into a metric space. We call $(M', \mathfrak{d}')$ a *metric subspace* or sometimes just a *subspace* of $(M, \mathfrak{d})$. □

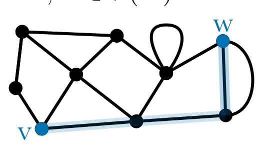
Example 2.1.6. Let $M \subset \mathbb{R}^d$. Endow $\mathbb{R}^d$ with the Euclidean metric. Viewing M as a metric subspace of $\mathbb{R}^d$, we get a metric space $(M, \mathfrak{d})$. We call $\mathfrak{d}$ the *Euclidean metric* on M. Many of the most important spaces in algebraic topology are either subspaces of $\mathbb{R}^d$ themselves or are constructed from subspaces of $\mathbb{R}^d$ by geometric operations (gluing, taking quotients, etc.). □

Example 2.1.7. Recall that a graph X is a set $\mathcal{V}(X)$ of vertices together with a set $\mathcal{E}(X)$ of edges, each of which connects two vertices in $\mathcal{V}(X)$. We allow edges to join a vertex to itself, and we also allow there to be more than one edge joining two vertices. Here are two examples:


and


An *edge path* in X connecting vertices $v, v' \in \mathcal{V}(X)$ is a sequence of vertices $v = v_0, v_1, \ldots, v_n = v'$ together with a choice of edge e_i joining v_{i-1} to v_i for all $1 \leq i \leq n$. The integer n is the *length* of the edge path. We allow $n = 0$, in which case the edge path has no edges and $v = v'$. A graph X is *connected* if $\mathcal{V}(X) \neq \emptyset$ and any two of its vertices can be connected by an edge path.

Let X be a connected graph. We can define a metric $\mathfrak{d}$ on $\mathcal{V}(X)$ as follows. For vertices $v, w \in \mathcal{V}(X)$, let $\mathfrak{d}(v, w)$ be the minimal length of an edge path connecting v and w. Since X is connected, the set of lengths of edge paths connecting v and w is a nonempty subset of $\mathbb{Z}_{\geq 0}$, so this minimum makes sense. For example, for $v, w \in \mathcal{V}(X)$ as below we have $\mathfrak{d}(v, w) = 3$:



The highlighted edge path realizes the minimal distance. The pair $(\mathcal{V}(X), \mathfrak{d})$ is a metric space (see Exercise 2.2). □

Example 2.1.8. Let S be a set and let $\mathrm{Seq}(S)$ be the set of sequences $\mathbf{s} = (s_m)_{m \geq 1}$ of elements of S. The *first-difference metric* $\mathfrak{d}$ on $\mathrm{Seq}(S)$ is as follows. Consider $\mathbf{s}, \mathbf{t} \in \mathrm{Seq}(S)$. Write $\mathbf{s} = (s_m)_{m \geq 1}$ and $\mathbf{t} = (t_m)_{m \geq 1}$. If $\mathbf{s} = \mathbf{t}$, then set $\mathfrak{d}(\mathbf{s}, \mathbf{t}) = 0$. Otherwise, let $k = \min\{m \geq 1 \mid s_m \neq t_m\}$ and set $\mathfrak{d}(\mathbf{s}, \mathbf{t}) = 1/k$. This is a metric (see Exercise 2.4). □

Example 2.1.9. Let ℓ^∞ be the set of sequences $\mathbf{x} = (x_m)_{m \geq 1}$ of real numbers that are bounded, i.e., such that there exists some $C \geq 0$ with $|x_m| \leq C$ for all $m \geq 1$. This can be endowed with the structure of a metric space via the following metric $\mathfrak{d}$:

$$\mathfrak{d}(\mathbf{x}, \mathbf{y}) = \sup_{m \geq 1} |x_m - y_m| \quad \text{for } \mathbf{x} = (x_m)_{m \geq 1} \text{ and } \mathbf{y} = (y_m)_{m \geq 1} \text{ in } \ell^\infty.$$

Since both $(x_m)_{m\geq 1}$ and $(y_m)_{m\geq 1}$ are bounded, this supremum is finite. The triangle inequality follows from the following calculation: for $\mathbf{x} = (x_m)_{m\geq 1}$ and $\mathbf{y} = (y_m)_{m\geq 1}$ and $\mathbf{z} = (z_m)_{m\geq 1}$ in ℓ^∞,

$$\begin{aligned}\mathfrak{d}(\mathbf{x},\mathbf{y}) &= \sup_{m\geq 1}|x_m - y_m| = \sup_{m\geq 1}|(x_m - z_m) - (y_m - z_m)| \\ &\leq \sup_{m\geq 1}(|x_m - z_m| + |y_m - z_m|) \leq \sup_{m\geq 1}|x_m - z_m| + \sup_{m\geq 1}|y_m - z_m| = \mathfrak{d}(\mathbf{x},\mathbf{z}) + \mathfrak{d}(\mathbf{z},\mathbf{y}). \quad \square\end{aligned}$$

EXAMPLE 2.1.10. Let $I = [0,1]$ be the closed interval and let $\mathcal{C}(I,\mathbb{R})$ be the set of all continuous functions $f\colon I \to \mathbb{R}$. Define a metric on $\mathcal{C}(I,\mathbb{R})$ as follows:

$$\mathfrak{D}(f,g) = \sup\{|f(t) - g(t)| \mid t \in I\} \quad \text{for continuous } f, g\colon I \to \mathbb{R}.$$

Since the closed interval I is sequentially compact, this supremum is finite and attained at some point of I (see[1] Theorem 1.11.2). The metric $\mathfrak{D}$ is called the *sup metric* on $\mathcal{C}(I,\mathbb{R})$. It makes $\mathcal{C}(I,\mathbb{R})$ into a metric space (see Exercise 2.2). $\square$

2.2. Products of metric spaces

Let $(M_1,\mathfrak{d}_1),\ldots,(M_d,\mathfrak{d}_d)$ be metric spaces. Set $X = M_1 \times \cdots \times M_d$. Define $\mathfrak{d}_{prod}\colon X \times X \to \mathbb{R}$ as follows:

$$\mathfrak{d}_{\text{prod}}(\mathbf{x},\mathbf{y}) = \sum_{k=1}^{d}\mathfrak{d}_k(x_k,y_k) \quad \text{for } \mathbf{x} = (x_1,\ldots,x_d) \text{ and } \mathbf{y} = (y_1,\ldots,y_d) \text{ in } X.$$

This is a metric on X that we will call the *product metric* (see Exercise 2.3). The metric space $(X,\mathfrak{d}_{\text{prod}})$ will be called the *metric product* of $(M_1,\mathfrak{d}_1),\ldots,(M_d,\mathfrak{d}_d)$.

EXAMPLE 2.2.1. Consider $\mathbb{R}$ with the standard Euclidean metric $\mathfrak{d}(x,y) = |x-y|$. The product metric on $\mathbb{R}^d = \mathbb{R}\times\cdots\times\mathbb{R}$ is then the ℓ^1-metric

$$\mathfrak{d}_1(\mathbf{x},\mathbf{y}) = |x_1 - y_1| + \cdots + |x_d - y_d| \quad \text{for } \mathbf{x} = (x_1,\ldots,x_d) \text{ and } \mathbf{y} = (y_1,\ldots,y_d) \text{ in } \mathbb{R}^d.$$

We remark that we could have defined the product metric to ensure that this was the Euclidean metric. However, the product metric as defined above is easier to work with. As we said earlier, in Chapter 4 we will explain why $\mathfrak{d}_1$ is equivalent to the Euclidean metric from the standpoint of topology, so we might as well use the easier product metric. $\square$

2.3. Continuity

Once we have defined metric spaces, we can define continuity by imitating the classical definition from Chapter 1:

DEFINITION 2.3.1. Let $(M,\mathfrak{d}_M)$ and $(N,\mathfrak{d}_N)$ be metric spaces and let $f\colon M \to N$ be a function. Then:

- f is *continuous at* $x_0 \in M$ if for all $\epsilon > 0$, there exists a $\delta > 0$ such that for all $x \in M$ with $\mathfrak{d}_M(x,x_0) < \delta$ we have $\mathfrak{d}_N(f(x),f(x_0)) < \epsilon$.
- f is *continuous* if it is continuous at all $x_0 \in M$. $\square$

EXAMPLE 2.3.2. Let $M \subset \mathbb{R}^d$ and $N \subset \mathbb{R}^e$ be equipped with their Euclidean metrics and let $f\colon M \to N$ be a function. In this special case, the above definition of what it means for f to be continuous is exactly the same as the definition from Chapter 1. $\square$

EXAMPLE 2.3.3. Let $(M,\mathfrak{d})$ be a metric space and let $p \in M$ be a point. Define $f\colon M \to \mathbb{R}$ via the formula

$$f(x) = \mathfrak{d}(x,p) \quad \text{for } x \in M.$$

Equipping $\mathbb{R}$ with the Euclidean metric, the function f is continuous. Indeed, consider $x_0 \in M$ and $\epsilon > 0$. Setting $\delta = \epsilon$, if $x \in M$ satisfies $\mathfrak{d}(x,x_0) < \delta = \epsilon$ then

$$|f(x) - f(x_0)| = |\,\mathfrak{d}(x,p) - \mathfrak{d}(x_0,p)|. \tag{2.3.1}$$

[1]Here we are using the fact that the function $\phi\colon I \to \mathbb{R}$ defined by $\phi(t) = |f(t) - g(t)|$ for $t \in I$ is continuous.

We must prove this is less than ϵ. The triangle inequality implies that

$$\mathfrak{d}(x,p) \le \mathfrak{d}(x,x_0) + \mathfrak{d}(x_0,p) \quad \text{and} \quad \mathfrak{d}(x_0,p) \le \mathfrak{d}(x_0,x) + \mathfrak{d}(x,p).$$

Rearranging these, we see that

$$\mathfrak{d}(x,p) - \mathfrak{d}(x_0,p) \le \mathfrak{d}(x,x_0) \quad \text{and} \quad \mathfrak{d}(x_0,p) - \mathfrak{d}(x,p) \le \mathfrak{d}(x_0,x) = \mathfrak{d}(x,x_0).$$

These imply that (2.3.1) is at most $\mathfrak{d}(x,x_0) < \epsilon$, as desired. □

Continuous functions between metric spaces have many of the same formal properties as continuous functions between subsets of Euclidean space. One important such property is as follows:

LEMMA 2.3.4. *Let $(M,\mathfrak{d}_M)$ and $(N,\mathfrak{d}_N)$ and $(P,\mathfrak{d}_P)$ be metric spaces. Let $f\colon M \to N$ and $g\colon N \to P$ be continuous functions. Then $g \circ f\colon M \to P$ is continuous.*

PROOF. Let $x_0 \in M$ and let $\epsilon > 0$. Since g is continuous at $f(x_0)$, there is some $\delta' > 0$ such that for all $y \in N$ with $\mathfrak{d}_N(y, f(x_0)) < \delta'$ we have $\mathfrak{d}_P(g(y), g(f(x_0))) < \epsilon$. Letting $\epsilon' = \delta'$, since f is continuous at x_0 there is some $\delta > 0$ such that for all $x \in M$ with $\mathfrak{d}_M(x,x_0) < \delta$ we have $\mathfrak{d}_N(f(x), f(x_0)) < \epsilon' = \delta'$. Combining these, for $x \in M$ with $\mathfrak{d}_M(x,x_0) < \delta$ we have $\mathfrak{d}_P(g(f(x)), g(f(x_0))) < \epsilon$, as desired. □

2.4. Open sets

Fix a metric space $(M,\mathfrak{d})$. For $x_0 \in M$ and $r > 0$, let

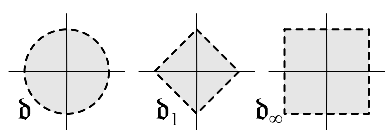
$$B_r(x_0) = \{x \in M \mid \mathfrak{d}(x,x_0) < r\}.$$

This is called the *open ball* of radius r around x_0.

EXAMPLE 2.4.1. On $\mathbb{R}^2$, recall that the Euclidean metric $\mathfrak{d}$, the ℓ^1-metric $\mathfrak{d}_1$, and the ℓ^∞-metric $\mathfrak{d}_\infty$ are as follows:

$$\begin{aligned}
\mathfrak{d}((x,y),(z,w)) &= \sqrt{(x-z)^2 + (y-w)^2}, \\
\mathfrak{d}_1((x,y),(z,w)) &= |x-z| + |y-w|, \\
\mathfrak{d}_\infty((x,y),(z,w)) &= \max\{|x-z|, |y-w|\}.
\end{aligned}$$

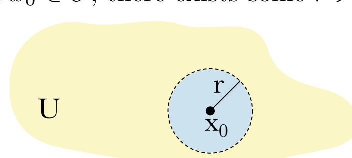
The balls of radius $r = 1$ around $x_0 = (0,0)$ in these three metrics are as follows:



The pictures are similar in higher dimensions, and it is enlightening to draw them in $\mathbb{R}^3$. □

EXAMPLE 2.4.2. Let S be a set, let $\operatorname{Seq}(S)$ be the set of sequences $\mathbf{s} = (s_m)_{m \ge 1}$ of elements of S, and let $\mathfrak{d}$ be the first-difference metric on $\operatorname{Seq}(S)$ (see Example 2.1.8). Consider a point $\mathbf{s} = (s_m)_{m\ge1}$ of $\operatorname{Seq}(S)$ and some $r > 0$. Let $n \ge 0$ be the largest integer that is less than or equal to $1/r$. A point $\mathbf{t} = (t_m)_{m\ge1}$ of $\operatorname{Seq}(S)$ lies in $B_r(\mathbf{s})$ precisely when $t_m = s_m$ for all $1 \le m \le n$ (see Exercise 2.4). □

A set $U \subset M$ is *open* if for all $x_0 \in U$, there exists some $r > 0$ such that $B_r(x_0) \subset U$:



EXAMPLE 2.4.3. The set $B_r(x_0)$ is open (see Exercise 2.6). □

EXAMPLE 2.4.4. An open interval in $\mathbb{R}$ is a set of the form (a,b) with $-\infty \le a < b \le \infty$. Open intervals are open, and each open set in $\mathbb{R}$ can be written as a union of open intervals. In fact, it turns out that open sets in $\mathbb{R}$ can be written as the union of countably[2] many *disjoint* open intervals (see Exercise 2.16). □

[2] Our convention is that a finite set is countable. A *countably infinite* set is a countable set that is infinite.

EXAMPLE 2.4.5. Let M be a metric space and let $X \subset M$ be a metric subspace. The open sets in X are exactly the sets of the form $U \cap X$ for an open set U of M (see Exercise 2.14). $\square$

If $A \subset M$ is an arbitrary subset then the *interior* of A, denoted $\mathrm{Int}(A)$, is the set of all $x_0 \in A$ such that there exists some $r > 0$ such that $B_r(x_0) \subset A$. We have:

LEMMA 2.4.6. *Let $(M, \mathfrak{d})$ be a metric space and let $A \subset M$. Then $\mathrm{Int}(A)$ is an open set contained in A. Moreover, $\mathrm{Int}(A)$ is the largest open set contained in A in the following sense: if $U \subset A$ is open, then $U \subset \mathrm{Int}(A)$.*

PROOF. Once we prove that $\mathrm{Int}(A)$ is open, the remaining conclusions will immediately follow. To see that $\mathrm{Int}(A)$ is open, consider $x_0 \in \mathrm{Int}(A)$. By definition, there is some $r > 0$ such that $B_r(x_0) \subset A$. We claim that $B_r(x_0) \subset \mathrm{Int}(A)$. Indeed, consider $x_1 \in B_r(x_0)$. We must prove that $x_1 \in \mathrm{Int}(A)$. Set $r' = r - \mathfrak{d}(x_1, x_0) > 0$. For $x \in B_{r'}(x_1)$, the triangle inequality implies that

$$\mathfrak{d}(x, x_0) \leq \mathfrak{d}(x, x_1) + \mathfrak{d}(x_1, x_0) < (r - \mathfrak{d}(x_1, x_0)) + \mathfrak{d}(x_1, x_0) = r,$$

so $x \in B_r(x_0) \subset A$. It follows that $B_{r'}(x_1) \subset A$, so $x_1 \in \mathrm{Int}(A)$, as desired. $\square$

EXAMPLE 2.4.7. The closed unit disk in $\mathbb{R}^d$ is

$$\mathbb{D}^d = \left\{(x_1, \ldots, x_d) \in \mathbb{R}^d \mid x_1^2 + \cdots + x_d^2 \leq 1\right\}.$$

We have

$$\mathrm{Int}(\mathbb{D}^d) = \left\{(x_1, \ldots, x_d) \in \mathbb{R}^d \mid x_1^2 + \cdots + x_d^2 < 1\right\} = B_1(\mathbf{0}). \qquad \square$$

Open subsets of metric spaces are closed under arbitrary unions and finite intersections:

LEMMA 2.4.8. *Let $(M, \mathfrak{d})$ be a metric space. Then:*

- *The empty set $\emptyset$ and the whole set M are open.*
- *If $\{U_i\}_{i \in I}$ is an arbitrary collection of open subsets of M, then $\cup_{i \in I} U_i$ is open.*
- *If $\{U_1, \ldots, U_n\}$ is a finite collection of open subsets of M, then $\cap_{i=1}^n U_i$ is open.*

PROOF. See Exercise 2.11. $\square$

REMARK 2.4.9. In Lemma 2.4.8, we allow the empty union (i.e., the union of no subsets) and the empty intersection (i.e., the intersection of no subsets). For subsets of M, our convention is that the empty union is $\emptyset$ and the empty intersection is M. $\square$

2.5. Limits and continuity

Fix a metric space $(M, \mathfrak{d})$. Consider a sequence $(x_n)_{n \geq 1}$ of points of M. We say that the sequence $(x_n)_{n \geq 1}$ *converges* to $x \in M$ if for all $\epsilon > 0$, there exists some $N \geq 1$ such that $\mathfrak{d}(x_n, x) < \epsilon$ for all $n \geq N$. Equivalently, the sequence of real numbers $\mathfrak{d}(x_n, x)$ converges to 0 as $n \to \infty$. If this holds, we write $\lim_{n \to \infty} x_n = x$ and say that $(x_n)_{n \geq 1}$ is a *convergent sequence*. The following shows that a sequence $(x_n)_{n \geq 1}$ can converge to at most one point of M, so this notation is unambiguous:

LEMMA 2.5.1. *Let $(M, \mathfrak{d})$ be a metric space and let $(x_n)_{n \geq 1}$ be a sequence of points of M that converges to points $x \in M$ and $x' \in M$. Then $x = x'$.*

PROOF. Let $\epsilon > 0$. Since $(x_n)_{n \geq 1}$ converges to x and x', we can find $N \geq 1$ such that for $n \geq N$ we have $\mathfrak{d}(x_n, x) < \epsilon/2$ and $\mathfrak{d}(x_n, x') < \epsilon/2$. Using the triangle inequality, we thus have

$$\mathfrak{d}(x, x') \leq \mathfrak{d}(x, x_N) + \mathfrak{d}(x_N, x') < \epsilon/2 + \epsilon/2 = \epsilon.$$

Since $\mathfrak{d}(x, x') < \epsilon$ for all $\epsilon > 0$, it follows that $\mathfrak{d}(x, x') = 0$ and thus $x = x'$, as desired. $\square$

A sequence $(x_n)_{n \geq 1}$ of points of M is a *Cauchy sequence* if for all $\epsilon > 0$, there exists some $N \geq 1$ such that for all $n, m \geq N$ we have $\mathfrak{d}(x_n, x_m) < \epsilon$. We have:

LEMMA 2.5.2. *Let $(M, \mathfrak{d})$ be a metric space and let $(x_n)_{n \geq 1}$ be a convergent sequence of points of M. Then $(x_n)_{n \geq 1}$ is a Cauchy sequence.*

PROOF. Consider $\epsilon > 0$. Setting $x = \lim_{n \to \infty} x_n$, there exists some $N \geq 1$ such that for all $n \geq N$ we have $\mathfrak{d}(x_n, x) < \epsilon/2$. For $n, m \geq N$, we have $\mathfrak{d}(x_n, x_m) \leq \mathfrak{d}(x_n, x) + \mathfrak{d}(x, x_m) < \epsilon/2 + \epsilon/2 = \epsilon$. $\square$

Here is an example showing that not all Cauchy sequences converge:

EXAMPLE 2.5.3. Let $M = (0,1)$ with the Euclidean metric. For $n \geq 1$, set $x_n = 1/(n+1) \in M$. The sequence $(x_n)_{n\geq 1}$ of points of M is a Cauchy sequence since in $\mathbb{R}$ it converges to 0 (see Lemma 2.5.2). However, since $0 \notin M$ it does not converge in M (see Lemma 2.5.1). □

REMARK 2.5.4. A metric space in which all Cauchy sequences converge is said to be *complete*. We will discuss this in Chapter 6. □

We can express continuity using limits as follows:

LEMMA 2.5.5. *Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces and let $f \colon M \to N$ be a function. Then f is continuous if and only if the following holds:*

(♠) *For all convergent sequences $(x_n)_{n\geq 1}$ of points of M, the sequence $(f(x_n))_{n\geq 1}$ of points of N is convergent and $\lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n)$.*

PROOF. We must prove the following two claims:

CLAIM. *Assume that f is continuous. Then (♠) holds.*

Let $(x_n)_{n\geq 1}$ be a convergent sequence of points of M. Set $x = \lim_{n\to\infty} x_n$. Our goal is to prove that $\lim_{n\to\infty} f(x_n) = f(x)$. Consider some $\epsilon > 0$. We must find some $N \geq 1$ such that for $n \geq N$ we have $\mathfrak{d}_N(f(x), f(x_n)) < \epsilon$. Since f is continuous at x, there is some $\delta > 0$ such that if $y \in M$ satisfies $\mathfrak{d}_M(y, x) < \delta$, then $\mathfrak{d}_N(f(y), f(x)) < \epsilon$. Choose $N \geq 1$ such that for $n \geq N$ we have $\mathfrak{d}_M(x, x_n) < \delta$. It follows that for $n \geq N$ we have $\mathfrak{d}_N(f(x), f(x_n)) < \epsilon$, as desired.

CLAIM. *Assume that (♠) holds. Then f is continuous.*

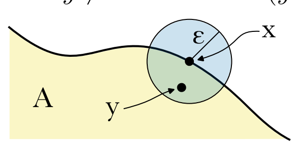
Assume for the sake of contradiction that f is not continuous at some point $x_0 \in M$. There thus exists some $\epsilon > 0$ such that for all $\delta > 0$, it is *not* the case that for all $x \in M$ with $\mathfrak{d}_M(x, x_0) < \delta$ we have $\mathfrak{d}_N(f(x), f(x_0)) < \epsilon$. For each $n \geq 1$, there therefore exists some $x_n \in M$ with $\mathfrak{d}_M(x_n, x_0) < 1/n$ and $\mathfrak{d}_N(f(x_n), f(x_0)) \geq \epsilon$. Since $\mathfrak{d}_M(x_n, x_0) < 1/n$ for all $n \geq 1$, it follows that

$$\lim_{n\to\infty} \mathfrak{d}_M(x_n, x_0) = 0.$$

This implies that $\lim_{n\to\infty} x_n = x_0$. It thus follows from (♠) that $\lim_{n\to\infty} f(x_n) = f(x_0)$. In particular, there exists some $N \geq 1$ such that for $n \geq N$ we have $\mathfrak{d}_N(f(x_n), f(x_0)) < \epsilon$, contradicting the fact that $\mathfrak{d}_N(f(x_n), f(x_0)) \geq \epsilon$. □

2.6. Limit points

Fix a metric space $(M, \mathfrak{d})$. Let $A \subset M$ be a set. A *limit point* of A is a point $x \in M$ such that for all $\epsilon > 0$ there exists some $y \in A$ with $y \neq x$ such that $\mathfrak{d}(y, x) < \epsilon$:



EXAMPLE 2.6.1. Let $M = \mathbb{R}$ and let $A = \mathbb{Q}$. All points of M are limit points of A. □

EXAMPLE 2.6.2. Let $M = \mathbb{R}$ and let $A = \mathbb{Z}$. No points of M are limit points of A. □

REMARK 2.6.3. In Example 2.6.2, each $a \in A = \mathbb{Z}$ is an *isolated point*, i.e., there exists $r > 0$ such that $B_r(a) \cap A = \{a\}$. By definition, isolated points of a set A cannot be limit points of A. □

The following lemma gives an alternate way of thinking about limit points:

LEMMA 2.6.4. *Let $(M, \mathfrak{d})$ be a metric space and let $A \subset M$. A point $x \in M$ is a limit point of A if and only if there exists a convergent sequence $(y_n)_{n\geq 1}$ of points of A with $y_n \neq x$ for all $n \geq 1$ such that $\lim_{n\to\infty} y_n = x$.*

PROOF. Fix $x \in M$. We must prove the following two claims:

Claim. *Assume that there exists a convergent sequence $(y_n)_{n\geq 1}$ of points of A with $y_n \neq x$ for all $n \geq 1$ such that $\lim_{n\to\infty} y_n = x$. Then x is a limit point of A.*

Consider $\epsilon > 0$. Since $\lim_{n\to\infty} y_n = x$, there exists some $N \geq 1$ such that $\mathfrak{d}(y_n, x) < \epsilon$ for $n \geq N$. In particular, $y_N \in A$ is a point satisfying $y_N \neq x$ and $\mathfrak{d}(y_N, x) < \epsilon$. We conclude that x is a limit point of A.

Claim. *Assume that x is a limit point of A. Then there exists a convergent sequence $(y_n)_{n\geq 1}$ of points of A with $y_n \neq x$ for all $n \geq 1$ such that $\lim_{n\to\infty} y_n = x$.*

For $n \geq 1$, we can therefore find $y_n \in A$ such that $y_n \neq x$ and $\mathfrak{d}(y_n, x) < 1/n$. Since $\mathfrak{d}(y_n, x) < 1/n$, it follows that $\lim_{n\to\infty} \mathfrak{d}(y_n, x) = 0$. This implies that $\lim_{n\to\infty} y_n = x$, as desired. □

Corollary 2.6.5. *Let $(M, \mathfrak{d})$ be a metric space, let $A \subset M$, and let x be a limit point of A. Then for all $r > 0$, there exist infinitely many $y \in B_r(x)$ with $y \in A$ and $y \neq x$.*

Proof. By Lemma 2.6.4, there exists a sequence $(y_n)_{n\geq 1}$ of points of A with $y_n \neq x$ for all $n \geq 1$ such that $\lim_{n\to\infty} y_n = x$. Fix some $r > 0$. We can find $n_1 \geq 1$ such that $\mathfrak{d}(y_{n_1}, x) < r$. We can then find $n_2 > n_1$ such that $\mathfrak{d}(y_{n_2}, x) < \mathfrak{d}(y_{n_1}, x)$. This process can be repeated over and over, and the result is a sequence $(y_{n_k})_{k\geq 1}$ of points of A such that

$$r > \mathfrak{d}(y_{n_1}, x) > \mathfrak{d}(y_{n_2}, x) > \mathfrak{d}(y_{n_3}, x) > \cdots .$$

Since the $\mathfrak{d}(y_{n_k}, x)$ are strictly decreasing, they are in particular distinct. This implies that the points $(y_{n_k})_{k\geq 1}$ are all distinct. The corollary follows. □

2.7. Closed sets

Let $(M, \mathfrak{d})$ be a metric space. A set $C \subset M$ is *closed* if it contains all of its limit points.

Example 2.7.1. Every finite subset $C \subset M$ is closed. Indeed, a finite subset of a metric space has no limit points, and thus vacuously contains all of its limit points. □

Example 2.7.2. The set

$$\mathbb{D}^d = \left\{(x_1, \ldots, x_d) \in \mathbb{R}^d \mid x_1^2 + \cdots + x_d^2 \leq 1\right\}$$

is a closed subset of $\mathbb{R}^d$. □

Example 2.7.3. Unlike open subsets of $\mathbb{R}$, closed subsets of $\mathbb{R}$ do not have a simple description. For instance, the classical *Cantor set* is as follows. Each $x \in I = [0, 1]$ can be written as[3]

$$x = \sum_{n=1}^{\infty} \frac{x_n}{3^n} \quad \text{with } x_n \in \{0, 1, 2\} \text{ for all } n \geq 1.$$

The Cantor set C consists of all $x \in I$ that can be written as

$$x = \sum_{n=1}^{\infty} \frac{x_n}{3^n} \quad \text{with } x_n \in \{0, 2\} \text{ for all } n \geq 1.$$

The set C is closed (see Exercise 2.17*). □

Example 2.7.4. Let M be a metric space and let $X \subset M$ be a metric subspace. The closed sets in X are exactly the sets of the form $D \cap X$ for a closed set D of M (see Exercise 2.14). □

Another way to think about closed sets is as follows:

Lemma 2.7.5. *Let $(M, \mathfrak{d})$ be a metric space and let $C \subset M$. Then C is closed if and only if for all sequences $(y_n)_{n\geq 1}$ of points of C that converge in M, we have $\lim_{n\to\infty} y_n \in C$.*

Proof. We must prove the following two claims:

Claim. *Assume that for all sequences $(y_n)_{n\geq 1}$ of points of C that converge in M, we have $\lim_{n\to\infty} y_n \in C$. Then C is closed.*

---

[3] This expression is not unique; for instance, $1/3 = 2/3^2 + 2/3^3 + 2/3^4 + \cdots$. This is the same ambiguity that occurs in ordinary base-10 decimal expansions and results in identities like $1 = 0.999\cdots$.

Let x be a limit point of C. We must prove that $x \in C$. By Lemma 2.6.4, there exists a sequence $(y_n)_{n\geq 1}$ of points of C with $y_n \neq x$ for all $n \geq 1$ such that $\lim_{n\to\infty} y_n = x$. This is even stronger than the condition on sequences in our assumption, which does not require that $y_n \neq x$ for all $n \geq 1$. By our assumption, $x = \lim_{n\to\infty} y_n \in C$.

Claim. *Assume that C is closed. Then for all sequences $(y_n)_{n\geq 1}$ of points of C that converge in M, we have $\lim_{n\to\infty} y_n \in C$.*

Consider a sequence $(y_n)_{n\geq 1}$ of points of C that converge in M. Set $x = \lim_{n\to\infty} y_n$. We must prove that $x \in C$. If x is a limit point of C, then since C is closed we must have $x \in C$. Assume, therefore, that x is not a limit point of C. There then exists some $\epsilon > 0$ such that there does not exist $y \in C$ with $y \neq x$ and $\mathfrak{d}(y, x) < \epsilon$. Since $\lim_{n\to\infty} y_n = x$, we can find some $N \geq 1$ such that for $n \geq N$ we have $\mathfrak{d}(y_n, x) < \epsilon$. Since $y_n \in C$ for all $n \geq 1$, the only way this can hold is if $y_n = x$ for all $n \geq N$. In particular, $x = y_N \in C$. $\square$

Let $A \subset M$ be a subset. The *closure* of A, denoted $\overline{A}$, is the union of A with its limit points. We have:

Lemma 2.7.6. *Let $(M, \mathfrak{d})$ be a metric space and let $A \subset M$. Then $\overline{A}$ is a closed set containing A. Moreover, $\overline{A}$ is the smallest closed set containing A in the following sense: if C is a closed set with $A \subset C$, then $\overline{A} \subset C$.*

Proof. Once we prove that $\overline{A}$ is closed, the remaining conclusions will immediately follow. Let $x \in M$ be a limit point of $\overline{A}$. We must prove that $x \in \overline{A}$. To do this, it is enough to prove that x is a limit point of A.

Consider $\epsilon > 0$. We must find $y \in A$ with $y \neq x$ and $\mathfrak{d}(y, x) < \epsilon$. Since x is a limit point of $\overline{A}$, there exists $z \in \overline{A}$ with $z \neq x$ and $\mathfrak{d}(z, x) < \epsilon/2$. If $z \in A$, then we can take $y = z$. Otherwise, z must be a limit point of A. By Corollary 2.6.5, there exist infinitely many $y \in A$ with $y \neq z$ and $\mathfrak{d}(y, z) < \epsilon/2$. Since there are infinitely many, we can find such a $y \in A$ with $y \neq x$ as well. By the triangle inequality, $\mathfrak{d}(y, x) \leq \mathfrak{d}(y, z) + \mathfrak{d}(z, x) < \epsilon/2 + \epsilon/2 = \epsilon$. $\square$

Closed subsets of metric spaces are closed under arbitrary intersections and finite unions (see Remark 2.4.9 for our conventions about the empty union and the empty intersection):

Lemma 2.7.7. *Let M be a metric space. Then:*

- *The empty set $\emptyset$ and the whole set M are closed.*
- *If $\{C_i\}_{i\in I}$ is an arbitrary collection of closed subsets of M, then $\cap_{i\in I} C_i$ is closed.*
- *If $\{C_1, \ldots, C_n\}$ is a finite collection of closed subsets of M, then $\cup_{i=1}^n C_i$ is closed.*

Proof. See Exercise 2.11. $\square$

2.8. Relating open and closed sets

We close this chapter by relating open and closed sets:

Lemma 2.8.1. *Let $(M, \mathfrak{d})$ be a metric space and let $A \subset M$ be a set. Then A is open if and only if $M \setminus A$ is closed.*

Proof. We must prove the following two claims:

Claim. *Assume that A is open. Then $M \setminus A$ is closed.*

Let $x \in M$ be a limit point of $M \setminus A$. We must prove that $x \in M \setminus A$. Assume otherwise. We then have $x \in A$, so since A is open there exists some $r > 0$ such that $B_r(x) \subset A$. Since $B_r(x)$ contains no points of $M \setminus A$, it follows that x is not a limit point of $M \setminus A$, contradicting the fact that it is such a limit point.

Claim. *Assume that $M \setminus A$ is closed. Then A is open.*

Consider $a \in A$. We must find some $r > 0$ such that $B_r(a) \subset A$. Since $M \setminus A$ is closed and $a \in A$, it must be the case that a is not a limit point of $M \setminus A$. There therefore exists some $r > 0$ such that $B_r(a)$ contains no points of $M \setminus A$, i.e., such that $B_r(a) \subset A$. $\square$

2.9. Exercises

Exercise 2.1. Let M be any set. Endow M with the discrete metric $\mathfrak{d}$, so

$$\mathfrak{d}(x,y) = \begin{cases} 1 & \text{if } x \neq y, \\ 0 & \text{if } x = y \end{cases} \quad \text{for } x, y \in M.$$

Prove that $(M, \mathfrak{d})$ is a metric space, and identify its open sets and its closed sets. □

Exercise 2.2. Prove that the following are metrics, verifying in particular the triangle inequality:

(a) On $\mathbb{R}^d$, the ℓ^1-metric

$$\mathfrak{d}_1(\mathbf{x}, \mathbf{y}) = \sum_{n=1}^{d} |x_n - y_n| \quad \text{for } \mathbf{x} = (x_1, \ldots, x_d) \text{ and } \mathbf{y} = (y_1, \ldots, y_d).$$

(b) On $\mathbb{R}^d$, the ℓ^∞-metric:

$$\mathfrak{d}_\infty(\mathbf{x}, \mathbf{y}) = \max\{|x_1 - y_1|, \ldots, |x_d - y_d|\} \quad \text{for } \mathbf{x} = (x_1, \ldots, x_d) \text{ and } \mathbf{y} = (y_1, \ldots, y_d).$$

(c) On the set $\mathcal{C}(I, \mathbb{R})$ of continuous functions $f \colon I \to \mathbb{R}$, the sup metric

$$\mathfrak{D}(f, g) = \sup\{|f(t) - g(t)| \mid t \in I\} \quad \text{for all continuous } f, g \colon I \to \mathbb{R}.$$

(d) For a connected graph X, the metric $\mathfrak{d}$ on the set $\mathcal{V}(X)$ of vertices where for $v, w \in \mathcal{V}(X)$ the value $\mathfrak{d}(v, w)$ is the minimal length of an edge path connecting v to w. □

Exercise 2.3. Let $(M_1, \mathfrak{d}_1), \ldots, (M_d, \mathfrak{d}_d)$ be metric spaces. Set $X = M_1 \times \cdots \times M_d$. Do the following:

(a) Recall that the product metric $\mathfrak{d}_{\text{prod}}$ on X is as follows:

$$\mathfrak{d}_{\text{prod}}(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^{d} \mathfrak{d}_k(x_k, y_k) \quad \text{for } \mathbf{x} = (x_1, \ldots, x_d) \text{ and } \mathbf{y} = (y_1, \ldots, y_d) \text{ in } X.$$

Prove that $\mathfrak{d}_{\text{prod}}$ is a metric on X.

(b) Let $(\mathbf{x}[n])_{n \geq 1}$ be a sequence of points in X. For each $n \geq 1$, write $\mathbf{x}[n] = (x[n]_1, \ldots, x[n]_d)$. Prove that $(\mathbf{x}[n])_{n \geq 1}$ converges to a point $\mathbf{y} = (y_1, \ldots, y_d)$ if and only if for all $1 \leq k \leq d$ we have $\lim_{n \to \infty} x[n]_k = y_k$. □

Exercise 2.4. Let S be a set, let $\operatorname{Seq}(S)$ be the set of sequences $\mathbf{s} = (s_m)_{m \geq 1}$ of elements of S, and let $\mathfrak{d}$ be the first-difference metric on $\operatorname{Seq}(S)$. Recall that $\mathfrak{d}$ is defined as follows. Consider $\mathbf{s}, \mathbf{t} \in \operatorname{Seq}(S)$. Write $\mathbf{s} = (s_m)_{m \geq 1}$ and $\mathbf{t} = (t_m)_{m \geq 1}$. If $\mathbf{s} = \mathbf{t}$, then set $\mathfrak{d}(\mathbf{s}, \mathbf{t}) = 0$. Otherwise, let $k = \min\{m \geq 1 \mid s_m \neq t_m\}$ and set $\mathfrak{d}(\mathbf{s}, \mathbf{t}) = 1/k$. Do the following:

(a) For $\mathbf{s}, \mathbf{t}, \mathbf{u} \in \operatorname{Seq}(S)$ prove that $\mathfrak{d}(\mathbf{s}, \mathbf{t}) \leq \max\{\mathfrak{d}(\mathbf{s}, \mathbf{u}), \mathfrak{d}(\mathbf{u}, \mathbf{t})\}$. This kind of inequality is sometimes called an *ultrametric inequality*.

(b) Prove that $\mathfrak{d}$ is a metric.

(c) Using the ultrametric inequality, prove that for all $\mathbf{s} \in \operatorname{Seq}(S)$ and $r > 0$, the open ball $B_r(\mathbf{s})$ is closed.

(d) Using the ultrametric inequality, prove that if B and B' are two open balls in $\operatorname{Seq}(S)$ with $B \cap B' \neq \emptyset$, then either $B \subset B'$ or $B' \subset B$.

(e) Consider a point $\mathbf{s} = (s_m)_{m \geq 1}$ of $\operatorname{Seq}(S)$ and some $r > 0$. Let $n \geq 0$ be the largest integer that is less than or equal to $1/r$. Prove that a point $\mathbf{t} = (t_m)_{m \geq 1}$ of $\operatorname{Seq}(S)$ lies in $B_r(\mathbf{s})$ if and only if $t_m = s_m$ for all $1 \leq m \leq n$.

(f) Let $(\mathbf{s}[n])_{n \geq 1}$ be a sequence of points in $\operatorname{Seq}(S)$. For $n \geq 1$, write $\mathbf{s}[n] = (s[n]_m)_{m \geq 1}$. Prove that $(\mathbf{s}[n])_{n \geq 1}$ converges to a point $\mathbf{t} = (t_m)_{m \geq 1}$ in $\operatorname{Seq}(S)$ if and only if for all $k \geq 1$, there exists some $N \geq 1$ such that for all $n \geq N$ we have $s[n]_i = t_i$ for all $1 \leq i \leq k$. □

Exercise 2.5. Let $(M, \mathfrak{d})$ be a metric space. Give $M \times M$ the product metric. Prove that the distance function $\mathfrak{d} \colon M \times M \to \mathbb{R}$ is continuous. We remark that this generalizes Example 2.3.3. □

Exercise 2.6. Let M be a metric space, let $x_0 \in M$, and let $r > 0$. Prove that $B_r(x_0)$ is open. □

EXERCISE 2.7. Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces and let $f\colon M \to N$ be a function. Assume that M is a finite set. Prove that f is continuous. $\square$

EXERCISE 2.8. Let $(M, \mathfrak{d})$ be a metric space and let $A \subset M$ be an arbitrary nonempty subset. Define a function $f\colon M \to \mathbb{R}$ via the formula

$$f(x) = \inf \{\mathfrak{d}(x, a) \mid a \in A\} \quad \text{for } x \in M.$$

Prove that f is continuous. $\square$

EXERCISE 2.9. Let $(M, \mathfrak{d})$ be a metric space. For $x_0 \in M$ and $r \geq 0$, define

$$B_r^{\mathrm{c}\ell}(x_0) = \{x \in M \mid \mathfrak{d}(x, x_0) \leq r\}.$$

Do the following:

(a) Prove that $B_r^{\mathrm{c}\ell}(x_0)$ is closed.
(b) Give an example with $r > 0$ to show that $B_r^{\mathrm{c}\ell}(x_0)$ can be different from the closure $\overline{B_r(x_0)}$ of the open ball $B_r(x_0)$. $\square$

EXERCISE 2.10. Let $(M, \mathfrak{d})$ be a metric space and let $f\colon M \to \mathbb{R}$ and $g\colon M \to \mathbb{R}$ be continuous functions. Prove the following:

(a) Define a function $\phi\colon M \to \mathbb{R}^2$ via the formula

$$\phi(x) = (f(x), g(x)) \quad \text{for all } x \in M.$$

Endow $\mathbb{R}^2$ with the Euclidean metric. Prove that ϕ is continuous.
(b) Use part (a) to prove that the sum and product of f and g are continuous. Hint: we already know from real analysis that the functions $s\colon \mathbb{R}^2 \to \mathbb{R}$ and $m\colon \mathbb{R}^2 \to \mathbb{R}$ defined by

$$s(x, y) = x + y \quad \text{and} \quad m(x, y) = xy \quad \text{for all } x, y \in \mathbb{R}$$

are continuous. Use the fact that the sum and product of f and g are $s \circ \phi$ and $m \circ \phi$.
(c) For any $\lambda \in \mathbb{R}$, prove that the constant function $M \to \mathbb{R}$ with value λ is continuous. Use this to prove that λf is continuous. $\square$

EXERCISE 2.11. Let M be a metric space. Prove the following (see Remark 2.4.9 for our conventions about the empty union and the empty intersection):

(a) The empty set $\emptyset$ and the whole set M are both open and closed.
(b) If $\{U_i\}_{i \in I}$ is an arbitrary collection of open subsets of M, then $\cup_{i \in I} U_i$ is open.
(c) If $\{C_i\}_{i \in I}$ is an arbitrary collection of closed subsets of M, then $\cap_{i \in I} C_i$ is closed.
(d) If $\{U_1, \ldots, U_n\}$ is a finite collection of open subsets of M, then $\cap_{i=1}^n U_i$ is open.
(e) If $\{C_1, \ldots, C_n\}$ is a finite collection of closed subsets of M, then $\cup_{i=1}^n C_i$ is closed.

Hint: you can use Lemma 2.8.1 to reduce the assertions about closed sets to the assertions about open sets. $\square$

EXERCISE 2.12. Let $(M, \mathfrak{d})$ be a metric space and let $f\colon M \to \mathbb{R}$ be a continuous function.

(a) For real numbers $a < b$, prove that the following sets are open:

$$\begin{aligned} U &= \{x \in M \mid a < f(x) < b\}, \\ U' &= \{x \in M \mid f(x) < b\}, \\ U'' &= \{x \in M \mid a < f(x)\}. \end{aligned}$$

(b) For real numbers $a \leq b$, prove that the following sets are closed:

$$\begin{aligned} C &= \{x \in M \mid a \leq f(x) \leq b\}, \\ C' &= \{x \in M \mid f(x) \leq b\}, \\ C'' &= \{x \in M \mid a \leq f(x)\}. \end{aligned}$$

(c) For $c \in \mathbb{R}$, prove that $f^{-1}(c)$ is closed.

Hint: this might seem like a lot of things to prove, but using Exercise 2.11 and Lemma 2.8.1 you can show that a small number of these assertions imply all the other ones. $\square$

EXERCISE 2.13. Let $\mathcal{C}(I,\mathbb{R})$ be the set of all continuous functions $f\colon I \to \mathbb{R}$, equipped with the sup metric:

$$\mathfrak{D}(f,g) = \sup\{|f(t)-g(t)| \mid t \in I\} \quad \text{for all continuous } f,g\colon I \to \mathbb{R}.$$

Do the following:

(a) For $s \in I$, define a function $\mathfrak{ev}_s\colon \mathcal{C}(I,\mathbb{R}) \to \mathbb{R}$ via the formula

$$\mathfrak{ev}_s(f) = f(s) \quad \text{for all continuous } f\colon I \to \mathbb{R}.$$

Prove that $\mathfrak{ev}_s$ is continuous.

(b) For $t_1,\ldots,t_n \in I$ and $c_1,\ldots,c_n \in \mathbb{R}$, define

$$C((t_1,c_1),\ldots,(t_n,c_n)) = \{f \in \mathcal{C}(I,\mathbb{R}) \mid f(t_i) = c_i \text{ for all } 1 \le i \le n\}.$$

Prove that $C((t_1,c_1),\ldots,(t_n,c_n))$ is closed. □

EXERCISE 2.14. Let $(M,\mathfrak{d}_M)$ be a metric space and let $(X,\mathfrak{d}_X)$ be a metric subspace of M. Do the following:

(a) Prove that a set $U \subset X$ is open in X if and only if there exists an open set $V \subset M$ such that $U = V \cap X$.

(b) Prove that a set $C \subset X$ is closed in X if and only if there exists a closed set $D \subset M$ such that $C = X \cap D$. □

EXERCISE 2.15. Let $I = [0,1]$ with the Euclidean metric. For each $n \ge 1$, construct a closed set $C \subset I$ with exactly n limit points. □

EXERCISE 2.16. Recall that an open interval in $\mathbb{R}$ is a set of the form (a,b) with $-\infty \le a < b \le \infty$. Let $U \subset \mathbb{R}$ be an open set. Prove that U can be written as the union of countably[4] many *disjoint* open intervals. Hint: First show that U can be written as the union of a collection of disjoint open intervals. The key fact to prove is that for $x,y \in U$, if I_x and I_y are the maximal open intervals in U containing x and y, then either $I_x = I_y$ or $I_x \cap I_y = \emptyset$. Conclude by using the fact that each open interval contains a rational number to prove that such a collection must be countable. □

EXERCISE 2.17*. Let $C \subset \mathbb{R}$ be the classical Cantor set, i.e., the set of all $x \in I = [0,1]$ that can be written as

$$x = \sum_{n=1}^{\infty} \frac{x_n}{3^n} \quad \text{with } x_n \in \{0,2\} \text{ for all } n \ge 1.$$

Prove that C is closed. □

EXERCISE 2.18. Let $\mathcal{C}(\mathbb{R},\mathbb{R})$ be the set of all continuous functions $f\colon \mathbb{R} \to \mathbb{R}$. Define the following metric on $\mathcal{C}(\mathbb{R},\mathbb{R})$:

$$\mathfrak{D}(f,g) = \sup\{\min(|f(t)-g(t)|,1) \mid t \in \mathbb{R}\} \quad \text{for all continuous } f,g\colon \mathbb{R} \to \mathbb{R}.$$

Do the following:

(a) Prove that $\mathfrak{D}$ is a metric.

(b) For $\lambda \in \mathbb{R}$, let $\mathfrak{c}_\lambda\colon \mathbb{R} \to \mathbb{R}$ be the constant function $\mathfrak{c}_\lambda(t) = \lambda$. Using the metric $\mathfrak{D}$ on $\mathcal{C}(\mathbb{R},\mathbb{R})$, prove that $\lim_{n\to\infty} \mathfrak{c}_{1/n} = \mathfrak{c}_0$.

(c) For $\lambda \in \mathbb{R}$, let $f_\lambda\colon \mathbb{R} \to \mathbb{R}$ be the function $f_\lambda(t) = \lambda t$. Using the metric $\mathfrak{D}$ on $\mathcal{C}(\mathbb{R},\mathbb{R})$, prove that $\lim_{n\to\infty} f_{1/n}$ does not exist.

(d)* We would like to have $\lim_{n\to\infty} f_{1/n} = \mathfrak{c}_0$, which requires a more subtle metric. Define the following:

$$\mathfrak{D}'(f,g) = \sum_{k=1}^{\infty} \frac{1}{2^k} \sup\{\min(|f(t)-g(t)|,1) \mid t \in [-k,k]\} \quad \text{for all continuous } f,g\colon \mathbb{R} \to \mathbb{R}.$$

Prove that $\mathfrak{D}'$ is a metric on $\mathcal{C}(\mathbb{R},\mathbb{R})$.

[4] Recall that our convention is that finite sets are countable.

(e)$*$ Using the metric $\mathfrak{D}'$ on $\mathcal{C}(\mathbb{R}, \mathbb{R})$, prove that for a sequence $(f_n)_{n\geq 1}$ of functions $f_n \in \mathcal{C}(\mathbb{R}, \mathbb{R})$ we have $\lim_{n\to\infty} f_n = f$ for a function $f \in \mathcal{C}(\mathbb{R}, \mathbb{R})$ if and only if for all closed intervals $[a, b] \subset \mathbb{R}$ the functions $f_n|_{[a,b]}$ converge to $f|_{[a,b]}$ with respect to the sup metric on $\mathcal{C}([a, b], \mathbb{R})$. In particular, $\lim_{n\to\infty} f_{1/n} = \mathfrak{c}_0$. $\square$

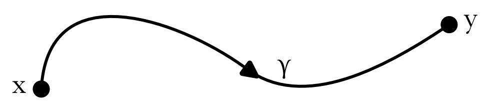
CHAPTER 3

Path connectedness

In this short chapter, we study how to move around continuously in a metric space.

3.1. Paths and path connectedness

A *path* in a metric space M from $x \in M$ to $y \in M$ is a continuous function $\gamma\colon [0,1] \to M$ with $\gamma(0) = x$ and $\gamma(1) = y$:



We say that γ starts at x and ends at y. The metric space M is *path connected* if it is nonempty[1] and for all $x, y \in M$ there exists a path in M from x to y.

Example 3.1.1. The metric space $\mathbb{R}^n$ is path connected. Indeed, for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ the function $\gamma\colon [0,1] \to \mathbb{R}^n$ defined by

$$\gamma(t) = (1-t)\mathbf{x} + t\mathbf{y} \quad \text{for } t \in [0,1]$$

is a path from x to y. This path traces out a straight line segment from $\mathbf{x}$ to $\mathbf{y}$. □

Example 3.1.2. A subspace $A \subset \mathbb{R}^n$ is *convex* if for all $\mathbf{x}, \mathbf{y} \in A$, the straight line segment from x to y lies in A, i.e., for all $t \in [0,1]$ we have $(1-t)\mathbf{x} + t\mathbf{y} \in A$. Just like in the previous example, all nonempty convex subspaces of $\mathbb{R}^n$ are path connected. □

Example 3.1.3. Let $d \geq 1$. Recall that the d-sphere is the subspace

$$\mathbb{S}^d = \left\{ \mathbf{x} \in \mathbb{R}^{d+1} \mid \|\mathbf{x}\| = 1 \right\}$$

of $\mathbb{R}^{d+1}$. Endow $\mathbb{S}^d$ with its Euclidean metric. The metric space $\mathbb{S}^d$ is path connected (see Exercise 3.7). □

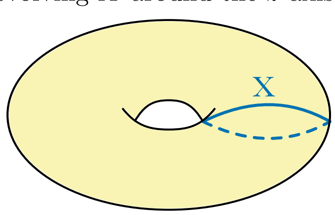
3.2. Images of path connected metric spaces

We next prove that images of path connected metric spaces under continuous functions are path connected, and use this to give more examples of path connected metric spaces.

Lemma 3.2.1. *Let M and N be metric spaces and let $f\colon M \to N$ be a continuous function. Assume that M is path connected. Then the subspace $f(M)$ of N is path connected.*

Proof. Consider points $x, y \in f(M)$. Choose $z, w \in M$ with $f(z) = x$ and $f(w) = y$. Since M is path connected, there exists a path $\gamma\colon [0,1] \to M$ from z to w. The function $f \circ \gamma\colon [0,1] \to f(M)$ is then a path from x to y. □

Example 3.2.2. Let $X \subset \mathbb{R}^3$ be the circle in the x-z plane with center $(2,0,0)$ and radius 1. Let $T \subset \mathbb{R}^3$ be the torus obtained by revolving X around the z-axis:



[1] This is a slightly controversial convention, but it avoids a large number of awkward edge cases.

We can define a surjective continuous function $f\colon \mathbb{R}^2 \to T$ via the formula

$$f(\theta,\phi) = ((2+\cos(\theta))\cos(\phi), (2+\cos(\theta))\sin(\phi), \sin(\theta)) \quad \text{for } \theta,\phi \in \mathbb{R}.$$

Since $\mathbb{R}^2$ is path connected, we can then apply Lemma 3.2.1 to deduce that $f(\mathbb{R}^2) = T$ is path connected. □

3.3. The intermediate value theorem

Our next goal is to give examples of spaces that are not path connected. The key is the following basic result from real analysis:

Theorem 3.3.1 (Intermediate value theorem). *Let $a \leq b$ be real numbers and let $f\colon [a,b] \to \mathbb{R}$ be a continuous function. Assume that $f(a) \leq f(b)$, and let $\lambda \in \mathbb{R}$ satisfy $f(a) \leq \lambda \leq f(b)$. There then exists some $c \in [a,b]$ such that $f(c) = \lambda$.*

Proof. If $f(a) = \lambda$ or $f(b) = \lambda$ this is trivial, so assume that $f(a) < \lambda < f(b)$. Set

$$c = \sup\{x \in [a,b] \mid f(x) < \lambda\} \in [a,b]. \tag{3.3.1}$$

We claim that $f(c) = \lambda$. Indeed, assume for the sake of contradiction that $f(c) \neq \lambda$. There are two cases:

- If $f(c) < \lambda$, then we must have $c < b$. Since f is continuous at c, there exists some $\delta > 0$ such that for $x \in [a,b]$ with $|x-c| < \delta$ we have $|f(x) - f(c)| < |\lambda - f(c)|$. Shrinking δ if necessary, we can assume that $c + \delta \leq b$. For all $x \in [c, c+\delta)$, we then have $f(x) < \lambda$. In particular, $c + \delta/2$ lies in the set in (3.3.1), contradicting the fact that c is the supremum of that set.
- If $f(c) > \lambda$, then we must have $c > a$. Since f is continuous at c, there exists some $\delta > 0$ such that for $x \in [a,b]$ with $|x-c| < \delta$ we have $|f(x) - f(c)| < |\lambda - f(c)|$. Shrinking δ if necessary, we can assume that $c - \delta \geq a$. For all $x \in (c-\delta, c]$, we then have $f(x) > \lambda$. By the definition of c we also have $f(x) \geq \lambda$ for all $x \in [a,b]$ with $x > c$. Thus $c - \delta$ is an upper bound for the set in (3.3.1), contradicting the fact that c is the supremum of that set. □

Here is an example of how this can be used:

Example 3.3.2. Let $x \in \mathbb{R}$ and set $M = \mathbb{R} \setminus x$, endowed with the Euclidean metric. Then there does not exist a path in M from $x-1$ to $x+1$. Indeed, by the intermediate value theorem (Theorem 3.3.1) a path $\gamma\colon [0,1] \to \mathbb{R}$ from $x-1$ to $x+1$ must have $\gamma(c) = x$ for some $c \in [0,1]$. It follows that M is not path connected. □

Remark 3.3.3. Another example that can be handled the same way is that $\mathbb{Q}$ is not path connected (see Exercise 3.11). □

3.4. Paths and isolated points

For a metric space M, recall that a point $x \in M$ is an *isolated point* if there exists some $r > 0$ such that $B_r(x) = \{x\}$. We have:

Lemma 3.4.1. *Let $(M, \mathfrak{d})$ be a metric space, let $x \in M$ be an isolated point, and let $\gamma\colon [0,1] \to M$ be a path starting at x. Then γ is constant.*

Proof. Let $r > 0$ be such that $\mathfrak{d}(y,x) \geq r$ for all $y \in M$ with $y \neq x$. Define $f\colon [0,1] \to \mathbb{R}$ via the formula $f(t) = \mathfrak{d}(x, \gamma(t))$ for $t \in [0,1]$. Recall that the map $M \to \mathbb{R}$ defined by $y \mapsto \mathfrak{d}(x,y)$ is continuous (see Example 2.3.3). Since f is the composition of this with γ, it follows that f is continuous (see Lemma 2.3.4). We have $f(0) = 0$ and $f(t) \geq 0$ for all $t \in [0,1]$, and to prove the lemma we must prove that f is constant.

Assume that f is not constant, so there exists some $0 < s \leq 1$ such that $f(s) > 0$. Choosing some $0 < \lambda < f(s)$ with $\lambda < r$, the intermediate value theorem (Theorem 3.3.1) implies that there exists some $t \in [0,s]$ with $f(t) = \lambda$. We then have $0 < \mathfrak{d}(x, \gamma(t)) = \lambda < r$, a contradiction. □

This has the following consequence:

Corollary 3.4.2. *Let M be a metric space with at least two points. Assume that M has an isolated point. Then M is not path connected.*

Proof. If $x \in M$ is an isolated point and $y \in M$ satisfies $y \neq x$, then Lemma 3.4.1 implies that there is not a path from x to y. □

This corollary implies, for example, that if M is a set with at least two elements and $\mathfrak{d}$ is the discrete metric on M, then $(M, \mathfrak{d})$ is not path connected.

3.5. Operations on paths

The following lemma summarizes some basic properties of paths:

Lemma 3.5.1. *Let M be a metric space. The following hold:*

(i) For all $x \in M$, there exists a path from x to x.
(ii) Let $x, y \in M$ be such that there exists a path from x to y. There then exists a path from y to x.
(iii) Let $x, y, z \in M$ be such that there exists a path from x to y and a path from y to z. There then exists a path from x to z.

Proof. We prove the three parts separately:

Claim. *For all $x \in M$, there exists a path from x to x.*

The desired path is the constant path $\mathfrak{c}_x \colon [0,1] \to M$ defined by $\mathfrak{c}_x(t) = x$ for $t \in [0,1]$.

Claim. *Let $x, y \in M$ be such that there exists a path from x to y. There then exists a path from y to x.*

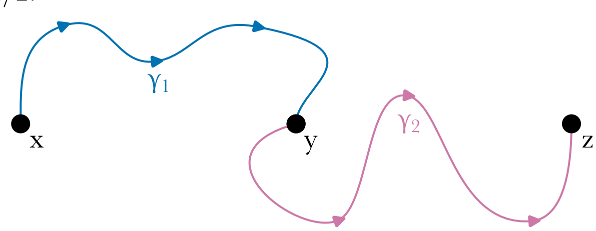
Let $\gamma \colon [0,1] \to M$ be a path from x to y. Let $\overline{\gamma} \colon [0,1] \to M$ be the path defined by $\overline{\gamma}(t) = \gamma(1-t)$ for $t \in [0,1]$. This path traverses γ in reverse order, and is a path from y to x.

Claim. *Let $x, y, z \in M$ be such that there exists a path from x to y and a path from y to z. There then exists a path from x to z.*

Let $\gamma_1 \colon [0,1] \to M$ be a path from x to y and let $\gamma_2 \colon [0,1] \to M$ be a path from y to z. Let $\gamma_1 \cdot \gamma_2 \colon [0,1] \to M$ be the path defined by

$$(\gamma_1 \cdot \gamma_2)(t) = \begin{cases} \gamma_1(2t) & \text{for } t \in [0,1/2], \\ \gamma_2(2t-1) & \text{for } t \in [1/2,1]. \end{cases}$$

This path traverses γ_1 at double speed and then traverses γ_2 at double speed. Since $\gamma_1(1) = y = \gamma_2(0)$, it is continuous at $t = 1/2$:



The path $\gamma_1 \cdot \gamma_2 \colon [0,1] \to M$ is thus a path from x to z. □

Remark 3.5.2. In Volume 2, we will use operations like those in the proof of Lemma 3.5.1 to extract a group called the fundamental group from the collection of paths in M. □

3.6. Path components

We will now need the notion of an equivalence relation. A reader who has not seen this should consult Appendix 3.I. Let M be a metric space. Let $\sim$ be the following relation on the points of M: for $x, y \in M$, we have $x \sim y$ if there is a path in M from x to y. Lemma 3.5.1 says that this is an equivalence relation. Its equivalence classes are the *path components* of M. Two points $x, y \in M$ are thus in the same path component exactly when there is a path in M from x to y.

The following summarizes the main properties of path components:

THEOREM 3.6.1. *Let M be a metric space and let $\{P_j\}_{j\in J}$ be the path components of M. The following hold:*

(i) The P_j are pairwise disjoint and $M = \sqcup_{j\in J} P_j$.
(ii) Each P_j is path connected.
(iii) If $N \subset M$ is a path connected subspace, then there exists a unique $j \in J$ such that $N \subset P_j$.

PROOF. Conclusion (i) is true for the equivalence classes of any equivalence relation. We prove conclusions (ii) and (iii) separately:

CLAIM. *For $j \in J$, the subspace P_j is path connected.*

Consider $x, y \in P_j$. Let $\gamma\colon [0,1] \to M$ be a path from x to y. To prove the claim, it is enough to prove that the image of γ lies in P_j. In other words, for a fixed $s \in [0,1]$ we must prove that $x \sim \gamma(s)$. The required path $\gamma_s\colon [0,1] \to M$ from $x = \gamma(0)$ to $\gamma(s)$ is given by the following formula:

$$\gamma_s(t) = \gamma(st) \quad \text{for } t \in [0,1].$$

CLAIM. *If $N \subset M$ is a path connected subspace, then there exists a unique $j \in J$ such that $N \subset P_j$.*

For $x, y \in N$ we have $x \sim y$. This implies that all points of N lie in the same equivalence class, i.e., there exists a unique $j \in J$ such that $N \subset P_j$. ☐

Another important observation is that continuous functions take path components to path components:

LEMMA 3.6.2. *Let M and N be metric spaces and let $f\colon M \to N$ be a continuous function. For all path components P of M, there exists a unique path component Q of N such that $f(P) \subset Q$.*

PROOF. Immediate from the fact that $f(P)$ is path connected (Lemma 3.2.1) together with conclusion (iii) of Theorem 3.6.1. ☐

Here are some examples:

EXAMPLE 3.6.3. If M is a path connected metric space, then the only path component of M is M itself. ☐

EXAMPLE 3.6.4. Let $M = \mathbb{R} \setminus 0$. The path components of M are $(-\infty, 0)$ and $(0, \infty)$. ☐

EXAMPLE 3.6.5. Let $M = \mathbb{Q}$. The path components of M are the one-point sets $\{q\}$ with $q \in \mathbb{Q}$ (see Exercise 3.11). ☐

3.7. Exercises

EXERCISE 3.1. A subspace $A \subset \mathbb{R}^d$ is *star-shaped* if there exists some $\mathbf{a}_0 \in A$ such that for all $\mathbf{x} \in A$, the straight line segment from $\mathbf{a}_0$ to $\mathbf{x}$ is contained in A. Do the following:

(a) Give an example of a star-shaped subspace of $\mathbb{R}^2$ that is not convex.
(b) Prove that all star-shaped subspaces of $\mathbb{R}^d$ are path connected. ☐

EXERCISE 3.2. Fix $n \geq 1$, and consider real numbers $a_1 < \cdots < a_n$. Determine the path components of $\mathbb{R} \setminus \{a_1, \ldots, a_n\}$. ☐

EXERCISE 3.3. Let M and N be metric spaces. Give $M \times N$ the product metric. Do the following:

(a) Assume that M and N are path connected. Prove that $M \times N$ is path connected.
(b) Prove that the path components of $M \times N$ are exactly the subspaces of the form $P \times Q$ with P a path component of M and Q a path component of N. ☐

EXERCISE 3.4. Let M be a metric space. Let $\{A_j\}_{j\in J}$ be a nonempty collection of path connected subspaces of M. Assume that $\cap_{j\in J} A_j \neq \emptyset$. Prove that $\cup_{j\in J} A_j$ is path connected. ☐

EXERCISE 3.5. Let $I = [0,1]$. As in Example 2.1.10, let $\mathcal{C}(I,\mathbb{R})$ be the set of all continuous functions $f: I \to \mathbb{R}$ and let $\mathfrak{D}$ be the sup metric on $\mathcal{C}(I,\mathbb{R})$:

$$\mathfrak{D}(f,g) = \sup\{|f(t) - g(t)| \mid t \in I\} \quad \text{for all continuous } f,g: I \to \mathbb{R}.$$

Prove that $\mathcal{C}(I,\mathbb{R})$ is path connected. □

EXERCISE 3.6. Fix some $d \geq 2$. Do the following:

(a) Let $S \subset \mathbb{R}^d$ be a finite set. Prove that $\mathbb{R}^d \setminus S$ is path connected.
(b) Prove that $\mathbb{R}^d \setminus \mathbb{Q}^d$ is path connected. □

EXERCISE 3.7. Let $d \geq 1$. Recall that the d-sphere is the subspace

$$\mathbb{S}^d = \left\{\mathbf{x} \in \mathbb{R}^{d+1} \mid \|\mathbf{x}\| = 1\right\}$$

of $\mathbb{R}^{d+1}$. Endow $\mathbb{S}^d$ with its Euclidean metric. Prove that $\mathbb{S}^d$ is path connected. Hint: We know that $\mathbb{R}^{d+1} \setminus 0$ is path connected (see Exercise 3.6). Construct a continuous surjection $f: \mathbb{R}^{d+1} \setminus 0 \to \mathbb{S}^d$ and use it to prove that $\mathbb{S}^d$ is path connected. □

EXERCISE 3.8. An *interval* in $\mathbb{R}$ is a set of one of the following forms:

- (a,b) for some $-\infty \leq a < b \leq \infty$; or
- $[a,b)$ for some $-\infty < a < b \leq \infty$; or
- $(a,b]$ for some $-\infty \leq a < b < \infty$; or
- $[a,b]$ for some $-\infty < a \leq b < \infty$.

Prove that the only path connected subspaces of $\mathbb{R}$ are the intervals. □

EXERCISE 3.9. For $d \geq 1$, let

$$\mathrm{PConf}_2(\mathbb{R}^d) = \left\{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d \mid x \neq y\right\}.$$

Give $\mathbb{R}^d \times \mathbb{R}^d$ the product metric, and endow $\mathrm{PConf}_2(\mathbb{R}^d)$ with the subspace metric. The metric space $\mathrm{PConf}_2(\mathbb{R}^d)$ is called the *pure*[2] *configuration space* of 2 points in $\mathbb{R}^d$. We will encounter this space and its generalizations many times in later volumes. Do the following:

(a) Prove that $\mathrm{PConf}_2(\mathbb{R})$ has exactly two path components.
(b) Prove that $\mathrm{PConf}_2(\mathbb{R}^d)$ is path connected for $d \geq 2$. □

EXERCISE 3.10. Fix some $d \geq 1$. Let $V \subset \mathbb{R}^d$ be a k-dimensional linear subspace for some $0 \leq k \leq d-1$. Do the following:

(a) If $d - k = 1$, then prove that $\mathbb{R}^d \setminus V$ has exactly two path components. Hint: Start by choosing a linear map $\phi: \mathbb{R}^d \to \mathbb{R}$ with $\ker(\phi) = V$.
(b)∗ If $d - k \geq 2$, then prove that $\mathbb{R}^d \setminus V$ is path connected. □

EXERCISE 3.11. Do the following:

(a) Prove that any path in $\mathbb{Q}$ must be constant. Deduce that $\mathbb{Q}$ is not path connected and that the path components of $\mathbb{Q}$ are the one-point sets $\{q\}$ for $q \in \mathbb{Q}$.
(b) Prove that $\mathbb{Q}^d$ is not path connected for any $d \geq 1$. Hint: reduce this to (a) by projecting $\mathbb{Q}^d$ surjectively to $\mathbb{Q}$. □

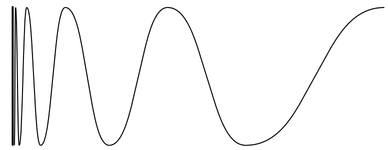
EXERCISE 3.12. As in Example 2.1.8, let S be a set, let $\mathrm{Seq}(S)$ be the set of sequences $\mathbf{s} = (s_m)_{m\geq 1}$ of elements of S, and let $\mathfrak{d}$ be the first-difference metric on $\mathrm{Seq}(S)$. Assume that S has at least two elements. Prove that $\mathrm{Seq}(S)$ is not path connected. Hint: Give S the discrete metric, so S is not path connected. Construct a continuous surjection $f: \mathrm{Seq}(S) \to S$. □

EXERCISE 3.13. Generalize Exercise 3.12 as follows. Let S be a set, let $\mathrm{Seq}(S)$ be the set of sequences $\mathbf{s} = (s_m)_{m\geq 1}$ of elements of S, and let $\mathfrak{d}$ be the first-difference metric on $\mathrm{Seq}(S)$. Let $\gamma: [0,1] \to \mathrm{Seq}(S)$ be a path. Prove that γ is constant. □

---

[2]Or ordered.

EXERCISE 3.14**. Let $X = \{(0, y) \mid -1 \leq y \leq 1\} \cup \{(x, \sin(1/x)) \mid x > 0\} \subset \mathbb{R}^2$.



The space X is often called the topologist's sine curve. Prove that X is not path connected and that its path components are

$$\begin{aligned} X_1 &= \{(0, y) \mid -1 \leq y \leq 1\}, \\ X_2 &= \{(x, \sin(1/x)) \mid x > 0\}. \end{aligned}$$

□

APPENDIX 3.I

Relations and equivalence relations

This appendix introduces relations and equivalence relations, which are needed for our discussion of path components. They will also show up later in this book. For instance, we will use them to construct completions of metric spaces.

3.I.1. Relations

We start with the following very general definition:

DEFINITION 3.I.1.1. Let S be a set. A *relation* on S is a subset $R \subset S \times S$. For $s, t \in S$, we say that s is related to t by R if $(s, t) \in R$, and we indicate this by writing sRt. □

EXAMPLE 3.I.1.2. On $\mathbb{Z}$, let $R = \{(n, m) \in \mathbb{Z}^2 \mid n \leq m\}$. For $n, m \in \mathbb{Z}$, we then have nRm precisely when $n \leq m$. □

EXAMPLE 3.I.1.3. Fix an integer $d \geq 2$. On $\mathbb{Z}$, let $R = \{(n, m) \in \mathbb{Z}^2 \mid n - m \text{ is divisible by } d\}$. For $n, m \in \mathbb{Z}$, we then have nRm precisely when $n \equiv m \pmod{d}$. □

3.I.2. Equivalence relations

We now focus on the following special kind of relation:

DEFINITION 3.I.2.1. A relation R on a set S is an *equivalence relation* if it satisfies the following three axioms:

- **Reflexive**. For all $s \in S$, we have sRs.
- **Symmetry**. For all $s, t \in S$ with sRt, we have tRs.
- **Transitivity**. For all $s, t, u \in S$ with sRt and tRu, we have sRu. □

EXAMPLE 3.I.2.2. The relation R on $\mathbb{Z}$ where nRm precisely when $n \leq m$ is not an equivalence relation since it is not symmetric. □

EXAMPLE 3.I.2.3. For an integer $d \geq 2$, the relation R on $\mathbb{Z}$ where nRm precisely when $n \equiv m \pmod{d}$ is an equivalence relation. □

REMARK 3.I.2.4. We will meet other kinds of relations later in this book; for instance, the relation R on $\mathbb{Z}$ where nRm precisely when $n \leq m$ will be an example of a total ordering. □

NOTATION 3.I.2.5. Rather than R, it is traditional to write equivalence relations using the symbol $\sim$, so for an equivalence relation on a set S we write $s \sim t$ to indicate that s is related to t. □

3.I.3. Equivalence classes

Associated to an equivalence relation are the following subsets of a set:

DEFINITION 3.I.3.1. If S is a set equipped with an equivalence relation $\sim$, then for $s \in S$ the *equivalence class* of s is the following subset of S:

$$[s] = \{t \in S \mid s \sim t\}.$$

Sets of the form $[s]$ for some $s \in S$ are called the *equivalence classes* of S. □

The following summarizes the key properties of equivalence classes:

LEMMA 3.I.3.2. *Let S be a set equipped with an equivalence relation $\sim$. The following then hold:*

(i) For all $s \in S$, we have $s \in [s]$.

(ii) *For all $s, t \in S$, we either have $[s] = [t]$ or $[s] \cap [t] = \emptyset$.*
(iii) *The set S is the disjoint union of its equivalence classes.*

PROOF. Item (iii) follows from (i) and (ii). For (i), since $\sim$ is reflexive we have $s \sim s$, so $s \in [s]$. For (ii), assume that $[s] \cap [t] \neq \emptyset$. We must prove that $[s] = [t]$. Since s and t play symmetric roles here, it is enough to prove that $[s] \subset [t]$ since this will also imply that $[t] \subset [s]$ and hence that $[s] = [t]$.

For this, consider $u \in [s]$, so $s \sim u$. We must prove that $u \in [t]$. By assumption there exists some $v \in [s] \cap [t]$, so $s \sim v$ and $t \sim v$. Since $s \sim u$ and $\sim$ is symmetric, we have $u \sim s$. Since $u \sim s$ and $s \sim v$, we can apply transitivity to see that $u \sim v$. Since $\sim$ is symmetric, this implies that $v \sim u$. Since $t \sim v$ and $v \sim u$, transitivity implies that $t \sim u$. This implies that $u \in [t]$, as desired. □

If S is a set equipped with an equivalence relation $\sim$, then the set of equivalence classes of S is called the *quotient* of S by $\sim$ and is denoted by $S/\sim$. The map $p\colon S \to S/\sim$ defined by $p(s) = [s]$ is called the *projection* to the quotient. To explain this terminology, consider the following example:

EXAMPLE 3.I.3.3. Let V be a vector space over a field and let $W < V$ be a subspace. Define an equivalence relation $\sim$ on V as follows: for $v, v' \in V$, write $v \sim v'$ if $v - v' \in W$. The equivalence classes of this equivalence relation are then the cosets of W in V:

$$[v] = v + W \quad \text{for } v \in V.$$

It follows that $V/\sim$ can be canonically identified with the quotient vector space V/W. □

Another important example is as follows:

EXAMPLE 3.I.3.4. Let S and T be sets and let $f\colon S \to T$ be a surjection. Define an equivalence relation $\sim$ on S as follows: for $s, s' \in S$, write $s \sim s'$ if $f(s) = f(s')$. The equivalence classes of this equivalence relation are the fibers of f:

- For $s \in S$ with $t = f(s)$, we have $[s] = f^{-1}(t)$.

Let $p\colon S \to S/\sim$ be the projection. The map $f\colon S \to T$ factors as $f = \phi \circ p$, where $\phi\colon S/\sim \to T$ is the following bijection: for an equivalence class $C \in S/\sim$, we define $\phi(C) \in T$ to be the common value of $f(s)$ as s ranges over elements of C. The map ϕ identifies $S/\sim$ with T. □

3.I.4. Exercises

EXERCISE 3.I.1. Let $S = \{1, 2, 3\}$. Do the following:

(a) Construct a relation on S that is symmetric and reflexive, but not transitive.
(b) Construct a relation on S that is symmetric and transitive, but not reflexive.
(c) Construct a relation on S that is reflexive and transitive, but not symmetric. □

EXERCISE 3.I.2. For each of the following relations, determine whether or not it is an equivalence relation. If it is an equivalence relation, describe its equivalence classes. If it is not an equivalence relation, identify the axiom(s) of an equivalence relation that do not hold.

(a) On $\mathbb{Z}$, let $m \sim n$ if $m + n$ is even.
(b) On $\mathbb{R} \setminus 0$, let $x \sim y$ if $xy > 0$.
(c) On $\mathbb{R}$, let $x \sim y$ if $x - y \in \mathbb{Z}$.
(d) On $\mathbb{R}$, let $x \sim y$ if $|x - y| < 1$. □

EXERCISE 3.I.3. Let X be a graph with vertices $\mathcal{V}(X)$ and edges $\mathcal{E}(X)$ (see Example 2.1.7). Define a relation $\sim$ on $\mathcal{V}(X)$ by letting $v \sim w$ if there exists an edge path from v to w. Do the following:

(a) Prove that $\sim$ is an equivalence relation and describe its equivalence classes.
(b) Prove that X is connected if and only if $\mathcal{V}(X)/\sim$ is a one-element set. □

EXERCISE 3.I.4. Let S be a set. A *partition* of S is a collection $\mathcal{P}$ of nonempty subsets of S such that for all $s \in S$, there exists a unique $P \in \mathcal{P}$ with $s \in P$. Let $\mathcal{P}$ be a partition of S. Define a relation $\sim_{\mathcal{P}}$ on S by letting $s \sim t$ if there exists some $P \in \mathcal{P}$ with $s, t \in P$. Do the following:

(a) Prove that $\sim_{\mathcal{P}}$ is an equivalence relation on S.
(b) Describe the equivalence classes of S with respect to $\sim_{\mathcal{P}}$.

(c) Let $\sim$ be an equivalence relation on S. It follows from Lemma 3.I.3.2 that the set of equivalence classes of S forms a partition $\mathcal{P}$ of S. Prove that $\sim_{\mathcal{P}}=\sim$.
(d) Conclude that there is a bijection between equivalence relations on S and partitions of S. $\square$

EXERCISE 3.I.5. Let S be a set equipped with an equivalence relation $\sim$ and let $p\colon S \to S/\sim$ be the projection. Let T be another set and let $f\colon S \to T$ be a function such that for $s, t \in S$ with $s \sim t$ we have $f(s) = f(t)$. Do the following:

(a) Prove that there exists a unique map $\overline{f}\colon S/\sim \to T$ such that $f = \overline{f} \circ p$.
(b) Prove that $\overline{f}$ is surjective if and only if f is surjective.
(c) Prove that $\overline{f}$ is injective if and only if for all $s, t \in S$ with $f(s) = f(t)$ we have $s \sim t$. $\square$

EXERCISE 3.I.6. Let S be a set and let R be a relation on S. Define a new relation $\sim$ on S as follows:

- For $s, s' \in S$, we have $s \sim s'$ if for some $n \geq 0$ there exists a sequence $s = s_0, s_1, \ldots, s_n = s'$ of elements of S such that for all $1 \leq i \leq n$, we either have $s_{i-1} R s_i$ or $s_i R s_{i-1}$.

Prove that $\sim$ is an equivalence relation. It is called the equivalence relation *generated* by the relation R. $\square$

EXERCISE 3.I.7*. For integers $1 \leq k \leq n$, let $a_{n,k}$ be the number of equivalence relations on $\{1, \ldots, n\}$ having exactly k equivalence classes. Do the following:

(a) For $n \geq 1$, prove that $a_{n,1} = a_{n,n} = 1$.
(b) For $2 \leq k \leq n-1$, prove that $a_{n,k} = a_{n-1,k-1} + k a_{n-1,k}$.
(c) Using this recurrence, calculate the number of equivalence relations on $\{1, \ldots, n\}$ for $1 \leq n \leq 4$. $\square$

CHAPTER 4

Morphisms between metric spaces

We now study several families of functions between metric spaces.

4.1. Philosophical remarks

To understand a class of mathematical objects, it is important to understand not just the internal structure of a single object, but also the structure of maps between different objects that preserve that internal structure. These maps are often called morphisms.

EXAMPLE 4.1.1. In linear algebra, the objects are vector spaces over some fixed field like $\mathbb{C}$. The morphisms between two vector spaces V and W are the linear maps $f: V \to W$. In many ways, these linear maps are more important than individual vector spaces. $\square$

The most important morphisms are the isomorphisms, which are invertible and allow us to identify two different objects. If there exists an isomorphism from one object to another, then we say that those objects are isomorphic.

EXAMPLE 4.1.2. If V and W are vector spaces, then an isomorphism from V to W is a linear map $f: V \to W$ that is invertible in the sense that there exists a linear map $g: W \to V$ with $g \circ f = \mathbb{1}_V$ and $f \circ g = \mathbb{1}_W$. Here $\mathbb{1}_V: V \to V$ and $\mathbb{1}_W: W \to W$ are the identity maps. One of the key theorems of linear algebra is that an isomorphism $f: V \to W$ exists if and only if V and W have the same dimension.[1] In particular, if we are working over the field $\mathbb{C}$ then for an integer $d \geq 0$ a vector space is isomorphic to $\mathbb{C}^d$ if and only if it has dimension d. $\square$

As we will see in this chapter, one thing that makes metric spaces a little subtle is that the metric encodes several different kinds of structure on a space, so there are several different classes of morphisms between metric spaces. Consequently, there are several different notions of isomorphism between metric spaces. We will discuss three different kinds of isomorphisms between metric spaces: isometries, homeomorphisms, and bi-Lipschitz equivalences.

REMARK 4.1.3. We could have introduced the language of category theory in this chapter and defined several different categories of metric spaces. We chose not to do this since at this point it would not really add anything to the discussion. $\square$

4.2. Isometric embeddings and isometries

An isometric embedding is a function between metric spaces that preserves the metric:

DEFINITION 4.2.1. Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces. An *isometric embedding* is a function $f: M \to N$ such that $\mathfrak{d}_N(f(x), f(y)) = \mathfrak{d}_M(x, y)$ for all $x, y \in M$. $\square$

EXAMPLE 4.2.2. Let $(M, \mathfrak{d}_M)$ be a metric space and let $A \subset M$ be a metric subspace, so the metric $\mathfrak{d}_A$ on A is $\mathfrak{d}_A(x, y) = \mathfrak{d}_M(x, y)$ for all $x, y \in A$. The inclusion $\iota: A \to M$ is then an isometric embedding. $\square$

Isometric embeddings are continuous and injective (see Exercise 4.3). We now introduce a notion of isomorphism based on isometric embeddings. For a metric space M, let $\mathbb{1}_M: M \to M$ be the identity function.

[1] This is true even if this dimension is infinite, in which case it should be regarded as a cardinal number.

DEFINITION 4.2.3. Let M and N be metric spaces. An *isometry* from M to N is an isometric embedding $f\colon M \to N$ that is a bijection. Equivalently, an isometry is an isometric embedding $f\colon M \to N$ that has a two-sided inverse $g\colon N \to M$, so $g \circ f = \mathbb{1}_M$ and $f \circ g = \mathbb{1}_N$. The function g is called the *inverse* of f, and is necessarily also an isometry. If there exists an isometry from M to N, then we say that M and N are *isometric*. $\square$

On any set $\mathcal{M}$ of metric spaces,[2] being isometric is an equivalence relation (see Exercise 4.5). To give examples of isometries, one natural place to look is at the isometries of a metric space to itself:

EXAMPLE 4.2.4. Give $\mathbb{R}$ its Euclidean metric. The following are both isometries from $\mathbb{R}$ to $\mathbb{R}$:

- For $a \in \mathbb{R}$, the translation function $\tau_a\colon \mathbb{R} \to \mathbb{R}$ defined by $\tau_a(x) = x + a$ for all $x \in \mathbb{R}$.
- For $a \in \mathbb{R}$, the reflection function $r_a\colon \mathbb{R} \to \mathbb{R}$ defined by $r_a(x) = a - (x - a) = 2a - x$ for $x \in \mathbb{R}$.

Every isometry from $\mathbb{R}$ to $\mathbb{R}$ is of one of these two forms (see Exercise 4.1). $\square$

EXAMPLE 4.2.5. Let S be a set and let $\mathrm{Seq}(S)$ be the set of all sequences $\mathbf{x} = (x_m)_{m\geq 1}$ of elements of S. Give $\mathrm{Seq}(S)$ the first-difference metric (see Example 2.1.8). Consider a sequence $\boldsymbol{\sigma} = (\sigma_m)_{m\geq 1}$ with $\sigma_m\colon S \to S$ a bijection for all $m \geq 1$. We can then define an isometry $\phi_{\boldsymbol{\sigma}}\colon \mathrm{Seq}(S) \to \mathrm{Seq}(S)$ via the formula

$$\phi_{\boldsymbol{\sigma}}((x_m)_{m\geq 1}) = (\sigma_m(x_m))_{m\geq 1} \quad \text{for all } (x_m)_{m\geq 1} \text{ in } \mathrm{Seq}(S).$$

If $|S| \geq 2$, then this gives uncountably many different isometries $\mathrm{Seq}(S) \to \mathrm{Seq}(S)$. However, there are many other isometries from $\mathrm{Seq}(S)$ to itself that are not of this form (see Exercise 4.1). $\square$

Isometric metric spaces share all properties that can be defined in terms of the metric. Here is one important example:

EXAMPLE 4.2.6. Let $(M, \mathfrak{d})$ be a metric space and let $A \subset M$ be a subspace. The *diameter* of A is

$$\mathrm{diam}(A) = \sup\left(\{\mathfrak{d}(x, y) \mid x, y \in A\} \cup \{0\}\right).$$

This might be infinity. It is immediate that if M is isometric to N, then $\mathrm{diam}(M) = \mathrm{diam}(N)$. $\square$

4.3. Continuous functions and homeomorphisms

We have already introduced the class of continuous functions between metric spaces. They are far easier to construct than isometric embeddings, and lead to the following notion of isomorphism:

DEFINITION 4.3.1. Let M and N be metric spaces. A *homeomorphism* from M to N is a continuous function $f\colon M \to N$ such that there exists a continuous function $g\colon N \to M$ with $g \circ f = \mathbb{1}_M$ and $f \circ g = \mathbb{1}_N$. The function g is called the *inverse* of f, and is also a homeomorphism. If there exists a homeomorphism from M to N, then M and N are said to be *homeomorphic*. We will write $M \cong N$ to indicate that M is homeomorphic to N. $\square$

On any set $\mathcal{M}$ of metric spaces, being homeomorphic is an equivalence relation (see Exercise 4.5). This is the most important notion introduced in this chapter. One of the main tasks of topology is to classify all metric spaces up to homeomorphism. Here are some examples:

EXAMPLE 4.3.2. Fix some $d \geq 1$. For $r > 0$, let $B_r = B_r(\mathbf{0})$ be the open ball of radius r in $\mathbb{R}^d$ around the origin $\mathbf{0} \in \mathbb{R}^d$. Endow B_r with the Euclidean metric. For $r, s > 0$, we can define a homeomorphism $f_{rs}\colon B_r \to B_s$ via the formula

$$f_{rs}(\mathbf{x}) = \frac{s}{r}\mathbf{x} \quad \text{for all } \mathbf{x} \in B_r.$$

The inverse of $f_{rs}\colon B_r \to B_s$ is the function $f_{sr}\colon B_s \to B_r$. The metric spaces B_r and B_s are thus homeomorphic, so $B_r \cong B_s$. Since $\mathrm{diam}(B_r) = 2r$ and $\mathrm{diam}(B_s) = 2s$, they are not isometric if $r \neq s$. $\square$

---

[2]There is no set of all metric spaces for the same reason that there is no set of all sets: it would lead to a version of Russell's paradox.

Example 4.3.3. Fix some $d \geq 1$ and let $B = B_1(\mathbf{0})$ be the open unit ball in $\mathbb{R}^d$. Endow B with the Euclidean metric. We can define a homeomorphism $f\colon B \to \mathbb{R}^d$ via the formula

$$f(\mathbf{x}) = \frac{\mathbf{x}}{1 - \|\mathbf{x}\|} \quad \text{for all } \mathbf{x} \in B.$$

Indeed, the inverse of $f\colon B \to \mathbb{R}^d$ is the function $g\colon \mathbb{R}^d \to B$ given by the formula

$$g(\mathbf{y}) = \frac{\mathbf{y}}{1 + \|\mathbf{y}\|} \quad \text{for all } \mathbf{y} \in \mathbb{R}^d.$$

This lands in B since $\|\mathbf{y}\|/(1 + \|\mathbf{y}\|) < 1$. To see that these are inverses to each other, observe that for $\mathbf{x} \in B$ and $\mathbf{y} \in \mathbb{R}^d$ we have

$$\begin{aligned} g(f(\mathbf{x})) &= \frac{1}{1 + \|f(\mathbf{x})\|} f(\mathbf{x}) = \left(\frac{1}{1 + \frac{\|\mathbf{x}\|}{1 - \|\mathbf{x}\|}} \right) \frac{\mathbf{x}}{1 - \|\mathbf{x}\|} = \left(\frac{1 - \|\mathbf{x}\|}{1} \right) \frac{\mathbf{x}}{1 - \|\mathbf{x}\|} = \mathbf{x}, \\ f(g(\mathbf{y})) &= \frac{1}{1 - \|g(\mathbf{y})\|} g(\mathbf{y}) = \left(\frac{1}{1 - \frac{\|\mathbf{y}\|}{1 + \|\mathbf{y}\|}} \right) \frac{\mathbf{y}}{1 + \|\mathbf{y}\|} = \left(\frac{1 + \|\mathbf{y}\|}{1} \right) \frac{\mathbf{y}}{1 + \|\mathbf{y}\|} = \mathbf{y}. \end{aligned}$$

We therefore have $B \cong \mathbb{R}^d$. This is even more striking than the previous example since $\operatorname{diam}(B) = 2$ but $\operatorname{diam}(\mathbb{R}^d) = \infty$. □

Example 4.3.4. For $d \geq 2$, the metric spaces $\mathbb{R}^d$ and $\mathbb{R}^1$ are not homeomorphic. To prove this, we must identify a property of metric spaces that is preserved by homeomorphisms and holds for $\mathbb{R}^d$ but not for $\mathbb{R}^1$. One property of metric spaces that is preserved by homeomorphisms is being path connected (see Exercise 4.10). Unfortunately, both $\mathbb{R}^d$ and $\mathbb{R}^1$ are path connected.

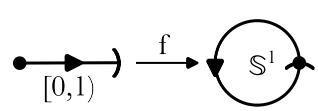
However, we can still use the notion of path connectedness to prove the result. Assume for the sake of contradiction that there exists a homeomorphism $f\colon \mathbb{R}^d \to \mathbb{R}^1$. Pick $x \in \mathbb{R}^d$, and let $y = f(x) \in \mathbb{R}^1$. The homeomorphism f restricts to a homeomorphism $f'\colon \mathbb{R}^d \setminus x \to \mathbb{R}^1 \setminus y$. Since $d \geq 2$, the metric space $\mathbb{R}^d \setminus x$ is path connected (see Exercise 3.6). Since $\mathbb{R}^d \setminus x$ is homeomorphic to $\mathbb{R}^1 \setminus y$, it follows that $\mathbb{R}^1 \setminus y$ is path connected (see Exercise 4.10). However, $\mathbb{R}^1 \setminus y$ is not path connected (see Example 3.3.2), a contradiction. □

Remark 4.3.5. In fact, for $d, e \geq 1$ with $d \neq e$ the metric spaces $\mathbb{R}^d$ and $\mathbb{R}^e$ are not homeomorphic. For $d, e \geq 2$ this cannot be proved in an elementary way. We will prove it using dimension theory in Essay D. □

Example 4.3.6. We warn the reader that metric spaces that look different can sometimes be homeomorphic. For example, it turns out that for all $d, e \geq 1$ the metric spaces $\mathbb{Q}^d$ and $\mathbb{Q}^e$ are homeomorphic. This is not easy to prove – both are countable, but most easily written down bijections between them are not continuous. We will prove a generalization of the fact that $\mathbb{Q}^d \cong \mathbb{Q}^e$ due to Sierpiński in Appendix 4.I*. □

The next example shows that not all continuous bijections are homeomorphisms:

Example 4.3.7. Identify $\mathbb{R}^2$ with $\mathbb{C}$, and regard $\mathbb{S}^1$ as a subspace of $\mathbb{C}$. Define a function $f\colon [0,1) \to \mathbb{S}^1$ via the formula $f(t) = e^{2\pi i t}$ for $t \in [0,1)$. The function f wraps $[0,1)$ once around $\mathbb{S}^1$ in the counterclockwise direction:



The function f is a continuous bijection. However, its inverse $g\colon \mathbb{S}^1 \to [0,1)$ is not continuous (see Exercise 4.8), so f is not a homeomorphism. □

Remark 4.3.8. In fact, $[0,1)$ is not homeomorphic to $\mathbb{S}^1$ (see Exercise 4.8). □

The following is an interesting and non-obvious example of a homeomorphism.

EXAMPLE 4.3.9. As in Example 2.1.8, let $\mathrm{Seq}(\{0,1\})$ be the set of sequences $\mathbf{s} = (s_m)_{m\geq 1}$ of elements of $\{0,1\}$ and let $\mathfrak{d}$ be the first-difference metric on $\mathrm{Seq}(\{0,1\})$. Finally, let C be the classical Cantor set, i.e., the set of all $x \in I = [0,1]$ that can be written as

$$x = \sum_{m=1}^{\infty} \frac{x_m}{3^m} \quad \text{with } x_m \in \{0,2\} \text{ for all } m \geq 1.$$

Define a function $f\colon \mathrm{Seq}(\{0,1\}) \to C$ as follows:

$$f((s_m)_{m\geq 1}) = \sum_{m=1}^{\infty} \frac{2s_m}{3^m} \quad \text{for } (s_m)_{m\geq 1} \in \mathrm{Seq}(\{0,1\}).$$

Then f is a homeomorphism (see Exercise 4.18). □

4.4. Equivalent metrics

Let M be a set. Two metrics $\mathfrak{d}$ and $\mathfrak{d}'$ on M are *equivalent* if the identity function $\mathbb{1}\colon M \to M$ is a homeomorphism from $(M,\mathfrak{d})$ to $(M,\mathfrak{d}')$. This can be reformulated in terms of balls as follows. For $x_0 \in M$ and $r > 0$ and a metric $\mathfrak{d}$ on M, let $B_r(x_0,\mathfrak{d})$ be the open ball around x_0 of radius r with respect to the metric $\mathfrak{d}$. Then:

LEMMA 4.4.1. *Let M be a set and let $\mathfrak{d}$ and $\mathfrak{d}'$ be metrics on M. Then $\mathfrak{d}$ is equivalent to $\mathfrak{d}'$ if and only if the following two things hold for all $x_0 \in M$:*

(i) for all $r > 0$, there exists some $r' > 0$ such that $B_{r'}(x_0,\mathfrak{d}') \subset B_r(x_0,\mathfrak{d})$; and
(ii) for all $r' > 0$, there exists some $r > 0$ such that $B_r(x_0,\mathfrak{d}) \subset B_{r'}(x_0,\mathfrak{d}')$.

PROOF. Translating from the ϵ-δ definition of continuity to the language of open balls, we see that condition (i) is equivalent to $\mathbb{1}\colon M \to M$ being a continuous function from $(M,\mathfrak{d}')$ to $(M,\mathfrak{d})$ at x_0 and condition (ii) is equivalent to $\mathbb{1}\colon M \to M$ being a continuous function from $(M,\mathfrak{d})$ to $(M,\mathfrak{d}')$ at x_0. The lemma follows. □

EXAMPLE 4.4.2. Let $(M,\mathfrak{d})$ be a metric space and let $\lambda > 0$. Let $\lambda\,\mathfrak{d}$ be the metric on M obtained by multiplying $\mathfrak{d}$ by λ. Then $\mathfrak{d}$ is equivalent to $\lambda\,\mathfrak{d}$. Indeed, for all $x_0 \in M$ and $r > 0$ we have

$$B_{\lambda r}(x_0,\lambda\,\mathfrak{d}) \subset B_r(x_0,\mathfrak{d}) \quad \text{and} \quad B_{r/\lambda}(x_0,\mathfrak{d}) \subset B_r(x_0,\lambda\,\mathfrak{d}).$$

In fact, both inclusions are actually equalities. The claim therefore follows from Lemma 4.4.1. □

EXAMPLE 4.4.3. For $d \geq 1$, consider the following metrics on $\mathbb{R}^d$:

- The Euclidean metric

$$\mathfrak{d}(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{n=1}^{d} |x_n - y_n|^2} \quad \text{for } \mathbf{x} = (x_1,\ldots,x_d) \text{ and } \mathbf{y} = (y_1,\ldots,y_d).$$

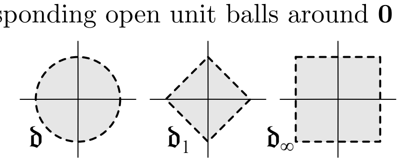
- The ℓ^1-metric

$$\mathfrak{d}_1(\mathbf{x},\mathbf{y}) = \sum_{n=1}^{d} |x_n - y_n| \quad \text{for } \mathbf{x} = (x_1,\ldots,x_d) \text{ and } \mathbf{y} = (y_1,\ldots,y_d).$$

- The ℓ^∞-metric

$$\mathfrak{d}_\infty(\mathbf{x},\mathbf{y}) = \max\{|x_1 - y_1|,\ldots,|x_d - y_d|\} \quad \text{for } \mathbf{x} = (x_1,\ldots,x_d) \text{ and } \mathbf{y} = (y_1,\ldots,y_d).$$

For $d = 2$, pictures of the corresponding open unit balls around $\mathbf{0} \in \mathbb{R}^2$ are as follows:



For all $\mathbf{x},\mathbf{y} \in \mathbb{R}^d$, we claim that

$$\mathfrak{d}_\infty(\mathbf{x},\mathbf{y}) \leq \mathfrak{d}(\mathbf{x},\mathbf{y}) \leq \mathfrak{d}_1(\mathbf{x},\mathbf{y}) \leq d\,\mathfrak{d}_\infty(\mathbf{x},\mathbf{y}).$$

Indeed, the only nontrivial inequality here is $\mathfrak{d}(\mathbf{x},\mathbf{y}) \leq \mathfrak{d}_1(\mathbf{x},\mathbf{y})$, which follows from the following calculation:

$$\mathfrak{d}(\mathbf{x},\mathbf{y})^2 = \sum_{n=1}^{d} |x_n - y_n|^2 \leq \left(\sum_{n=1}^{d} |x_n - y_n|\right)^2 = \mathfrak{d}_1(\mathbf{x},\mathbf{y})^2.$$

The above chain of inequalities implies that for $\mathbf{x}_0 \in \mathbb{R}^d$ and $r > 0$, we have the following containment relations among open balls in these metrics:[3]

$$B_{r/d}(\mathbf{x}_0, \mathfrak{d}_\infty) \subset B_r(\mathbf{x}_0, \mathfrak{d}_1) \subset B_r(\mathbf{x}_0, \mathfrak{d}) \subset B_r(\mathbf{x}_0, \mathfrak{d}_\infty).$$

Applying Lemma 4.4.1, we see that $\mathfrak{d}$ and $\mathfrak{d}_1$ and $\mathfrak{d}_\infty$ are all equivalent metrics. See also Exercise 4.16*, which gives a vast generalization of this. □

The following example shows that every metric is equivalent to a metric of diameter at most 1:

EXAMPLE 4.4.4. Let $(M, \mathfrak{d})$ be a metric space. Define $\mathfrak{d}' \colon M \times M \to \mathbb{R}$ via the formula

$$\mathfrak{d}'(x,y) = \min\{\mathfrak{d}(x,y), 1\} \quad \text{for } x, y \in M.$$

This is a metric on M that is equivalent to $\mathfrak{d}$, and $(M, \mathfrak{d}')$ is a metric space of diameter at most 1 (see Exercise 4.9). □

We now explain two senses in which equivalent metrics are the "same". The first is that they have the same open sets:

LEMMA 4.4.5. *Let M be a set and let $\mathfrak{d}$ and $\mathfrak{d}'$ be metrics on M. Then $\mathfrak{d}$ is equivalent to $\mathfrak{d}'$ if and only if the open sets of $(M, \mathfrak{d})$ are the same as the open sets of $(M, \mathfrak{d}')$.*

PROOF. We must prove the following two claims:

CLAIM. *Assume that $\mathfrak{d}$ is equivalent to $\mathfrak{d}'$. Then the open sets of $(M, \mathfrak{d})$ are the same as the open sets of $(M, \mathfrak{d}')$.*

Since our hypotheses are symmetric in $\mathfrak{d}$ and $\mathfrak{d}'$, it is enough to prove that every set that is open in $(M, \mathfrak{d})$ is also open in $(M, \mathfrak{d}')$. Let U be an open set in $(M, \mathfrak{d})$. Consider $x_0 \in U$. We must find some $r' > 0$ such that $B_{r'}(x_0, \mathfrak{d}') \subset U$. Since U is open in $(M, \mathfrak{d})$, there exists some $r > 0$ such that $B_r(x_0, \mathfrak{d}) \subset U$. Applying condition (i) of Lemma 4.4.1, we get $r' > 0$ such that $B_{r'}(x_0, \mathfrak{d}') \subset B_r(x_0, \mathfrak{d}) \subset U$, as desired.

CLAIM. *Assume that the open sets of $(M, \mathfrak{d})$ are the same as the open sets of $(M, \mathfrak{d}')$. Then $\mathfrak{d}$ is equivalent to $\mathfrak{d}'$.*

We must verify conditions (i) and (ii) of Lemma 4.4.1. Since our hypotheses are symmetric in $\mathfrak{d}$ and $\mathfrak{d}'$, it is enough to verify (i). Consider some $x_0 \in M$ and $r > 0$. Since $B_r(x_0, \mathfrak{d})$ is open in $(M, \mathfrak{d})$, it is also open in $(M, \mathfrak{d}')$. It follows that there is some $r' > 0$ such that $B_{r'}(x_0, \mathfrak{d}') \subset B_r(x_0, \mathfrak{d})$, as desired. □

The second is that they have the same limits. More precisely:

LEMMA 4.4.6. *Let M be a set and let $\mathfrak{d}$ and $\mathfrak{d}'$ be metrics on M. Then $\mathfrak{d}$ is equivalent to $\mathfrak{d}'$ if and only if the following holds:*

- *For every sequence $(x_n)_{n\geq 1}$ in M and every $x \in M$, we have $\lim_{n\to\infty} x_n = x$ in $(M, \mathfrak{d})$ if and only if $\lim_{n\to\infty} x_n = x$ in $(M, \mathfrak{d}')$.*

PROOF. To keep our notation straight, let $X = (M, \mathfrak{d})$ and $Y = (M, \mathfrak{d}')$ and let $f \colon X \to Y$ and $g \colon Y \to X$ both be given by the identity function on the underlying set M. The two metrics are equivalent if and only if both f and g are continuous. By Lemma 2.5.5, these two maps will be continuous if and only if the following two conditions hold:

- For all convergent sequences $(x_n)_{n\geq 1}$ of points of X, we have $\lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n)$.
- For all convergent sequences $(y_n)_{n\geq 1}$ of points of Y, we have $\lim_{n\to\infty} g(y_n) = g(\lim_{n\to\infty} y_n)$.

Translating these back to statements about M and $\mathfrak{d}$ and $\mathfrak{d}'$, they are exactly the indicated limit condition. □

[3]Note that the chain of inclusions among balls reverses the order of the chain of inequalities.

4.5. Topological properties

A property $\mathcal{P}$ of metric spaces is said to be a *topological property* if it is preserved by homeomorphisms. More precisely, it should satisfy the following:

- Let X and Y be homeomorphic metric spaces. Then X has property $\mathcal{P}$ if and only if Y has property $\mathcal{P}$.

EXAMPLE 4.5.1. For a fixed integer $c \geq 0$, one easy topological property of a metric space M is "the underlying set M has cardinality c".[4] Indeed, a homeomorphism $f\colon M \to N$ is in particular a bijection of sets, so if M and N are homeomorphic then necessarily the sets M and N have the same cardinality. □

EXAMPLE 4.5.2. Being path connected is a topological property (see Exercise 4.10). □

EXAMPLE 4.5.3. A metric space $(M, \mathfrak{d})$ is *discrete* if for all $x \in M$, there exists some $r > 0$ such that $B_r(x) = \{x\}$. In other words, there does not exist any $y \in M$ with $y \neq x$ such that $\mathfrak{d}(x, y) < r$. For instance, for $d \geq 1$ the subspace $\mathbb{Z}^d$ of $\mathbb{R}^d$ is discrete. Being discrete is a topological property (see Exercise 4.11). □

REMARK 4.5.4. If M and N are discrete metric spaces, then M and N are homeomorphic if and only if M and N have the same cardinality (see Exercise 4.11). □

EXAMPLE 4.5.5. A metric space is *bounded* if its diameter is finite. Since the unbounded metric space $\mathbb{R}$ is homeomorphic to the bounded metric space $(0, 1)$, being bounded is not a topological property. In fact, as we noted in Example 4.4.4 every metric space is homeomorphic to a bounded metric space. □

4.6. Lipschitz functions and bi-Lipschitz equivalences

We now introduce a class of morphisms between metric spaces intermediate between isometric embeddings and continuous functions. The definition is as follows:

DEFINITION 4.6.1. Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces. A function $f\colon M \to N$ is *Lipschitz* if there exists some $L \geq 0$ such that $\mathfrak{d}_N(f(x), f(y)) \leq L\,\mathfrak{d}_M(x, y)$ for all $x, y \in M$. The constant L is called a *Lipschitz constant* for f. □

These are always continuous:

LEMMA 4.6.2. *Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces and let $f\colon M \to N$ be a Lipschitz function. Then f is continuous.*

PROOF. Let $L \geq 0$ be a Lipschitz constant for f. Increasing L if necessary, we can assume that $L \neq 0$. Consider some $x_0 \in M$ and $\epsilon > 0$. We must find some $\delta > 0$ such that for all $x \in M$ with $\mathfrak{d}_M(x, x_0) < \delta$ we have $\mathfrak{d}_N(f(x), f(x_0)) < \epsilon$. Set $\delta = \epsilon/L$, and consider some $x \in M$ with $\mathfrak{d}_M(x, x_0) < \delta$. Since f has a Lipschitz constant L, we have

$$\mathfrak{d}_N(f(x), f(x_0)) \leq L\,\mathfrak{d}_M(x, x_0) < L\delta = L(\epsilon/L) = \epsilon.$$ □

Lipschitz functions interact well with many geometric quantities. Here is one example of this:

LEMMA 4.6.3. *Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces and let $f\colon M \to N$ be a Lipschitz function with a Lipschitz constant $L \geq 0$. Then for all subsets $A \subset M$ with* $\operatorname{diam}(A) < \infty$ *we have* $\operatorname{diam}(f(A)) \leq L \operatorname{diam}(A)$.

PROOF. Consider $x, y \in A$. By definition, we have

$$\mathfrak{d}_N(f(x), f(y)) \leq L\,\mathfrak{d}_M(x, y) \leq L \operatorname{diam}(A).$$

This implies that $\operatorname{diam}(f(A)) \leq L \operatorname{diam}(A)$, as desired. □

[4] Here c could more generally be a cardinal number.

REMARK 4.6.4. Let $U \subset \mathbb{R}^d$ be an open set and let $f\colon U \to \mathbb{R}^e$ be a function. We can write $f = (f_1, \ldots, f_e)$ for functions $f_1, \ldots, f_e\colon U \to \mathbb{R}$, and f is Lipschitz if and only if each f_i is Lipschitz (see Exercise 4.15). Moreover, if U is convex[5] and each $f_i\colon U \to \mathbb{R}$ is differentiable, then there is a simple criterion for each f_i to be Lipschitz. Indeed:

- For an open convex set $U \subset \mathbb{R}^d$, a differentiable function $g\colon U \to \mathbb{R}$ is Lipschitz if and only if there exists some constant $C \geq 0$ such that
$$\left|\frac{\partial g}{\partial x_j}(\mathbf{x})\right| \leq C \quad \text{for all } \mathbf{x} \in U \text{ and } 1 \leq j \leq d.$$

See Exercise 4.15. □

Associated to Lipschitz functions is the following notion of isomorphism between metric spaces:

DEFINITION 4.6.5. Let M and N be metric spaces. A function $f\colon M \to N$ is a *bi-Lipschitz equivalence* if f is a bijection and both $f\colon M \to N$ and its inverse $g\colon N \to M$ are Lipschitz. If there exists a bi-Lipschitz equivalence from M to N, we say that M is *bi-Lipschitz equivalent* to N. □

These are always homeomorphisms:

LEMMA 4.6.6. *Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces. All bi-Lipschitz equivalences $f\colon M \to N$ are homeomorphisms.*

PROOF. Immediate from the fact that Lipschitz functions are continuous (Lemma 4.6.2). □

REMARK 4.6.7. We warn the reader that there exist Lipschitz homeomorphisms whose inverses are not Lipschitz (see Exercise 4.12). □

The following is a useful characterization of bi-Lipschitz equivalences. The constant $L \geq 1$ in it is called a *bi-Lipschitz constant* of the bi-Lipschitz equivalence $f\colon M \to N$:

LEMMA 4.6.8. *Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces and let $f\colon M \to N$ be a function. Then f is a bi-Lipschitz equivalence if and only if it is a surjection and there exists some $L \geq 1$ such that*
$$\frac{1}{L}\,\mathfrak{d}_M(x, y) \leq \mathfrak{d}_N(f(x), f(y)) \leq L\,\mathfrak{d}_M(x, y) \quad \textit{for all } x, y \in M.$$

PROOF. See Exercise 4.13. □

We now give some examples.

EXAMPLE 4.6.9. Fix some $d \geq 1$. For $r > 0$, let $B_r = B_r(\mathbf{0})$ be the open ball of radius r in $\mathbb{R}^d$ around the origin $\mathbf{0} \in \mathbb{R}^d$. Endow B_r with the Euclidean metric. For $r, s > 0$, define a homeomorphism $f_{rs}\colon B_r \to B_s$ via the formula
$$f_{rs}(\mathbf{x}) = \frac{s}{r}\mathbf{x} \quad \text{for all } \mathbf{x} \in B_r.$$
The inverse of $f_{rs}\colon B_r \to B_s$ is the function $f_{sr}\colon B_s \to B_r$. The functions f_{rs} and f_{sr} are both Lipschitz (see Exercise 4.14), so f_{rs} is a bi-Lipschitz equivalence. □

EXAMPLE 4.6.10. Fix some $d \geq 1$ and let $B = B_1(\mathbf{0})$ be the open unit ball in $\mathbb{R}^d$. Endow B with the Euclidean metric. We can define a homeomorphism $f\colon B \to \mathbb{R}^d$ via the formula
$$f(\mathbf{x}) = \frac{\mathbf{x}}{1 - \|\mathbf{x}\|} \quad \text{for all } \mathbf{x} \in B.$$
The function f is *not* Lipschitz (see Exercise 4.14). In fact, B and $\mathbb{R}^d$ are not bi-Lipschitz equivalent (see Lemma 4.6.11 below). □

Bi-Lipschitz equivalences preserve many metric properties. One such property that is easy to state is as follows:

LEMMA 4.6.11. *Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be two metric spaces that are bi-Lipschitz equivalent. Then M has bounded diameter if and only if N has bounded diameter.*

[5] That is, for all $x, y \in U$ the set U contains the straight line segment connecting x and y.

PROOF. It is enough to prove that if M has bounded diameter then N has bounded diameter. Assume, therefore, that $\operatorname{diam}(M)$ is finite. Let $f\colon M \to N$ be a bi-Lipschitz equivalence with a bi-Lipschitz constant $L \geq 1$. Since $f(M) = N$, Lemma 4.6.3 implies that $\operatorname{diam}(N) \leq L \operatorname{diam}(M) < \infty$, as desired. □

4.7. Bi-Lipschitz equivalent metrics

Let M be a set. We close by describing the Lipschitz analogue of two metrics on a set being equivalent. Two metrics $\mathfrak{d}$ and $\mathfrak{d}'$ on M are *bi-Lipschitz equivalent* if the identity function $\mathbb{1}\colon M \to M$ is a bi-Lipschitz equivalence from $(M, \mathfrak{d})$ to $(M, \mathfrak{d}')$. This can be reformulated as follows:

LEMMA 4.7.1. *Let M be a set and let $\mathfrak{d}$ and $\mathfrak{d}'$ be metrics on M. Then $\mathfrak{d}$ is bi-Lipschitz equivalent to $\mathfrak{d}'$ if and only if there is some $L \geq 1$ such that for all $x, y \in M$ we have*

$$\frac{1}{L}\,\mathfrak{d}(x,y) \leq \mathfrak{d}'(x,y) \leq L\,\mathfrak{d}(x,y).$$

PROOF. Letting $\mathbb{1}\colon M \to M$ be the identity function, Lemma 4.6.8 says that $\mathbb{1}\colon (M, \mathfrak{d}) \to (M, \mathfrak{d}')$ is a bi-Lipschitz equivalence if and only if there exists some $L \geq 1$ such that the following inequality holds:

$$\frac{1}{L}\,\mathfrak{d}(x,y) \leq \mathfrak{d}'(\mathbb{1}(x), \mathbb{1}(y)) \leq L\,\mathfrak{d}(x,y) \quad \text{for all } x, y \in M.$$

Since $\mathbb{1}(x) = x$ and $\mathbb{1}(y) = y$, this is exactly the inequality in the statement of the lemma. □

EXAMPLE 4.7.2. Let $(M, \mathfrak{d})$ be a metric space and let $\lambda > 0$. Let $\lambda\,\mathfrak{d}$ be the metric on M obtained by multiplying $\mathfrak{d}$ by λ. Then $\mathfrak{d}$ is bi-Lipschitz equivalent to $\lambda\,\mathfrak{d}$. Indeed, swapping the roles of $\mathfrak{d}$ and $\lambda\,\mathfrak{d}$ if necessary it is enough to prove this for $\lambda \leq 1$. Set $L = 1/\lambda \geq 1$. For $x, y \in M$, since $1/L = \lambda$ and $\lambda \leq L$ we have

$$\frac{1}{L}\,\mathfrak{d}(x,y) \leq (\lambda\,\mathfrak{d})(x,y) \leq L\,\mathfrak{d}(x,y).$$

In fact, the first inequality is an equality. The claim therefore follows from Lemma 4.7.1. □

EXAMPLE 4.7.3. For $d \geq 1$, consider the following metrics on $\mathbb{R}^d$:

- The Euclidean metric $\mathfrak{d}$.
- The ℓ^1-metric

$$\mathfrak{d}_1(\mathbf{x}, \mathbf{y}) = \sum_{n=1}^{d} |x_n - y_n| \quad \text{for } \mathbf{x} = (x_1, \ldots, x_d) \text{ and } \mathbf{y} = (y_1, \ldots, y_d).$$

- The ℓ^∞-metric

$$\mathfrak{d}_\infty(\mathbf{x}, \mathbf{y}) = \max\{|x_1 - y_1|, \ldots, |x_d - y_d|\} \quad \text{for } \mathbf{x} = (x_1, \ldots, x_d) \text{ and } \mathbf{y} = (y_1, \ldots, y_d).$$

For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, the chain of inequalities

$$\mathfrak{d}_\infty(\mathbf{x}, \mathbf{y}) \leq \mathfrak{d}(\mathbf{x}, \mathbf{y}) \leq \mathfrak{d}_1(\mathbf{x}, \mathbf{y}) \leq d\,\mathfrak{d}_\infty(\mathbf{x}, \mathbf{y})$$

from Example 4.4.3 implies that the hypotheses of Lemma 4.7.1 are satisfied for any pair of these metrics. For instance, they imply that

$$\frac{1}{d}\,\mathfrak{d}(\mathbf{x}, \mathbf{y}) \leq \mathfrak{d}_\infty(\mathbf{x}, \mathbf{y}) \leq \mathfrak{d}(\mathbf{x}, \mathbf{y}).$$

We conclude that all three of these metrics are bi-Lipschitz equivalent. See Exercise 4.16*, which gives a vast generalization of this. □

One source of non-examples is as follows:

EXAMPLE 4.7.4. Let $(M, \mathfrak{d})$ be an unbounded metric space. Define $\mathfrak{d}'\colon M \times M \to \mathbb{R}$ via the formula

$$\mathfrak{d}'(x,y) = \min\{\mathfrak{d}(x,y), 1\} \quad \text{for } x, y \in M.$$

This is a metric on M that is equivalent to $\mathfrak{d}$, and $(M, \mathfrak{d}')$ is a bounded metric space (see Exercise 4.9). Since $(M, \mathfrak{d})$ is unbounded but $(M, \mathfrak{d}')$ is bounded, the metrics $\mathfrak{d}$ and $\mathfrak{d}'$ cannot be bi-Lipschitz equivalent (see Lemma 4.6.11). □

4.8. Summary of chapter

In this chapter, we introduced three kinds of functions between metric spaces. Their names and the implications between them are:

$$\text{isometric embedding} \quad \Rightarrow \quad \text{Lipschitz function} \quad \Rightarrow \quad \text{continuous function.}$$

Neither implication can be reversed (see below). They lead to three kinds of equivalences between metric spaces. Their names and the implications between them are:

$$\text{isometry} \quad \Rightarrow \quad \text{bi-Lipschitz equivalence} \quad \Rightarrow \quad \text{homeomorphism.}$$

Again, neither implication can be reversed (see below).

Here is a table with examples showing that the above implications cannot be reversed. In it, all the subspaces of $\mathbb{R}$ are given their Euclidean metrics.

Property	**Example**
Lipschitz but not isometric embedding	$\mathbb{R} \to \mathbb{R}$, $x \mapsto 2x$
Continuous but not Lipschitz	$[0,1] \to [0,1]$, $x \mapsto \sqrt{x}$ (see Exercise 4.12)
Bi-Lipschitz equivalence but not isometry	$[0,1] \to [0,2]$, $x \mapsto 2x$
Homeomorphism but not bi-Lipschitz equivalence	$(-1,1) \to \mathbb{R}$, $x \mapsto x/(1-\lvert x\rvert)$

4.9. Exercises

Exercise 4.1. Do the following:

(a) Endow $\mathbb{R}$ with the Euclidean metric, and let $f\colon \mathbb{R} \to \mathbb{R}$ be an isometry. Prove that there is some $a \in \mathbb{R}$ such that either $f = \tau_a$ or $f = r_a$, where τ_a and r_a are as follows:
- $\tau_a\colon \mathbb{R} \to \mathbb{R}$ is the function defined by $\tau_a(x) = x + a$ for all $x \in \mathbb{R}$.
- $r_a\colon \mathbb{R} \to \mathbb{R}$ is the function defined by $r_a(x) = a - (x - a) = 2a - x$ for $x \in \mathbb{R}$.

(b)$*$ Let S be a set and let $\mathrm{Seq}(S)$ be the set of all sequences $\mathbf{x} = (x_m)_{m\geq 1}$ of elements of S. Give $\mathrm{Seq}(S)$ the first-difference metric (see Example 2.1.8). We can define isometries from $\mathrm{Seq}(S)$ to itself as follows:
- Consider a sequence $\boldsymbol{\sigma} = (\sigma_m)_{m\geq 1}$ with $\sigma_m\colon S \to S$ a bijection for all $m \geq 1$. We can then define an isometry $\phi_{\boldsymbol{\sigma}}\colon \mathrm{Seq}(S) \to \mathrm{Seq}(S)$ via the formula

$$\phi_{\boldsymbol{\sigma}}((x_m)_{m\geq 1}) = (\sigma_m(x_m))_{m\geq 1} \quad \text{for all } (x_m)_{m\geq 1} \text{ in } \mathrm{Seq}(S).$$

Assume that $|S| \geq 2$. Construct uncountably many isometries from $\mathrm{Seq}(S)$ to itself that are not of this form. Hint: Allow the permutations applied to one coordinate to depend on previous coordinates. $\square$

Exercise 4.2. Let $(M, \mathfrak{d}_M)$ be a metric space, let N be a set, and let $f\colon M \to N$ be a bijection of sets. Prove that there exists a unique metric $\mathfrak{d}_N$ on N such that f is an isometry from $(M, \mathfrak{d}_M)$ to $(N, \mathfrak{d}_N)$. $\square$

Exercise 4.3. Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces and let $f\colon M \to N$ be an isometric embedding. Do the following:

(a) Prove that f is injective.
(b) Prove that f is Lipschitz, and deduce that f is continuous. $\square$

Exercise 4.4. Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ and $(P, \mathfrak{d}_P)$ be metric spaces. Do the following:

(a) Let $f\colon M \to N$ and $g\colon N \to P$ be isometric embeddings. Prove that $g \circ f\colon M \to P$ is an isometric embedding.
(b) Let $f\colon M \to N$ and $g\colon N \to P$ be Lipschitz functions. Prove that $g \circ f\colon M \to P$ is a Lipschitz function.

We remark that we proved in Chapter 2 that the composition of continuous functions is continuous (see Lemma 2.3.4). $\square$

Exercise 4.5. Let $\mathcal{M}$ be any set of metric spaces. Do the following:

(a) Prove that being isometric is an equivalence relation on $\mathcal{M}$.
(b) Prove that being homeomorphic is an equivalence relation on $\mathcal{M}$.
(c) Prove that being bi-Lipschitz equivalent is an equivalence relation on $\mathcal{M}$.

We remark that in all three cases the most important equivalence relation axiom to verify is transitivity. $\square$

EXERCISE 4.6. For each $d \geq 1$, prove that $\mathbb{R}^d \setminus \{\mathbf{0}\}$ and $\mathbb{S}^{d-1} \times \mathbb{R}$ are homeomorphic. Here $\mathbb{R}^d \setminus \{\mathbf{0}\}$ and $\mathbb{S}^{d-1}$ and $\mathbb{R}$ are endowed with their Euclidean metrics and $\mathbb{S}^{d-1} \times \mathbb{R}$ is endowed with its product metric. $\square$

EXERCISE 4.7. Fix some $d \geq 1$. Do the following:

(a) Let $\mathbb{D}^d = \left\{\mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x}\| \leq 1\right\}$ be the closed unit ball in $\mathbb{R}^d$ and let $C = [-1,1]^d$ be the cube in $\mathbb{R}^d$. Give both $\mathbb{D}^d$ and C their Euclidean metrics. Prove that $\mathbb{D}^d$ and C are homeomorphic. Hint: Try to find a homeomorphism $f \colon \mathbb{D}^d \to C$ of the form

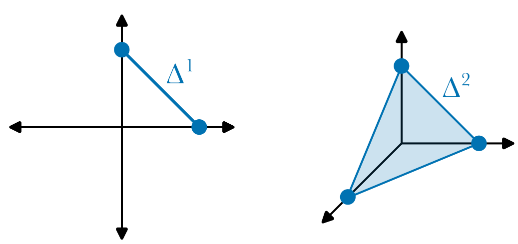
$$f(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} = 0 \\ \lambda(\mathbf{x})x & \text{if } \mathbf{x} \neq 0 \end{cases} \qquad (\mathbf{x} \in \mathbb{D}^d)$$

for some continuous function $\lambda \colon \mathbb{D}^d \setminus \{\mathbf{0}\} \to \mathbb{R}$. We remark that while it is possible to do this for a λ that extends continuously over $\mathbf{0}$, it is a little easier not to require this. However, you should make sure that f is continuous at $\mathbf{0}$.

(b)$*$ Let $C = [-1,1]^d$ be the cube in $\mathbb{R}^d$ and let

$$\Delta^d = \left\{(t_0, \ldots, t_d) \in \mathbb{R}^{d+1} \mid t_i \geq 0 \text{ for all } 0 \leq i \leq d \text{ and } t_0 + \cdots + t_d = 1\right\}.$$

The set Δ^d is called the d-simplex:



Give both C and Δ^d their Euclidean metrics. Prove that C and Δ^d are homeomorphic.

In fact, the metric spaces $\mathbb{D}^d$ and C and Δ^d are all bi-Lipschitz equivalent to each other. Indeed, there is a good chance that the functions you wrote down to solve (a) and (b) are bi-Lipschitz equivalences, but verifying this is a little challenging, so we decided to not include it. $\square$

EXERCISE 4.8. Do the following:

(a) As in Example 4.3.7, identify $\mathbb{R}^2$ with $\mathbb{C}$ and regard $\mathbb{S}^1$ as a subspace of $\mathbb{C}$. Define a function $f \colon [0,1) \to \mathbb{S}^1$ via the formula $f(t) = e^{2\pi i t}$ for $t \in [0,1)$. This is a continuous bijection. Prove that its inverse $g \colon \mathbb{S}^1 \to [0,1)$ is not continuous.
(b) Prove that $[0,1)$ is not homeomorphic to $\mathbb{S}^1$. Hint: try to generalize the argument from Example 4.3.4. $\square$

EXERCISE 4.9. Let $(M, \mathfrak{d})$ be a metric space. In this exercise, you will prove that $\mathfrak{d}$ is equivalent to a metric $\mathfrak{d}'$ such that $(M, \mathfrak{d}')$ is a bounded metric space. Define $\mathfrak{d}' \colon M \times M \to \mathbb{R}$ as follows:

$$\mathfrak{d}'(x,y) = \min\{\mathfrak{d}(x,y), 1\} \quad \text{for all } x, y \in M.$$

Do the following:

(a) Prove that $\mathfrak{d}'$ is a metric on M and that $(M, \mathfrak{d}')$ is bounded.
(b) Prove that $\mathfrak{d}'$ is equivalent to $\mathfrak{d}$. $\square$

EXERCISE 4.10. Prove that being path connected is a topological property. $\square$

EXERCISE 4.11. Do the following:

(a) Prove that being discrete is a topological property. In other words, prove that if M and N are homeomorphic metric spaces, then M is discrete if and only if N is discrete.
(b) Prove that all finite metric spaces are discrete.
(c) Let M and N be discrete metric spaces such that the sets M and N have the same cardinality. Prove that M and N are homeomorphic.
(d) In (c), assume now that M and N are both finite. Prove that M and N are bi-Lipschitz equivalent.

We remark that in Exercise 4.19* you will construct a counterexample to (d) if you assume merely that M and N are both countably infinite. □

EXERCISE 4.12. Give $I = [0, 1]$ its Euclidean metric. Do the following:

(a) Define $f : I \to I$ via the formula $f(x) = x^2$ for all $x \in I$. Prove that f is a Lipschitz homeomorphism.
(b) Define $g : I \to I$ via the formula $g(x) = \sqrt{x}$ for all $x \in I$. Prove that g is a homeomorphism that is not Lipschitz. Conclude that f is a Lipschitz homeomorphism that is not a bi-Lipschitz equivalence. □

EXERCISE 4.13. Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces and let $f : M \to N$ be a function. Prove that f is a bi-Lipschitz equivalence if and only if it is a surjection and there exists some $L \geq 1$ such that

$$\frac{1}{L}\,\mathfrak{d}_M(x, y) \leq \mathfrak{d}_N(f(x), f(y)) \leq L\,\mathfrak{d}_M(x, y) \quad \text{for all } x, y \in M.$$ □

EXERCISE 4.14. Fix some $d \geq 1$. For $r > 0$, let $B_r = B_r(\mathbf{0})$ be the open ball of radius r in $\mathbb{R}^d$ around the origin $\mathbf{0} \in \mathbb{R}^d$. Endow B_r with the Euclidean metric. Do the following:

(a) For $r, s > 0$, define a homeomorphism $f_{rs} : B_r \to B_s$ via the formula

$$f_{rs}(\mathbf{x}) = \frac{s}{r}\mathbf{x} \quad \text{for all } \mathbf{x} \in B_r.$$

Prove that f_{rs} is Lipschitz.
(b) Define a homeomorphism $f : B_1 \to \mathbb{R}^d$ via the formula

$$f(\mathbf{x}) = \frac{\mathbf{x}}{1 - \|\mathbf{x}\|} \quad \text{for all } \mathbf{x} \in B_1.$$

Prove that f is not Lipschitz. □

EXERCISE 4.15. Let $U \subset \mathbb{R}^d$ be an open set. Do the following:

(a) Let $f : U \to \mathbb{R}^e$ be a function. Write $f = (f_1, \ldots, f_e)$ for functions $f_1, \ldots, f_e : U \to \mathbb{R}$. Prove that f is Lipschitz if and only if each f_i is Lipschitz.
(b) Assume now that U is open and convex. Let $g : U \to \mathbb{R}$ be a differentiable function. Prove that g is Lipschitz if and only if there exists some constant $C \geq 0$ such that

$$\left|\frac{\partial g}{\partial x_j}(\mathbf{x})\right| \leq C \quad \text{for all } \mathbf{x} \in U \text{ and } 1 \leq j \leq d.$$

Hint: use the mean value theorem. □

EXERCISE 4.16*. Fix some $d \geq 1$. Recall that a *norm* on $\mathbb{R}^d$ is a function $\mathfrak{n} : \mathbb{R}^d \to \mathbb{R}$ satisfying the following three properties:

- For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, we have $\mathfrak{n}(\mathbf{x} + \mathbf{y}) \leq \mathfrak{n}(\mathbf{x}) + \mathfrak{n}(\mathbf{y})$.
- For all $\mathbf{x} \in \mathbb{R}^d$ and $\lambda \in \mathbb{R}$, we have $\mathfrak{n}(\lambda\mathbf{x}) = |\lambda|\mathfrak{n}(\mathbf{x})$.
- For all $\mathbf{x} \in \mathbb{R}^d$, we have $\mathfrak{n}(\mathbf{x}) \geq 0$ with equality if and only if $\mathbf{x} = 0$.

Prove the following:

(a) Let $\mathfrak{n}$ be a norm on $\mathbb{R}^d$. Its associated metric is the function $\mathfrak{d}_{\mathfrak{n}} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined by $\mathfrak{d}_{\mathfrak{n}}(\mathbf{x}, \mathbf{y}) = \mathfrak{n}(\mathbf{x} - \mathbf{y})$. Prove that $\mathfrak{d}_{\mathfrak{n}}$ is a metric.
(b) Let $\mathfrak{n}$ be a norm on $\mathbb{R}^d$ and let $\| - \|$ be the Euclidean norm, so

$$\|\mathbf{x}\| = \sqrt{x_1^2 + \cdots + x_d^2} \quad \text{for all } \mathbf{x} = (x_1, \ldots, x_d) \text{ in } \mathbb{R}^d.$$

Prove that there is some constant $E > 0$ such that $\mathfrak{n}(\mathbf{x}) \leq E\|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{R}^d$. Hint: Let $\{\vec{e}_1, \ldots, \vec{e}_d\}$ be the standard vector space basis for $\mathbb{R}^d$. For $\mathbf{x} = (x_1, \ldots, x_d)$, start by observing that

$$\mathfrak{n}(\mathbf{x}) = \mathfrak{n}(\sum_{i=1}^{d} x_i \vec{e}_i) \leq \sum_{i=1}^{d} |x_i| \mathfrak{n}(\vec{e}_i).$$

Now use the Cauchy–Schwarz inequality.

(c) Let $\mathfrak{n}$ be a norm on $\mathbb{R}^d$. Prove that the function $\mathfrak{n}\colon \mathbb{R}^d \to \mathbb{R}$ is continuous. Hint: (b) will be helpful.

(d) Let $\mathfrak{n}$ be a norm on $\mathbb{R}^d$ with associated metric $\mathfrak{d}_{\mathfrak{n}}$. Prove that $\mathfrak{d}_{\mathfrak{n}}$ is bi-Lipschitz equivalent to the Euclidean metric $\mathfrak{d}$. Hint: By (c), the function $\mathfrak{n}\colon \mathbb{R}^d \to \mathbb{R}$ is continuous. Since $\mathbb{S}^{d-1}$ is sequentially compact (Theorem 1.10.5), the restriction of $\mathfrak{n}$ to $\mathbb{S}^{d-1}$ is bounded and achieves its infimum and supremum (Theorem 1.11.2). Set

$$C = \min\left\{\mathfrak{n}(\mathbf{x}) \mid \mathbf{x} \in \mathbb{S}^{d-1}\right\} \quad \text{and} \quad D = \max\left\{\mathfrak{n}(\mathbf{x}) \mid \mathbf{x} \in \mathbb{S}^{d-1}\right\}.$$

Prove that $C > 0$, and then prove that $C\|\mathbf{x}\| \leq \mathfrak{n}(\mathbf{x}) \leq D\|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{R}^d$. Use this to prove the result. □

EXERCISE 4.17*. For $p \geq 1$, let $S_p = \{0, 1, 2, \ldots, p\}$. As in Example 2.1.8, let $\mathrm{Seq}(S_p)$ be the set of sequences $\mathbf{s} = (s_m)_{m\geq 1}$ of elements of S_p and let $\mathfrak{d}$ be the first-difference metric on $\mathrm{Seq}(S_p)$. We recall that this is as follows:

- Consider $\mathbf{s}, \mathbf{t} \in \mathrm{Seq}(S_p)$. Write $\mathbf{s} = (s_m)_{m\geq 1}$ and $\mathbf{t} = (t_m)_{m\geq 1}$. If $\mathbf{s} = \mathbf{t}$, then set $\mathfrak{d}(\mathbf{s}, \mathbf{t}) = 0$. Otherwise, let $k = \min\{m \geq 1 \mid s_m \neq t_m\}$ and set $\mathfrak{d}(\mathbf{s}, \mathbf{t}) = 1/k$.

For $p, q \geq 1$, prove that $\mathrm{Seq}(S_p)$ is bi-Lipschitz equivalent to $\mathrm{Seq}(S_q)$. Hint: First argue that it is enough to prove that $\mathrm{Seq}(S_p)$ is bi-Lipschitz equivalent to $\mathrm{Seq}(S_1)$ for all $p \geq 1$. Define a function $f\colon \mathrm{Seq}(S_p) \to \mathrm{Seq}(S_1)$ in the following way. Consider $\mathbf{s} = (s_m)_{m\geq 1}$ in $\mathrm{Seq}(S_p)$. Define $f(\mathbf{s}) \in \mathrm{Seq}(S_1)$ to be the sequence obtained by replacing each term $s_m \in S_p = \{0, \ldots, p\}$ of $\mathbf{s}$ by the following finite sequence of elements of $S_1 = \{0, 1\}$:

- If $s_m = 0$, then replace s_m by the 1-term sequence (0).
- If $1 \leq s_m \leq p - 1$, then replace s_m by the $(s_m + 1)$-term sequence $(1, \ldots, 1, 0)$ with s_m ones and 1 zero.
- If $s_m = p$, then replace s_m by the p-term sequence $(1, \ldots, 1)$ consisting of p ones.

First prove that f is a bijection, and then prove that f and its inverse are both Lipschitz. □

EXERCISE 4.18. As in Example 2.1.8, let $\mathrm{Seq}(\{0, 1\})$ be the set of sequences $\mathbf{s} = (s_m)_{m\geq 1}$ of elements of $\{0, 1\}$ and let $\mathfrak{d}$ be the first-difference metric on $\mathrm{Seq}(\{0, 1\})$ (see Exercise 4.17*). Finally, let C be the classical Cantor set, i.e., the set of all $x \in I = [0, 1]$ that can be written as

$$x = \sum_{m=1}^{\infty} \frac{x_m}{3^m} \quad \text{with } x_m \in \{0, 2\} \text{ for all } m \geq 1.$$

Define a function $f\colon \mathrm{Seq}(\{0, 1\}) \to C$ as follows:

$$f((s_m)_{m\geq 1}) = \sum_{m=1}^{\infty} \frac{2s_m}{3^m} \quad \text{for } (s_m)_{m\geq 1} \in \mathrm{Seq}(\{0, 1\}).$$

Do the following:

(a) Prove that f is a homeomorphism.

(b) Prove that f is not a bi-Lipschitz equivalence.

(c)* Prove that $\mathrm{Seq}(\{0, 1\})$ is not bi-Lipschitz equivalent to C. Hint: Assume that they are bi-Lipschitz equivalent, and let $g\colon \mathrm{Seq}(\{0, 1\}) \to C$ be a bi-Lipschitz equivalence with a bi-Lipschitz constant $L \geq 1$. For $n \geq 1$, let A_n be the following 2^n-element subset of $\mathrm{Seq}(\{0, 1\})$:

$$A_n = \{(a_1, \ldots, a_n, 0, 0, \ldots) \mid a_1, \ldots, a_n \in \{0, 1\}\}.$$

Any two distinct elements of A_n have distance at least $1/n$. It follows that the 2^n elements of $g(A_n) \subset C \subset [0,1]$ have pairwise distances at least $1/(Ln)$. Derive a contradiction from this. $\square$

EXERCISE 4.19*. For $d, e \geq 1$, give $\mathbb{Z}^d$ and $\mathbb{Z}^e$ the Euclidean metrics. Since $\mathbb{Z}^d$ and $\mathbb{Z}^e$ are countably infinite discrete metric spaces, they are homeomorphic (see Exercise 4.11). In this exercise, you will prove that they are not bi-Lipschitz equivalent to each other if $d \neq e$.

(a) Let $(M, \mathfrak{d}_M)$ be a metric space that is bi-Lipschitz equivalent to $\mathbb{Z}^d$ and let $x_0 \in M$. Prove that there exists some $C \geq 1$ such that for all $r \geq 1$ we have $|B_r(x_0)| \leq Cr^d$. In particular, the number of points in the ball $B_r(x_0)$ is finite.

(b) For $d, e \geq 1$ with $d \neq e$, prove that $\mathbb{Z}^d$ is not bi-Lipschitz equivalent to $\mathbb{Z}^e$.

(c)** Using the ideas from parts (a) and (b), prove that for $d, e \geq 1$ with $d \neq e$ the metric spaces $\mathbb{R}^d$ and $\mathbb{R}^e$ are not bi-Lipschitz equivalent. Hint: assume that $d > e$, and for a bi-Lipschitz equivalence $f \colon \mathbb{R}^d \to \mathbb{R}^e$ think about what growth properties $f(\mathbb{Z}^d)$ would have. One useful intermediate thing to prove is that the points of $f(\mathbb{Z}^d)$ are *uniformly separated* in $\mathbb{R}^e$, i.e., there exists some constant $\lambda > 0$ such that $\|f(\mathbf{z}_1) - f(\mathbf{z}_2)\| \geq \lambda$ for all distinct $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{Z}^d$.

We remark that $\mathbb{R}^d$ is not homeomorphic to $\mathbb{R}^e$, but this is much harder to prove. We will prove it in Essay D using dimension theory. $\square$

APPENDIX 4.I*

Sierpiński's classification of countable metric spaces with no isolated points

This appendix proves a theorem of Sierpiński that says that any two nonempty countable metric spaces without isolated points are homeomorphic. Since the proof is a little complicated and we will not need it later in the book, we marked this appendix with an $*$ to indicate that a first-time reader might want to skip it.

4.I.1. Clopen sets

We start with some preliminary results. Let M be a metric space. A subset $E \subset M$ is *clopen* if E is both open and closed.

EXAMPLE 4.I.1.1. For a metric space M, the whole set M and the empty set $\emptyset$ are both clopen. □

The following lemma summarizes some basic properties of clopen sets:

LEMMA 4.I.1.2. *Let M be a metric space. The following hold:*

(a) A set $E \subset M$ is clopen if and only if $M \setminus E$ is clopen.

(b) If $E_1, \ldots, E_n \subset M$ are clopen, then $\cup_{i=1}^n E_i$ and $\cap_{i=1}^n E_i$ are clopen.

PROOF. Conclusion (a) follows from the fact that a set $U \subset M$ is open if and only if $M \setminus U$ is closed (see Lemma 2.8.1). Conclusion (b) follows from the fact that open sets (resp. closed sets) in M are closed under finite unions and intersections (see Lemmas 2.4.8 and 2.7.7). □

Countable metric spaces have many clopen subsets:

LEMMA 4.I.1.3. *Let $(M, \mathfrak{d})$ be a countable metric space, let $U \subset M$ be open, and let $x_0 \in U$. Then for all $\epsilon > 0$, there exists some $r > 0$ with $r < \epsilon$ such that $B_r(x_0)$ is clopen and contained in U.*

PROOF. Decreasing $\epsilon > 0$ if necessary, we can assume that $B_\epsilon(x_0) \subset U$. Let R be the set of all $r > 0$ such that $\{x \in M \mid \mathfrak{d}(x, x_0) = r\} = \emptyset$ and such that $r < \epsilon$. Since M is countable, the set R contains all but countably many $r \in \mathbb{R}$ with $0 < r < \epsilon$. In particular, it is nonempty. Letting $r \in R$, we claim that $B_r(x_0)$ is clopen. Since open balls are open, it is enough to prove that $B_r(x_0)$ is closed. By our choice of r, we have

$$B_r(x_0) = B_r(x_0) \cup \{x \in M \mid \mathfrak{d}(x, x_0) = r\} = \{x \in M \mid \mathfrak{d}(x, x_0) \leq r\}.$$

This is a closed set (see Exercise 2.9, where this set is denoted $B_r^{c\ell}(x_0)$). □

Recall that a point $x \in M$ is an *isolated point* if there exists some $r > 0$ such that $B_r(x) = \{x\}$. In this case, the one-point set $B_r(x) = \{x\}$ is clopen. Later it will be important for us to produce infinite clopen sets. The following lemma ensures this:

LEMMA 4.I.1.4. *Let $(M, \mathfrak{d})$ be a metric space and let $U \subset M$ be an open set that contains a non-isolated point of M. Then U is infinite. Note that this applies in particular if U is clopen.*

PROOF. Let $x \in U$ be a non-isolated point. Pick $r > 0$ such that $B_r(x) \subset U$. Since $B_r(x) \neq \{x\}$, we can find $x_1 \in B_r(x)$ with $x_1 \neq x$. Set $r(1) = \mathfrak{d}(x_1, x)$. Since $B_{r(1)}(x) \neq \{x\}$, we can find $x_2 \in B_{r(1)}(x)$ with $x_2 \neq x$. Set $r(2) = \mathfrak{d}(x_2, x)$. Etc. Repeating this process over and over, we produce a sequence $(x_n)_{n \geq 1}$ of points of $B_r(x) \subset U$ such that

$$\mathfrak{d}(x_1, x) > \mathfrak{d}(x_2, x) > \mathfrak{d}(x_3, x) > \cdots.$$

This implies that the x_n are all distinct, so $\{x_n \mid n \geq 1\}$ is an infinite set of points of U. □

4.I.2. Spaces of sequences

As in Example 2.1.8, for a set S let $\mathrm{Seq}(S)$ be the set of sequences $\mathbf{s} = (s_m)_{m\geq 1}$ of elements of S and let $\mathfrak{d}$ be the first-difference metric on $\mathrm{Seq}(S)$. We will recall the definition of $\mathfrak{d}$ later when we do calculations with it. These contain many clopen sets:

LEMMA 4.I.2.1. *Let S be a set. Give $\mathrm{Seq}(S)$ the first-difference metric. Consider a point $\mathbf{t} = (t_m)_{m\geq 1}$ in $\mathrm{Seq}(S)$ and $r > 0$. The following hold:*

(i) The open ball $B_r(\mathbf{t})$ in $\mathrm{Seq}(S)$ is clopen.
(ii) Let $n \geq 0$ be the largest integer such that $n \leq 1/r$. Then $B_r(\mathbf{t})$ consists of all $\mathbf{s} = (s_m)_{m\geq 1}$ in $\mathrm{Seq}(S)$ such that $s_m = t_m$ for $1 \leq m \leq n$.

PROOF. See parts (c) and (e) of Exercise 2.4. □

We now restrict to $S = \{0, 1\}$. Let $\mathrm{Seq}_0(\{0, 1\})$ be the subspace of $\mathrm{Seq}(\{0, 1\})$ consisting of sequences $\mathbf{s} = (s_m)_{m\geq 1}$ such that there exists some $D \geq 1$ with $\mathbf{s}_m = 0$ for all $m \geq D$. The following summarizes some properties of $\mathrm{Seq}_0(\{0, 1\})$:

LEMMA 4.I.2.2. *The following hold:*

(a) $\mathrm{Seq}_0(\{0, 1\})$ is countable.
(b) $\mathrm{Seq}_0(\{0, 1\})$ has no isolated points.
(c) For all $\mathbf{t} \in \mathrm{Seq}_0(\{0, 1\})$ and $r > 0$, the open ball $B_r(\mathbf{t})$ in $\mathrm{Seq}_0(\{0, 1\})$ is clopen.

PROOF. See Exercise 4.I.2. □

REMARK 4.I.2.3. We will use $\mathrm{Seq}_0(\{0, 1\})$ in our proof of Sierpiński's theorem, but we will not use Lemma 4.I.2.2. That lemma should be regarded more as motivation for our proof. □

4.I.3. Sierpiński's theorem

The main result we will prove in this appendix is as follows. It implies, for instance, that $\mathbb{Q}^d$ and $\mathbb{Q}^e$ are homeomorphic for all $d, e \geq 1$.

THEOREM 4.I.3.1 (Sierpiński). *Any two nonempty countable metric spaces that contain no isolated points are homeomorphic. In particular, any such metric space is homeomorphic to $\mathbb{Q}$.*

PROOF. Let $(M, \mathfrak{d}_M)$ be a nonempty countable metric space with no isolated points. By Lemma 4.I.1.4, the metric space M is countably infinite. Giving $\mathrm{Seq}_0(\{0, 1\})$ the first-difference metric $\mathfrak{d}$, it is enough to prove that M is homeomorphic to $\mathrm{Seq}_0(\{0, 1\})$.

For the proof, let $\mathrm{FinSeq}(\{0, 1\})$ be the set of all finite sequences $\boldsymbol{\sigma} = (\sigma_m)_{m=1}^n$ with $n \geq 0$ and $\sigma_m \in \{0, 1\}$ for $1 \leq m \leq n$. We call n the *length* of $\boldsymbol{\sigma}$. For $n = 0$, this is the empty sequence, which we will also denote by $()$. Note that we regard $\mathrm{FinSeq}(\{0, 1\})$ as a set, not as a metric space. For an element $\boldsymbol{\sigma} = (\sigma_m)_{m=1}^n$ of $\mathrm{FinSeq}(\{0, 1\})$ and $b \in \{0, 1\}$, let $\boldsymbol{\sigma} * b \in \mathrm{FinSeq}(\{0, 1\})$ be the finite sequence obtained by appending b to $\boldsymbol{\sigma}$, so $\boldsymbol{\sigma} * b = (\sigma'_m)_{m=1}^{n+1}$ with

$$\sigma'_m = \begin{cases} \sigma_m & \text{if } 1 \leq m \leq n, \\ b & \text{if } m = n + 1. \end{cases}$$

The proof will have four steps. In the preliminary Step 1 we will construct for each $\boldsymbol{\sigma} \in \mathrm{FinSeq}(\{0, 1\})$ a clopen set $E[\boldsymbol{\sigma}]$ of M. Next, in Step 2 we use these to construct for each $\mathbf{s} \in \mathrm{Seq}_0(\{0, 1\})$ a decreasing sequence

$$M = E[\mathbf{s}, 0] \supset E[\mathbf{s}, 1] \supset E[\mathbf{s}, 2] \supset \cdots$$

of clopen subsets of M. These clopen subsets will satisfy

$$\bigcap_{n=0}^{\infty} E[\mathbf{s}, n] = \{x[\mathbf{s}]\} \quad \text{for some } x[\mathbf{s}] \in M,$$

and in Step 3 we will define $f\colon \mathrm{Seq}_0(\{0, 1\}) \to M$ via the formula $f(\mathbf{s}) = x[\mathbf{s}]$. That step will also prove that f is a bijection. Finally, we will prove that f is a homeomorphism in Step 4.

Fix an enumeration $M = \{x_1, x_2, \ldots\}$. For points $x, x' \in M$, we write $x < x'$ if $x = x_k$ and $x' = x_\ell$ with $k < \ell$. Similarly, we will write $x \leq x'$ if either $x < x'$ or $x = x'$. For a nonempty subset $S \subset M$, define the *minimal point* of S to be the point x_k with $k = \min\{k \geq 1 \mid x_k \in S\}$.

STEP 1. *For each $\boldsymbol{\sigma} \in \mathrm{FinSeq}(\{0,1\})$, we construct a nonempty clopen set $E[\boldsymbol{\sigma}]$ of M.*

Our construction of $E[\boldsymbol{\sigma}]$ will be inductive. Start by setting $E[()] = M$. Next, assume that for some $n \geq 0$ we have constructed $E[\boldsymbol{\sigma}]$ for all $\boldsymbol{\sigma} \in \mathrm{FinSeq}(\{0,1\})$ of length n. Consider some $\boldsymbol{\sigma} \in \mathrm{FinSeq}(\{0,1\})$ of length n. We must define $E[\boldsymbol{\sigma} * 0]$ and $E[\boldsymbol{\sigma} * 1]$. Let $x[\boldsymbol{\sigma}]$ be the minimal point of $E[\boldsymbol{\sigma}]$. Since $E[\boldsymbol{\sigma}]$ is nonempty and open, it is an infinite set (see Lemma 4.I.1.4). We can therefore choose $\rho > 0$ such that $B_\rho(x[\boldsymbol{\sigma}])$ is a proper subset of $E[\boldsymbol{\sigma}]$. Using Lemma 4.I.1.3, we can choose some $r[\boldsymbol{\sigma}] > 0$ with the following properties:

- $B_{r[\boldsymbol{\sigma}]}(x[\boldsymbol{\sigma}])$ is clopen; and
- $B_{r[\boldsymbol{\sigma}]}(x[\boldsymbol{\sigma}]) \subset E[\boldsymbol{\sigma}]$; and
- $r[\boldsymbol{\sigma}] < 1/(n+1)$ and $r[\boldsymbol{\sigma}] < \rho$. Since $r[\boldsymbol{\sigma}] < \rho$, it follows that $B_{r[\boldsymbol{\sigma}]}(x[\boldsymbol{\sigma}]) \neq E[\boldsymbol{\sigma}]$ and thus that $E[\boldsymbol{\sigma}] \setminus B_{r[\boldsymbol{\sigma}]}(x[\boldsymbol{\sigma}])$ is nonempty and clopen (see Lemma 4.I.1.2).

We then define

$$E[\boldsymbol{\sigma} * 0] = B_{r[\boldsymbol{\sigma}]}(x[\boldsymbol{\sigma}]),$$
$$E[\boldsymbol{\sigma} * 1] = E[\boldsymbol{\sigma}] \setminus B_{r[\boldsymbol{\sigma}]}(x[\boldsymbol{\sigma}]).$$

By construction, these are nonempty clopen subsets of M.

STEP 2. *For each point $\mathbf{s} = (s_m)_{m \geq 1}$ in $\mathrm{Seq}_0(\{0,1\})$ and each $n \geq 0$, we construct a nonempty clopen set $E[\mathbf{s}, n]$ of M. These sets have the following properties:*

(a) $E[\mathbf{s}, 0] \supset E[\mathbf{s}, 1] \supset E[\mathbf{s}, 2] \supset \cdots$*; and*
(b) $\lim_{n \to \infty} \mathrm{diam}(E[\mathbf{s}, n]) = 0$*; and*
(c) $\cap_{n=0}^{\infty} E[\mathbf{s}, n] = \{x[\mathbf{s}]\}$ *for a point* $x[\mathbf{s}] \in M$*.*

Consider a point $\mathbf{s} = (s_m)_{m \geq 1}$ of $\mathrm{Seq}_0(\{0,1\})$. For $n \geq 0$, let $\mathbf{s}\langle \leq n \rangle$ be the length n finite sequence $(s_m)_{m=1}^n$. Define

$$E[\mathbf{s}, n] = E[\mathbf{s}\langle \leq n \rangle] \quad \text{for } n \geq 0.$$

Conclusion (a) is immediate. For the other two conclusions, by definition there exists some $D \geq 0$ such that $s_m = 0$ for $m \geq D + 1$. For $n \geq D + 1$, we therefore have

$$E[\mathbf{s}, n] = B_{r[\mathbf{s}\langle \leq n-1 \rangle]}(x[\mathbf{s}\langle \leq n-1 \rangle]).$$

The point $x[\mathbf{s}\langle \leq n-1 \rangle]$ is the minimal point of $E[\mathbf{s}, n-1]$, so it follows that it is also the minimal point of $E[\mathbf{s}, n]$. We conclude that

$$x[\mathbf{s}\langle \leq D \rangle] = x[\mathbf{s}\langle \leq D+1 \rangle] = x[\mathbf{s}\langle \leq D+2 \rangle] = \cdots.$$

Let $x[\mathbf{s}]$ be this common value, so

$$E[\mathbf{s}, n] = B_{r[\mathbf{s}\langle \leq n-1 \rangle]}(x[\mathbf{s}]) \quad \text{for } n \geq D+1.$$

By construction, we have $r[\mathbf{s}\langle \leq n-1 \rangle] < 1/n$. It follows that these are balls around a common point $x[\mathbf{s}]$ with radii tending to 0 as n goes to infinity. Items (b) and (c) follow.

STEP 3. *We construct a bijective function $f\colon \mathrm{Seq}_0(\{0,1\}) \to M$.*

Define f via the formula $f(\mathbf{s}) = x[\mathbf{s}]$ for $\mathbf{s} \in \mathrm{Seq}_0(\{0,1\})$. Consider some $x \in M$. We will prove two things:

(i) There is a unique way to choose $s_m \in \{0,1\}$ such that for $\mathbf{s} = (s_m)_{m \geq 1}$ we have $x \in E[\mathbf{s}\langle \leq n \rangle]$ for all $n \geq 0$.
(ii) This $\mathbf{s} = (s_m)_{m \geq 1}$ lies in $\mathrm{Seq}_0(\{0,1\})$.

Items (i) and (ii) will imply that there is a unique point $\mathbf{s} = (s_m)_{m \geq 1}$ in $\mathrm{Seq}_0(\{0,1\})$ such that $f(\mathbf{s}) = x$. From this, we will be able to conclude that f is a bijection.

We now prove (i). We will construct the s_m inductively, and use $\boldsymbol{\sigma}$ to denote the finite sequence consisting of the s_m we have chosen so far. For some $n \geq 0$, assume that we have proved that there is a unique way to choose $s_m \in \{0,1\}$ for $1 \leq m \leq n$ such that for $\boldsymbol{\sigma} = (s_m)_{m=1}^n$ we have $x \in E[\boldsymbol{\sigma}]$.

We remark that this is vacuous in the base case $n = 0$. We will prove that there exists a unique $s_{n+1} \in \{0,1\}$ such that $x \in E[\boldsymbol{\sigma} * s_{n+1}]$. To see this, note that the fact that $x \in E[\boldsymbol{\sigma}]$ implies that x lies in exactly one of $B_{r[\boldsymbol{\sigma}]}(x[\boldsymbol{\sigma}])$ or $E[\boldsymbol{\sigma}] \setminus B_{r[\boldsymbol{\sigma}]}(x[\boldsymbol{\sigma}])$. We therefore must choose $s_{n+1} = 0$ if $x \in B_{r[\boldsymbol{\sigma}]}(x[\boldsymbol{\sigma}])$ and $s_{n+1} = 1$ if $x \in E[\boldsymbol{\sigma}] \setminus B_{r[\boldsymbol{\sigma}]}(x[\boldsymbol{\sigma}])$. Item (i) follows.

Letting $\mathbf{s} = (s_m)_{m\geq 1}$, we now prove (ii). To avoid awkward notation, we will use notation like $E[\mathbf{s}, n]$ even though technically this has not been defined since we do not yet know that $\mathbf{s} \in \mathrm{Seq}_0(\{0,1\})$. The point x lies in $M = \{x_1, x_2, \ldots\}$. Write $x = x_p$ for some $p \geq 1$. For all $n \geq 0$, we have $x_p \in E[\mathbf{s}, n]$. Since $x[\mathbf{s}, n]$ is the minimal point of $E[\mathbf{s}, n]$, it follows that $x[\mathbf{s}, n] \leq x_p$ for all $n \geq 0$. By construction,

$$x[\mathbf{s}, 0] \leq x[\mathbf{s}, 1] \leq x[\mathbf{s}, 2] \leq \cdots .$$

Since each of these points is less than or equal to x_p, it must be the case that there is some $D \geq 1$ such that $x[\mathbf{s}, n] = x[\mathbf{s}, n+1]$ for all $n \geq D$. For $n \geq D$, we must therefore have $s_{n+1} = 0$; indeed, otherwise we would have

$$E[\mathbf{s}, n+1] = E[\mathbf{s}, n] \setminus B_{r[\mathbf{s},n]}(x[\mathbf{s}, n]),$$

so $E[\mathbf{s}, n+1]$ would not contain $x[\mathbf{s}, n] = x[\mathbf{s}, n+1]$. Conclusion (ii) follows.

Before we move on to prove that f is a homeomorphism, we make the following observation:

- (♠) Consider some $\mathbf{t} = (t_m)_{m\geq 1}$ in $\mathrm{Seq}_0(\{0,1\})$ and some $n \geq 1$. The $\mathbf{s} \in \mathrm{Seq}_0(\{0,1\})$ such that $f(\mathbf{s}) \in E[\mathbf{t}, n]$ are exactly those $\mathbf{s} = (s_m)_{m\geq 1}$ such that $s_m = t_m$ for $1 \leq m \leq n$.

Indeed, this is immediate from the construction underlying (i) above.

STEP 4. *We prove that the function $f\colon \mathrm{Seq}_0(\{0,1\}) \to M$ constructed in Step 3 is a homeomorphism.*

We must prove that both f and f^{-1} are continuous. Since this involves the metric $\mathfrak{d}$ on $\mathrm{Seq}_0(\{0,1\})$, we recall its definition. Consider $\mathbf{s}, \mathbf{t} \in \mathrm{Seq}_0(\{0,1\})$. Write $\mathbf{s} = (s_m)_{m\geq 1}$ and $\mathbf{t} = (t_m)_{m\geq 1}$. If $\mathbf{s} = \mathbf{t}$, then $\mathfrak{d}(\mathbf{s}, \mathbf{t}) = 0$. Otherwise, letting $k = \min\{m \geq 1 \mid s_m \neq t_m\}$ we have $\mathfrak{d}(\mathbf{s}, \mathbf{t}) = 1/k$.

CLAIM. *The function $f\colon \mathrm{Seq}_0(\{0,1\}) \to M$ is continuous.*

Consider a point $\mathbf{t} = (t_m)_{m\geq 1}$ of $\mathrm{Seq}_0(\{0,1\})$ and $\epsilon > 0$. We must find some $\delta > 0$ such that for $\mathbf{s} \in \mathrm{Seq}_0(\{0,1\})$ with $\mathfrak{d}(\mathbf{s}, \mathbf{t}) < \delta$ we have $\mathfrak{d}_M(f(\mathbf{s}), f(\mathbf{t})) < \epsilon$. Since $\lim_{n\to\infty} \mathrm{diam}(E[\mathbf{t}, n]) = 0$, we can find some $n \geq 1$ such that $\mathrm{diam}(E[\mathbf{t}, n]) < \epsilon$. Set $\delta = 1/n$, and consider $\mathbf{s} = (s_m)_{m\geq 1}$ in $\mathrm{Seq}_0(\{0,1\})$ such that $\mathfrak{d}(\mathbf{s}, \mathbf{t}) < \delta$. This implies that $s_m = t_m$ for $1 \leq m \leq n$, so $E[\mathbf{s}, n] = E[\mathbf{t}, n]$. We have $f(\mathbf{s}) \in E[\mathbf{s}, n] = E[\mathbf{t}, n]$ and $f(\mathbf{t}) \in E[\mathbf{t}, n]$, so we conclude that

$$\mathfrak{d}_M(f(\mathbf{s}), f(\mathbf{t})) \leq \mathrm{diam}(E[\mathbf{t}, n]) < \epsilon,$$

as desired.

CLAIM. *Letting $g = f^{-1}$, the function $g\colon M \to \mathrm{Seq}_0(\{0,1\})$ is continuous.*

Consider a point $y \in M$ and $\epsilon > 0$. We must find some $\delta > 0$ such that for $x \in M$ with $\mathfrak{d}_M(x, y) < \delta$ we have $\mathfrak{d}(g(x), g(y)) < \epsilon$. Setting $\mathbf{t} = g(y)$, we must equivalently find some $\delta > 0$ such that for $\mathbf{s} \in \mathrm{Seq}_0(\{0,1\})$ with $\mathfrak{d}_M(f(\mathbf{s}), f(\mathbf{t})) < \delta$ we have $\mathfrak{d}(\mathbf{s}, \mathbf{t}) < \epsilon$.

Pick $n \geq 1$ such that $1/n < \epsilon$. The set $E[\mathbf{t}, n]$ is a clopen set containing $f(\mathbf{t})$, so we can find some $\delta > 0$ such that $B_\delta(f(\mathbf{t})) \subset E[\mathbf{t}, n]$. For $\mathbf{s} \in \mathrm{Seq}_0(\{0,1\})$ with $\mathfrak{d}_M(f(\mathbf{s}), f(\mathbf{t})) < \delta$, we therefore have $f(\mathbf{s}) \in E[\mathbf{t}, n]$. Write $\mathbf{s} = (s_m)_{m\geq 1}$. By the observation (♠) at the end of the proof of Step 3 above, the fact that $f(\mathbf{s}) \in E[\mathbf{t}, n]$ implies that $s_m = t_m$ for $1 \leq m \leq n$. This implies that $\mathfrak{d}(\mathbf{s}, \mathbf{t}) \leq 1/n < \epsilon$, as desired. □

4.I.4. Exercises

EXERCISE 4.I.1. Do the following:

(a) Giving $\mathbb{Q}$ its Euclidean metric, for $r > 0$ and $x \in \mathbb{Q}$ prove that $B_r(x) \subset \mathbb{Q}$ is clopen if and only if r is irrational.

(b) Giving $\mathbb{R}$ its Euclidean metric, for all $r > 0$ and $x \in \mathbb{R}$ prove that $B_r(x) \subset \mathbb{R}$ is not clopen. □

EXERCISE 4.I.2. Do the following:

(a) Prove directly (that is, without citing Theorem 4.I.3.1) that $\mathrm{Seq}_0(\{0,1\})$ is countable.
(b) Prove that $\mathrm{Seq}_0(\{0,1\})$ has no isolated points.
(c) For all $\mathbf{t} \in \mathrm{Seq}_0(\{0,1\})$ and $r > 0$, prove that the open ball $B_r(\mathbf{t})$ in $\mathrm{Seq}_0(\{0,1\})$ is clopen. □

EXERCISE 4.I.3. Let M be a countably infinite metric space and let $\epsilon > 0$. Prove that we can write

$$M = \bigcup_{n=1}^{\infty} E_n,$$

where the E_n are pairwise disjoint clopen sets with $\mathrm{diam}(E_n) < \epsilon$ for all $n \geq 1$. Hint: start by finding a sequence $E_1', E_2', \ldots$ of clopen sets such that $M = \cup_{n=1}^{\infty} E_n'$ and such that $\mathrm{diam}(E_n') < \epsilon$ for all $n \geq 1$. Try to shrink the E_n' one at a time to make them disjoint. □

EXERCISE 4.I.4. Do the following:

(a) Let M be a nonempty countable metric space with no isolated points and $U \subset M$ be a nonempty open set. Prove that M is homeomorphic to U.
(b) For every finite set $F \subset \mathbb{Q}$, prove that $\mathbb{Q}$ is homeomorphic to $\mathbb{Q} \setminus F$. □

EXERCISE 4.I.5. Let M be a nonempty countable metric space with no isolated points. For $x, y \in M$, prove that there exists a homeomorphism $f\colon M \to M$ such that $f(x) = y$. □

EXERCISE 4.I.6. Give an example to show that the following generalization of Sierpiński's theorem is false:

- Let M and N be countably infinite metric spaces. Let $M_{\mathrm{iso}} \subset M$ and $N_{\mathrm{iso}} \subset N$ be their sets of isolated points. Assume that M_{iso} and N_{iso} have the same cardinality. Then M is homeomorphic to N. □

CHAPTER 5

Sequential compactness, total boundedness, and separability

We now introduce three properties that each say that a metric space is not too "large": sequential compactness, total boundedness, and separability. The most important of these conditions is sequential compactness, which generalizes to metric spaces the concept of sequential compactness that was discussed for subsets of Euclidean space in Chapter 1.

5.1. Definition and basic examples

Let M be a metric space. If $(x_n)_{n\geq 1}$ is a sequence of points of M, then a *subsequence* of $(x_n)_{n\geq 1}$ is a sequence $(y_m)_{m\geq 1}$ such that there exists an increasing sequence of positive integers

$$1 \leq n_1 < n_2 < \cdots$$

such that $y_m = x_{n_m}$ for all $m \geq 1$. For instance, $(y_m)_{m\geq 1}$ might be the sequence $(x_{2m})_{m\geq 1}$. We make the following definition:

Definition 5.1.1. Let M be a metric space. We say that M is *sequentially compact* if all sequences $(x_n)_{n\geq 1}$ in M have convergent subsequences.[1] □

Example 5.1.2. If M is a metric space with M a finite set, then M is sequentially compact (see Exercise 5.1). □

Example 5.1.3. By the Heine–Borel theorem (Theorem 1.10.5), a subspace $K \subset \mathbb{R}^d$ is sequentially compact if and only if it is closed and bounded. In particular, for $d \geq 1$ the $(d-1)$-sphere $\mathbb{S}^{d-1} \subset \mathbb{R}^d$ is sequentially compact, but $\mathbb{R}^d$ itself is not. □

Remark 5.1.4. We will give some more sophisticated examples of sequentially compact metric spaces in Appendix 5.II*. □

Recall that a metric space $(M, \mathfrak{d})$ is bounded if it has finite diameter, i.e., there exists some $C \geq 0$ such that $\mathfrak{d}(x, y) \leq C$ for all $x, y \in M$. Every sequentially compact metric space is bounded (see Exercise 5.3). However, the Heine–Borel theorem does not generalize in a simple way beyond subspaces of $\mathbb{R}^d$ (see Remark 5.2.3 below). We illustrate this as follows:

Example 5.1.5. Let M be an infinite set. Equip M with the metric

$$\mathfrak{d}(x, y) = \begin{cases} 1 & \text{if } x \neq y, \\ 0 & \text{if } x = y \end{cases} \quad \text{for } x, y \in M.$$

Then $(M, \mathfrak{d})$ is a bounded metric space that is not sequentially compact. Indeed, any sequence $(x_n)_{n\geq 1}$ of distinct points of M has no convergent subsequence. □

5.2. Sequential compactness is a topological property

We now prove that sequential compactness is a topological property, i.e., that it is preserved by homeomorphisms. The key to this is the following lemma:

Lemma 5.2.1. *Let M and N be metric spaces and let $f\colon M \to N$ be a continuous function. Assume that M is sequentially compact. Then the subspace $f(M)$ of N is sequentially compact.*

[1] Here we emphasize that convergent sequences must converge to points of M – unlike for subsets of $\mathbb{R}^d$, there is no specified ambient space containing points to which they could converge.

Proof. Let $(x_n)_{n\geq 1}$ be a sequence of points of $f(M)$. We must prove that $(x_n)_{n\geq 1}$ has a subsequence that converges to a point of $f(M)$. For each $n \geq 1$, we can choose $z_n \in M$ with $f(z_n) = x_n$. Since M is sequentially compact, the sequence $(z_n)_{n\geq 1}$ in M has a convergent subsequence $(w_m)_{m\geq 1}$. For $m \geq 1$, set $y_m = f(w_m)$. The sequence $(y_m)_{m\geq 1}$ is a subsequence of $(x_n)_{n\geq 1}$. We claim that it converges to a point of $f(M)$. Indeed, since f is continuous we have

$$\lim_{m\to\infty} y_m = \lim_{m\to\infty} f(w_m) = f(\lim_{m\to\infty} w_m) \in f(M);$$

see Lemma 2.5.5. The lemma follows. □

Corollary 5.2.2. *Sequential compactness is a topological property.*

Proof. Let M and N be homeomorphic metric spaces with M sequentially compact. We must prove that N is sequentially compact. This is immediate from Lemma 5.2.1: if $f\colon M \to N$ is a homeomorphism, then since f is surjective Lemma 5.2.1 implies that $N = f(M)$ is sequentially compact. □

Remark 5.2.3. Recall from Example 5.1.5 that there exist bounded metric spaces that are not sequentially compact. This is not an isolated phenomenon. Indeed, recall that every metric space (sequentially compact or not) is homeomorphic to a bounded metric space (see Exercise 4.9). Since being sequentially compact is a topological property, this illustrates that being bounded is not anywhere close to a sufficient condition for being sequentially compact. Later in this chapter we will introduce a condition called being totally bounded that is very close to a sufficient condition. □

5.3. Sequential compactness and closed sets

Our next goal is to prove that a subspace of a sequentially compact metric space is sequentially compact if and only if it is closed:

Lemma 5.3.1. *Let M be a sequentially compact metric space and let $N \subset M$ be a subspace. Then N is sequentially compact if and only if N is a closed subset of M.*

Proof. The implication

$$N \text{ is sequentially compact} \quad \Rightarrow \quad N \text{ is closed}$$

holds in more generality, so we prove it separately in Lemma 5.3.2 below. To prove the converse, assume that N is closed. We must prove that N is sequentially compact. Let $(x_n)_{n\geq 1}$ be a sequence of points of N. We must prove that $(x_n)_{n\geq 1}$ has a subsequence that converges to a point of N. Since M is sequentially compact, $(x_n)_{n\geq 1}$ has a subsequence $(y_m)_{m\geq 1}$ that converges to a point of M. Since N is closed, we have $\lim_{m\to\infty} y_m \in N$ (see Lemma 2.7.5). The lemma follows. □

The following lemma was invoked in the above proof. It says that sequentially compact subspaces of metric spaces are necessarily closed:

Lemma 5.3.2. *Let M be a metric space and let $N \subset M$ be a subspace that is sequentially compact. Then N is a closed subset of M.*

Proof. We must prove that N contains all of its limit points. Let $x \in M$ be a limit point of N. By Lemma 2.6.4, there is a sequence $(y_n)_{n\geq 1}$ of points of N such that $x = \lim_{n\to\infty} y_n$. Since N is sequentially compact, a subsequence of $(y_n)_{n\geq 1}$ converges to a point of N. Since passing to a subsequence does not change the limit of a convergent sequence, this implies that $x \in N$, as desired. □

The above has the following useful corollary. For metric spaces M and N, say that a function $f\colon M \to N$ is a *closed map* if for all closed sets $C \subset M$, the image $f(C)$ is closed in N. Open maps are defined similarly, though we will not need them now. We have:

Corollary 5.3.3. *Let M and N be metric spaces and let $f\colon M \to N$ be a continuous function. Assume that M is sequentially compact. Then f is a closed map.*

Proof. Let $C \subset M$ be closed. Since M is sequentially compact, the subspace C of M is sequentially compact (Lemma 5.3.1). It follows that $f(C)$ is sequentially compact (Lemma 5.2.1), and hence is a closed subset of N (Lemma 5.3.2). □

5.4. Sequential compactness and continuous bijections

To motivate the next consequence of sequential compactness, we repeat Example 4.3.7, which shows that continuous bijections need not be homeomorphisms:

EXAMPLE 5.4.1. Identify $\mathbb{R}^2$ with $\mathbb{C}$, and regard $\mathbb{S}^1$ as a subspace of $\mathbb{C}$. Define a function $f\colon [0,1) \to \mathbb{S}^1$ via the formula $f(t) = e^{2\pi i t}$ for $t \in [0,1)$. The function f is a continuous bijection. However, its inverse $g\colon \mathbb{S}^1 \to [0,1)$ is not continuous, so f is not a homeomorphism. □

In Example 5.4.1, the domain of the continuous bijection $f\colon [0,1) \to \mathbb{S}^1$ is not sequentially compact. The following basic result shows that continuous bijections with sequentially compact domains are necessarily homeomorphisms:

LEMMA 5.4.2. *Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces and let $f\colon M \to N$ be a continuous bijection. Assume that M is sequentially compact. Then f is a homeomorphism.*

PROOF. Let $g\colon N \to M$ be the inverse to f. We must prove that g is continuous. Let $y_0 \in N$, and consider some $\epsilon > 0$. We must find some $\delta > 0$ such that if $y \in N$ satisfies $\mathfrak{d}_N(y, y_0) < \delta$, then $\mathfrak{d}_M(g(y), g(y_0)) < \epsilon$. Rephrasing this in terms of balls, we must find $\delta > 0$ such that $g(B_\delta(y_0)) \subset B_\epsilon(g(y_0))$. Since f is the inverse to g, this is equivalent to

$$B_\delta(y_0) \subset f(B_\epsilon(g(y_0))). \tag{5.4.1}$$

The open ball $B_\epsilon(g(y_0))$ is an open set in M, so $M \setminus B_\epsilon(g(y_0))$ is closed (Lemma 2.8.1). Since M is sequentially compact, the function $f\colon M \to N$ is a closed map (Corollary 5.3.3). It follows that $f(M \setminus B_\epsilon(g(y_0)))$ is closed. Since f is a bijection, this closed set equals

$$f(M \setminus B_\epsilon(g(y_0))) = N \setminus f(B_\epsilon(g(y_0))).$$

We deduce that $f(B_\epsilon(g(y_0)))$ is open. This open set contains $f(g(y_0)) = y_0$, so there is some $\delta > 0$ such that (5.4.1) holds, as desired. □

For metric spaces M and N, a *topological embedding* of M into N is a continuous map $f\colon M \to N$ that is a homeomorphism from M onto its image $f(M)$. The following is a useful immediate corollary of Lemma 5.4.2:

COROLLARY 5.4.3. *Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces and let $f\colon M \to N$ be a continuous injection. Assume that M is sequentially compact. Then f is a topological embedding whose image $f(M)$ is closed.*

PROOF. Since f is a closed map (Corollary 5.3.3), its image $f(M)$ is closed. Replacing N with $f(M)$, we can assume that f is surjective and hence bijective. Lemma 5.4.2 now implies that f is a homeomorphism, as desired. □

5.5. Extreme value theorem

Let M be a nonempty metric space and $f\colon M \to \mathbb{R}$ be a continuous function. The function f is *bounded* if there exist $a \leq b$ such that $a \leq f(x) \leq b$ for all $x \in M$. Assume that f is bounded. The *infimum* of f, denoted $\inf(f)$, is the infimum of the image

$$f(M) = \{f(x) \mid x \in M\} \subset \mathbb{R}$$

of f, i.e., its greatest lower bound. We say that the infimum of f is *attained* if there exists some $x \in M$ such that $f(x) = \inf(f)$, in which case we write $\min(f) = \inf(f)$. Similarly, the *supremum* of f, denoted $\sup(f)$, is the supremum of $f(M)$, i.e., its least upper bound. We say that the supremum of f is *attained* if there exists some $x \in M$ such that $f(x) = \sup(f)$, in which case we write $\max(f) = \sup(f)$.

The following is a generalization to sequentially compact metric spaces of Theorem 1.11.2, which is about real-valued functions on sequentially compact subsets of Euclidean space. We could prove it using exactly the same proof as Theorem 1.11.2, but to illustrate our tools we will give a shorter and more abstract proof:

THEOREM 5.5.1 (Extreme value theorem, metric spaces). *Let M be a nonempty sequentially compact metric space and let $f\colon M \to \mathbb{R}$ be a continuous function. Then:*

(i) The function f is bounded.
(ii) The infimum and supremum of f are attained.

PROOF. Since M is sequentially compact, its image $f(M)$ is a sequentially compact subset of $\mathbb{R}$ (Lemma 5.2.1). Lemma 1.11.3 implies that there exist $a, b \in f(M)$ with $a \leq b$ such that $a = \inf(f(M))$ and $b = \sup(f(M))$. The theorem follows. □

5.6. Uniform continuity and the Heine–Cantor theorem

Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces and let $f\colon M \to N$ be a continuous function. For each $x \in M$ and $\epsilon > 0$, the continuity of f at x gives some $\delta > 0$ such that if $y \in M$ satisfies $\mathfrak{d}_M(x, y) < \delta$, then $\mathfrak{d}_N(f(x), f(y)) < \epsilon$. The value of δ might depend on both x and ϵ. We say that f is *uniformly continuous* if we can choose $\delta > 0$ such that it only depends on ϵ:

DEFINITION 5.6.1. Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces and let $f\colon M \to N$ be a function. We say that f is *uniformly continuous* if for all $\epsilon > 0$, there exists some $\delta > 0$ such that if $x, y \in M$ satisfy $\mathfrak{d}_M(x, y) < \delta$, then $\mathfrak{d}_N(f(x), f(y)) < \epsilon$. □

Every Lipschitz function is uniformly continuous (see Exercise 5.10). However, there are uniformly continuous functions that are not Lipschitz:

EXAMPLE 5.6.2. Let $f\colon [0, 1] \to [0, 1]$ be the function defined by $f(x) = \sqrt{x}$ for $x \in [0, 1]$. The function f is not Lipschitz (see Exercise 4.12). However, f is uniformly continuous. For instance, this follows from Theorem 5.6.5 below. □

It is easy to construct examples of continuous functions that are not uniformly continuous. For instance:

EXAMPLE 5.6.3. Let $f\colon \mathbb{R} \to \mathbb{R}$ be the function given by $f(x) = x^2$ for all $x \in \mathbb{R}$. Then f is not uniformly continuous (see Exercise 5.9). □

EXAMPLE 5.6.4. Let $f\colon (0, 1) \to \mathbb{R}$ be the function given by $f(x) = 1/x$ for all $x \in (0, 1)$. Then f is not uniformly continuous (see Exercise 5.9). □

However, if the domain M is sequentially compact then all continuous functions $f\colon M \to N$ are uniformly continuous:

THEOREM 5.6.5 (Heine–Cantor theorem). *Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces and let $f\colon M \to N$ be a continuous function. Assume that M is sequentially compact. Then f is uniformly continuous.*

PROOF. Assume that f is not uniformly continuous. There must therefore exist some $\epsilon > 0$ such that the following holds:

- For all $\delta > 0$, there exists $x, y \in M$ with $\mathfrak{d}_M(x, y) < \delta$ but $\mathfrak{d}_N(f(x), f(y)) \geq \epsilon$.

For each $n \geq 1$, we can therefore find $x_n, y_n \in M$ with $\mathfrak{d}_M(x_n, y_n) < 1/n$ and $\mathfrak{d}_N(f(x_n), f(y_n)) \geq \epsilon$. Since M is sequentially compact, we can find an increasing sequence of natural numbers

$$n_1 < n_2 < n_3 < \cdots$$

such that the sequence $(x_{n_k})_{k \geq 1}$ of points of M converges. Set $x = \lim_{k\to\infty} x_{n_k}$.

Since f is continuous at x, we can find some $\delta > 0$ such that if $y \in M$ satisfies $\mathfrak{d}_M(x, y) < \delta$, then $\mathfrak{d}_N(f(x), f(y)) < \epsilon/2$. Choose $k \geq 1$ large enough such that $\mathfrak{d}_M(x_{n_k}, x) < \delta/2$ and $1/n_k < \delta/2$. We therefore have the following two things:

$$\begin{aligned} \mathfrak{d}_M(x_{n_k}, x) &< \delta/2 < \delta, \\ \mathfrak{d}_M(y_{n_k}, x) &\leq \mathfrak{d}_M(y_{n_k}, x_{n_k}) + \mathfrak{d}_M(x_{n_k}, x) < \delta/2 + \delta/2 = \delta. \end{aligned}$$

It follows that

$$\mathfrak{d}_N(f(x_{n_k}), f(y_{n_k})) \leq \mathfrak{d}_N(f(x_{n_k}), f(x)) + \mathfrak{d}_N(f(x), f(y_{n_k})) < \epsilon/2 + \epsilon/2 = \epsilon,$$

contradicting the fact that $\mathfrak{d}_N(f(x_{n_k}), f(y_{n_k})) \geq \epsilon$. □

5.7. Total boundedness

We now discuss a property that will turn out to be a useful weakening of sequential compactness. This requires the following preliminary definition:

DEFINITION 5.7.1. Let $(M, \mathfrak{d})$ be a metric space and let $\epsilon > 0$. An *ϵ-net* in M is a subset $\Lambda \subset M$ such that for all $x \in M$, there exists some $z \in \Lambda$ with $\mathfrak{d}(x, z) < \epsilon$. □

For instance, for $d \geq 1$ the integer points $\mathbb{Z}^d$ are a $\sqrt{d}$-net in $\mathbb{R}^d$. We can now define our property:

DEFINITION 5.7.2. A metric space M is *totally bounded* if for every $\epsilon > 0$ there exists a finite ϵ-net in M, that is, an ϵ-net Λ with Λ a finite set. □

Every totally bounded metric space is bounded (see Exercise 5.11). The converse is false:

EXAMPLE 5.7.3. Let M be an infinite set. Endow M with the discrete metric $\mathfrak{d}$:

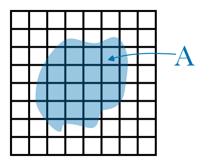
$$\mathfrak{d}(x, y) = \begin{cases} 1 & \text{if } x \neq y, \\ 0 & \text{if } x = y \end{cases} \quad \text{for } x, y \in M.$$

The metric space $(M, \mathfrak{d})$ is bounded, but it is not totally bounded. Indeed, the only 1-net Λ in M is $\Lambda = M$, which is infinite. □

Here is a key example:

LEMMA 5.7.4. *Let $d \geq 1$ and let $A \subset \mathbb{R}^d$ be a bounded subspace. Then A is totally bounded.*

PROOF. Consider $\epsilon > 0$. We must find a finite ϵ-net Λ in A. Since A is bounded, we can find an integer $N \geq 1$ such that A is contained in the cube $[-N, N]^d \subset \mathbb{R}^d$. For some $k \geq 1$, we can write $[-N, N]^d = C_1 \cup \cdots \cup C_k$, where each C_i is a small cube of diameter less than ϵ:



Let $\Lambda \subset A$ be a finite set containing a point in $C_i \cap A$ for each C_i with $C_i \cap A \neq \emptyset$. We claim that Λ is a finite ϵ-net in A. To see this, consider $x \in A$. Since $A \subset [-N, N]^d$, we can find some $1 \leq j \leq k$ such that $x \in C_j$. Letting $z \in \Lambda$ be a point in $C_j \cap A$, since $x, z \in C_j$ and C_j has diameter less than ϵ we have $\mathfrak{d}(x, z) < \epsilon$, as desired. □

We next prove that all sequentially compact metric spaces are totally bounded:

LEMMA 5.7.5. *Let $(M, \mathfrak{d})$ be a sequentially compact metric space. Then M is totally bounded.*

PROOF. This is trivial if $M = \emptyset$, so assume that M is nonempty. Assume that M is not totally bounded. There thus exists some $\epsilon > 0$ such that M has no finite ϵ-net. We will construct a sequence $(x_n)_{n \geq 1}$ of points of M such that for all distinct $n_1, n_2 \geq 1$ we have $\mathfrak{d}(x_{n_1}, x_{n_2}) \geq \epsilon$. Such a sequence has no subsequences that are Cauchy sequences, and in particular has no convergent subsequences (see Lemma 2.5.2). This will contradict the fact that M is sequentially compact.

It remains to construct $(x_n)_{n \geq 1}$. The construction is inductive. Start by letting $x_1 \in M$ be arbitrary. Assume now that we have constructed points $\{x_1, \ldots, x_n\}$ of M such that $\mathfrak{d}(x_{n_1}, x_{n_2}) \geq \epsilon$ for all distinct $1 \leq n_1, n_2 \leq n$. The set $\{x_1, \ldots, x_n\}$ cannot be an ϵ-net since M has no finite ϵ-net. We can therefore find some $x_{n+1} \in M$ such that $\mathfrak{d}(x_k, x_{n+1}) \geq \epsilon$ for all $1 \leq k \leq n$, as desired. □

Not all totally bounded metric spaces are sequentially compact:

EXAMPLE 5.7.6. Since the interval $(0, 1)$ is a bounded subset of $\mathbb{R}$, Lemma 5.7.4 implies that it is totally bounded. However, since $(0, 1)$ is not a closed subset of $\mathbb{R}$ it is not sequentially compact. □

However, it turns out that being totally bounded is not far from being sequentially compact. Recall that a sequence $(x_n)_{n \geq 1}$ of points in a metric space $(M, \mathfrak{d})$ is a *Cauchy sequence* if for all $\epsilon > 0$, there exists some $N \geq 1$ such that for all $n, m \geq N$ we have $\mathfrak{d}(x_n, x_m) < \epsilon$. All convergent sequences are Cauchy sequences (see Lemma 2.5.2), but the converse is false in general. We then have:

LEMMA 5.7.7. *Let $(M, \mathfrak{d})$ be a metric space. Then M is totally bounded if and only if each sequence $(x_n)_{n \geq 1}$ of points of M has a Cauchy subsequence.*

PROOF. Assume that M is totally bounded. We will prove that each sequence $(x_n)_{n \geq 1}$ of points of M has a Cauchy subsequence. The converse is left as an exercise (see Exercise 5.12).

Let $(x_n)_{n \geq 1}$ be a sequence in M. We must prove that $(x_n)_{n \geq 1}$ has a Cauchy subsequence. Since M is totally bounded, for each $k \geq 1$ we can find a finite $1/k$-net $\Lambda[k]$ in M. We inductively construct nested infinite sets of natural numbers

$$\mathcal{I}[0] \supset \mathcal{I}[1] \supset \mathcal{I}[2] \supset \mathcal{I}[3] \supset \cdots$$

as follows. Start by setting $\mathcal{I}[0] = \{1, 2, 3, \ldots\}$. Next, assume that for some $k \geq 1$ we have constructed $\mathcal{I}[k-1]$. Since $\Lambda[k]$ is finite and each x_n has distance less than $1/k$ from some point of $\Lambda[k]$, there must exist some $z_k \in \Lambda[k]$ such that the set

$$\mathcal{I}[k] = \{n \in \mathcal{I}[k-1] \mid \mathfrak{d}(x_n, z_k) < 1/k\}$$

is infinite. This constructs $\mathcal{I}[k]$. Its key property is that for $n, m \in \mathcal{I}[k]$ we have

$$\mathfrak{d}(x_n, x_m) \leq \mathfrak{d}(x_n, z_k) + \mathfrak{d}(z_k, x_m) < 1/k + 1/k = 2/k. \tag{5.7.1}$$

Since the sets $\mathcal{I}[k]$ are infinite, we can find

$$n_1 < n_2 < n_3 < \cdots$$

such that $n_k \in \mathcal{I}[k]$ for all $k \geq 1$: first choose $n_1 \in \mathcal{I}[1]$ arbitrarily, then choose $n_2 \in \mathcal{I}[2]$ with $n_2 > n_1$, then choose $n_3 \in \mathcal{I}[3]$ with $n_3 > n_2$, etc. We claim that $(x_{n_k})_{k \geq 1}$ is a Cauchy sequence. Indeed, consider some $\epsilon > 0$. Pick $K \geq 1$ such that $2/K < \epsilon$. For $k, \ell \geq K$, we have $n_k \in \mathcal{I}[k] \subset \mathcal{I}[K]$ and $n_\ell \in \mathcal{I}[\ell] \subset \mathcal{I}[K]$. Applying (5.7.1), we see that $\mathfrak{d}(x_{n_k}, x_{n_\ell}) < 2/K < \epsilon$, as desired. □

REMARK 5.7.8. In Chapter 6, we will introduce the notion of completeness for metric spaces and prove that metric spaces are sequentially compact if and only if they are complete and totally bounded (see Theorem 6.3.1). □

5.8. Separability

The final condition we will discuss in this chapter is as follows:

DEFINITION 5.8.1. Let M be a metric space. A set $A \subset M$ is *dense* if its closure equals M, i.e., if $\overline{A} = M$. We say that M is *separable* if it has a countable dense subset. □

Most of the metric spaces we have encountered so far are separable:

EXAMPLE 5.8.2. All metric spaces $(M, \mathfrak{d})$ with M finite are separable. For the countable dense subset, we can just take all of M. □

EXAMPLE 5.8.3. The metric space $\mathbb{R}^d$ is separable. Indeed, $\mathbb{Q}^d$ is a countable dense subset. □

EXAMPLE 5.8.4. All subspaces $M \subset \mathbb{R}^d$ are separable. In fact, more generally all subspaces of separable metric spaces are separable. We outline a proof of this in Exercise 5.15. This is a slightly subtle fact; for instance, though $\mathbb{Q}^d$ is a countable dense subset of $\mathbb{R}^d$, we might have $M \cap \mathbb{Q}^d = \emptyset$. □

Here is an important non-example:

EXAMPLE 5.8.5. Let $\ell^\infty_{[0,1]}$ be the set of all sequences $\mathbf{x} = (x_m)_{m \geq 1}$ of real numbers such that $0 \leq x_m \leq 1$ for all $m \geq 1$. Endow $\ell^\infty_{[0,1]}$ with the sup metric

$$\mathfrak{d}(\mathbf{x}, \mathbf{y}) = \sup\{|x_m - y_m| \mid m \geq 1\} \quad \text{for } \mathbf{x} = (x_m)_{m \geq 1} \text{ and } \mathbf{y} = (y_m)_{m \geq 1} \text{ in } \ell^\infty_{[0,1]}.$$

Then $\ell^\infty_{[0,1]}$ is not separable (see Exercise 5.17). □

Just like being sequentially compact, being separable is a topological property (see Exercise 5.14). We now prove that all totally bounded metric spaces are separable. This implies in particular that all sequentially compact metric spaces are separable (see Lemma 5.7.5).

LEMMA 5.8.6. *Let $(M, \mathfrak{d})$ be a totally bounded metric space. Then M is separable.*

PROOF. This is trivial if $M = \emptyset$, so assume that $M \neq \emptyset$. Since M is totally bounded, for each $m \geq 1$ we can find a finite $1/m$-net $S[m]$ in M. Set $S = \cup_{m=1}^{\infty} S[m]$. The set S is countable, and we claim that it is dense. Indeed, let $x \in M$ and let $\epsilon > 0$. We must find some $s \in S$ with $\mathfrak{d}(s, x) < \epsilon$. For this, pick $m \geq 1$ large enough such that $1/m < \epsilon$. Since $S[m]$ is a $1/m$-net in M, we can find some $s \in S[m]$ such that $\mathfrak{d}(s, x) < 1/m < \epsilon$, as desired. □

5.9. Summary of chapter

In this chapter, we introduced three conditions that ensure that a metric space is not too large:

$$\text{sequentially compact} \quad \Rightarrow \quad \text{totally bounded} \quad \Rightarrow \quad \begin{cases} \text{separable} \\ \text{bounded.} \end{cases}$$

Here are some simple (non)-examples of these concepts showing that in particular none of these implications can be reversed. In this table, all the subspaces of $\mathbb{R}$ are given their Euclidean metrics.

Property	**Example**
totally bounded but not sequentially compact	$(0, 1)$
separable but not totally bounded	$\mathbb{Z}$
bounded and separable but not totally bounded	countably infinite set M with discrete metric
bounded but not separable	uncountable set M with discrete metric

REMARK 5.9.1. Being sequentially compact and being separable are both topological properties (see Corollary 5.2.2 and Exercise 5.14). However, neither being totally bounded nor being bounded is a topological property. Indeed, $(0, 1)$ is totally bounded (and hence bounded), but $(0, 1)$ is homeomorphic to the unbounded metric space $\mathbb{R}$. □

5.10. Exercises

EXERCISE 5.1. Let $(M, \mathfrak{d})$ be a metric space with M a finite set. Prove that M is sequentially compact. □

EXERCISE 5.2. Let $(M, \mathfrak{d})$ be a metric space and let $N_1, \ldots, N_k \subset M$ be subspaces. Assume that each N_i is sequentially compact, and set $N = N_1 \cup \cdots \cup N_k$. Prove that N is sequentially compact. □

EXERCISE 5.3. Let $(M, \mathfrak{d})$ be a sequentially compact metric space. Prove that M is bounded, that is, there exists some $C \geq 0$ such that $\mathfrak{d}(x, y) \leq C$ for all $x, y \in M$. □

EXERCISE 5.4. Let M be a metric space. Prove that M is sequentially compact if and only if every infinite subset of M has a limit point in M. □

EXERCISE 5.5. Let $I = [0, 1]$. As in Example 2.1.10, let $\mathcal{C}(I, \mathbb{R})$ be the set of all continuous functions $f\colon I \to \mathbb{R}$ and let $\mathfrak{D}$ be the sup metric on $\mathcal{C}(I, \mathbb{R})$:

$$\mathfrak{D}(f, g) = \sup\{|f(t) - g(t)| \mid t \in I\} \quad \text{for all continuous } f, g\colon I \to \mathbb{R}.$$

Do the following:

(a) Prove that $(\mathcal{C}(I, \mathbb{R}), \mathfrak{D})$ itself is not sequentially compact.
(b) Let $\mathcal{C}(I, I)$ be the subspace of $\mathcal{C}(I, \mathbb{R})$ consisting of continuous functions $f\colon I \to \mathbb{R}$ such that $f(I) \subset I$. Prove that $\mathcal{C}(I, I)$ is bounded but not sequentially compact. □

EXERCISE 5.6. Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be sequentially compact metric spaces. Let $\mathfrak{d}_{\text{prod}}$ be the product metric on $M \times N$, so

$$\mathfrak{d}_{\text{prod}}((x, y), (z, w)) = \mathfrak{d}_M(x, z) + \mathfrak{d}_N(y, w) \quad \text{for all } (x, y), (z, w) \in M \times N.$$

Prove that $(M \times N, \mathfrak{d}_{\text{prod}})$ is sequentially compact. □

EXERCISE 5.7. Let $(M, \mathfrak{d})$ be a metric space and let N_1 and N_2 be two nonempty sequentially compact subspaces of M. Do the following:

(a) Prove that there are points $x_1 \in N_1$ and $x_2 \in N_2$ such that $\mathfrak{d}(x_1, x_2) \leq \mathfrak{d}(y_1, y_2)$ for all $y_1 \in N_1$ and $y_2 \in N_2$.
(b) Prove that $N_1 \cap N_2 = \emptyset$ if and only if $\inf \{\mathfrak{d}(x_1, x_2) \mid x_1 \in N_1,\, x_2 \in N_2\} > 0$.
(c) Give an example of a metric space $(M, \mathfrak{d})$ containing two disjoint closed subsets $X_1, X_2 \subset M$ with $\inf \{\mathfrak{d}(x_1, x_2) \mid x_1 \in X_1,\, x_2 \in X_2\} = 0$. □

EXERCISE 5.8. Let $(M, \mathfrak{d})$ be a metric space. For $k \geq 1$, let $N_k \subset M$ be a nonempty subspace that is sequentially compact. Assume that

$$N_1 \supset N_2 \supset N_3 \supset \cdots .$$

Prove that $\cap_{k=1}^{\infty} N_k$ is nonempty. □

EXERCISE 5.9. Do the following:

(a) Let $f \colon \mathbb{R} \to \mathbb{R}$ be the function given by $f(x) = x^2$ for all $x \in \mathbb{R}$. Prove that f is not uniformly continuous.
(b) Let $f \colon (0,1) \to \mathbb{R}$ be the function given by $f(x) = 1/x$ for all $x \in (0,1)$. Prove that f is not uniformly continuous. □

EXERCISE 5.10. Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces and let $f \colon M \to N$ be a Lipschitz function. Prove that f is uniformly continuous. □

EXERCISE 5.11. Let $(M, \mathfrak{d})$ be a totally bounded metric space. Prove that M is bounded. Hint: start by choosing a finite 1-net Λ in M. □

EXERCISE 5.12. Let $(M, \mathfrak{d})$ be a metric space. Do the following:

(a) Assume that every sequence $(x_n)_{n \geq 1}$ of points of M has a Cauchy subsequence. Prove that M is totally bounded. Hint: prove the contrapositive.
(b) Prove that M is totally bounded if and only if every subset $S \subset M$ such that there exists some $\epsilon > 0$ with $\mathfrak{d}(s,t) \geq \epsilon$ for distinct $s, t \in S$ is finite. □

EXERCISE 5.13. Let $(M, \mathfrak{d})$ be a totally bounded metric space and let $N \subset M$ be a metric subspace. Prove that N is totally bounded. □

EXERCISE 5.14. Prove that being separable is a topological property. In other words, prove that if $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ are homeomorphic metric spaces, then M is separable if and only if N is separable. □

EXERCISE 5.15. Let $(M, \mathfrak{d})$ be a separable metric space and let N be a subspace of M. Prove that N is separable. Hint: If $N = \emptyset$ there is nothing to prove, so assume that $N \neq \emptyset$. Let S be a countable dense subset of M. For every pair $(s, n) \in S \times \mathbb{N}$ such that $B_{1/n}(s) \cap N \neq \emptyset$, choose one point $t_{s,n} \in B_{1/n}(s) \cap N$. Prove that the set T of all such $t_{s,n}$ is a countable dense subset of N. □

EXERCISE 5.16. Let $(M, \mathfrak{d})$ be a metric space. For each $k \geq 1$, let $N_k \subset M$ be a subspace such that N_k is separable. Assume that $M = \cup_{k=1}^{\infty} N_k$. Prove that M is separable. □

EXERCISE 5.17. Let $\ell^{\infty}_{[0,1]}$ be the set of all sequences $\mathbf{x} = (x_m)_{m \geq 1}$ of real numbers such that $0 \leq x_m \leq 1$ for all $m \geq 1$. Endow $\ell^{\infty}_{[0,1]}$ with the metric

$$\mathfrak{d}(\mathbf{x}, \mathbf{y}) = \sup \{|x_m - y_m| \mid m \geq 1\} \quad \text{for } \mathbf{x} = (x_m)_{m \geq 1} \text{ and } \mathbf{y} = (y_m)_{m \geq 1} \text{ in } \ell^{\infty}_{[0,1]}.$$

Do the following:

(a) Prove that $\ell^{\infty}_{[0,1]}$ is not sequentially compact.
(b)* Prove that $\ell^{\infty}_{[0,1]}$ is not separable. □

EXERCISE 5.18*. Let $I = [0,1]$ be the closed interval and let $\mathcal{C}(I, \mathbb{R})$ be the set of all continuous functions $f \colon I \to \mathbb{R}$. Endow $\mathcal{C}(I, \mathbb{R})$ with the sup metric:

$$\mathfrak{D}(f, g) = \sup \{|f(t) - g(t)| \mid t \in I\} \quad \text{for all continuous } f, g \colon I \to \mathbb{R}.$$

Do the following:

(a) For an interval $J \subset \mathbb{R}$, a function $h\colon J \to \mathbb{R}$ is *affine linear* if it is of the form $h(x) = cx + d$ for some $c, d \in \mathbb{R}$. Let $\mathcal{L} \subset \mathcal{C}(I, \mathbb{R})$ be the set of all continuous functions $g\colon I \to \mathbb{R}$ with the following property: for some $k \geq 1$, there exists a partition

$$0 = a_0 < a_1 < \cdots < a_k = 1$$

such that the following hold:

- the function $g|_{[a_{i-1}, a_i]}$ is affine linear for all $1 \leq i \leq k$; and
- we have $a_i \in \mathbb{Q}$ and $g(a_i) \in \mathbb{Q}$ for all $0 \leq i \leq k$.

Prove that $\mathcal{L}$ is dense in $\mathcal{C}(I, \mathbb{R})$.

(b) Prove that $\mathcal{C}(I, \mathbb{R})$ is separable. $\square$

APPENDIX 5.I*

Compactness vs sequential compactness in metric spaces

When we discuss compactness for general topological spaces, we will define it in terms of open covers. This is a genuinely different notion from sequential compactness, and for general topological spaces neither condition implies the other. This appendix introduces compactness for metric spaces and proves that a metric space is compact if and only if it is sequentially compact. A reader learning this material for the first time might want to skip this and postpone the general notion of compactness until they study it for topological spaces, so we have marked this appendix with an $*$.

5.I.1. Open covers

Let M be a metric space. A *cover* of M is a set $\mathfrak{B}$ of subsets of M such that

$$M = \bigcup_{B \in \mathfrak{B}} B.$$

An *open cover* of M is a cover $\mathfrak{U}$ such that each $U \in \mathfrak{U}$ is open. Here are some examples:

Example 5.I.1.1. Consider $\mathbb{R}$ with its Euclidean metric. The set $\mathfrak{U} = \{(n-1, n+1) \mid n \in \mathbb{Z}\}$ of open intervals is an open cover of $\mathbb{R}$. This would not work if these open intervals did not overlap; for instance, the set $\{(n-0.5, n+0.5) \mid n \in \mathbb{Z}\}$ is not a cover of $\mathbb{R}$. $\square$

Example 5.I.1.2. Let M be a metric space. For each $x \in M$, choose a real number $r(x) > 0$. Then $\mathfrak{U} = \left\{B_{r(x)}(x) \mid x \in M\right\}$ is an open cover of M. $\square$

5.I.2. Subcovers and compactness

A cover $\mathfrak{B}$ of a metric space M is a *finite cover* if $\mathfrak{B}$ is a finite set. A *subcover* of a cover $\mathfrak{B}$ is a subset $\mathfrak{B}' \subset \mathfrak{B}$ that is itself a cover. We now come to the main definition of this appendix:

Definition 5.I.2.1. A metric space M is *compact* if each open cover $\mathfrak{U}$ of M has a finite subcover. $\square$

Example 5.I.2.2. Every finite metric space M is compact; indeed, if M is finite, then since there are only finitely many subsets of M each open cover of M is already a finite cover. $\square$

Example 5.I.2.3. Consider $\mathbb{R}$ with its Euclidean metric. The metric space $\mathbb{R}$ is not compact. Indeed, the open cover $\mathfrak{U} = \{(n-1, n+1) \mid n \in \mathbb{Z}\}$ has no finite subcover. To see this, consider a nonempty finite subset

$$\mathfrak{U}' = \{(n_1 - 1, n_1 + 1), \ldots, (n_m - 1, n_m + 1)\}$$

of $\mathfrak{U}$. Setting $N = \max\{n_1, \ldots, n_m\}$, the union of the intervals in $\mathfrak{U}'$ does not contain the point $N + 1 \in \mathbb{R}$. $\square$

Generalizing this example, the following easy lemma says that all compact metric spaces are bounded:

Lemma 5.I.2.4. *Let M be a compact metric space. Then M is bounded.*

Proof. This is trivial if $M = \emptyset$, so assume that M is nonempty. Let $x_0 \in M$, and consider the open cover $\mathfrak{U} = \{B_r(x_0) \mid r > 0\}$. Since M is compact, this has a finite subcover. For some $r_1, \ldots, r_n > 0$, we therefore have

$$M = B_{r_1}(x_0) \cup \cdots \cup B_{r_n}(x_0).$$

Setting $R = \max\{r_1, \ldots, r_n\}$, each $B_{r_i}(x_0)$ is contained in $B_R(x_0)$, so $M = B_R(x_0)$ and hence M is bounded. $\square$

5.I.3. Compact vs sequentially compact

Our main goal in this appendix is to prove the following:

THEOREM 5.I.3.1. *Let M be a metric space. Then M is compact if and only if M is sequentially compact.*

We will prove this in four steps. In §5.I.4, we will study intersections of closed subsets of compact metric spaces. In §5.I.5, we will use this to prove that all compact metric spaces are sequentially compact. Next, in §5.I.6 we will prove a result called the Lebesgue number lemma. Finally, in §5.I.7 we will use the Lebesgue number lemma to prove that all sequentially compact metric spaces are compact.

5.I.4. Intersections of closed subsets of compact metric spaces

In this section, we prove the following. It should be compared to the similar Exercise 5.8 for sequentially compact metric spaces.

LEMMA 5.I.4.1. *Let $(M, \mathfrak{d})$ be a compact metric space. For $k \geq 1$, let $C_k \subset M$ be a nonempty closed subset such that*

$$C_1 \supset C_2 \supset C_3 \supset \cdots .$$

Then $\cap_{k=1}^{\infty} C_k$ is nonempty.

PROOF. Assume for the sake of contradiction that $\cap_{k=1}^{\infty} C_k = \emptyset$. For $k \geq 1$, set $U_k = M \setminus C_k$. Since C_k is closed, U_k is open (Lemma 2.8.1). Since $C_1 \supset C_2 \supset \cdots$, we have

$$U_1 \subset U_2 \subset U_3 \subset \cdots .$$

Finally, since $\cap_{k=1}^{\infty} C_k = \emptyset$, it follows that $\cup_{k=1}^{\infty} U_k = M$. In other words, $\mathfrak{U} = \{U_k \mid k \geq 1\}$ is an open cover of M. Since M is compact, $\mathfrak{U}$ has a finite subcover $\mathfrak{U}' = \{U_{k_1}, \ldots, U_{k_n}\}$. Setting $K = \max\{k_1, \ldots, k_n\}$, we have $U_{k_i} \subset U_K$ for all $1 \leq i \leq n$. Since $\mathfrak{U}'$ is a cover of M, this implies that $M = U_K$. Since $U_K = M \setminus C_K$, we conclude that $C_K = \emptyset$, contradicting the fact that C_K is nonempty. □

5.I.5. Compact metric spaces are sequentially compact

We now prove that all compact metric spaces are sequentially compact:

THEOREM 5.I.5.1. *Let $(M, \mathfrak{d})$ be a metric space. If M is compact, then M is sequentially compact.*

PROOF. Let $(x_n)_{n \geq 1}$ be a sequence of points of M. We must prove that $(x_n)_{n \geq 1}$ has a convergent subsequence. For $n \geq 1$, set

$$C_n = \overline{\{x_m \mid m \geq n\}}.$$

Each C_n is nonempty and closed, and we have

$$C_1 \supset C_2 \supset C_3 \supset \cdots .$$

Applying Lemma 5.I.4.1, we find a point $y \in M$ such that $y \in C_n$ for all $n \geq 1$. For each $\epsilon > 0$ and $n \geq 1$, we can therefore find some $m \geq n$ such that $\mathfrak{d}(x_m, y) < \epsilon$. We now inductively construct a subsequence of $(x_n)_{n \geq 1}$ as follows:

- Start by choosing some $n_1 \geq 1$ such that $\mathfrak{d}(x_{n_1}, y) < 1$.
- Assume now that for some $k \geq 1$ we have constructed n_k. Choose some $n_{k+1} \geq n_k + 1$ such that $\mathfrak{d}(x_{n_{k+1}}, y) < 1/(k+1)$.

By construction, $(x_{n_k})_{k \geq 1}$ is a subsequence of $(x_n)_{n \geq 1}$ such that $\lim_{k \to \infty} \mathfrak{d}(x_{n_k}, y) = 0$. It follows that $\lim_{k \to \infty} x_{n_k} = y$, as desired. □

5.I.6. The Lebesgue number lemma

Before proving that all sequentially compact metric spaces are compact, we must prove an important result. If M is a metric space and $\mathfrak{U}$ is an open cover of M, then a *Lebesgue number* of $\mathfrak{U}$ is a real number $r > 0$ such that for all $x \in M$ there exists some $U \in \mathfrak{U}$ such that $B_r(x) \subset U$. Not all open covers have Lebesgue numbers:

EXAMPLE 5.I.6.1. Consider the following open cover of the open interval $(0, 1)$:

$$\mathfrak{U} = \{(1/n, 1) \mid n \geq 2\}.$$

This open cover has no Lebesgue number. Indeed, consider $r > 0$. Set $x = \min\{r/2, 1/2\}$. The ball $B_r(x)$ contains points arbitrarily close to 0, so $B_r(x)$ is not contained in any interval of the form $(1/n, 1)$ with $n \geq 2$. □

The following basic result says that all open covers of sequentially compact metric spaces have Lebesgue numbers:

LEMMA 5.I.6.2 (Lebesgue number lemma). *Let $(M, \mathfrak{d})$ be a sequentially compact metric space and let $\mathfrak{U}$ be an open cover of M. Then $\mathfrak{U}$ has a Lebesgue number.*

PROOF. Assume that $\mathfrak{U}$ does not have a Lebesgue number. For each $n \geq 1$, we can therefore find some $x_n \in M$ such that $B_{1/n}(x_n)$ is not contained in any $U \in \mathfrak{U}$. Since M is sequentially compact, the sequence $(x_n)_{n\geq 1}$ has a convergent subsequence $(x_{n_k})_{k\geq 1}$. Set $y = \lim_{k\to\infty} x_{n_k}$. Pick $U \in \mathfrak{U}$ such that $y \in U$. Since U is open, we can find some $r > 0$ such that $B_r(y) \subset U$. Pick $k \geq 1$ such that $\mathfrak{d}(x_{n_k}, y) < r/2$ and $n_k > 2/r$. By the triangle inequality, for $z \in B_{r/2}(x_{n_k})$ we have

$$\mathfrak{d}(z, y) \leq \mathfrak{d}(z, x_{n_k}) + \mathfrak{d}(x_{n_k}, y) < r/2 + r/2 = r.$$

It follows that

$$B_{r/2}(x_{n_k}) \subset B_r(y) \subset U.$$

However, by assumption $B_{1/n_k}(x_{n_k})$ is not contained in U. Since $1/n_k < r/2$, this is a contradiction. □

5.I.7. Sequentially compact metric spaces are compact

We now prove that sequentially compact metric spaces are compact. Since we have already proved the converse (Theorem 5.I.5.1), this will imply that for metric spaces being compact is the same as being sequentially compact (see Theorem 5.I.3.1).

THEOREM 5.I.7.1. *Let $(M, \mathfrak{d})$ be a metric space. If M is sequentially compact, then M is compact.*

PROOF. This is trivial if $M = \emptyset$, so assume that M is nonempty. Let $\mathfrak{U}$ be an open cover of M. We must prove that $\mathfrak{U}$ has a finite subcover. Let $r > 0$ be a Lebesgue number for $\mathfrak{U}$, so for all $x \in M$ there exists some $U \in \mathfrak{U}$ such that $B_r(x) \subset U$. Since M is sequentially compact, it is totally bounded (Lemma 5.7.5). We can therefore find a finite r-net $\{x_1, \ldots, x_n\}$ in M. By definition, this means that

$$M = \cup_{k=1}^n B_r(x_k).$$

For each $1 \leq k \leq n$, we can find some $U_k \in \mathfrak{U}$ such that $B_r(x_k) \subset U_k$. We have

$$M = \cup_{k=1}^n B_r(x_k) \subset \cup_{k=1}^n U_k,$$

so we conclude that $\{U_1, \ldots, U_n\}$ is a finite subcover of $\mathfrak{U}$, as desired. □

5.I.8. Final remarks

Since metric spaces are compact if and only if they are sequentially compact, we can replace "sequentially compact" with "compact" in all our previous results. We collect some of these results here:

THEOREM 5.I.8.1 (Heine–Borel theorem, cf. Theorem 1.10.5). *A set $K \subset \mathbb{R}^d$ is compact if and only if it is closed and bounded.*

Lemma 5.I.8.2 (cf. Lemma 5.2.1). *Let M and N be metric spaces and let $f\colon M \to N$ be a continuous function. Assume that M is compact. Then the subspace $f(M)$ of N is compact.*

Lemma 5.I.8.3 (cf. Lemma 5.3.1). *Let M be a compact metric space and let $N \subset M$ be a subspace. Then N is compact if and only if N is a closed subset of M.*

Lemma 5.I.8.4 (cf. Lemma 5.4.2). *Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces and let $f\colon M \to N$ be a continuous bijection. Assume that M is compact. Then f is a homeomorphism.*

Theorem 5.I.8.5 (Extreme value theorem, cf. Theorem 5.5.1). *Let M be a nonempty compact metric space and let $f\colon M \to \mathbb{R}$ be a continuous function. Then:*

(i) The function f is bounded.
(ii) The infimum and supremum of f are attained.

Theorem 5.I.8.6 (Heine–Cantor theorem, cf. Theorem 5.6.5). *Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces and let $f\colon M \to N$ be a continuous function. Assume that M is compact. Then f is uniformly continuous.*

Lemma 5.I.8.7 (Lebesgue number lemma, cf. Lemma 5.I.6.2). *Let $(M, \mathfrak{d})$ be a compact metric space and let $\mathfrak{U}$ be an open cover of M. Then $\mathfrak{U}$ has a Lebesgue number.*

5.I.9. Exercises

Exercise 5.I.1. Let $(M, \mathfrak{d})$ be a compact metric space and let $\mathfrak{U}$ be an open cover of M. Using the open cover definition of compactness, prove that $\mathfrak{U}$ has a Lebesgue number. Hint: start by using the fact that M is compact to find $x_1, \ldots, x_n \in M$ and $\epsilon_1, \ldots, \epsilon_n > 0$ such that the $B_{\epsilon_i}(x_i)$ cover M and such that for all $1 \leq i \leq n$ there exists some $U_i \in \mathfrak{U}$ such that $B_{2\epsilon_i}(x_i) \subset U_i$. □

Exercise 5.I.2. Let M be a metric space. Prove that M is compact if and only if the following holds:

- Let $\mathfrak{B}$ be an open cover of M such that each $B \in \mathfrak{B}$ is of the form $B_r(x)$ for some $x \in M$ and $r > 0$. Then $\mathfrak{B}$ has a finite subcover. □

Exercise 5.I.3. Let M be a metric space and let $K \subset M$ be a subspace. Prove that K is compact if and only if for all collections $\mathfrak{U}$ of open sets in M such that

$$K \subset \bigcup_{U \in \mathfrak{U}} U,$$

there exists a finite subset $\mathfrak{U}' \subset \mathfrak{U}$ such that $K \subset \cup_{U \in \mathfrak{U}'} U$. □

Exercise 5.I.4. Let M be a compact metric space. Using the open cover definition of compactness, prove that M is totally bounded. □

Exercise 5.I.5. Say that a metric space M has the *Lebesgue number property* if every open cover of M has a Lebesgue number. Do the following:

(a) Prove that M is compact if and only if it is totally bounded and has the Lebesgue number property.
(b) Give an example of a metric space that is not compact but has the Lebesgue number property.
(c) Give an example of a metric space that is totally bounded but does not have the Lebesgue number property. □

Exercise 5.I.6. Let M be a compact metric space and let N be another metric space. Using the Lebesgue number lemma, give an alternate proof that every continuous function $f\colon M \to N$ is uniformly continuous. □

APPENDIX 5.II*

Examples: discrete sequences and the Hilbert cube

This appendix describes two examples of sequentially compact metric spaces: the space of sequences of elements of a finite set, and the Hilbert cube. The proofs that they are sequentially compact use a diagonalization argument reminiscent of Cantor's proof of the uncountability of the real numbers. In Chapter 21, we will prove a general theorem called Tychonoff's theorem that will include the sequential compactness of both spaces as a special case. We have marked this appendix with an $*$ to indicate that a first-time reader might want to skip it.

5.II.1. Spaces of discrete sequences

Recall the following definitions from Example 2.1.8. Consider a set S. Let $\mathrm{Seq}(S)$ be the set of sequences $\mathbf{s} = (s_m)_{m \geq 1}$ of elements of S. The first-difference metric $\mathfrak{d}$ on $\mathrm{Seq}(S)$ is as follows. Consider $\mathbf{s}, \mathbf{t} \in \mathrm{Seq}(S)$. Write $\mathbf{s} = (s_m)_{m \geq 1}$ and $\mathbf{t} = (t_m)_{m \geq 1}$. If $\mathbf{s} = \mathbf{t}$, then set $\mathfrak{d}(\mathbf{s}, \mathbf{t}) = 0$. Otherwise, let $k = \min\{m \geq 1 \mid s_m \neq t_m\}$ and set $\mathfrak{d}(\mathbf{s}, \mathbf{t}) = 1/k$. This is a metric (see Exercise 2.4). If S is finite, then $(\mathrm{Seq}(S), \mathfrak{d})$ is sequentially compact:

THEOREM 5.II.1.1. *Let S be a finite set and let $\mathfrak{d}$ be the first-difference metric on* $\mathrm{Seq}(S)$. *Then* $(\mathrm{Seq}(S), \mathfrak{d})$ *is sequentially compact.*

PROOF. Let $(\mathbf{s}[n])_{n \geq 1}$ be a sequence of points of $\mathrm{Seq}(S)$. We must prove that $(\mathbf{s}[n])_{n \geq 1}$ has a convergent subsequence. For $n \geq 1$, write

$$\mathbf{s}[n] = (s[n]_m)_{m \geq 1} \quad \text{with } s[n]_m \in S \text{ for all } m \geq 1.$$

The key to the proof is the following claim:

CLAIM. *There exists $\mathbf{t} = (t_m)_{m \geq 1} \in \mathrm{Seq}(S)$ such that for all $m \geq 1$, there exist infinitely many values of $n \geq 1$ with $s[n]_i = t_i$ for $1 \leq i \leq m$.*

PROOF OF CLAIM. We construct the t_m inductively. The base case $m = 1$ and the inductive case are proved identically, so we give a unified proof of them. For some $m \geq 1$, assume that we have constructed $t_1, \ldots, t_{m-1}$. This is vacuous if $m = 1$. Let

$$\mathcal{I}[m-1] = \{n \geq 1 \mid s[n]_i = t_i \text{ for } 1 \leq i \leq m-1\}.$$

Note that in the edge case $m = 1$ we have $\mathcal{I}[m-1] = \{1, 2, 3, \ldots\}$. By the inductive hypothesis, $\mathcal{I}[m-1]$ is infinite. Since S is finite, we can choose $t_m \in S$ such that the set

$$\mathcal{I}[m] = \{n \in \mathcal{I}[m-1] \mid s[n]_m = t_m\}$$

is infinite. Since $\mathcal{I}[m] \subset \mathcal{I}[m-1]$, for $n \in \mathcal{I}[m]$ we have $s[n]_i = t_i$ for $1 \leq i \leq m$, as desired. □

As in the proof of the claim, for $m \geq 1$ let

$$\mathcal{I}[m] = \{n \geq 1 \mid s[n]_i = t_i \text{ for } 1 \leq i \leq m\}.$$

By construction, these are infinite sets satisfying

$$\mathcal{I}[1] \supset \mathcal{I}[2] \supset \mathcal{I}[3] \supset \cdots.$$

We can therefore find an increasing sequence of integers

$$n_1 < n_2 < n_3 < \cdots$$

such that $n_k \in \mathcal{I}[k]$ for all $k \geq 1$. We claim that the subsequence $(\mathbf{s}[n_k])_{k \geq 1}$ converges to $\mathbf{t}$. To see this, note that by construction the first k terms of $\mathbf{s}[n_k]$ agree with the first k terms of $\mathbf{t}$. It follows that $\mathfrak{d}(\mathbf{s}[n_k], \mathbf{t}) < 1/k$. In particular, $\lim_{k \to \infty} \mathfrak{d}(\mathbf{s}[n_k, \mathbf{t}) = 0$, so $\lim_{k \to \infty} \mathbf{s}[n_k] = \mathbf{t}$, as claimed. □

Remark 5.II.1.2. Another way to prove Theorem 5.II.1.1 is as follows. If S is empty then $\mathrm{Seq}(S) = \emptyset$, and if S has one point then $\mathrm{Seq}(S)$ is a single point. These are both sequentially compact, so we can assume that S has at least 2 points. For $p \geq 1$, let $S_p = \{0, \ldots, p\}$. If S has $(p+1)$ points, then $\mathrm{Seq}(S) \cong \mathrm{Seq}(S_p)$. Exercise 4.17* says that $\mathrm{Seq}(S_p) \cong \mathrm{Seq}(S_1)$, and Exercise 4.18 says that $\mathrm{Seq}(S_1)$ is homeomorphic to the Cantor set C. We deduce that $\mathrm{Seq}(S) \cong C$.

The Cantor set is a closed (see Exercise 2.17*) and bounded subset of $\mathbb{R}$, so by the Heine–Borel theorem (Theorem 1.10.5) C is sequentially compact. Since being sequentially compact is a topological property (Corollary 5.2.2) and $\mathrm{Seq}(S) \cong C$, we conclude that $\mathrm{Seq}(S)$ is sequentially compact. □

5.II.2. The Hilbert cube

We now introduce an important space $\mathcal{H}$ called the Hilbert cube. Just like the space $\ell^\infty_{[0,1]}$ from Example 5.8.5, the points of $\mathcal{H}$ are sequences $(x_m)_{m \geq 1}$ of elements of the interval $[0,1]$. However, the metric on $\mathcal{H}$ is different from the metric on $\ell^\infty_{[0,1]}$. This results in very different metric spaces. Indeed, while $\ell^\infty_{[0,1]}$ is not sequentially compact (see Exercise 5.17), the Hilbert cube will turn out to be sequentially compact.

Let $\mathcal{H}$ be the set of sequences $\mathbf{x} = (x_m)_{m \geq 1}$ of real numbers such that $0 \leq x_m \leq 1$ for all $m \geq 1$. For $\mathbf{x}, \mathbf{y} \in \mathcal{H}$ with $\mathbf{x} = (x_m)_{m \geq 1}$ and $\mathbf{y} = (y_m)_{m \geq 1}$, define

$$\mathfrak{d}(\mathbf{x}, \mathbf{y}) = \sup \left\{ \frac{|x_m - y_m|}{m} \mid m \geq 1 \right\}.$$

Since $|x_m - y_m| \leq 2$ for all $m \geq 1$, the above set is bounded above and this supremum is finite. This is a metric on $\mathcal{H}$ (see Exercise 5.II.1). The metric space $(\mathcal{H}, \mathfrak{d})$ is called the *Hilbert cube*. The following lemma characterizes convergent sequences in the Hilbert cube:

Lemma 5.II.2.1. *Let $(\mathbf{x}[n])_{n \geq 1}$ be a sequence of points of the Hilbert cube $(\mathcal{H}, \mathfrak{d})$. For $n \geq 1$, write $\mathbf{x}[n] = (x[n]_m)_{m \geq 1}$. For a point $\mathbf{y} = (y_m)_{m \geq 1}$ of $\mathcal{H}$, we have $\lim_{n \to \infty} \mathbf{x}[n] = \mathbf{y}$ if and only if*

$$\lim_{n \to \infty} x[n]_m = y_m \quad \textit{for all } m \geq 1.$$

Proof. We divide the proof into two claims.

Claim. *Assume that for some $m_0 \geq 1$ the sequence of real numbers $(x[n]_{m_0})_{n \geq 1}$ does not converge to y_{m_0}. Then the sequence $(\mathbf{x}[n])_{n \geq 1}$ in $\mathcal{H}$ does not converge to $\mathbf{y}$.*

For $n \geq 1$, observe that

$$\mathfrak{d}(\mathbf{x}[n], \mathbf{y}) = \sup \left\{ \frac{|x[n]_m - y_m|}{m} \mid m \geq 1 \right\} \geq \frac{|x[n]_{m_0} - y_{m_0}|}{m_0}.$$

Since the sequence $(x[n]_{m_0})_{n \geq 1}$ does not converge to y_{m_0}, the right-hand side does not approach 0 as n goes to ∞. It follows that $\mathfrak{d}(\mathbf{x}[n], \mathbf{y})$ does not converge to 0 as n goes to ∞, so $(\mathbf{x}[n])_{n \geq 1}$ does not converge to $\mathbf{y}$.

Claim. *For each $m \geq 1$, assume that the sequence of real numbers $(x[n]_m)_{n \geq 1}$ converges to y_m. Then $\lim_{n \to \infty} \mathbf{x}[n] = \mathbf{y}$.*

Fix $\epsilon > 0$. Our goal is to find some $N \geq 1$ such that for $n \geq N$ we have $\mathfrak{d}(\mathbf{x}[n], \mathbf{y}) < \epsilon$. What we will prove is that there exists some $N \geq 1$ such that for $n \geq N$ we have $|x[n]_m - y_m|/m < \epsilon/2$ for all $m \geq 1$. This will imply that for $n \geq N$ we have

$$\mathfrak{d}(\mathbf{x}[n], \mathbf{y}) = \sup \left\{ \frac{|x[n]_m - y_m|}{m} \mid m \geq 1 \right\} \leq \epsilon/2 < \epsilon,$$

as desired.

It remains to find the desired N. Pick $M \geq 1$ such that for $m \geq M+1$ we have $1/m < \epsilon/4$. Since $\lim_{n \to \infty} x[n]_m = y_m$ for all $m \geq 1$, we can find $N \geq 1$ such that for $n \geq N$ and $1 \leq m \leq M$ we have $|x[n]_m - y_m| < m\epsilon/2$. For all $n \geq N$ and $m \geq 1$, we claim that $|x[n]_m - y_m|/m < \epsilon/2$. Indeed:

- if $m \geq M+1$, then $|x[n]_m - y_m|/m \leq 2/m < 2\epsilon/4 = \epsilon/2$; and
- if $1 \leq m \leq M$, then $|x[m]_m - y_m|/m < (m\epsilon/2)/m = \epsilon/2$.

The claim follows. □

Our main result about the Hilbert cube is that it is sequentially compact:

THEOREM 5.II.2.2. *The Hilbert cube* $(\mathcal{H}, \mathfrak{d})$ *is sequentially compact.*

PROOF. Let $(\mathbf{x}[n])_{n\geq 1}$ be a sequence of points of $\mathcal{H}$. We must prove that $(\mathbf{x}[n])_{n\geq 1}$ has a convergent subsequence. For $n \geq 1$, write

$$\mathbf{x}[n] = (x[n]_m)_{m\geq 1} \quad \text{with } 0 \leq x[n]_m \leq 1 \text{ for all } m \geq 1.$$

The key to the proof is the following claim:

CLAIM. *There exists a nested sequence*

$$\mathcal{J}[1] \supset \mathcal{J}[2] \supset \mathcal{J}[3] \supset \cdots$$

of infinite sets of natural numbers such that the following holds for all $m \geq 1$*:*

- *Enumerate* $\mathcal{J}[m]$ *as*

$$\mathcal{J}[m] = \{n_{m,1} < n_{m,2} < n_{m,3} < \cdots\}.$$

Then the sequence $(x[n_{m,k}]_m)_{k\geq 1}$ *of points of* $[0,1]$ *converges.*

PROOF OF CLAIM. We construct the $\mathcal{J}[m]$ inductively. Start by setting $\mathcal{J}[0] = \{1, 2, 3, \ldots\}$. For some $m \geq 1$, assume that we have constructed $\mathcal{J}[m-1]$. Enumerate $\mathcal{J}[m-1]$ as

$$\mathcal{J}[m-1] = \{n_{m-1,1} < n_{m-1,2} < n_{m-1,3} < \cdots\}.$$

We therefore have a sequence $(x[n_{m-1,k}]_m)_{k\geq 1}$ of points of $[0,1]$. Since $[0,1]$ is sequentially compact, this has a convergent subsequence. We can therefore find a strictly increasing sequence of indices

$$k_1 < k_2 < k_3 < \cdots$$

such that $(x[n_{m-1,k_j}]_m)_{j\geq 1}$ converges and let

$$\mathcal{J}[m] = \{n_{m-1,k_1} < n_{m-1,k_2} < n_{m-1,k_3} < \cdots\}. \qquad \square$$

For $m \geq 1$, let $\mathcal{J}[m]$ be as in the above claim. Enumerate it as

$$\mathcal{J}[m] = \{n_{m,1} < n_{m,2} < n_{m,3} < \cdots\}.$$

The sequence $(x[n_{m,k}]_m)_{k\geq 1}$ of points of $[0,1]$ converges. Define

$$y_m = \lim_{k\to\infty} x[n_{m,k}]_m \in [0,1],$$

and set $\mathbf{y} = (y_m)_{m\geq 1} \in \mathcal{H}$. We will prove that our sequence $(\mathbf{x}[n])_{n\geq 1}$ of points of $\mathcal{H}$ has a subsequence that converges to $\mathbf{y}$.

Since the $\mathcal{J}[m]$ are infinite and satisfy

$$\mathcal{J}[1] \supset \mathcal{J}[2] \supset \mathcal{J}[3] \supset \cdots, \tag{5.II.2.1}$$

we can find an increasing sequence

$$\ell_1 < \ell_2 < \ell_3 < \cdots$$

of natural numbers such that $\ell_m \in \mathcal{J}[m]$ for all $m \geq 1$. By (5.II.2.1), for each $m \geq 1$ the sequence

$$\ell_m < \ell_{m+1} < \ell_{m+2} < \ell_{m+3} < \cdots$$

is a subsequence of the sequence obtained by enumerating $\mathcal{J}[m]$ in increasing order:

$$\mathcal{J}[m] = \{n_{m,1} < n_{m,2} < n_{m,3} < \cdots\}.$$

Neither passing to a subsequence nor deleting finitely many terms from a sequence changes its limit, so we deduce that for $m \geq 1$ we have

$$\lim_{k\to\infty} x[\ell_k]_m = \lim_{k\to\infty} x[n_{m,k}]_m = y_m.$$

Applying the criterion from Lemma 5.II.2.1, we conclude that the sequence

$$(\mathbf{x}[\ell_k])_{k\geq 1} = ((x[\ell_k]_m)_{m\geq 1})_{k\geq 1}$$

of points of $\mathcal{H}$ converges to $\mathbf{y} = (y_m)_{m\geq 1}$, as desired. $\square$

The Hilbert cube has many remarkable properties. One that is striking but easy to state is as follows:

THEOREM 5.II.2.3. *Let M be a sequentially compact metric space. Then M is homeomorphic to a closed subspace of the Hilbert cube.*

PROOF. We will outline a proof in Exercise 5.II.4* below. □

REMARK 5.II.2.4. It is also the case that every separable metric space M is homeomorphic to a subspace of the Hilbert cube. That subspace is closed if and only if M is sequentially compact. The proof is similar to the one we give in Exercise 5.II.4* below, but a bit more work is needed to ensure that the injective function from M to the Hilbert cube is a homeomorphism. □

5.II.3. Exercises

EXERCISE 5.II.1. Let $\mathcal{H}$ be the Hilbert cube, that is, the set of sequences $\mathbf{x} = (x_m)_{m \geq 1}$ of real numbers such that $0 \leq x_m \leq 1$ for all $m \geq 1$. Recall that the metric $\mathfrak{d}$ on $\mathcal{H}$ is defined as follows:

$$\mathfrak{d}(\mathbf{x}, \mathbf{y}) = \sup \left\{ \frac{|x_m - y_m|}{m} \mid m \geq 1 \right\} \quad \text{for } \mathbf{x} = (x_m)_{m \geq 1} \text{ and } \mathbf{y} = (y_m)_{m \geq 1} \text{ in } \mathcal{H}.$$

Prove that this is a metric. □

EXERCISE 5.II.2. Do the following:

(a) Let S be a nonempty finite set and let $\mathfrak{d}$ be the first-difference metric on $\operatorname{Seq}(S)$. Since $\operatorname{Seq}(S)$ is sequentially compact, we know it is separable. Construct an explicit countable dense subset of $\operatorname{Seq}(S)$.

(b) Let $(\mathcal{H}, \mathfrak{d})$ be the Hilbert cube. Since $\mathcal{H}$ is sequentially compact, we know it is separable. Construct an explicit countable dense subset of $\mathcal{H}$. Hint: a first idea is to consider the set of points $\mathbf{x} = (x_m)_{m \geq 1}$ with $x_m \in [0, 1] \cap \mathbb{Q}$ for all $m \geq 1$. However, this does not quite work since this is an uncountable set. Try to modify it to get a countable set. □

EXERCISE 5.II.3. Recall that $\ell^\infty_{[0,1]}$ denotes the set of all sequences $(x_m)_{m \geq 1}$ of real numbers such that $0 \leq x_m \leq 1$, equipped with the sup metric. Construct an isometric embedding of $\mathcal{H}$ into $\ell^\infty_{[0,1]}$. □

EXERCISE 5.II.4*. Let $(M, \mathfrak{d})$ be a sequentially compact metric space and let $(\mathcal{H}, \mathfrak{d}_H)$ be the Hilbert cube. Prove that M is homeomorphic to a closed subspace of $\mathcal{H}$. Hint: If $M = \emptyset$ there is nothing to prove, so we can assume that $M \neq \emptyset$. By Lemmas 5.7.5 and 5.8.6, the metric space M is separable. Let S be a countable dense subset of M. Enumerate S as $S = \{s_1, s_2, s_3, \ldots\}$, repeating terms to give infinitely many terms if S is finite. By Exercise 5.3, the metric space M is bounded. Let $C > 0$ be such that $\mathfrak{d}(y, z) \leq C$ for all $y, z \in M$. Recalling that points of $\mathcal{H}$ are sequences $(x_m)_{m \geq 1}$ of real numbers satisfying $0 \leq x_m \leq 1$, define a function $\phi \colon M \to \mathcal{H}$ as follows:

$$\phi(z) = \left(\frac{\mathfrak{d}(z, s_1)}{C}, \frac{\mathfrak{d}(z, s_2)}{C}, \frac{\mathfrak{d}(z, s_3)}{C}, \ldots \right) \quad \text{for } z \in M.$$

Prove that ϕ is a continuous injective function, and then prove that ϕ is a homeomorphism between M and the subspace $\phi(M) \subset \mathcal{H}$. Finally, prove that $\phi(M)$ is closed. □

CHAPTER 6

Complete metric spaces

We next discuss complete metric spaces.

6.1. Complete metric spaces

Let $(M, \mathfrak{d})$ be a metric space. Recall that a sequence $(x_n)_{n \geq 1}$ of points of M is a *Cauchy sequence* if for all $\epsilon > 0$, there exists some $N \geq 1$ such that for all $n, m \geq N$ we have $\mathfrak{d}(x_n, x_m) < \epsilon$. All convergent sequences are Cauchy sequences (see Lemma 2.5.2), but the converse is false in general. This leads to the following definition:

Definition 6.1.1. A metric space $(M, \mathfrak{d})$ is *complete* if all Cauchy sequences in M converge. $\square$

Example 6.1.2. Euclidean space $\mathbb{R}^d$ is complete (Theorem 1.4.2). As we will prove in Lemma 6.1.6, this implies that all closed subspaces of $\mathbb{R}^d$ are also complete. $\square$

Example 6.1.3. The open unit interval $(0, 1)$ with the Euclidean metric is not complete. Indeed, for $n \geq 1$ let $x_n = 1/(n + 1) \in (0, 1)$. The sequence $(x_n)_{n \geq 1}$ of points of $(0, 1)$ is a Cauchy sequence, but it does not converge in $(0, 1)$. The issue here is that the point 0 does not lie in $(0, 1)$. $\square$

Remark 6.1.4. Since $\mathbb{R}$ and $(0, 1)$ are homeomorphic metric spaces with $\mathbb{R}$ complete and $(0, 1)$ not complete, being complete is *not* a topological property. However, being complete is preserved by bi-Lipschitz equivalences (see Exercise 6.11). $\square$

Remark 6.1.5. The non-completeness of $(0, 1)$ has something to do with the "missing" points $\{0, 1\}$. Later we will show how to add points to a metric space to make it complete. $\square$

Many other examples are provided by the following observation:

Lemma 6.1.6. *Let $(M, \mathfrak{d})$ be a complete metric space and let $X \subset M$ be a subspace. Assume that X is a closed subset of M. Then X is complete.*

Proof. Let $(x_n)_{n \geq 1}$ be a Cauchy sequence in X. We must prove that $(x_n)_{n \geq 1}$ converges to a point of X. Since M is complete, the sequence $(x_n)_{n \geq 1}$ converges to a point $x \in M$. Since X is closed, we have $x \in X$ (see Lemma 2.7.5). The claim follows. $\square$

6.2. Completeness of sequentially compact metric spaces

We now prove the following:

Lemma 6.2.1. *Let $(M, \mathfrak{d})$ be a sequentially compact metric space. Then M is complete.*

Proof. Let $(x_n)_{n \geq 1}$ be a Cauchy sequence in M. We must prove that $(x_n)_{n \geq 1}$ converges. Since M is sequentially compact, $(x_n)_{n \geq 1}$ has a convergent subsequence. In other words, there is an increasing sequence of natural numbers

$$n_1 < n_2 < n_3 < \cdots$$

such that $(x_{n_k})_{k \geq 1}$ converges. Set $x = \lim_{k \to \infty} x_{n_k}$. We will prove that $(x_n)_{n \geq 1}$ converges to x. Letting $\epsilon > 0$, we must find some $N \geq 1$ such that for $n \geq N$ we have $\mathfrak{d}(x_n, x) < \epsilon$. Since $(x_n)_{n \geq 1}$ is a Cauchy sequence, we can find some $N \geq 1$ such that for $n, m \geq N$ we have $\mathfrak{d}(x_n, x_m) < \epsilon/2$. Consider some $n \geq N$. Since $(x_{n_k})_{k \geq 1}$ converges to x, we can find some $k \geq 1$ with $n_k \geq N$ such that $\mathfrak{d}(x_{n_k}, x) < \epsilon/2$. Since $n, n_k \geq N$, it follows that

$$\mathfrak{d}(x_n, x) \leq \mathfrak{d}(x_n, x_{n_k}) + \mathfrak{d}(x_{n_k}, x) < \epsilon/2 + \epsilon/2 = \epsilon,$$

as desired. $\square$

EXAMPLE 6.2.2. Since the Hilbert cube $\mathcal{H}$ is sequentially compact (Theorem 5.II.2.2), Lemma 6.2.1 implies that $\mathcal{H}$ is complete. □

EXAMPLE 6.2.3. For a set S, let $\mathrm{Seq}(S)$ be the space of sequences of elements of S with the first-difference metric (see §5.II.1). If S is finite, then $\mathrm{Seq}(S)$ is sequentially compact (Theorem 5.II.1.1), so Lemma 6.2.1 implies that $\mathrm{Seq}(S)$ is complete. In fact, $\mathrm{Seq}(S)$ is complete for arbitrary sets S (see Exercise 6.7). □

6.3. Complete totally bounded metric spaces are sequentially compact

Recall that a metric space $(M, \mathfrak{d})$ is totally bounded if for all $\epsilon > 0$ it has a finite ϵ-net, i.e., there is a finite subset $\Lambda \subset M$ such that for all $x \in M$ there is some $z \in \Lambda$ with $\mathfrak{d}(x, z) < \epsilon$. Sequentially compact metric spaces are totally bounded (Lemma 5.7.5) and complete (Lemma 6.2.1). It turns out that these two properties characterize sequentially compact metric spaces:

THEOREM 6.3.1 (Generalized Heine–Borel theorem). *Let $(M, \mathfrak{d})$ be a metric space. Then M is sequentially compact if and only if it is complete and totally bounded.*

6.4. Spaces of bounded functions

Let $(M, \mathfrak{d})$ be a metric space and let S be a set. Let $\mathrm{Map}(S, M)$ be the set of all functions $f\colon S \to M$. For a set $T \subset M$, define

$$\mathrm{diam}(T) = \sup\left(\{\mathfrak{d}(t_1, t_2) \mid t_1, t_2 \in T\} \cup \{0\}\right).$$

We might have $\mathrm{diam}(T) = \infty$. A function $f\colon S \to M$ is *bounded* if $\mathrm{diam}(f(S)) < \infty$. Let $\mathrm{Map}_b(S, M)$ be the set of all bounded functions $f\colon S \to M$. The sup metric on $\mathrm{Map}_b(S, M)$ is defined as follows:

$$\mathfrak{D}(f, g) = \sup\left(\{\mathfrak{d}(f(s), g(s)) \mid s \in S\} \cup \{0\}\right) \quad \text{for all } f, g \in \mathrm{Map}_b(S, M).$$

We will prove below that $\mathfrak{D}(f, g) < \infty$. Once we have done this, it is straightforward to see that $\mathfrak{D}$ satisfies the axioms of a metric. The proof that $\mathfrak{D}(f, g)$ is finite is as follows:

LEMMA 6.4.1. *Let S be a set, let $(M, \mathfrak{d})$ be a metric space, and let $f, g \in \mathrm{Map}_b(S, M)$. Then $\mathfrak{D}(f, g) < \infty$.*

PROOF. This is immediate if $S = \emptyset$, so assume that $S \neq \emptyset$. Fix some $s_0 \in S$. For $s \in S$, we have

$$\begin{aligned}\mathfrak{d}(f(s), g(s)) &\le \mathfrak{d}(f(s), f(s_0)) + \mathfrak{d}(f(s_0), g(s_0)) + \mathfrak{d}(g(s_0), g(s)) \\ &\le \mathrm{diam}(f(S)) + \mathfrak{d}(f(s_0), g(s_0)) + \mathrm{diam}(g(S)).\end{aligned}$$

Taking the supremum as s ranges over elements of S, we conclude that

$$\mathfrak{D}(f, g) \le \mathrm{diam}(f(S)) + \mathfrak{d}(f(s_0), g(s_0)) + \mathrm{diam}(g(S)) < \infty.$$ □

If M is a complete metric space, the sup metric makes $\mathrm{Map}_b(S, M)$ into a complete metric space:

THEOREM 6.4.2. *Let S be a set, let $(M, \mathfrak{d})$ be a complete metric space, and let $\mathfrak{D}$ be the sup metric on $\mathrm{Map}_b(S, M)$. Then $(\mathrm{Map}_b(S, M), \mathfrak{D})$ is a complete metric space.*

PROOF. This is immediate if $S = \emptyset$, so assume that $S \neq \emptyset$. Let $(f_n)_{n \ge 1}$ be a Cauchy sequence in $\mathrm{Map}_b(S, M)$. We must prove that $(f_n)_{n \ge 1}$ converges to a point of $\mathrm{Map}_b(S, M)$. Define a function $f\colon S \to M$ as follows:

- Consider $s \in S$. For $n, m \ge 1$, we have $\mathfrak{d}(f_n(s), f_m(s)) \le \mathfrak{D}(f_n, f_m)$. Since $(f_n)_{n \ge 1}$ is a Cauchy sequence in $\mathrm{Map}_b(S, M)$, this implies that $(f_n(s))_{n \ge 1}$ is a Cauchy sequence in M. Since M is complete this Cauchy sequence converges. Define $f(s) = \lim_{n \to \infty} f_n(s)$.

To prove the theorem, it is enough to prove the following two things:

CLAIM. *The function $f\colon S \to M$ is bounded.*

Consider $s, t \in S$. We must give an upper bound on $\mathfrak{d}(f(s), f(t))$ that does not depend on s and t. Since $(f_n)_{n\geq 1}$ is a Cauchy sequence, we can find $N \geq 1$ such that for all $n \geq N$ we have $\mathfrak{D}(f_n, f_N) \leq 1$. This N does not depend on s or t. Since $(f_n(s))_{n\geq 1}$ converges to $f(s)$ and $(f_n(t))_{n\geq 1}$ converges to $f(t)$, we can find some $n \geq N$ such that $\mathfrak{d}(f_n(s), f(s)) < 1$ and $\mathfrak{d}(f_n(t), f(t)) < 1$. We have

$$\begin{aligned}\mathfrak{d}(f(s), f(t)) \leq& \mathfrak{d}(f(s), f_n(s)) + \mathfrak{d}(f_n(s), f_N(s)) \\ &+ \mathfrak{d}(f_N(s), f_N(t)) + \mathfrak{d}(f_N(t), f_n(t)) + \mathfrak{d}(f_n(t), f(t)) \\ \leq& 1 + 1 + \mathfrak{d}(f_N(s), f_N(t)) + 1 + 1 \leq \operatorname{diam}(f_N(S)) + 4.\end{aligned}$$

This upper bound $\operatorname{diam}(f_N(S)) + 4$ only depends on N, which as we noted above is independent of s and t, as desired.

Claim. *The sequence $(f_n)_{n\geq 1}$ converges to f.*

Consider $\epsilon > 0$. We must find some $N \geq 1$ such that for $n \geq N$ we have $\mathfrak{D}(f_n, f) < \epsilon$. Since $\mathfrak{D}(f_n, f)$ is the supremum of $\mathfrak{d}(f_n(s), f(s))$ as s ranges over elements of S, it is enough to find some $N \geq 1$ such that for $n \geq N$ and $s \in S$ we have $\mathfrak{d}(f_n(s), f(s)) < \epsilon/2$. Taking the supremum, we will then be able to conclude that $\mathfrak{D}(f_n, f) \leq \epsilon/2 < \epsilon$.

Since $(f_n)_{n\geq 1}$ is a Cauchy sequence, we can find some $N \geq 1$ such that for $n, m \geq N$ we have $\mathfrak{D}(f_n, f_m) < \epsilon/4$. Consider some $s \in S$ and $n \geq N$. To see that $\mathfrak{d}(f_n(s), f(s)) < \epsilon/2$, observe that since $f(s) = \lim_{m\to\infty} f_m(s)$ we can find some $m \geq N$ such that $\mathfrak{d}(f_m(s), f(s)) < \epsilon/4$. We then have

$$\begin{aligned}\mathfrak{d}(f_n(s), f(s)) \leq& \mathfrak{d}(f_n(s), f_m(s)) + \mathfrak{d}(f_m(s), f(s)) \\ \leq& \mathfrak{D}(f_n, f_m) + \mathfrak{d}(f_m(s), f(s)) < \epsilon/4 + \epsilon/4 = \epsilon/2.\end{aligned}$$ □

Recall that ℓ^∞ is the set of bounded sequences of real numbers (see Example 2.1.9). We have:

Lemma 6.4.3. *Let $S = \{1, 2, 3, \ldots\}$. Then $\mathrm{Map}_b(S, \mathbb{R})$ with the sup metric is isometric to ℓ^∞ with the sup metric.*

Proof. Define a map $\phi\colon \mathrm{Map}_b(S, \mathbb{R}) \to \ell^\infty$ as follows. Consider $f \in \mathrm{Map}_b(S, \mathbb{R})$, so f is a bounded function $f\colon S \to \mathbb{R}$. For $m \geq 1$, let $x_m = f(m)$. We then define $\phi(f) = (x_m)_{m\geq 1} \in \ell^\infty$. It is immediate from the definitions that ϕ is a bijection and an isometric embedding, and hence an isometry. □

In light of Theorem 6.4.2, this implies the following:

Corollary 6.4.4. *The metric space ℓ^∞ with the sup metric is complete.*

6.5. Spaces of continuous functions

Let $(N, \mathfrak{d}_N)$ and $(M, \mathfrak{d}_M)$ be metric spaces. Let $\mathcal{C}_b(N, M)$ be the set of all bounded functions $f\colon N \to M$ that are continuous. This is a subset of $\mathrm{Map}_b(N, M)$, which is a metric space with respect to the sup metric. We will prove that it is a closed subset:

Lemma 6.5.1. *Let $(N, \mathfrak{d}_N)$ and $(M, \mathfrak{d}_M)$ be metric spaces. Let $\mathfrak{D}$ be the sup metric on $\mathrm{Map}_b(N, M)$. Then $\mathcal{C}_b(N, M)$ is a closed subset of $(\mathrm{Map}_b(N, M), \mathfrak{D})$.*

Proof. Let $(f_n)_{n\geq 1}$ be a sequence of elements of $\mathcal{C}_b(N, M)$ that converges to $f \in \mathrm{Map}_b(N, M)$. To prove that $\mathcal{C}_b(N, M)$ is closed, we must prove that $f \in \mathcal{C}_b(N, M)$ (see Lemma 2.7.5). In other words, we must prove that f is continuous. Consider a point $x_0 \in N$ and some $\epsilon > 0$. Our goal is to find some $\delta > 0$ such that for $x \in N$ with $\mathfrak{d}_N(x, x_0) < \delta$ we have $\mathfrak{d}_M(f(x), f(x_0)) < \epsilon$.

Since $(f_n)_{n\geq 1}$ converges to f, we can find some $n_0 \geq 1$ such that $\mathfrak{D}(f_{n_0}, f) < \epsilon/3$. This implies that $\mathfrak{d}_M(f_{n_0}(x), f(x)) < \epsilon/3$ for all $x \in N$. Since $f_{n_0}\colon N \to M$ is continuous at x_0, we can find $\delta > 0$ such that for $x \in N$ with $\mathfrak{d}_N(x, x_0) < \delta$ we have $\mathfrak{d}_M(f_{n_0}(x), f_{n_0}(x_0)) < \epsilon/3$. For $x \in N$ with $\mathfrak{d}_N(x, x_0) < \delta$ we then have

$$\begin{aligned}\mathfrak{d}_M(f(x), f(x_0)) \leq& \mathfrak{d}_M(f(x), f_{n_0}(x)) + \mathfrak{d}_M(f_{n_0}(x), f_{n_0}(x_0)) + \mathfrak{d}_M(f_{n_0}(x_0), f(x_0)) \\ <& \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon.\end{aligned}$$ □

We will call the restriction to $\mathcal{C}_b(N, M)$ of the sup metric on $\mathrm{Map}_b(N, M)$ the sup metric on $\mathcal{C}_b(N, M)$. Combining Lemma 6.5.1 with Theorem 6.4.2 and Lemma 6.1.6, we deduce the following:

COROLLARY 6.5.2. *Let $(N, \mathfrak{d}_N)$ and $(M, \mathfrak{d}_M)$ be metric spaces such that M is complete. Let $\mathfrak{D}$ be the sup metric on $\mathcal{C}_b(N, M)$. Then $(\mathcal{C}_b(N, M), \mathfrak{D})$ is a complete metric space.*

REMARK 6.5.3. Let $I = [0, 1]$. Since I is sequentially compact, all continuous functions $f : I \to \mathbb{R}$ are bounded (see Theorem 5.5.1), so $\mathcal{C}(I, \mathbb{R}) = \mathcal{C}_b(I, \mathbb{R})$. It thus follows from Corollary 6.5.2 that $\mathcal{C}(I, \mathbb{R})$ is complete. □

6.6. Exercises

EXERCISE 6.1. Let $(M, \mathfrak{d})$ be a metric space and let $(x_n)_{n\geq 1}$ be a Cauchy sequence in M. Do the following:

(a) Prove that $(x_n)_{n\geq 1}$ is bounded in the sense that there exists some $C \geq 0$ such that $\mathfrak{d}(x_n, x_m) \leq C$ for all $n, m \geq 1$.
(b) Prove that every subsequence of $(x_n)_{n\geq 1}$ is a Cauchy sequence. □

EXERCISE 6.2. Let $(M, \mathfrak{d})$ be a metric space and let $(x_n)_{n\geq 1}$ be a Cauchy sequence in M. Let $\mathfrak{e} : \mathbb{N} \to \mathbb{R}$ be a function such that $\mathfrak{e}(n) > 0$ for all $n \geq 1$. Prove that $(x_n)_{n\geq 1}$ has a subsequence $(y_n)_{n\geq 1}$ such that $\mathfrak{d}(y_n, y_m) < \mathfrak{e}(n)$ for all $n, m \geq 1$ with $m \geq n$. In the literature, when $\mathfrak{e}(n)$ is chosen to decay quickly to 0 as $n \to \infty$ such sequences $(y_n)_{n\geq 1}$ are often called *fast Cauchy sequences*. Common choices of quickly-decaying $\mathfrak{e}(n)$ are $\mathfrak{e}(n) = 1/2^n$ and $\mathfrak{e}(n) = 1/n$. □

EXERCISE 6.3. Let $(M, \mathfrak{d})$ be a metric space and let $X \subset M$ be a subspace. Assume that X is not a closed subset of M. Prove that X is not complete. □

EXERCISE 6.4. Let $(M, \mathfrak{d})$ be a complete metric space. Define $\mathfrak{d}'$ as follows:

$$\mathfrak{d}'(x, y) = \min\{\mathfrak{d}(x, y), 1\} \quad \text{for } x, y \in M.$$

In Exercise 4.9, you proved that $\mathfrak{d}'$ is a metric and that $(M, \mathfrak{d})$ is homeomorphic to $(M, \mathfrak{d}')$. Prove that $(M, \mathfrak{d}')$ is complete. □

EXERCISE 6.5. Let $(M, \mathfrak{d})$ be a metric space. Do the following:

(a) Let $N_1, \ldots, N_k$ be subspaces of M. Assume that each N_i is complete. Prove that $N_1 \cup \cdots \cup N_k$ is complete.
(b) Let $N_1, N_2, N_3, \ldots$ be subspaces of M. Assume that each N_i is complete. Prove that $\cap_{i=1}^{\infty} N_i$ is complete.
(c) Is (a) true for unions of countably many subspaces? Either prove this or find a counterexample. □

EXERCISE 6.6. Let $(M, \mathfrak{d})$ be a metric space. For $x_0 \in M$ and $r \geq 0$, recall that the closed ball around x_0 of radius r is

$$B_r^{c\ell}(x_0) = \{x \in M \mid \mathfrak{d}(x, x_0) \leq r\}.$$

This is a closed subspace of M (see Exercise 2.9). Do the following:

(a) Assume that there exists some $r > 0$ such that $B_r^{c\ell}(x_0)$ is complete for all $x_0 \in M$. Prove that M is complete.
(b) Give an example of a non-complete metric space $(M, \mathfrak{d})$ such that for all $x_0 \in M$, there exists some $r > 0$ such that $B_r^{c\ell}(x_0)$ is complete. □

EXERCISE 6.7. Let S be an arbitrary set. Recall (see §5.II.1) that $\mathrm{Seq}(S)$ is the set of sequences $\mathbf{s} = (s_m)_{m\geq 1}$ of elements of S. Also, the first-difference metric $\mathfrak{d}$ on $\mathrm{Seq}(S)$ is as follows. Consider $\mathbf{s}, \mathbf{t} \in \mathrm{Seq}(S)$. Write $\mathbf{s} = (s_m)_{m\geq 1}$ and $\mathbf{t} = (t_m)_{m\geq 1}$. If $\mathbf{s} = \mathbf{t}$, then $\mathfrak{d}(\sigma, \tau) = 0$. Otherwise, letting $k = \min\{m \geq 1 \mid s_m \neq t_m\}$ we have $\mathfrak{d}(\mathbf{s}, \mathbf{t}) = 1/k$. Prove that $(\mathrm{Seq}(S), \mathfrak{d})$ is complete. □

EXERCISE 6.8*. Let $(M, \mathfrak{d})$ be a metric space. For a subset $A \subset M$, recall that the diameter of A is

$$\mathrm{diam}(A) = \sup\left(\{\mathfrak{d}(x, y) \mid x, y \in A\} \cup \{0\}\right).$$

This is finite if and only if A is bounded. Prove that M is complete if and only if it satisfies the following condition:

- For all nested sequences of nonempty closed and bounded sets
$$C_1 \supset C_2 \supset C_3 \supset \cdots$$
such that $\lim_{n\to\infty} \operatorname{diam}(C_n) = 0$, the intersection $\cap_{k=1}^{\infty} C_k$ consists of a single point. □

EXERCISE 6.9. Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces. Let $\mathfrak{D}$ be the sup metric on $\mathrm{Map}_b(M, N)$.

(a) Let $f\colon M \to N$ be a bounded function that is not continuous at some $x_0 \in M$. Prove that there exists some $\epsilon > 0$ such that if $g\colon M \to N$ is a bounded function with $\mathfrak{D}(f, g) < \epsilon$, then g is also not continuous at x_0.
(b) Use (a) to give an alternate proof of Lemma 6.5.1, which says that $\mathcal{C}_b(M, N)$ is a closed subset of $\mathrm{Map}_b(M, N)$. Hint: argue that $\mathrm{Map}_b(M, N) \setminus \mathcal{C}_b(M, N)$ is open (see Lemma 2.8.1). □

EXERCISE 6.10. Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces. Endow $\mathcal{C}_b(M, N)$ with the sup metric $\mathfrak{D}$, and give $M \times \mathcal{C}_b(M, N)$ the product metric $\mathfrak{d}_{\text{prod}}$, so
$$\mathfrak{d}_{\text{prod}}((x, f), (y, g)) = \mathfrak{d}_M(x, y) + \mathfrak{D}(f, g) \quad \text{for all } (x, f), (y, g) \in M \times \mathcal{C}_b(M, N).$$
Define $E\colon M \times \mathcal{C}_b(M, N) \to N$ to be the evaluation map, so
$$E(x, f) = f(x) \quad \text{for all } (x, f) \in M \times \mathcal{C}_b(M, N).$$
Prove that E is continuous. □

EXERCISE 6.11. Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces. Recall that a function $f\colon M \to N$ is said to be Lipschitz if there exists some $L \geq 0$ such that $\mathfrak{d}_N(f(x), f(y)) \leq L\, \mathfrak{d}_M(x, y)$ for all $x, y \in M$. All Lipschitz functions are continuous, but the converse is not true. Do the following:

(a) Let $f\colon M \to N$ be a Lipschitz function and let $(x_n)_{n\geq 1}$ be a Cauchy sequence in M. Prove that $(f(x_n))_{n\geq 1}$ is a Cauchy sequence in N.
(b) Give a counterexample to (a) if f is merely assumed to be continuous.
(c) Let $f\colon M \to N$ be a bi-Lipschitz equivalence, i.e., a bijection such that both f and its inverse $g\colon N \to M$ are Lipschitz. Prove that M is complete if and only if N is complete. □

APPENDIX 6.I*

Space-filling curves

CHAPTER 7*

Completions of metric spaces

In this chapter, we prove that every metric space M can be uniquely embedded as a dense subspace of a complete metric space $\widehat{M}$. Since the proof of this is a little more abstract than some of our other arguments and we will not need it in later chapters, we have marked this chapter with an $*$ to indicate that a first-time reader might want to skip it.

7.1. Definition of completion

Let M be a metric space. A *completion* of M is a complete metric space $\widehat{M}$ together with an isometric embedding $\iota\colon M \to \widehat{M}$ whose image is dense. Here we recall that:

- For metric spaces $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$, an isometric embedding from M to N is a map $\iota\colon M \to N$ such that $d_M(x,y) = d_N(\iota(x), \iota(y))$ for all $x, y \in M$. Isometric embeddings are continuous. Moreover, they are necessarily injective; indeed, if $\iota(x) = \iota(y)$, then $d_M(x,y) = d_N(\iota(x), \iota(y)) = 0$, so $x = y$.
- If N is a metric space and $A \subset N$ is a subspace, then A is *dense* in N if the closure $\overline{A}$ of A equals N.

Example 7.1.1. The closed interval $[0,1]$ is a completion of the open interval $(0,1)$. The associated isometric embedding $\iota\colon (0,1) \to [0,1]$ is the inclusion. $\square$

Example 7.1.2. Endow $\mathbb{Q}$ with its Euclidean metric. The real numbers $\mathbb{R}$ are a completion of $\mathbb{Q}$. The associated isometric embedding $\iota\colon \mathbb{Q} \to \mathbb{R}$ is the inclusion. $\square$

Remark 7.1.3. Like in the above examples, for a metric space M we will typically find a completion of M by identifying a complete metric space $\widehat{M}$ that contains M as a dense subspace. The associated isometric embedding $\iota\colon M \to \widehat{M}$ will then be the inclusion. $\square$

7.2. Uniqueness of completion

Later in this chapter we will prove that every metric space has a completion. In this section, we prove that if it exists such a completion is unique up to isometry. This requires a small lemma:

Lemma 7.2.1. *Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces and let $f, g\colon M \to N$ be continuous functions. Let $A \subset M$ be a dense subset, and assume that $f(a) = g(a)$ for all $a \in A$. Then $f = g$.*

Proof. Let $x \in M$. We must prove that $f(x) = g(x)$, i.e., that $\mathfrak{d}_N(f(x), g(x)) = 0$. Letting $\epsilon > 0$ be arbitrary, it is enough to prove that $\mathfrak{d}_N(f(x), g(x)) < \epsilon$. Since f and g are both continuous at x, there exists some $\delta > 0$ such that for all $y \in M$ with $\mathfrak{d}_M(x,y) < \delta$ we have $\mathfrak{d}_N(f(x), f(y)) < \epsilon/2$ and $\mathfrak{d}_N(g(x), g(y)) < \epsilon/2$. Since A is dense in M, we can find $a \in A$ with $\mathfrak{d}_M(x,a) < \delta$. We then have

$$\mathfrak{d}_N(f(x), g(x)) \leq \mathfrak{d}_N(f(x), f(a)) + \mathfrak{d}_N(f(a), g(a)) + \mathfrak{d}_N(g(a), g(x)) < \epsilon/2 + 0 + \epsilon/2 = \epsilon,$$

as desired. $\square$

Our uniqueness result for completions is as follows:

Theorem 7.2.2 (Uniqueness of completions). *Let $(M, \mathfrak{d})$ be a metric space and let $(\widehat{M}_1, \widehat{\mathfrak{d}}_1)$ and $(\widehat{M}_2, \widehat{\mathfrak{d}}_2)$ be two completions of M. Let $\iota_1\colon M \to \widehat{M}_1$ and $\iota_2\colon M \to \widehat{M}_2$ be the corresponding isometric embeddings. There is then a unique isometry $\phi\colon \widehat{M}_1 \to \widehat{M}_2$ such that for all $x \in M$ we have $\phi(\iota_1(x)) = \iota_2(x)$.*

REMARK 7.2.3. If M is given as a subspace of $\widehat{M_1}$ and $\widehat{M_2}$, then the isometry $\phi\colon \widehat{M_1} \to \widehat{M_2}$ in Theorem 7.2.2 will satisfy $\phi(x) = x$ for all $x \in M$. □

PROOF OF THEOREM 7.2.2. The uniqueness of ϕ follows from the fact that the image of ι_1 is dense in $\widehat{M_1}$ (see Lemma 7.2.1). We therefore must only prove existence.

Consider $y \in \widehat{M_1}$. Our goal is to determine $\phi(y) \in \widehat{M_2}$. Since $\iota_1(M)$ is dense in $\widehat{M_1}$, there exists a sequence $(x_n)_{n\geq 1}$ of points of M such that $\lim_{n\to\infty} \iota_1(x_n) = y$. The sequence $(\iota_1(x_n))_{n\geq 1}$ will be a Cauchy sequence in $\widehat{M_1}$, so since ι_1 is an isometric embedding the sequence $(x_n)_{n\geq 1}$ will be a Cauchy sequence in M. Applying the isometric embedding ι_2, we get a Cauchy sequence $(\iota_2(x_n))_{n\geq 1}$ in $\widehat{M_2}$. Since $\widehat{M_2}$ is complete, this has a limit. We would like to define $\phi(y) = \lim_{n\to\infty} \iota_2(x_n) \in \widehat{M_2}$.

To do this, we must prove that this does not depend on the sequence $(x_n)_{n\geq 1}$. This follows from the following claim:

CLAIM. *Let $(x_n)_{n\geq 1}$ and $(x'_n)_{n\geq 1}$ be sequences in M such that $\lim_{n\to\infty} \iota_1(x_n) = \lim_{n\to\infty} \iota_1(x'_n)$. Then $\lim_{n\to\infty} \iota_2(x_n) = \lim_{n\to\infty} \iota_2(x'_n)$.*

PROOF OF CLAIM. The same argument we gave before the claim shows that both $(\iota_2(x_n))_{n\geq 1}$ and $(\iota_2(x'_n))_{n\geq 1}$ are convergent Cauchy sequences. We can thus talk about their limits. Since $\lim_{n\to\infty} \iota_1(x_n) = \lim_{n\to\infty} \iota_1(x'_n)$, we must have

$$\lim_{n\to\infty} \widehat{\mathfrak{d}}_1(\iota_1(x_n), \iota_1(x'_n)) = 0.$$

Since ι_1 and ι_2 are isometric embeddings, we deduce that

$$0 = \lim_{n\to\infty} \widehat{\mathfrak{d}}_1(\iota_1(x_n), \iota_1(x'_n)) = \lim_{n\to\infty} \mathfrak{d}(x_n, x'_n) = \lim_{n\to\infty} \widehat{\mathfrak{d}}_2(\iota_2(x_n), \iota_2(x'_n)).$$

We conclude that $\lim_{n\to\infty} \iota_2(x_n) = \lim_{n\to\infty} \iota_2(x'_n)$, as desired. □

The claim shows that the above recipe gives a well-defined map $\phi\colon \widehat{M_1} \to \widehat{M_2}$. We must prove that ϕ is an isometry such that $\phi(\iota_1(x)) = \iota_2(x)$ for all $x \in M$. We do this in three steps:

STEP 1. *For all $x \in M$, we have $\phi(\iota_1(x)) = \iota_2(x)$.*

Set $y = \iota_1(x)$. To define $\phi(y)$, we must first find a sequence in M whose image under ι_1 converges to y. For this, we can just take the constant sequence $(x)_{n\geq 1}$. Following our recipe, we have

$$\phi(y) = \lim_{n\to\infty} \iota_2(x) = \iota_2(x).$$

STEP 2. *The map $\phi\colon \widehat{M_1} \to \widehat{M_2}$ is an isometric embedding.*

Consider $y, y' \in \widehat{M_1}$. Our goal is to prove that $\widehat{\mathfrak{d}}_1(y, y') = \widehat{\mathfrak{d}}_2(\phi(y), \phi(y'))$. Let $(x_n)_{n\geq 1}$ and $(x'_n)_{n\geq 1}$ be sequences in M such that $(\iota_1(x_n))_{n\geq 1}$ converges to y and $(\iota_1(x'_n))_{n\geq 1}$ converges to y'. Using the fact that ι_1 and ι_2 are isometric embeddings, we have

$$\widehat{\mathfrak{d}}_1(y, y') = \lim_{n\to\infty} \widehat{\mathfrak{d}}_1(\iota_1(x_n), \iota_1(x'_n)) = \lim_{n\to\infty} \mathfrak{d}(x_n, x'_n) = \lim_{n\to\infty} \widehat{\mathfrak{d}}_2(\iota_2(x_n), \iota_2(x'_n)) = \widehat{\mathfrak{d}}_2(\phi(y), \phi(y')).$$

Here the first and last equalities use the fact that the distance function in a metric space is continuous (see Exercise 2.5).

STEP 3. *The map $\phi\colon \widehat{M_1} \to \widehat{M_2}$ is a bijection, and hence an isometry.*

We already know that $\phi\colon \widehat{M_1} \to \widehat{M_2}$ is an isometric embedding such that $\phi(\iota_1(x)) = \iota_2(x)$ for all $x \in M$. Applying the same construction, we can find an isometric embedding $\phi'\colon \widehat{M_2} \to \widehat{M_1}$ such that $\phi'(\iota_2(x)) = \iota_1(x)$ for all $x \in M$.

The composition $\phi' \circ \phi\colon \widehat{M_1} \to \widehat{M_1}$ is therefore an isometric embedding that fixes $\iota_1(x)$ for all $x \in M$. Since the image of ι_1 is dense in $\widehat{M_1}$, this implies that $\phi' \circ \phi = \mathbb{1}_{\widehat{M_1}}$ (see Lemma 7.2.1). Here we recall that $\mathbb{1}_{\widehat{M_1}}$ is the identity of $\widehat{M_1}$. In a similar way, we get that $\phi \circ \phi' = \mathbb{1}_{\widehat{M_2}}$. We conclude that ϕ and ϕ' are inverse isometries. In particular, ϕ is a bijection. □

7.3. Testing completeness on dense subspaces

Before we can prove that every metric space has a completion, we must prove a lemma:

LEMMA 7.3.1. *Let $(X, \mathfrak{d})$ be a metric space and let $A \subset X$ be a dense subspace. Assume that every Cauchy sequence in A converges to a point of X. Then X is complete.*

PROOF. Let $(x_n)_{n\geq 1}$ be a Cauchy sequence in X. Our goal is to prove that $(x_n)_{n\geq 1}$ converges. Since A is dense in X, for each $n \geq 1$ we can find some $a_n \in A$ with $\mathfrak{d}(a_n, x_n) < 1/n$. We then have:

CLAIM. *The sequence $(a_n)_{n\geq 1}$ is a Cauchy sequence.*

PROOF OF CLAIM. Consider $\epsilon > 0$. Since $(x_n)_{n\geq 1}$ is a Cauchy sequence, we can find some $N_1 \geq 1$ such that for $n, m \geq N_1$ we have $\mathfrak{d}(x_n, x_m) < \epsilon/3$. Choose $N_2 \geq 1$ such that for $n \geq N_2$ we have $1/n < \epsilon/3$, and set $N = \max(N_1, N_2)$. For $n, m \geq N$, we have

$$\mathfrak{d}(a_n, a_m) \leq \mathfrak{d}(a_n, x_n) + \mathfrak{d}(x_n, x_m) + \mathfrak{d}(x_m, a_m) < \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon,$$

as desired. □

Since every Cauchy sequence in A converges to a point of X, the claim implies that $(a_n)_{n\geq 1}$ is a convergent sequence in X. Set $z = \lim_{n\to\infty} a_n$. We claim that $\lim_{n\to\infty} x_n = z$. To see this, note that for $n \geq 1$ we have

$$0 \leq \mathfrak{d}(x_n, z) \leq \mathfrak{d}(x_n, a_n) + \mathfrak{d}(a_n, z) \leq 1/n + \mathfrak{d}(a_n, z).$$

Since $(a_n)_{n\geq 1}$ converges to z, the right-hand side of this converges to $0 + 0 = 0$. This implies that $\lim_{n\to\infty} \mathfrak{d}(x_n, z) = 0$, so $\lim_{n\to\infty} x_n = z$, as desired. □

7.4. Existence of completions

Our final goal in this chapter is to prove that every metric space has a completion. The construction uses the notion of an equivalence relation, which we discuss in Appendix 3.I. Our theorem is as follows:

THEOREM 7.4.1 (Existence of completions). *Every metric space $(M, \mathfrak{d})$ has a completion.*

EXAMPLE 7.4.2. Though $\mathbb{Z}$ with its Euclidean metric is discrete, it has other interesting metrics that reflect its arithmetic properties. Let p be a prime. For a nonzero $x \in \mathbb{Z}$, there exists a unique $\nu_p(x) \in \mathbb{Z}_{\geq 0}$ such that $x = p^{\nu_p(x)} y$ with $y \in \mathbb{Z}$ not divisible by p. The function $\nu_p \colon \mathbb{Z} \setminus \{0\} \to \mathbb{Z}_{\geq 0}$ is called the *p-adic valuation*. Set

$$\|x\|_p = \begin{cases} p^{-\nu_p(x)} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

This is called the *p-adic norm*. Define $\mathfrak{d}_p \colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$ via the formula $\mathfrak{d}_p(x, y) = \|x - y\|_p$. This is a metric (see Exercise 7.3*), and is called the *p-adic metric* on $\mathbb{Z}$. An integer $x \in \mathbb{Z}$ is close to 0 in the p-adic metric if x is divisible by a large power of p. In particular, with respect to the p-adic metric we have $\lim_{k\to\infty} p^k = 0$. Even better, the partial sums $\sigma_n = \sum_{k=1}^n p^k$ form a Cauchy sequence $(\sigma_n)_{n\geq 1}$ in $\mathbb{Z}$ (see Exercise 7.3*). If $p = 2$ this Cauchy sequence converges to -2 (see Exercise 7.3*), but for $p \geq 3$ it does not have a limit in $\mathbb{Z}$. Its limit lies in the completion of $\mathbb{Z}$ with respect to the metric $\mathfrak{d}_p$.

The metric completion of $\mathbb{Z}$ with respect to $\mathfrak{d}_p$ is called the *p-adic integers*, and[1] is denoted $\mathbb{Z}_p$. Identifying $\mathbb{Z}$ with its image in $\mathbb{Z}_p$, the multiplication and addition operations on $\mathbb{Z}$ extend to continuous multiplication and addition operators on $\mathbb{Z}_p$ (see Exercise 7.4*). This makes $\mathbb{Z}_p$ into what is called a topological ring. □

REMARK 7.4.3. It is tempting to use Theorem 7.4.1 to construct $\mathbb{R}$ as the completion of $\mathbb{Q}$ under its Euclidean metric. However, this would be circular since the existence of $\mathbb{R}$ is needed for our definition of a metric space. The proof of Theorem 7.4.1 is inspired by one of the standard constructions of $\mathbb{R}$ from $\mathbb{Q}$. □

PROOF OF THEOREM 7.4.1. Let $(M, \mathfrak{d})$ be a metric space. We start with the following:

[1] It is common to write $\mathbb{Z}_p$ for the cyclic group C_p of order p or for the ring $\mathbb{Z}/p$, but we avoid this since we believe that $\mathbb{Z}_p$ should always be p-adic integers.

CLAIM 1. *Let $(x_n)_{n\geq 1}$ and $(y_n)_{n\geq 1}$ be Cauchy sequences in M. Then $(\mathfrak{d}(x_n, y_n))_{n\geq 1}$ is a convergent sequence of real numbers.*

PROOF OF CLAIM. Since $\mathbb{R}$ is complete (Theorem 1.4.2), it is enough to prove that $(\mathfrak{d}(x_n, y_n))_{n\geq 1}$ is a Cauchy sequence. Consider $\epsilon > 0$. We must find some $N \geq 1$ such that for $n, m \geq N$ we have

$$|\,\mathfrak{d}(x_n, y_n) - \mathfrak{d}(x_m, y_m)| < \epsilon. \tag{7.4.1}$$

Since $(x_n)_{n\geq 1}$ and $(y_n)_{n\geq 1}$ are both Cauchy sequences, we can find some $N \geq 1$ such that for $n, m \geq N$ we have $\mathfrak{d}(x_n, x_m) < \epsilon/2$ and $\mathfrak{d}(y_n, y_m) < \epsilon/2$. For $n, m \geq N$, we then have

$$\mathfrak{d}(x_n, y_n) \leq \mathfrak{d}(x_n, x_m) + \mathfrak{d}(x_m, y_m) + \mathfrak{d}(y_m, y_n) < \epsilon/2 + \mathfrak{d}(x_m, y_m) + \epsilon/2.$$

It follows that $\mathfrak{d}(x_n, y_n) - \mathfrak{d}(x_m, y_m) < \epsilon$. Similarly, we have $\mathfrak{d}(x_m, y_m) - \mathfrak{d}(x_n, y_n) < \epsilon$. The inequality (7.4.1) follows. □

Let $C(M)$ be the set of Cauchy sequences $\sigma = (x_n)_{n\geq 1}$ in M. By Claim 1, we can define a map $\widehat{\mathfrak{d}}_0 \colon C(M) \times C(M) \to \mathbb{R}$ as follows:

- For Cauchy sequences $\sigma = (x_n)_{n\geq 1}$ and $\tau = (y_n)_{n\geq 1}$ of M, set $\widehat{\mathfrak{d}}_0(\sigma, \tau) = \lim_{n\to\infty} \mathfrak{d}(x_n, y_n)$.

This almost defines a metric on $C(M)$:

CLAIM 2. *The following hold:*

(i) For $\sigma, \tau \in C(M)$, we have $\widehat{\mathfrak{d}}_0(\sigma, \tau) \geq 0$.
(ii) For $\sigma, \tau \in C(M)$, we have $\widehat{\mathfrak{d}}_0(\sigma, \tau) = \widehat{\mathfrak{d}}_0(\tau, \sigma)$.
(iii) For $\sigma, \tau, \mu \in C(M)$, we have $\widehat{\mathfrak{d}}_0(\sigma, \tau) \leq \widehat{\mathfrak{d}}_0(\sigma, \mu) + \widehat{\mathfrak{d}}_0(\mu, \tau)$.

PROOF OF CLAIM. Conclusions (i) and (ii) are immediate from the definitions. As for (iii), write $\sigma = (x_n)_{n\geq 1}$ and $\tau = (y_n)_{n\geq 1}$ and $\mu = (z_n)_{n\geq 1}$. We then have

$$\widehat{\mathfrak{d}}_0(\sigma, \tau) = \lim_{n\to\infty} \mathfrak{d}(x_n, y_n) \leq \lim_{n\to\infty} (\mathfrak{d}(x_n, z_n) + \mathfrak{d}(z_n, y_n)) = \widehat{\mathfrak{d}}_0(\sigma, \mu) + \widehat{\mathfrak{d}}_0(\mu, \tau). \qquad \square$$

The only property of a metric that is missing is $\widehat{\mathfrak{d}}_0(\sigma, \tau) = 0$ if and only if $\sigma = \tau$, which need not hold. The issue is that different Cauchy sequences in M can be asymptotic, and thus should represent the same point of the completion we are constructing. The solution is to quotient $C(M)$ by an appropriate equivalence relation:

CLAIM 3. *Let the notation be as above. For $\sigma, \tau \in C(M)$, let $\sigma \sim \tau$ if $\widehat{\mathfrak{d}}_0(\sigma, \tau) = 0$. Then the relation $\sim$ is an equivalence relation.*

PROOF OF CLAIM. There are three conditions we must check:

- Reflexivity. Consider $\sigma = (x_n)_{n\geq 1}$ in $C(M)$. We must prove that $\widehat{\mathfrak{d}}_0(\sigma, \sigma) = 0$. Indeed, $\widehat{\mathfrak{d}}_0(\sigma, \sigma) = \lim_{n\to\infty} \mathfrak{d}(x_n, x_n) = 0$.
- Symmetry. Consider $\sigma, \tau \in C(M)$ with $\sigma \sim \tau$. We must prove that $\tau \sim \sigma$. Since $\sigma \sim \tau$, we have $\widehat{\mathfrak{d}}_0(\sigma, \tau) = 0$. By Claim 2.(ii), we have $\widehat{\mathfrak{d}}_0(\tau, \sigma) = \widehat{\mathfrak{d}}_0(\sigma, \tau) = 0$, so $\tau \sim \sigma$.
- Transitivity. Consider $\sigma, \tau, \mu \in C(M)$ with $\sigma \sim \tau$ and $\tau \sim \mu$. We must prove that $\sigma \sim \mu$. Since $\sigma \sim \tau$ and $\tau \sim \mu$, we have $\widehat{\mathfrak{d}}_0(\sigma, \tau) = \widehat{\mathfrak{d}}_0(\tau, \mu) = 0$. By Claims 2.(i) and 2.(iii), we have
$$0 \leq \widehat{\mathfrak{d}}_0(\sigma, \mu) \leq \widehat{\mathfrak{d}}_0(\sigma, \tau) + \widehat{\mathfrak{d}}_0(\tau, \mu) = 0 + 0 = 0,$$
so $\widehat{\mathfrak{d}}_0(\sigma, \mu) = 0$ and $\sigma \sim \mu$. □

Let $\sim$ be the equivalence relation from Claim 3. Define $\widehat{M} = C(M)/\sim$. We wish to define a metric $\widehat{\mathfrak{d}}$ on $\widehat{M}$. The key is the following:

CLAIM 4. *Let the notation be as above. Let $\sigma, \sigma', \tau, \tau' \in C(M)$ be such that $\sigma \sim \sigma'$ and $\tau \sim \tau'$. Then $\widehat{\mathfrak{d}}_0(\sigma, \tau) = \widehat{\mathfrak{d}}_0(\sigma', \tau')$.*

PROOF OF CLAIM. Using Claim 2.(iii), we have

$$\widehat{\mathfrak{d}}_0(\sigma, \tau) \leq \widehat{\mathfrak{d}}_0(\sigma, \sigma') + \widehat{\mathfrak{d}}_0(\sigma', \tau) = 0 + \widehat{\mathfrak{d}}_0(\sigma', \tau),$$

so $\widehat{\mathfrak{d}}_0(\sigma, \tau) \leq \widehat{\mathfrak{d}}_0(\sigma', \tau)$. The same argument shows that $\widehat{\mathfrak{d}}_0(\sigma', \tau) \leq \widehat{\mathfrak{d}}_0(\sigma, \tau)$, so $\widehat{\mathfrak{d}}_0(\sigma, \tau) = \widehat{\mathfrak{d}}_0(\sigma', \tau)$. Applying this same proof to the second input of $\widehat{\mathfrak{d}}_0(\sigma', \tau)$, we get that $\widehat{\mathfrak{d}}_0(\sigma', \tau) = \widehat{\mathfrak{d}}_0(\sigma', \tau')$. Combining all of this, we conclude that $\widehat{\mathfrak{d}}_0(\sigma, \tau) = \widehat{\mathfrak{d}}_0(\sigma', \tau')$. □

For $[\sigma], [\tau] \in \widehat{M}$, define $\widehat{\mathfrak{d}}([\sigma], [\tau]) = \widehat{\mathfrak{d}}_0(\sigma, \tau)$. Claim 4 implies that this is well-defined; indeed, we have $[\sigma] = [\sigma']$ precisely when $\sigma \sim \sigma'$, and similarly for τ. It is immediate from Claim 2 and the definition of $\sim$ that $\widehat{\mathfrak{d}}$ is a metric.

The next thing we must prove is as follows:

Claim 5. *There is an isometric embedding $\iota\colon M \to \widehat{M}$ whose image is dense.*

Proof of claim. Define a map $\iota\colon M \to \widehat{M}$ as follows. Consider $x \in M$. Let $\mathfrak{c}_x$ be the constant Cauchy sequence $(x_n)_{n\geq 1}$ with $x_n = x$ for all $n \geq 1$. Define $\iota(x) = [\mathfrak{c}_x]$. We must prove two things:

(a) The map $\iota\colon M \to \widehat{M}$ is an isometric embedding, that is, an injection such that for all $x, y \in M$ we have $\widehat{\mathfrak{d}}(\iota(x), \iota(y)) = \mathfrak{d}(x, y)$.
(b) The image of ι is a dense subspace of $\widehat{M}$.

We first check (a). Consider $x, y \in M$. By definition,

$$\widehat{\mathfrak{d}}(\iota(x), \iota(y)) = \widehat{\mathfrak{d}}([\mathfrak{c}_x], [\mathfrak{c}_y]) = \widehat{\mathfrak{d}}_0(\mathfrak{c}_x, \mathfrak{c}_y) = \lim_{n\to\infty} \mathfrak{d}(x, y) = \mathfrak{d}(x, y).$$

This calculation also shows that ι is injective: if $\iota(x) = \iota(y)$ then $0 = \widehat{\mathfrak{d}}(\iota(x), \iota(y)) = \mathfrak{d}(x, y)$, so $x = y$.

We conclude by checking (b). Consider $[\sigma] \in \widehat{M}$ and $\epsilon > 0$. We must prove that there is some $z \in M$ such that $\widehat{\mathfrak{d}}(\iota(z), [\sigma]) < \epsilon$. Write $\sigma = (w_n)_{n\geq 1}$. This is a Cauchy sequence, so we can find $N \geq 1$ such that for $n, m \geq N$ we have $\mathfrak{d}(w_n, w_m) < \epsilon/2$. Set $z = w_N$. We then have

$$\widehat{\mathfrak{d}}(\iota(z), [\sigma]) = \widehat{\mathfrak{d}}([\mathfrak{c}_{w_N}], [(w_n)_{n\geq 1}]) = \widehat{\mathfrak{d}}_0(\mathfrak{c}_{w_N}, (w_n)_{n\geq 1}) = \lim_{n\to\infty} \mathfrak{d}(w_N, w_n).$$

For $n \geq N$, the term $\mathfrak{d}(w_N, w_n)$ is less than $\epsilon/2$. It follows that $\lim_{n\to\infty} \mathfrak{d}(w_N, w_n) \leq \epsilon/2 < \epsilon$. □

We are almost done. The only remaining thing we must prove is that $\widehat{M}$ is complete. We recall the statement of Lemma 7.3.1:

- Let $(X, \mathfrak{d}_X)$ be a metric space and let $A \subset X$ be a dense subspace. Assume that every Cauchy sequence in A converges to a point of X. Then X is complete.

Since $\iota\colon M \to \widehat{M}$ is an isometric embedding with dense image, this lemma reduces us to proving the following:

Claim 6. *Let $\sigma = (x_n)_{n\geq 1}$ be a Cauchy sequence in M. Then $\lim_{m\to\infty} \iota(x_m) = [\sigma]$.*

Proof of claim. It is enough to prove that $\lim_{m\to\infty} \widehat{\mathfrak{d}}(\iota(x_m), [\sigma]) = 0$. Observe that

$$\lim_{m\to\infty} \widehat{\mathfrak{d}}(\iota(x_m), [\sigma]) = \lim_{m\to\infty} \widehat{\mathfrak{d}}_0(\mathfrak{c}_{x_m}, (x_n)_{n\geq 1}) = \lim_{m\to\infty} \lim_{n\to\infty} \mathfrak{d}(x_m, x_n).$$

This limit is 0. Indeed, since $(x_n)_{n\geq 1}$ is a Cauchy sequence, for all $\epsilon > 0$ we can find some $N \geq 1$ such that for $n, m \geq N$ we have $\mathfrak{d}(x_m, x_n) < \epsilon$. This implies that the above limit is at most ϵ. Since $\epsilon > 0$ was arbitrary and $\mathfrak{d}(x_m, x_n) \geq 0$ for all $n, m \geq 1$, the limit must be 0, as desired. □

This completes the proof that every metric space has a completion. □

7.5. Exercises

Exercise 7.1. Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces and $f\colon M \to N$ be a uniformly continuous function. Assume that N is complete and that M is a dense subspace of a metric space $(M', \mathfrak{d}_{M'})$. Do the following:

(a) Prove that there is a unique continuous function $F\colon M' \to N$ such that $F(x) = f(x)$ for all $x \in M$.
(b) If f is Lipschitz, prove that F is Lipschitz. □

Exercise 7.2. Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces. Give $M \times N$ the product metric $\mathfrak{d}_{\text{prod}}$. Do the following:

(a) Assume that M and N are both complete. Prove that $M \times N$ is complete.
(b) Assume that $A \subset M$ is a dense subset and $B \subset N$ is a dense subset. Prove that $A \times B$ is dense in $M \times N$. □

EXERCISE 7.3*. Let p be a prime. Recall that the p-adic metric $\mathfrak{d}_p$ on $\mathbb{Z}$ is as follows. For a nonzero $x \in \mathbb{Z}$, there exists a unique $\nu_p(x) \in \mathbb{Z}_{\geq 0}$ such that $x = p^{\nu_p(x)}y$ with $y \in \mathbb{Z}$ not divisible by p. The function ν_p is called the p-adic valuation. Set

$$\|x\|_p = \begin{cases} p^{-\nu_p(x)} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

This is called the p-adic norm. Define $\mathfrak{d}_p \colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$ via the formula $\mathfrak{d}_p(x, y) = \|x - y\|_p$. This is called the p-adic metric. Do the following:

(a) Prove that the p-adic norm satisfies the following inequality, which is called the *ultrametric inequality*:
$$\|x + y\|_p \leq \max\{\|x\|_p, \|y\|_p\} \quad \text{for all } x, y \in \mathbb{Z}.$$
Deduce that the p-adic metric $\mathfrak{d}_p$ is a metric.

(b) Prove that the p-adic norm is multiplicative in the sense that $\|xy\|_p = \|x\|_p\|y\|_p$ for all $x, y \in \mathbb{Z}$.

(c) Let $(x_k)_{k\geq 1}$ be a sequence of integers such that $\lim_{k\to\infty} x_k = 0$ in the p-adic metric. For instance, we might have $x_k = p^k$. For $n \geq 1$, set $\sigma_n = \sum_{k=1}^n x_k$. Prove that $(\sigma_n)_{n\geq 1}$ is a Cauchy sequence of integers in the p-adic metric.[2] The completion $\mathbb{Z}_p$ of $\mathbb{Z}$ with respect to this metric thus contains an element $\sum_{k=1}^\infty x_k$.

(d) Identify $\mathbb{Z}$ with its image in $\mathbb{Z}_p$. If $p = 2$, then prove that $\sum_{k=1}^\infty p^k = -2$. However, if $p \geq 3$ then prove that $\sum_{k=1}^\infty p^k$ does not lie in $\mathbb{Z}$. □

EXERCISE 7.4*. Let p be a prime. As in Exercise 7.3*, let $\mathfrak{d}_p$ be the p-adic metric on $\mathbb{Z}$. Let $(\mathbb{Z}_p, \widehat{\mathfrak{d}}_p)$ be the completion of $\mathbb{Z}$ with respect to the p-adic metric. This comes with an isometric embedding $\iota \colon \mathbb{Z} \to \mathbb{Z}_p$ with dense image, and we will identify $\mathbb{Z}$ with its image in $\mathbb{Z}_p$. Do the following:

(a) Give $\mathbb{Z} \times \mathbb{Z}$ its p-adic product metric $\mathfrak{d}_{\text{prod}}$, so
$$\mathfrak{d}_{\text{prod}}((x, y), (z, w)) = \mathfrak{d}_p(x, z) + \mathfrak{d}_p(y, w) \quad \text{for } (x, y), (z, w) \in \mathbb{Z}^2.$$
Prove that the addition map $a \colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ defined by $a(x, y) = x + y$ is Lipschitz with respect to the p-adic metric.

(b) Prove that $a \colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ extends uniquely to a continuous map $\overline{a} \colon \mathbb{Z}_p \times \mathbb{Z}_p \to \mathbb{Z}_p$. Hint: Exercise 7.2 will be useful here.

(c) Prove that the multiplication map $m \colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ defined by $m(x, y) = xy$ is Lipschitz with respect to the p-adic metric.

(d) Prove that $m \colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ extends uniquely to a continuous map $\overline{m} \colon \mathbb{Z}_p \times \mathbb{Z}_p \to \mathbb{Z}_p$.

(e) Use $\overline{a}$ and $\overline{m}$ to define addition and multiplication in $\mathbb{Z}_p$, so for $x, y \in \mathbb{Z}_p$ we have
$$x + y = \overline{a}(x, y),$$
$$xy = \overline{m}(x, y).$$
Prove that these operations make $\mathbb{Z}_p$ into a ring. Hint: Use the fact that the ring axioms hold in $\mathbb{Z}$ and $\mathbb{Z}$ is dense in $\mathbb{Z}_p$. To verify that additive inverses exist, for $x \in \mathbb{Z}_p$ prove that $(-1)x$ is the additive inverse of x. Here $(-1)x$ is the product of $-1 \in \mathbb{Z} \subset \mathbb{Z}_p$ and $x \in \mathbb{Z}_p$ using the multiplication in $\mathbb{Z}_p$. □

[2] This should be a little surprising since such a thing would be false for series of real numbers. For instance, $\sum_{k=1}^\infty 1/k$ is a divergent series even though its terms approach 0 in $\mathbb{R}$.

CHAPTER 8*

The Baire category theorem for metric spaces

CHAPTER 9*

The Banach fixed point theorem

Our next topic is the Banach fixed point theorem along with two of its classic applications: the inverse function theorem, and the Picard–Lindelöf theorem giving solutions to first-order ordinary differential equations.

9.1. Contraction mappings and the Banach fixed point theorem

Let $(M, \mathfrak{d})$ be a metric space. A function $T\colon M \to M$ is a *contraction mapping* if there exists a constant $0 \leq C < 1$ such that $\mathfrak{d}(T(x), T(y)) \leq C\,\mathfrak{d}(x, y)$ for all $x, y \in M$. This implies that T is continuous.

EXAMPLE 9.1.1. Endow $\mathbb{R}^d$ with its Euclidean metric $\mathfrak{d}$. Fix $\mathbf{x}_0 \in \mathbb{R}^d$ and $0 \leq C < 1$. We can then define a contraction mapping $T\colon \mathbb{R}^d \to \mathbb{R}^d$ via the formula

$$T(\mathbf{x}) = \mathbf{x}_0 + C(\mathbf{x} - \mathbf{x}_0) \quad \text{for } \mathbf{x} \in \mathbb{R}^d.$$

The fact that this is a contraction mapping follows from the following calculation: for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, we have

$$\begin{aligned}\mathfrak{d}(T(\mathbf{x}), T(\mathbf{y})) &= \|T(\mathbf{x}) - T(\mathbf{y})\| = \|(\mathbf{x}_0 + C(\mathbf{x} - \mathbf{x}_0)) - (\mathbf{x}_0 + C(\mathbf{y} - \mathbf{x}_0))\| \\ &= \|C\mathbf{x} - C\mathbf{y}\| = C\|\mathbf{x} - \mathbf{y}\| = C\,\mathfrak{d}(\mathbf{x}, \mathbf{y}).\end{aligned} \qquad \square$$

A *fixed point* of a function $T\colon M \to M$ is a point $x_0 \in M$ such that $T(x_0) = x_0$. In Example 9.1.1, the point $\mathbf{x}_0 \in \mathbb{R}^d$ is a fixed point for T. It turns out that if M is complete, then all contraction mappings $T\colon M \to M$ have a unique fixed point:

THEOREM 9.1.2 (Banach fixed point theorem). *Let $(M, \mathfrak{d})$ be a complete nonempty metric space and let $T\colon M \to M$ be a contraction mapping. Then T has a unique fixed point.*

PROOF. Let $0 \leq C < 1$ be such that $\mathfrak{d}(T(x), T(y)) \leq C\,\mathfrak{d}(x, y)$ for all $x, y \in M$. We start by observing that if a fixed point exists, then it is necessarily unique. Indeed, if $x_0, x_0' \in M$ are distinct fixed points, then

$$\mathfrak{d}(x_0, x_0') = \mathfrak{d}(T(x_0), T(x_0')) \leq C\,\mathfrak{d}(x_0, x_0').$$

Since $x_0 \neq x_0'$, we can divide by $\mathfrak{d}(x_0, x_0') \neq 0$ and see that $1 \leq C$, contradicting the fact that $C < 1$.

We now turn to constructing a fixed point. Choose some arbitrary point $x_1 \in M$. We then define a sequence $\{x_n\}_{n \geq 1}$ of points of M by repeatedly applying T to x_1:

$$x_2 = T(x_1), \quad x_3 = T(x_2), \quad x_4 = T(x_3), \quad \text{etc.}$$

In general, for $n \geq 1$ we have $x_{n+1} = T(x_n)$. In the following two claims, we will prove that $\{x_n\}_{n \geq 1}$ converges to a fixed point of T:

CLAIM. *The sequence $\{x_n\}_{n \geq 1}$ converges.*

Since M is complete, it is enough to prove that $\{x_n\}_{n \geq 1}$ is a Cauchy sequence. Consider $\epsilon > 0$. We must find some $N \geq 1$ such that for $n, m \geq N$ we have $\mathfrak{d}(x_n, x_m) < \epsilon$. In fact, it is enough to prove that this holds when $m > n \geq N$. For $k \geq 2$, we have

$$\mathfrak{d}(x_k, x_{k+1}) = \mathfrak{d}(T(x_{k-1}), T(x_k)) \leq C\,\mathfrak{d}(x_{k-1}, x_k).$$

Iterating this, we see that

$$\mathfrak{d}(x_k, x_{k+1}) \leq C^{k-1}\,\mathfrak{d}(x_1, x_2) \quad \text{for } k \geq 1.$$

For $n, m \geq 1$ with $m > n$, we deduce that

$$\mathfrak{d}(x_n, x_m) \leq \mathfrak{d}(x_n, x_{n+1}) + \mathfrak{d}(x_{n+1}, x_{n+2}) + \cdots + \mathfrak{d}(x_{m-1}, x_m) \leq \sum_{k=n}^{\infty} \mathfrak{d}(x_k, x_{k+1}) \leq \sum_{k=n}^{\infty} C^{k-1}\, \mathfrak{d}(x_1, x_2). \tag{9.1.1}$$

Here the second inequality holds since each term of this sum is nonnegative. Since $0 \leq C < 1$, the sum $\sum_{k=1}^{\infty} C^{k-1}\, \mathfrak{d}(x_1, x_2)$ is a convergent geometric series. It follows that there exists $N \geq 1$ such that for $n \geq N$ we have $\sum_{k=n}^{\infty} C^{k-1}\, \mathfrak{d}(x_1, x_2) < \epsilon$. Combining this with (9.1.1), we conclude that for $n, m \geq N$ with $m > n$ we have $\mathfrak{d}(x_n, x_m) < \epsilon$, as desired.

Claim. *The point $x_0 = \lim_{n\to\infty} x_n$ is a fixed point for T.*

We must prove that $\mathfrak{d}(T(x_0), x_0) = 0$. Fixing $\epsilon > 0$, it is enough to prove that $\mathfrak{d}(T(x_0), x_0) < \epsilon$. Choose $N \geq 1$ such that for all $n \geq N$ we have $\mathfrak{d}(x_n, x_0) < \epsilon/3$ and $\mathfrak{d}(x_{n+1}, x_n) < \epsilon/3$. For $n \geq N$, since $T(x_n) = x_{n+1}$ we have

$$\begin{aligned}\mathfrak{d}(T(x_0), x_0) &\leq \mathfrak{d}(T(x_0), T(x_n)) + \mathfrak{d}(T(x_n), x_n) + \mathfrak{d}(x_n, x_0)\\ &\leq C\, \mathfrak{d}(x_0, x_n) + \mathfrak{d}(x_{n+1}, x_n) + \mathfrak{d}(x_n, x_0) < C\epsilon/3 + \epsilon/3 + \epsilon/3.\end{aligned}$$

Since $C < 1$, this is at most ϵ, as desired. □

9.2. Application: inverse function theorem

9.3. Application: solving ordinary differential equations

We now give a classic application of the Banach fixed point theorem. A first-order ordinary differential equation (ODE) is an equation involving a function and its first derivative. To solve it, we must find the unknown function. It is an initial value problem if the value of the function at a single point is specified. Here is an example:

Example 9.3.1. On the closed interval $[0, 1]$, consider the first-order differential equation $f'(x) = 2f(x) + 6$ with initial value $f(0) = 5$. A standard ODE computation shows that all solutions to $f'(x) = 2f(x) + 6$ are of the form $f(x) = Ce^{2x} - 3$ for some constant $C \in \mathbb{R}$. Plugging in $f(0) = 5$, we get $C = 8$, so the solution to this initial value problem is $f(x) = 8e^{2x} - 3$. □

It is unusual that there is an exact formula for the solution of the initial value problem in Example 9.3.1. For more typical initial value problems, there is no solution that can be expressed via a formula and the best one can hope for is an existence theorem together with tools to approximate the solution. The Picard–Lindelöf theorem says that under mild conditions such initial value problems have unique solutions, at least locally. Moreover, its proof is explicit enough that it is easy to follow it to construct approximations to the solution via what is called Picard iteration.

Let $U \subset \mathbb{R}^2$ be open. A function $G\colon U \to \mathbb{R}$ is *locally Lipschitz* in its second variable if for all $(x_0, y_0) \in U$, there exists an open set $V \subset U$ with $(x_0, y_0) \in V$ and some $L \geq 0$ such that

$$|G(x, y) - G(x, y')| \leq L|y - y'| \quad \text{for all } (x, y), (x, y') \in V.$$

This holds, for instance, if G is continuously differentiable (see Exercise 9.1). We then have:

Theorem 9.3.2 (Picard–Lindelöf theorem). *Let $U \subset \mathbb{R}^2$ be open and let $G\colon U \to \mathbb{R}$ be continuous and locally Lipschitz in its second variable. Then for all $(x_0, y_0) \in U$, there exists $\delta > 0$ such that the following holds. There exists a unique differentiable function $f\colon [x_0 - \delta, x_0 + \delta] \to \mathbb{R}$ with $(x, f(x)) \in U$ for all $x \in [x_0 - \delta, x_0 + \delta]$ such that $f(x_0) = y_0$ and $f'(x) = G(x, f(x))$ for all $x \in [x_0 - \delta, x_0 + \delta]$.*

Remark 9.3.3. If G is continuous but not locally Lipschitz, then solutions to $f'(x) = G(x, f(x))$ still exist locally but might no longer be unique. This is the Peano existence theorem; see [**1**]. □

Proof of Theorem 9.3.2. To avoid technicalities, we will prove the theorem in the special case where $U = \mathbb{R}^2$ and for some fixed $L \geq 0$ we have

$$|G(x, y) - G(x, y')| \leq L|y - y'| \quad \text{for all } (x, y), (x, y') \in \mathbb{R}^2.$$

The general case is similar, but requires a little more care (see Exercise 9.2). Choose $\delta > 0$ such that $\delta L < 1$. Set $J = [x_0 - \delta, x_0 + \delta]$. By the fundamental theorem of calculus, a continuous function $f\colon J \to \mathbb{R}$ is differentiable and satisfies $f(x_0) = y_0$ and $f'(x) = G(x, f(x))$ if and only if it satisfies the integral equation

$$f(t) = y_0 + \int_{x_0}^{t} G(x, f(x))\,\mathrm{dx} \quad \text{for all } t \in J. \tag{9.3.1}$$

Consider the set $\mathcal{C}(J,\mathbb{R})$ of continuous functions $f\colon J \to \mathbb{R}$. Equip $\mathcal{C}(J,\mathbb{R})$ with the sup metric:

$$\mathfrak{D}(f,g) = \sup\{|f(t) - g(t)| \mid t \in J\}.$$

Since $J = [x_0 - \delta, x_0 + \delta]$ is a sequentially compact, each $f \in \mathcal{C}(J,\mathbb{R})$ is bounded (Theorem 1.11.2). It follows that $\mathcal{C}(J,\mathbb{R}) = \mathcal{C}_b(J,\mathbb{R})$, so by Corollary 6.5.2 the metric $\mathfrak{D}$ makes $\mathcal{C}(J,\mathbb{R})$ into a complete metric space. We can therefore apply the Banach fixed point theorem (Theorem 9.1.2) to it.

Define a function $T\colon \mathcal{C}(J,\mathbb{R}) \to \mathcal{C}(J,\mathbb{R})$ as follows:

- For a continuous function $f\colon J \to \mathbb{R}$, define $T(f)\colon J \to \mathbb{R}$ to be the function

$$T(f)(t) = y_0 + \int_{x_0}^{t} G(x, f(x))\,\mathrm{dx} \quad \text{for all } t \in J = [x_0 - \delta, x_0 + \delta].$$

A continuous function $f\colon J \to \mathbb{R}$ satisfies (9.3.1) if and only if it is a fixed point of T, i.e., it satisfies $T(f) = f$. Our goal, therefore, is to prove that T has a unique fixed point. By the Banach fixed point theorem (Theorem 9.1.2), to prove this it is enough to prove that T is a contraction mapping. We remark that the proof of the Banach fixed point theorem shows that we can approximate a fixed point by repeatedly applying T to any fixed initial $f\colon J \to \mathbb{R}$. This process is called *Picard iteration.*

Consider continuous functions $f, g\colon J \to \mathbb{R}$. We estimate $\mathfrak{D}(T(f), T(g))$ as follows. Consider $t \in J = [x_0 - \delta, x_0 + \delta]$. Assume first that $t \in [x_0, x_0 + \delta]$. We have

$$\begin{aligned}
|T(f)(t) - T(g)(t)| =& \left| \left(y_0 + \int_{x_0}^{t} G(x, f(x))\,\mathrm{dx}\right) - \left(y_0 + \int_{x_0}^{t} G(x, g(x))\,\mathrm{dx}\right) \right| \\
\leq& \int_{x_0}^{t} |G(x, f(x)) - G(x, g(x))|\,\mathrm{dx} \leq L \int_{x_0}^{t} |f(x) - g(x)|\,\mathrm{dx} \\
\leq& L \int_{x_0}^{t} \mathfrak{D}(f,g)\,\mathrm{dx} = L(t - x_0)\,\mathfrak{D}(f,g) \leq L\delta\,\mathfrak{D}(f,g).
\end{aligned}$$

Set $C = L\delta$, which by our assumptions on δ is less than 1. The above implies that

$$|T(f)(t) - T(g)(t)| \leq C\,\mathfrak{D}(f,g) \quad \text{for } t \in [x_0, x_0 + \delta].$$

A similar argument shows $|T(f)(t) - T(g)(t)| \leq C\,\mathfrak{D}(f,g)$ also holds for $t \in [x_0 - \delta, x_0]$. It follows that

$$\mathfrak{D}(T(f), T(g)) = \max\{|T(f)(t) - T(g)(t)| \mid t \in J\} \leq C\,\mathfrak{D}(f,g),$$

showing that T is indeed a contraction mapping. $\square$

9.4. Exercises

Exercise 9.1. WRITE IT!!! $\square$

Exercise 9.2. WRITE IT!!! $\square$

Bibliography

[1] W. Hurewicz, *Lectures on ordinary differential equations*, reprint of the 1964 edition, Dover, New York, 1990. (Cited on page 98.)

CHAPTER 10*

The Arzelá–Ascoli theorem

CHAPTER 11

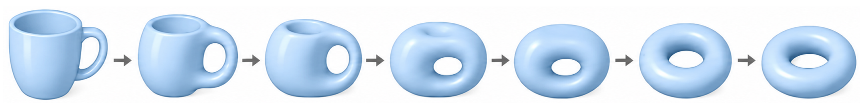
Why are metric spaces insufficient for topology?

11.1. What is topology?

Roughly speaking, in topology we study topological properties of metric spaces.[1] The most basic question in the field is as follows:

QUESTION 11.1.1. Can we classify metric spaces up to homeomorphism? □

One traditional joke is that a topologist cannot tell the difference between a donut and a coffee cup since these are homeomorphic spaces. Here is an illustration showing how to see this homeomorphism:



QUESTION 11.1.2. Fix metric spaces M and N. We say that M can be *embedded* in N if there is a subspace $X \subset N$ with $M \cong X$. Can we determine whether M can be embedded in N? □

General metric spaces are far too wild for questions like these to have reasonable answers. Typically topologists restrict to special classes of spaces; for instance, subspaces of Euclidean spaces. Even with this restriction, this quickly leads to difficult questions. For instance, let $\mathbb{S}^d$ and $\mathbb{D}^d$ be the d-sphere and the d-disk:

$$\mathbb{S}^d = \left\{(x_1, \ldots, x_{d+1}) \in \mathbb{R}^{d+1} \mid x_1^2 + \cdots + x_{d+1}^2 = 1\right\},$$
$$\mathbb{D}^d = \left\{(x_1, \ldots, x_d) \in \mathbb{R}^d \mid x_1^2 + \cdots + x_d^2 \leq 1\right\}.$$

Consider the following assertions:

(i) For all $d, e \geq 0$, the metric spaces $\mathbb{S}^d$ and $\mathbb{D}^e$ are not homeomorphic.

(ii) For all $d, e \geq 0$ with $d \neq e$, the following hold:
  - $\mathbb{S}^d$ is not homeomorphic to $\mathbb{S}^e$; and
  - $\mathbb{D}^d$ is not homeomorphic to $\mathbb{D}^e$.

(iii) For $d > e$, the following hold:
  - $\mathbb{D}^d$ cannot be embedded in $\mathbb{D}^e$ or $\mathbb{S}^e$; and
  - $\mathbb{S}^d$ cannot be embedded in $\mathbb{D}^e$ or $\mathbb{S}^e$.

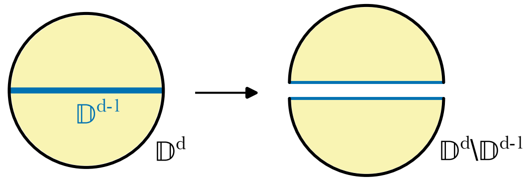
Each of these is in fact true, but aside from some degenerate cases none of them have straightforward proofs. Using tools we will develop in this book, we will be able to prove all of them:

- We will develop a notion of "dimension" for metric spaces in Essay D. The spaces $\mathbb{S}^d$ and $\mathbb{D}^d$ will both have dimension d. Metric spaces of different dimensions cannot be homeomorphic, and higher-dimensional metric spaces cannot embed into lower-dimensional ones. This handles (ii) and (iii) and most cases of (i).
- For (i), dimension theory will prove everything but the fact that $\mathbb{S}^d$ is not homeomorphic to $\mathbb{D}^d$. To tell $\mathbb{S}^d$ and $\mathbb{D}^d$ apart, we must identify a topological property that holds for one but not the other. These spaces have different cardinalities for $d = 0$, so it is enough to handle the case $d \geq 1$.

  We will develop a notion of a space having multiple "pieces" called components. Essay F will prove that there does not exist a subspace $X \subset \mathbb{S}^d$ with $X \cong \mathbb{D}^k$ for some $k \geq 0$ such

[1] In fact, topologists generally focus on the more general notion of a topological space we will introduce later in this book. However, most of the key examples of topological spaces that show up in topology come from metric spaces.

that $\mathbb{S}^d \setminus X$ has two components.[2] On the other hand, letting $X = \mathbb{D}^d \cap (\mathbb{R}^{d-1} \times \{\mathbf{0}\}) \cong \mathbb{D}^{d-1}$ the space $\mathbb{D}^d \setminus X$ has two components; see the following figure:



This will imply that $\mathbb{S}^d$ and $\mathbb{D}^d$ cannot be homeomorphic.

Remark 11.1.3. In later volumes of this book, we will develop tools from algebraic topology like homology groups and homotopy groups. Once these tools are in place, results like (i)–(iii) will become routine. $\square$

[2] This result is called the Jordan non-separation theorem.

Topological spaces

CHAPTER 12

Topological spaces: definition and basic properties

Since continuity for metric spaces can be described entirely in terms of open sets, it is natural to abstract the notion of "open sets".

12.1. HOLDING PEN FOR MOTIVATION

LEMMA 12.1.1. *Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces and let $f\colon M \to N$ be a function. Then f is continuous if and only if for all open sets $U \subset N$ the preimage $f^{-1}(U)$ is open in M.*

PROOF. We must prove the following two claims:

CLAIM. *Assume that f is continuous and $U \subset N$ is open. Then $f^{-1}(U) \subset M$ is open.*

Let $x_0 \in f^{-1}(U)$. To prove that $f^{-1}(U)$ is open, we must find some $\delta > 0$ such that $B_\delta(x_0) \subset f^{-1}(U)$. Since U is open, we can find some $\epsilon > 0$ such that $B_\epsilon(f(x_0)) \subset U$. Since f is continuous at x_0, there is some $\delta > 0$ such that for $x \in M$ with $\mathfrak{d}_M(x, x_0) < \delta$ we have $\mathfrak{d}_N(f(x), f(x_0)) < \epsilon$. In other words, f takes all $x \in B_\delta(x_0)$ to a point of $B_\epsilon(f(x_0)) \subset U$, so $B_\delta(x_0) \subset f^{-1}(U)$, as desired.

CLAIM. *Assume that for all open sets $U \subset N$ the set $f^{-1}(U) \subset M$ is open. Then f is continuous.*

Let $x_0 \in M$ and $\epsilon > 0$. Since $B_\epsilon(f(x_0)) \subset N$ is open (see Exercise 2.6), the set $f^{-1}(B_\epsilon(f(x_0))) \subset M$ is open. Since $x_0 \in f^{-1}(B_\epsilon(f(x_0)))$, there thus exists some $\delta > 0$ such that $B_\delta(x_0) \subset f^{-1}(B_\epsilon(f(x_0)))$. In other words, for $x \in M$ with $\mathfrak{d}_M(x, x_0) < \delta$ we have $\mathfrak{d}_N(f(x), f(x_0)) < \epsilon$, as desired. □

12.2. Definition of topological space

A *topological space* is a set X equipped with a collection of subsets of X called the *open sets*. These open sets should satisfy the following three properties:

- The whole space X and the empty set $\emptyset$ are both open.
- The collection of open sets is closed under arbitrary unions: if $\{U_i\}_{i \in I}$ is any collection of open sets, then $\cup_{i \in I} U_i$ is open.
- The collection of open sets is closed under finite intersections: if $U_1, \ldots, U_n$ are open sets, then $U_1 \cap \cdots \cap U_n$ is open.

We call the collection of open sets on X a *topology* on X. A key example is:

EXAMPLE 12.2.1. If $(M, \mathfrak{d})$ is a metric space, then Lemma 2.4.8 implies that the collection of open sets in M makes M into a topological space. We say that this topology on M is *induced* by the metric $\mathfrak{d}$. □

CONVENTION 12.2.2. Whenever we draw a figure in $\mathbb{R}^n$, we give it the topology it inherits as a metric space via the Euclidean metric on $\mathbb{R}^n$ discussed in Example 2.1.6. □

Topologies arising from metrics will play a key role in this book, so we introduce the following terminology. A topological space X is *metrizable* if there exists a distance function $\mathfrak{d}\colon X \times X \to \mathbb{R}$ making $(X, \mathfrak{d})$ into a metric space such that the topology on X is the one induced by $\mathfrak{d}$. The metric $\mathfrak{d}$ inducing the topology on X is not unique (see Exercise 12.8).

We will give many more examples of topological spaces in §12.6 and §12.7 below after introducing some more terminology. First, however, we prove an important lemma:

LEMMA 12.2.3. *Let X be a topological space and let $V \subset X$ be a subset. Assume that for all $p \in V$, there exists an open set $U \subset V$ with $p \in U$. Then V is open.*

PROOF. For $p \in V$, let $U_p \subset V$ be an open set with $p \in U_p$. We have $V = \cup_{p \in V} U_p$, so since open sets are closed under arbitrary unions it follows that V is open. $\square$

REMARK 12.2.4. Lemma 12.2.3 allows reasoning like we did for metric spaces where the role of the open set U is played by open balls $B_r(p)$. $\square$

12.3. Closed sets and limit points

Let X be a topological space. A set $C \subset X$ is *closed* if $X \setminus C$ is open. By Lemma 2.8.1, this agrees with our previous definition if X is metrizable. Since the collection of open sets is closed under arbitrary unions and finite intersections, it follows that the collection of closed sets is closed under finite unions and arbitrary intersections. The whole subject could be developed using closed sets instead of open ones (see Exercise 12.6).

Let $A \subset X$ be a set. A *limit point* of A is a point $p \in X$ such that for all open sets U with $p \in U$, there exists some $a \in U \cap A$ with $a \neq p$. Just like for metric spaces, we can characterize closed sets using limit points:

LEMMA 12.3.1. *Let X be a topological space and let $A \subset X$. Then A is closed if and only if all limit points of A lie in A.*

PROOF. We divide the proof into two steps:

STEP 1. *Assume that A is closed. Then all limit points of A lie in A.*

Consider $p \in X \setminus A$. We must prove that p is not a limit point of A. Since X is closed, $U = X \setminus A$ is open. It follows that U is an open set with $p \in U$ and $U \cap A = \emptyset$, so p is not a limit point of A.

STEP 2. *Assume that all limit points of A lie in A. Then A is closed.*

We must prove that $X \setminus A$ is open. Consider $p \in X \setminus A$. By Lemma 12.2.3, is enough to find some open set $U \subset X \setminus X$ with $p \in U$. Since all limit points of A lie in A, the point p is not a limit point of A. There is thus an open set U with $p \in U$ such that $U \cap A$ contains not points of A except for possibly p. Since $p \notin A$, it follows that $U \cap A = \emptyset$, i.e., $U \subset X \setminus A$, as desired. $\square$

12.4. Interiors, closures, and neighborhoods

For a subset $A \subset X$, we define the interior $\mathrm{Int}(A)$ and the closure $\overline{A}$ as follows:

- The interior $\mathrm{Int}(A)$ is the union of all open sets U with $U \subset A$. Since the collection of open sets is closed under arbitrary unions, $\mathrm{Int}(A)$ is open. The set $\mathrm{Int}(A)$ is the largest open set contained in A.
- The closure $\overline{A}$ is the intersection of all closed sets C with $A \subset C$. Since the collection of closed sets is closed under arbitrary intersections, $\overline{A}$ is closed. The set $\overline{A}$ is the smallest closed set containing A.

These agree with our previous definitions if X is metrizable (see Exercise 12.7). For $p \in X$, a *neighborhood* of p is a set A with $p \in \mathrm{Int}(A)$. More generally, for a set $B \subset X$, a *neighborhood* of B is a set A with $B \subset \mathrm{Int}(A)$. The most important special case of this terminology is an open neighborhood of $B \subset X$, which is an open set U with $B \subset U = \mathrm{Int}(U)$.

12.5. Continuity

A map $f\colon X \to Y$ between topological spaces is *continuous* if for all $U \subset Y$ open, its preimage $f^{-1}(U) \subset X$ is open. By Lemma 12.1.1, this is equivalent to the usual ϵ-δ definition if X and Y are metric spaces. The following lemma shows that there is a category `Top` whose objects are topological spaces and whose morphisms are continuous maps. This category is the natural home for topology.

LEMMA 12.5.1. *The following hold:*

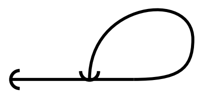
(i) If X is a topological space, then the identity map $\mathbb{1}_X\colon X \to X$ is continuous.
(ii) If $f\colon X \to Y$ and $g\colon Y \to Z$ are continuous maps between topological spaces, then $g \circ f\colon X \to Z$ is continuous.

Proof. Conclusion (i) is obvious. As for conclusion (ii), let $U \subset Z$ be open. Since f is continuous the set $g^{-1}(U) \subset Y$ is open, so since g is continuous the set $f^{-1}(g^{-1}(U)) = (g \circ f)^{-1}(U) \subset X$ is open, as desired. □

If $\mathbf{C}$ is a category, then a morphism $m\colon A \to B$ in $\mathbf{C}$ is an *isomorphism* if there exists a morphism $m^{-1}\colon B \to A$ such that $m^{-1} \circ m = \mathbb{1}_A$ and $m \circ m^{-1} = \mathbb{1}_B$. If there exists an isomorphism from A to B, then we say that A and B are *isomorphic*. An isomorphism in $\mathtt{Top}$ is called a *homeomorphism*. Unwinding the definition, a homeomorphism between topological spaces X and Y is a continuous map $f\colon X \to Y$ that is bijective and whose inverse $f^{-1}\colon Y \to X$ is continuous. If there exists a homeomorphism from X to Y, then we say that X and Y are *homeomorphic* and write $X \cong Y$.

Here is an example to show that the continuity of f^{-1} is not immediate:

Example 12.5.2. Consider the injective map $f\colon (0,1) \to \mathbb{R}^2$ whose image X is as follows:



The map $f\colon (0,1) \to X$ is continuous and bijective, but the inverse map $f^{-1}\colon X \to (0,1)$ is not continuous (see Exercise 12.13). □

Convention 12.5.3. Henceforth, we will use the word "space" as a synonym for "topological space". Also, unless otherwise specified all maps between spaces are assumed to be continuous. □

Remark 12.5.4. For metric spaces, we also characterized continuity using limits (see §2.5). The notion of limits can be generalized to topological spaces, though with some subtleties (for instance, limits of sequences need not be unique). However, without some additional assumptions it would give a different notion of continuity. See §15.2 for the definition of a limit in a topological space and Lemma 16.3.1 for the conditions needed for continuity to be described in terms of limits. □

12.6. Other examples

The notion of a topological space is extremely general. Here are a few more examples.

Example 12.6.1. Let X be a set. The *discrete topology* on X is the one where all sets are open. The *trivial topology* on X is the one where the only open sets are $\emptyset$ and X. Another topology that can be put on an arbitrary set X is the *cofinite topology* whose open sets are those of the form $X \setminus F$ with F finite. The fact that this is a topology follows from the fact that finite sets are closed under finite unions and arbitrary intersections. □

Example 12.6.2. Let $\mathbf{k}$ be a field; for instance, $\mathbf{k}$ might be $\mathbb{C}$ or $\mathbb{R}$. For a polynomial $f \in \mathbf{k}[z_1,\ldots,z_n]$, define the *vanishing* and *non-vanishing* loci of f to be

$$V(f) = \{(x_1,\ldots,x_n) \in \mathbf{k}^n \mid f(x_1,\ldots,x_n) = 0\} \subset \mathbf{k}^n \quad \text{and} \quad NV(f) = \mathbf{k}^n \setminus V(f).$$

The *Zariski topology* on $\mathbf{k}^n$ is the topology whose open sets are the nonvanishing loci $NV(f)$ as f ranges over elements of $\mathbf{k}[z_1,\ldots,z_n]$ (see Exercise 12.9). The closed sets are thus the vanishing loci $Z(f)$. For $n = 1$, the vanishing locus of a polynomial in $\mathbf{k}[z_1]$ can be any finite subset of $\mathbf{k}^1$, so the Zariski topology on $\mathbf{k}^1$ is the cofinite topology. □

Remark 12.6.3. For $\mathbf{k}$ equal to $\mathbb{C}$ or $\mathbb{R}$, we have now seen two topologies on $\mathbf{k}^n$:

- the classical topology obtained by regarding $\mathbf{k}^n$ as a metric space; and
- the Zariski topology.

Every open set in the Zariski topology is open in the classical topology. We say that the classical topology is *finer* or *stronger* than the Zariski topology, and that the Zariski topology is *coarser* or *weaker* than the classical topology. □

Example 12.6.4. A finite set X can be endowed with many topologies. For instance, if $X = \{a, b\}$ then the following are all topologies on X:

- the discrete topology: $\emptyset, \{a\}, \{b\}, \{a, b\}$
- $\emptyset, \{a\}, \{a, b\}$

- $\emptyset, \{b\}, \{a, b\}$
- the topology: $\emptyset, \{a, b\}$

The number of topologies on a finite set X grows very quickly as the size of X grows. Some of these spaces can be given geometric interpretations, but most of them are purely combinatorial objects. $\square$

REMARK 12.6.5. Because the notion of a topological space is so general, there is almost nothing nontrivial that can be said about an arbitrary topological space. They are thus almost never studied for their own sake. Rather, they provide a minimal framework and language for studying continuity as it appears throughout mathematics. $\square$

12.7. Basis for a topology

A *basis* for a topology on a set X consists of a set $\mathfrak{B}$ of subsets of X such that:

(i) all points of X lie in some $U \in \mathfrak{B}$; and
(ii) for all $U, V \in \mathfrak{B}$, the intersection $U \cap V$ can be written as a union of sets in $\mathfrak{B}$.

Given such a basis, the corresponding topology is the one where a set $U \subset X$ is open if and only if U is a union of sets in $\mathfrak{B}$. The following shows that this is indeed a topology:

LEMMA 12.7.1. *Let X and $\mathcal{B}$ be as above. Then the indicated collection of open sets forms a topology on X.*

PROOF. The empty set $\emptyset$ is open since it is the union of no elements of $\mathcal{B}$. The whole set X is open because of (i). It is immediate that the indicated collection of open sets is closed under arbitrary unions, so the only nontrivial thing we must check is that it is closed under finite intersections. By induction, it is enough to prove that if U and V are open then so is $U \cap V$. Write

$$U = \bigcup_{i \in I} U_i \quad \text{and} \quad V = \bigcup_{j \in J} V_j$$

with each U_i and V_j in $\mathcal{B}$. We then have

$$U \cap V = \bigcup_{\substack{i \in I \\ j \in J}} U_i \cap V_j.$$

By (ii), each $U_i \cap V_j$ is the union of sets in $\mathcal{B}$, so $U \cap V$, as desired. $\square$

Here are several examples:

EXAMPLE 12.7.2. Let M be a metric space. The set

$$\mathfrak{B} = \{B_r(p) \mid p \in M \text{ and } r > 0\}$$

of open balls in M is a basis for the metric space topology on M. It is immediate from the definitions that a set $U \subset M$ is open in the metric space topology if and only if it is a union of elements of $\mathfrak{B}$. The only thing we must therefore check is that $\mathfrak{B}$ satisfies the axioms of a basis. For this, the nontrivial thing to verify is that if $B_r(p), B_{r'}(p') \in \mathfrak{B}$, then $B_r(p) \cap B_{r'}(p')$ is a union of elements of $\mathfrak{B}$. Let $x \in B_r(p) \cap B_{r'}(p')$. It is enough to find some $s > 0$ with $B_s(x) \subset B_r(p) \cap B_{r'}(p')$. Set

$$s = \min(r - \mathfrak{d}(x, p), r' - \mathfrak{d}(x, p')) > 0.$$

For $y \in B_s(x)$, we have

$$\mathfrak{d}(y, p) \leq \mathfrak{d}(y, x) + \mathfrak{d}(x, p) < (r - \mathfrak{d}(x, p)) + \mathfrak{d}(x, p) = r,$$
$$\mathfrak{d}(y, p') \leq \mathfrak{d}(y, x) + \mathfrak{d}(x, p') < (r' - \mathfrak{d}(x, p')) + \mathfrak{d}(x, p') = r'.$$

It follows that $y \in B_r(p) \cap B_{r'}(p')$, so $B_s(x) \subset B_r(p) \cap B_{r'}(p')$, as desired. $\square$

EXAMPLE 12.7.3. Let S be a set with a total ordering $\leq$. For $s_1, s_2 \in S$ with $s_1 < s_2$, let $(s_1, s_2) = \{s \in S \mid s_1 < s < s_2\}$. The set

$$\mathfrak{B} = \{(s_1, s_2) \mid s_1, s_2 \in S,\ s_1 < s_2\}$$

is a basis for a topology on S (see Exercise 12.10) called the *order topology*. For instance, if $S = \mathbb{R}$ with its usual ordering the order topology is the same as the standard topology on $\mathbb{R}$. $\square$

EXAMPLE 12.7.4. Let X and Y be topological spaces. Let $\mathcal{B}$ be the set of subsets of $X \times Y$ of the form $U \times V$ with $U \subset X$ and $V \subset Y$ open. This is a basis for a topology on $X \times Y$ (see Exercise 12.11). We call this the *product topology*, and will discuss it extensively in Chapter 21. □

EXAMPLE 12.7.5. Consider the real line $\mathbb{R}$. Let $\mathcal{B}$ be the set $\{[a, b) \mid a < b\}$ of all half-open intervals in $\mathbb{R}$. This is the basis for a topology on $\mathbb{R}$ called the *lower limit topology* (see Exercise 12.12). The real line $\mathbb{R}$ equipped with the lower limit topology is called the *Sorgenfrey line*. □

REMARK 12.7.6. There is a also the weaker notion of a subbasis; see §23.1. □

12.8. Subspaces, embeddings, and open/closed maps

Let X be a space and let $Y \subset X$ be a subset. We would like to make Y into a space. Letting $\iota\colon Y \to X$ be the inclusion, the topology we impose on Y should make ι into a continuous function. For an open set $U \subset X$, we therefore need $\iota^{-1}(U) = U \cap Y$ to be open in Y. This suggests the following: the *subspace topology* on Y is the topology whose open sets $V \subset Y$ are the sets of the form $V = U \cap Y$ for an open set $U \subset X$. Unless we say otherwise, all subspaces are given the subspace topology.

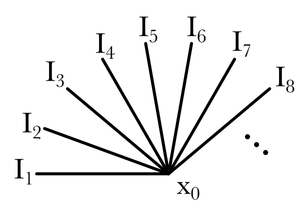
Let $f\colon X \to Y$ be a continuous map. By definition, $f^{-1}(U) \subset X$ is open if $U \subset Y$ is open, and also $f^{-1}(C) \subset X$ is closed if $C \subset Y$ is closed. We say that the map f is *open* if $f(V) \subset Y$ is open for all $V \subset X$ open. Similarly, the map f is *closed* if $f(D) \subset Y$ is closed for all $D \subset X$ closed. Homeomorphisms are both open and closed.

A map $f\colon X \to Y$ is an *embedding* if it is a homeomorphism onto its image. In other words, an embedding f is a continuous injection onto a subspace $f(X)$ of Y such that the inverse map $f^{-1}\colon f(X) \to X$ is continuous. An injection which is either open or closed is automatically an embedding. For a subspace A of X, the inclusion $\iota\colon A \to X$ is an embedding. If $A \subset X$ is open (resp. closed), then ι is an open embedding (resp. a closed embedding).

12.9. Gluing intervals

the notion of a topological space fixes its pathological behavior.

EXAMPLE 12.9.1. For each integer $n \geq 1$, let I_n be a copy of the interval $I = [0, 1]$. Let M be the set obtained from the disjoint union of all the I_n by identifying the points $0 \in I_n$ together to a single point x_0:



Endow M with the following topology:

- A set $U \subset M$ is open if and only if $U \cap I_n$ is open for all $n \geq 1$. It is easy to see that this does indeed give a topology (see Exercise 12.4).

It is immediate from this definition that a function $f\colon M \to \mathbb{R}$ is continuous if and only if $f|_{I_n}\colon I_n \to \mathbb{R}$ is continuous for all $n \geq 1$. In defined via the formula $f(p) = np$ for $p \in I_n$ is continuous. We will later see that M is not metrizable (see Exercise 15.4). □

REMARK 12.9.2. is an example of an identification space topology. See Chapter 13 below for more details about this. □

12.10. Exercises

EXERCISE 12.1. Give an example of a topological space X such that there exist open subsets $\{U_i \mid i \geq 1\}$ of X for which $\cap_{i=1}^{\infty} U_i$ is not open. □

EXERCISE 12.2. Let X be a topological space and let $A \subset X$. Assume that for each $a \in A$ there exists a neighborhood N of a with $N \subset A$. Note that we are not assuming that N is open. Prove that A is open. □

EXERCISE 12.3. Let X and Y be topological spaces and let $f\colon X \to Y$ be a set map. Prove that f is continuous if the following holds:

- For all $x \in X$ and open neighborhoods $V \subset Y$ of $f(x)$, there exists an open neighborhood U of x with $f(U) \subset V$. □

EXERCISE 12.4. Let Y be a set. For each $i \in I$, let X_i be a subset of Y. Assume that X_i is endowed with a topology. Prove that the following gives a topology on Y:

- A set $U \subset Y$ is open if and only $U \cap X_i$ is open for all $i \in I$. □

EXERCISE 12.5. Let X be a space and $\mathfrak{A}$ be a collection of subsets of X such that $X = \cup_{A\in\mathfrak{A}} A$. Let Y be another space and let $f\colon X \to Y$ be a map of sets (not necessarily continuous) such that $f|_A$ is continuous for all $A \in \mathfrak{A}$.

(a) Assume that each $A \in \mathfrak{A}$ is open. Prove that f is continuous.
(b) Assume that each $A \in \mathfrak{A}$ is closed and that $\mathfrak{A}$ is finite. Prove that f is continuous.
(c) Construct an example where each $A \in \mathfrak{A}$ is closed but f is *not* continuous. □

EXERCISE 12.6. Formulate a set of axioms for a collection of closed sets to be the closed sets of a topology. Prove that this gives the same notion of a topological space as the axiomization we gave using open sets. □

EXERCISE 12.7. Let M be a metrizable space and let $A \subset M$ be a subset. Prove that the interior $\mathrm{Int}(A)$ and closure $\overline{A}$ defined in §12.4 agrees with the definitions for metric spaces given in §2.4 and §2.7. □

EXERCISE 12.8. Prove the following:

(a) Let $(M, \mathfrak{d})$ be a metric space. Define $\mathfrak{d}'\colon M \times M \to \mathbb{R}$ via the formula $\mathfrak{d}'(p,q) = \min\{\mathfrak{d}(p,q), 1\}$. Prove that $\mathfrak{d}'$ is a metric on M that induces the same topology on M that $\mathfrak{d}$ does.
(b) Let $\|-\|$ be the following standard norm on $\mathbb{R}^n$:
$$\|(x_1,\ldots,x_n)\| = \sqrt{x_1^2 + \cdots + x_n^2} \quad \text{for all } (x_1,\ldots,x_n) \in \mathbb{R}^n.$$
This induces the metric $\mathfrak{d}(p,q) = \|p-q\|$ on $\mathbb{R}^n$. Now let $\|-\|'$ be an arbitrary norm on the vector space $\mathbb{R}^n$. Define a function $\mathfrak{d}'\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ via the formula
$$\mathfrak{d}'(p,q) = \|p-q\|' \quad \text{for } p,q \in \mathbb{R}^n.$$
Prove that $\mathfrak{d}'$ is a metric on $\mathbb{R}^n$ and that $\mathfrak{d}'$ induces the same topology on $\mathbb{R}^n$ as $\mathfrak{d}$. □

EXERCISE 12.9. Let $\mathbf{k}$ be a field. Prove that the Zariski topology on $\mathbf{k}^n$ described in Example 12.6.2 is a topology. □

EXERCISE 12.10. Let S be a set with a total ordering $\leq$. For $s_1, s_2 \in S$ with $s_1 < s_2$, let $(s_1, s_2) = \{s \in S \mid s_1 < s < s_2\}$. Prove that the collection of all sets of the form (s_1, s_2) forms a basis for a topology on S. □

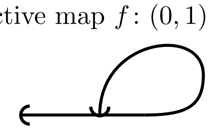
EXERCISE 12.11. Let X and Y be topological spaces. Let $\mathcal{B}$ be the set of subsets of $X \times Y$ of the form $U \times V$ with $U \subset X$ and $V \subset Y$ open. Prove that $\mathcal{B}$ is the basis for a topology on $X \times Y$. This is called the *product topology*. □

EXERCISE 12.12. Consider the real line $\mathbb{R}$.

(a) Let $\mathcal{B}$ be the set $\{[a,b) \mid a < b\}$ of all half-open intervals in $\mathbb{R}$. Prove that $\mathcal{B}$ is the basis for a topology on $\mathbb{R}$ called the *lower limit topology*.
(b) Prove that the lower limit topology on $\mathbb{R}$ is finer than the standard topology.

The real line $\mathbb{R}$ equipped with the lower limit topology is called the *Sorgenfrey line*. □

EXERCISE 12.13. Consider the injective map $f\colon (0,1) \to \mathbb{R}^2$ whose image X is as follows:



Regard f as a bijective map $f\colon (0,1) \to X$. Prove that $f^{-1}\colon X \to (0,1)$ is not continuous. □

CHAPTER 13

Identification spaces and the quotient topology

We now explain how to construct a new space from a collection of existing ones by identifying certain points together. This generalizes the construction we gave in Example 12.9.1.

13.1. Identification spaces

Let $\{X_i\}_{i \in I}$ be a collection of spaces. An *identification space* is a topological space Y equipped with maps $f_i \colon X_i \to Y$ for each $i \in I$ such that:

- each $y \in Y$ is in the image of some f_i; and
- a set $U \subset Y$ is open if and only if $f_i^{-1}(U) \subset X_i$ is open for all $i \in I$.

The second condition ensures that each $f_i \colon X_i \to Y$ is continuous. It also ensures that for a space Z a map of sets $\phi \colon Y \to Z$ if continuous if and only if $\phi \circ f_i \colon X_i \to Z$ is continuous for all $i \in I$ (we will say more about this in §13.4 below).

Example 13.1.1. Let $\{X_i\}_{i \in I}$ be a collection of spaces. As a set, let Y be the disjoint union $\sqcup_{i \in I} X_i$ of the X_i. Topologize Y by letting $U \subset Y$ be open if and only if $U \cap X_i$ is open for all $i \in I$ (see Exercise 12.4). The space Y is an identification space under the inclusion maps $f_i \colon X_i \to Y$. □

Now assume that the $\{X_i\}_{i \in I}$ are spaces and $\sim$ is an equivalence relation on the disjoint union $\sqcup_{i \in I} X_i$. Define

$$Y = \sqcup_{i \in I} X_i / \sim .$$

Informally, Y is obtained from the X_i by identifying some of their points together. Let $f_i \colon X_i \to Y$ be the projections. Give Y the following topology:

- A set $U \subset Y$ is open if and only if $f^{-1}(U) \subset X_i$ is open for all $i \in I$. This is easily seen to give a topology (see Exercise 13.1).

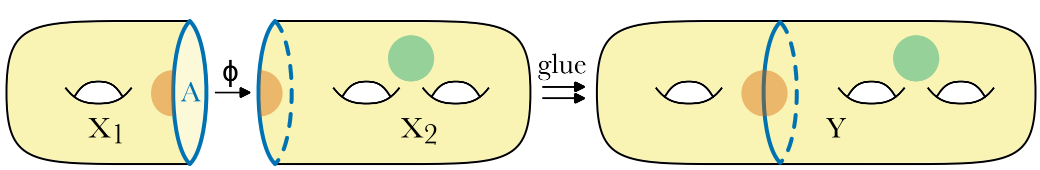
We call this the *identification space topology* on Y. It makes Y into an identification space. If we have a construction of a purported "space" from the points of the X_i, then this gives a canonical way of turning our purported "space" into a topological space.

13.2. Examples

The above discussion is a little abstract. Here are some examples.

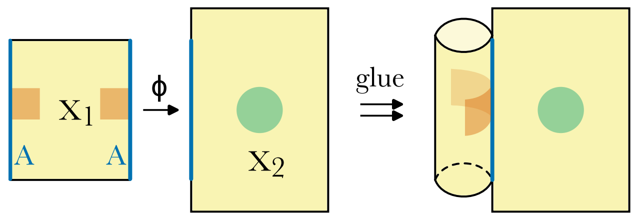
Example 13.2.1 (Gluing). Let X_1 and X_2 be spaces, $A \subset X_1$ be a subspace, and $\phi \colon A \to X_2$ be a map. As a set, let Y be the disjoint union of X_1 and X_2 modulo the equivalence relation that identifies each $a \in A \subset X_1$ with $\phi(a) \in X_2$. There are natural maps $f_1 \colon X_1 \to Y$ and $f_2 \colon X_2 \to Y$, and we give Y the identification space topology. We call Y the space obtained by *gluing* X_1 to X_2 via the gluing map ϕ. If $A \subset X_1$ is closed, then $f_2 \colon X_2 \to Y$ is a closed embedding and $f_1|_{X_1 \setminus A} \colon X_1 \setminus A \to Y$ is an open embedding (see Exercise 13.2). If A is not closed, then things can be much more complicated. □

Example 13.2.2. Here is one easy example of gluing. Let X_1 and X_2 and Y be the following surfaces:



As is shown in this figure, Y is obtained by gluing the boundary component $A \cong \mathbb{S}^1$ of X_1 to the boundary component of X_2 via a homeomorphism. Two open sets on Y are drawn together with their preimages in X_1 and X_2. In this example, ϕ is a homeomorphism onto its image and both X_1 and X_2 are subspaces of Y. □

Example 13.2.3. Here is another example of gluing where the gluing map is not injective:

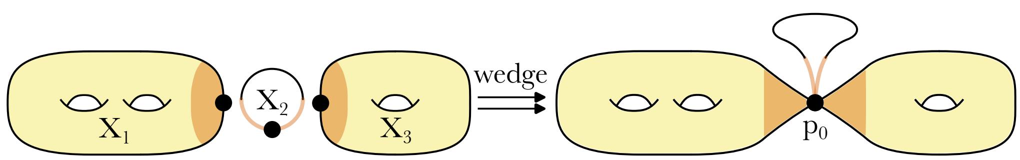


In this figure, ϕ identifies the two blue vertical edges of the rectangle X_1 with a single segment in the left-hand vertical edge of X_2. □

Remark 13.2.4. The most important class of spaces in algebraic topology are CW complexes, which are constructed from collections of disks $\mathbb{D}^n$ of various dimensions n by gluing their boundaries together. See Essay H for more details. □

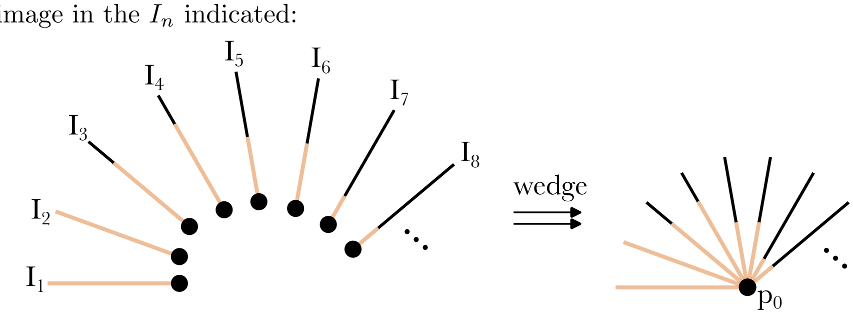
Example 13.2.5 (Wedge product). Let $\{X_i\}_{i\in I}$ be a collection of topological spaces. Assume that each X_i has a distinguished basepoint $x_i \in X_i$. The *wedge product* of the X_i, denoted $\vee_{i\in I} X_i$, is the space obtained by identifying all the x_i together to a single point p_0. There are inclusions $f_i\colon X_i \to \vee_{i\in I} X_i$, and we give $\vee_{i\in I} X_i$ the identification space topology. The maps f_i are all embeddings (see Exercise 13.3). □

Example 13.2.6. Here is an example of a wedge product:



This figure shows an open neighborhood of p_0 together with its preimage in the X_i. □

Example 13.2.7. Example 12.9.1 is the wedge product of countably many intervals $I_n = I$ equipped with the basepoints $0 \in I_n$. Here is a picture of this, with an open neighborhood p_0 together with its preimage in the I_n indicated:

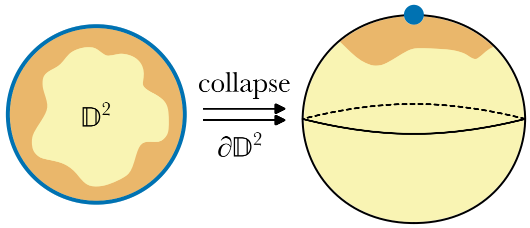


See Exercise 15.4 for an explanation of why this topology does not come from a metric. □

Example 13.2.8 (Collapsing subspace). In an identification space, we allow there to only be a single space X. As an example of this, let X be a space and let $A \subset X$ be a subspace. Denote by X/A the result of collapsing A to a single point. The points of X/A are thus the points of $X \setminus A$ together with a single point $[A]$ corresponding to A. Letting $f\colon X \to X/A$ be the projection, we can endow X/A with the identification space topology. If A is a closed (resp. open) set, then the restriction of f to $X \setminus A$ is an open (resp. closed) embedding (see Exercise 13.4). □

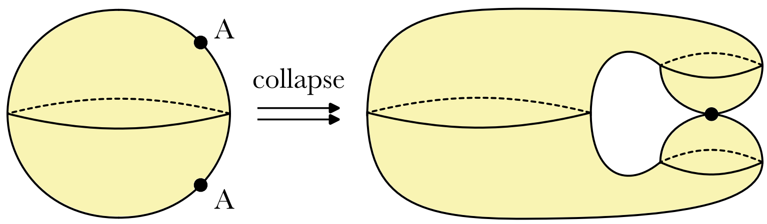
Remark 13.2.9. Collapsing a subspace $A \subset X$ will typically yield a pathological space X/A if A is not closed. For instance, in most natural spaces single points are closed, and this will only hold for the point $[A]$ of X/A if $A \subset X$ is closed. As an example of how pathological this can be, collapsing the subspace $\mathbb{Q}$ of $\mathbb{R}$ gives a terrible space $\mathbb{R}/\mathbb{Q}$. □

EXAMPLE 13.2.10. Consider the boundary $\partial\mathbb{D}^n = \mathbb{S}^{n-1}$. As the following shows, $\mathbb{D}^n/\partial\mathbb{D}^n \cong \mathbb{S}^n$, with the blue $\partial\mathbb{D}^n$ mapping to the north pole of $\mathbb{S}^n$:



As this figure shows, a neighborhood of the north pole in $\mathbb{S}^n$ lifts to a neighborhood of $\partial\mathbb{D}^n$. □

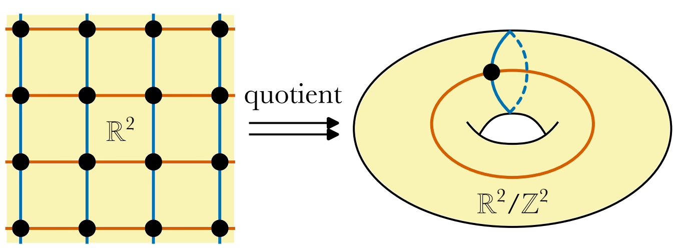
EXAMPLE 13.2.11. The set A need not be connected. Here is an example:



In this example, $X = \mathbb{S}^2$ and A is two points on X. □

EXAMPLE 13.2.12 (Quotienting by group action). Let X be a space and let G be a group acting on X. As a set, X/G consists of the orbits of X under the action of G. Letting $f\colon X \to X/G$ be the quotient map, we endow X/G with the identification space topology. The point-set topological properties of X/G depend both on the properties of X and the qualities of the group action G. We explore this in Essay J. □

EXAMPLE 13.2.13. Let the group $\mathbb{Z}^2$ act on $\mathbb{R}^2$ by translations. The quotient $\mathbb{R}^2/\mathbb{Z}^2$ is homeomorphic to the 2-torus:



The orange and blue loops on $\mathbb{R}^2/\mathbb{Z}^2$ lift to the orange and blue parallel lines on $\mathbb{R}^2$. □

13.3. Quotient topology

A map $f\colon X \to Y$ is a *quotient map* if f is surjective and $U \subset Y$ is open if and only if $f^{-1}(U) \subset X$ is open. Given a space X and a surjection of sets $f\colon X \to Y$, the *quotient topology* on Y is the topology making $f\colon X \to Y$ a quotient map. We call Y a *quotient space* of X.

Of course, this is a special case of an identification space. Moreover, given a collection of spaces $\{X_i\}_{i\in}$ and an identification space Y of the X_i with maps $f_i\colon X_i \to Y$, we can realize Y as a quotient space in the following way. Let $\sqcup_{i\in I}X_i$ be the disjoint union of the X_i (see Example 13.1.1). The maps $f_i\colon X_i \to Y$ then assemble to a quotient map $F\colon \sqcup_{i\in I} X_i \to Y$.

REMARK 13.3.1. Many treatments of point-set topology only talk about quotient spaces, but we find it convenient to use the slightly more general notion of identification spaces since like in the examples from earlier in this chapter, we often use them to build a space out of several spaces, not just one. □

13.4. Universal mapping property

Let $f\colon X \to Y$ be a quotient map. Let $\sim$ be the equivalence relation on the set X where $p \sim q$ if and only if $f(p) = f(q)$. The equivalence classes of $\sim$ are the fibers $f^{-1}(y)$ for $y \in Y$. Letting Z be another space and $\phi\colon Y \to Z$ be a continuous map, the composition $\Phi = \phi \circ f$ is a continuous map $\Phi\colon X \to Z$ that is $\sim$-invariant, i.e., that has $\Phi(p) = \Phi(q)$ whenever $p \sim q$. Conversely, if $\Phi\colon X \to Z$ is a continuous $\sim$-invariant map, then there is a set map $\phi\colon Y \to Z$ such that $\Phi = \phi \circ f$ and the quotient topology on Y is set up to ensure that ϕ is continuous.

The above discussion shows that composition with f gives a bijection between continuous maps $\phi\colon Y \to Z$ and $\sim$-invariant continuous maps $\Phi\colon X \to Z$. This is an example of a *universal mapping property*, and we will describe it informally by saying that a map $\phi\colon Y \to Z$ is the same as a $\sim$-invariant map $\Phi\colon X \to Z$. This universal mapping property characterizes quotient spaces (see Exercise 13.8).

Example 13.4.1 (Disjoint union). Let $\{X_i\}_{i\in I}$ be a collection of spaces. Let $Y = \sqcup_{i\in I} X_i$ with the disjoint union topology discussed in Example 13.1.1. For a space Z, maps $\phi\colon \vee_{i\in I} X_i \to Z$ are the same as collections of maps $\Phi_i\colon X_i \to Z$ for each $i \in I$. $\square$

Remark 13.4.2. If $\mathbf{C}$ is a category and $\{C_i\}_{i\in I}$ are objects of $\mathbf{C}$, then a *categorical sum* of the C_i is an object D of $\mathbf{C}$ together with morphisms $\{f_i\colon C_i \to D\}_{i\in I}$ such that the following holds:

- For all objects E of $\mathbf{C}$, there is a bijection between morphisms $\phi\colon D \to E$ and collections of morphisms $\{\Phi_i\colon C_i \to E\}_{i\in I}$ taking a morphism $\phi\colon D \to E$ to the collection of morphisms $\{\phi \circ f_i\colon C_i \to E\}_{i\in I}$.

Categorical sums might or might not exist, but if they do exist they are unique up to isomorphism (see Exercise 13.9). For spaces $\{X_i\}_{i\in I}$, the disjoint union $U = \sqcup_{i\in I} X_i$ together with the natural inclusions $\{\iota_i\colon X_i \hookrightarrow U\}_{i\in I}$ is therefore the categorical sum of the $\{X_i\}_{i\in I}$ in the category `Top`. See Exercise 13.10 for what the categorical sum means in the category of abelian groups. $\square$

Remark 13.4.3. There are also categorical products, which we will discuss in Chapter 21. $\square$

Example 13.4.4 (Wedge product). Let $\{X_i\}_{i\in I}$ be a collection of topological spaces. Assume that each X_i has a distinguished basepoint $x_i \in X_i$. For a space Z, maps $\phi\colon \vee_{i\in I} X_i \to Z$ are the same as collections of maps $\Phi_i\colon X_i \to Z$ such that $\Phi_i(x_i) = \Phi_j(x_j)$ for all $i, j \in I$. In particular, this explains why the quotient topology is the right one to ensure the real-valued function in Example 12.9.1 is continuous. $\square$

Remark 13.4.5. Let $\mathtt{Top}_*$ be the following category:

- The objects of $\mathtt{Top}_*$ are pairs (X, x_0) with X a space and $x_0 \in X$ a basepoint.
- The morphism in $\mathtt{Top}_*$ from (X, x_0) to (Y, y_0) are continuous maps $f\colon X \to Y$ with $f(x_0) = y_0$.

The universal mapping property from Example 13.4.4 says that the wedge product is the categorical sum in the category $\mathtt{Top}_*$. $\square$

Example 13.4.6 (Collapsing subspace). Let X be a space and let $A \subset X$ be a subspace. For a space Z, maps $\phi\colon X/A \to Z$ are the same as maps $\Phi\colon X \to Z$ such that $\Phi(A)$ is a single point. $\square$

Example 13.4.7 (Quotienting by group action). Let X be a space and let G be a group acting on X. For a space Z, maps $\phi\colon X/G \to Z$ are the same as maps $\Phi\colon X \to Z$ that are G-invariant in the sense that $\Phi(g{\cdot}x) = \Phi(x)$ for all $x \in X$ and $g \in G$. $\square$

13.5. Exercises

Exercise 13.1. Let $\{X_i\}_{i\in I}$ be a collection of spaces. Let Y be a set, and for all $i \in I$ let $f_i\colon X_i \to Y$ be a set map. Prove that the following gives a topology on Y:

- A set $U \subset Y$ is open if and only $f_i^{-1}(U) \cap X_i$ is open for all $i \in I$. $\square$

Exercise 13.2. Let X_1 and X_2 be spaces, let $A \subset X_1$ be a closed subspace, and let $\phi\colon A \to X_2$ is a map. Let Y be the space obtained by gluing X_1 to X_2 via the gluing map ϕ, and let $f_1\colon X_1 \to Y$ and $f_2\colon X_2 \to Y$ be the natural maps.

(a) Prove that $f_2\colon X_2 \to Y$ is a closed embedding.
(b) Prove that $f_1|_{X_1 \setminus A}\colon X_1 \setminus A \to Y$ is an open embedding. $\square$

EXERCISE 13.3. Let $\{X_i\}_{i \in I}$ be a collection of topological spaces. Assume that each X_i has a distinguished basepoint $x_i \in X_i$, and consider the edge product $\vee_{i\in I} X_i$. For each $i \in I$, let $f_i\colon X_i \to \vee_{i\in I} X_i$ be the inclusion.

(a) For each $i \in I$, prove that the map f_i is an embedding.
(b) If each basepoint $x_i \in X_i$ is closed (as a one-point subspace of X_i) then prove that each f_i is a closed embedding. $\square$

EXERCISE 13.4. Let X be a space, let $A \subset X$ be a subspace, and let $f\colon X \to X/A$ be the projection.

(a) If A is closed, then prove that the restriction of f to $X \setminus A$ is an open embedding.
(b) If A is open, then prove that the restriction of f to $X \setminus A$ is a closed embedding. $\square$

EXERCISE 13.5. Let $f\colon X \to Y$ be a quotient map. Let $B \subset Y$ be either open or closed and let $A = f^{-1}(B)$. Prove that $f|_A\colon A \to B$ is a quotient map. $\square$

EXERCISE 13.6. Let $f\colon X \to Y$ be a surjective continuous map that is either open or closed. Prove that f is a quotient map. $\square$

EXERCISE 13.7. Let $\sim_1$ and $\sim_2$ be the following equivalence relations on $\mathbb{R}^2$:

$$\begin{aligned} (x,y) \sim_1 (z,w) \quad &\text{if } x + y^2 = z + w^2, \\ (x,y) \sim_2 (z,w) \quad &\text{if } x^2 + y^2 = z^2 + w^2. \end{aligned}$$

For $i = 1, 2$, let X_i be the quotient of $\mathbb{R}^2$ by $\sim_i$. Identify the spaces X_1 and X_2. $\square$

EXERCISE 13.8. Let X be a space and let $\sim$ be an equivalence relation on X. As a set, let $Y = X/\sim$ and let $f\colon X \to Y$ be the projection. Endow Y with the quotient topology, so $f\colon X \to Y$ is a quotient map. Let Y' be another space and let $f'\colon X \to Y'$ be a map such that the following holds:

- For all spaces Z, composition with f' gives a bijection between continuous maps $\phi\colon Y' \to Z$ and $\sim$-invariant continuous maps $\Phi\colon X \to Z$.

Prove that there is a homeomorphism $g\colon Y \to Y'$ such that $f' = g \circ f$. In other words, the above universal mapping property characterizes the quotient space Y. $\square$

EXERCISE 13.9. Let $\mathbf{C}$ be a category and let $\{C_i\}_{i\in I}$ be objects of $\mathbf{C}$. For $k = 1, 2$, let $D(k)$ be a categorical sum of the $\{C_i\}_{i\in I}$ with morphisms $\{f_i(k)\colon C_i \to D(k)\}_{i\in I}$. Prove that there exists an isomorphism $\lambda\colon D(1) \to D(2)$ such that $f_i(2) = \lambda \circ f_i(2)$ for all $i \in I$. This can be interpreted as saying that categorical sums are unique up to isomorphism. $\square$

EXERCISE 13.10. Let $\mathtt{AbGrp}$ be the category of abelian groups and group homomorphisms. Let $\{A_i\}_{i\in I}$ be a collection of abelian groups. Let $\oplus_{i\in I} A_i$ be the sum of the A_i, so

$$\bigoplus_{i\in I} A_i = \left\{ (a_i)_{i\in I} \in \prod_{i\in I} A_i \mid a_i = 0 \text{ for all but finitely many } i \in I \right\}.$$

Prove that $S = \oplus_{i\in I} A_i$ together with the natural inclusions $\{\iota_i\colon A_i \hookrightarrow S\}_{i\in I}$ is the categorical sum of the $\{A_i\}_{i\in I}$ in the category $\mathtt{AbGrp}$. We remark that in the exercises to Chapter 21 we will prove that $\prod_{i\in I} A_i$ is the categorical product of the $\{A_i\}_{i\in I}$ (which in particular requires defining the categorical product). $\square$

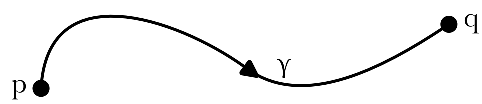
CHAPTER 14

Connectivity properties

Our next topic is connectivity and path connectivity.

14.1. Path connectivity

Recall that $I = [0,1]$. A *path* in a space X from $p \in X$ to $q \in X$ is a map $\gamma\colon I \to X$ with $\gamma(0) = p$ and $\gamma(1) = q$:



The space X is *path connected* if it is nonempty and for all $p, q \in X$ there exists a path in X from p to q.

EXAMPLE 14.1.1. The space $\mathbb{R}^n$ is path connected. Indeed, for $p, q \in \mathbb{R}^n$ the map $\gamma\colon I \to \mathbb{R}^n$ defined by

$$\gamma(t) = (1-t)p + tq \quad \text{for } t \in I$$

is a path from p to q. $\square$

14.2. Connectivity

We now turn to connectivity. It is easier to say what it means for a space to be disconnected. A space X is *disconnected* if we can write $X = U \sqcup V$ with $U, V \subset X$ disjoint nonempty open subsets of X. Since $X \setminus U = V$ and $X \setminus V = U$, the sets U and V are necessarily closed as well as open. Sets that are both open and closed are called *clopen sets*.[1]

EXAMPLE 14.2.1. The space $\mathbb{R} \setminus 0$ is disconnected. Indeed, $\mathbb{R} \setminus 0 = (-\infty, 0) \sqcup (0, \infty)$. $\square$

EXAMPLE 14.2.2. The space $\mathbb{Q}$ is disconnected. Indeed, for $U \subset \mathbb{R}$ let $U_{\mathbb{Q}} = U \cap \mathbb{Q}$. We then have $\mathbb{Q} = (-\infty, \sqrt{2})_{\mathbb{Q}} \sqcup (\sqrt{2}, \infty)_{\mathbb{Q}}$. $\square$

A space X is *connected* if it is nonempty and not disconnected. Another way of saying this is that a nonempty space X is connected if whenever $U, V \subset X$ are disjoint open sets with $X = U \sqcup V$, then either $U = X$ or $V = X$ (see Exercise 14.1).

14.3. Basic properties of connectivity

Here are some basic properties of connected spaces:

LEMMA 14.3.1. *Let X be a space and let $Y \subset X$ be a connected subspace. Then the closure $\overline{Y}$ of Y is connected.*

PROOF. Replacing X with $\overline{Y}$, we can assume that $\overline{Y} = X$. Our goal now is to prove that X is connected. Let $U, V \subset X$ be disjoint open subsets such that $X = U \sqcup V$. We must prove that either $U = X$ or $V = X$. Since Y is connected and $Y = (U \cap Y) \sqcup (V \cap Y)$, we must have either $U \cap Y = Y$ or $V \cap Y = Y$. Reordering U and V, we can assume that $U \cap Y = Y$, and hence $Y \subset X \setminus V$. Since $X \setminus V$ is closed, we have $\overline{Y} \subset X \setminus V$. Since $\overline{Y} = X$ and $X = U \sqcup V$, this implies that $U = X$, as desired. $\square$

LEMMA 14.3.2. *Let $f\colon X \to Y$ be a continuous map from a connected space X to a space Y. Then $f(X)$ is connected.*

[1] This is a terrible term, but is the standard word for this.

Proof. Replacing Y with $f(X)$, we can assume that $f(X) = Y$. Our goal now is to prove that Y is connected. Assume that $U, V \subset Y$ are disjoint open sets such that $Y = U \sqcup V$. We must prove that either $U = Y$ or $V = Y$. Since $f(X) = Y$, this is equivalent to showing that either $f^{-1}(U) = X$ or $f^{-1}(V) = X$. Since $Y = U \sqcup V$, we have that $f^{-1}(U)$ and $f^{-1}(V)$ are disjoint open subsets of X such that $X = f^{-1}(U) \sqcup f^{-1}(V)$. Since X is connected, we must therefore have either $f^{-1}(U) = X$ or $f^{-1}(V) = X$, as desired. □

Lemma 14.3.3. *Let X be a space and let $\{Y_j\}_{j \in J}$ be a nonempty collection of subspaces of X. Assume that:*

- *each Y_j is connected; and*
- *for all $j, k \in J$, the space $Y_j \cap Y_k$ is nonempty; and*
- *$X = \cup_{j \in J} Y_j$.*

Then X is connected.

Proof. Assume that $U, V \subset X$ are disjoint open sets such that $X = U \sqcup V$. We prove that either $U = X$ or $V = X$. Pick some $j_0 \in J$. Since Y_{j_0} is connected, either $U \cap Y_{j_0} = Y_{j_0}$ or $V \cap Y_{j_0} = Y_{j_0}$. Reordering U and V, assume that $U \cap Y_{j_0} = Y_{j_0}$. Now consider an arbitrary $j \in J$. Since $Y_j \cap Y_{j_0}$ is nonempty, we have $U \cap Y_j \neq \emptyset$. Since Y_j is connected, this implies that $U \cap Y_j = Y_j$. Since $U \cap Y_j = Y_j$ for all $j \in J$, we conclude that $U = X$, as desired. □

14.4. Path connected spaces are connected

We now prove:

Lemma 14.4.1. *Let X be a path connected space. Then X is connected.*

Proof. This is trivial if $X = \emptyset$, so assume that $X \neq \emptyset$. Fix a point $p \in X$. For each $q \in X$, pick a path $\gamma_q \colon I \to X$ from p to q. Set $Y_q = \gamma_q(I)$. Below in Lemma 14.4.2 we will prove that I is connected. Lemma 14.3.2 therefore implies that Y_q is connected. The space X is the union of the Y_q, and for $q, q' \in X$ we have $p \in Y_q \cap Y_{q'}$. By Lemma 14.3.3, we conclude that X is connected. □

The above proof required:

Lemma 14.4.2. *The space $I = [0, 1]$ is connected.*

Proof. Assume for the sake of contradiction that I is disconnected. We can therefore write $I = U \sqcup V$ with $U, V \subset I$ disjoint nonempty open sets. Define a set map $f \colon I \to \mathbb{R}$ via the formula

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

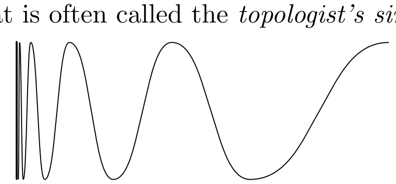
This is continuous; indeed, if $W \subset \mathbb{R}$ is any open set then $f^{-1}(W)$ is either $\emptyset$ or U or V or $U \cup V = I$. Since the image of f contains both 0 and 1, the intermediate value theorem implies that the image of f contains $1/2$. Since the image of f is $\{0, 1\}$, this is a contradiction. □

The converse of Lemma 14.4.1 is not true:

Example 14.4.3. Let X be the following subset of $\mathbb{R}^2$:

$$X = \{(0, y) \mid -1 \leq y \leq 1\} \cup \{(x, \sin(1/x) \mid x > 0\}.$$

This is a closed subset of $\mathbb{R}^2$ that is often called the *topologist's sine curve*:



The space X is *not* path connected; indeed, there is no path connecting $(0, 0)$ and $(x, \sin(1/x))$ for any $x > 0$ (see Exercise 14.4). However, X is connected (see Exercise 14.4). □

14.5. Path components

Let X be a space. Let $\sim$ be the following relation on the points of X: for $p, q \in X$, we have $p \sim q$ if there is a path in X from p to q. This is an equivalence relation on the points of X (see Exercise 14.2), and its equivalence classes are the *path components* of X. It is immediate from the definition that the path components of X are path connected and that X is the disjoint union of its path components.

EXAMPLE 14.5.1. Let X be the topologist's sine-curve from Example 14.4.3. The path components of X are as follows (see Exercise 14.4):

$$X_1 = \{(0, y) \mid -1 \leq y \leq 1\},$$
$$X_2 = \{(x, \sin(1/x) \mid x > 0\}.$$ □

14.6. Connected components

Continue to let X be a space. Now let $\sim$ be the following relation on the points of X: for $p, q \in X$, we have $p \sim q$ if there is a connected subspace $Y \subset X$ with $p, q \in Y$. Just like for path connectedness, we have:

LEMMA 14.6.1. *Letting the notation be as above, $\sim$ is an equivalence relation on the points of X.*

PROOF. It is clear that $\sim$ is reflexive, i.e., that $p \sim p$ for all $p \in X$. It is also clear that $\sim$ is symmetric, i.e., that if $p \sim q$ for some $p, q \in X$ then $q \sim p$. We must therefore only prove that $\sim$ is transitive. Consider $p, q, r \in X$ such that $p \sim q$ and $q \sim r$. We must prove that $p \sim r$. Since $p \sim q$ and $q \sim r$, there are connected subspaces $Y, Z \subset X$ with $p, q \in Y$ and $q, r \in Z$. Since Y and Z are connected subspaces with $q \in Y \cap Z$, it follows from Lemma 14.3.3 that $Y \cup Z$ is connected. Since $p, r \in Y \cup Z$, we conclude that $p \sim r$, as desired. □

The equivalence classes of $\sim$ are the *connected components* of X. The space X is the disjoint union of its connected components. They have the following properties:

LEMMA 14.6.2. *Let X be a space and let A be a connected component of X. Then A is closed and connected.*

PROOF. We first prove that A is connected. Fix some $a \in A$. For $p \in A$, there exists a connected subspace Y_p of X with $a, p \in Y_p$. Since Y_p is connected we have $a \sim q$ for all $q \in Y_p$, so $Y_p \subset A$. Since $A = \cup_{p \in Y} Y_p$ and each Y_p contains the point a, Lemma 14.3.3 impies that A is connected, as desired. Since A is connected, Lemma 14.3.1 implies that $\overline{A}$ is connected. It follows that $A = \overline{A}$, i.e., that A is closed. □

REMARK 14.6.3. Example 14.5.1 shows that path components need not be closed. □

14.7. Local connectivity

Since a space X is disconnected if we can write $X = U \sqcup V$ with $U, V \subset X$ disjoint nonempty clopen subsets, it is natural to hope that the connected components of X are clopen. Unfortunately, this need not hold:

EXAMPLE 14.7.1. Let $X = \mathbb{Q}$. The connected components of X and the path components of X both consist of the one-points sets $\{q\}$ for $q \in \mathbb{Q}$. □

REMARK 14.7.2. A space like $\mathbb{Q}$ whose connected components consist of single points is called *totally disconnected*. □

As this example suggests, the cause of this is pathological local behavior. A space X is *locally connected* at $p \in X$ if for all open neighborhoods U of p, there is a connected open neighborhood V of p with $V \subset U$. The space X is *locally connected* if it is locally connected at all $p \in X$. Similarly, a space X is *locally path connected* at $p \in X$ if for all open neighborhoods U of p, there is a path connected open neighborhood V of p with $V \subset U$. The space X is *locally path connected* if it is locally path connected at all $p \in X$. We then have:

LEMMA 14.7.3. *Let X be a space. Then:*

- *If X is locally connected, then all connected components of X are clopen.*
- *If X is locally path connected, then all path components of X are clopen.*

PROOF. The proofs for components and path components are similar, so we will prove it for path components.[2] Assume that X is locally path connected. Let Y be a path component of X. For $p \in Y$, since X is locally connected we can find a path connected open neighborhood V of p. By definition, p in the same path component as all the points of V, so $V \subset Y$. We deduce that Y is open. Since $X \setminus Y$ is the union of the path components of X that are different from Y and all those path components are open, it follows that $X \setminus Y$ is open and hence that Y is closed. ☐

COROLLARY 14.7.4. *Let X be a locally path connected space. Then the connected components and path components of X coincide.*

PROOF. Let Y be a connected component of X. The subspace Y is the disjoint union of a collection of path components. To prove that it is actually a path component, it is enough to prove that Y is path connected. Assume otherwise. We can then write $Y = Y_1 \cup Y_2$ with each Y_i a nonempty union of path components and $Y_1 \cap Y_2 = \emptyset$. Lemma 14.7.3 implies that each path component is clopen, so both Y_1 and Y_2 are also clopen. Since $Y = Y_1 \sqcup Y_2$, we deduce that Y is disconnected, contradicting the fact that it is connected. ☐

REMARK 14.7.5. As our examples show, not all metric spaces (or even subspaces of $\mathbb{R}^n$) are locally connected or locally path connected. However, most spaces that appear in algebraic topology are locally path connected. ☐

14.8. Exercises

EXERCISE 14.1. Let X be a space with $X \neq \emptyset$. Prove the following:

(a) The space X is connected if whenever $U, V \subset X$ are disjoint open sets with $X = U \sqcup V$, then either $U = X$ or $V = X$.
(b) The space X is connected if whenever $U, V \subset X$ are disjoint open sets with $X = U \sqcup V$, then either $U = \emptyset$ or $V = \emptyset$.
(c) The space X is connected if whenever $U, V \subset X$ are open sets with $X = U \sqcup V$, then $U \cap V \neq \emptyset$. ☐

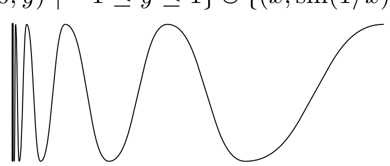
EXERCISE 14.2. Let X be a space. Prove that the following is an equivalence relation on the points of X:

(a) For $p, q \in X$, the relation where p is equivalent to q if there is a path in X from p to q. ☐

EXERCISE 14.3. Let X and Y be spaces. Give $X \times Y$ the product topology (see Example 12.7.4).

(a) If X and Y are connected, prove that $X \times Y$ is connected.
(b) If X and Y are path connected, prove that $X \times Y$ is path connected. Be careful to prove that your paths are continuous! ☐

EXERCISE 14.4. Let $X = \{(0, y) \mid -1 \leq y \leq 1\} \cup \{(x, \sin(1/x) \mid x > 0\} \subset \mathbb{R}^2$.



This space is called the topologist's sine curve. You proved in Exercise 3.14** that X is not path connected. Prove that X is connected. ☐

EXERCISE 14.5. Let $I = [0, 1]$ and let $\leq$ be the dictionary order topology on I^2, so $(x, y) \leq (z, w)$ if $x \leq z$ or if $x = z$ and $y \leq w$. Let X the topological space on the set I^2 with the corresponding order topology (see Example 12.7.3). Prove the following:

(a) X is connected.

[2]In fact, this case is slightly harder since in general path components need not be closed.

(b) X is not path connected. $\square$

EXERCISE 14.6. Let C be the classical Cantor set (see Example 2.7.3), so C consists of all $x \in I = [0,1]$ of the form

$$x = \sum_{n=1}^{\infty} \frac{x_n}{3^n} \quad \text{with } x_n \in \{0,2\} \text{ for all } n \geq 1.$$

Prove that C is totally disconnected. $\square$

EXERCISE 14.7. Let S be the Sorgenfrey line (see Example 12.7.5), so S is $\mathbb{R}$ with the topology given by the basis $\{[a,b) \mid a < b\}$ of all half-open intervals in $\mathbb{R}$. Prove that S is totally disconnected. $\square$

EXERCISE 14.8. Let X and Y be spaces such that there exists a homeomorphism $f\colon \mathbb{R} \to X \times Y$. Here we topologize $X \times Y$ using the product topology as in Example 12.7.4. Prove that either X or Y is a point. $\square$

CHAPTER 15

Countability properties

This chapter discussed properties that ensure a topological space is not "too large".

15.1. First countability

Let X be a space. A *neighborhood basis* for X at a point $p \in X$ is a collection $\mathfrak{B}_p$ of open neighborhoods of p such that:

- For all open neighborhoods V of p, we have $U \subset V$ for some $U \in \mathfrak{B}_p$.

The space X is *first countable* if it has a countable neighborhood basis at each point $p \in X$. All metrizable spaces have this property:

LEMMA 15.1.1. *Let M be a metrizable space. Then M is first countable.*

PROOF. Fix a metric on M inducing its topology. Recall that $B_r(p)$ is the open ball of radius $r > 0$ around $p \in M$. For $p \in M$, the set $\{B_r(p) \mid r > 0 \text{ rational}\}$ is a countable neighborhood basis for X at p. □

15.2. Sequences

Let X be a space. A *sequence* in X is an ordered collection $\{x_n\}_{n \geq 1}$ of points of X. Given such a sequence, a point $y \in X$ is its *limit* if for all open neighborhoods U of y there is some $N \geq 1$ such that $x_n \in U$ for $n \geq N$. If y is a limit of $\{x_n\}_{n \in X}$, then we write $\lim_{n \mapsto \infty} x_n = y$ and say that $\{x_n\}_{n \geq 1}$ *converges* to y. If such a y exists, then we say that $\{x_n\}_{n \geq 1}$ is a *convergent sequence*. If X is metrizable, then this agrees with our previous definition (see Exercise 15.1).

REMARK 15.2.1. Be warned that a sequence can have multiple distinct limits. This only happens for fairly pathological spaces. In the next chapter, we introduce a property of spaces called being Hausdorff that forces convergent sequences to have unique limits. □

15.3. Closure

Sequences are most useful for first countable spaces. For instance, if X is first countable, then for $A \subset X$ we can construct the closure $\overline{A}$ using limits:

LEMMA 15.3.1. *Let X be a first countable space and let $A \subset X$. Then $\overline{A}$ is the set of all $y \in A$ such that there exists a sequence $\{a_n\}_{n \geq 1}$ of points of A such that $\lim_{n \mapsto \infty} a_n = y$.*

PROOF. Let B be the set of limits of sequences of points of A. We first prove that $B \subset \overline{A}$. Let $b \in B$ and let $C \subset X$ be a closed set with $A \subset C$. We must prove that $b \in C$. Indeed, if $b \notin C$ then we can find an open neighborhood U of b such that $U \subset X \setminus C$. However, since $b \in B$ there must exist points of $A \subset C$ in U, contradicting the fact that U is disjoint from C.

We next prove that $\overline{A} \subset B$. This uses first countability. Consider a point $p \in \overline{A}$. Each open neighborhood V of p must contain a point of A. Let $\mathfrak{B}_p = \{U_1, U_2, \ldots\}$ be a countable neighborhood basis at p. For each $n \geq 1$, choose $x_n \in U_n$ with $x_n \in A$. We then have $\lim_{n \mapsto \infty} x_n = p$, so $p \in B$. □

REMARK 15.3.2. Though metrizable spaces are first countable, not all spaces that appear in algebraic topology are first countable. This is why arguments using limits are mostly avoided in this book. There are generalizations of sequences and limits (nets, filters, etc.) that work for spaces that are not first countable (see [1]), but in practice they do not simplify arguments in algebraic topology. □

15.4. Second countability

A space X is *second countable* if there is a countable basis for its topology. It is clear that all second countable spaces are first countable. It is not true that all metrizable spaces are second countable, but all subspaces of $\mathbb{R}^n$ are second countable:

Lemma 15.4.1. *Let X be a subspace of $\mathbb{R}^n$. Then X is second countable.*

Proof. For all $p \in \mathbb{R}^n$ and $r > 0$, let $B_r(p) \subset \mathbb{R}^n$ be the open ball around p. Then X has the countable basis $\{B_r(p) \cap X \mid p \in \mathbb{Q}^n \text{ and } r > 0 \text{ rational}\}$. □

15.5. Separability

There is one further countability condition that occasionally shows up. For a space X, a set $A \subset X$ is *dense* if its closure $\overline{A}$ equals X. The space X is *separable* if X has a countable dense subset. This is slightly weaker than second countability:

Lemma 15.5.1. *Let X be a second countable space. Then X is separable.*

Proof. Let $\mathfrak{B} = \{U_1, U_2, \ldots, \}$ be a countable basis for the topology of X. Pick $x_n \in U_n$. Then the set $\{x_n \mid n \geq 1\}$ is a countable dense set in X. □

For metrizable spaces, these two notions coincide:

Lemma 15.5.2. *Let M be a separable metrizable space. Then M is second countable.*

Proof. Fixing a metric on M inducing its topology, the proof is similar to that of Lemma 15.4.1: if $A \subset M$ is a countable dense set, then $\left\{B_{1/n}(a) \mid a \in A,\, n \geq 1\right\}$ is a countable basis for the topology on M. □

15.6. Exercises

Exercise 15.1. Let M be a metric space. Prove that the notion of limit from §15.2 agrees with the notion of limit for metric spaces from §2.5. □

Exercise 15.2. Let X be second countable. Prove the following:

(a) Let $\mathfrak{U}$ be a collection of open subsets of X with $X = \cup_{U \in \mathfrak{U}} U$. Prove that there is a countable subset $\mathfrak{U}' \subset \mathfrak{U}$ such that $X = \cup_{U \in \mathfrak{U}} U$.

(b) Let $\mathfrak{B}$ be any basis for the topology on X. Prove that there is a countable subset $\mathfrak{B}' \subset \mathfrak{B}$ that is also a basis for the topology on X. □

Exercise 15.3. Let X be second countable and let $\mathfrak{U}$ be a collection of disjoint open subsets of X. Prove that $\mathfrak{U}$ is countable. □

Exercise 15.4. As in Example 13.2.7, let X be the wedge product of countably many intervals $I_n = I$ equipped with the basepoints $0 \in I_n$. Prove that X is not first countable, and deduce that X is not metrizable. □

Exercise 15.5. Let $I = [0, 1]$ and let $\leq$ be the dictionary order topology on I^2, so $(x, y) \leq (z, w)$ if $x \leq z$ or if $x = z$ and $y \leq w$. Let X the topological space on the set I^2 with the corresponding order topology (see Example 12.7.3). Prove the following:

(a) X is first countable.

(b) X is not second countable. □

Exercise 15.6. Let S be the Sorgenfrey line (see Example 12.7.5), so S is $\mathbb{R}$ with the topology given by the basis $\{[a, b) \mid a < b\}$ of all half-open intervals in $\mathbb{R}$. Prove the following:

(a) S is first countable.

(b) S is separable.

(c) S is not second countable. □

Bibliography

[1] J. Kelley, *General topology*, D. Van Nostrand Co., Inc., Toronto-New York-London, 1955. (Cited on page 125.)

CHAPTER 16

Separation properties and the Tietze extension theorem

This chapter discusses properties that are necessary to ensure that continuous functions have the properties one would expect.

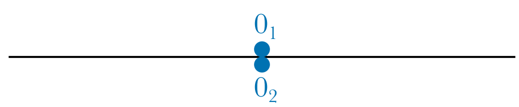
16.1. Pathology

Consider maps $f, g\colon X \to Y$. If $A \subset X$ is dense and $f|_A = g|_A$, then it is natural to expect that $f = g$. Unfortunately, this need not hold:

Example 16.1.1 (Line with two origins). As a set, let $Y = (\mathbb{R} \setminus \{0\}) \sqcup \{0_1, 0_2\}$. For $i = 1, 2$, let $f_i\colon \mathbb{R} \to Y$ be the map defined by $f_i(x) = x$ for $x \in \mathbb{R} \setminus \{0\}$ and $f_i(0) = 0_i$. Give Y the identification space topology, so:

- a set $U \subset Y$ is open if and only if $f_1^{-1}(U)$ and $f_2^{-1}(U)$ are open in $\mathbb{R}$.

Here is a picture of Y:



With this topology, the subspaces $Y \setminus \{0_2\} = f_1(\mathbb{R})$ and $Y \setminus \{0_1\} = f_2(\mathbb{R})$ are both homeomorphic to $\mathbb{R}$. The maps $f_1, f_2\colon \mathbb{R} \to Y$ are continuous and agree on the dense set $\mathbb{R} \setminus \{0\}$. However, $f_1 \neq f_2$. $\square$

16.2. Hausdorff spaces

The issue with the line with two origins from Example 16.1.1 is that the points 0_1 and 0_2 do not have disjoint open neighborhoods. To rule this out, say that a space X is *Hausdorff* if for all distinct points $p, q \in X$, there are open neighborhoods U of p and V of q with $U \cap V = \emptyset$. This has a number of nice consequences (see Exercise 16.3):

- All points in X are closed, i.e., for all $p \in X$ the one-point set $\{p\}$ is closed.
- If Z is another space and $f, g\colon Z \to X$ are two maps, then the subset $\{z \in Z \mid f(z) = g(z)\}$ of points in Z where f and g are equal is closed. In particular, if f and g agree on a dense subset of Z, then $f = g$.
- Limits in X are unique in the following sense. Let $\{x_n\}_{n \geq 1}$ be a sequence in X and let $y_1, y_2 \in X$ be such that $\lim_{n \mapsto \infty} x_n = y_1$ and $\lim_{n \mapsto \infty} x_n = y_2$. Then $y_1 = y_2$.

Most geometrically natural spaces are Hausdorff. In particular:

Lemma 16.2.1. *Let M be a metrizable space. Then M is Hausdorff.*

Proof. Fix a metric on M inducing its topology. For distinct $p, q \in M$, let $\epsilon = \mathfrak{d}(p, q)/2$. The open balls $B_\epsilon(p)$ and $B_\epsilon(q)$ are disjoint. $\square$

Remark 16.2.2. For an infinite field $\mathbf{k}$, an important non-example is given by the Zariski topology on $\mathbf{k}^n$. See Exercise 16.4. $\square$

16.3. Continuity

For first countable Hausdorff spaces, we can characterize continuity with sequences:

Lemma 16.3.1. *Let X be a first countable Hausdorff space, let Y be a Hausdorff space, and let $f\colon X \to Y$ be a map of sets. Then f is continuous if and only if the following holds:*[1]

[1]There is a version of this result that is true without the Hausdorff assumption, but it is awkward to state since in non-Hausdorff spaces limits need not be unique.

- *Let* $\{x_n\}_{n\geq 1}$ *be a convergent sequence in* X*. Then* $\{f(x_n)\}_{n\geq 1}$ *is a convergent sequence in* Y *and* $\lim_{n\mapsto\infty} f(x_n) = f(\lim_{n\mapsto\infty} x_n)$.

Proof. See Exercise 16.2. □

16.4. Normal spaces

In fact, most geometrically natural spaces have even stronger separation properties. A space X is *normal* if it satisfies the following two conditions:

- for all disjoint closed sets $C, D \subset X$, there exist open neighborhoods U of C and V of D with $U \cap V = \emptyset$; and
- all points in X are closed.[2]

All normal spaces are Hausdorff. The key example is:

Lemma 16.4.1. *Let* M *be a metrizable space. Then* M *is normal.*

Proof. Fix a metric $\mathfrak{d}$ on M inducing its topology Since M is Hausdorff, all points in M are closed. Consider disjoint closed sets $C, D \subset M$. For $z \in M$, let

$$r(z) = \inf\{\mathfrak{d}(z,c) \mid c \in C\} \quad \text{and} \quad s(z) = \inf\{\mathfrak{d}(z,d) \mid d \in D\}.$$

Since C and D are disjoint closed sets, we have $r(d) > 0$ for $d \in D$ and $s(c) > 0$ for $c \in C$. Define

$$U = \bigcup_{c\in C} B_{s(c)/3}(c) \quad \text{and} \quad V = \bigcup_{d\in D} B_{r(d)/3}(d).$$

The sets U and V are open, and $C \subset U$ and $D \subset V$. To prove the lemma, it is enough to show that $U \cap V = \emptyset$. Assume this is false, and let $x \in U \cap V$. We can therefore find $c_0 \in C$ and $d_0 \in D$ such that $\mathfrak{d}(c_0, x) < s(c_0)/3$ and $\mathfrak{d}(d_0, x) < r(d_0)/3$. We either have $s(c_0) \leq r(d_0)$ or $r(d_0) \leq s(c_0)$. Both cases lead to a similar contradiction, so we will give the details for $s(c_0) \leq r(d_0)$. This implies that

$$\mathfrak{d}(c_0, d_0) \leq \mathfrak{d}(c_0, x) + \mathfrak{d}(x, d_0) < s(c_0)/3 + r(d_0)/3 \leq r(d_0)/3 + r(d_0)/3 = \frac{2}{3} r(d_0).$$

However, we also have $\mathfrak{d}(c_0, d_0) \geq \inf\{\mathfrak{d}(d_0, c) \mid c \in C\} = r(d_0)$, a contradiction. □

The following characterization of normality is often useful. Recall that $\overline{V}$ is the closure of V.

Lemma 16.4.2. *A space* X *is normal if and only if all points in* X *are closed and:*

(♠) *For all closed sets* $C \subset X$ *and all open neighborhoods* U *of* C*, there exists an open neighborhood* V *of* C *with* $\overline{V} \subset U$.

Proof. Assume first that X is normal. To verify (♠), let $C \subset X$ be closed and let U be an open neighborhood of C. The set $D = X \setminus U$ is then a closed set that is disjoint from C, so by normality there exist disjoint open neighborhoods V and W of C and D. Since $X \setminus W$ is a closed subset of U containing V, it follows that $\overline{V} \subset U$.

Assume now that all points in X are closed and (♠) holds. To verify normality, let $C, D \subset X$ be disjoint closed sets. Applying (♠) to the open neighborhood $U = X \setminus D$ of C, we obtain an open neighborhood V of C with $\overline{V} \subset U$. It follows that V and $W = X \setminus \overline{V}$ are disjoint open neighborhoods of C and D. □

16.5. Urysohn's Lemma

A key feature of normal spaces is that they have a rich supply of continuous real-valued functions. For our first example of this, we need a definition. The *support* of a function $f\colon X \to \mathbb{R}$, denoted $\operatorname{supp}(f)$, is the closure of the set $\{p \in X \mid f(p) \neq 0\}$. We then have:

Theorem 16.5.1 (Urysohn's Lemma). *Let* X *be a normal space, let* $C \subset X$ *be closed, and let* $U \subset X$ *be an open neighborhood of* C*. Then there exists a map* $f\colon X \to I = [0,1]$ *such that* $f|_C = 1$ *and* $\operatorname{supp}(f) \subset U$*.*

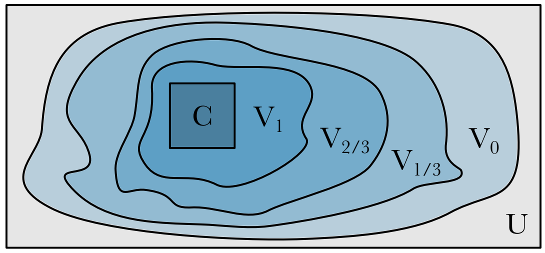
Proof. We must use the open sets provided by normality to construct f. The key is:

[2] This is not always included in the definition of normality, but it ensures that normal spaces are Hausdorff.

Claim. *There exist open sets V_α for all $\alpha \in \mathbb{Q}$ with the following properties:*

(i) For rational $r < s$, we have $\overline{V}_s \subset V_r$.
(ii) $C \subset V_1$ and $\overline{V}_0 \subset U$.
(iii) $V_\alpha = \emptyset$ for $\alpha > 1$ and $V_\alpha = X$ for $\alpha < 0$.

Proof of claim. The picture is as follows:



Define V_α for $\alpha > 1$ and $\alpha < 0$ as in (iii). Next, using Lemma 16.4.2 choose an open neighborhood V_0 of C with $\overline{V}_0 \subset U$ and an open neighborhood V_1 of C with $\overline{V}_1 \subset \overline{V}_0$. Conditions (ii) and (iii) hold, and we inductively construct the remaining V_α satisfying (i) as follows. Enumerate the rational numbers in I as $\{\alpha_0, \alpha_1, \ldots\}$ with $\alpha_0 = 0$ and $\alpha_1 = 1$. We have already constructed V_0 and V_1, so assume that $n \geq 2$ and that we have constructed V_{α_m} for $0 \leq m \leq n-1$ satisfying (i). We construct V_{α_n} as follows. Let

$$r = \max\{\alpha_m \mid 0 \leq m \leq n-1,\ \alpha_m < \alpha_n\} \quad \text{and} \quad s = \min\{\alpha_m \mid 0 \leq m \leq n-1,\ \alpha_m > \alpha_n\}.$$

We thus have $r < \alpha_n < s$, and by (iii) we have $\overline{V}_s \subset V_r$. Using Lemma 16.4.2, we can then find an open neighborhood V_{α_n} of $\overline{V}_s$ such that $\overline{V}_{\alpha_n} \subset V_r$. □

We now define a set map $f\colon X \to \mathbb{R}$ via the formula

$$f(p) = \sup\{\alpha \in \mathbb{Q} \mid p \in V_\alpha\}.$$

By (iii) we have $f(p) \in I$ for all $p \in X$. Also, by (ii) we have $f(p) = 1$ for $p \in C$ and $\operatorname{supp}(f) \subset U$. All that remains is to check that f is continuous.

Let $W \subset \mathbb{R}$ be open. We must prove that $f^{-1}(W)$ is open. Consider $p \in f^{-1}(W)$. Choose rational $r < s$ such that $[r,s] \subset W$ and $f(p) \in [r,s]$. By (iii), we have $\overline{V}_s \subset V_r$. To prove that $f^{-1}(W)$ is open, it is enough to prove that the open set $V_r \setminus \overline{V}_s$ is contained in $f^{-1}(W)$. To do this, it is enough to prove that f maps $V_r \setminus \overline{V}_s$ into $[r,s] \subset W$. This follows from the following two facts, both of which are immediate from (iii):

- for $q \in V_r$, we have $f(q) \geq r$; and
- for $q \notin V_s$, we have $f(q) \leq s$. □

16.6. Converse to Urysohn

The following lemma shows that the conclusion of Urysohn's lemma characterizes normality:

Lemma 16.6.1. *Let X be a space such that all points in X are closed. For every closed $C \subset X$ and every neighborhood U of C, assume that there exists a continuous map[3] $f\colon X \to \mathbb{R}$ with $f|_C = 1$ and $\operatorname{supp}(f) \subset U$. Then X is normal.*

Proof. Let C and D be disjoint closed sets in X. By assumption, there is a continuous map $f\colon X \to \mathbb{R}$ with $f|_C = 1$ and $\operatorname{supp}(f) \subset X \setminus D$. The sets $U = f^{-1}((1/2,\infty))$ and $V = X \setminus \operatorname{supp}(f)$ are then disjoint open neighborhoods of C and D. □

[3] In Urysohn's lemma, the target of f is $I = [0,1]$. Here we relax this.

16.7. Perfectly normal spaces

Say that a space X is *perfectly normal* if points in X are closed and for all closed $C \subset X$ and all open neighborhoods U of C, there exists a continuous map $f: X \to I = [0,1]$ such that $f^{-1}(1) = C$ and $\operatorname{supp}(f) \subset U$. Lemma 16.6.1 implies that perfectly normal spaces are normal.

The definition of a perfectly normal space resembles the conclusion of Urysohn's lemma, but there is a small difference: in a perfectly normal space we have $f^{-1}(1) = C$, while in the conclusion of Urysohn's lemma we only have $C \subset f^{-1}(1)$. Most geometrically natural spaces are perfectly normal. In particular:

LEMMA 16.7.1. *Let M be a metrizable space. Then M is perfectly normal.*

PROOF. Fix a metric $\mathfrak{d}$ on M inducing its topology. Lemma 16.4.1 implies that M is normal, and in particular points are closed. Consider $C \subset X$ closed and U an open neighborhood of C. By Urysohn's Lemma, there exists a continuous map $f: X \to I$ such that $f|_C = 1$ and $\operatorname{supp}(f) \subset U$. We want to modify f to ensure it is less than 1 at all points that do not lie in C. Let $g: X \to \mathbb{R}$ be the function

$$g(p) = \inf \{\mathfrak{d}(p,c) \mid c \in C\} \quad \text{for } p \in X$$

and let $h: X \to I$ be the function

$$h(p) = \min(g(p), 1) \quad \text{for } p \in X.$$

Both g and h are continuous and satisfy $g^{-1}(0) = h^{-1}(0) = C$. The function $f': X \to I$ defined by

$$f'(p) = (1 - h(p)) \cdot f(p) \quad \text{for all } p \in M$$

then satisfies $(f')^{-1}(1) = C$ and $\operatorname{supp}(f') \subset U$. □

REMARK 16.7.2. We have introduced the notion of a space being Hausdorff, being normal, and being perfectly normal. These are called *separation axioms*. It is common to call a Hausdorff space a T_2-*space*, a normal space a T_4-*space*, and a perfectly normal space a T_6-*space*. As this terminology suggests, there are many other separation axioms as well.[4] For instance, we will discuss regular spaces (i.e., T_3-spaces) below in §16.10. The vast majority of spaces considered in algebraic topology are perfectly normal. □

16.8. Uniform limits of functions

Our next goal is to prove the Tietze extension theorem, which says that continuous real-valued functions on closed subsets of normal spaces can be extended to the whole space. The extension we construct will be a limit of functions constructed using Urysohn's Lemma. We therefore need a way to certify that such functions are continuous.

Let X be a space. A sequence of functions $f_n: X \to \mathbb{R}$ is said to *converge uniformly* to a function $f: X \to \mathbb{R}$ if the following holds:

- for all $\epsilon > 0$, there exists some $N \geq 1$ such that $|f(p) - f_n(p)| < \epsilon$ for all $n \geq N$ and $p \in X$.

We then have the following, which generalizes a familiar fact from real analysis:

LEMMA 16.8.1. *Let X be a space and let $f_n: X \to \mathbb{R}$ be a sequence of continuous functions converging uniformly to a function $f: X \to \mathbb{R}$. Then f is continuous.*

PROOF. This can be proved using an argument similar to the one used to prove the analogous fact for functions defined on $X = \mathbb{R}$. See Exercise 16.10. □

16.9. Tietze Extension Theorem

We can now prove the Tietze Extension Theorem:

THEOREM 16.9.1 (Tietze Extension Theorem). *Let X be a normal space, let $C \subset X$ be closed, and let $f: C \to \mathbb{R}$ be a continuous function. Then f can be extended to a continuous function $F: X \to \mathbb{R}$. Moreover, if the image of f lies in a closed interval $[a,b]$ then F can be chosen such that its image also lies in $[a,b]$.*

[4] In fact, not only are there T_k-spaces for $0 \leq k \leq 6$, but there are even $T_{2.5}$-spaces and $T_{3.5}$-spaces.

PROOF. We first prove the case where f is bounded, and then derive the unbounded case.

CASE 1. *The theorem holds if the image of f lies in a closed interval $[a, b]$.*

Since $[a, b] \cong [-1, 1]$, we can assume without loss of generality that $[a, b] = [-1, 1]$. For $n \geq 1$, we will construct continuous functions $G_n \colon X \to \mathbb{R}$ such that letting $F_n = G_1 + \cdots + G_n$, we have:

(i) The function F_n satisfies $|f(p) - F_n(p)| \leq (2/3)^n$ for all $p \in C$.
(ii) The function G_n satisfies $|G_n(p)| \leq (1/3)(2/3)^{n-1}$ for all $p \in X$.

Condition (ii) will imply that the functions $F_n = G_1 + \cdots + G_n$ converge uniformly to a function F such that

$$|F(p)| \leq \frac{1}{3}\left(1 + (2/3) + (2/3)^2 + \cdots\right) = \frac{1}{3}\left(\frac{1}{1 - 2/3}\right) = 1 \quad \text{for all } p \in X.$$

Lemma 16.8.1 implies that $F \colon X \to [-1, 1]$ is continuous, and condition (i) implies that $F|_C = f$.

It remains to construct the G_n. Assume that $n \geq 1$ and we have constructed $G_1, \ldots, G_{n-1}$ satisfying (ii) such that letting $F_{n-1} = G_1 + \cdots + G_{n-1}$, we have

$$|f(p) - F_{n-1}(p)| \leq (2/3)^{n-1} \quad \text{for all } p \in C. \tag{16.9.1}$$

This is vacuous for $n = 1$. We will construct G_n as follows. Let

$$L = \left\{p \in C \mid f(p) - F_{n-1}(p) \leq -(1/3)(2/3)^{n-1}\right\}$$
$$R = \left\{p \in C \mid f(p) - F_{n-1}(p) \geq (1/3)(2/3)^{n-1}\right\}.$$

The sets L and R are disjoint closed sets. Using Urysohn's lemma, we can find:

- a continuous map $h_L \colon X \to I$ with $h_L|_L = 1$ and $\operatorname{supp}(h_L) \subset X \setminus R$; and
- a continuous map $h_R \colon X \to I$ with $h_R|_R = 1$ and $\operatorname{supp}(h_R) \subset X \setminus L$.

Let $G_n \colon X \to [-(1/3)(2/3)^{n-1}, (1/3)(2/3)^{n-1}]$ be the map

$$G_n = -(1/3)(2/3)^{n-1} h_L + (1/3)(2/3)^{n-1} h_R.$$

By construction, G_n satisfies (ii). To show that $F_n = F_{n-1} + G_n$ satisfies (i), consider some $p \in C$. There are three cases:

- If $p \in L$, then by (16.9.1) we have

$$|f(p) - F_n(p)| = |f(p) - F_{n-1}(p) + (1/3)(2/3)^{n-1}| \leq (2/3)^{n-1} - (1/3)(2/3)^{n-1} = (2/3)^n.$$

- If $p \in R$, then by (16.9.1) we have

$$|f(p) - F_n(p)| = |f(p) - F_{n-1}(p) - (1/3)(2/3)^{n-1}| \leq (2/3)^{n-1} - (1/3)(2/3)^{n-1} = (2/3)^n.$$

- If $p \notin L \cup R$, then by definition we have $|f(p) - F_{n-1}(p)| \leq (1/3)(2/3)^{n-1}$, so since $|G_n(p)| \leq (1/3)(2/3)^{n-1}$ we have

$$|f(p) - F_n(p)| = |f(p) - F_{n-1}(p) - G_n(p)| \leq (1/3)(2/3)^{n-1} + (1/3)(2/3)^{n-1} = (2/3)^n.$$

In all three cases, (ii) is satisfied. The theorem follows.

CASE 2. *The theorem holds in general.*

Since $\mathbb{R} \cong (-1, 1)$, it is enough to prove that every continuous function $f \colon C \to (-1, 1)$ can be extended to a continuous function $F \colon X \to (-1, 1)$. By Case 1, we can extend f to a continuous function $F' \colon X \to [-1, 1]$. Our goal is to modify F' such that its image does not contain -1 or 1. Set $U = (F')^{-1}((-1, 1))$. Applying Urysohn's Lemma (Theorem 16.5.1), there exists a continuous function $g \colon X \to I$ with $g|_C = 1$ and $\operatorname{supp}(g) \subset U$. The product $F = g \cdot F'$ then still extends f and satisfies $F(X) \subset (-1, 1)$. $\square$

16.10. Regular spaces

We close this chapter by discussing one further separation property that will play a small technical role at several points throughout the rest of this book. A space X is *regular* if it satisfies the following two conditions:

- for all points $p \in X$ and closed sets $C \subset X$ with $p \notin C$, there exist open neighborhoods U of p and V of C with $U \cap V = \emptyset$; and
- all points in X are closed.[5]

All normal spaces are regular and all regular spaces are Hausdorff. There exist Hausdorff spaces all of whose points are closed that are not regular (see Exercise 16.5). There also exist regular spaces that are not normal, but these are fairly pathological; see [**1**]. The following lemma shows that many regular spaces are normal:

LEMMA 16.10.1. *Let X be a space that is regular and second countable. Then X is normal.*

PROOF. Let $A, B \subset X$ be disjoint closed sets. To prove that X is normal, we must find disjoint open neighborhoods U and V of A and B. The key to this is:

CLAIM. *There exist countable collections of open sets $\{U_n\}_{n\geq 1}$ and $\{V_n\}_{n\geq 1}$ such that:*

- $A \subset \cup_{n=1}^{\infty} U_n$ *and* $B \subset \cup_{n=1}^{\infty} V_n$*; and*
- *For all $n \geq 1$, we have $\overline{U}_n \cap B = \emptyset$ and $\overline{V}_n \cap A = \emptyset$.*

PROOF OF CLAIM. The constructions of the $\{U_n\}_{n\geq 1}$ and the $\{V_n\}_{n\geq 1}$ are identical, so we will explain how to construct the $\{U_n\}_{n\geq 1}$. Let $\mathcal{B}$ be a countable basis for X. For $a \in A$, since X is regular we can find disjoint open neighborhoods U_a' of a and V_a' of B. Choose $U_a \in \mathcal{B}$ such that U_a is an open neighborhood of a with $U_a \subset U_a'$. We then have $\overline{U}_a \subset \overline{U}_a' \subset B \setminus V_a'$, so $\overline{U}_a \cap B = \emptyset$. For the $\{U_n\}_{n\geq 1}$ we can then take an enumeration of the countable set $\{U_a \mid a \in A\} \subset \mathcal{B}$. □

The naive thing to do now would be to take our open neighborhoods of A and B to be $\cup_{n=1}^{\infty} U_n$ and $\cup_{n=1}^{\infty} V_n$, but these would not necessarily be disjoint. To fix this, for $n \geq 1$ define

$$U_n' = U_n \setminus \bigcup_{k=1}^{n} \overline{V}_k \quad \text{and} \quad V_n' = V_n \setminus \bigcup_{k=1}^{n} \overline{U}_k.$$

Since each $\overline{V}_k$ is disjoint from A, we still have $A \subset \cup_{n=1}^{\infty} U_n'$. Similarly, we still have $B \subset \cup_{n=1}^{\infty} V_n'$. Now, however, we have $U_n' \cap V_m' = \emptyset$ for all $n, m \geq 1$. It follows that $U = \cup_{n=1}^{\infty} U_n$ and $V = \cup_{n=1}^{\infty} V_n$ are disjoint open neighborhoods of A and B, as desired. □

16.11. Exercises

EXERCISE 16.1. Let X be Hausdorff and let $x_1, \ldots, x_n \in X$ be distinct points. Prove that there exist open neighborhoods U_i of the x_i such that $U_i \cap U_j = \emptyset$ for all $1 \leq i, j \leq n$ distinct. □

EXERCISE 16.2. Let X be a first countable Hausdorff space, let Y be a Hausdorff space, and let $f\colon X \to Y$ be a map of sets. Then f is continuous if and only if the following holds:

- Let $\{x_n\}_{n\geq 1}$ be a convergent sequence in X. Then $\{f(x_n)\}_{n\geq 1}$ is a convergent sequence in Y and $\lim_{n\mapsto\infty} f(x_n) = f(\lim_{n\mapsto\infty} x_n)$. □

EXERCISE 16.3. Let X be a Hausdorff space. Prove the following:

(a) All points in X are closed, i.e., for all $p \in X$ the one-point set $\{p\}$ is closed.
(b) If Z is another space and $f, g\colon Z \to X$ are two maps, then the subset $\{z \in Z \mid f(z) = g(z)\}$ of points in Z where f and g are equal is closed. In particular, if f and g agree on a dense subset of Z, then $f = g$.
(c) Let $\{x_n\}_{n\geq 1}$ be a sequence in X and let $y_1, y_2 \in X$ be such that $\lim_{n\mapsto\infty} x_n = y_1$ and $\lim_{n\mapsto\infty} x_n = y_2$. Then $y_1 = y_2$. □

EXERCISE 16.4. Let $\mathbf{k}$ be a field. Prove that the Zariski topology on $\mathbf{k}^n$ described in Example 12.6.2 is Hausdorff if and only if $\mathbf{k}$ is a finite field. □

[5]This is not always included in the definition of regularity, but it ensures that regular spaces are Hausdorff.

Exercise 16.5. As a set, let $X = \mathbb{R}$. Let $\mathcal{B}$ be the set of open subsets of $X = \mathbb{R}$ in the standard topology on $\mathbb{R}$ and let

$$\mathcal{B}' = \mathcal{B} \cup \{U \cap \mathbb{Q} \mid U \in \mathcal{B}\}.$$

Prove the following:

- The set $\mathcal{B}'$ is the basis for a topology on X that we will call the $\mathcal{B}'$-topology.
- The $\mathcal{B}'$-topology on X is Hausdorff and all points are closed.
- The $\mathcal{B}'$-topology on X is not regular. □

Exercise 16.6. Let X be a Hausdorff space and let $Y \subset X$ be a subspace. Prove that Y is Hausdorff. □

Exercise 16.7. Let X be a normal space and let $Y \subset X$ be a closed subspace. Prove that Y is normal. We remark that this need not hold if Y is not closed. See [**1**] for examples. □

Exercise 16.8. Say that a space X is *completely normal* if every subspace of X is normal. Prove that X is completely normal if and only if for every pair of subsets $A, B \subset X$ with $\overline{A} \cap B = \emptyset$ and $A \cap \overline{B} = \emptyset$, there exist disjoint open neighborhoods U and V of A and B. □

Exercise 16.9. Let $f\colon X \to Y$ be a quotient map that is also closed. Assume that X is normal. Prove that Y is normal. □

Exercise 16.10. Let X be a space and let $f_n\colon X \to \mathbb{R}$ be a sequence of continuous functions converging uniformly to a function $f\colon X \to \mathbb{R}$. Prove that f is continuous. □

Exercise 16.11. Let X be a connected normal space containing more than one point. Prove that X has uncountably many points. □

Exercise 16.12. Let S be the Sorgenfrey line (see Example 12.7.5), so S is $\mathbb{R}$ with the topology given by the basis $\{[a, b) \mid a < b\}$ of all half-open intervals in $\mathbb{R}$. Prove that S is normal. □

Bibliography

[1] L. Steen & J. Seebach Jr., *Counterexamples in topology*, second edition, Springer, New York-Heidelberg, 1978. (Cited on pages 132 and 133.)

CHAPTER 17

Compactness and the Heine–Borel theorem

We now introduce the key concept of compactness, which generalizes the notion of compactness for subsets of $\mathbb{R}$ and $\mathbb{R}^n$ from real analysis.

17.1. Compactness

Let X be a space and let $K \subset X$. An *open cover* of K is a collection $\mathfrak{U}$ of open sets in X such that $K \subset \cup_{U \in \mathfrak{U}} U$. The open cover $\mathfrak{U}$ is *finite* if it consists of finitely many open sets. A *subcover* of an open cover $\mathfrak{U}$ is a subset $\mathfrak{U}' \subset \mathfrak{U}$ that is still a cover. The subspace K is *compact* if every open cover of K has a finite subcover. In particular, X itself is compact if every open cover of X has a finite subcover.

17.2. Compactness and closed sets

Compactness behaves best for Hausdorff spaces. In fact, in some treatments of point-set topology a space is said to be quasi-compact if each open cover has a finite subcover, and a compact space is a space that is Hausdorff and quasi-compact. For Hausdorff spaces, we have:

Lemma 17.2.1. *Let X be a Hausdorff space and let $K \subset X$ be compact. Then K is closed.*

Proof. We must prove that $X \setminus K$ is open. Consider $p \in X \setminus K$. Since X is Hausdorff, for each $k \in K$ there are disjoint open neighborhoods U_k and V_k of p and k. Since K is compact, we can find finitely many points $k_1, \ldots, k_n \in K$ such that $\{V_{k_1}, \ldots, V_{k_n}\}$ is an open cover of K. Letting $U = U_{k_1} \cap \cdots \cap U_{k_n}$, the set U is an open neighborhood of p that is disjoint from K, as desired. □

For all spaces, we have:

Lemma 17.2.2. *Let X be a space, let $K \subset X$ be compact, and let C be a closed subset of X with $C \subset K$. Then C is compact.*

Proof. Let $\mathfrak{U}$ be an open cover of $C \subset X$. The set $\{X \setminus C\} \cup \mathfrak{U}$ is an open cover of K. Since K is compact, it has a finite subcover. Removing $X \setminus C$ from this finite subcover if necessary, we obtain a finite subcover of $\mathfrak{U}$. □

As another indication of how strong an assumption being compact Hausdorff is, we have:

Lemma 17.2.3. *Let X be a compact Hausdorff space. Then X is normal.*

Proof. See Exercise 17.2. □

17.3. Compactness and functions

Continuous maps take compact sets to compact sets:

Lemma 17.3.1. *Let $f\colon X \to Y$ be a map of spaces and let $K \subset X$ be compact. Then $f(K)$ is compact.*

Proof. See Exercise 17.4 □

This has the following corollary:

Corollary 17.3.2. *Let $f\colon X \to Y$ be a map of spaces with X compact and Y Hausdorff. Then f is a closed map.*

PROOF. Let $C \subset X$ be closed. Since X is compact, C is compact. It follows that $f(C)$ is compact, so since Y is Hausdorff $f(C)$ is closed. □

Another important property of compact sets is that real-valued functions on them are bounded and attain maximum and minimum values:

LEMMA 17.3.3. *Let X be a compact space and let $f\colon X \to \mathbb{R}$ be a map. Then there exist real numbers $m \leq M$ such that:*

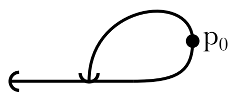
- *for all $p \in X$, we have $m \leq f(p) \leq M$; and*
- *there exists $p_0, q_0 \in X$ such that $m = f(p_0)$ and $M = f(q_0)$.*

PROOF. By Lemma 17.3.1, the image $K = f(X)$ is a compact subset of $\mathbb{R}$. The lemma now follows from the following standard fact about compact subsets of $\mathbb{R}$: there exist $m, M \in K$ such that $m \leq k \leq M$ for all $k \in K$ (see Exercise 17.5). □

17.4. Compactness and injective maps

For general spaces X and Y, an injective map $f\colon X \to Y$ need not be an embedding, i.e., a homeomorphism onto its image. Here is an example:

EXAMPLE 17.4.1. Consider the injective map $f\colon (0,1) \to \mathbb{R}^2$ whose image X is as follows:



This is not an embedding; indeed, for every $p \in (0,1)$ the space $(0,1) \setminus \{p\}$ is disconnected but for the indicated point $p_0 \in X$ we have $X \setminus \{p_0\}$ connected. □

However, if X is compact and Y is Hausdorff this pathology does not occur:

LEMMA 17.4.2. *Let X be a compact space, let Y be a Hausdorff space, and let $f\colon X \to Y$ be an injective map. Then f is a closed embedding.*

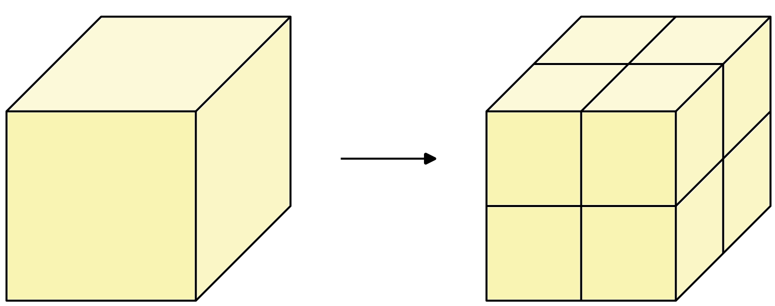
PROOF. Immediate from the fact that f is injective and closed (see Corollary 17.3.2). □

17.5. Heine–Borel Theorem

Let $(M, \mathfrak{d})$ be a metric space. A subset $K \subset M$ is *bounded* if there is some $R \geq 0$ such that $\mathfrak{d}(p,q) \leq R$ for all $p, q \in K$. The following theorem gives a large supply of compact spaces:

THEOREM 17.5.1 (Heine–Borel Theorem). *Let $K \subset \mathbb{R}^n$ be closed and bounded. Then K is compact.*

PROOF. For some $D \gg 0$, the set K is contained in the cube $[-D, D]^n$. By Lemma 17.2.2, it is enough to prove that $[-D, D]^n$ is compact. Since all cubes in $\mathbb{R}^n$, are homeomorphic, it is actually enough to prove that the unit cube $C_1 = I^n$ is compact. Let $\mathfrak{U}$ be an open cover of C_1. For the sake of contradiction, assume that it has no finite subcover. Divide C_1 into 2^n subcubes with side lengths $1/2$:



The cover $\mathfrak{U}$ is a cover of each of these subcubes. Since no finite subset of $\mathfrak{U}$ covers C_1, it must the case that among these 2^n subcubes there is a subcube C_2 such that no finite subset of $\mathfrak{U}$ covers C_2. This process can then be repeated: C_2 can be divided into 2^n subcubes with side length $1/2^2$, and there among these there must exist a subcube C_3 such that no finite subset of $\mathfrak{U}$ covers it. We then divide C_3 into 2^n subcubes with side lengths $1/2^3$, etc. This procedure gives a nested sequence

$$C_1 \supset C_2 \supset C_3 \supset \cdots$$

of cubes with the following properties:

- the cube C_n has side lengths $1/2^n$; and
- no finite subset of $\mathfrak{U}$ covers any of the C_n.

By the completeness of $\mathbb{R}$, the intersection $\cap_{n=1}^{\infty} C_n$ must consist of a single point p. Pick $U \in \mathfrak{U}$ such that $p \in U$. Since U is open, for some $\epsilon > 0$ the ϵ-ball around p must be contained in U. This implies that for $n \gg 0$ we have $C_n \subset U$, contradicting the fact that no finite subset of $\mathfrak{U}$ covers any C_n. □

REMARK 17.5.2. A metric space in which closed and bounded subsets are compact is called a *proper* metric space. This is a property of the metric, not just of the topology (see Exercise 17.6). □

17.6. Compactness and intersections of closed sets

The following is a useful rephrasing of the definition of compactness:

LEMMA 17.6.1. *Let X be a space. The X is compact if and only if the following holds for all sets $\mathfrak{C}$ of closed subsets of X:*

(∗) *If for all finite subsets $\mathfrak{C}' \subset \mathfrak{C}$ we have $\bigcap_{C \in \mathfrak{C}'} C \neq \emptyset$, then $\bigcap_{C \in \mathfrak{C}} C \neq \emptyset$.*

PROOF. The condition (∗) is equivalent to:

(∗′) If $\bigcap_{C \in \mathfrak{C}} C = \emptyset$, then there exists a finite subset $\mathfrak{C}' \subset \mathfrak{C}$ such that $\bigcap_{C \in \mathfrak{C}'} C = \emptyset$.

There is a bijection between sets of closed subsets of X and sets of open subsets of X taking a set $\mathfrak{C}$ of closed subsets to $\mathfrak{U}(\mathfrak{C}) = \{X \setminus C \mid C \in \mathfrak{C}\}$. A set $\mathfrak{C}$ of closed subsets of X has empty intersection exactly when $\mathfrak{U}(\mathfrak{C})$ covers X. It follows (∗′) is equivalent to saying that if $\mathfrak{U}(\mathfrak{C})$ is a cover of X, then $\mathfrak{U}(\mathfrak{C})$ has a finite subcover. □

This has the following immediate corollary:

COROLLARY 17.6.2. *Let X be a space and let $C_1 \supset C_2 \supset \cdots$ be a nested sequence of nonempty compact subspaces of X. Then $\cap_{n \geq 1} C_n \neq \emptyset$.*

17.7. Lebesgue number

If M is a metric space and $\mathfrak{U}$ is an open cover of M, then a *Lebesgue number* for $\mathfrak{U}$ is an $\epsilon > 0$ such that for all $p \in M$ there exists some $U \in \mathfrak{U}$ such that the ϵ-ball $B_\epsilon(p)$ is contained in U. The following basic result shows that these always exist if M is compact:

LEMMA 17.7.1 (Lebesgue number lemma). *Let M be a compact metric space and let $\mathfrak{U}$ be an open cover of M. Then $\mathfrak{U}$ has a Lebesgue number.*

PROOF. Since M is compact, we can write M as

$$M = B_{\epsilon_1}(p_1) \cup \cdots \cup B_{\epsilon_n}(p_n) \quad \text{for some } p_1, \ldots, p_n \in M \text{ and } \epsilon_1, \ldots, \epsilon_n > 0$$

such that for each $1 \leq i \leq n$ there is some $U \in \mathfrak{U}$ with $B_{2\epsilon_i}(p_i) \subset U$. Set $\epsilon = \min(\epsilon_1, \ldots, \epsilon_n)$, and consider $p \in M$. We have $p \in B_{\epsilon_i}(p_i)$ for some $1 \leq i \leq n$. By assumption, there is some $U \in \mathfrak{U}$ with $B_{2\epsilon_i}(p_i) \subset U$. The triangle inequality implies that $B_\epsilon(p) \subset B_{2\epsilon_i}(z_i)$ and thus $B_\epsilon(p) \subset U$. □

17.8. Compactness and limits

If X is a space and $\{x_n\}_{n \geq 1}$ is a sequence in X, then a *subsequence* of $\{x_n\}_{n \geq 1}$ is a sequence of the form $\{x_{n_i}\}_{i \geq 1}$ with $n_1 < n_2 < \cdots$ a strictly increasing sequence of natural numbers. A subspace $K \subset X$ is *sequentially compact* if every sequence in K has a subsequence that converges to a point of K. With appropriate countability assumptions, this is equivalent to compactness. We divide this into two results:

LEMMA 17.8.1. *Let X be a first countable space and let $K \subset X$ be compact. Then K is sequentially compact.*

Proof. See Exercise 17.3. □

Lemma 17.8.2. *Let X be a second countable space and let $K \subset X$ be sequentially compact. Then K is compact.*

Proof. See Exercise 17.3. □

Similarly, for metrizable spaces compactness and sequential compactness are the same:

Lemma 17.8.3. *Let M be a metrizable space and let $K \subset X$. Then K is compact if and only if K is sequentially compact.*

Proof. Fix a metric $\mathfrak{d}$ on M inducing its topology. Since M is first first countable, Lemma 17.8.1 implies that compact subsets of M are sequentially compact. For the converse, we can replace M by the subspace in question and prove that if M is sequentially compact, then M is compact. By Lemma 17.8.2 it is enough to prove that M is second countable, which by Lemma 15.5.2 is equivalent to proving that M is separable, i.e., that M has a countable dense subset.

Since M is sequentially compact, it cannot contain an infinite discrete subspace. In particular, for each $n \geq 1$ there does not exist an infinite subset $T \subset M$ with $\mathfrak{d}(t_1, t_2) \geq 1/n$ for all distinct $t_1, t_2 \in T$. For each $n \geq 1$, we can therefore find a finite set S_n such that for all $p \in M$ there exists some $s \in S_n$ with $\mathfrak{d}(p, s) < 1/n$. The set $\cup_{n \geq 1} S_n$ is then a countable dense subset of M. □

17.9. Exercises

Exercise 17.1. Let X be a space and let $C_1, \ldots, C_n \subset X$ be compact subspaces. Prove that $C_1 \cup \cdots \cup C_n$ is compact. □

Exercise 17.2. Let X be a compact Hausdorff space. Prove that X is normal. □

Exercise 17.3. Let X be a space and $K \subset X$ be a subspace. Prove:

(a) If X is first countable and K is compact, then K is sequentially compact.
(b) If X is second countable and K is sequentially compact, then K is compact. □

Exercise 17.4. Let $f \colon X \to Y$ be a map of spaces and let $K \subset X$ be compact. Prove that $f(K)$ is compact. □

Exercise 17.5. Let $K \subset \mathbb{R}$ be compact. Prove that there exist $m, M \in K$ such that $m \leq k \leq M$ for all $k \in K$. □

Exercise 17.6. Construct a metric on $\mathbb{R}^n$ inducing it usual topology such that the closed 1-ball around the origin is not compact. In the terminology of Remark 17.5.2, with this metric $\mathbb{R}^n$ is not a proper metric space. □

Exercise 17.7. Let $(M, \mathfrak{d}_M)$ and $(N, \mathfrak{d}_N)$ be metric spaces and let $f \colon M \to N$ be a map. We say that f is *uniformly continuous* if for all $\epsilon > 0$ there exists some $\delta > 0$ such that if $p, q \in M$ satisfy $\mathfrak{d}_M(p, q) < \delta$ then $\mathfrak{d}_N(f(p), f(q)) < \epsilon$. Use the Lebesgue number lemma to prove that if M is compact then all continuous maps $f \colon M \to N$ are uniformly continuous. □

Exercise 17.8. Let X and Y be spaces. Give $X \times Y$ the product topology (see Example 12.7.4). Let $U \subset X \times Y$ be open. Let $A \subset X$ and $K \subset Y$ be such that $A \times K \subset U$. Assume that K is compact. Prove that there exists an open neighborhood V of A such that $V \times K \subset U$. We remark that this is often called the "tube lemma". □

Exercise 17.9. Let $f \colon X \to Y$ be a map of spaces (not necessarily continuous) with Y compact Hausdorff. Give $X \times Y$ the product topology (see Example 12.7.4). Define the *graph* of f to be

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

Prove that f is continuous if and only if Γ_f is a closed subset of $X \times Y$. □

Exercise 17.10. Let X and Y be spaces with Y compact. Give $X \times Y$ the product topology (see Example 12.7.4). Let $\pi \colon X \times Y \to X$ be the projection onto the first factor. Prove that π is a closed map. □

EXERCISE 17.11. Let S be the Sorgenfrey line (see Example 12.7.5), so S is $\mathbb{R}$ with the topology given by the basis $\{[a, b) \mid a < b\}$ of all half-open intervals in $\mathbb{R}$. Prove the following:

(a) Let C be a compact subset of S. Prove that C is countable. Hint: first prove that C is compact in the classical topology on $\mathbb{R}$, then prove that there is no strictly increasing sequence $x_1 < x_2 < \cdots$ of points of C, and then prove that C is countable.

(b) Prove that every open cover of S has a countable subcover (spaces with this property are called *Lindelöf space*). □

CHAPTER 18

Local compactness and the Baire category theorem

We now turn to the local version of compactness.

18.1. Local compactness

Let X be a space. Recall that a general neighborhood of $p \in X$ is a set $Z \subset X$ with $p \in \text{Int}(Z)$. The space X is *locally compact* if the following holds for all $p \in X$:

- For all open neighborhoods U of p, there exists a compact neighborhood K of p with $K \subset U$.

For Hausdorff spaces, this is easier to understand:

LEMMA 18.1.1. *Let X be a Hausdorff space. Then X is locally compact if and only if for all $p \in X$, there exists a compact neighborhood K of p. In particular, if X is compact then X is locally compact.*

PROOF. See Exercise 18.2. □

REMARK 18.1.2. Local compactness is poorly behaved for non-Hausdorff spaces, and not all sources agree on the right definition for non-Hausdorff spaces. □

EXAMPLE 18.1.3. If X is either an open or a closed subspace of $\mathbb{R}^n$, then the Heine–Borel Theorem (Theorem 1.10.5) implies that X is locally compact. □

18.2. Regularity and normality

We now prove that locally compact Hausdorff spaces are regular:

LEMMA 18.2.1. *Let X be a locally compact Hausdorff space. Then X is regular.*

PROOF. Consider a point $p \in X$ and a closed set $C \subset X$ with $p \notin C$. We must find disjoint open neighborhoods of p and C. Since X is locally compact, there exists a compact neighborhood K of p with $K \subset X \setminus C$. Since X is Hausdorff the compact set K is closed. The desired open neighborhoods are then $\text{Int}(K)$ and $X \setminus K$. □

Since second countable regular spaces are normal (Lemma 16.10.1), this has the following corollary:

COROLLARY 18.2.2. *Let X be a locally compact Hausdorff space that is second countable. Then X is normal.*

18.3. One-point compactification

Let X be a space. A *compactification* of X is a compact space $\widehat{X}$ containing X as an open dense subspace.

EXAMPLE 18.3.1. The space $\mathbb{S}^n$ is a compactification of $\mathbb{R}^n$. Indeed, for all $p_0 \in \mathbb{S}^n$ we have $\mathbb{S}^n \setminus p_0 \cong \mathbb{R}^n$ (see Exercise 18.6). □

If X is locally compact Hausdorff, there is a natural way to compactify X that generalizes the compactification $\mathbb{S}^n$ of $\mathbb{R}^n$. As a set, let $X^* = X \sqcup \{\infty\}$ with ∞ a formal symbol that does not lie in X. Say that $U \subset X^*$ is open if either:

- U is an open subset of X; or

- $U = (X \setminus C) \cup \{\infty\}$, where $C \subset X$ is compact.

This is a topology (see Exercise 18.5), and the space X^* is called the *one-point compactification* of X. The following shows that it is indeed a compactification:

LEMMA 18.3.2. *Let X be locally compact Hausdorff and let X^* be the one-point compactification of X. Then X^* is compact Hausdorff, and X^* is a compactification of X.*

PROOF. See Exercise 18.5. $\square$

18.4. σ-compactness

A space X is *σ-compact* if it is the union of countably many compact subspaces. This condition will be important in the next chapter when we discuss paracompactness and partitions of unity. Here we prove:

LEMMA 18.4.1. *Let X be a Hausdorff space that is second countable and locally compact. Then X is σ-compact.*

PROOF. Let $\mathfrak{B}$ be a countable basis for the topology of X. Set $\mathfrak{U} = \big\{U \in \mathfrak{B} \mid \overline{U} \text{ is compact}\big\}$, so $\mathfrak{U}$ is a countable collection of open sets of X. It is enough to prove that $\mathfrak{U}$ covers X. Indeed, consider $p \in X$. We must find some $U \in \mathfrak{U}$ with $p \in U$. By Lemma 18.1.1, there is a compact neighborhood K of p. Since $p \in \mathrm{Int}(K)$, we can find $U \in \mathfrak{B}$ such that $p \in U$ and $U \subset K$. Since X is Hausdorff the compact set K is closed, so $\overline{U} \subset K$. Since $\overline{U}$ is a closed subset of the compact set K, it follows that $\overline{U}$ is compact and $U \in \mathfrak{U}$, as desired. $\square$

EXAMPLE 18.4.2. If X is either an open or a closed subspace of $\mathbb{R}^n$, then the Heine–Borel Theorem (Theorem 1.10.5) implies that X is σ-compact. $\square$

18.5. Baire category theorem

The following is a surprisingly powerful tool for proving existence theorems:

THEOREM 18.5.1 (Baire category theorem). *Let X be a locally compact Hausdorff space and let $\{U_n\}_{n\geq 1}$ be a collection of open dense subsets of X. Then $\cap_{n\geq 1} U_n$ is dense.*

PROOF. Let $V_0 \subset X$ be a nonempty open set. We must prove that V_0 intersects $\cap_{n\geq 1} U_n$. Since U_1 is open and dense, the set $V_0 \cap U_1$ is open and nonempty. Since X is locally compact and Hausdorff, we can find a nonempty open set V_1 with $\overline{V}_1$ compact such that $\overline{V}_1 \subset V_0 \cap U_1$. The same argument shows that there exists a nonempty open set V_2 with $\overline{V}_2$ compact such that $\overline{V}_2 \subset V_1 \cap U_2$. Repeating this over and over, we find nonempty open sets $\{V_n\}_{n\geq 1}$ with the following property for all $n \geq 1$:

- $\overline{V}_n$ is compact and $\overline{V}_{n+1} \subset V_n \cap U_{n+1}$.

Applying Corollary 17.6.2 to the nested sequence $\overline{V}_1 \supset \overline{V}_2 \supset \overline{V}_3 \supset \cdots$ of nonempty compact subspaces of X, we see that their intersection must be nonempty, i.e., there exists some p with $p \in \overline{V}_n$ for all $n \geq 1$. By construction, p lies in both V_0 and $\cap_{n\geq 1} U_n$, as desired. $\square$

REMARK 18.5.2. The word "category" in the Baire category theorem has nothing to do with category theory. Instead, it refers to the following archaic terminology: a space X is of the *first category* if it is the union of countably many nowhere dense[1] sets, and is of the *second category* otherwise. The conclusion of the Baire category theorem then is equivalent to saying that every nonempty open set in X is of the second category. $\square$

18.6. Complete metric spaces

A space X is a *Baire space* if all countable intersections of open dense subsets of X are dense. Theorem 18.5.1 says that locally compact Hausdorff spaces are Baire spaces. For another useful class of such spaces, consider a metric space $(M, \mathfrak{d})$. A *Cauchy sequence* in M is a sequence $\{p_n\}_{n\geq 1}$ such that for all $\epsilon > 0$ there exists some $N \geq 1$ such that $\mathfrak{d}(p_n, p_m) < \epsilon$ for all $n, m \geq N$. The metric space M is *complete* if all Cauchy sequences in M have limits. For instance, $\mathbb{R}^n$ is complete (Exercise 18.8). We have:

[1] A subset A of a topological space is *nowhere dense* if $\overline{A}$ contain no nonempty open sets, i.e., if $\mathrm{Int}(\overline{A}) = \emptyset$.

THEOREM 18.6.1 (Baire category theorem′). *Let M be a complete metric space. Then M is a Baire space.*

PROOF. This is similar to the proof of Theorem 18.5.1, so we leave it as Exercise 18.9. □

18.7. Application: nowhere differentiable functions

To illustrate how the Baire category theorem can be used, we prove the following classic result:

THEOREM 18.7.1. *Let $\mathcal{C}(I,\mathbb{R})$ be the set of continuous functions $f\colon I \to \mathbb{R}$. Let $\mathfrak{d}(f,g) = \max\{|f(x)-g(x)| \mid x \in I\}$ be the standard metric on $\mathcal{C}(I,\mathbb{R})$. Then the set of nowhere-differentiable functions on is dense in $\mathcal{C}(I,R)$.*

PROOF. For each $n \geq 1$, let U_n be the set of all continuous functions $f\colon I \to \mathbb{R}$ satisfying:

(♠) There exists $0 < \delta < 1/n$ and $\lambda > 0$ such that for all $x \in I$, there exists some $y \in I$ with $\delta < |x-y| < 1/n$ and $\left|\frac{f(x)-f(y)}{x-y}\right| > n + \lambda$.

In the three steps below, we will prove that U_n is open (Step 1), we will construct a family of function in U_n (Step 2), and we will show that U_n is dense (Step 3). Since $\mathcal{C}(I,\mathbb{R})$ is a complete metric space, Theorem 18.6.1 will then apply and show that $\Lambda = \cap_{n\geq 1} U_n$ is dense in $\mathcal{C}(I,\mathbb{R})$. Each $f \in \Lambda$ is nowhere differentiable; indeed, for $x \in I$ the condition (♠) forces $\lim_{y\mapsto x} \frac{f(x)-f(y)}{x-y}$ to not exist.

STEP 1. *For all $n \geq 1$, the set U_n is open in $\mathcal{C}(I,\mathbb{R})$.*

Consider $f \in U_n$. Let $0 < \delta < 1/n$ and $\lambda > 0$ be the constants for f from (♠). Let $g \in \mathcal{C}(I,\mathbb{R})$ be such that $\mathfrak{d}(f,g) < \lambda\delta/4$. We claim that $g \in U_n$. Indeed, consider $x \in I$. Choose $y \in I$ such that $\delta < |x-y| < 1/n$ and $\left|\frac{f(x)-f(y)}{x-y}\right| > n+\lambda$. We then have

$$\begin{aligned}\left|\frac{g(x)-g(y)}{x-y}\right| \geq& \left|\frac{f(x)-f(y)}{x-y}\right| - \left|\frac{g(x)-f(x)}{x-y}\right| - \left|\frac{g(y)-f(y)}{x-y}\right| \\ >&(n+\lambda) - 2\frac{\lambda\delta/4}{\delta} = n + \lambda/2.\end{aligned}$$

It follows that g satisfies (♠) with the constants δ and $\lambda/2$, so $g \in U_n$.

STEP 2. *For some $n \geq 1$, let $g\colon I \to \mathbb{R}$ be a piecewise-linear continuous function such that $|g'(x)| > n$ for all $x \in I$ where g is differentiable. Then $g \in U_n$.*

Let $0 = a_0 < a_1 < \cdots < a_m = 1$ be a partition of I such that $g|_{[a_i,a_{i+1}]}$ is linear for all $0 \leq i < m$. For each $0 \leq i < m$, let $c_i, d_i \in \mathbb{R}$ be the constants such that $g(x) = c_i x + d_i$ for all $x \in [a_i, a_{i+1}]$. By assumption, $|c_i| > n$ for all $0 \leq i < m$. Pick $\lambda > 0$ such that $|c_i| > n + \lambda$ for all $0 \leq i < m$. Also, pick $0 < \delta < 1/n$ such that $\delta < (a_{i+1} - a_i)/2$ for all $0 \leq i < m$. Consider some $x \in I$. We have $x \in [a_{i_0}, a_{i_0+1}]$ for some $0 \leq i_0 < m$. Since $0 < \delta < (a_{i_0+1} - a_{i_0})/2$, we can choose some $y \in [a_{i_0}, a_{i_0+1}]$ such that $\delta < |x-y| < 1/n$. It follows that

$$\left|\frac{g(x)-g(y)}{x-y}\right| = \left|\frac{(c_i x + d_i) - (c_i y + d_i)}{x-y}\right| = |c_i| > n + \lambda,$$

proving that g satisfies (♠) and thus $g \in U_n$.

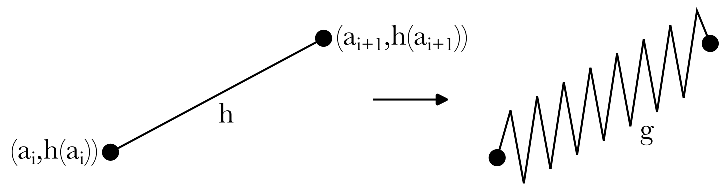
STEP 3. *For all $n \geq 1$, the set U_n is dense in $\mathcal{C}(I,\mathbb{R})$.*

Consider $f \in \mathcal{C}(I,\mathbb{R})$ and $\epsilon > 0$. We must find some $g \in U_n$ such that $\mathfrak{d}(f,g) < \epsilon$. Since f is uniformly continuous on I, we can choose a partition $0 = a_0 < a_1 < \cdots < a_m = 1$ of I such that for all $0 \leq i < m$ and $x \in [a_i, a_{i+1}]$ we have $|f(x) - f(a_i)| < \epsilon/4$. Let $h\colon I \to \mathbb{R}$ be the piecewise-linear continuous function that is linear on each $[a_i, a_{i+1}]$ and satisfies $h(a_i) = f(a_i)$ for $0 \leq i \leq m$. For

$x \in [a_i, a_{i+1}]$, we therefore have

$$\begin{aligned}|h(x) - f(x)| &= \left| \frac{f(a_{i+1}) - f(a_i)}{a_{i+1} - a_i}(x - a_i) + f(a_i) - f(x) \right| \\ &\leq \left| \frac{f(a_{i+1}) - f(a_i)}{a_{i+1} - a_i} \right| |x - a_i| + |f(a_i) - f(x)| \\ &\leq |f(a_{i+1}) - f(a_i)| + |f(a_i) - f(x)| < \epsilon/4 + \epsilon/4 = \epsilon/2.\end{aligned}$$

It follows that $\mathfrak{d}(f, h) < \epsilon/2$. As in the following figure, we can find a piecewise-linear continuous function $g\colon I \to \mathbb{R}$ with $\mathfrak{d}(g, h) < \epsilon/2$ and $|g'(x)| > n$ for all $x \in I$ where g is differentiable by changing h on each interval $[a_i, a_{i+1}]$ to a function whose graph is a rapidly osculating sawtooth:



We have $\mathfrak{d}(f, g) \leq \mathfrak{d}(f, h) + \mathfrak{d}(h, g) < \epsilon$, and by Step 2 we have $g \in U_n$. □

18.8. Exercises

EXERCISE 18.1. Find an example of a locally compact Hausdorff space X that has a subspace $Y \subset X$ that is not locally compact. □

EXERCISE 18.2. Let X be a Hausdorff space. Prove that X is locally compact if and only if for all $p \in X$, there exists a compact neighborhood K of p. □

EXERCISE 18.3. Let X be a locally compact space. Let $K \subset X$ be compact and $U \subset X$ be open with $K \subset U$. Prove that there is a compact neighborhood L of K with $L \subset U$. □

EXERCISE 18.4. Let X be a locally compact Hausdorff space, let $K \subset X$ be a compact subspace, and let $f\colon K \to \mathbb{R}$ be a continuous function. Prove that f can be extended to a continuous function $F\colon X \to \mathbb{R}$ with $\mathrm{supp}(F)$ compact. We remark that X might not be normal, so you can't just apply the Tietze extension theorem. □

EXERCISE 18.5. Let X be a locally compact Hausdorff space. Recall that the one-point compactification of X is the set $X^* = X \sqcup \{\infty\}$ with the following topology: a set $U \subset X^*$ is open if either:

- U is an open subset of X; or
- $U = (X \setminus C) \cup \{\infty\}$, where $C \subset X$ is compact.

Prove the following:

(a) This is a topology.
(b) The space X^* is compact Hausdorff.
(c) The space X^* is a compactification of X. □

EXERCISE 18.6. Prove the following:

(a) For all $p_0 \in \mathbb{S}^n$ we have $\mathbb{S}^n \setminus p_0 \cong \mathbb{R}^n$.
(b) The space $\mathbb{S}^n$ is a compactification of $\mathbb{R}^n$.
(c) The space $\mathbb{S}^n$ is homeomorphic to the one-point compactification of X. □

EXERCISE 18.7. Let X be the one-point compactification of $\mathbb{Z}$. Prove that X is homeomorphic to $\{0\} \cup \{1/n \mid n \in \mathbb{Z} \text{ nonzero}\} \subset \mathbb{R}$. □

EXERCISE 18.8. Prove that $\mathbb{R}^n$ with its standard metric is complete, i.e., that all Cauchy sequences in $\mathbb{R}^n$ have limits. □

EXERCISE 18.9. Let M be a complete metric space. Prove that M is a Baire space, i.e., that the following holds. Let $\{U_n\}_{n \geq 1}$ be a collection of open dense subsets of X. Then $\cap_{n \geq 1} U_n$ is dense. □

EXERCISE 18.10. Use the Baire category theorem to prove that if $S \subset \mathbb{R}^n$ is a countable collection of points, then $\mathbb{R}^n \setminus S$ is path-connected. □

EXERCISE 18.11. Recall that a subset A of a space is nowhere dense if $\overline{A}$ contain no nonempty open sets, i.e., if $\mathrm{Int}(\overline{A}) = \emptyset$. Prove that $\mathbb{R} \setminus \mathbb{Q}$ cannot be written as a countable union of nowhere dense sets. □

EXERCISE 18.12. Let S be the Sorgenfrey line (see Example 12.7.5), so S is $\mathbb{R}$ with the topology given by the basis $\{[a, b) \mid a < b\}$ of all half-open intervals in $\mathbb{R}$. Prove the following:

(a) S is not locally compact.
(b) S is not σ-compact.
(c) S is a Baire space.

Hint for (a) and (b): use Exercise 17.11. □

CHAPTER 19

Proper maps

This brief chapter covers proper maps, which play an important role in both group actions and the study of embeddings.

19.1. Definition and examples

A map $f\colon X \to Y$ is *proper* if it is closed and $f^{-1}(y)$ is compact for all $y \in Y$. The topological meaning of this is a little subtle, and will become more clear as we give examples and alternate characterizations of these maps.

EXAMPLE 19.1.1. A closed embedding $f\colon X \to Y$ is trivially proper. $\square$

EXAMPLE 19.1.2. If X is compact and Y is Hausdorff, then any map $f\colon X \to Y$ is proper (see Exercise 19.1). $\square$

19.2. Preimages of compact sets

The following is a key property of proper maps:

LEMMA 19.2.1. *Let $f\colon X \to Y$ be a proper map and let $K \subset Y$ be compact. Then $f^{-1}(K)$ is compact.*

PROOF. Set $L = f^{-1}(K)$ and let $\mathfrak{U}$ be an open cover of L. Consider some $k \in K$. Set $L_k = f^{-1}(k)$. By assumption, L_k is compact. We can therefore find a finite subset $\mathfrak{U}_k \subset \mathfrak{U}$ that covers L_k. Let $U_k = \cup_{U \in \mathfrak{U}_k} U$. Since f is a closed map, the set $f(X \setminus U_k)$ is closed. We can therefore find an open neighborhood V_k of k with $V_k \subset Y \setminus f(X \setminus U_k)$, i.e., with $f^{-1}(V_k) \subset U_k$. Since K is compact, we can find $k_1, \ldots, k_n \in K$ such that $K \subset V_{k_1} \cup \cdots \cup V_{k_n}$. We therefore have $L \subset f^{-1}(V_{k_1}) \cup \cdots \cup f^{-1}(V_{k_n})$. The finite subset $\mathfrak{U}_{k_1} \cup \cdots \mathfrak{U}_{k_n}$ of $\mathfrak{U}$ therefore covers $L = f^{-1}(K)$. $\square$

The conclusion of Lemma 19.2.1 is often taken as the definition of a proper map. The following shows that this alternate definition is equivalent to the one we gave for maps between spaces that are reasonable. For instance, it applies if the spaces in question are metrizable.

LEMMA 19.2.2. *Let $f\colon X \to Y$ be a continuous map satisfying the following:*

- *For all compact subsets $K \subset Y$, the preimage $f^{-1}(K) \subset X$ is compact.*

Assume that X and Y are Hausdorff and both are either first countable or locally compact. Then f is proper.

PROOF. We must prove that f is a closed map. We will prove this when X and Y are first countable; see Exercise 19.2 for the case where they are locally compact. Assume that X and Y are first countable and that $C \subset X$ is closed. We must prove that $f(C)$ is closed, i.e., that $\overline{f(C)} = f(C)$. Since Y is first countable, we can calculate the closure $\overline{f(C)}$ using limits; see Lemma 15.3.1. Letting $y \in \overline{f(C)}$, we can therefore find a sequence $\{c_n\}_{n \geq 1}$ of points of C such that $\lim_{n \mapsto \infty} f(c_n) = y$. Set

$$K = \{y\} \cup \{f(c_n) \mid n \geq 1\}.$$

Since $\lim_{n \mapsto \infty} f(c_n) = y$, the set K is compact. By assumption, $f^{-1}(K)$ is compact. Since X is first countable, it follows that $f^{-1}(K)$ is sequentially compact; see Lemma 17.8.1. After possibly replacing $\{c_n\}_{n \geq 1}$ by a subsequence, we can therefore assume that there is some $x \in f^{-1}(K)$ with $\lim_{n \mapsto \infty} c_n = x$. Since Y is Hausdorff, we have $f(x) = f(\lim_{n \mapsto \infty} c_n) = \lim_{n \mapsto \infty} f(c_n) = y$. $\square$

19.3. Properness and strongly divergent sequences

Let X be a space and let $\{x_n\}_{n\geq 1}$ be a sequence of points of X. We say that $\{x_n\}_{n\geq 1}$ is *strongly divergent* if for all compact $K \subset X$ there exists some $N \geq 1$ such that $x_n \notin K$ for all $n \geq N$. This implies that $\{x_n\}_{n\geq 1}$ has no convergent subsequence (see Exercise 19.3).

EXAMPLE 19.3.1. Let $(M, \mathfrak{d})$ be a metric space. Assume that M is a proper metric space, which we recall means that closed and bounded subsets of M are compact. A sequence $\{x_n\}_{n\geq 1}$ in M is strongly divergent if and only if for all $p \in M$ we have $\lim_{n\mapsto\infty} \mathfrak{d}(x_n, p) = \infty$ (see Exercise 19.4). ☐

Proper maps take strongly divergent sequences to strongly divergent sequences:

LEMMA 19.3.2. *Let $f\colon X \to Y$ be a proper map and let $\{x_n\}_{n\geq 1}$ be a strongly divergent sequence in X. Then $\{f(x_n)\}_{n\geq 1}$ is a strongly divergent sequence in Y.*

PROOF. See Exercise 19.5. ☐

For metrizable spaces, the following lemma shows that the conclusion of Lemma 19.3.2 characterizes proper maps. This gives an important source of intuition about the meaning of properness.

LEMMA 19.3.3. *Let $f\colon X \to Y$ be a map between metrizable spaces. Assume that the following holds:*

- *For all strongly divergent sequences $\{x_n\}_{n\geq 1}$ in X, the sequence $\{f(x_n)\}_{n\geq 1}$ in Y is strongly divergent.*

Then f is proper.

PROOF. Let $K \subset Y$ be compact. By Lemma 19.2.2, it is enough to show that $L = f^{-1}(K)$ is compact. Since X is metrizable, to prove that L is compact it suffices to prove that L is sequentially compact. Let $\{x_n\}_{n\geq 1}$ be a sequence of points of L. Since all points of $\{f(x_n)\}_{n\geq 1}$ lie in the compact set K, it follows that $\{f(x_n)\}_{n\geq 1}$ is not strongly divergent. By assumption, it follows that $\{x_n\}_{n\geq 1}$ is not strongly divergent.

Passing to a subsequence, we can therefore assume that there is some compact $C \subset X$ such that $x_n \in C$ for all $n \geq 1$. The sequence $\{x_n\}_{n\geq 1}$ thus lies in $C \cap L$. Since $C \subset X$ and $K \subset Y$ are compact it follows that they are closed, so $L = f^{-1}(K)$ and $C \cap L$ are closed. The closed subset $C \cap L$ of the compact set C is compact and hence sequentially compact, so we conclude that $\{x_n\}_{n\geq 1}$ has a subsequence that converges to a point of $C \cap L \subset L$, as desired. ☐

19.4. Application: fundamental theorem of algebra

To illustrate the geometric meaning of proper maps, we use them to prove the fundamental theorem of algebra. One way of stating this theorem is that every degree-n complex polynomial has n roots, at least if you count these roots with multiplicity. The nontrivial part of this is that every nonconstant polynomial has a root, so this is what we will prove:

THEOREM 19.4.1 (Fundamental theorem of algebra). *Let $f(z) \in \mathbb{C}[z]$ be a nonconstant polynomial. Then there exists some $z_0 \in \mathbb{C}$ such that $f(z_0) = 0$.*

PROOF. We will prove more generally that regarded as a map $f\colon \mathbb{C} \to \mathbb{C}$, the polynomial $f(z)$ is surjective. In fact, it is slightly easier to prove a variant of this. Let $f'(z)$ be the derivative of $f(z)$ and $C = \{z \in \mathbb{C} \mid f'(z) = 0\}$. Define $B = f(C)$ and $A = f^{-1}(X)$. Both A and B are finite sets, and f restricts to a map $F\colon \mathbb{C} \setminus A \to \mathbb{C} \setminus B$. To prove that f is surjective, it is enough to prove that F is surjective. Since B is a finite set, its complement $\mathbb{C} \setminus B$ is connected. This reduces us to showing that $F(\mathbb{C} \setminus B)$ is both open and closed. We will prove that F is both an open map and a closed map.[1]

Since $\lim_{z\mapsto\infty} f(z) = \infty$, the map f takes stronger divergent sequences to strongly divergent sequences. By Lemma 19.3.2, the map f is proper and thus so is F (see Exercise 19.6). In particular, F is a closed map. For $z \in \mathbb{C} \setminus A$, since $f'(z) \neq 0$ the inverse function theorem implies that f is a *local homeomorphism* at z, i.e., there is an open neighborhood U of z such that $f|_U\colon U \to \mathbb{C}$ is an open embedding. This implies that F is also a local homeomorphism at z. Since F is a local homeomorphism at all points in its domain, it follows that F is an open map (see Exercise 19.7). ☐

[1] In fact, f itself is both an open and closed map, but it is a little easier to prove this for F.

19.5. Exercises

EXERCISE 19.1. Let $f\colon X \to Y$ be a map from a compact space X to a Hausdorff space Y. Prove that f is proper. $\square$

EXERCISE 19.2. Let $f\colon X \to Y$ be a continuous map satisfying the following:

- For all compact subsets $K \subset Y$, the preimage $f^{-1}(K) \subset X$ is compact.

Assume that X and Y are Hausdorff and locally compact. Prove that f is a closed map, and hence is proper. $\square$

EXERCISE 19.3. Let X be a space and let $\{x_n\}_{n\geq 1}$ be a strongly divergent sequence of points of X. Prove that $\{x_n\}_{n\geq 1}$ has no convergent subsequence. $\square$

EXERCISE 19.4. Let $(M, \mathfrak{d})$ be a proper metric space. Prove that a sequence $\{x_n\}_{n\geq 1}$ in M is strongly divergent if and only if for all $p \in M$ we have $\lim_{n\mapsto\infty} \mathfrak{d}(x_n, p) = \infty$. $\square$

EXERCISE 19.5. Let $f\colon X \to Y$ be a proper map and let $\{x_n\}_{n\geq 1}$ be a strongly divergent sequence in X. Prove that $\{f(x_n)\}_{n\geq 1}$ is a strongly divergent sequence in Y. $\square$

EXERCISE 19.6. Let $f\colon X \to Y$ be a proper map, let $Y' \subset Y$ be any subspace, and let $X' = f^{-1}(Y')$. Prove that $f|_{X'}\colon X' \to Y'$ is a proper map. $\square$

EXERCISE 19.7. Let $f\colon X \to Y$ be a map. For each $x \in X$, assume that there is an open neighborhood U of x such that $f|_U\colon U \to Y$ is an open map. Prove that f is an open map. $\square$

CHAPTER 20

Paracompactness and partitions of unity

We now turn to paracompactness, which is a condition that ensure the existence of what are called partitions of unity. These play a basic role in algebraic topology, especially in the theory of manifolds.

20.1. Paracompactness

Let X be a space and let $\mathfrak{Z}$ be a collection of subsets X. We say that $\mathfrak{Z}$ is *locally finite* if for all $p \in X$, there exists a neighborhood U such that only finitely many $Z \in \mathfrak{Z}$ satisfy $U \cap Z \neq \emptyset$. This implies in particular that for all $p \in X$ there are only finitely many $Z \in \mathfrak{Z}$ with $p \in Z$, but is stronger (see Exercise 20.1). One nice property of locally finite collections of subset is:

LEMMA 20.1.1. *Let X be a space and let $\mathfrak{Z}$ be a locally finite collection of subsets of X. Then*

$$\overline{\bigcup_{Z \in \mathfrak{Z}} Z} = \bigcup_{Z \in \mathfrak{Z}} \overline{Z}.$$

PROOF. See Exercise 20.2. In that exercise, you will also show that this is false without the local finiteness assumption. □

Now let $\mathfrak{U}$ be an open cover of X. A *refinement* of $\mathfrak{U}$ is an open cover $\mathfrak{V}$ such that for all $V \in \mathfrak{V}$, there exists some $U \in \mathfrak{U}$ with $V \subset U$. A space X is *paracompact* if it is Hausdorff and every open cover of X admits a locally finite refinement. We will prove that this has strong consequences for the topology of X. In particular, X must be normal (see Lemma 20.4.1).

20.2. Locally compact Hausdorff spaces that are σ-compact are paracompact

The easiest examples of paracompact spaces are compact Hausdorff spaces, where every open cover admits a finite cover (not just a locally finite one). Our next goal is to prove the following generalization of this:

THEOREM 20.2.1. *Let X be a locally compact Hausdorff space that is σ-compact. Then X is paracompact.*

Before we prove this, we note that in light of Lemma 18.4.1 it implies:

COROLLARY 20.2.2. *Let X be a locally compact Hausdorff space that is second countable. Then X is paracompact. In particular, both open and closed subspaces of $\mathbb{R}^n$ are paracompact.*

PROOF OF THEOREM 20.2.1. We start by proving:

CLAIM. *There exists a countable open cover $\{W_1, W_2, \ldots\}$ of X such that for all $n \geq 1$ the set $\overline{W}_n$ is compact and satisfies $\overline{W}_n \subset W_{n+1}$.*

PROOF OF CLAIM. Since X is σ-compact, we can write $X = \cup_{n \geq 1} K_n$ with K_n compact. We will inductively construct open sets W_n of X such that $W_0 = \emptyset$ and for all $n \geq 0$ we have:

- $\overline{W}_n$ is compact; and
- W_{n+1} contains $\overline{W}_n \cup K_{n+1}$.

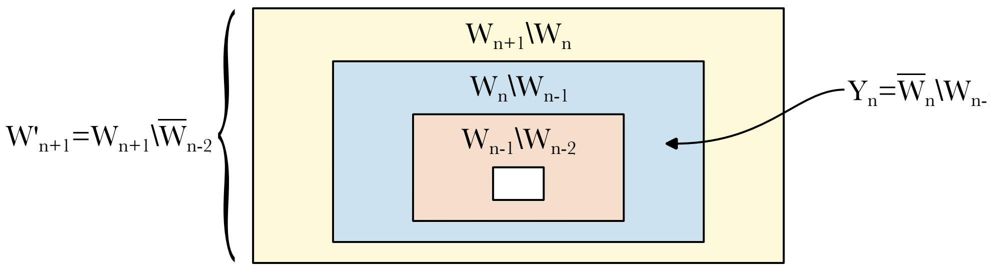
Since $X = \cup_{n \geq 1} K_n$, this will be a open open cover of X with the properties indicated in the claim. Start by setting $W_0 = \emptyset$, and assume we have constructed $W_0, \ldots, W_n$. For $p \in \overline{W}_n \cup K_{n+1}$, local compactness gives an open neighborhood $W_{n+1}(p)$ of p with $\overline{W}_{n+1}(p)$ compact. Since $\overline{W}_n \cup K_{n+1}$ is compact, we can find $p_1, \ldots, p_m \in \overline{W}_n \cup K_{n+1}$ such that $\{W_{n+1}(p_1), \ldots, W_{n+1}(p_m)\}$ covers

$\overline{W}_n \cup K_{n+1}$. We can then define $W_{n+1} = W_{n+1}(p_1) \cup \cdots \cup W_{n+1}(p_m)$. The set $\overline{W}_{n+1}$ is compact since $\overline{W}_{n+1} = \overline{W}_{n+1}(p_1) \cup \cdots \cup \overline{W}_{n+1}(p_m)$ (see Lemma 20.1.1). □

We now prove that X is paracompact. Let $\mathfrak{U}$ be an open cover of X. Let $\{W_n\}_{n \geq 1}$ be as in the claim. Set $W_m = \emptyset$ for $m \leq 0$. For $n \in \mathbb{Z}$, define $Y_n = \overline{W}_n \setminus W_{n-1}$ and $W'_{n+1} = W_{n+1} \setminus \overline{W}_{n-2}$. These satisfy:

(i) Y_n is a compact subset of the open set W'_{n+1}; and
(ii) $X = \cup_{n=1}^{\infty} Y_n$; and
(iii) $W'_{n_1} \cap W'_{n_2} = \emptyset$ whenever $|n_1 - n_2| \geq 3$.

See here:



For each $n \geq 1$, the set $\left\{U \cap W'_{n+1} \mid U \in \mathfrak{U}\right\}$ is an open cover of compact set Y_n, so there is a finite subset $\mathfrak{U}(n) \subset \mathfrak{U}$ such that $\left\{U \cap W'_{n+1} \mid U \in \mathfrak{U}(n)\right\}$ covers Y_n. Let

$$\mathfrak{V} = \left\{U \cap W'_{n+1} \mid n \geq 1 \text{ and } U \in \mathfrak{U}(n)\right\}.$$

The set $\mathfrak{V}$ is an open cover of each Y_n, so by (ii) it follows that $\mathfrak{V}$ is an open cover of X. By construction, $\mathfrak{V}$ refines $\mathfrak{U}$. Using (iii) together with the fact that only finitely many $V \in \mathfrak{V}$ are contained in each W'_n, the open cover $\mathfrak{V}$ is locally finite. The theorem follows. □

20.3. Metrizable spaces are paracompact

We next prove that metrizable spaces are paracompact. This was originally proved by Stone [**2**], but the proof we give is a much easier argument of Rudin [**1**]. The proof uses some set-theoretic technology. A *well-ordered set* is a set I equipped with a total ordering $\leq$ such that every nonempty subset $S \subset I$ has a minimal element. We call $\leq$ a *well-ordering*. A remarkable consequence of the axiom of choice is that every set can be equipped with a well-ordering.

Example 20.3.1. All total orderings on finite sets are well-orderings, and the standard ordering on $\mathbb{N}$ is a well-ordering. □

We will develop the theory of well-ordered sets more in Chapter 21 when we introduce transfinite induction, which is needed to prove that arbitrary products of compact spaces are compact (Tychonoff's Theorem). Here we use them to prove the following:

Lemma 20.3.2. *Let M be a metrizable space. Then M is paracompact.*

Proof. Let $\mathfrak{d}$ be a distance function on M inducing its topology. Consider an open cover $\mathfrak{U}$ of M. Choose a well-ordering on $\mathfrak{U}$. For each $x \in M$, let U_x be the minimal element of the set $\{U \in \mathfrak{U} \mid x \in U\}$. For each $U \in \mathfrak{U}$ and $n \geq 1$, we inductively construct a subset $U[n] \subset U$ as follows. Assume that $n \geq 1$ and that for all $1 \leq m < n$ we have constructed $V[m] \subset V$ for all $V \in \mathfrak{U}$ (a vacuous assumption if $n = 1$). Consider some $U \in \mathfrak{U}$. We define $U[n]$ to be the union of all open balls $B_x(1/2^n)$ where $x \in M$ satisfies the following three conditions:

(i) $U_x = U$; and
(ii) $x \notin V[m]$ for any $V \in \mathfrak{U}$ and $1 \leq m < n$; and
(iii) $B_x(3/2^n) \subset U$.

By (iii) we have $U[n] \subset U$. Define

$$\mathfrak{V} = \{U[n] \mid U \in \mathfrak{U} \text{ and } n \geq 1\}.$$

This clearly refines $\mathfrak{U}$. We will prove that it covers M and is locally finite:

Claim 1. *The set $\mathfrak{V}$ covers M.*

Consider some $x \in M$. Set $U = U_x$. Since U is an open neighborhood of x, there is some $n \geq 1$ such that $B_x(3/2^n) \subset U$. We then either have $x \in V[m]$ for some $V \in \mathfrak{U}$ and $1 \leq m < n$ or we have $B_x(1/2^n) \subset U[n]$ and thus $x \in U[n]$. In either case, x lies in some element of $\mathfrak{V}$.

Claim 2. *The set $\mathfrak{V}$ is locally finite.*

Consider some $x \in M$. Let U be the smallest element of $\mathfrak{U}$ such that there exists some $n \geq 1$ with $x \in U[n]$. Letting $n \geq 1$ be the minimal element such that $x \in U[n]$, choose $d \geq 1$ such that $B_x(1/2^d) \subset U[n]$. We will prove that the open neighborhood $B_x(1/2^{n+d})$ of x only intersects finitely many elements of $\mathfrak{V}$. In fact, we will prove the following:

(a) For all $V \in \mathfrak{U}$ and $m \geq n + d$, the element $V[m]$ of $\mathfrak{V}$ does not intersect $B_x(1/2^{n+d})$.
(b) For all $1 \leq \ell < n + d$, at most one element of $\mathfrak{V}$ of the form $W[\ell]$ can intersect $B_x(1/2^{n+d})$.

Together these will imply that at most $n + d - 1$ elements of $\mathfrak{V}$ can intersect $B_x(1/2^{n+d})$.

We start with (a). Consider some $V \in \mathfrak{U}$ and $m \geq n + d$. We wish to prove that $V[m]$ is disjoint from $B_x(1/2^{n+d})$. The open set $V[m]$ is the union of open balls $B_y(1/2^m)$ where $y \in M$ satisfies:

(i) $U_y = V$; and
(ii) $y \notin W[k]$ for any $W \in \mathfrak{U}$ and $1 \leq k < m$.

We do not include condition (iii) since it is not needed for the proof. Letting $y \in M$ satisfy these conditions, we must show that $B_y(1/2^m)$ is disjoint from $B_x(1/2^{n+d})$, i.e., that $\mathfrak{d}(x, y) \geq 1/2^{n+d} + 1/2^m$. Since $m \geq n + d > n$, condition (ii) implies that $y \notin U[n]$. Since $B_x(1/2^d) \subset U[n]$, this implies that

$$\mathfrak{d}(x, y) \geq 1/2^d = 1/2^{d+1} + 1/2^{d+1} \geq 1/2^{n+d} + 1/2^m,$$

as desired.

We now prove (b). Fix some $1 \leq \ell < n + d$, and let $W, W' \in \mathfrak{U}$ be such that $W[\ell]$ and $W'[\ell]$ both intersect $B_x(1/2^{n+d})$. We must prove that $W = W'$. The open set $W[\ell]$ is the union of open balls $B_z(1/2^\ell)$ where $z \in M$ satisfies:

(i) $U_z = W$; and
(iii) $B_z(3/2^\ell) \subset W$.

We do not include condition (ii) since it is not needed for the proof. We can therefore find some $z \in M$ satisfying (i) and (iii) such that $B_z(1/2^\ell)$ intersects $B_x(1/2^{n+d})$. Choose a point p in this intersection. Similarly, we can find $z' \in M$ satisfying the analogues of (i) and (iii) for W' such that $B_{z'}(1/2^\ell)$ intersects $B_x(1/2^{n+d})$. Choose a point p' in this intersection. We have

$$\mathfrak{d}(z, z') \leq \mathfrak{d}(z, p) + \mathfrak{d}(p, p') + \mathfrak{d}(p, z') < 1/2^\ell + 2/2^{n+d} + 1/2^\ell \leq 1/2^\ell + 1/2^\ell + 1/2^\ell = 3/2^\ell.$$

Applying (iii), we deduce that

$$z \in B_{z'}(3/2^\ell) \subset W' \quad \text{and} \quad z' \in B_z(3/2^\ell) \subset W.$$

Conditition (i) says that W is the minimal element of $\mathfrak{U}$ with $z \in W$, so $W \leq W'$. Similarly, W' is the minimal element of $\mathfrak{U}$ with $z' \in W'$, so $W' \leq W$. We conclude that $W = W'$, as desired. □

20.4. Normality

Our next goal is to prove that paracompact spaces are normal:

Lemma 20.4.1. *Let X be a paracompact space. Then X is normal.*

Proof. Recall that paracompact spaces are assumed to be Hausdorff. We start by proving tha X is regular (see §16.10), i.e., that the following holds:

Claim. *For $p \in X$ and $C \subset X$ closed with $p \notin C$, there exist disjoint open neighborhoods of p and C.*

Proof of claim. For each $q \in C$, since X is Hausdorff there exist open neighborhoods U_{qp} of q and U'_{qp} of p such that $U_{qp} \cap U'_{qp} = \emptyset$. Since X is paracompact, the open cover $\{X \setminus C\} \cup \{U_{qp} \mid q \in C\}$ admits a locally finite refinement. Let $\mathfrak{V}$ be the open sets in this locally finite refinement that are not contained in $X \setminus C$. For each $V \in \mathfrak{V}$, there is some $q \in C$ such that $V \subset U_{qp}$. Since U'_{qp} is an open neighborhood of p that is disjoint from U_{qp}, we deduce that $p \notin \overline{V}$ for all $V \in \mathfrak{V}$. Set $W = \cup_{V\in\mathfrak{V}} V$. The set W is an open neighborhood of C, and by local finiteness and Lemma 20.1.1 we have

$$\overline{W} = \bigcup_{V\in\mathfrak{V}} \overline{V}.$$

Since $p \notin \overline{V}$ for all $V \in \mathfrak{V}$, we deduce that $p \notin \overline{W}$. It follows that $X \setminus \overline{W}$ and W are disjoint open neighborhoods of p and C. $\square$

To prove that X is normal, let C and D be disjoint closed subsets of X. We can find disjoint open neighborhoods of C and D by the same argument we used to prove the above claim. Simply substitute the above claim for X being Hausdorff and replace every occurrence of the point p by the closed set D. $\square$

20.5. Strong refinements

Let $\mathfrak{U}$ be an open cover of a space X. Enumerate $\mathfrak{U}$ as $\mathfrak{U} = \{U_i\}_{i\in I}$. A *strong refinement* of $\mathfrak{U}$ consists of an open cover $\{V_i\}_{i\in I}$ such that $\overline{V}_i \subset U_i$ for all $i \in I$. We have:

LEMMA 20.5.1. *Let X be a paracompact space and let $\mathfrak{U}$ be an open cover of X. Then there exists a locally finite strong refinement of $\mathfrak{U}$.*

PROOF. Enumerate $\mathfrak{U}$ as $\mathfrak{U} = \{U_i\}_{i\in I}$. Let

$$\mathfrak{W}' = \left\{ W' \mid W' \text{ open set with } \overline{W}' \subset U_i \text{ for some } i \in I \right\}.$$

The set $\mathfrak{W}'$ is an open cover of X; indeed, since X is normal for all $p \in X$ and all $i \in I$ with $p \in U_i$ we can find an open neighborhood W' of p with $\overline{W}' \subset U_i$. Since X is paracompact, we can find a locally finite refinement $\mathfrak{W}$ of $\mathfrak{W}'$. For each $W \in \mathfrak{W}$, there is some $i \in I$ with $\overline{W} \subset U_i$. For $i \in I$, let $\mathfrak{W}(i) = \{W \in \mathfrak{W} \mid \overline{W} \subset U_i\}$ and $V_i = \cup_{W\in\mathfrak{W}(i)} W$. Since $\mathfrak{W}(i)$ is a locally finite collection of open sets, Lemma 20.1.1 implies that

$$\overline{V}_i = \bigcup_{W\in\mathfrak{W}(i)} \overline{W} \subset U_i.$$

The open cover $\mathfrak{V} = \{V_i\}_{i\in I}$ is thus a locally finite strong refinement of $\mathfrak{U} = \{U_i\}_{i\in I}$. $\square$

20.6. Partitions of unity

We now come to the most important property of paracompact spaces. Let X be a space. Recall that for a continuous function $f\colon X \to \mathbb{R}$, the support of f is $\mathrm{supp}(f) = \overline{\{p \in X \mid f(p) \neq 0\}}$. A *partition of unity* subordinate to an open cover $\mathfrak{U}$ of X consists of continuous functions $f_U\colon X \to I = [0,1]$ for each $U \in \mathfrak{U}$ satisfying the following three conditions:

(a) For all $U \in \mathfrak{U}$, we have $\mathrm{supp}(f) \subset U$.
(b) The set $\{\mathrm{supp}(f_U) \mid U \in \mathfrak{U}\}$ is locally finite.
(c) For all $p \in X$, we have $\sum_{U\in\mathfrak{U}} f_U(p) = 1$. Note that (b) implies that only finitely many terms of this sum are nonzero, so this sum makes sense.

We have:

THEOREM 20.6.1. *Let X be a paracompact space and let $\mathfrak{U}$ be an open cover of X. Then there exists a partition of unity subordinate to $\mathfrak{U}$.*

PROOF. Enumerate $\mathfrak{U}$ as $\mathfrak{U} = \{U_i\}_{i\in I}$. By Lemma 20.5.1, we can find a locally finite strong refinement $\{V_i\}_{i\in I}$ of $\{U_i\}_{i\in I}$. Applying this lemma again, we obtain a locally finite strong refinement $\{W_i\}_{i\in I}$ of $\{V_i\}_{i\in I}$. Lemma 20.4.1 says that X is normal, so we can apply Urysohn's Lemma (Theorem 16.5.1) to X. For $i \in I$, since $\overline{W}_i \subset V_i$ Urysohn's Lemma (Theorem 16.5.1) implies that there is a continuous function $f_i'\colon X \to I$ such that $f_i'|_{\overline{W}_i} = 1$ and $\mathrm{supp}(f_i') \subset V_i$. Since $\{V_i\}_{i\in I}$ is locally finite and $\mathrm{supp}(f_i') \subset \overline{W}_i \subset V_i$ for each $i \in I$, we can define $g\colon X \to [0,\infty)$ via the formula

$$g(p) = \sum_{i\in I} f_i'(p) \quad \text{for } p \in X.$$

The function $g\colon X \to [0,\infty)$ is continuous (see Exercise 20.6). Each $p \in X$ lies in some W_i, so since $f'_i|_{\overline{W_i}} = 1$ it follows that $g(p) > 0$ for all $p \in X$. For $i \in I$, we can therefore define $f_i\colon X \to [0,\infty)$ via the formula

$$f_i(p) = \frac{1}{g(p)} f'_i(p) \quad \text{for } p \in X.$$

For $p \in X$, we have

$$\sum_{i \in I} f_i(p) = \frac{1}{g(p)} \sum_{i \in I} f'_i(p) = \frac{1}{g(p)} g(p) = 1.$$

Since $f_i(p) \in [0,\infty)$ for all $i \in I$, this implies that the image of each f_i lies in I and that the f_i form a partition of unity subordinate to $\mathfrak{U} = \{U_i\}_{i \in I}$. $\square$

20.7. Application: extending functions

Here is a typical application of partitions of unity:

LEMMA 20.7.1. *Let X be a paracompact space, let $A \subset X$ be a subspace, and let $f\colon A \to \mathbb{R}$ be continuous. For all $a \in A$, assume that there is a neighborhood U_a of a and an extension of $f|_{U_a \cap A}$ to $F_a\colon U_a \to \mathbb{R}$. Set $U = \cup_{a \in A} U_a$. Then f can be extended to a continuous function $F\colon U \to \mathbb{R}$.*

REMARK 20.7.2. If A is closed, then the Tietze extension theorem (Theorem 16.9.1) says that f can be extended to the whole space X. This can fail for non-closed subspaces. For instance, consider the subspace $\mathbb{Q}$ of $\mathbb{R}$. The continuous function $f\colon \mathbb{Q} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} -1 & \text{if } x < \sqrt{2}, \\ 1 & \text{if } x > \sqrt{2} \end{cases} \quad \text{for } x \in \mathbb{Q}.$$

can be extended to a continuous function on the open set $\mathbb{R} \setminus \{\sqrt{2}\}$, but cannot be extended to a continuous function on $\mathbb{R}$. $\square$

PROOF. Replacing X by U, we can assume that $\mathfrak{U} = \{U_a \mid a \in A\}$ is an open cover of X. Let $\{\phi_{U_a}\colon X \to \mathbb{R} \mid a \in A\}$ be a partition of unity subordinate to $\mathfrak{U}$. Since $\operatorname{supp}(\phi_{U_a}) \subset U_a$, the function $F_a \phi_{U_a}\colon U_a \to \mathbb{R}$ can be extended to a continuous function $G_a\colon X \to \mathbb{R}$ by letting $G_a(x) = 0$ for $x \in X \setminus U_a$. We have $\operatorname{supp}(G_a) \subset \operatorname{supp}(\phi_a)$ for $a \in A$, so since the set of supports of the ϕ_a are locally finite we can define $F\colon X \to \mathbb{R}$ via the formula $F = \sum_{a \in A} G_a$. For $a \in A$, we have

$$F(a) = \sum_{a \in A} F_a(a) \phi_{U_a}(a) = f(a) \sum_{a \in A} \phi_{U_a}(a) = f(a),$$

so F is an extension of f. $\square$

20.8. Exercises

EXERCISE 20.1. Give an example of a space X and a non-locally finite open cover $\mathfrak{U}$ of X such that for all $p \in X$ there are only finitely many $U \in \mathfrak{U}$ with $p \in U$. $\square$

EXERCISE 20.2. Let X be a space and let $\mathfrak{Z}$ be a collection of subsets of X.

(a) If $\mathfrak{Z}$ is locally finite, prove that

$$\overline{\bigcup_{Z \in \mathfrak{Z}} Z} = \bigcup_{Z \in \mathfrak{Z}} \overline{Z}.$$

(b) Give an example to show that local finiteness is needed in the previous part. $\square$

EXERCISE 20.3. Let X be a space, let $\mathfrak{U}$ be an open cover of X, and let $\mathfrak{V}$ be an open cover of X that refines $\mathfrak{U}$. Assume that $\mathfrak{V}$ has a finite subcover. Prove that $\mathfrak{U}$ has a finite subcover. $\square$

EXERCISE 20.4. Let X be paracompact and let $A \subset X$ be closed. Topologize X/A using the quotient topology (see Example 13.2.8). Prove that X/A is paracompact. $\square$

EXERCISE 20.5. Let X and Y be paracompact spaces, let $A \subset X$ be closed, and let $\phi\colon A \to Y$ be a closed map. Let Z be the result of gluing X to Y using ϕ (see Example 13.2.1). Prove that Z is paracompact. $\square$

Exercise 20.6. Let X be a space and let $\{V_i\}_{i \in I}$ be a locally finite collection of open subsets of X. For each $i \in I$, let $h_i\colon X \to \mathbb{R}$ be a continuous function such that $\operatorname{supp}(h_i) \subset V_i$. Define $h\colon X \to \mathbb{R}$ via the formula

$$h(p) = \sum_{i \in I} h_i(p) \quad \text{for } p \in X.$$

Prove that $h\colon X \to \mathbb{R}$ is continuous. □

Exercise 20.7. Let S be the Sorgenfrey line (see Example 12.7.5), so S is $\mathbb{R}$ with the topology given by the basis $\{[a,b) \mid a < b\}$ of all half-open intervals in $\mathbb{R}$. Prove that S is paracompact. Warning: this is harder than most of our other exercises. Exercises 16.12 and 17.11 might be helpful. These exercises show that S is a normal Lindelöf space, and in fact all normal Lindelöf spaces are paracompact.[1] □

Bibliography

[1] M. Rudin, A new proof that metric spaces are paracompact, Proc. Amer. Math. Soc. 20 (1969), 603. (Cited on page 152.)

[2] A. Stone, Paracompactness and product spaces, Bull. Amer. Math. Soc. 54 (1948), 977–982. (Cited on page 152.)

[1] In fact, only regularity is needed here since all regular Lindelöf spaces are normal.

CHAPTER 21

Products and Tychonoff's theorem

We now discuss products of spaces, generalizing the topology on $X \times Y$ from Example 12.7.4.

21.1. Finite products

Let $X_1, \dots, X_n$ be spaces. As a set, $X_1 \times \cdots \times X_n$ consists of tuples $(x_1, \dots, x_n)$ with $x_i \in X_i$ for $1 \le i \le n$. Give this the topology with the basis consisting of products $U_1 \times \cdots \times U_n$ with $U_i \subset X_i$ open for $1 \le i \le n$. We will call these the *basic open sets* of the product. A general open set $V \subset X_1 \times \cdots \times X_n$ can therefore be written a union of basic open sets. Equivalently, $V \subset X_1 \times \cdots \times X_n$ is open if and only if for all $(p_1, \dots, p_n) \in V$, there exist open neighborhoods $U_i \subset X_i$ of each p_i such that $U_1 \times \cdots \times U_n \subset V$.

EXAMPLE 21.1.1. This gives the usual topology on $\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$ (see Exercise 21.1). $\square$

21.2. Finite universal property

Let $\pi_i \colon X_1 \times \cdots \times X_n \to X_i$ be the projection. This is continuous: if $U_i \subset X_i$ is open, then

$$\pi_i^{-1}(U_i) = X_1 \times \cdots \times X_{i-1} \times U_i \times X_{i+1} \times \cdots \times X_n.$$

Now let Y be another space, and for $1 \le i \le n$ let $f_i \colon Y \to X_i$ be a continuous map. Let $f_1 \times \cdots \times f_n \colon Y \to X_1 \times \cdots \times X_n$ be the map where $f_1 \times \cdots \times f_n(y) = (f_1(y), \dots, f_n(y))$ for $y \in Y$. This is continuous; indeed, if $U_i \subset X_i$ is open for $1 \le i \le n$ then

$$(f_1 \times \cdots \times f_n)^{-1}(U_1 \times \cdots \times U_n) = U_1 \cap \cdots \cap U_n.$$

Conversely, if $F \colon Y \to X_1 \times \cdots \times X_n$ is a continuous map, then letting $f_i = \pi_i \circ F$ we have $F = f_1 \times \cdots \times f_n$. We summarize this informally as:

- A continuous map $F \colon Y \to X_1 \times \cdots \times X_n$ is the same thing as a collection of continuous maps $f_i \colon Y \to X_i$ for all $1 \le i \le n$.

Just like for quotient spaces in §13.4, this is an example of a universal mapping property. We will say more about it in §21.6 below.

21.3. Homotopies, products, and quotient maps

One place where products show up in algebraic topology is in the definition of a homotopy. Roughly speaking, a homotopy is a continuous deformation of a map. The precise definition is as follows. Let $f_0, f_1 \colon Y \to Z$ be maps. A *homotopy* from f_0 to f_1 is a map $H \colon Y \times I \to Z$ such that $H(y, 0) = f_0(y)$ and $H(y, 1) = f_1(y)$ for $y \in Y$. If such a homotopy exists, we say that f_0 and f_1 are *homotopic*

For $t \in I$, we can let $f_t \colon Y \to Z$ be the map defined by $f_t(y) = H(y, t)$ for $y \in Y$. The maps $f_t \colon Y \to Z$ can be viewed informally as a continuous family of maps connecting f_0 to f_1. See §23.7 for how to topologize the space of maps $Y \to Z$ and make this precise.

EXAMPLE 21.3.1. Any two maps $f_0, f_1 \colon Y \to \mathbb{R}^n$ are homotopic via the homotopy $H \colon Y \times I \to \mathbb{R}^n$ defined by $H(y, t) = (1 - t) f_0(y) + t f_1(y)$. $\square$

EXAMPLE 21.3.2. Let $Y = \{*\}$ be a one-point space. Two maps $f_0, f_1 \colon Y \to Z$ are homotopic if and only if $f_0(*)$ and $f_1(*)$ lie in the same path component of Z. $\square$

Now assume that $q\colon X \to Y$ is a quotient map (see §13.3), so q is surjective and $U \subset Y$ is open if and only if $q^{-1}(U) \subset X$ is open. Given $f_0, f_1\colon Y \to Z$, it is natural to try to construct a homotopy from f_0 to f_1 as follows:

- Define $g_0 = f_0 \circ q$ and $g_1 = f_1 \circ q$. Construct a homotopy $\widetilde{H}\colon X \times I \to Z$ from g_0 to g_1.
- Next, use the universal property of the quotient map from §13.4 to show that $\widetilde{H}$ descends to a homotopy $H\colon Y \times I \to Z$.

Here are an example of how this might work:

Example 21.3.3. We have $\mathbb{D}^n/\partial\mathbb{D}^n \cong \mathbb{S}^n$ (see Example 13.2.10). A map $f\colon \mathbb{S}^n \to Z$ is thus the same as a map $g\colon \mathbb{D}^n \to Z$ such that $g|_{\partial\mathbb{D}^n}$ is constant. Given $f_0, f_1\colon \mathbb{S}^n \to Z$, let $g_0, g_1\colon \mathbb{D}^n \to Z$ be the corresponding maps. To construct a homotopy from f_0 to f_1, it is natural to instead try to construct a homotopy g_t from g_0 to g_1 such that $g_t|_{\partial\mathbb{D}^n}$ is constant for all t. □

However, there is a flaw in the above reasoning: if $q\colon X \to Y$ is a quotient map, it not clear that $q \times \mathbb{1}\colon X \times I \to Y \times I$ is a quotient map. Indeed, there are counterexamples if I is replaced by a more complicated space. However, for nice spaces like I this is not a problem. More generally:

Lemma 21.3.4. *Let $q\colon X \to Y$ be a quotient map and let Z be a locally compact space. Then the map $q \times \mathbb{1}\colon X \times Z \to Y \times Z$ is a quotient map.*

Proof.[1] The map $q \times \mathbb{1}\colon X \times Z \to Y \times Z$ is continuous, so for every open set $U \subset Y \times Z$ we have $q^{-1}(U)$ open. We must prove the converse. In other words, letting $U \subset Y \times Z$ be a set such that $q^{-1}(U)$ is open, we must prove that U is open. Letting $(y, z) \in U$, it is enough to find an open neighborhood of (y, z) that is contained in U.

Pick $x \in X$ with $q(x) = y$. We have $(x, z) \in q^{-1}(U)$. Since $q^{-1}(U) \subset X \times Z$ is open and Z is locally compact, we can find an open neighborhood $V_1 \subset X$ of x and a compact neighborhood $K \subset Z$ of z such that $V_1 \times K \subset q^{-1}(U)$. We have

$$(y, z) \in q(V_1 \times \mathrm{Int}(K)) = q(V_1) \times \mathrm{Int}(K) \subset U.$$

If $q(V_1) \subset Y$ were open, then $q(V_1) \times \mathrm{Int}(K)$ would be an open neighborhood of (y, z) contained in U and we would be done.

Unfortunately, $q(V_1)$ might not be open since $q^{-1}(q(V_1))$ might be larger than V_1. We do have $q^{-1}(q(V_1)) \times K \subset q^{-1}(U)$. Since K is compact and $q^{-1}(U)$ is open, we can find an open neighborhood V_2 of $q^{-1}(q(V_1))$ with $V_2 \times K \subset q^{-1}(U)$ (see Exercise 17.8; this is often called the "tube lemma"). Just like for V_1, there is no reason to expect $q(V_2) \subset Y$ to be open since $q^{-1}(q(V_2))$ might be larger than V_2. However, we can iterate the procedure we used to find V_2. The result is an increasing sequence $V_1 \subset V_2 \subset \cdots$ of open subsets of Y such that for all $n \geq 1$ we have:

- $V_n \times K \subset q^{-1}(U)$ and $q^{-1}(q(V_n)) \subset V_{n+1}$.

The set $V = \cup_{n \geq 1} V_n$ is then an open subset of X with $V \times K \subset q^{-1}(U)$ and $q^{-1}(q(V)) = V$. It follows that $q(V)$ is an open subset of Y, so $q(V) \times \mathrm{Int}(K)$ is an open neighborhood of (y, z) with $q(V) \times \mathrm{Int}(K) \subset U$, as desired. □

21.4. Tychonoff's theorem, finite case

We have the following basic result:

Theorem 21.4.1 (Tychonoff's theorem, finite case). *Let $X_1, \ldots, X_n$ be compact spaces. Then $X_1 \times \cdots \times X_n$ is compact.*

Proof. By induction, it is enough to prove this for $n = 2$. Let $\mathfrak{U}$ be an open cover of $X_1 \times X_2$. We must prove that $\mathfrak{U}$ has a finite subcover. In fact, it is enough to prove that some refinement of $\mathfrak{U}$ has a finite subcover (see Exercise 20.3). Each element of $\mathfrak{U}$ is a union of basic open sets. Letting $\mathfrak{V}$ be the set of all basic open sets V such that there exists some $U \in \mathfrak{U}$ with $V \subset U$, it is therefore enough to prove that $\mathfrak{V}$ has a finite subcover.

[1] We will give an alternate proof in Chapter 23 which is shorter but more abstract. See §23.8.

For $p \in X_1$, let $Z(p) = p \times X_2$. By assumption, $Z(p) \cong X_2$ is compact. We can therefore find a finite subset $\mathfrak{V}(p)$ of $\mathfrak{V}$ that covers $Z(p)$. Since $\mathfrak{V}$ consists of basic open sets, we can write

$$\mathfrak{V}(p) = \{V_1(p) \times V_1'(p), \ldots, V_{m_p}(p) \times V_{m_p}'(p)\}$$

with $V_i(p) \subset X_1$ and $V_i'(p) \subset X_2$ for $1 \leq i \leq m_p$. Discarding unneeded terms if necessary, we can assume that $p \in V_i(p)$ for all $1 \leq i \leq m_p$. Letting $V(p) = V_1(p) \cap \cdots \cap V_{m_p}(p)$, it follows that $V(p)$ is an open neighborhood of p and $\mathfrak{V}(p)$ covers $V(p) \times X_2$.

The set $\{V(p) \mid p \in X_1\}$ is an open cover of the compact space X_1, so we can find $p_1, \ldots, p_d \in X_1$ such that $X_1 = V(p_1) \cup \cdots \cup V(p_d)$. Since $\mathfrak{V}(p_i)$ is a finite cover of $V(p_i) \times X_2$ for $1 \leq i \leq d$, we conclude that $\mathfrak{V}(p_1) \cup \cdots \mathfrak{V}(p_d)$ is a finite subset of $\mathfrak{V}$ that covers $X_1 \times X_2$. $\square$

21.5. Infinite products

Now let $\{X_i\}_{i \in I}$ be an arbitrary collection of spaces. As a set, the product $\prod_{i \in I} X_i$ consists of tuples $(x_i)_{i \in I}$ with $x_i \in X_i$ for $i \in I$. The obvious first guess for a topology on $\prod_{i \in I} X_i$ is the one with basis the collection of products $\prod_{i \in I} U_i$ with $U_i \subset X_i$ open for all $i \in I$. However, this topology turns out to be pathological. The issue is that it has too many open sets, and there are maps into it that should be continuous but are not. Here is a key example:

EXAMPLE 21.5.1. Let X be a space and let I be an infinite indexing set. Consider the diagonal map $\Delta\colon X \to \prod_{i \in I} X$, so $\Delta(x) = (x)_{i \in I}$ for all $x \in X$. If $U_i \subset X$ is an open set for all $i \in I$, then

$$\Delta^{-1}(\prod_{i \in I} U_i) = \bigcap_{i \in I} U_i.$$

Since the collection of open sets is *not* closed under infinite intersections, this is not always open. It follows that Δ will generally not be continuous if all such sets of the form $\prod_{i \in I} U_i$ are open. $\square$

To eliminate this pathology, we must avoid infinite intersections of open sets. This can be done as follows. A *basic open set* in $\prod_{i \in I} X_i$ is a product $\prod_{i \in I} U_i$ such that:

- $U_i \subset X_i$ is open for all $i \in I$; and
- $U_i = X_i$ for all but finitely many $i \in I$.

The *product topology* on $\prod_{i \in I} X_i$ is the topology with basis the basic open sets, so a subset of $\prod_{i \in I} X_i$ is open if and only if it is a union of basic open sets. To simplify our notation when talking about these infinite products, we introduce the following convention:

CONVENTION 21.5.2. We regard the indexing set I as being unordered, and thus if $I = J \sqcup K$ we identify

$$\left(\prod_{j \in J} X_j\right) \times \left(\prod_{k \in K} X_k\right) \quad \text{and} \quad \prod_{i \in I} X_i$$

in the obvious way. $\square$

With this notational convention, the basic open sets in $\prod_{i \in I} X_i$ are those that for some distinct $j_1, \ldots, j_n \in I$ can be written as

$$U_{j_1} \times \cdots \times U_{j_n} \times \prod_{i \in I \setminus \{j_1, \ldots, j_n\}} X_i \quad \text{with } U_{j_k} \subset X_{j_k} \text{ open for } 1 \leq k \leq n.$$

REMARK 21.5.3. The topology on $\prod_{i \in I} X_i$ with basis arbitrary products $\prod_{i \in I} U_i$ with $U_i \subset X_i$ open is sometimes called the *box topology*. It is rarely useful. $\square$

21.6. Infinite universal property

Continue to let $\{X_i\}_{i \in I}$ be an arbitrary collection of spaces. For $j \in I$, let $\pi_j\colon \prod_{i \in I} X_i \to X_j$ be the projection. The map π_j is continuous; indeed, if $U_j \subset X_i$ is open, then

$$\pi_j^{-1}(U_j) = U_j \times \prod_{i \in I \setminus \{j\}} X_i.$$

Now let Y be another space, and for $i \in I$ let $f_i\colon Y \to X_i$ be a continuous map. Let $\prod_{i\in I} f_i\colon Y \to \prod_{i\in I} X_i$ be the map

$$\left(\prod_{i\in I} f_i\right)(y) = (f_i(y))_{i\in I} \quad \text{for } y \in Y.$$

This map is continuous; indeed, if $\prod_{i\in I} U_i$ is a basic open set then

$$(\left(\prod_{i\in I} f_i\right)^{-1}(\prod_{i\in I} U_i) = \bigcap_{i\in I} f_i^{-1}(U_i).$$

This is open since $f_i^{-1}(U_i) = f_i^{-1}(X_i) = Y$ for all but finitely many $i \in I$, so this intersection is actually a finite intersection. Conversely, if $F\colon Y \to \prod_{i\in I} X_i$ is a continuous map, then letting $f_i = \pi_i \circ F$ we have $F = \prod_{i\in I} f_i$. We summarize this informally as:

- A continuous map $F\colon Y \to \prod_{i\in I} X_i$ is the same thing as a collection of continuous maps $f_i\colon Y \to X_i$ for all $i \in I$.

Having this universal property is one of the reasons we defined the product topology like we did. It follows from Exercise 21.3 that this universal property categorizes $\prod_{i\in I} X_i$ up to homeomorphism.

21.7. Categorical interpretation

If $\mathbf{C}$ is a category and $\{C_i\}_{i\in I}$ are objects of $\mathbf{C}$, then a *categorical product* of the C_i is an object D of $\mathbf{C}$ together with morphisms $\{f_i\colon C_i \to D\}_{i\in I}$ such that the following holds:

- For all objects E of $\mathbf{C}$, there is a bijection between morphisms $\phi\colon E \to D$ and collections of morphisms $\{\Phi_i\colon E \to C_i\}_{i\in I}$ taking a morphism $\phi\colon E \to D$ to the collection of morphisms $\{f_i\colon \phi\colon E \to C_i\}_{i\in I}$.

Categorical products might or might not exist, but if they do exist they are unique up to isomorphism (see Exercise 21.3). The universal property from §21.6 shows that if $\{X_i\}_{i\in I}$ is a collection of spaces, the product $P = \prod_{i\in I} X_i$ equipped with the projections $\{\pi_i\colon P \to X_i\}_{i\in I}$ is the categorical product of the $\{X_i\}_{i\in I}$ in the category $\mathtt{Top}$ of topological spaces. See Exercise 21.4 for the product in the category of abelian groups.

Remark 21.7.1. This should be compared to the categorical sum from Remark 13.4.2, which we now recall. If $\mathbf{C}$ is a category and $\{C_i\}_{i\in I}$ are objects of $\mathbf{C}$, then a *categorical sum* of the C_i is an object D of $\mathbf{C}$ together with morphisms $\{f_i\colon C_i \to D\}_{i\in I}$ such that the following holds:

- For all objects E of $\mathbf{C}$, there is a bijection between morphisms $\phi\colon D \to E$ and collections of morphisms $\{\Phi_i\colon C_i \to E\}_{i\in I}$ taking a morphism $\phi\colon D \to E$ to the collection of morphisms $\{\phi \circ f_i\colon C_i \to E\}_{i\in I}$.

As we discussed in Remark 13.4.2, the categorical sum of a collection of spaces $\{X_i\}_{i\in I}$ is the disjoint union $\sqcup_{i\in i} X_i$. □

21.8. Metrics on countable products

Arbitrary products of metrizable spaces need not be metrizable. However, it turns out that countable products of metricizable spaces are metrizable. This would not be true if we used the box topology. This is easy for finite products (see Exercise 21.6), so we focus on the countably infinite case:

Lemma 21.8.1. *For each $n \geq 1$, let M_n be a metrizable space. Then $\prod_{n=1}^{\infty} M_n$ is metrizable.*

Proof. Let $\mathfrak{d}_n$ be a metric on M_n inducing its topology. Let $\mathfrak{d}'_n$ be the metric on M_n defined by $\mathfrak{d}'_n(p,q) = \min\{\mathfrak{d}_n(p,q), 1\}$. This induces the same topology on M_n as $\mathfrak{d}_n$ (see Exercise 12.8). We can then define a two-variable real-valued function on $\prod_{n=1}^{\infty} M_n$ via the formula

$$\mathfrak{d}((p_n)_{n\geq 1}, (q_n)_{n\geq 1}) = \sum_{n=1}^{\infty} \frac{1}{2^n}\, \mathfrak{d}'_n(p_n, q_n).$$

This is a metric on $\prod_{n=1}^{\infty} M_n$ that induces the product topology (see Exercise 21.7). □

21.9. Tychonoff's theorem, countable case

Tychonoff's theorem generalizes to arbitrary products of compact spaces. We start by proving this for countable products. The proof of the general case is similar, but requires more set theoretic technology.

THEOREM 21.9.1 (Tychonoff's theorem, countable case). *Let $\{X_i\}_{i \geq 1}$ be a countable collection of compact spaces. Then $\prod_{i \geq 1} X_i$ is compact.*

PROOF. Unlike in the finite case, we cannot prove this by induction. However, we will see that the argument we gave in the finite case is almost enough. Only one new idea is needed. Let $\mathfrak{U}$ be an open cover of $\prod_{i \geq 1} X_i$. We must prove that $\mathfrak{U}$ has a finite subcover. In fact, it is enough to prove that some refinement of $\mathfrak{U}$ has a finite subcover (see Exercise 20.3). Each element of $\mathfrak{U}$ is a union of basic open sets. Letting $\mathfrak{V}$ be the set of all basic open sets V such that there exists some $U \in \mathfrak{U}$ with $V \subset U$, it is therefore enough to prove that $\mathfrak{V}$ has a finite subcover.

Assume for the sake of contradiction that $\mathfrak{V}$ has no finite subcover. The proof now has two steps:

STEP 1. *For all $i \geq 1$, there exists some $p_i \in X_i$ such that no finite subset of $\mathfrak{V}$ covers $p_1 \times \cdots \times p_n \times \prod_{i \geq n+1} X_i$ for any $n \geq 1$.*

We construct the p_i inductively. Assume that for some $n \geq 1$ we have found $p_i \in X_i$ for $1 \leq i \leq n-1$ such that no finite subset of $\mathfrak{V}$ covers $p_1 \times \cdots \times p_{n-1} \times \prod_{i \geq n} X_i$. For $n = 1$, this is simply our assumption that the open cover $\mathfrak{V}$ of $\prod_{i \geq 1} X_i$ has no finite subcover. We find $p_n \in X_n$ as follows. For $p \in X_n$, let

$$Z(p) = p_1 \times \cdots \times p_{n-1} \times p \times \prod_{i \geq n+1} X_i.$$

Assume for the sake of contradiction that for all $p \in X_n$, there exists a finite subset $\mathfrak{V}(p)$ of $\mathfrak{V}$ that covers $Z(p)$. Since $\mathfrak{V}$ consists of basic open sets, we can write

$$\mathfrak{V}(p) = \left\{ \prod_{i \geq 1} V_{i,j}(p) \mid 1 \leq j \leq m_p \right\}$$

with $V_{i,j}(p) \subset X_i$ for all $i \geq 1$ and $1 \leq j \leq m_p$. Discarding unneeded terms if necessary, we can assume that $p_i \in V_{i,j}(p)$ for all $1 \leq i \leq n-1$ and $1 \leq j \leq m_p$, and also that $p \in V_{n,j}(p)$ for all $1 \leq j \leq m_p$. Define

$$V_i(p) = \bigcap_{j=1}^{m_p} V_{i,j}(p) \quad \text{for } 1 \leq i \leq n,$$
$$V(p) = V_1(p) \times \cdots \times V_n(p).$$

It follows that $V(p)$ is an open neighborhood of $(p_1, \ldots, p_{n-1}, p) \in X_1 \times \cdots \times X_n$ and that $\mathfrak{V}(p)$ covers $V(p) \times \prod_{i \geq n+1} X_i$.

The set $\{V(p) \mid p \in X_n\}$ is an open cover of the compact space $p_1 \times \cdots \times p_{n-1} \times X_n$, so we can find $q_1, \ldots, q_d \in X_n$ such that

$$p_1 \times \cdots \times p_{n-1} \times X_n \subset V(q_1) \cup \cdots \cup V(q_d).$$

Since $\mathfrak{V}(q_k)$ is a finite cover of $V(q_k) \times \prod_{i \geq n+1} X_i$ for $1 \leq k \leq d$, we conclude that $\mathfrak{V}(q_1) \cup \cdots \mathfrak{V}(q_d)$ is a finite subset of $\mathfrak{V}$ that covers $p_1 \times \cdots \times p_{n-1} \times \prod_{i \geq n} X_i$, contradicting the fact that no such finite cover exists.

STEP 2. *No finite subset of $\mathfrak{V}$ covers $\prod_{i \geq 1} X_i$.*

Pick $V \in \mathfrak{V}$ such that $(p_i)_{i \geq 1} \in V$. Since $\mathfrak{V}$ consists of basic open sets, we can write $V = \prod_{i \geq 1} V_i$ with $V_i \subset X_i$ open for all $i \geq 1$. Moreover, we have $V_i = X_i$ for all but finitely many $i \geq 1$. This implies that there exists some $n \geq 1$ such that $V_i = X_i$ for $i \geq n+1$. It follows that

$$p_1 \times \cdots \times p_n \times \prod_{i \geq n+1} X_i \subset V \in \mathfrak{V}.$$

This contradicts the fact that no finite subset of $\mathfrak{V}$ covers $p_1 \times \cdots \times p_n \times \prod_{i \geq n+1} X_i$. □

21.10. Well-ordered sets

To generalize the above proof of Tychonoff's theorem to arbitrary products, we need some set-theoretic technology. Recall from §20.3 that a well-ordered set is a set I equipped with a total ordering $\leq$ such that every nonempty subset $S \subset I$ has a minimal element. We call $\leq$ a well-ordering. A remarkable consequence of the axiom of choice is that every set can be equipped with a well-ordering.

Example 21.10.1. All total orderings on finite sets are well-orderings. The canonical example of an infinite well-ordered set is $\mathbb{N}$ with its standard ordering. See Exercise 21.8 for more examples. $\square$

If I is a well-ordered set with ordering $\leq$, then an *initial segment* of I is a subset $J \subset I$ such that for all $j \in J$ and $i \in I$ with $i \leq j$ we have $i \in J$. If $J_1, J_2 \subset I$ are initial segments, then either $J_1 \subset J_2$ or $J_2 \subset J_1$. Indeed, assume that J_1 is *not* a subset of J_2 and pick $j_1 \in J_1 \setminus J_2$. For $j_2 \in J_2$, we cannot have $j_1 \leq j_2$ since $j_1 \notin J_2$. It follows that $j_2 \leq j_1$, so $j_2 \in J_1$ and thus $J_2 \subset J_1$. The initial segments of I are thus totally ordered under inclusion. They fall into three classes:

- The empty set $\emptyset$, which is the unique initial segment that is contained in all initial segments.
- The *successor segments*, which are initial segments $J \subset I$ of the form $J = J' \sqcup \{n\}$ for some initial segment $J' \subsetneq J$ and some $n \in J \setminus J'$.
- The *limit segments*, which are nonempty initial segments $J \subset I$ that are not successor segments. These J are the union of the initial segments $J' \subsetneq J$.

Example 21.10.2. For $\mathbb{N}$, the successor segments are of the form $\{1, \ldots, n\}$ and the whole set $\mathbb{N}$ is the only limit segment. $\square$

21.11. Transfinite induction

Assume now that I is a well-ordered set and for each $i \in I$ we have a set X_i. Our goal is to construct some $p_i \in X_i$ for all $i \in I$. For each initial segment $J \subset I$, we want some property $\mathcal{P}(J)$ to hold that only refers to the $p_i \in X_i$ for $i \in J$. To simplify our exposition, assume that if $\mathcal{P}(J)$ holds then so does $\mathcal{P}(J')$ for all initial segments $J' \subset J$.

We can construct the $p_i \in X_i$ by *transfinite induction*.[2] For this, we must prove three things:

(0) The property $\mathcal{P}(\emptyset)$ holds. This makes sense since $\mathcal{P}(\emptyset)$ makes no reference to any p_i.
(1) Let J be a successor segment of the form $J = J' \sqcup \{n\}$ for some initial segment $J' \subsetneq J$. Assume that we have already constructed $p_i \in X_i$ for all $i \in J'$ such that $\mathcal{P}(J')$ holds. We must show how to construct $p_n \in X_n$ such that $\mathcal{P}(J)$ holds.
(2) Let J be a limit segment. Assume that we have constructed $p_i \in X_i$ for all $i \in J$ such that $\mathcal{P}(J')$ holds for all initial segments $J' \subsetneq J$. We must prove that $\mathcal{P}(J)$ holds.

We can then construct $p_i \in X_i$ for all $i \in I$ such that $\mathcal{P}(J)$ holds for all initial segments $J \subset I$. Indeed, let $\mathfrak{J}$ be the set of all initial segments $J \subset I$ for which we can construct $p_i \in X_i$ for each $i \in J$ such that $\mathcal{P}(J)$ holds. The set $\mathfrak{J}$ is linearly ordered by inclusion and nonempty since $\emptyset \in \mathfrak{J}$. Let $J_0 = \cup_{J \in \mathfrak{J}} J$. By (1) and (2), we have $J_0 \in \mathfrak{J}$. We must prove that $J_0 = I$. Indeed, assume that $J_0 \subsetneq I$. Since I is well-ordered, there is a minimal $n \in I \setminus J_0$. It follows that $J_0 \sqcup \{n\}$ is an initial segment, and by (1) we have $J_0 \sqcup \{n\} \in \mathfrak{J}$, contradicting the fact that $J \subset J_0$ for all $J \in \mathfrak{J}$.

Remark 21.11.1. Isomorphism classes of well-ordered sets are called *ordinals*. Any set of ordinals has a well-ordering where $\mathcal{O}_1 \leq \mathcal{O}_2$ when $\mathcal{O}_1$ is isomorphic to an initial segment of $\mathcal{O}_2$. Transfinite induction is typically discussed using ordinals. $\square$

21.12. Tychonoff's theorem, general case

We now use transfinite induction to prove the general case of Tychonoff's theorem:

Theorem 21.12.1 (Tychonoff's theorem). *Let $\{X_i\}_{i \in I}$ be a collection of compact spaces. Then $\prod_{i \in I} X_i$ is compact.*

[2] Since we constructing things, this is sometimes called *transfinite recursion*.

Proof. The proof will be almost identical to proof in the countable case, but with some small complications due to the need for transfinite induction. Let $\mathfrak{U}$ be an open cover of $\prod_{i\in I} X_i$. We must prove that $\mathfrak{U}$ has a finite subcover. In fact, it is enough to prove that some refinement of $\mathfrak{U}$ has a finite subcover (see Exercise 20.3). Each element of $\mathfrak{U}$ is a union of basic open sets. Letting $\mathfrak{V}$ be the set of all basic open sets V such that there exists some $U \in \mathfrak{U}$ with $V \subset U$, it is therefore enough to prove that $\mathfrak{V}$ has a finite subcover.

Assume for the sake of contradiction that $\mathfrak{V}$ has no finite subcover. Choose a well-ordering on the indexing set I. By transfinite induction, for each $i \in I$ we will construct some $p_i \in X_i$ such that the following holds for all initial segments $J \subset I$:

($\spadesuit_J$) No finite subset of $\mathfrak{V}$ covers $Y(J) = \prod_{j\in J} p_j \times \prod_{i\in I\setminus J} X_i$.

The special case ($\spadesuit_I$) says that no finite subset of $\mathfrak{V}$ covers the one-point set $Y(I) = \prod_{i\in I} p_i$, which will be our contradiction. We have ($\spadesuit_\emptyset$) from our assumption that no finite subset of $\mathfrak{V}$ covers $Y(\emptyset) = \prod_{i\in I} X_i$. According to the transfinite induction scheme discussed in §21.11, to prove that ($\spadesuit_J$) holds for all initial segments $J \subset I$ we must prove:

Step 1. *Let $J \subset I$ be a successor segment, so $J = J' \sqcup \{n\}$ for some initial segment $J' \subset J$ and $n \in J \setminus J'$. Assume that we have constructed $p_i \in X_i$ for all $i \in J'$ such that ($\spadesuit_{J'}$) holds. We can then construct $p_n \in X_n$ such that ($\spadesuit_J$) holds.*

For $p \in X_n$, let
$$Z(p) = p \times \prod_{j'\in J'} p_{j'} \times \prod_{i\in I\setminus J} X_i.$$
Assume for the sake of contradiction that for all $p \in X_n$, there exists a finite subset $\mathfrak{V}(p)$ of $\mathfrak{V}$ that covers $Z(p)$. Since $\mathfrak{V}$ consists of basic open sets, we can write
$$\mathfrak{V}(p) = \left\{ \prod_{i\in I} V_{i,k}(p) \mid 1 \le k \le m_p \right\}$$
with $V_{i,k}(p) \subset X_i$ for all $i \in I$ and $1 \le k \le m_p$. Discarding unneeded terms if necessary, we can assume that $p_{j'} \in V_{j',k}(p)$ for all $j' \in J'$ and $1 \le k \le m_p$, and also that $p \in V_{n,k}(p)$ for all $1 \le k \le m_p$. Keeping in mind that $J = J' \sqcup \{n\}$, define
$$V_j(p) = \bigcap_{k=1}^{m_p} V_{j,k}(p) \quad \text{for } j \in J,$$
$$V(p) = V_n(p) \times \prod_{j'\in J'} V_{j'}(p).$$
It follows that $V(p)$ is an open neighborhood of $p \times \prod_{j'\in J} p_{j'}$ and that $\mathfrak{V}(p)$ covers $V(p) \times \prod_{i\ge I\setminus J} X_i$.

The set $\{V(p) \mid p \in X_n\}$ is an open cover of the compact space $X_n \times \prod_{j'\in J'} p_{j'}$, so we can find $q_1, \dots, q_d \in X_n$ such that
$$X_n \times \prod_{j'\in J'} p_{j'} \subset V(q_1) \cup \cdots \cup V(q_d).$$
Since $\mathfrak{V}(q_\ell)$ is a finite cover of $V(q_\ell) \times \prod_{i\in I\setminus J} X_i$ for $1 \le \ell \le d$, we conclude that $\mathfrak{V}(q_1) \cup \cdots \mathfrak{V}(q_d)$ is a finite subset of $\mathfrak{V}$ that covers
$$X_n \times \prod_{j'\in J'} p_{j'} \times \prod_{i\in I\setminus J} X_i = \prod_{j'\in J'} p_{j'} \times \prod_{i\in I\setminus J'} X_i = Y(J'),$$
contradicting the fact that no such finite cover exists.

Step 2. *Let $J \subset I$ be a limit segment. Assume that we have constructed p_i for all $i \in J$ such that ($\spadesuit_{J'}$) holds for all initial segments $J' \subsetneq J$. Then ($\spadesuit_J$) holds.*

Assume for the sake of contradiction that a finite subset $\{V_1, \dots, V_d\}$ of $\mathfrak{V}$ covers $Y(J)$. Each V_k is a basic open set, so we can write
$$V_k = \prod_{i\in I} V_{k,i} \quad \text{with } V_{k,i} \subset X_i \text{ open for all } i \in I.$$

Moreover, we have $V_{k,i} = X_i$ for all but finitely many $i \in I$. For $1 \leq k \leq d$, let $J(k) = \{j \in J \mid V_{k,j} \neq X_j\}$. Set $\widehat{J} = J(1) \cup \cdots\cdots \cup J(d)$. Let J' be the smallest initial segment containing $\widehat{J}$. Since $\widehat{J}$ is a finite subset of J, we have $J' \subsetneq J$. Since $V_{k,j} = X_j$ for all $1 \leq k \leq d$ and $j \in J \setminus J'$, the fact that $\{V_1, \ldots, V_d\}$ covers $Y(J)$ implies that it alsoo covers $Y(J')$. This contradicts the fact that no finite subset of $\mathfrak{V}$ covers $Y(J')$. □

21.13. Exercises

Exercise 21.1. Prove that the product topology on $\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$ is the same as the metric space topology. □

Exercise 21.2. Let $\{X_i\}_{i \in I}$ be spaces. For each $i \in I$, let $A_i \subset X_i$ be a nonempty subspace. We therefore have a nonempty subspace $\prod_{i \in I} A_i$ of $\prod_{i \in I} X_i$. Prove that $\overline{\prod_{i \in I} A_i} = \prod_{i \in I} \overline{A}_i$. □

Exercise 21.3. Let $\mathbf{C}$ be a category and let $\{C_i\}_{i \in I}$ be objects of $\mathbf{C}$. For $k = 1, 2$, let $D(k)$ be a categorical product of the $\{C_i\}_{i \in I}$ with morphisms $\{f_i(k) \colon D(k) \to C_i\}_{i \in I}$. Prove that there exists an isomorphism $\lambda \colon D(1) \to D(2)$ such that $f_i(2) = f_i(1) \circ \lambda$ for all $i \in I$. This can be interpreted as saying that categorical products are unique up to isomorphism. □

Exercise 21.4. Let $\{A_i\}_{i \in I}$ be a collection of abelian groups. Prove that $P = \prod_{i \in I} A_i$ together with the projections $\{\pi_i \colon P \to A_i\}$ is the categorical product of the A_i in the category $\mathtt{Ab}$Grp of abelian groups. We remark that Exercise 13.10 showed that the direct sum

$$\bigoplus_{i \in I} A_i = \left\{ (a_i)_{i \in I} \in \prod_{i \in I} A_i \mid a_i = 0 \text{ for all but finitely many } i \in I \right\}.$$

is the categorical sum of the $\{A_i\}_{i \in I}$ in $\mathtt{Ab}$Grp. □

Exercise 21.5. Let $\{X_i\}_{i \in I}$ be a collection of spaces. For each $i \in I$, let $\{p(i)_n\}_{n \geq 1}$ be a sequence of points in X_i that converges to $p(i) \in X_i$. For $n \geq 1$, let $p_n = (p(i)_n)_{i \in I} \in \prod_{i \in I} X_i$. Prove that $\{p_n\}_{n \geq 1}$ converges to $(p(i))_{i \in I} \in \prod_{i \in I} X_i$. □

Exercise 21.6. For $1 \leq n \leq N$, let $(M_n, \mathfrak{d}_n)$ be a metric space. Define a two-variable real-valued function on $\prod_{n=1}^N M_n$ via the formula

$$\mathfrak{d}((p_1, \ldots, p_N), (q_1, \ldots, q_N)) = \mathfrak{d}_1(p_1, q_1) + \cdots + \mathfrak{d}_N(p_N, q_N).$$

Prove that this is a metric on $\prod_{n=1}^N M_n$ that induces the product topology. □

Exercise 21.7. For each $n \geq 1$, let $(M_n, \mathfrak{d}_n)$ be a metric space. For each $n \geq 1$, assume that $\mathfrak{d}_n(p, q) \leq 1$ for all $p, q \in M_n$. Define a two-variable real-valued function on $\prod_{n=1}^\infty M_n$ via the formula

$$\mathfrak{d}((p_n)_{n \geq 1}, (q_n)_{n \geq 1}) = \sum_{n=1}^{\infty} \frac{1}{2^n} \mathfrak{d}'_n(p_n, q_n).$$

Prove that this is a metric on $\prod_{n=1}^\infty M_n$ that induces the product topology. □

Exercise 21.8. This exercise gives more examples of well-ordered sets.

(a) Let A and B be two well-ordered sets. Let $A \sqcup B$ be their disjoint union, and let $\leq$ be the total ordering on $A \sqcup B$ that restricts to the given orderings on A and B and has $a < b$ for all $a \in A$ and $b \in B$. Prove that $A \sqcup B$ is a well-ordered set.
(b) Let A and B be two well-ordered sets. Give $A \times B$ the dictionary ordering, so $(a_1, b_1) < (a_2, b_2)$ if $a_1 < a_2$ or if $a_1 = a_2$ and $b_1 < b_2$. Prove that $A \times B$ is a well-ordered set.
(c) For some $N \geq 1$, give $\{1, \ldots, N\} \times \mathbb{N}$ the dictionary ordering. Identify the initial segments, and say which are successor segments and which are limit segments.
(d) Give $\mathbb{N} \times \mathbb{N}$ the dictionary ordering. Identify the initial segments, and say which are successor segments and which are limit segments. □

EXERCISE 21.9. Let C be the classical Cantor set, i.e., the set of all $x \in I = [0,1]$ of the form

$$x = \sum_{n=1}^{\infty} \frac{x_n}{3^n} \quad \text{with } x_n \in \{0,2\} \text{ for all } n \geq 1.$$

Let X be the discrete 2-point space $X = \{0,2\}$. Prove the following:

(a) Define a set map $\Psi\colon \prod_{n=1}^{\infty} X \to C$ via the formula

$$\Psi((x_n)_{n\geq 1}) = \sum_{n=1}^{\infty} \frac{x_n}{3^n}.$$

Prove that Ψ is a homeomorphism.

(b) Prove that for all $p, q \in C$ there exists a homeomorphism $f\colon C \to C$ such that $f(p) = q$. In other words, every two points of C "look the same". The technical term for this is that C is *homogeneous*.

(c) Define a set map $\Phi\colon \prod_{n=1}^{\infty} X \to I$ via the formula

$$\Phi((x_n)_{n\geq 1}) = \sum_{n=1}^{\infty} \frac{x_n}{2^n}.$$

Prove that Φ is a continuous surjection.

(d) For each $d \geq 1$, construct a homeomorphism $\lambda_d\colon \prod_{n=1}^{\infty} X \to (\prod_{n=1}^{\infty} X)^{\times d}$.

(e) For each $d \geq 1$, let $f_d\colon C \to I^d$ be the composition

$$C \xrightarrow{\Psi^{-1}} \prod_{n=1}^{\infty} X \xrightarrow{\lambda_d} \left(\prod_{n=1}^{\infty} X\right)^{\times d} \xrightarrow{\prod_{i=1}^{d} \Phi} I^d.$$

Prove that f_d is a continuous surjection that can be extended to a continuous surjection $g_d\colon I \to I^d$ (a "space-filling curve"). □

CHAPTER 22

Metrization theorems

This chapter discusses topological properties that ensure that a space is metrizable.

22.1. Nagata–Smirnov metrization theorem

We have already observed that all metrizable spaces are normal (Lemma 16.4.1) and first countable (Lemma 15.1.1). These conditions are not sufficient. It turns out all metrizable spaces satisfy a condition that is intermediate between first and second countability. A space is first countable if it has a countable neighborhood basis at each point, and is second countable if it has a countable basis. Say that a basis $\mathcal{B}$ is *countably locally finite* if it can be written as $\mathcal{B} = \cup_{n=1}^{\infty} \mathcal{B}_n$ with each $\mathcal{B}_n$ locally finite. We are not requiring the $\mathcal{B}_n$ to be countable, so this is weaker than being second countable.

The Nagata–Smirnov metrization theorem says that a space X is metrizable if and only if it is regular and has a countably locally finite basis. We remark that these conditions imply that X is normal; see Exercise 22.2. See [**1**, Chapter 6] for a proof of the Nagata–Smirnov metrization theorem.

22.2. Urysohn metrization theorem

We prove the following weakening of the Nagata–Smirnov metrization theorem, which is often called the Urysohn metrization theorem:

Theorem 22.2.1. *Let X be a space that is regular and second countable. Then X is metrizable.*

Proof. Since second countable regular spaces are normal (Lemma 16.10.1), our space X is normal and we can apply Urysohn's lemma to it. Let $\mathcal{B}$ be a countable basis for X. Set

$$\mathcal{D} = \left\{(U, V) \mid U, V \in \mathcal{B} \text{ and } \overline{U} \subset V\right\}.$$

Since $\mathcal{B}$ is countable, the set $\mathcal{D}$ is also countable. Recalling that $I = [0, 1]$, define

$$Z = \prod_{(U,V)\in\mathcal{D}} I.$$

Since countable products of metrizable spaces are metrizable (Lemma 21.8.1), it follows that Z is metrizable. It is therefore enough to construct an embedding $F \colon X \to Z$.

For each $(U, V) \in \mathcal{D}$, apply Urysohn's Lemma (Theorem 16.5.1) to find a continuous function $f_{UV} \colon X \to I$ such that $f_{UV}|_{\overline{U}} = 1$ and such that $\operatorname{supp}(f_{UV}) \subset V$. Define a function

$$F \colon X \longrightarrow Z = \prod_{(U,V)\in\mathcal{D}} I$$

via the formula

$$F(x) = (f_{UV}(x))_{(U,V)\in\mathcal{D}} \quad \text{for } x \in X.$$

This is continuous by the universal property of the product topology (see §21.6). To prove that F is an embedding, we must prove that F is injective and that F takes open sets of X to open sets of $F(X) \subset Z$. We divide this into two steps:

Step 1. *The map F is injective.*

Consider distinct points $x, y \in X$. Since X is normal it is in particular Hausdorff. We can therefore find an open neighborhood $V \in \mathcal{B}$ with $x \in V$ and $y \notin V$. Applying normality, we can find an open neighborhood $U \in \mathcal{B}$ of x with $\overline{U} \subset V$. It follows that $f_{UV}(x) = 1$ and $f_{UV}(y) = 0$. The (U, V)-coordinates of $F(x)$ and $F(y)$ therefore differ, so $F(x) \neq F(y)$.

STEP 2. *The map F takes open sets of X to open sets of $F(X) \subset Z$.*

Consider $V_0 \in \mathcal{B}$. It is enough to prove that $F(V_0)$ is an open set of $F(X) \subset Z$. Consider some $x_0 \in V_0$. To prove that $F(V_0)$ is an open set of $F(X) \subset Z$, it is enough to prove that there is an open neighborhood $W_0 \subset Z$ of $F(x_0)$ with $W_0 \cap F(X) \subset F(V_0)$. Since X is normal, we can find an open neighborhood U_0 of x_0 with $\overline{U}_0 \subset V_0$. We therefore have $f_{U_0 V_0}(x_0) = 1$. Let

$$W_0 \subset Z = \prod_{(U,V) \in \mathcal{D}} I$$

be the open subset consisting of points whose (U_0, V_0)-coordinate is nonzero. Since $f_{U_0 V_0}(x_0) = 1$, we have $F(x_0) \in W_0$. We claim that $W_0 \cap F(X) \subset F(V_0)$. Indeed, consider a point $x_1 \in X$ with $F(x_1) \in W_0$. We must prove that $x_1 \in V_0$. To see this, note that since $F(x_1) \in W_0$ it must be the case that $f_{U_0 V_0}(x_1) \neq 0$. Since $\operatorname{supp}(f_{U_0 V_0}) \subset V_0$, it follows that $x_1 \in V_0$, as desired. □

22.3. Exercises

EXERCISE 22.1. Give an example of space X that metrizable but not second countable. □

EXERCISE 22.2. Prove the following variant of Lemma 16.10.1: if X is a regular space that has a countably locally finite basis, then X is normal. □

EXERCISE 22.3. Let M be a metrizable space. Prove that X has a countably locally finite basis. Hint: the proof uses the fact that metrizable spaces are paracompact. □

EXERCISE 22.4. Let X be a locally compact Hausdorff space and let X^* be the one-point compactification of X (see §18.3). Prove that X^* is metrizable if and only if X is second countable. □

EXERCISE 22.5. Let X be a paracompact space that is *locally metrizable*, i.e., each $p \in X$ has a metrizable open neighborhood.

(a) Prove that X has a countably locally finite basis. Hint: use the fact that metrizable spaces are countably locally finite.
(b) Use the Nagata–Smirnov metrization theorem to deduce that X is metrizable.

This is often called the *Smirnov metrization theorem*. It is enlightening to prove it directly, though the proof is too hard for an exercise. □

Bibliography

[1] J. Munkres, *Topology*, 2nd edition, Prentice Hall, Upper Saddle River, NJ, 2000. (Cited on page 167.)

CHAPTER 23

Function spaces and the compact-open topology

Let X and Y be spaces and let[1] $\mathcal{C}(X,Y)$ be the set of all continuous maps $f\colon X \to Y$. In this chapter we explain how to turn $\mathcal{C}(X,Y)$ into a space.

23.1. Subbasis

Let X be a set and let $\mathfrak{B}$ be a set of subsets of X. We would like to topologize X with the smallest collection of open sets possible to make each $U \in \mathfrak{B}$ open. If for all $U, V \in \mathfrak{B}$ the intersection $U \cap V$ could be written as a union of sets in $\mathfrak{B}$, then $\mathfrak{B}$ would be a basis for a topology as in §12.7. In that case, we could topologize X by saying that $U \subset X$ is open precisely when U is the union of sets in $\mathfrak{B}$.

However, if $\mathfrak{B}$ does not form a basis then this does not work since in the resulting "topology" the collection of open sets is not closed under finite intersections. To fix this, let $\mathfrak{B}'$ be the set of all finite intersections of elements of $\mathfrak{B}$. Here we interpret the intersection of zero sets as X, so $X \in \mathfrak{B}'$. The set $\mathfrak{B}'$ does form a basis for a topology on X. In this case, we say that $\mathfrak{B}$ is a *subbasis* for this topology.

23.2. Compact-open topology

For sets $A, B \subset X$, define

$$B(A,B) = \{f\colon X \to Y \mid f(K) \subset U\} \subset \mathcal{C}(X,Y).$$

The *compact-open topology* on $\mathcal{C}(X,Y)$ is the topology with subbasis the collection of all $B(K,U)$ with $K \subset X$ compact and $U \subset Y$ open. In other words, a set $V \subset \mathcal{C}(X,Y)$ is open if for all $f \in V$ there exist $K_1, \ldots, K_n \subset X$ compact and $U_1, \ldots, U_n \subset Y$ open such that

$$f \in B(K_1, U_1) \cap \cdots \cap B(K_n, U_n) \subset V.$$

23.3. Metrics

If $(Y, \mathfrak{d})$ is a metric space, then it is also natural to try to topologize $\mathcal{C}(X,Y)$ using $\mathfrak{d}$. This is easiest for X compact, in which case we can define a metric $\mathfrak{D}$ on $\mathcal{C}(X,Y)$ by letting

$$\mathfrak{D}(f,g) = \max\{\mathfrak{d}(f(x_1), f(x_1)) \mid x_1, x_2 \in X\} \quad \text{for } f, g\colon X \to Y. \tag{23.3.1}$$

This makes sense since X is compact, which implies that $f(X)$ and $g(X)$ are compact subsets of the metric space Y and thus that the above maximum is finite and attained. We have:

LEMMA 23.3.1. *Let X be a compact space and let $(Y, \mathfrak{d})$ be a metric space. The compact-open topology on $\mathcal{C}(X,Y)$ and the metric topology on $\mathcal{C}(X,Y)$ coming from (23.3.1) are the same.*

PROOF. We divide the proof into two steps:

STEP 1. *Every open set in the compact-open topology is open in the metric topology.*

Let $K \subset X$ be compact and $U \subset Y$ be open. We must prove that $B(K,U)$ is open in the metric topology. Indeed, consider $f \in B(K,U)$, so $f(K) \subset U$. Since $f(K)$ is a compact subset of U, we can find some $\epsilon > 0$ such that the ϵ-neighborhood of $f(K)$ is contained in U. For $g \in \mathcal{C}(X,Y)$ with $\mathfrak{D}(f,g) < \epsilon$, since $\mathfrak{d}(g(k), f(k)) < \epsilon$ for all $k \in K$ it follows that $g(K)$ is contained in the ϵ-neighborhood of $f(K)$. We thus have $g(K) \subset U$, so $g \in B(K,U)$. We conclude that the ϵ-ball around f is contained in $B(K,U)$, so $B(K,U)$ is open in the metric topology.

[1]It is also common to call this space Y^X, but we think the notation $\mathcal{C}(X,Y)$ is easier to understand.

STEP 2. *Every open set in the metric topology is open in the compact-open topology.*

Let $f \in \mathcal{C}(X,Y)$ and let $\epsilon > 0$. Let

$$B_\epsilon(f) = \{g \in \mathcal{C}(X,Y) \mid d(g(x), f(x)) < \epsilon \text{ for all } x \in X\}$$

be the open ball around f in the metric topology. It is enough to find compact sets $K_1, \ldots, K_n \subset X$ and open sets $U_1, \ldots, U_n \subset Y$ such that

$$f \in B(K_1, U_1) \cap \cdots \cap B(K_n, U_n) \subset B_\epsilon(f).$$

Since $f(X)$ is a compact subset of Y, we can find $x_1, \ldots, x_n \in X$ such that

$$f(X) \subset B_{\epsilon/3}(f(x_1)) \cup \cdots \cup B_{\epsilon/3}(f(x_n)). \tag{23.3.2}$$

For $1 \leq i \leq n$, let $K_i = f^{-1}(\overline{B_{\epsilon/3}(f(x_i))})$ and $U_i = B_{\epsilon/2}(f(x_i))$. Since K_i is a closed subset of the compact space X, it follows that K_i is closed. By (23.3.2), the sets K_i cover X. Finally, by construction

$$f \in B(K_1, U_1) \cap \cdots \cap B(K_n, U_n).$$

Now consider some $g \in B(K_1, U_1) \cap \cdots \cap B(K_n, U_n)$. We must prove that $g \in B_\epsilon(f)$. In other words, letting $x \in X$ we must prove that $\mathfrak{d}(f(x), g(x)) < \epsilon$. We have $x \in K_i$ for some $1 \leq i \leq n$, so $f(x), g(x) \in U_i$. It follows that $\mathfrak{d}(f(x), g(x))$ is at most the diameter ϵ of $U_i = B_{\epsilon/2}(f(x_i))$. □

REMARK 23.3.2. If $(Y, \mathfrak{d})$ is a metric space but X is not compact, then the metric $\mathfrak{d}$ induces a topology on $\mathcal{C}(X,Y)$ as follows. For $f \in \mathcal{C}(X,Y)$ and a compact subset $K \subset X$ and $\epsilon > 0$, let

$$B(f, K, \epsilon) = \{g \in \mathcal{C}(X,Y) \mid \mathfrak{d}(f(x), g(x)) < \epsilon \text{ for all } x, y \in K\}.$$

These sets form the basis for a topology on $\mathcal{C}(X,Y)$ called the *topology of compact convergence*, and this is the same as the compact-open topology (see Exercise 23.1). □

23.4. Composition

For spaces X and Y and Z, there is a composition map $\mathfrak{c}\colon \mathcal{C}(Y,Z) \times \mathcal{C}(X,Y) \to \mathcal{C}(X,Z)$ defined by $\mathfrak{c}(g,f) = g \circ f$ for $g \in \mathcal{C}(Y,Z)$ and $f \in \mathcal{C}(X,Z)$. It is natural to hope that this is continuous. Unfortunately, this does not hold in general. However, it does hold if Y is locally compact:

LEMMA 23.4.1. *Let X and Y and Z be spaces with Y locally compact. Then the composition map $\mathfrak{c}\colon \mathcal{C}(Y,Z) \times \mathcal{C}(X,Y) \to \mathcal{C}(X,Z)$ is continuous.*

PROOF. Let $K \subset X$ be compact and $U \subset Z$ be open. We must prove that $\mathfrak{c}^{-1}(B(K,U))$ is open. Let $(g,f) \in \mathcal{C}(Y,Z) \times \mathcal{C}(X,Y)$ satisfy $\mathfrak{c}(g,f) \in B(K,U)$. It is enough to find an open neighborhood of (g,f) that is mapped by $\mathfrak{c}$ into $B(K,U)$. Since $g \circ f \in B(K,U)$, we have $f(K) \subset g^{-1}(U)$. Since $f(K)$ is a compact subset of the open subset $g^{-1}(U) \subset Y$ and Y is locally compact, there is a compact neighborhood L of $f(K)$ with $L \subset g^{-1}(U)$ (see Exercise 18.3). It follows that $\mathfrak{c}$ takes the open neighborhood $B(L,U) \times B(K, g^{-1}(U))$ of (g,f) into $B(K,U)$, as desired. □

23.5. Evaluation

For spaces X and Y, there is an evaluation map $\mathfrak{e}\colon \mathcal{C}(X,Y) \times X \to Y$ defined by $\mathfrak{e}(f,x) = f(x)$ for $f \in \mathcal{C}(X,y)$ and $x \in X$. Just like for the composition map, to ensure this is continuous we need to assume that X is locally compact:

LEMMA 23.5.1. *Let X and Y be spaces with X locally compact. Then the evaluation map $\mathfrak{e}\colon \mathcal{C}(X,Y) \times X \to Y$ is continuous.*

PROOF. Let p_0 be a one-point space. We have $\mathcal{C}(p_0, X) = X$ and $\mathcal{C}(p_0, Y) = Y$. Applying these identities, the evaluation map becomes the composition map $\mathcal{C}(X,Y) \times \mathcal{C}(p_0, X) \to \mathcal{C}(p_0, Y)$, which is continuous by Lemma 23.4.1. □

23.6. Parameterized maps

Let X and Y and Z be spaces. It is natural to expect maps $\phi\colon X \times Z \to Y$ and $\Phi\colon Z \to \mathcal{C}(X, Y)$ to be closely related. Indeed, if we were working with sets rather than spaces then such maps would be in bijection with each other: a map $\Phi\colon Z \to \mathcal{C}(X, Y)$ would correspond to the map $\phi\colon X \times Z \to Y$ defined by $\phi(x, z) = \Phi(z)(x)$. The following shows that this holds topologically if X is locally compact:

LEMMA 23.6.1. *Let X and Y and Z be spaces. The following holds:*

(i) Let $\phi\colon X \times Z \to Y$ be continuous. Define $\Phi\colon Z \to \mathcal{C}(X, Y)$ to be the map that takes $z \in Z$ to the map $X \to Y$ taking $x \in X$ to $\phi(x, z) \in Y$. Then Φ is continuous.

(ii) Assume that X is locally compact. Let $\Psi\colon Z \to \mathcal{C}(X, Y)$ be continuous. Define $\psi\colon X \times Z \to Y$ to be the map taking $(x, z) \in X \times Z$ to $\Psi(z)(x) \in Y$. Then ψ is continuous.

PROOF. For (i), let $\phi\colon X \times Z \to Y$ be continuous and define $\Phi\colon Z \to \mathcal{C}(X, Y)$ as in (i). Let $K \subset X$ be compact and $U \subset Y$ be open. We must prove that $\Phi^{-1}(B(K, U)) \subset Z$ is open. Let $z_0 \in \Phi^{-1}(B(K, U))$, so $K \times z_0 \subset \phi^{-1}(U)$. Since $K \subset X$ is compact and $\phi^{-1}(U)$ is an open neighborhood of $K \times z_0$, Exercise 17.8 (the "tube lemma") gives an open neighborhood $V \subset Z$ of z_0 with $K \times V \subset \phi^{-1}(U)$. It follows that V is an open neighborhood of z_0 with $V \subset \Phi^{-1}(B(K, U))$, as desired.

We now prove (ii). Assume that X is locally compact and that $\Psi\colon Z \to \mathcal{C}(X, Y)$ is continuous. The map $\psi\colon X \times Z \to Y$ defined in (ii) is the composition

$$X \times Z \xrightarrow{\mathbb{1} \times \Psi} X \times \mathcal{C}(X, Y) \xrightarrow{\mathfrak{e}} Y,$$

where $\mathfrak{e}\colon X \times \mathcal{C}(X, Y) \to Y$ is the evaluation map $\mathfrak{e}(x, f) = f(x)$. Lemma 23.5.1 implies that $\mathfrak{e}$ is continuous, so we conclude that ψ is continuous. $\square$

23.7. Homotopies and the compact-open topology

Let $f_0, f_1\colon X \to Y$ be maps. Recall that a homotopy from f_0 to f_1 is a continuous map $H\colon X \times I \to Y$ with $H(x, 0) = f_0(x)$ and $H(x, 1) = f_1(x)$ for all $x \in X$. Lemma 23.6.1 implies that such a homotopy gives a map $h\colon I \to \mathcal{C}(X, Y)$. This map h can be viewed as a path from $h(0) = f_0$ to $h(1) = f_1$. Conversely, if X is locally compact then Lemma 23.6.1 implies that a path in $\mathcal{C}(X, Y)$ from f_0 to f_1 gives a homotopy from f_0 to f_1.

23.8. Quotient maps and the compact-open topology

As an application of our results, we give another proof of the following result from §21.3:

LEMMA 21.3.4. *Let $q\colon X \to Y$ be a quotient map and let Z be a locally compact space. Then the map $q \times \mathbb{1}\colon X \times Z \to Y \times Z$ is a quotient map.*

PROOF. As we discussed in §13.4, the quotient map $q\colon X \to Y$ satisfies the following universal property. Let $\sim$ be the equivalence relation on X where $x_1 \sim x_2$ if $q(x_1) = q(x_2)$. A map $F\colon X \to W$ is $\sim$-invariant if $F(x_1) = F(x_2)$ whenever $x_1 \sim x_2$. For all spaces W, the following holds:

- There is a bijection between maps $f\colon Y \to W$ and $\sim$-invariant maps $F\colon X \to W$ taking $f\colon Y \to W$ to $f \circ q\colon X \to W$.

In fact, this universal properties characterizes the quotient topology (see Exercise 13.8). We must therefore verify the analogue of it for the map $q \times \mathbb{1}\colon X \times Z \to Y \times Z$.

Consider a space W and a map $G\colon X \times Z \to W$ that is $\sim$-invariant in the sense that $G(x_1, z) = G(x_2, z)$ for all $x_1, x_2 \in X$ and $z \in Z$ with $x_1 \sim x_2$. We must construct a map $g\colon Y \times Z \to W$ such that $G = g \circ (q \times \mathbb{1})$. Let $F\colon X \to \mathcal{C}(Z, W)$ be the map defined by

$$F(x)(z) = G(x, z) \quad \text{for all } x \in X \text{ and } z \in Z.$$

By Lemma 23.6.1, the map F is continuous. Since G is $\sim$-invariant, so is F. It follows that there is a map $f\colon Y \to \mathcal{C}(Z, W)$ with $F = f \circ q$. Let $g\colon Y \times Z \to W$ be the map defined by

$$g(y, z) = f(y)(z) \quad \text{for all } y \in Y \text{ and } z \in Z.$$

Since Z is locally compact, Lemma 23.6.1 says that g is continuous. By construction we have $G = g \circ (q \times \mathbb{1})$, as desired. $\square$

23.9. Parameterized maps, II

Let X and Y and Z be spaces with X locally compact. Lemma 23.6.1 gives a bijection between $\mathcal{C}(X \times Z, Y)$ and $\mathcal{C}(Z, \mathcal{C}(X, Y))$. The following lemma says that this bijection is a homeomorphism if X and Z are Hausdorff:

Lemma 23.9.1. *Let X and Y and Z be spaces with X locally compact Hausdorff and Z Hausdorff. Let $\lambda\colon \mathcal{C}(X \times Z, Y) \to \mathcal{C}(Z, \mathcal{C}(X, Y))$ be the map taking $\phi\colon X \times Z \to Y$ to the map $\Phi\colon Z \to \mathcal{C}(X, Y)$ defined by*

$$\Phi(z)(x) = \phi(x, z) \in Y \quad \text{for all } z \in Z \text{ and } x \in X.$$

Then λ is a homeomorphism.

Proof. Lemma 23.6.1 says that λ is a bijection. For $K \subset X$ and $L \subset Z$ compact and $U \subset Y$ open the map λ restricts to a bijection between $B(K \times L, U)$ and $B(L, B(K, U))$. To prove the lemma, it is enough to prove that open sets of these forms are subbases for the topologies on $\mathcal{C}(X \times Z, Y)$ and $\mathcal{C}(Z, \mathcal{C}(X, Y))$:

- For $\mathcal{C}(X \times Z, Y)$, we prove this in Lemma 23.9.2 below.
- For $\mathcal{C}(Z, \mathcal{C}(X, Y))$, in Lemma 23.9.3 below we prove more generally that if $\mathcal{B}$ is any subbasis for the topology on a space W, then sets of the form $B(L, V)$ with $L \subset Z$ compact and $V \in \mathcal{B}$ form a subbasis for $\mathcal{C}(Z, W)$. $\square$

The above proof used the following two results:

Lemma 23.9.2. *Let X and Y and Z be spaces with X and Z Hausdorff. Then the set of all $B(K \times L, U)$ with $K \subset X$ compact and $L \subset Z$ compact and $U \subset Y$ open forms a subbasis for the compact-open topology on $\mathcal{C}(X \times Z, Y)$.*

Proof. Let $C \subset X \times Z$ be compact and $U \subset Y$ be open. We must prove that $B(C, U)$ is open in the topology with the indicated subbasis. Consider $f \in B(C, U)$. It is enough to find $K_1, \ldots, K_n \subset X$ compact and $L_1, \ldots, L_n \subset Z$ compact such that

$$f \in B(K_1 \times L_1, U) \cap \cdots \cap B(K_n \times L_n, U) \subset B(C, U).$$

Unwrapping this, we need the K_i and L_i to satisfy the following:

- $C \subset \cup_{i=1}^n K_i \times L_i$; and
- $K_i \times L_i \subset f^{-1}(U)$ for all $1 \leq i \leq n$.

Let $C(X) \subset X$ and $C(Z) \subset Z$ be the projections of $C \subset X \times Z$. Both $C(X)$ and $C(Z)$ are compact Hausdorff spaces, and $C \subset C(X) \times C(Z)$. Replacing X with $C(X)$ and Z with $C(Z)$, we can therefore assume without loss of generality that X and Z are compact Hausdorff spaces. The space $X \times Z$ is thus also a compact Hausdorff space, and in particular is normal (see Lemma 17.2.3).

The set $f^{-1}(U)$ is an open neighborhood of C. Since $X \times Z$ is normal, for each $c \in C$ we can find open sets $V_c \subset X$ and $W_c \subset Z$ such that $c \in V_c \times W_c$ and $\overline{V}_c \times \overline{W}_c \subset f^{-1}(U)$. Since C is compact, we can find $c_1, \ldots, c_n$ such that $C \subset \cup_{i=1}^n V_{c_i} \times W_{c_i}$. Let $K_i = \overline{V}_{c_i} \subset X$ and $L_i = \overline{W}_{c_i} \subset Z$, so $K_i \times L_i \subset f^{-1}(U)$. Since X and Z are compact, the closed sets K_i and L_i are also compact. By construction we have $C \subset \cup_{i=1}^n K_i \times L_i$, as desired. $\square$

Lemma 23.9.3. *Let Z and W be spaces with Z Hausdorff and let $\mathcal{B}$ be a subbasis for the topology on W. Then the set of all $B(K, V)$ with $K \subset Z$ compact and $V \in \mathcal{B}$ forms a subbasis for the compact-open topology on $\mathcal{C}(Z, W)$.*

Proof. See Exercise 23.2. $\square$

Remark 23.9.4. It is a little annoying that the above results require local compactness. Unfortunately, they are false in general. There is a way around this using the theory of compactly generated spaces, which we describe in Essay I. $\square$

23.10. Exercises

EXERCISE 23.1. Let X be a space and let $(Y, \mathfrak{d})$ be a metric space. For $f \in Y^X$ and a compact subset $K \subset X$ and $\epsilon > 0$, let

$$B(f, K, \epsilon) = \left\{ g \in Y^X \mid \mathfrak{d}(f(x), g(x)) < \epsilon \text{ for all } x, y \in K \right\}.$$

Prove that these sets form the basis for a topology on Y^X, and this topology is the same as the compact-open topology. □

EXERCISE 23.2. Let Z and W be spaces with Z Hausdorff and let $\mathcal{B}$ be a subbasis for the topology on W. Prove that the set of all $B(K, V)$ with $K \subset Z$ compact and $V \in \mathcal{B}$ forms a subbasis for the compact-open topology on $\mathcal{C}(Z, W)$. □

EXERCISE 23.3. Topologize $\mathrm{GL}_n(\mathbb{R})$ by identifying it as a subspace of $\mathrm{Mat}_n(\mathbb{R}) \cong \mathbb{R}^{n^2}$. For each $M \in \mathrm{GL}_n(\mathbb{R})$, multiplication by M gives a linear map $\phi_M \colon \mathbb{R}^n \to \mathbb{R}^n$. Prove that the map $\iota\colon \mathrm{GL}_n(\mathbb{R}) \to \mathcal{C}(\mathbb{R}^n, \mathbb{R}^n)$ defined by $\iota(M) = \phi_M$ is a closed embedding. □

EXERCISE 23.4. Let $(M, \mathfrak{d})$ be a compact metric space. An isometry of M is a bijection $f \colon M \to M$ such that $\mathfrak{d}(f(p), f(q)) = \mathfrak{d}(p, q)$ for all $p, q \in M$. Let $\mathrm{Isom}(M)$ be the group of isometries of M. Topologize $\mathrm{Isom}(M)$ using the compact-open topology, i.e., by identifying $\mathrm{Isom}(M)$ with a subspace of $\mathcal{C}(M, M)$. Prove that $\mathrm{Isom}(M)$ is compact. □

EXERCISE 23.5. Let $\mathrm{Homeo}(\mathbb{S}^1)$ be the set of homeomorphisms $f \colon \mathbb{S}^1 \to \mathbb{S}^1$. Topologize $\mathrm{Homeo}(\mathbb{S}^1)$ using the compact-open topology, i.e., by identifying $\mathrm{Homeo}(\mathbb{S}^1)$ with a subspace of $\mathcal{C}(\mathbb{S}^1, \mathbb{S}^1)$. Prove that $\mathrm{Homeo}(\mathbb{S}^1)$ is not locally compact. □

Part 2

Essays

Essays on manifolds and related topics

ESSAY A

Topological manifolds

In this essay, we use the tools we have developed to study manifolds, which are perhaps the most important class of spaces in topology.

A.1. Basic definitions

An *n-dimensional manifold* (or simply an n-manifold) is a second countable Hausdorff space M^n that is locally homeomorphic to $\mathbb{R}^n$ in the following sense:

- For all $p \in M^n$, there exists an open neighborhood U of p that is homeomorphic to an open subset of $\mathbb{R}^n$.

A *chart* on M^n is a homeomorphism $\phi\colon U \to V$ with $U \subset M^n$ and $V \subset \mathbb{R}^n$ open sets. If U is an open neighborhood of $p \in M^n$, we call this chart $\phi\colon U \to V$ a chart around p. An *atlas* for M^n is a collection of charts $\{\phi_i\colon U_i \to V_i\}_{i \in I}$ such that the U_i cover M^n.

Here are several basic examples:

Example A.1.1. The whole space $\mathbb{R}^n$ is an n-manifold with an atlas consisting of a single chart $\mathbb{1}\colon \mathbb{R}^n \to \mathbb{R}^n$. More generally, an open set $U \subset \mathbb{R}^n$ is an n-manifold, again with an atlas consisting of a single chart $\mathbb{1}\colon U \to U$. □

Example A.1.2. More generally, if M^n is an n-manifold and $W \subset M^n$ is open, then W is an n-manifold. Indeed, for $p \in W$ let $\phi\colon U \to V$ be a chart around p for M^n. Letting $U' = U \cap W$ and $V' = \phi(U')$, the homeomorphism $\phi|_{U'}\colon U' \to V'$ is a chart around p for W. □

Example A.1.3. Let $\mathbb{S}^n$ be the n-sphere, so

$$\mathbb{S}^n = \left\{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} \mid x_1^2 + \cdots + x_{n+1}^2 = 1\right\}.$$

This is an n-manifold. Indeed, for $1 \leq k \leq n+1$ let

$$U_{x_k>0} = \{(x_1, \ldots, x_{n+1}) \in \mathbb{S}^n \mid x_k > 0\},$$
$$U_{x_k<0} = \{(x_1, \ldots, x_{n+1}) \in \mathbb{S}^n \mid x_k < 0\}.$$

Letting $B = B_1(0) \subset \mathbb{R}^n$ be the open unit ball, we have homeomorphisms $\phi_{x_k>0}\colon U_{x_k>0} \to B$ and $\phi_{x_k<0}\colon U_{x_k<0} \to B$ taking a point $(x_1, \ldots, x_{n+1})$ to $(x_1, \ldots, \widehat{x_k}, \ldots, x_{n+1}) \in B$, where the hat in $\widehat{x_j}$ indicates that this coordinate is being omitted. The set

$$\{\phi_{x_k>0}\colon U_{x_k>0} \to B,\ \phi_{x_k<0}\colon U_{x_k<0} \to B \mid 1 \leq k \leq n+1\}$$

is an atlas for $\mathbb{S}^n$. □

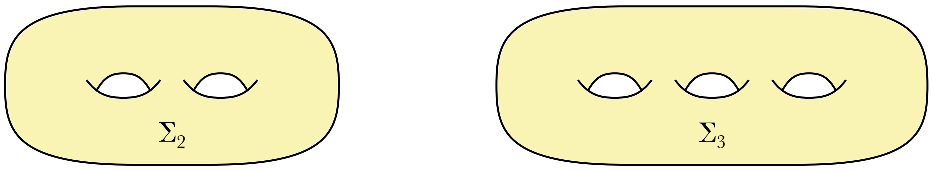
Example A.1.4. Let $\mathbb{RP}^n$ be the set of lines though the origin in $\mathbb{R}^{n+1}$. There is a projection map $q\colon \mathbb{R}^{n+1} \setminus 0 \to \mathbb{RP}^n$ taking $x \in \mathbb{R}^{n+1} \setminus 0$ to the line through 0 and x. We endow $\mathbb{RP}^n$ with the quotient topology from this projection, so $U \subset \mathbb{RP}^n$ is open if and only if $q^{-1}(U) \subset \mathbb{R}^{n+1} \setminus 0$ is open. The space $\mathbb{RP}^n$ is known as the n-dimensional real projective space. As notation, for $(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} \setminus 0$ we write $[x_1, \ldots, x_{n+1}]$ for the corresponding point of $\mathbb{RP}^n$, so for $\lambda \in \mathbb{R}$ nonzero we have $[\lambda x_1, \ldots, \lambda x_{n+1}] = [x_1, \ldots, x_{n+1}]$.

The space $\mathbb{RP}^n$ is an n-manifold. Unlike our previous examples, it is not totally obvious that it is second countable and Hausdorff, so we leave this as an exercise (Exercise A.1). We prove it is locally Euclidean by exhibiting an atlas as follows. For $1 \leq k \leq n+1$, let $U_k = \{[x_1, \ldots, x_{n+1}] \in \mathbb{RP}^n \mid x_k \neq 0\}$. This set is well-defined, and the map $\phi_k\colon U_k \to \mathbb{R}^n$ defined by

$$\phi_k([x_1, \ldots, x_{n+1}]) = (x_1/x_k, \ldots, \widehat{x_k/x_k}, \ldots, x_{n+1}/x_k) \quad \text{for } [x_1, \ldots, x_{n+1}] \in \mathbb{RP}^n$$

is a well-defined homeomorphism (see Exercise A.1). The set $\{\phi_k \colon U_k \to \mathbb{R}^n \mid 1 \leq k \leq n+1\}$ is an atlas for $\mathbb{RP}^n$. $\square$

Remark A.1.5. It is clear that the only connected 0-manifold is a single point. It turns out that $\mathbb{R}$ and $\mathbb{S}^1$ are the only connected 1-manifolds. There is also a very beautiful classification of compact connected 2-manifolds. Here are two examples of such 2-manifolds:



We describe the classification of 2-manifolds in Essay C. The exercises in that essay also outline a proof of the classification of 1-manifolds. In higher dimensions, things are much more complicated. $\square$

Remark A.1.6. The requirement that manifolds be second countable and Hausdorff is needed to rule out various pathological examples. Without them, even 1-manifolds would not have a simple classification. We describe some of these pathological examples later in this chapter. $\square$

A.2. Basic properties

The following summarizes some of the basic point-set topological properties of manifolds:

Lemma A.2.1. *Let M^n be an n-manifold. Then:*

- *M^n is perfectly normal.*
- *M^n is locally compact.*
- *M^n is paracompact.*
- *M^n is metrizable.*
- *M^n is locally path connected, so its path components and connected components coincide and are clopen.*

Proof. Since M^n is locally homeomorphic to $\mathbb{R}^n$, the fact that M^n is locally compact and locally path connected follows immediately from the fact that $\mathbb{R}^n$ is locally compact and locally path connected. Since M^n is second countable, Hausdorff, and locally compact, it follows that M^n is paracompact (see Corollary 20.2.2). This implies that M^n is normal (see Lemma 20.4.1). We can therefore use the Urysohn metrization theorem (Theorem 22.2.1) to see that M^n is metrizable and hence perfectly normal (Lemma 16.7.1). $\square$

Remark A.2.2. One basic property of manifolds we do not list above is that their dimension is well-defined. In fact, it is true that if M is both an n-manifold and an m-manifold then $n = m$, but this is a difficult theorem called the *invariance of domain*. The most natural proof of invariance of domain uses homology. $\square$

A.3. Embedding manifolds into Euclidean space

Many n-manifolds are constructed as subspaces of some $\mathbb{R}^d$, but some manifolds like $\mathbb{RP}^n$ do not have obvious embeddings into any Euclidean space. However, it turns out that all manifolds can be embedded in some $\mathbb{R}^d$:

Theorem A.3.1. *Let M^n be an n-manifold. Then for some $N \gg 0$ there exists a proper embedding $\iota \colon M^n \hookrightarrow \mathbb{R}^N$. Since ι is proper it is a closed map, so this implies that M^n is homeomorphic to a closed subspace of $\mathbb{R}^N$.*

In fact, in Essay D we will use dimension theory to embed M^n into $\mathbb{R}^{2n+1}$. We include the proof Theorem A.3.1 here both because it is easier than the dimension theory proof and because its details will be useful in a later volume when we study smooth manifolds.

Proof of Theorem A.3.1. The first step is to show that M^n has a finite atlas

$$\{\phi_k \colon U_k \to V_k \mid 1 \leq k \leq m\}$$

such that each $\phi_k\colon U_k \to V_k$ extends to a homeomorphism $\overline{\phi}_k\colon \overline{U}_k \to \overline{V}_k$. This is obvious if M^n is compact, while for M^n non-compact it follows from Lemma A.4.1 below. This might seem paradoxical if M^n is non-compact, but note that we are not requiring the U_k to be connected. In fact, our construction in Lemma A.4.1 will in general give an atlas in which each U_k has infinitely many path components.

Since M^n is paracompact, there is a strong refinement of the cover $\{U_1, \ldots, U_m\}$, i.e., a cover $\{W_1, \ldots, W_m\}$ such that $\overline{W}_k \subset U_k$ for each $1 \leq k \leq m$ (see Lemma 20.5.1). Since M^n is perfectly normal, for each $1 \leq k \leq m$ we can find a continuous map $f_k\colon X \to I$ such that $f_k^{-1}(1) = \overline{W}_k$ and $\operatorname{supp}(f_k) \subset U_k$; see §16.7. Multiplying ϕ_k by f_k, we get a map $f_k\phi_k\colon U_k \to \mathbb{R}^n$. Since $\operatorname{supp}(f_k) \subset U_k$, we can extend $f_k\phi_k\colon U_k \to \mathbb{R}^n$ to a continuous map $G_k\colon M^n \to \mathbb{R}^n$ with $G_k(p) = 0$ for $p \notin U_k$. Let $\iota\colon M^n \to \mathbb{R}^{nm+m}$ be the map defined by

$$\iota(p) = (G_1(p), f_1(p), \ldots, G_m(p), f_m(p)) \in (\mathbb{R}^n \times \mathbb{R}^1)^{\times m} = \mathbb{R}^{nm+m} \quad \text{for } p \in M^n.$$

We must prove that ι is injective and proper:

Step 1. *The map ι is injective.*

Consider $p, q \in M^n$ with $\iota(p) = \iota(q)$. Pick $1 \leq k \leq m$ with $p \in \overline{W}_k$. We thus have $f_k(p) = 1$, so since $\iota(p) = \iota(q)$ we also have $f_k(q) = 1$. Since $f_k^{-1}(1) = \overline{W}_k$, it follows that $q \in \overline{W}_k$. On $\overline{W}_k$ we have $G_k = \phi_k$, so since $p, q \in \overline{w}_k$ and $\iota(p) = \iota(q)$ we have $\phi_k(p) = \phi_k(q)$. Since $\phi_k\colon U_k \to V_k$ is a homeomorphism and $\overline{W}_k \subset U_k$, we conclude that $p = q$.

Step 2. *The map ι is a proper map. We remark that this is trivial if M^n is compact, so this step is only needed in the non-compact case.*

Let $L \subset \mathbb{R}^{nm+m}$ be compact. Applying Lemma 19.2.2, we must prove that $\iota^{-1}(L)$ is compact. Since

$$\iota^{-1}(L) = \left(\iota^{-1}(L) \cap \overline{W}_1\right) \cup \cdots \cup \left(\iota^{-1}(L) \cap \overline{W}_m\right),$$

it is enough to prove that each $\iota^{-1}(L) \cap \overline{W}_k = (\iota|_{\overline{W}_k})^{-1}(L)$ is compact. We will prove this for $k = 1$; the other cases are identical up to changes in notation. Let $\pi\colon \mathbb{R}^{nm+m} \to \mathbb{R}^n$ be the projection onto the first n coordinates and let $L' = \pi(L)$, so $L' \subset \mathbb{R}^n$ is compact. We have

$$(\iota|_{\overline{W}_1})^{-1}(L) \subset (\pi \circ \iota|_{\overline{W}_1})^{-1}(L').$$

Since $(\iota|_{\overline{W}_1})^{-1}(L)$ is a closed subspace of $\overline{W}_1$, we see that it is enough to prove that $(\pi \circ \iota|_{\overline{W}_1})^{-1}(L')$ is compact. On $\overline{W}_1$, we have $f_1 = 1$ and $G_1 = \phi_1$. It follows that $\pi \circ \iota|_{\overline{W}_1} = \phi_1|_{\overline{W}_1}$, so what we must prove is that $(\phi_1|_{\overline{W}_1})^{-1}(L')$ is compact. We assumed at the beginning of the proof that the homeomorphism $\phi_1\colon U_1 \to V_1$ extends to a homeomorphism $\overline{\phi}_1\colon \overline{U}_1 \to \overline{V}_1$. The closed subspace $\overline{V}_1 \cap L'$ of the compact subspace L' is compact, so $\overline{\phi}_1^{-1}(L')$ is compact. Since $(\phi_1|_{\overline{W}_1})^{-1}(L') = \overline{\phi}_1^{-1}(L') \cap \overline{W}_1$, we conclude that $(\phi_1|_{\overline{W}_1})^{-1}(L')$ is compact, as desired. □

A.4. Finite atlases for noncompact manifolds

At least in the non-compact case, the above proof of Theorem A.3.1 depended on the following:

Lemma A.4.1. *All n-manifolds M^n have finite atlases $\{\phi_k\colon U_k \to V_k \mid 1 \leq k \leq m\}$ such that each homeomorphism $\phi_k\colon U_k \to V_k$ extends to a homeomorphism $\overline{\phi}_k\colon \overline{U}_k \to \overline{V}_k$.*

Proof. Since M^n is second countable, we can find an atlas $\{\phi_k\colon U_k \to V_k \mid k \geq 1\}$ such that each homeomorphism $\phi_k\colon U_k \to V_k$ extends to a homeomorphism $\overline{\phi}_k\colon \overline{U}_k \to \overline{V}_k$ and each $\overline{V}_k$ is a bounded subset of $\mathbb{R}^n$. We now invoke a result from dimension theory we will prove in Essay D below:

- We can shrink each U_k to a possibly smaller open set and ensure that for each $p \in M^n$ the set $\{k \mid p \in U_k\}$ has cardinality at most[1] $n+1$. In the terminology we will introduce in Essay D, this means that the cover $\{U_k \mid k \geq 1\}$ has order at most $n+1$.

[1] Of course, this reflects the fact that M^n is n-dimensional. We remark that the precise constant $n+1$ will not matter for our proof here, and could be replaced by any N that does not depend on the point $p \in M^n$.

Since M^n is paracompact, there is a partition of unity subordinate to the cover $\{U_k \mid k \geq 1\}$, i.e., a collection of continuous functions $\{f_k \colon M^n \to I \mid k \geq 1\}$ such that $\operatorname{supp}(f_k) \subset U_k$ for each $k \geq 1$ and such that for all $p \in M^n$ we have

$$\sum_{k=1}^{\infty} f_k(p) = 1.$$

Since each p lies in at most $n+1$ of the U_k, all but finitely many terms in this sum are 0. For each unordered tuple $\{k_1, \ldots, k_d\}$ of distinct positive integers, let

$$U_{k_1,\ldots,k_d} = \left\{p \in M^n \mid f_i(p) < f_{k_j}(p) \text{ for all } i \geq 1 \text{ and } 1 \leq j \leq d \text{ with } i \notin \{k_1, \ldots, k_d\}\right\}.$$

These are open sets (see Exercise A.2). Since the inequality in the definition of $U_{k_1,\ldots,k_d}$ is strict, it follows that for $p \in U_{k_1,\ldots,k_d}$ we have $f_{k_j}(p) > 0$ for all $1 \leq j \leq d$. We thus have

$$U_{k_1,\ldots,k_d} \subset \bigcap_{j=1}^{d} U_{k_j}.$$

Each ϕ_{k_j} therefore restricts to a homeomorphism between $U_{k_1,\ldots,k_d}$ and an open set in $\mathbb{R}^n$. Let $V_{k_1,\ldots,k_d}$ be the image of one of these restrictions and let $\phi_{k_1,\ldots,k_d} \colon U_{k_1,\ldots,k_d} \to V_{k_1,\ldots,k_d}$ be the restricted homeomorphism. From the ϕ_{k_j} these inherit the property that they extend to homeomorphisms $\overline{\phi}_{k_1,\ldots,k_d} \colon \overline{U}_{k_1,\ldots,k_d} \to \overline{V}_{k_1,\ldots,k_d}$ and that $\overline{V}_{k_1,\ldots,k_d}$ is a bounded subset of $\mathbb{R}^n$. We now prove two key properties of the $U_{k_1,\ldots,k_d}$:

Claim 1. *For some $d \geq 1$, let $\{k_1, \ldots, k_d\}$ and $\{\ell_1, \ldots, \ell_d\}$ be distinct unordered d-tuples of positive integers. Then $U_{k_1,\ldots,k_d} \cap U_{\ell_1,\ldots,\ell_d} = \emptyset$.*

Proof of claim. Assume for the sake of contradiction that there is a point p in this intersection. After reodering the k_j and ℓ_j, we can assume that k_1 does not appear in $(\ell_1, \ldots, \ell_d)$ and that ℓ_1 does not appear in $(k_1, \ldots, k_d)$. We then have that $f_{\ell_1}(p) < f_{k_1}(p)$ and that $f_{k_1}(p) < f_{\ell_1}(p)$, a contradiction. □

Claim 2. *The collection of open sets $\{U_{k_1,\ldots,k_d} \mid 1 \leq d \leq n+1 \text{ and } k_1, \ldots, k_d \geq 1\}$ covers M^n.*

Proof of claim. Let $p \in M^n$. By assumption, p lies in at most $n+1$ of the U_k. Enumerating the set $\{k \mid f_k(p) > 0\}$ as $\{k_1, \ldots, k_d\}$, we therefore have $1 \leq d \leq n+1$. Since $f_i(p) = 0$ for all $i \notin \{k_1, \ldots, k_d\}$, it follows that $p \in U_{k_1,\ldots,k_d}$. □

In the rest of the proof, we not need the definition of the $U_{k_1,\ldots,k_d}$ but only the above two claims and the fact that we have homeomorphisms $\phi_{k_1,\ldots,k_d} \colon U_{k_1,\ldots,k_d} \to V_{k_1,\ldots,k_d}$ that extend to homeomorphisms $\overline{\phi}_{k_1,\ldots,k_d} \colon \overline{U}_{k_1,\ldots,k_d} \to \overline{V}_{k_1,\ldots,k_d}$. Since M^n is paracompact, we can shrink the $U_{k_1,\ldots,k_d}$ so that they still cover M^n but now satisfy the following strengthening of Claim 1:

(i) For some $d \geq 1$, let $\{k_1, \ldots, k_d\}$ and $\{\ell_1, \ldots, \ell_d\}$ be distinct unordered d-tuples of positive integers. Then $\overline{U}_{k_1,\ldots,k_d} \cap \overline{U}_{\ell_1,\ldots,\ell_d} = \emptyset$.

Moreover, by moving the bounded sets $\overline{V}_{k_1,\ldots,k_d} \subset \mathbb{R}^n$ around we can ensure that:

(ii) For some $d \geq 1$, let $\{k_1, \ldots, k_d\}$ and $\{\ell_1, \ldots, \ell_d\}$ be distinct unordered d-tuples of positive integers. Then $\overline{V}_{k_1,\ldots,k_d} \cap \overline{V}_{\ell_1,\ldots,\ell_d} = \emptyset$.

For each $1 \leq d \leq n+1$, let $A_d \subset M^n$ be the union of the disjoint sets $\{U_{k_1,\ldots,k_d} \mid k_1, \ldots, k_g \geq 1\}$ and let $B_d \subset \mathbb{R}^n$ be the union of the disjoint sets $\{V_{k_1,\ldots,k_d} \mid k_1, \ldots, k_g \geq 1\}$. The homeomorphisms $\phi_{k_1,\ldots,k_d} \colon U_{k_1,\ldots,k_d} \to V_{k_1,\ldots,k_d}$ assemble into a homeomorphism $\psi_d \colon A_d \to B_d$. By (i) and (ii) above, the homeomorphism ψ_d extends to a homeomorphism $\overline{\psi}_d \colon \overline{A}_d \to \overline{B}_d$. The desired finite atlas is then $\{\psi_d \colon A_d \to B_d \mid 1 \leq d \leq n+1\}$. □

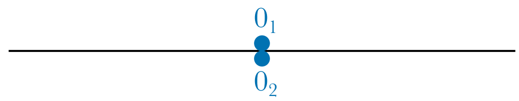
A.5. Non-Hausdorff manifolds

Recall that we require manifolds to be Hausdorff and second countable. Removing these hypotheses gives many exotic generalized manifolds, even in dimension 1. We have already seen one example of a non-Hausdorff 1-manifold, namely the line with two origins from Example 16.1.1. We recall the construction:

EXAMPLE A.5.1. As a set, let $Y = (\mathbb{R} \setminus \{0\}) \sqcup \{0_1, 0_2\}$. For $i = 1, 2$, let $f_i \colon \mathbb{R} \to Y$ be the map defined by $f_i(x) = x$ for $x \in \mathbb{R} \setminus \{0\}$ and $f_i(0) = 0_i$. Give Y the identification space topology, so:

- a set $U \subset Y$ is open if and only if $f_1^{-1}(U)$ and $f_2^{-1}(U)$ are open in $\mathbb{R}$.

Here is a picture of this:



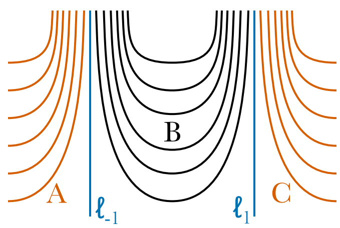
With this topology, the subspaces $Y \setminus \{0_2\} = f_1(\mathbb{R})$ and $Y \setminus \{0_1\} = f_2(\mathbb{R})$ are open subsets of Y that are both homeomorphic to $\mathbb{R}$. It follows that Y a second-countable non-Hausdorff 1-manifold. □

This example might not seem very geometrically interesting. The theory of foliations of the plane gives non-Hausdorff 1-manifolds with a closer connection to geometry. See [**1**] for a beautiful discussion of this. We content ourselves here with one example:

EXAMPLE A.5.2. For $c \in \mathbb{R}$, let $X_c = \left\{(x, y) \mid (x^2 - 1)e^y = c\right\} \subset \mathbb{R}^2$. Define

$$\mathfrak{F} = \{L \mid L \text{ is a connected component of } X_c \text{ for some } c \in \mathbb{R}\}.$$

The set $\mathfrak{F}$ is what is called a foliation of $\mathbb{R}^2$. Each $L \in \mathfrak{L}$ is called a *leaf* of the foliation. Here is a picture of $\mathfrak{F}$:

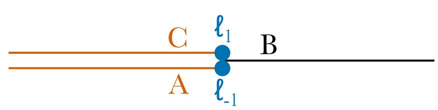


Each leaf L is homeomorphic to $\mathbb{R}$, and $\mathbb{R}^2$ is the disjoint union of the $L \in \mathfrak{F}$. The set X_0 consists of two vertical lines ℓ_{-1} and ℓ_1 where $x = \pm 1$. For $c > 0$, the set X_c consists of two arcs, one lying in the region to the left of ℓ_{-1} labeled A and one lying in the region to the right of ℓ_1 labeled C. For $c < 0$, the set X_c consists of a single arc in the region between ℓ_{-1} and ℓ_1 labeled B.

Let $\mathcal{L}$ be the quotient space of $\mathbb{R}^2$ obtained by collapsing each $L \in \mathfrak{F}$ to a point. This is called the leaf space of the foliation $\mathfrak{F}$. The space $\mathcal{L}$ is a non-Hausdorff 1-manifold. To describe it, let R_1 and R_2 be copies of $\mathbb{R}$. The space $\mathcal{L}$ is obtained by gluing R_1 to R_2 so as to identify each $t \in R_1$ with $t > 0$ with the corresponding $t \in R_2$. The various types of leaves correspond to the following points:

- The points $0 \in R_1$ and $0 \in R_2$ correspond to ℓ_{-1} and ℓ_1.
- The points $t \in R_1$ with $t < 0$ correspond to the arcs in the region A.
- The points $t \in R_2$ with $t < 0$ correspond to the arcs in region C.
- The points $t \in R_1$ and $t \in R_2$ with $t > 0$ that are glued together correspond to the arcs in the region B.

The picture is as follows:



This space is non-Hausdorff since the points corresponding to ℓ_{-1} and ℓ_1 do not have disjoint neighborhoods. You will verify all of this in Exercise A.3. □

A.6. Non-second countable manifolds

The theory of non-second countable manifolds has a set-theoretic flavor. It turns out that in dimension one there is a single example of a connected non-second countable Hausdorff 1-manifold called the *long line* L. We close this chapter with a brief discussion of it. The space L has the following seemingly paradoxical properties:

- L is a path-connected Hausdorff non-second-countable 1-manifold.

- Like $\mathbb{R}$, the points of L are endowed with a total ordering.
- For $x, y \in L$ with $x < y$, the "interval"
$$[x, y] = \{z \in L \mid x \leq z \leq y\}$$
is homeomorphic to the closed interval $I = [0, 1]$. This accounts for L being path connected.
- On the other hand, since L is not second countable it contains uncountably many subspaces homeomorphic to the open interval $(0, 1)$.

Before we can construct L, we need to discuss some more details about well-ordered sets, which we introduced in §21.10 to set up the process of transfinite induction.

A.6.1. Minimal uncountable well-ordered set. Let S be an uncountable set. Pick a well-ordering on S. Let $\mathfrak{C}$ be the set of all initial segments of S that are either finite or countably infinite. The set $\mathfrak{C}$ is nonempty since $\emptyset \in \mathfrak{C}$. In fact, by starting with $\emptyset$ and repeatedly adding the minimal element we have not yet chosen we see that there exists a countably infinite set in $\mathfrak{C}$. As we discussed in §21.10, the initial segments of S are totally ordered under inclusion. Let
$$S_\Omega = \bigcup_{J \in \mathfrak{C}} J.$$
The set S_Ω is an initial segment of S. By construction, all initial segments J with $J \subsetneq S_\Omega$ are countable. We claim that S_Ω is not countable. Indeed, let s_0 be the minimal element of $S \setminus S_\Omega$. The initial segment $S_\Omega \sqcup \{s_0\}$ cannot lie in $\mathfrak{C}$, so $S_\Omega \sqcup \{s_0\}$ is uncountable. This implies that S_Ω is uncountable. The totally ordered set S_Ω is called the *minimal uncountable well-ordered set.*[2] It is unique up to isomorphism, but we will not need this. All we need to know about S_Ω is that it is uncountable but all proper initial segments of S_Ω are finite or countably infinite.

A.6.2. Constructing the long line. Let $\widehat{L} = S_\Omega \times [0, 1)$. Both S_Ω and $[0, 1)$ have total orderings. Give $\widehat{L}$ the dictionary ordering, so $(s, x) \leq (s', x')$ if $s < s'$ or if $s = s'$ and $x < x'$. An *open interval* in $\widehat{L}$ is a set of the form $(\theta_1, \theta_2) = \{\nu \mid \theta_1 < \nu < \theta_2\}$ for some $\theta_1, \theta_2 \in \widehat{L}$ with $\theta_1 < \theta_2$. This is a basis for a topology called the *order topology* (see Example 12.7.3). We endow $\widehat{L}$ with the order topology.

To form the long line L, let $s_0 \in S_\omega$ be the minimal element. It follows that $(s_0, 0) \in \widehat{L}$ is the minimal element of $\widehat{L}$. Define $L = \widehat{L} \setminus \{(s_0, 0)\}$. As you will verify in Exercise A.4, this has the properties claimed in §A.6.

A.7. Exercises

EXERCISE A.1. Prove the following:

(a) The space $\mathbb{RP}^n$ is Hausdorff and second countable.
(b) Letting $U_k = \{[x_1, \ldots, x_{n+1}] \in \mathbb{RP}^n \mid x_k \neq 0\}$, the map $\phi_k \colon U_k \to \mathbb{R}^n$ defined by
$$\phi_k([x_1, \ldots, x_{n+1}]) = (x_1/x_k, \ldots, \widehat{x_k/x_k}, \ldots, x_{n+1}/x_k) \quad \text{for } [x_1, \ldots, x_{n+1}] \in \mathbb{RP}^n$$
is a well-defined homeomorphism. □

EXERCISE A.2. Let M^n be an n-manifold, let $\{U_k \mid k \geq 1\}$ be an open cover of M^n, and let $\{f_k \colon M^n \to I \mid k \geq 1\}$ be a partition of unity subordinate to $\{U_k \mid k \geq 1\}$. For an unordered tuple $\{k_1, \ldots, k_d\}$ of distinct positive integers, let
$$U_{k_1,\ldots,k_d} = \left\{p \in M^n \mid f_i(p) < f_{k_j}(p) \text{ for all } i \geq 1 \text{ and } 1 \leq j \leq d \text{ with } i \notin \{k_1, \ldots, k_d\}\right\}.$$
Prove that $U_{k_1,\ldots,k_d}$ is open. □

EXERCISE A.3. Verify the description of $\mathcal{L}$ in Example A.5.2. □

EXERCISE A.4. Let L be the long line constructed in §A.6.2. Prove the following:

[2] Or the minimal uncountable ordinal, but we have chosen not to use that terminology.

(a) For $x, y \in L$ with $x < y$, the closed interval
$$[x, y] = \{z \in L \mid x \leq z \leq y\}$$
is homeomorphic to the closed interval $I = [0, 1]$.

(b) The space L is path-connected.

(c) The space L contains uncountably many subspaces homeomorphic to the open interval $(0, 1)$.

(d) The space L is a Hausdorff non-second-countable 1-manifold. $\square$

Bibliography

[1] A. Haefliger & G. Reeb, One dimensional non-Hausdorff manifolds and foliations of the plane, in *Geometric methods in group theory—papers dedicated to Ruth Charney*, 225–241, Sémin. Congr., 34, Soc. Math. France, Paris. (Cited on page 183.)

ESSAY B

Classification of 1-manifolds

ESSAY C

Classification of surfaces

C.1. Introduction

An enormous amount of algebraic topology was developed to help classify manifolds up to homeomorphism. This classification is easy in dimensions 0 and 1, where the only connected examples are a point, a circle $\mathbb{S}^1$, and the real line $\mathbb{R}$ (see Exercise C.9). The first interesting dimension is 2, i.e., surfaces. Here there are infinitely many examples, but there is an elegant and easy-to-state classification (at least in the compact case) whose origins go back to 19th century work of Möbius.

C.1.1. Our goal. In this essay, we prove the classification of surfaces. Our goal is to emphasize geometric reasoning. There is a large expository gulf between the geometric topology literature and accounts of the classification of surfaces, which are typically aimed at beginning students and involve elaborate manipulations of triangulations. We include many examples and pictures, but some of our proofs and definitions are a little informal. Making them rigorous will (hopefully) be routine to readers who are experienced with smooth manifolds.

C.1.2. History and sources. The idea of our proof goes back to Zeeman [**13**]. Here are other accounts geared to students earlier in their education:

- See [**9**] or [**12**] for the classical combinatorial proof. I first learned this material from [**9**] when I was an undergraduate.
- See [**2**] for a proof similar to the one we give.

There are other possible proofs of this result. One that is particularly charming is Conway's "ZIP" proof, which can be found in [**4**]. For a history of the classification, see [**5**].

C.1.3. Assumed results. To avoid getting bogged down with foundational results, we will carefully state but not prove two important results:

- The existence of triangulations of surfaces. Actually, we will use the more flexible notion of "polygonal decompositions".
- The fact that the Euler characteristic of a surface is a topological invariant. This result is very easy once the theory of homology is introduced, so we see little point in giving a combinatorial proof that uses special features of surfaces.

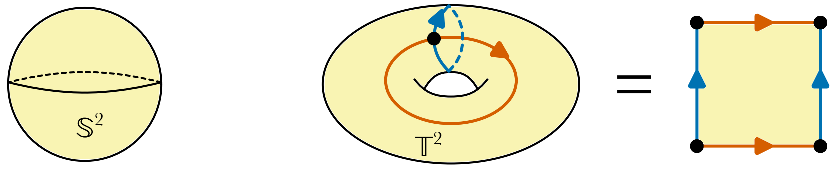
We will also freely use standard results about smooth manifolds, often without mentioning them explicitly.

C.1.4. Outline. In §C.2 we give examples of surfaces and state a first version of the classification theorem. Next, as a warm-up to the proof in §C.3 we discuss graphs and their Euler characteristics. We then introduce polygonal decompositions and prove some basic results about the Euler characteristic in §C.4, which ends with a refined version of the classification. We prove the classification in the next two sections: §C.5 proves the "Poincaré conjecture" characterizing the sphere, and §C.6 proves the rest of the classification. Finally, §C.7 gives some extensions and generalizations of the classification.

C.2. Examples of surfaces

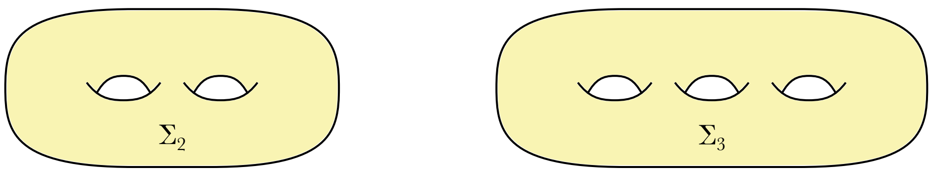
A *surface* is a 2-dimensional manifold, possibly with boundary. Our focus will be on surfaces that are connected and *closed*, that is, compact and without boundary. This section focuses on examples.

C.2.1. Basic examples. The most familiar surfaces are the 2-sphere $\mathbb{S}^2$ and the 2-torus $\mathbb{T}^2 = (\mathbb{S}^1)^{\times 2}$:



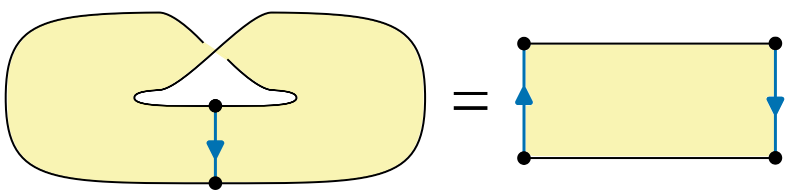
As is shown here, $\mathbb{T}^2$ can be obtained from $\mathbb{D}^2$ by identifying $\mathbb{D}^2$ with a square and identifying parallel sides. The four vertices of the square are all identified to a single point. The sphere $\mathbb{S}^2$ can also be obtained from $\mathbb{D}^2$ by identifying the entire boundary $\partial\mathbb{D}^2 = \mathbb{S}^1$ to a single point.

The torus is the surface of an ordinary donut. More generally, a *genus-g surface*, denoted Σ_g is the surface of a donut with g holes:



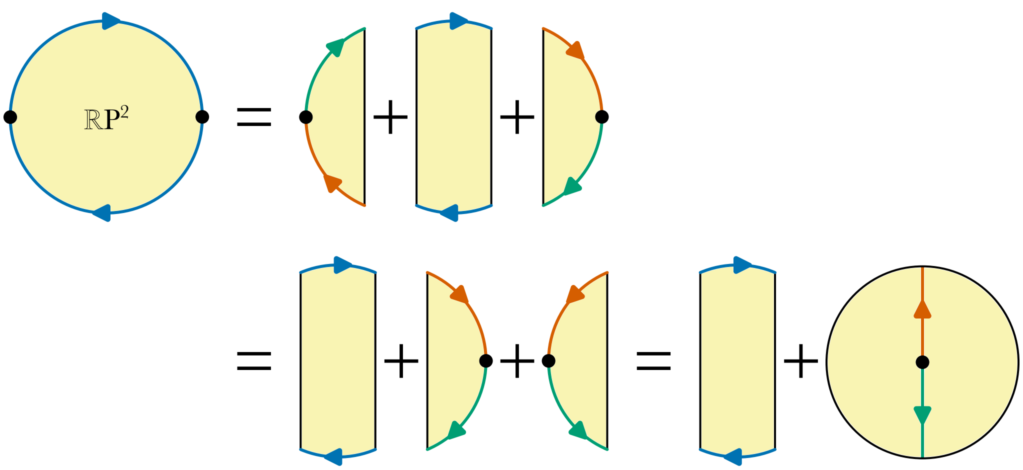
We therefore have $\Sigma_0 \cong \mathbb{S}^2$ and $\Sigma_1 \cong \mathbb{T}^2$. As we will discuss in §C.4 below, for $g \geq 1$ the surface Σ_g can be obtained from a $4g$-gon by identifying sides in an appropriate way.

C.2.2. Möbius band and real projective plane. The surfaces Σ_g are all orientable. We will assume that this notion is familiar from theory of manifolds. The most basic example of a non-orientable surface is a Möbius band:



The Möbius band has one boundary component. To obtain a closed surface, we glue a disk $\mathbb{D}^2$ to this boundary component to form the *real projective plane* $\mathbb{RP}^2$. You might worry that the result depends on the choice of a homeomorphism between $\partial\mathbb{D}^2 = \mathbb{S}^1$ and the boundary component of the Möbius band, but it turns out that the result is independent of the gluing. This holds in great generality; see Exercise C.11. We will use this fact silently throughout this essay.

Pictures of $\mathbb{RP}^2$ are not particularly enlightening,[1] but as the following shows it can be obtained from $\mathbb{D}^2$ by identifying antipodal points on the boundary $\partial\mathbb{D}^2$:



[1]It cannot be embedded in $\mathbb{R}^3$, but only in $\mathbb{R}^4$. There is a way of drawing it in $\mathbb{R}^3$ with self-intersections called the "Boy's Surface", but this picture does not shed much light on its nature.

Another way of viewing $\mathbb{RP}^2$ is as the space of lines through the origin in $\mathbb{R}^3$. To connect this with the above picture, note that every such line intersects the upper hemisphere

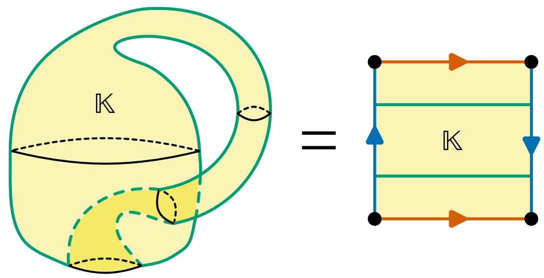
$$U = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1,\ z \geq 0\} \cong \mathbb{D}^2.$$

This intersection is unique except for lines lying in the xy-plane, which intersect

$$\partial U = \{(x, y, 0) \in \mathbb{R}^3 \mid x^2 + y^2 = 1\} \cong \partial\mathbb{D}^2$$

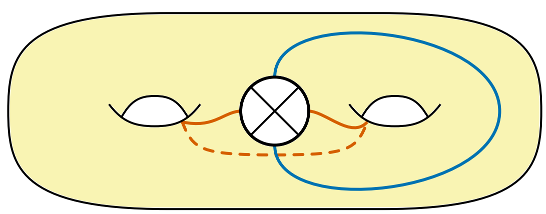
in two antipodal points. The space of lines through the origin can thus be identified with $U \cong \mathbb{D}^2$, but with antipodal points on $\partial U \cong \partial\mathbb{D}^2$ identified.

C.2.3. Klein bottle. Another important example of a non-orientable surface is the Klein bottle $\mathbb{K}$, which is obtained by gluing two Möbius bands together along their boundary. Unlike $\mathbb{RP}^2$, there is a somewhat enlightening way of drawing the $\mathbb{K}$, though necessarily this picture has self-intersections. See here for this and also how to get $\mathbb{K}$ by identifying the sides of a rectangle:

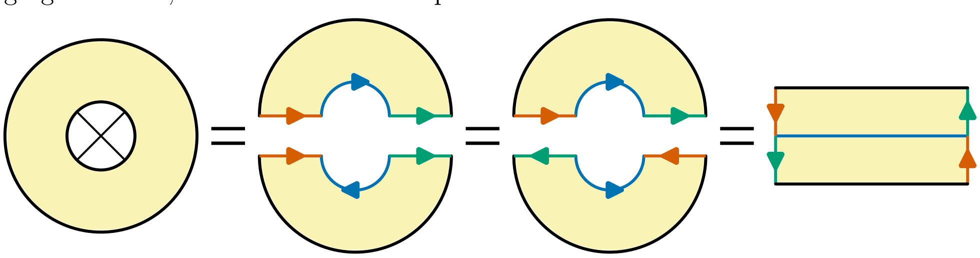


The green curve in both figures is a circle, and when you cut either open along it you get two Möbius bands. This shows that these two surfaces are indeed homeomorphic.

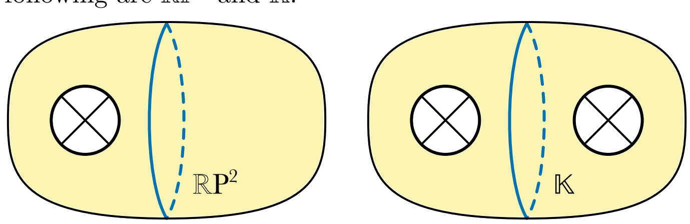
C.2.4. Cross caps. The surfaces $\mathbb{RP}^2$ and $\mathbb{K}$ are the first two elements of an infinite family of non-orientable surfaces. To explain this, we must introduce the notion of a *cross-cap*. A cross-cap on a surface is obtained by removing the interior of a disk and then identifying antipodal points. We denote this by drawing a disk with a cross in it like this:



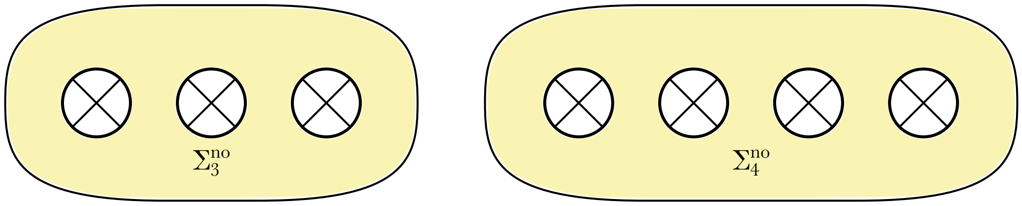
In this figure, the blue and orange arcs are actually disjoint circles embedded in the surface. As the following figure shows, a disk with a cross-cap in it is a Möbius band:



Since $\mathbb{RP}^2$ is a Möbius band with a disk glued to it and $\mathbb{K}$ is two Möbius bands glued together along their boundary, the following are $\mathbb{RP}^2$ and $\mathbb{K}$:



On the left the blue loop divides $\mathbb{RP}^2$ into a Möbius band and a disk, and on the right the blue loop divides $\mathbb{K}$ into two Möbius bands. These pictures suggest the general pattern: the *genus-n nonorientable surface*, denoted Σ_n^{no}, is a sphere with n cross-caps on it:



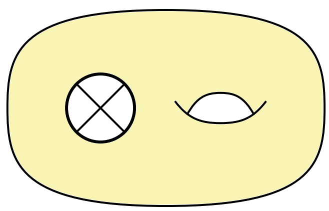
Thus $\Sigma_1^{\text{no}} \cong \mathbb{RP}^2$ and $\Sigma_2^{\text{no}} \cong \mathbb{K}$.

C.2.5. Classification theorem, first version. We can now state a first version of the classification theorem:

THEOREM C.2.1 (Classification of surfaces, weak). *Let Σ be a closed connected surface. Then:*

- *If Σ is orientable, then $\Sigma \cong \Sigma_g$ for a unique $g \geq 0$.*
- *If Σ is non-orientable, then $\Sigma \cong \Sigma_n^{no}$ for a unique $n \geq 1$.*

In some ways this is a very satisfying result, but one weakness is that it does not give an effective way to recognize a given surface. Since it is easy to write down surfaces that do not fit into the above classification in an obvious way, this is a real problem. For instance, consider the following non-orientable surface:

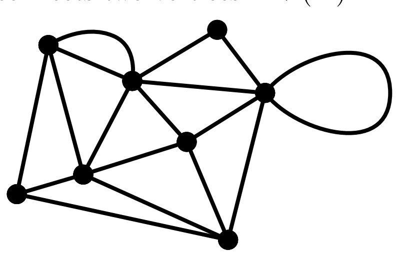


By Theorem C.2.1, this must be homeomorphic to Σ_n^{no} for some $n \geq 1$. However, it is not at all obvious which Σ_n^{no} it is. We will later give a refined classification theorem that will make it clear that the above surface is Σ_3^{no}; see Theorem C.4.12. Before reading this, it is worth trying to prove it for yourself.

C.3. Graphs and their Euler characteristics

As a warm-up before proving the classification of surfaces, this section discusses aspects of graph theory that can be viewed as a one-dimensional analogue of this classification.

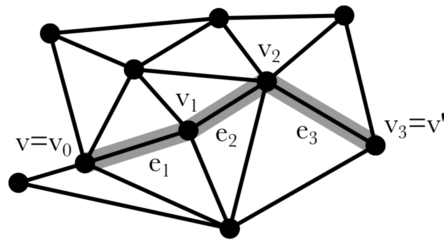
C.3.1. Basic definitions. Recall that a *graph* X is a collection of vertices $V(X)$ and a collection of edges $E(X)$. Each $e \in E(X)$ connects two vertices in $V(X)$. These vertices need not be distinct:



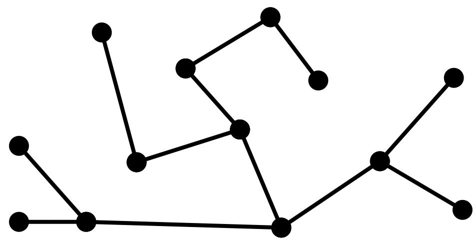
In this essay, we will only consider graphs with finitely many vertices and edges, which we call *finite graphs*. A finite graph is a topological space in a straightforward way. An *edge-path* in a graph from a vertex $v \in V(X)$ to a vertex $v' \in V(X)$ is a sequence of edges $e_1, \ldots, e_n$ such that there exist vertices

$$v = v_0, v_1, \ldots, v_n = v'$$

such that e_i connects v_{i-1} and v_i for all $1 \leq i \leq n$:

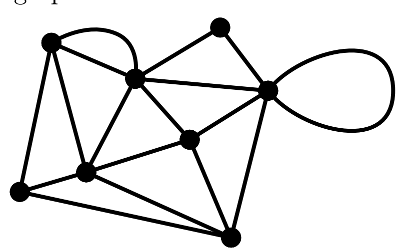


Associated to an edge-path is a continuous map $\gamma\colon I \to X$, and we will often confuse an edge-path with the associated map. The edge-path is *closed* if $v = v'$, in which case the associated path is a loop that we can regard as a continuous map $\gamma\colon \mathbb{S}^1 \to X$. The edge-path is a *cycle* if it is closed, $n \geq 1$, and all the e_i are distinct. A graph is *connected* if all distinct $v, v' \in V(X)$ are connected by an edge-path. This is equivalent to the graph being path-connected as a topological space. A *tree* is a nonempty connected graph with no cycles:



C.3.2. Euler characteristic of graphs. If X is a finite graph, then the *Euler characteristic* of X is $\chi(X) = |V(X)| - |E(X)|$.

Example C.3.1. If X is the graph

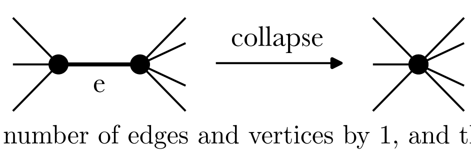


then the Euler characteristic of X is $\chi(X) = |V(X)|/ - |E(X)| = 8 - 17 = -9$. □

C.3.3. Poincaré conjecture for graphs. The importance of the Euler characteristic is illustrated by the following result, which we think of as the "Poincaré conjecture"[2] for graphs:

Lemma C.3.2 (Poincaré conjecture for graphs). *Let X be a finite nonempty connected graph. Then $\chi(X) \leq 1$, with equality if and only if X is a tree.*

Proof. If e is an edge of X connecting two distinct vertices, then we can collapse e without changing whether or not X is a tree:



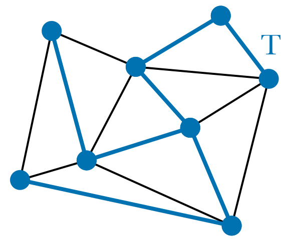
Such a collapse decreases the number of edges and vertices by 1, and thus does not change the Euler characteristic. Collapsing such edges repeatedly, we can therefore assume without loss of generality that all edges of X are loops. Since X is nonempty and connected, this implies that X has one vertex. We therefore have

$$\chi(X) = |V(X)| - |E(X)| = 1 - |E(X)| \leq 1$$

[2]For manifolds, the Poincaré conjecture is a topological characterization of a sphere. Once we have defined the Euler characteristic for surfaces, the two-dimensional Poincaré conjecture will say that a compact connected surface Σ is homeomorphic to $\mathbb{S}^2$ if and only if its Euler characteristic is 2.

with equality if and only if $|E(X)| = 0$, i.e., if and only if X is a tree.[3] □

C.3.4. Maximal trees. If X is a connected nonempty graph, then a maximal tree in X is a subtree T of X containing all the vertices. See here:



These always exist:

Lemma C.3.3. *Let X be a connected nonempty graph. Then X has a maximal tree.*

The proof for finite graphs X is a little easier, and this is the only case we need. We therefore restrict to this case:

Proof of Lemma C.3.3 for finite graphs. Assume that X has n vertices. Inductively define subtrees

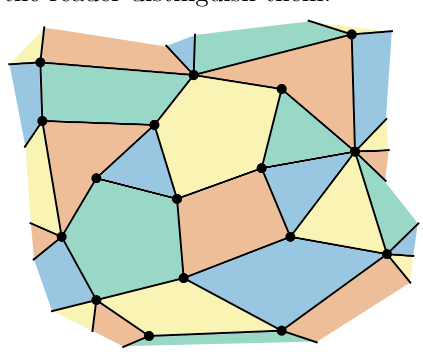
$$T_1 \subset T_2 \subset \cdots \subset T_n$$

of X in the following way. Start by choosing a vertex v_0 of X and letting $T_1 = v_0$. Now assume that T_k has been constructed for some $k < n$. Since X has n vertices, it must be the case that T_k does not contain all the vertices of X. Since X is connected, this implies that there must be an edge e of X that connects a vertex of T_k to a vertex that does not lie in T_k. Let T_{k+1} be the result of adding e to T_k. This process stops at T_n, which contains n vertices and hence is a maximal tree in X. □

C.4. Polygonal decompositions and the Euler characteristic

Our proof of the classification of surfaces will depend on a decomposition of the surface that is a sort of two-dimensional analogue of the decomposition of a graph into vertices and edges.

C.4.1. Basic definitions. A surface equipped with a *polygonal decomposition* is a compact surface Σ (possibly with boundary) together with a finite graph X embedded in Σ such that each path component F of $\Sigma \setminus X$ is homeomorphic to an open disk $\mathrm{Int}(\mathbb{D}^2)$. We will call such an F a *face* of the polygonal decomposition.[4] Here is a picture of part of a polygonal decomposition, with the faces in different colors to help the reader distinguish them:



Here is some terminology for polygonal decompositions:

- The graph X will be called the 1-*skeleton*.
- The vertices and edges of X will be called the vertices and edges of the polygonal decomposition, and the sets of vertices and edges will be written $V(\Sigma)$ and $E(\Sigma)$, respectively.
- The set of faces of the polygonal decomposition will be written $F(\Sigma)$.
- The *Euler characteristic* of the polygonal decomposition is $\chi(\Sigma) = |V(\Sigma)| - |E(\Sigma)| + |F(\Sigma)|$.

[3] Since X has only one vertex, the only way it can be a tree is if it has no edges.

[4] For non-compact surfaces, one would also need to require that the closure of each face is compact.

C.4.2. Existence. The following theorem will play a basic role in our proof:

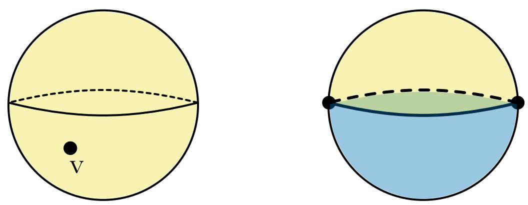
THEOREM C.4.1. *Let Σ be a compact surface, possibly with boundary. Then Σ has a polygonal decomposition.*

See Exercise C.7 for the analogous fact in dimension 1. We will not prove Theorem C.4.1, which requires a long detour into point-set topology. It was originally proved by Radó [**11**]. See [**1**] and [**10**] for modern versions of Radó's proof. I remark that I first learned this proof from [**1**]. A recent and elegant proof along very different lines can be found in [**6**]. Amazingly, the proof in [**6**] uses smooth manifold techniques (even though the surface is not assumed to be smooth), and avoids doing any serious point-set work.

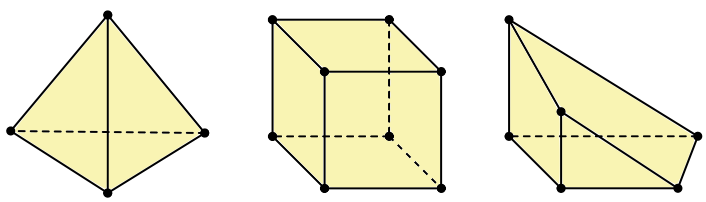
REMARK C.4.2. All of the above sources actually prove the slightly stronger fact that Σ has a triangulation, that is, a polygonal decomposition where the boundaries of each face are length-3 edge-paths in the 1-skeleton. It is easy to subdivide a general polygonal decomposition and turn it into a triangulation. □

C.4.3. Examples of polygonal decompositions. Here are a number of examples:

EXAMPLE C.4.3. Here are two easy polygonal decompositions of $\mathbb{S}^2$:

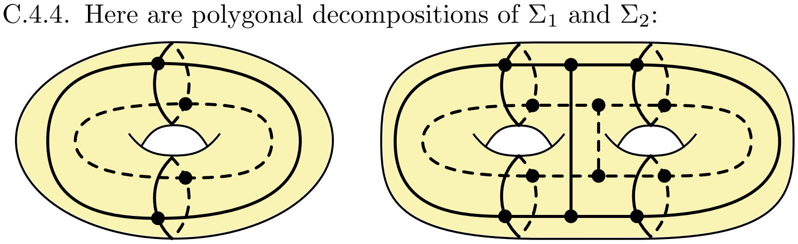


The first has a single vertex v, no edges, and one face $\mathbb{S}^2 \setminus v \cong \mathbb{R}^2 \cong \mathrm{Int}(\mathbb{D}^2)$. Its Euler characteristic is $1 - 0 + 1 = 2$. The second has two vertices, two edges, and two faces. Its Euler characteristic is $2 - 2 + 2 = 2$. Other polygonal decompositions of $\mathbb{S}^2$ can be obtained by identifying the boundaries of polyhedra in $\mathbb{R}^3$ with $\mathbb{S}^2$. For instance:



Here we have stopped trying to draw the faces in different colors. As the reader will check, in each of these cases the Euler characteristic is 2. For instance, the left-most polygonal decomposition has 4 vertices, 6 edges, and 4 faces, so its Euler characteristic is $4 - 6 + 4 = 2$. All these examples reflect a theorem we will discuss below that says that all polygonal decompositions of the same surface have the same Euler characteristic. □

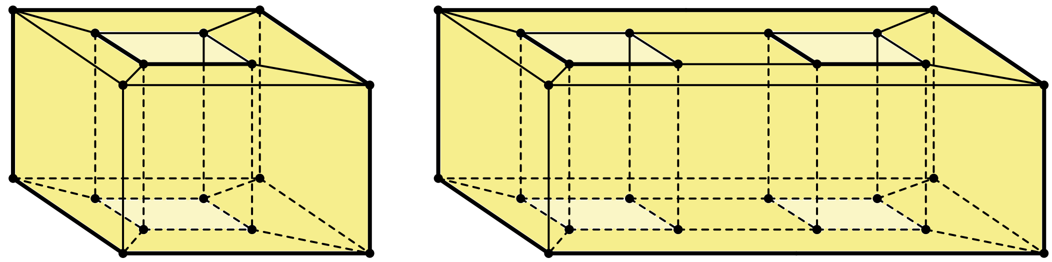
EXAMPLE C.4.4. Here are polygonal decompositions of Σ_1 and Σ_2:



The Euler characteristics of these polygonal decompositions are

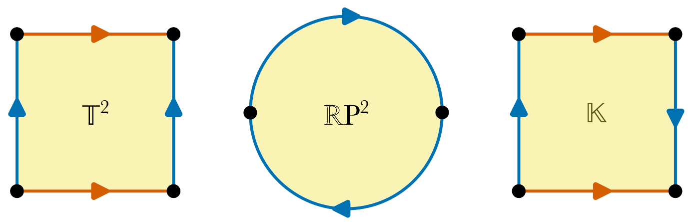
$$\chi(\Sigma_1) = 4 - 8 + 4 = 0,$$
$$\chi(\Sigma_2) = 12 - 22 + 8 = -2.$$

By regarding Σ_1 and Σ_2 as cubes with holes drilled through their centers, we can obtain two other polygonal decompositions for which it is a little easier to see that the faces are open disks:



Again, the reader can verify that these have Euler characteristic 0 and -2. □

Example C.4.5. When we gave examples of surfaces in §C.2, many of our surface came presented as polygons or disks with sides identified. This gives a polygonal decomposition of the surface. For instance, here are pictures of the torus $\mathbb{T}^2$, the real projective plane $\mathbb{RP}^2$, and the Klein bottle $\mathbb{K}$:



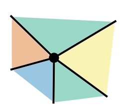
When we identify the boundary points as shown, the boundary becomes a graph and the interior of the disk/polygon gives a single face. In the above examples:

- $\mathbb{T}^2$ has one vertex and two edges and one face, so $\chi(\mathbb{T}^2) = 1 - 2 + 1 = 0$; and
- $\mathbb{RP}^2$ has one vertex and one edge and one face, so $\chi(\mathbb{RP}^2) = 1 - 1 + 1 = 1$; and
- $\mathbb{K}$ has one vertex and two edges and one face, so $\chi(\mathbb{K}) = 1 - 2 + 1 = 0$.

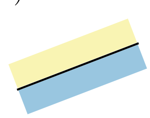
We will give more examples of this later in §C.4.10. □

C.4.4. Local structure of polygonal decompositions. Let Σ be a closed surface equipped with a polygonal decomposition. We now discuss the local structure of this polygonal decomposition. This local structure follows from the definition of a polygonal decomposition. However, the proof is a little technical and we will omit it.[5] For a reader who cannot prove it on their own, we suggest adding these local results to the definition. The various proofs that polygonal decompositions exist (Theorem C.4.1) give polygonal decompositions where this local structure definitely holds.

Our statements will be informal, but will be sufficient to understand the proof of the classification. First, a small neighborhood of a vertex looks like this:

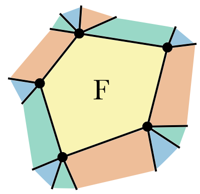


Each of the shaded regions is part of a face. These faces need not all be distinct. Next, consider an edge e. We have $e \cong [0,1]$ and $\mathrm{Int}(e) \cong (0,1)$. A small neighborhood of $\mathrm{Int}(e)$ looks like this:



Again, the two shaded regions are parts of two faces, though these faces might be the same. We now come to a face F. Recall that $F \cong \mathrm{Int}(\mathbb{D}^2)$. There exists a homeomorphism $\phi\colon \mathrm{Int}(\mathbb{D}^2) \to F$ that extends to a continuous map $\phi\colon \mathbb{D}^2 \to F$. The restriction of ϕ to $\partial\mathbb{D}^2 = \mathbb{S}^1$ is usually a closed edge-path in the 1-skeleton:

[5] See Exercise C.8 for analogous results in dimension 1.



The edges in the 1-skeleton traversed by this closed edge path need not be distinct. There is one case where this does not hold: for the polygonal decomposition of $\mathbb{S}^2$ with a single vertex and face and no edges (cf. Example C.4.3), the restriction of ϕ to $\partial\mathbb{D}^2 = \mathbb{S}^1$ is the constant map to this single vertex.

Remark C.4.6. Polygonal decompositions also are useful for surfaces with boundary, but the local structure described above needs to be modified for vertices and edges that are contained in the boundary. □

C.4.5. Well-definedness of Euler characteristic. In our examples above, the Euler characteristics of different polygonal decompositions of the same surface were always the same. This always holds:

Theorem C.4.7. *Let Σ be a compact surface, possibly with boundary. Then the Euler characteristics of any two polygonal decompositions of Σ are the same.*

The most conceptual proof of this theorem uses homology. For any reasonable compact space X, that theory produces integers $b_i(X) \geq 0$ for each $i \geq 0$ called the *Betti numbers* of X. The Betti number of X are manifestly invariants of X, and for any polygonal decomposition of a compact surface Σ we have

$$\chi(\Sigma) = b_0(\Sigma) - b_1(\Sigma) + b_2(\Sigma).$$

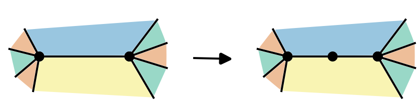
In fact, there are higher-dimensional versions of polygonal decompositions called *CW complex structures*. For a compact space X equipped with a CW complex structure, we have $b_i(X) = 0$ for $i \gg 0$, so the a priori infinite alternating sum

$$\chi(X) = b_0(X) - b_1(X) + b_2(X) - \cdots + (-1)^i b_i(X) + \cdots$$

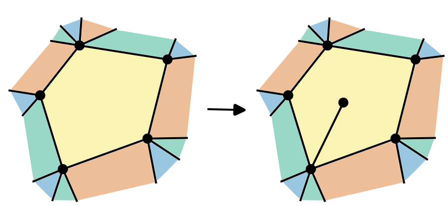
is a finite sum, called the Euler characteristic of X. One of the basic structural theorems about Betti numbers is that $\chi(X)$ also equals an alternating sum of the i-dimensional cells of X. We will prove this more general version in Volume 2 when we discuss homology.

For surfaces, there is also an alternate approach as follows. Along with the existence of polygonal decompositions (Theorem C.4.1), there is also a uniqueness statement saying that any two polygonal decompositions of a compact surface Σ have what is called a *common subdivision*. What this means is that after applying a sequence of the following three moves, any two polygonal decompositions differ by a homeomorphism of Σ:

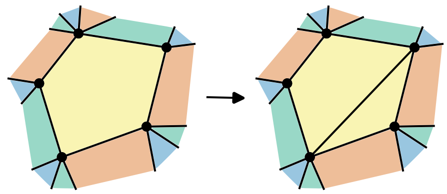
- Subdivide an edge as follows:



- Add a vertex and edge in the interior of a face as follows:



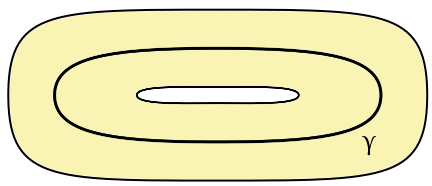
- Add an edge connecting two vertices of a face as follows:



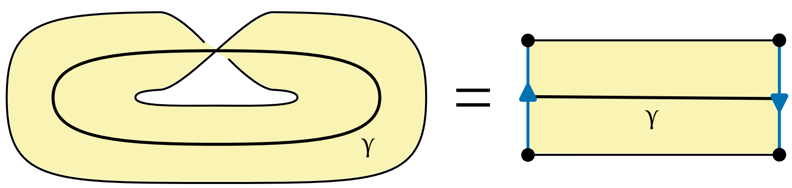
To prove that the Euler characteristic does not depend on the polygonal decomposition, it is thus enough to observe that the above three moves do not change it. For instance, subdividing an edge add one vertex and one edge, and these cancel out when calculating the Euler characteristic.

C.4.6. Cutting. We now explore the effect on the Euler characteristic of cutting along simple closed curves. Let Σ be a compact surface, possibly with boundary. Let γ be a simple closed curve lying in the interior of Σ. From the theory of manifolds, γ has what is called a *tubular neighborhood.* There are two cases:

- This tubular neighborhood is an annulus with γ in its interior like this:

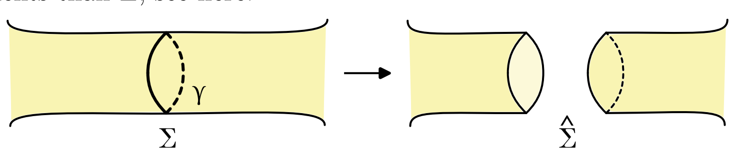


  In this case, we say that γ is a *two-sided curve.* If Σ is orientable, all γ are two-sided.
- This tubular neighborhood is a Möbius band with γ in its interior like this:



  In this case, we say that γ is a *one-sided curve.*

Cutting Σ open along γ turns Σ into a surface $\widehat{\Sigma}$. If γ is two-sided, the surface $\widehat{\Sigma}$ has two more boundary components than Σ, see here:

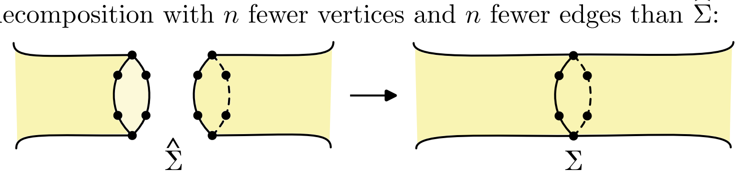


On the other hand, if γ is one-sided then $\widehat{\Sigma}$ has only more more boundary component. We remark that $\widehat{\Sigma}$ might be disconnected.

We will prove:

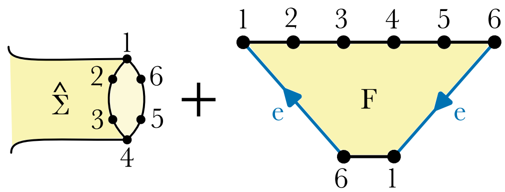
Lemma C.4.8. *Let Σ be a compact surface, possibly with boundary. Let γ be a simple closed curve in the interior of Σ and let $\widehat{\Sigma}$ be the result of cutting Σ open along γ. Then $\chi(\widehat{\Sigma}) = \chi(\Sigma)$.*

Proof. Fix a polygonal decomposition of $\widehat{\Sigma}$. Assume first that γ is two-sided. In this case, there are boundary components ∂_1 and ∂_2 of $\widehat{\Sigma}$ such that Σ can be obtained by gluing ∂_1 to ∂_2. Necessarily ∂_1 and ∂_2 are cycles in the 1-skeleton of $\widehat{\Sigma}$. Subdividing edges in the ∂_i if necessary, we can assume that ∂_1 and ∂_2 contain the same number of n of vertices. As the following shows, Σ then has a polygonal decomposition with n fewer vertices and n fewer edges than $\widehat{\Sigma}$:



This figure does not include the part of the polygonal decomposition lying in the interior of $\widehat{\Sigma}$, which is irrelevant for this calculation. These changes in the numbers of vertices and edges cancel out when we calculate the Euler characteristic, giving that $\chi(\widehat{\Sigma}) = \chi(\Sigma)$.

Now assume that γ is one-sided. In this case, there is a boundary component ∂ of $\widehat{\Sigma}$ such that Σ can be obtained by gluing a Möbius band to ∂. Necessarily ∂ lies in the 1-skeleton of $\widehat{\Sigma}$. Assume that ∂ has n vertices. As the following shows, with respect to an appropriate polygonal decomposition Σ has 1 more edge (labeled e) and 1 more face (labeled F) than $\widehat{\Sigma}$:



Here we have drawn the Möbius band in a skewed way to emphasize that like on ∂ its vertices are equally spaced around its single boundary component, and again we did not include the part of the polygonal decomposition lying in the interior of $\widehat{\Sigma}$. These changes in the numbers of vertices and edges again cancel out when we calculate the Euler characteristic, giving that $\chi(\widehat{\Sigma}) = \chi(\Sigma)$. □

C.4.7. Capping. We now study the effect on the Euler characteristic of gluing a disk to a boundary component. Recall that the result does not depend on the gluing map (Exercise C.11).

LEMMA C.4.9. *Let Σ be a compact surface with boundary and let ∂ be a boundary component of Σ. Let $\overline{\Sigma}$ be the result of gluing a disk to ∂. Then $\chi(\overline{\Sigma}) = \chi(\Sigma) + 1$.*

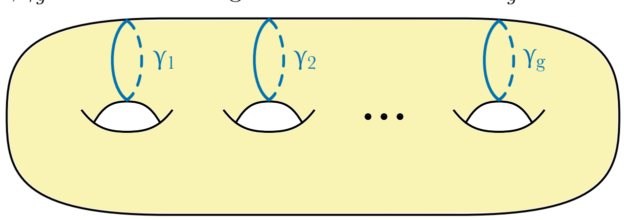
PROOF. Choose a polygonal decomposition of Σ. The disk we glued to ∂ can serve as a face of a polygonal decomposition of $\overline{\Sigma}$, giving a polygonal decomposition of $\overline{\Sigma}$ with the same number of vertices and edges as Σ and one more face. It follows that

$$\chi(\overline{\Sigma}) = |V(\overline{\Sigma})| - |E(\overline{\Sigma})| + |F(\overline{\Sigma})| = |V(\Sigma)| - |E(\Sigma)| + |F(\Sigma)| + 1 = \chi(\Sigma) + 1. \qquad \square$$

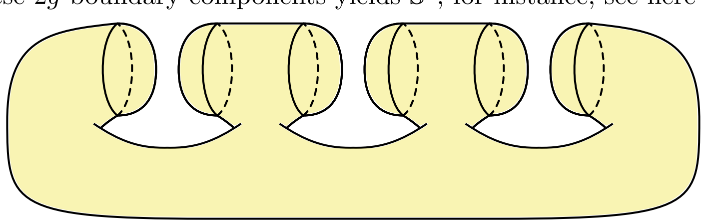
C.4.8. Euler characteristic calculations. Our results about cutting and capping make it easy to calculate the Euler characteristics of surfaces without needing to find explicit polygonal decompositions of them. As examples of this, we calculate the Euler characteristics of the genus-g surface Σ_g and the nonorientable genus-n surface Σ_n^{no}:

LEMMA C.4.10. *For $g \geq 0$, we have $\chi(\Sigma_g) = 2 - 2g$.*

PROOF. Let $\gamma_1, \ldots, \gamma_g$ be the following two-sided curves on Σ_g:



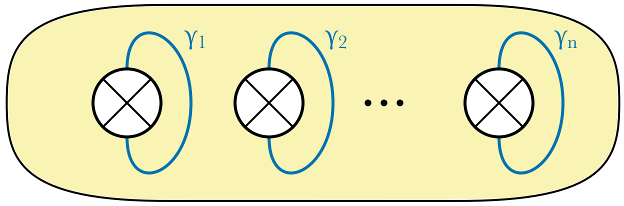
Cutting Σ_g open along the γ_i yields a connected surface $\widehat{\Sigma}$ with $2g$ boundary components. Gluing disks to each of these $2g$ boundary components yields $\mathbb{S}^2$; for instance, see here for the case $g = 3$:



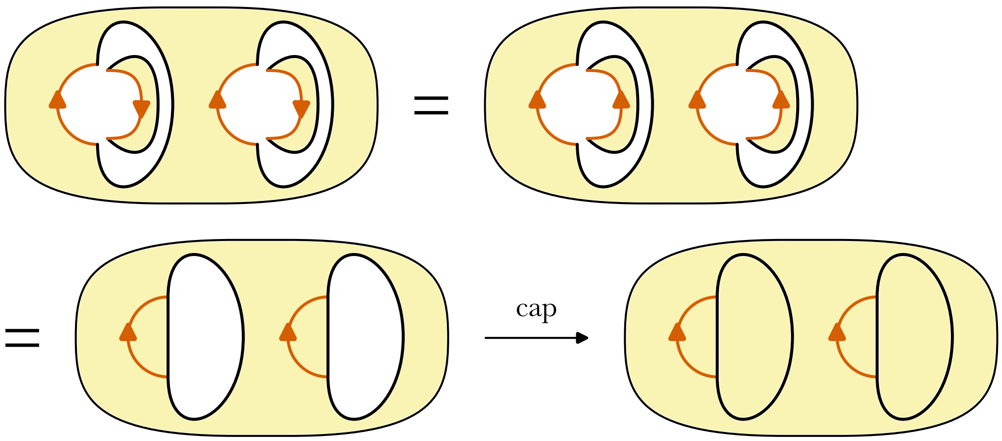
Applying Lemmas C.4.8 and C.4.9, we deduce that $\chi(\Sigma_g) = \chi(\widehat{\Sigma}) = \chi(\mathbb{S}^2) - 2g = 2 - 2g$. □

LEMMA C.4.11. *For $n \geq 1$, we have $\chi(\Sigma_n^{no}) = 2 - n$.*

PROOF. Let $\gamma_1, \ldots, \gamma_n$ be the following one-sided curves on Σ_n^{no}:



Cutting Σ_n^{no} open along the γ_i yields a connected surface $\widehat{\Sigma}$ with n boundary components. Gluing disks to each of these n boundary components yields $\mathbb{S}^2$; for instance, see here for the case $n = 2$:



Applying Lemmas C.4.8 and C.4.9, we deduce that $\chi(\Sigma_n^{\mathrm{no}}) = \chi(\widehat{\Sigma}) = \chi(\mathbb{S}^2) - n = 2 - n$. □

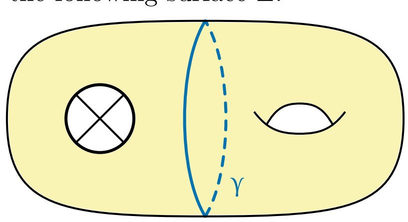
C.4.9. Refined classification. Recall that the classification theorem for surfaces says that every closed connected surface is homeomorphic to some Σ_g or Σ_n^{no}. Since $\chi(\Sigma_g) = 2 - 2g$ and $\chi(\Sigma_n^{\mathrm{no}}) = 2 - n$, this will imply that the homeomorphism type of a closed connected surface is completely determined by its Euler characteristic and whether or not it is orientable. We state this refined version of the classification as follows:

THEOREM C.4.12 (Classification of surfaces). *Let Σ be a closed connected surface. Then:*

- *If Σ is orientable then $\Sigma \cong \Sigma_g$, where $g \geq 0$ satisfies $\chi(\Sigma) = 2 - 2g$. In particular, $\chi(\Sigma)$ is even.*
- *If Σ is non-orientable then $\Sigma \cong \Sigma_n^{no}$, where $n \geq 1$ satisfies $\chi(\Sigma) = 2 - n$.*

Using this, we can answer the question we posed after stating the first version of the classification (Theorem C.2.1):

EXAMPLE C.4.13. Consider the following surface Σ:



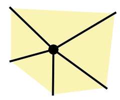
This surface is non-orientable since it contains a cross-cut. To calculate its Euler characteristic, let γ be the curve drawn above. The curve γ is two-sided, and when we cut Σ open along it and cap off the resulting two boundary components we get Σ_1 and Σ_1^{no}. It follows that

$$\chi(\Sigma) = \chi(\Sigma_1) + \chi(\Sigma_1^{\mathrm{no}}) - 2 = 0 + 1 - 2 = -1.$$

By Theorem C.4.12, we deduce that $\Sigma \cong \Sigma_3^{\mathrm{no}}$. □

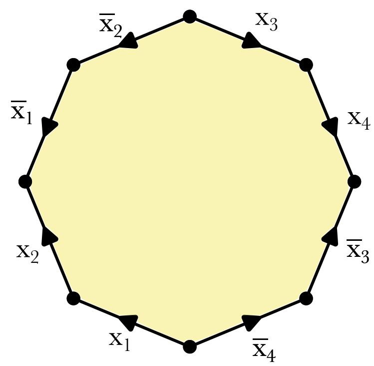
REMARK C.4.14. Many proofs of the classification of surfaces require the homeomorphism $\Sigma \cong \Sigma_3^{no}$ from Example C.4.13, which must be proved by hand. As we will see, our proof does not need this fact, so the above argument is not circular. □

C.4.10. Polygons with sides identified. We will start the proof of Theorem C.4.12 soon, but first we give a few more examples of how it can be used. One natural way to build a surface is to take a polygon (or several polygons) in $\mathbb{R}^2$ and glue its sides together. We saw several examples of this already in Example C.4.5. If each side is glued to exactly one other side, then the result will always be a surface. Indeed, the only place where it is not obviously a surface is at the vertices, and a neighborhood of a vertex looks like this:

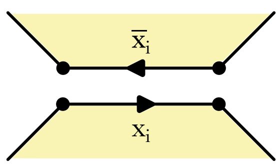


This surface has a natural polygonal decomposition where the glued-together boundary forms the 1-skeleton and the polygon is a single face (or multiple faces if there are multiple polygons). Here are some example of how to use Theorem C.4.12 to identify the resulting surface:

EXAMPLE C.4.15. Let Σ be an octagon with sides identified as follows:



Here we have labeled the sides with letters and oriented them to show how they should be glued. The bars over the letters indicate a reversed orientation on the edge. The surface Σ is orientable since the gluing respects the orientation of the plane:



All the vertices are identified to a single vertex, and the boundary edges glue to 4 edges. Since there is a single face, we see that $\chi(\Sigma) = 1 - 4 + 1 = -2$. It follows that $\Sigma \cong \Sigma_2$. □

REMARK C.4.16. A nice way to give a gluing pattern on the boundary of a $2n$-gon is to label the paired edges by letters $x_1, \ldots, x_n$. You then give a word in letters $\{x_1, \overline{x}_1, \ldots, x_n, \overline{x}_n\}$ obtained by going around the polygon clockwise starting from some vertex and recording which edge-labels appear, with a bar indicating that the edge is traversed in the opposite orientation. For instance, in Example C.4.15 the word would be $x_1 x_2 \overline{x}_1 \overline{x}_2 x_3 x_4 \overline{x}_3 \overline{x}_4$. For each $1 \leq i \leq n$, the letter x_i should appear twice (each time possibly with a bar over it). □

EXAMPLE C.4.17. Generalizing Example C.4.15, let Σ be a $4g$-gon with sides identified according to the pattern

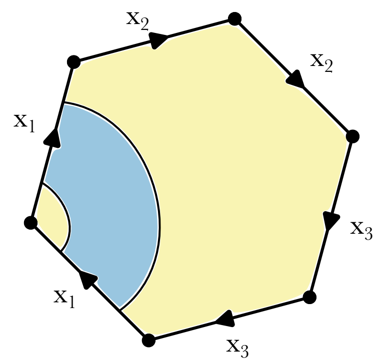
$$x_1 x_2 \overline{x}_1 \overline{x}_2 \cdots x_{2g-1} x_{2g} \overline{x}_{2g-1} \overline{x}_{2g}.$$

Again, all the vertices get identified to a single vertex and there are $2g$ edges and one face, so $\chi(\Sigma) = 1 - 2g + 1 = 2 - 2g$. We thus have $\Sigma \cong \Sigma_g$. □

EXAMPLE C.4.18. Let Σ be a $2n$-gon with sides identified according to the pattern

$$x_1 x_1 x_2 x_2 \cdots x_n x_n.$$

For instance, for $n = 3$ this is



The blue strip here glues up to a Möbius band, so Σ is non-orientable. All the vertices get identified to a single vertex and there are n edges and one face, so $\chi(\Sigma) = 1 - n + 1 = 2 - n$. We thus have $\Sigma \cong \Sigma_n^{\mathrm{no}}$. □

C.5. The two-dimensional Poincaré conjecture

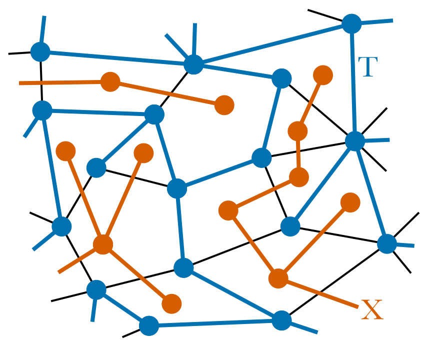
We now start the proof of the classification by proving the two-dimensional Poincaré conjecture, which says that $\mathbb{S}^2$ is the unique connected closed surface with Euler characteristic 2:

Theorem C.5.1 (Two-dimensional Poincaré conjecture). *Let Σ be a closed connected surface. Then $\chi(\Sigma) \leq 2$, with equality if and only if $\Sigma \cong \mathbb{S}^2$.*

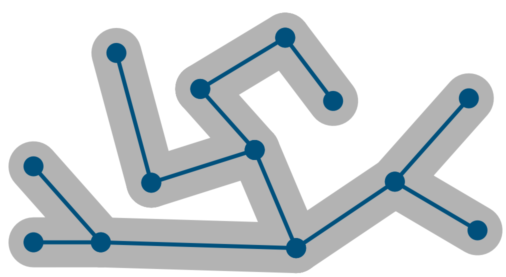
Proof. Fix a polygonal decomposition of Σ. Its 1-skeleton is a finite graph embedded in Σ. Let T be a maximal tree in the 1-skeleton. Next, let X the following graph, which one can view as a sort of "dual graph" to T:

- Put a vertex of X in the center of each polygon of our polygonal decomposition.
- Connect two vertices of X if their associated polygons share an edge that does not lie in T.

See here:



We claim that the graph X is connected. Equivalently, removing T does not disconnect Σ. The key to this is the fact that a small neighborhood U of T is homeomorphic to $\mathbb{D}^2$:

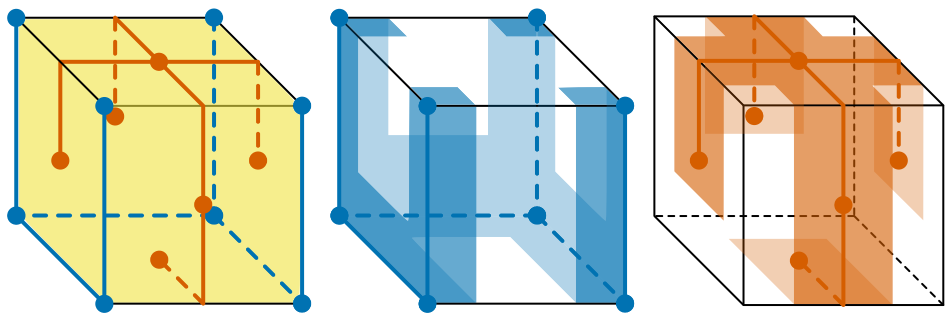


From this, we see that a path in Σ that crosses T can be re-routed to turn and follow the boundary of U until it gets to the other side of T, proving that T does not disconnect the surface.

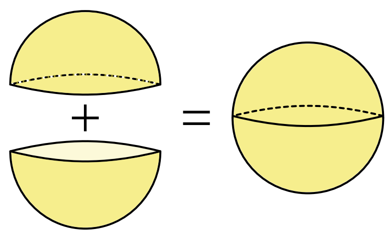
Lemma C.3.2 implies that $\chi(T) = 1$ and that $\chi(X) \leq 1$ with equality if and only if X is a tree. Since each vertex of Σ is a vertex of T, each polygon of Σ contains a unique vertex of X, and each edge of Σ is either an edge of T or is crossed by a unique edge of X, we have

$$\begin{aligned}\chi(\Sigma) &= |V(\Sigma)| - |E(\Sigma)| + |F(\sigma)| = |V(T)| - (|E(T)| + |E(X)|) + |V(X)| \\ &= \chi(T) + \chi(X) \leq 1 + 1 = 2.\end{aligned}$$

This proves half of the theorem. Equality holds if and only if $\chi(X) = 1$, i.e, if X is a tree. Assume, therefore, that X is a tree. We must prove that $\Sigma \cong \mathbb{S}^2$. To see this, note that just like above if we slightly thicken T and X we get disks D_1 and D_2. Choosing these thickenings carefully, we can ensure that D_1 and D_2 intersect in their boundaries. To help the reader understand this, here is an example of a polynomial decomposition of a surface where T and X are trees, along with D_1 and D_2:



We deduce that Σ can be decomposed into two disks meeting along their boundaries:



It follows that $\Sigma \cong \mathbb{S}^2$. $\square$

Before continuing with the classification, we pause to extract the following from the above proof:

Corollary C.5.2. *Let Σ be a closed connected surface such that $\chi(\Sigma) < 2$. Then there exists a simple closed curve γ on Σ that is nonseparating, i.e., such that $\Sigma \setminus \gamma$ is connected.*

Proof. Fix a polygonal decomposition of Σ, and let T and X be as in the proof of Theorem C.5.1. Since $\chi(\Sigma) < 2$, it follows from the proof of Theorem C.5.1 that X is a connected graph that is *not* a tree. It therefore contains a cycle γ. We claim that γ is nonseparating.

In fact, even more is true: $\Sigma \setminus X$ is connected. To see this, note that any point of $\Sigma \setminus X$ can be connected by a path in $\Sigma \setminus X$ to a vertex of the polygonal decomposition. This vertex lies in the maximal tree T, and since T is connected we can follow a path in T to any other vertex of the polygonal decomposition. The claim follows. $\square$

C.6. The classification of surfaces in general

We now come to the proof of the classification of surfaces, whose statement we recall:

Theorem C.4.12 (Classification of surfaces). *Let Σ be a closed connected surface. Then:*

- *If Σ is orientable then $\Sigma \cong \Sigma_g$, where $g \geq 0$ satisfies $\chi(\Sigma) = 2 - 2g$. In particular, $\chi(\Sigma)$ is even.*
- *If Σ is non-orientable then $\Sigma \cong \Sigma_n^{no}$, where $n \geq 1$ satisfies $\chi(\Sigma) = 2 - n$.*

Proof. Theorem C.5.1 says that $\chi(\Sigma) \leq 2$. The proof will be by reverse induction on $\chi(\Sigma)$. The base case is when $\chi(\Sigma) = 2$, in which case Theorem C.5.1 says that $\Sigma \cong \mathbb{S}^2 \cong \Sigma_0$. In particular, Σ must be orientable in this case, as claimed in the theorem.

Assume now that $\chi(\Sigma) < 2$ and that the theorem is true for closed connected surfaces with larger Euler characteristics. Corollary C.5.2 implies that there is a nonseparating simple closed curve γ on

Σ. As we discussed in §C.4.6, the curve γ is either two-sided or one-sided. Let $\widehat{\Sigma}$ be the connected surface with boundary obtained by cutting Σ open along γ. There are four cases, with the first case being the only one that occurs for Σ orientable:

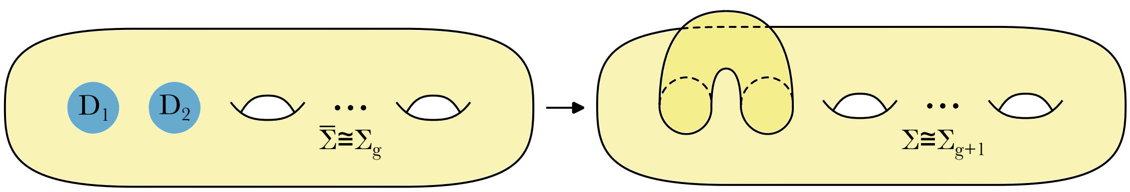
CASE 1. *γ is two-sided and $\widehat{\Sigma}$ is orientable.*

Since γ is two-sided, $\widehat{\Sigma}$ has two boundary components. Let $\overline{\Sigma}$ be the closed connected surface obtained from $\widehat{\Sigma}$ by gluing disks to both of its boundary components. Using Lemmas C.4.8 and C.4.9, we have

$$\chi(\overline{\Sigma}) = \chi(\widehat{\Sigma}) + 2 = \chi(\Sigma) + 2.$$

We can therefore apply our inductive hypothesis to $\overline{\Sigma}$. Since $\widehat{\Sigma}$ is orientable, so is $\overline{\Sigma}$. It follows that $\overline{\Sigma} \cong \Sigma_g$, where $g \geq 0$ satisfies $\chi(\overline{\Sigma}) = 2 - 2g$. Since $\chi(\Sigma) = \chi(\Sigma) - 2$, we have $\chi(\Sigma) = 2 - 2(g+1)$. Our goal, therefore, is to prove that $\Sigma \cong \Sigma_{g+1}$.

To see this, note that Σ is obtained from $\overline{\Sigma} \cong \Sigma_g$ by removing two open disks D_1 and D_2 whose closures are disjoint and gluing together the resulting boundary components. In other words, Σ is obtained by attached a handle as follows:



It follows that $\Sigma \cong \Sigma_{g+1}$, as desired.

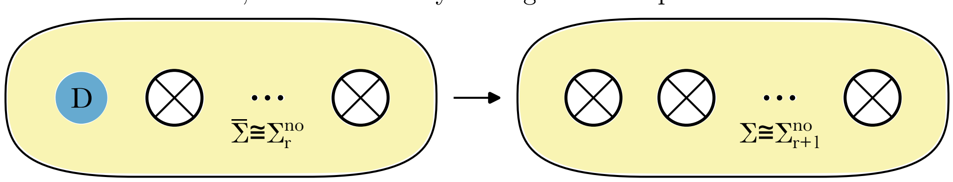
CASE 2. *γ is one-sided and $\widehat{\Sigma}$ is non-orientable.*

Since γ is one-sided, $\widehat{\Sigma}$ has one boundary component. Let $\overline{\Sigma}$ be the closed connected surface obtained from $\widehat{\Sigma}$ by gluing disks to its boundary component. Using Lemmas C.4.8 and C.4.9, we have

$$\chi(\overline{\Sigma}) = \chi(\widehat{\Sigma}) + 1 = \chi(\Sigma) + 1.$$

We can therefore apply our inductive hypothesis to $\overline{\Sigma}$. Since $\widehat{\Sigma}$ is non-orientable, so is $\overline{\Sigma}$. It follows that $\overline{\Sigma} \cong \Sigma_n^{\text{no}}$, where $n \geq 1$ satisfies $\chi(\overline{\Sigma}) = 2 - n$. Since $\chi(\Sigma) = \chi(\Sigma) - 1$, we have $\chi(\Sigma) = 2 - (n+1)$. Our goal, therefore, is to prove that $\Sigma \cong \Sigma_{n+1}^{\text{no}}$.

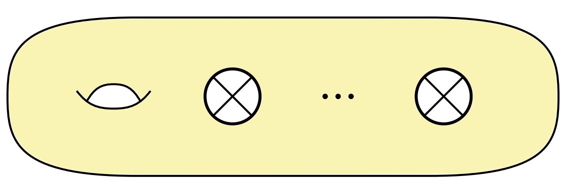
To see this, note that Σ is obtained from $\overline{\Sigma} \cong \Sigma_n^{\text{no}}$ by removing an open disk D and gluing in a Möbius band. In other words, Σ is obtained by adding a cross-cap as follows:



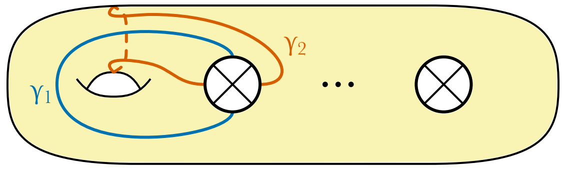
It follows that $\Sigma \cong \Sigma_{n+1}^{\text{no}}$, as desired.

CASE 3. *γ is two-sided and $\widehat{\Sigma}$ is non-orientable.*

Following the argument in the previous two cases, the surface Σ has one genus and $n \geq 1$ cross-caps as follows:



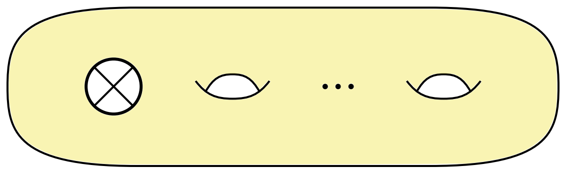
What we did wrong in this case was choose the wrong curve γ to cut along. Let γ_1 and γ_2 be as follows:



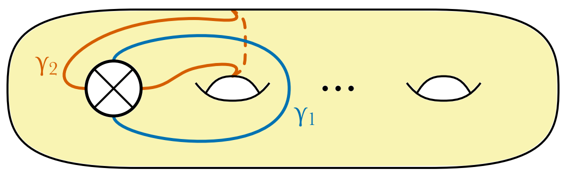
The curves γ_1 and γ_2 are both one-sided. Since γ_2 is one-sided and disjoint from γ_1, it follows that the surface $\widehat{\Sigma}'$ obtained by cutting along γ_1 is non-orientable. This reduces us to Case 2.

Case 4. *γ is one-sided and $\widehat{\Sigma}$ is orientable.*

This time, if we follow the argument from Cases 1 and 2 we get that Σ has 1 cross-cap and $g \geq 0$ genus as follows:



If $g = 0$ then $\Sigma \cong \Sigma_1^{\mathrm{no}}$ and we are done. Otherwise, we can use the same trick we used in Case 3. Namely, let γ_1 and γ_2 be as follows:

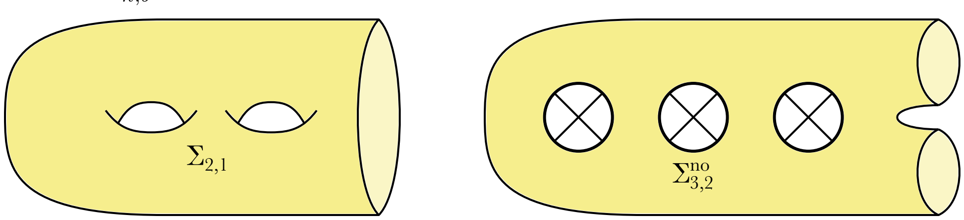


The curves γ_1 and γ_2 are both one-sided. Since γ_2 is one-sided and disjoint from γ_1, it follows that the surface $\widehat{\Sigma}'$ obtained by cutting along γ_1 is non-orientable. This reduces us to Case 2. $\square$

C.7. Extensions of the classification of surfaces

We close this essay by describing two extensions of the classification of surfaces: to compact surfaces with boundary, and to non-compact surfaces.

C.7.1. Compact surfaces with boundary. Let $\Sigma_{g,b}$ be genus-g surface Σ_g with b open disks removed and let $\Sigma_{n,b}^{\mathrm{no}}$ be a non-orientable genus-n surface with b open disks removed. For instance,



Both $\Sigma_{g,b}$ and $\Sigma_{n,b}$ are compact surfaces with b boundary components. These are the only surfaces with boundary:

Theorem C.7.1 (Classification of surfaces with boundary). *Let Σ be a compact connected surface with $b \geq 0$ boundary components. Then:*

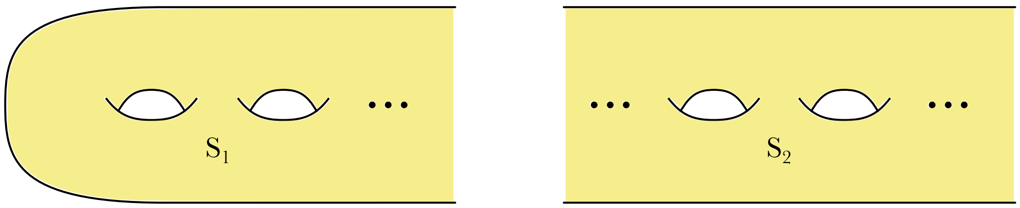
- *If Σ is orientable then $\Sigma \cong \Sigma_{g,b}$, where $g \geq 0$ satisfies $\chi(\Sigma) = 2 - 2g - b$.*
- *If Σ is non-orientable then $\Sigma \cong \Sigma_{n,b}^{no}$, where $n \geq 1$ satisfies $\chi(\Sigma) = 2 - n - b$.*

Proof. Gluing disks to all the boundary components of Σ gives a closed connected surface $\overline{\Sigma}$ with $\chi(\overline{\Sigma}) = \chi(\Sigma) + b$. By the classification of surfaces (Theorem C.4.12), we either have $\overline{\Sigma} \cong \Sigma_g$ with $\chi(\overline{\Sigma}) = 2 - 2g$ or $\overline{\Sigma} \cong \Sigma_n^{\mathrm{no}}$ with $\chi(\overline{\Sigma}) = 2 - n$ depending on whether or not $\overline{\Sigma}$ (and hence Σ) is orientable. The theorem follows. $\square$

C.7.2. Noncompact surfaces. One way to get a noncompact surface is to remove a finite number of points from the interior of a compact surface with boundary. This gives what is called a surface of *finite type*. However, non-compact surfaces can be much more complicated than this. Here are some examples.

EXAMPLE C.7.2. Let C be a Cantor set embedded in $\mathbb{S}^2$. Then $\mathbb{S}^2 \setminus C$ is a very complicated non-compact surface. □

EXAMPLE C.7.3. Consider the following infinite-genus surfaces S_1 and S_2:



The difference between them is that S_1 has genus going off to infinity only to the right, while S_2 has genus going off to infinity in both directions. These surfaces are not homeomorphic. One way to distinguish them is to note that for every compact subset $K_1 \subset S_1$, there is only one component C of $S_1 \setminus K_1$ such that $\overline{C}$ is noncompact. However, there exist compact subsets $K_2 \subset S_2$ such that $S_2 \setminus K_2$ has two such components. This can be formalized using the theory of what are called "ends" of a space. □

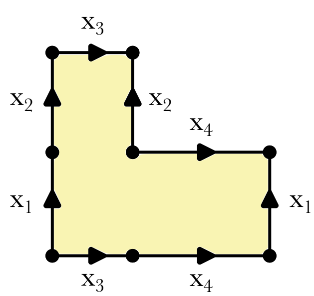
This might lead the reader to think that there is no hope of classifying noncompact surfaces. However, there is such a classification making use of "end data". It was first stated by Kerékjártó [**7**, Chapter 5]. His proof had gaps, and the first correct proof was found by Richards [**8**].

REMARK C.7.4. Noncompact surfaces with boundary are even more complicated, especially if they have noncompact boundary components. However, a classification of them was found by Brown–Messer [**3**]. □

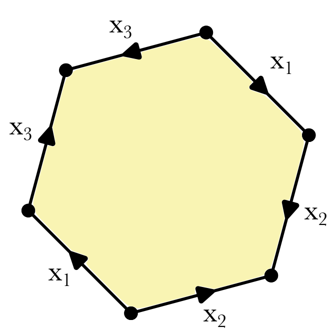
C.8. Exercises

EXERCISE C.1. Determine the surfaces by identifying sides of polygons as follows:

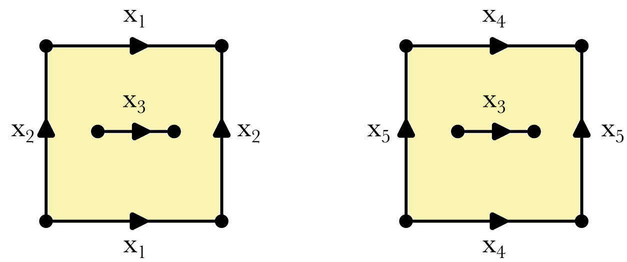
(a) The L-shaped polygon with opposite sides identified here:



(b) The hexagon with sides identified here:

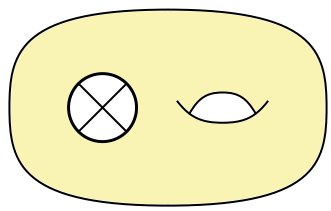


(c) The two squares with slits here:



For this, make sure you are working with a polygonal decomposition. □

Exercise C.2. As we discussed after stating Theorem C.4.12, the following surface is homeomorphic to Σ_3^{no}:

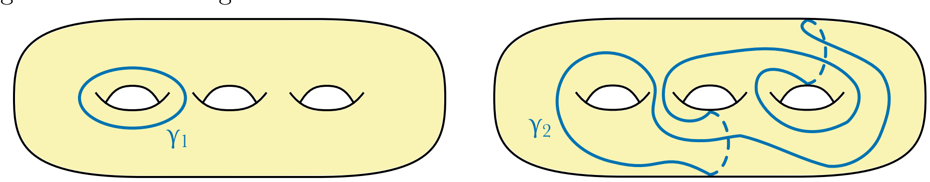


Prove this directly by decomposing both Σ_3^{no} and the above surface into the union of three Möbius bands and $\Sigma_{0,3}$ glued along their boundaries. □

Exercise C.3. Using the fact that the Euler characteristic of $\mathbb{S}^2$ is 2, prove that any regular polyhedron in $\mathbb{R}^3$ is either a tetrahedron, a cube, an octahedron, a dodecahedron, or an icosahedron. This goes back to the ancient greeks, and appears in Euclid's Elements. □

Exercise C.4. Using the fact that the Euler characteristic of $\mathbb{S}^2$ is 2, determine the number of components of the complement of n great circles on $\mathbb{S}^2$ no three of which pass though a common point. □

Exercise C.5. Let γ_1 and γ_2 be two nonseparating simple closed curves on Σ_g. For instance, the γ_i might be the following:



Prove that there is a homeomorphism $f\colon \Sigma_g \to \Sigma_g$ such that $f(\gamma_1) = \gamma_2$. Hint: use the classification of surfaces with boundary to prove that the surfaces you get by cutting Σ_g open along the γ_i are homeomorphic. □

Exercise C.6. Let $f\colon \widetilde{\Sigma} \to \Sigma$ be a degree-d cover between closed connected surfaces. Prove that $\chi(\widetilde{\Sigma}) = d\chi(\Sigma)$. Hint: Figure out how to lift a polygonal decomposition of Σ to one of $\widetilde{\Sigma}$. □

Exercise C.7. Let M be a 1-manifold. A *triangulation* of M is a closed discrete subset $V \subset M$ call the *vertices* such that each path-component E of $M \setminus V$ is homeomorphic to $(0,1)$ and has compact closure. We will call these E the *edges* of the triangulation. Prove that M has a triangulation by following these steps:

(a) First prove that there is a countable open cover $\{U_i \mid i \in I\}$ of M with the following properties:
   - For each $i \in I$, the closure $\overline{U}_i$ is a closed subset of M homeomorphic to $[0,1]$ via a homeomorphism taking U_i to $(0,1)$.
   - The set $\{\overline{U}_i \mid i \in I\}$ is locally finite, i.e., for each compact subset $K \subset M$ the set $\{i \in I \mid \overline{U}_i \cap K \neq \emptyset\}$ is finite.

   We remark that this will use the fact that M is second countable and Hausdorff.

(b) Letting $\{U_i \mid i \in I\}$ be as in (a), set
$$V = \bigcup_{i \in I} \overline{U}_i \setminus U_i.$$
Prove that V is the set of vertices of a triangulation of M. We remark that this will also use the fact that M is Hausdorff. □

EXERCISE C.8. Let M be a 1-manifold. Fix a triangulation of M as in Exercise C.7. We will verify that M has local properties similar to those of polygonal decompositions of surfaces; cf. §C.4.4.

(a) Prove that for each vertex v of the triangulation, there exists an open neighborhood U of v along with a homeomorphism $f\colon (-1,1) \to U$ such that $f(0) = v$ and such that f takes both $(-1,0)$ and $(0,1)$ into open subsets of edges of the triangulation. These edges need not be distinct.

(b) Let E be an edge of the triangulation. Prove that there is a homeomorphism $g\colon (0,1) \to E$ that extends to a continuous map $G\colon [0,1] \to M$ with $G(0)$ and $G(1)$ vertices of the triangulation. Hint: This exercise will require you to use the fact that manifolds are Hausdorff. □

EXERCISE C.9. Let M be a connected 1-manifold. Prove that M is homeomorphic to either $\mathbb{S}^1$ or $\mathbb{R}$. Hint: use a triangulation as in Exercises C.7 and C.8. We remark that this relies on the fact that manifolds are second countable and Hausdorff. Without these conditions there are many more examples of connected 1-manifolds. Even assuming second countability, as far as I am aware there is no reasonable classification of non-Hausdorff 1-manifolds. □

EXERCISE C.10. Let Σ be either $\Sigma_{g,b}$ or $\Sigma^{\text{no}}_{n,b}$ with $b \geq 1$, let ∂ be a boundary component of Σ, and let $f\colon \partial \to \partial$ be a homeomorphism. In this exercise, you will prove that f extends to a homeomorphism $F\colon \Sigma \to \Sigma$.

(a) Prove that if $\phi\colon \mathbb{S}^1 \to \mathbb{S}^1$ is an orientation-preserving homeomorphism, then there is a homotopy $\phi_t\colon \mathbb{S}^1 \to \mathbb{S}^1$ such that $\phi_0 = \phi$ and $\phi_1 = \mathbb{1}_{\mathbb{S}^1}$ and such that each ϕ_t is a homeomorphism. Hint: First reduce to the case where $\phi(1) = 1$. Identify ϕ with a path $\psi\colon I \to \mathbb{S}^1$ with $\psi(0) = \psi(1) = 1$, and then try lifting ϕ to the universal cover $p\colon \mathbb{R} \to \mathbb{S}^1$ of $\mathbb{S}^1$.

(b) Using (a), prove the exercise for f orientation-preserving. Hint: Use the fact[6] that ∂ has a *collar neighborhood*, i.e., an embedding $\iota\colon \partial \times [0,1] \to \Sigma$ such that $\iota(x,0) = x$ for all $x \in \partial$.

(c) Conclude by proving the exercise for f orientation-reversing. Hint: using (b), show that it is enough to exhibit a single orientation-reversing homeomorphism $\phi\colon \Sigma \to \Sigma$ with $\phi(\partial) = \partial$. □

EXERCISE C.11. Let Σ and Σ' be compact connected surfaces with boundary. Assume that Σ is either $\Sigma_{g,b}$ or $\Sigma^{\text{no}}_{n,b}$ with $b \geq 1$. Let ∂ be a boundary component of Σ and let ∂' be a boundary component of Σ'. Let $f_1, f_2\colon \partial \to \partial'$ be two homeomorphisms and let S_i be the result of gluing $\partial \subset \Sigma$ to $\partial' \subset \Sigma'$ using f_i. Prove that S_1 is homeomorphic to S_2. Hint: Exercise C.10 will be helpful. □

Bibliography

[1] L. V. Ahlfors and L. Sario, *Riemann surfaces*, Princeton Mathematical Series, No. 26, Princeton Univ. Press, Princeton, NJ, 1960. (Cited on page 195.)

[2] M. A. Armstrong, *Basic topology*, corrected reprint of the 1979 original, Undergraduate Texts in Mathematics, Springer, New York-Berlin, 1983. (Cited on page 189.)

[3] E. M. Brown and R. Messer, The classification of two-dimensional manifolds, Trans. Amer. Math. Soc. 255 (1979), 377–402. (Cited on page 206.)

[4] G. K. Francis and J. R. Weeks, Conway's ZIP proof, Amer. Math. Monthly 106 (1999), no. 5, 393–399 (Cited on page 189.)

[5] J. Gallier and D. Xu, A guide to the classification theorem for compact surfaces, Geometry and Computing, 9, Springer, Heidelberg, 2013. (Cited on page 189.)

[6] A. Hatcher, The Kirby torus trick for surfaces, to appear in Enseign. Math. (Cited on page 195.)

[6] This fact clearly holds for the Σ in the exercise. In fact, a standard theorem in manifold topology shows that collar neighborhoods exist for boundary components of arbitrary compact manifolds.

[7] B. Kerékjártó, Vorlesungen über Topologie. I, Springer, Berlin, 1923. (Cited on page 206.)
[8] I. Richards, On the classification of noncompact surfaces, Trans. Amer. Math. Soc. 106 (1963), 259–269. (Cited on page 206.)
[9] W. S. Massey, *Algebraic topology: an introduction*, reprint of the 1967 edition, Graduate Texts in Mathematics, Vol. 56, Springer, New York-Heidelberg, 1977. (Cited on page 189.)
[10] E. E. Moise, *Geometric topology in dimensions* 2 *and* 3, Graduate Texts in Mathematics, Vol. 47, Springer, New York-Heidelberg, 1977. (Cited on page 195.)
[11] T. Radó, Uber den Begriff der Riemannschen Fläche, Acta Sci. Math. Szeged. 2 (1925), 101–121. (Cited on page 195.)
[12] H. Seifert and W. Threlfall, *Seifert and Threlfall: a textbook of topology*, translated from the German edition of 1934 by Michael A. Goldman Translated from the German by Wolfgang Heil, Pure and Applied Mathematics, 89, Academic Press, New York-London, 1980. (Cited on page 189.)
[13] E. Zeeman, An introduction to topology: The classification theorem for surfaces, 1966. https://www.maths.ed.ac.uk/~v1ranick/surgery/zeeman.pdf (Cited on page 189.)

ESSAY D

Dimension theory (to be written)

Essays on classical geometric results

ESSAY E

The Brouwer fixed point theorem

This chapter gives an elementary proof of the Brouwer fixed point theorem based on Sperner's Lemma, which is a classic result about the combinatorics of triangulations.

E.1. Simplices and triangulations

Let $P \subset \mathbb{R}^d$ be a nonempty finite subset. Enumerate P as $P = \{\mathbf{p}_1, \ldots, \mathbf{p}_k\}$. An *affine linear combination* of the points of P is a point in $\mathbb{R}^d$ of the form

$$\lambda_1 \mathbf{p}_1 + \cdots + \lambda_k \mathbf{p}_k \in \mathbb{R}^d \quad \text{with } \lambda_1, \ldots, \lambda_k \in \mathbb{R} \text{ such that } \lambda_1 + \cdots + \lambda_k = 1.$$

The *affine hull* of P, denoted $\mathrm{Aff}(P)$, is the collection of all points that are affine linear combinations of the points of P. We say that P is *affinely dependent* if a point can be expressed as an affine linear combination of points of P in two different ways. Here are two equivalent ways of expressing this (see Exercise E.1):

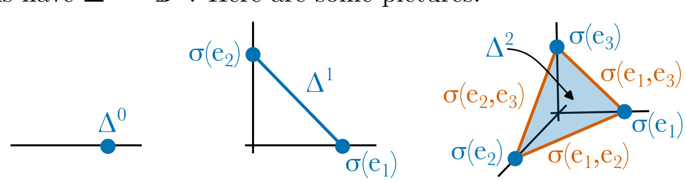
- There exist $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ with $\lambda_1 + \cdots + \lambda_k = 0$ and $(\lambda_1, \ldots, \lambda_k) \neq (0, \ldots, 0)$ such that $\lambda_1 \mathbf{p}_1 + \cdots + \lambda_k \mathbf{p}_k = 0$.
- The vectors $\{\mathbf{p}_2 - \mathbf{p}_1, \ldots, \mathbf{p}_k - \mathbf{p}_1\}$ in $\mathbb{R}^d$ are linearly dependent.

If P is affinely dependent, then there exists some $1 \leq i \leq k$ such that $\mathrm{Aff}(P \setminus p_i) = \mathrm{Aff}(P)$. If P is not affinely dependent, then we say that P is *affinely independent*. The *convex hull* of P is the subspace

$$\sigma(P) = \{\lambda_1 \mathbf{p}_1 + \cdots + \lambda_k \mathbf{p}_k \mid \lambda_1, \ldots, \lambda_1 \geq 0 \text{ and } \lambda_1 + \cdots + \lambda_k = 1\}$$

of $\mathrm{Aff}(P)$. A *k-simplex* is a subspace of $\mathbb{R}^d$ of the form $\sigma(P)$ with P an affinely independent set with $|P| = k+1$. We will just call this a *simplex* if we do not want to specify k. A k-simplex is homeomorphic to $\mathbb{D}^k$ (see Exercise E.2). The *faces* of $\sigma(P)$ are the subspaces of the form $\sigma(P')$ with $P' \subset P$ nonempty. The *dimension* of a face $\sigma(P')$ is its dimension as a simplex, namely $|P'| - 1$.

Letting $B = \{\mathbf{e}_1, \ldots, \mathbf{e}_{n+1}\}$ be the standard basis for $\mathbb{R}^{n+1}$, the *standard n-simplex* is the simplex $\Delta^n = \sigma(B)$. We thus have $\Delta^n \cong \mathbb{D}^n$. Here are some pictures:

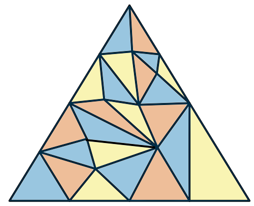


The different faces are indicated.

A *triangulation* of a subspace X of $\mathbb{R}^d$ is a decomposition X into a union of simplices with the following three properties:

- If σ is a simplex of the triangulation, then so are all the faces of σ.
- If σ and σ' are two simplices of the triangulation, then $\sigma \cap \sigma'$ is either empty or is a face of both σ and σ'.
- For each point $p \in X$, there is an open neighborhood U of p that only intersects finitely many simplices in the triangulation. This is automatic if the triangulation has only finitely many simplices, which will be the case in all our examples in this chapter.

Here is an example of a triangulation of a 2-simplex:



Let $\mathfrak{d}$ be the standard distance funtion on $\mathbb{R}^d$. The *diameter* of a simplex σ is

$$\operatorname{diam}(\sigma) = \max\left\{\mathfrak{d}(x,y) \mid x, y \in \sigma\right\}.$$

The *mesh size* of the triangulation is the maximal diameter of any of its simplices. We have:

LEMMA E.1.1. *Fix $n \geq 1$. For all $\epsilon > 0$, there exists a triangulation of Δ^n whose mesh size is at most ϵ.*

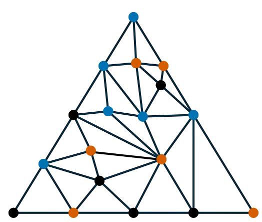
PROOF. See Exercise E.3. □

E.2. Sperner's Lemma

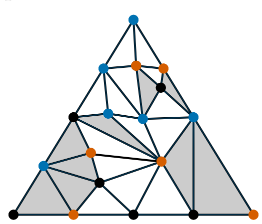
The key to our proofs is a simple combinatorial property of triangulations of simplices. Let $B = \{\mathbf{e}_1, \ldots, \mathbf{e}_{n+1}\}$ be the standard basis for $\mathbb{R}^{n+1}$, so $\Delta^n = \sigma(B)$. Fix a triangulation of Δ^n, and let V be the set of 0-simplices of this triangulation (the *vertices*). A *Sperner coloring* of this triangulation is a set map $s\colon V \to \{1, \ldots, n+1\}$ such that the following holds for each $v \in V$:

- Assume that v lies in a face $\sigma(B')$ for some $B' \subset B$. Write $B' = \{\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_k}\}$. Then $s(v) \in \{i_1, \ldots, i_k\}$.

We have $\mathbf{e}_i \in V$, and this rule implies in particular that $s(\mathbf{e}_i) = i$. We view the labels $\{1, \ldots, n+1\}$ as "colors" for the vertices. Here is an example of a Sperner coloring of a triangulation of Δ^2, with the vertices colored with the colors black and blue and orange to mean that they are labeled with 1 and 2 and 3:



A *rainbow n-simplex* of s is an n-simplex $\sigma(v_1, \ldots, v_{n+1})$ such that $\{s(v_1), \ldots, s(v_{n+1})\} = \{1, \ldots, n+1\}$. In the following the rainbow n-simplices are shaded:



Sperner's lemma says that rainbow n-simplices always exist. In fact, even more is true:

LEMMA E.2.1 (Sperner's Lemma). *Fix a triangulation of Δ^n with vertex set V. Let $s\colon V \to \{1, \ldots, n+1\}$ be a Sperner coloring. Then there are an odd number of rainbow n-simplices of s.*

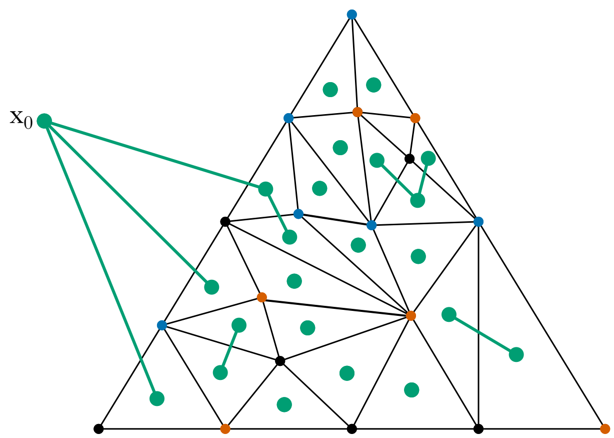
PROOF. The proof will be by induction on n. The base case $n = 0$ is trivial since Δ^0 consists of a single point. Assume now that $n \geq 1$ and that the lemma is true for Δ^{n-1}.

Call an $(n-1)$-simplex $\sigma(v_1, \ldots, v_n)$ of the triangulation a rainbow $(n-1)$-simplex if we have $\{s(v_1), \ldots, s(v_n)\} = \{1, \ldots, n\}$. In particular, none of the $s(v_i)$ equal $n+1$. Identify $\Delta^{n-1} \subset \mathbb{R}^n$ with the face of $\Delta^n \subset \mathbb{R}^{n+1}$ that is the convex hull of the first n standard basis vectors of $\mathbb{R}^{n+1}$. Our triangulation restricts to a triangulation of Δ^{n-1}, and the Sperner coloring of our triangulation restricts to a Sperner coloring of Δ^{n-1}. Our inductive hypothesis thus implies that there are an odd number of rainbow $(n-1)$-simplices in Δ^{n-1}.

Recall that a graph G is a collection of vertices $\mathcal{V}(G)$ connected by edges $\mathcal{E}(G)$. Let G be the following graph:

- The vertices $\mathcal{V}(G)$ consist of the n-simplices σ of the triangulation plus a single additional vertex x_0.
- The edges $\mathcal{E}(G)$ consist of the rainbow $(n-1)$-simplices of the triangulation. The edge corresponding to a rainbow $(n-1)$-simplex f connects the following two vertices:
  - If f is a face of two n-simplices σ and σ' of the triangulation (so f lies in the "interior" of Δ^n), then f connects the vertices corresponding to σ and σ'.
  - If f is the face of only one n-simplex σ of the triangulation (so f lies in the "boundary" of Δ^n), then f connects the vertex corresponding to σ to x_0.

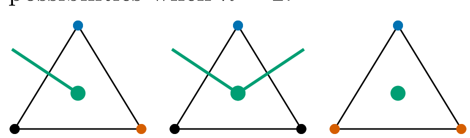
This graph can be visualized by placing a vertex of G in the center of each n-simplex of the triangulation and the vertex x_0 outside of Δ^n and then connecting the vertices according to the above rule. Here is an example, with G drawn in green:



The reader will notice that many vertices of G have no edges coming out of them. As we will soon see, this is not atypical.

Recall that the *degree* of a vertex x of a graph, denoted $\deg(x)$, is the number of edges coming out of x. By definition, $\deg(x_0)$ is the number of rainbow $(n-1)$-simplices f that are the face of exactly one simplex of the triangulation. Since our coloring is a Sperner coloring, a rainbow $(n-1)$-simplex f must lie on the face Δ^{n-1} of Δ^n. As we discussed above, our inductive hypothesis therefore implies that $\deg(x_0)$ is odd.

Let $\mathcal{V}_n(G)$ be the set of vertices corresponding to n-simplices, so $\mathcal{V}(G) = \mathcal{V}_n(G) \cup \{x_0\}$. For $\sigma \in \mathcal{V}_n(G)$, we have $\deg(\sigma) \in \{0, 1, 2\}$ with $\deg(\sigma) = 1$ exactly when σ is a rainbow n-simplex (see Exercise E.4). Here are some possibilities when $n = 2$:



From this, we see that to show there are an odd number of rainbow n-simplices, we must prove that

$$\sum_{\sigma \in \mathcal{V}_n(G)} \deg(\sigma) \equiv 1 \pmod{2}.$$

Since each edges connects two vertices, we have

$$\deg(x_0) + \sum_{\sigma \in \mathcal{V}_n(G)} \deg(\sigma) = \sum_{x \in \mathcal{V}(G)} \deg(x) = 2|\mathcal{E}(G)| \equiv 0 \pmod{2}.$$

It follows that

$$\sum_{\sigma \in \mathcal{V}_n(G)} \deg(\sigma) \equiv \deg(x_0) \equiv 1 \pmod{2},$$

as desired. $\square$

E.3. Brouwer fixed point theorem

We now use Sperner's Lemma to prove the following classic result:

THEOREM E.3.1 (Brouwer fixed point theorem). *Let $f\colon \mathbb{D}^n \to \mathbb{D}^n$ be a map. Then there exists some $x \in \mathbb{D}^n$ such that $f(x) = x$.*

PROOF. Identify $\mathbb{D}^n$ with the standard n-simplex

$$\Delta^n = \{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^n \mid x_1, \ldots, x_{n+1} \geq 0 \text{ and } x_1 + \cdots + x_{n+1} = 1\};$$

see Exercise E.2. Under this identification, f is a map $f\colon \Delta^n \to \Delta^n$. Assume for the sake of contradiction that $f(\mathbf{x}) \neq \mathbf{x}$ for all $\mathbf{x} \in \Delta^n$. We will use this assumption to find $\mathbf{z} \in \Delta^n$ with $f(\mathbf{z}) = \mathbf{z}$ (which is a contradiction, but also what we were trying to find in the first place!).

Define a non-continuous set map

$$s\colon \Delta^n \to \{1, \ldots, n+1\}$$

as follows. Consider $\mathbf{x} = (x_1, \ldots, x_{n+1}) \in \Delta^n$, and let $f(\mathbf{x}) = (y_1, \ldots, y_{n+1})$. Since $f(\mathbf{x}) \neq \mathbf{x}$ and the x_i and y_j are nonnegative real numbers satisfying

$$x_1 + \cdots + x_{n+1} = y_1 + \cdots + y_{n+1} = 1,$$

there must be some $1 \leq i_0 \leq n+1$ such that $y_{i_0} < x_{i_0}$. Choose some such i_0, and define $s(\mathbf{x}) = i_0$.

The map s is sort of like a Sperner coloring in the following sense. Let $B = \{\mathbf{e}_1, \ldots, \mathbf{e}_{n+1}\}$ be the standard basis for $\mathbb{R}^{n+1}$. Then:

(♠) Consider $\mathbf{x} \in \Delta^n$ such that $\mathbf{x}$ lies in a face $\sigma(B')$ for some $B' \subset B$. Write $B' = \{\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_k}\}$. Then $s(\mathbf{x}) \in \{i_1, \ldots, i_k\}$.

Indeed, for $j \notin \{i_1, \ldots, i_k\}$ the $\mathbf{e}_j$-coordinate of $\mathbf{x}$ is 0, so f cannot decrease the $\mathbf{e}_j$-coordinate of $\mathbf{x}$.

For each triangulation of Δ^n, it follows from (♠) that the restriction of s to the vertices of the triangulation is a Sperner coloring. Sperner's Lemma (Lemma E.2.1) implies that our triangulation has a rainbow n-simplex. For each $\ell \geq 1$, apply this to a triangulation with mesh size less than $1/\ell$ (see Lemma E.1.1) and let $\mathbf{x}(\ell)$ be a point in this rainbow n-simplex. Letting $\mathfrak{d}$ be the standard distance function on $\mathbb{R}^{n+1}$, for all $\ell \geq 1$ we then have:

(♣) For each $1 \leq i \leq n+1$, there is a point $\mathbf{x}' \in \Delta^n$ with $\mathfrak{d}(\mathbf{x}(\ell), \mathbf{x}') < 1/\ell$ such that $s(\mathbf{x}') = i$.

By the Heine–Borel theorem (Theorem 1.10.5), the n-simplex Δ^n is compact and thus sequentially compact. Some subsequence of the sequence $\{\mathbf{x}(\ell)\}_{\ell \geq 1}$ therefore converges to a point $\mathbf{z}$. For all $\epsilon > 0$ and $1 \leq i \leq n+1$, it follows from (♣) that the following holds:

(♦) There is a point $\mathbf{z}' \in \Delta^n$ with $\mathfrak{d}(\mathbf{z}, \mathbf{z}') < \epsilon$ such that $s(\mathbf{z}') = i$, i.e., such that the $\mathbf{e}_i$-coordinate of $f(\mathbf{z}')$ is less than the $\mathbf{e}_i$-coordinate of $\mathbf{z}'$.

Let $\mathbf{z} = (z_1, \ldots, z_{n+1})$ and $f(\mathbf{z}) = (w_1, \ldots, w_{n+1})$. From (♦), we deduce that $w_i \leq z_i$ for all $1 \leq i \leq n+1$. Since the z_i and w_i are nonnegative numbers satisfying

$$z_1 + \cdots + z_{n+1} = w_1 + \cdots + w_{n+1} = 1,$$

this implies that $w_i = z_i$ for all $1 \leq i \leq n+1$, i.e., that $f(\mathbf{z}) = \mathbf{z}$, as desired. □

If X is a space and $Y \subset X$ is a subspace, then a *retraction* of X to Y is a map $r\colon X \to Y$ such that $r(y) = y$ for all $y \in Y$, i.e., such that $r|_Y = \mathbb{1}_Y$. Note that $\mathbb{S}^{n-1} \subset \mathbb{D}^n$. The Brouwer fixed point theorem implies the following:

COROLLARY E.3.2. *For $n \geq 1$, there does not exist a retraction $r\colon \mathbb{D}^n \to \mathbb{S}^{n-1}$.*

PROOF. Assume a retraction $r\colon \mathbb{D}^n \to \mathbb{S}^{n-1}$ exists. Let $\alpha\colon \mathbb{S}^{n-1} \to \mathbb{S}^{n-1}$ be the antipodal map:

$$\alpha(\mathbf{x}) = -\mathbf{x} \quad \text{for all } \mathbf{x} \in \mathbb{S}^{n-1} \subset \mathbb{R}^n.$$

Since α has no fixed points, it follows that the map $\alpha \circ r\colon \mathbb{D}^n \to \mathbb{D}^n$ has no fixed points, contradicting the Brouwer fixed point theorem. □

REMARK E.3.3. Almost every tool in algebraic topology can be used to prove the Brouwer fixed point theorem. Most of those proofs actually first establish Corollary E.3.2. See Exercise E.5 for how to derive the Brouwer fixed point theorem from Corollary E.3.2. □

E.4. Exercises

Exercise E.1. Let $P = \{p_1, \ldots, p_k\}$ be a finite set of points in $\mathbb{R}^n$. Prove that the following are equivalent:

- The set P is affinely dependent.
- There exist $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ with $\lambda_1 + \cdots + \lambda_k = 0$ and $(\lambda_1, \ldots, \lambda_k) \neq (0, \ldots, 0)$ such that $\lambda_1 \mathbf{p}_1 + \cdots + \lambda_k \mathbf{p}_k = 0$.
- The vectors $\{\mathbf{p}_2 - \mathbf{p}_1, \ldots, \mathbf{p}_k - \mathbf{p}_1\}$ in $\mathbb{R}^n$ are linearly dependent. □

Exercise E.2. Prove that a k-simplex is homeomorphic to $\mathbb{D}^k$. □

Exercise E.3. Fix $n \geq 1$ and $\epsilon > 0$. Prove that there exists a triangulation of Δ^n whose mesh size is at most ϵ. □

Exercise E.4. Fix a Sperner coloring of a triangulation of Δ^n. Let σ be an n-simplex of the triangulation, and let d be the number of $(n-1)$-dimensional faces of σ that are rainbow $(n-1)$-simplices. Prove that $d \in \{0, 1, 2\}$ with $d = 1$ exactly when σ is a rainbow n-simplex. □

Exercise E.5. Prove that Corollary E.3.2 implies the Brouwer fixed point theorem by constructing from a map $f \colon \mathbb{D}^n \to \mathbb{D}^n$ with $f(\mathbf{x}) \neq \mathbf{x}$ for all $\mathbf{x} \in \mathbb{D}^n$ a retraction $r \colon \mathbb{D}^n \to \mathbb{S}^{n-1}$. Hint: for $\mathbf{x} \in \mathbb{D}^n$, define $r(\mathbf{x})$ to be the intersection point with $\mathbb{S}^{n-1}$ of a ray starting at $f(\mathbf{x})$ and passing through $\mathbf{x}$. □

ESSAY F

The Jordan separation theorem

This essay proves the Jordan separation theorem. As an application, we prove that disks and spheres are never homeomorphic.

F.1. Jordan non-separation theorem

The easiest special case of the Jordan non-separation theorem is that no subspace $X \subset \mathbb{S}^n$ with $X \cong \mathbb{D}^0$ can separate $\mathbb{S}^n$ into multiple path components. Since $\mathbb{D}^0$ is a single point, another way to state this is that for all $p_0 \in \mathbb{S}^n$, the space $\mathbb{S}^n \setminus p_0$ is path connected. In fact, even more is true: $\mathbb{S}^n \setminus p_0 \cong \mathbb{R}^n$ (see Exercise F.2).

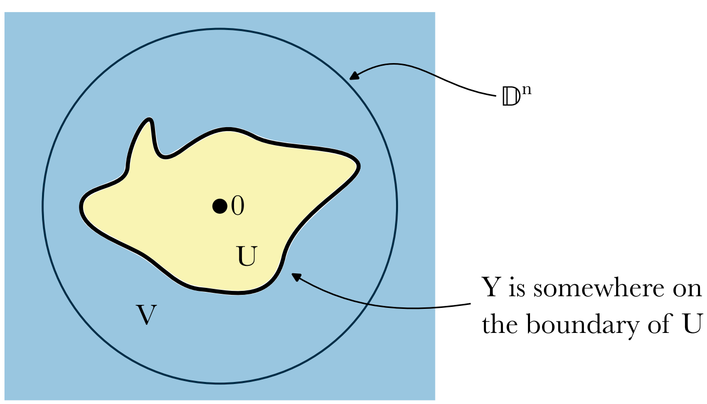
The general case allows X to be homeomorphic to an arbitrary disk. Since a disk can be embedded into $\mathbb{S}^n$ in many ways, this is a far more non-trivial fact. The proof is as follows:

Theorem F.1.1 (Jordan non-separation theorem). *For some $n \geq 1$ and $m \geq 0$, let $X \subset \mathbb{S}^n$ be a closed subspace with $X \cong \mathbb{D}^m$. Then $\mathbb{S}^n \setminus X$ is path-connected.*

Proof. Assume that $\mathbb{S}^n \setminus X$ is not path-connected and let U' be a path component of $\mathbb{S}^n \setminus X$. Set $V' = (\mathbb{S}^n \setminus X) \setminus U'$. Both U' and V' are open subsets of $\mathbb{S}^n$. Pick some $q_0 \in V'$. We have $\mathbb{S}^n \setminus q_0 \cong \mathbb{R}^n$ (see Exercise F.2). Pick a homeomorphism $h\colon \mathbb{S}^n \setminus q_0 \to \mathbb{R}^n$ and set $Y = h(X)$ and $U = h(U')$ and $V = h(V' \setminus q_0)$. The following hold (see Exercise F.2):

- Y is a closed subset of $\mathbb{R}^n$ satisfying $Y \cong \mathbb{D}^m$; and
- U is a path component of $\mathbb{R}^n \setminus Y$ that is bounded[1] and $V = (\mathbb{R}^n \setminus Y) \setminus U$.

Translating everything and rescaling, we can assume that $0 \in U$ and that U is contained in the unit disk $\mathbb{D}^n$. The picture is as follows:



Since $\mathbb{R}^n$ is locally path connected and $\mathbb{R}^n \setminus Y$ is an open subspace of $\mathbb{R}^n$, it follows that $\mathbb{R}^n \setminus Y$ is locally path-connected. This implies that the path components of $\mathbb{R}^n \setminus Y$ are the same as the connected components and that the connected components are both open and closed. It follows that U is both open and closed in $\mathbb{R}^n \setminus Y$. In particular, there is a closed set $C \subset \mathbb{R}^n$ such that

$$U = C \cap (\mathbb{R}^n \setminus Y) = C \setminus Y.$$

From this we see that

$$A = U \cup Y = (C \setminus Y) \cup Y = C \cup Y$$

[1] That is, U is contained in some large ball around the origin.

is closed in $\mathbb{R}^n$. Similarly, the set $B = V \cup Y$ is closed. By construction, A and B are nonempty closed sets with $\mathbb{R}^n = A \cup B$ and $A \cap B = Y$.

Recall that $I = [0, 1]$. Since $Y \cong \mathbb{D}^m$ and $\mathbb{D}^m \cong I^m$ (see Exercise F.1), a map $f\colon \mathbb{R}^n \to Y$ can be viewed as an m-tuple of maps $\mathbb{R}^n \to I$. The space $\mathbb{R}^n$ is normal, so the Tietze extension applies to it. Applying the Tietze extension theorem m times, we can therefore extend the identity map $\mathbb{1}_Y\colon Y \to Y$ to a continuous map $f\colon \mathbb{R}^n \to Y$. Define $r\colon \mathbb{R}^n \to \mathbb{R}^n$ to be the map

$$r(p) = \begin{cases} f(p) & \text{if } p \in A, \\ p & \text{if } p \in B. \end{cases}$$

For $p \in A \cap B = Y$, we have $f(p) = p$, so this definition makes sense and defines a continuous function (see Exercise 12.5). By construction, r is a retraction from $\mathbb{R}^n$ to B. There also exists a retraction $r'\colon \mathbb{D}^n \setminus 0 \to \mathbb{S}^{n-1}$ (see Exercise F.3). The composition

$$\mathbb{D}^n \xrightarrow{r|_{\mathbb{D}^n}} B \cap \mathbb{D}^n \xrightarrow{r'} \mathbb{S}^{n-1}$$

is then a retraction. Using the Brouwer fixed point theorem, we proved in Essay E that no such retraction exists (see Corollary E.3.2), so this is a contradiction. □

Remark F.1.2. Deeper than Theorem F.1.1 is the Jordan separation theorem, which says that for any closed subspace $X \subset \mathbb{S}^n$ with $X \cong \mathbb{S}^{n-1}$ the space $\mathbb{S}^n \setminus X$ has two path components. It is not hard to see that this implies a similar result with $\mathbb{S}^n$ replaced by $\mathbb{R}^n$. A special case of this is the Jordan curve theorem, which says that any embedded circle in $\mathbb{R}^2$ separates $\mathbb{R}^2$ into two path components.

I am not aware of any proof of the general high-dimensional Jordan separation theorem that does not use tools from algebraic topology. However, there are elementary proofs of the Jordan curve theorem. We will give one in Essay G. □

F.2. Distinguishing spheres and disks

F.3. Exercises

Exercise F.1. Prove that $\mathbb{D}^k$ is homeomorphic to I^k. □

Exercise F.2. WRITE IT!!! □

Exercise F.3. WRITE IT!!! □

ESSAY G

The Jordan curve theorem (to be written)

Essays on classes of spaces

ESSAY H

CW complexes (to be written)

ESSAY I

Compactly generated spaces (to be written)

Essays on groups and group actions

ESSAY J

Quotients by group actions

This essay discusses the point-set topology of group actions. Our exposition is influenced by the paper [1], to which we refer the reader for many more results along these lines.

J.1. Group actions

Let G be a group and X be a space. In this chapter, actions of G on X are always assumed to be continuous. This means that a (left) action of G on X consists of a left action of G on the points of X such that for all $g \in G$ the left-multiplication map $m_g \colon X \to X$ defined by

$$m_g(x) = g \cdot x \quad \text{for } g \in G \text{ and } x \in X$$

is continuous. Since $m_1 = m_{g^{-1}g} = m_{g^{-1}} m_g$, the map m_g is necessarily a homeomorphism.

For a subset $Y \subset X$, let

$$G \cdot Y = \{ g \cdot y \mid g \in G \text{ and } y \in Y \}.$$

A G-orbit of X is a set of the form $G \cdot x_0$ for some $x_0 \in X$. For $Y \subset X$, the set $G \cdot Y$ is a union of G-orbits. We write X/G for the space of orbits of G acting on X, equipped with the quotient topology. Letting $\pi \colon X \to X/G$ be the projection, a set $U \subset X/G$ is thus open if and only if $\pi^{-1}(U) \subset X$ is open. Here are some examples:

Example J.1.1. The group $\mathbb{Z}$ acts on $\mathbb{R}$ by integer translations, and $\mathbb{R}/\mathbb{Z} \cong \mathbb{S}^1$. □

Example J.1.2. The cyclic group C_2 of order 2 acts on $\mathbb{S}^n$ via the antipodal map. In other words, if $t \in C_2$ is the generator and $p \in \mathbb{S}^n \subset \mathbb{R}^{n+1}$, then $t \cdot p = -p$. The quotient $\mathbb{S}^n / C_2$ is the space of pairs of antipodal points on $\mathbb{S}^n$. A line in $\mathbb{R}^{n+1}$ though the origin intersects $\mathbb{S}^n$ in a pair of antipodal points, so we can identify $\mathbb{S}^n / C_2$ with the space of lines through the origin in $\mathbb{R}^{n+1}$. This space of lines is called *real projective space*, and is denoted $\mathbb{RP}^n = \mathbb{S}^n / C_2$. □

J.2. Easy facts about quotients

Many of our results will depend on the following observation:

Lemma J.2.1. *Let G be a group acting on a space X. Then the projection $\pi \colon X \to X/G$ is an open map.*

Proof. See Exercise J.2. □

Using this, we see that many properties of spaces pass to quotients by group actions:

Lemma J.2.2. *Let G be a group acting on a space X. The following hold:*

(a) If X is path connected (resp. connected), then X/G is path-connected (resp. connected).
(b) If X is first countable (resp. second countable, resp. separable), then X/G is first countable (resp. second countable, resp. separable).
(c) If X is compact, then X/G is compact.
(d) If X is locally compact, then X/G is locally compact.

Proof. Let $\pi \colon X \to X/G$ be the projection. These are all consequences of the following facts:

- The map π is a continuous open map (Lemma J.2.1).
- Like all continuous maps, π takes compact sets to compact sets.

See Exercise J.3. □

Lemma J.2.2 does *not* say anything about Hausdorff or normal spaces, and as we will see even fairly tame group actions do not always preserve these properties.

J.3. Finite group actions

We start by observing that these separation properties are always preserved if G is finite:

LEMMA J.3.1. *Let X be a space and let G be a finite group acting on X. Then:*

(i) If X is Hausdorff, then X/G is Hausdorff.
(ii) If X is normal, then X/G is normal.

PROOF. We prove (i) and leave (ii) as Exercise J.4. Assume that X is Hausdorff. Let $\pi\colon X \to X/G$ be the quotient map. Consider $x, y \in X$ such that $\pi(x) \neq \pi(y)$. We must find disjoint open neighborhoods of $\pi(x)$ and $\pi(y)$. Let

$$S = G{\cdot}x = \{g{\cdot}x \mid g \in G\} \quad \text{and} \quad T = G{\cdot}y = \{g{\cdot}y \mid g \in G\}\,.$$

By assumption, S and T are disjoint finite subsets of X. Since X is Hausdorff, we can find open neighborhoods U of S and V of T with $U \cap V = \emptyset$ (see Exercise 16.1). For all $g \in G$, the open set $g{\cdot}U$ is an open neighborhood of S and the open set $g{\cdot}V$ is an open neighborhood of T. Letting

$$U' = \bigcap_{g\in G} g{\cdot}U \quad \text{and} \quad V' = \bigcap_{g\in G} g{\cdot}V,$$

it follows that U' is an open neighborhood of S and V' is an open neighborhood of T. By construction, we have $U' \cap V' = \emptyset$ and also $g{\cdot}U' = U'$ and $g{\cdot}V' = V'$ for all $g \in G$. Since the quotient map $\pi\colon X \to X/G$ is an open map (Lemma J.2.1), we conclude that $\pi(U')$ and $\pi(V')$ are disjoint open neighborhoods of $\pi(x)$ and $\pi(y)$, as desired. $\square$

J.4. Free actions, and why they can be pathological

Nothing like Lemma J.3.1 can be true for actions of infinite groups. Let G be a group acting on a space X. For $x \in X$, let G_x be the stabilizer of x, i.e.,

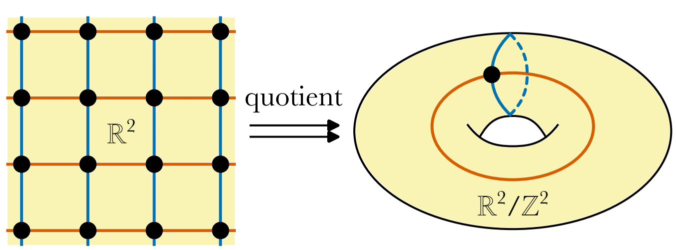
$$G_x = \{g \in G \mid g{\cdot}x = x\}\,.$$

One natural condition to impose on an action of a group G on a space X is that it is *free*, that is, that $G_x = 1$ for all $x \in X$. Here is an easy example showing that this can still be pathological:

EXAMPLE J.4.1. Let $\mathbb{Q}$ act by $\mathbb{R}$ by translations. Then $\mathbb{R}/\mathbb{Q}$ is not Hausdorff (see Exercise J.5). $\square$

A reader might think that the source of this pathology is that $\mathbb{Q}$ is "too large". For instance, though $\mathbb{Q}$ is countable it is not finitely generated. Here is an example showing that even for the simplest infinite group $\mathbb{Z}$ we can still have pathological free quotients:

EXAMPLE J.4.2. Consider the 2-torus $\mathbb{T}^2 = (\mathbb{S}^1)^{\times 2}$. As the following figure shows, we can identify $\mathbb{T}^2$ with $\mathbb{R}^2/\mathbb{Z}^2$ with $\mathbb{Z}^2$ acting on $\mathbb{R}^2$ by integer translations:



Of course, this action is not pathological. Let $G = \mathbb{Z}$ with generator $t = 1$. Fix an irrational number $\alpha \in \mathbb{R}$, and define an action of G on $\mathbb{R}^2$ via the formula

$$t^n{\cdot}(x, y) = (x + n, y + n\alpha) \quad \text{for all } (x, y) \in \mathbb{R}^2 \text{ and } n \in \mathbb{Z}.$$

This action commutes with the action of $\mathbb{Z}^2$, and thus descends to an action of G on $\mathbb{R}^2/\mathbb{Z}^2 = \mathbb{T}^2$. This action is free (see Exercise J.6). It turns out that all the G-orbits in $\mathbb{T}^2$ are dense and that $\mathbb{T}^2/G$ is not Hausdorff (see Exercise J.6). $\square$

J.5. Covering space actions, and why they can be pathological

From the discussion above, we see that we have to avoid dense orbits. One way to exclude them is as follows. An action of a group G on a space X is a *covering space action*[1] if for all $x \in X$ the following holds:

- There exists an open neighborhood U of x such that $U \cap g{\cdot}U = \emptyset$ for all $g \in G$ with $g \neq 1$. Equivalently, $g_1{\cdot}U \cap g_2{\cdot}U = \emptyset$ for all distinct $g_1, g_2 \in G$.

Covering space actions are clearly free. We have already seen many covering space actions:

- the action of $\mathbb{Z}$ on $\mathbb{R}$ by translations with $\mathbb{R}/\mathbb{Z} = \mathbb{S}^1$ (Example J.1.1); and
- the action of C_2 on $\mathbb{S}^n$ with $\mathbb{S}^n/C_2 = \mathbb{RP}^n$ (Example J.1.2); and
- the action of $\mathbb{Z}^2$ on $\mathbb{R}^2$ by translations with $\mathbb{R}^2/\mathbb{Z}^2 = \mathbb{T}^2$ (Example J.4.2).

These examples might suggest that these actions are not pathological. Unfortunately, they can still have pathological quotients. Here is an example:

EXAMPLE J.5.1. Let $X = \mathbb{R}^2 \setminus 0$ and let $G = \mathbb{Z}$ with generator $t = 1$. Define an action of G on X as follows:

$$t^n{\cdot}(x, y) = (2^n x, 2^{-n} y) \quad \text{for all } (x, y) \in X \text{ and } n \in \mathbb{Z}.$$

This is a covering space action (see Exercise J.8). Let $\pi\colon X \to X/G$ be the quotient. The points $\pi(1, 0)$ and $\pi(0, 1)$ then do *not* have disjoint open neighborhoods (see Exercise J.8), so X/G is not Hausdorff. □

REMARK J.5.2. Let $(M, \mathfrak{d})$ be a metric space and let G be a group acting on M. We say that G acts by *isometries* if

$$\mathfrak{d}(g{\cdot}x, g{\cdot}y) = \mathfrak{d}(x, y) \quad \text{for all } x, y \in M \text{ and } g \in G.$$

Below in Lemma J.7.2 we will prove that if G acts by isometries via a covering space action, then M/G is Hausdorff. In other words, we can avoid pathological examples like Example J.5.1 if we insist that our groups preserve a metric. □

J.6. Proper actions

Consider a group G acting on a space X. Regard G as a discrete topological space, and consider the map $\sigma\colon G \times X \to X \times X$ defined by

$$\sigma(g, x) = (x, g{\cdot}x) \quad \text{for all } g \in G \text{ and } x \in X.$$

We call σ the *action map* of the group action. In terms of the action map, we have the following criterion for X/G to be Hausdorff:

LEMMA J.6.1. *Let G be a group acting on a space X. Assume that the image of the action map $\sigma\colon G \times X \to X \times X$ is closed. Then X/G is Hausdorff.*[2]

PROOF. Let $\pi\colon X \to X/G$ be the projection. Consider $x, y \in X$ with $\pi(x) \neq \pi(y)$. We must find disjoint open neighborhoods of $\pi(x)$ and $\pi(y)$. Let $E \subset X \times X$ be the image of σ, so E is closed. Since $\pi(x) \neq \pi(y)$, we have $(x, y) \notin E$. Since E is closed, we can find open neighborhoods U of x and V of y such that the open neighborhood $U \times V$ of (x, y) is disjoint from E. Since $U \times V$ is disjoint from E, the sets $\pi(U)$ and $\pi(V)$ are disjoint. Since π is an open map (Lemma J.2.1), the sets $\pi(U)$ and $\pi(V)$ are disjoint open neighborhoods of $\pi(x)$ and $\pi(y)$. □

The action of G on X is a *proper action* if its action map is a proper map. We also say that G acts *properly* on X. As we will see below, there are simple criteria that imply this. Since proper maps are closed, Lemma J.6.1 implies:

LEMMA J.6.2. *Let G be a group acting properly on a space X. Then X/G is Hausdorff.*

The author of this book does not know if the following holds:

QUESTION J.6.3. Let G be a group acting properly on a normal space X. Is it necessarily the case that X/G is normal? □

[1] The terminology will be explained in a later volume when we discuss covering spaces.

[2] It is a little strange that we do not need to assume that X is Hausdorff.

However, the following result gives something stronger in many cases.

Corollary J.6.4. *Let G be a group acting properly on a second countable locally compact metrizable space X. Then X/G is a second countable locally compact metrizable space.*

Proof. Lemma J.2.2 implies that X/G is second countable and locally compact. To see that it is metrizable, note that Lemma J.6.2 implies that X/G is Hausdorff. Since second countable locally compact Hausdorff spaces are regular (Lemma 18.2.1), the space X is regular and the Urysohn metrization theorem (Theorem 22.2.1) applies and shows it is metrizable. $\square$

Remark J.6.5. This applies in particular if X is either an open or a closed subspace of $\mathbb{R}^n$. $\square$

Remark J.6.6. The local compactness assumption can be removed if for some metric on X the group G acts by isometries. See Exercise J.9. $\square$

J.7. Criteria for proper actions

We now give two criteria that ensure that a group action is proper. Confusingly, these different criteria are both called "proper discontinuity" in the literature. The first criterion is:

Lemma J.7.1. *Let X be a Hausdorff space that is either first countable or locally compact. Let G be a group acting on X such that the following holds:*

- *For all compact subsets $K \subset X$, the set $\{g \in G \mid K \cap g{\cdot}K \neq \emptyset\}$ is finite.*

Then G acts properly on X. Consequently, X/G is Hausdorff.

Proof. Let $\sigma\colon G \times X \to X \times X$ be the action map and let $L \subset X \times X$ be compact. By Lemma 19.2.2, it is enough to show that $\sigma^{-1}(L)$ is compact. Let $\pi_1\colon X \times X \to X$ and $\pi_2\colon X \times X \to X$ be the two projections. Set $K = \pi_1(L) \cup \pi_2(L)$. The set K is compact and $L \subset K \times K$. Since X is Hausdorff, the set L is closed and thus $\sigma^{-1}(L)$ is a closed subset of $\sigma^{-1}(K \times K)$. It is therefore enough to prove that $\sigma^{-1}(K \times K)$ is compact. For this, it follows from the definition of σ that $\sigma^{-1}(K \times K) = \Lambda \times K$ with

$$\Lambda = \{g \in G \mid K \cap g{\cdot}K \neq \emptyset\}.$$

By assumption Λ is finite, so since K is compact it follows that $\Lambda \times K$ is compact, as desired. $\square$

For our second result, recall that a group G acts on a metric space $(M, \mathfrak{d})$ by isometries if

$$\mathfrak{d}(g{\cdot}x, g{\cdot}y) = \mathfrak{d}(x, y) \quad \text{for all } x, y \in M \text{ and } g \in G.$$

Also, for $x \in M$ and $r > 0$ recall that $B_r(x)$ denotes the open r-ball around x. We then have:

Lemma J.7.2. *Let $(M, \mathfrak{d})$ be a metric space. Let G be a group acting on M by isometries such that the following holds:*

- *For all $x \in M$, there exists some $\epsilon > 0$ such that $\{g \in G \mid B_\epsilon(x) \cap g{\cdot}B_\epsilon(x) \neq \emptyset\}$ is finite.*

Then G acts properly on M. Consequently, M/G is Hausdorff.

Remark J.7.3. Lemma J.7.2 applies in particular to actions by isometries on metric spaces that are covering space actions. $\square$

Proof of Lemma J.7.2. Let $K \subset M$ be compact. Set $\Lambda = \{g \in G \mid K \cap g{\cdot}K \neq \emptyset\}$. By Lemma J.7.1, it is enough to prove that Λ is finite. For the sake of contradiction, assume that Λ is infinite. We can thus find a sequence $\{x_n\}_{n\geq 1}$ of points of K and a sequence $\{g_n\}_{n\geq 1}$ of distinct elements of Λ such that $g_n{\cdot}x_n \in K$ for all $n \geq 1$. Since K is a compact subset of the metric space M, it follows that K is sequentially compact. Passing to subsequences, we can therefore assume that there are $y, z \in K$ such that

$$\lim_{n\mapsto\infty} x_n = y \quad \text{and} \quad \lim_{n\mapsto\infty} g_n{\cdot}x_n = z.$$

Using our assumption, we can find some $\epsilon > 0$ such that

$$\{g \in G \mid B_\epsilon(y) \cap g{\cdot}B_\epsilon(y) \neq \emptyset\} \tag{J.7.1}$$

is finite. Choose $N > 0$ such that for $n \geq N$ we have $\mathfrak{d}(x_n, y) < \epsilon/2$ and $\mathfrak{d}(g_n{\cdot}x_n, z) < \epsilon/4$. For $n \geq N$, we thus have $x_n \in B_\epsilon(y)$. Also, using the fact that G acts by isometries we have

$$\begin{aligned}\mathfrak{d}(g_N^{-1}g_n{\cdot}x_n, y) &\leq \mathfrak{d}(g_N^{-1}g_n{\cdot}x_n, x_N) + \mathfrak{d}(x_N, y)\\ &= \mathfrak{d}(g_n{\cdot}x_n, g_N{\cdot}x_N) + \mathfrak{d}(x_N, y)\\ &\leq \mathfrak{d}(g_n{\cdot}x_n, z) + \mathfrak{d}(z, g_N{\cdot}x_N) + \mathfrak{d}(x_N, y)\\ &< \epsilon/4 + \epsilon/4 + \epsilon/2 = \epsilon,\end{aligned}$$

so $g_N^{-1}g_n{\cdot}x_n \in B_\epsilon(y)$. In other words, for $n \geq N$ we have $B_\epsilon(y) \cap g_N^{-1}g_n{\cdot}B_\epsilon(y) \neq \emptyset$. By assumption the $g_N^{-1}g_n$ are all distinct, so this contradicts the fact that (J.7.1) is finite. □

J.8. Exercises

EXERCISE J.1. For the following groups G acting on spaces X, identify X/G:

(a) The cyclic group $G = C_2$ of order 2 generated by $t \in C_2$ acting on $\mathbb{R}$ via
$$t{\cdot}x = -x \quad \text{for all } x \in \mathbb{R}.$$

(b) The group $G = \mathbb{Z}$ generated by $t = 1$ acting on $X = \mathbb{R}^n \setminus 0$ by
$$t^n{\cdot}x = 2^n x \quad \text{for all } n \in \mathbb{Z} \text{ and } x \in X.$$

(c) The cyclic group $G = C_2$ of order 2 generated by $t \in C_2$ acting on $X = \mathbb{S}^n$ via
$$t{\cdot}(x_1, \ldots, x_{n+1}) = (-x_1, x_2, \ldots, x_{n+1}) \quad \text{for all } (x_1, \ldots, x_{n+1}) \in \mathbb{S}^n \subset \mathbb{R}^{n+1}.$$ □

EXERCISE J.2. Let X be a space and let G be a group acting on X. Prove that the projection $\pi\colon X \to X/G$ is an open map. □

EXERCISE J.3. Let G be a group acting on a space X. Prove the following:

(a) If X is path connected, then X/G is path-connected.
(b) If X is connected, then X/G is connected.
(c) If X is first countable, then X/G is first countable.
(d) If X is second countable, then X/G is second countable.
(e) If X is separable, then X/G is separable.
(f) If X is compact, then X/G is compact.
(g) If X is locally compact, then X/G is locally compact. □

EXERCISE J.4. Let X be a normal space and let G be a finite group acting on X. Prove that X/G is normal. □

EXERCISE J.5. Let $\mathbb{Q}$ act by $\mathbb{R}$ by translations. Prove that $\mathbb{R}/\mathbb{Q}$ is not Hausdorff. □

EXERCISE J.6. Let $G = \mathbb{Z}$ with generator $t = 1$. Fix an irrational number $\alpha \in \mathbb{R}$, and define an action of G on $\mathbb{R}^2$ via the formula
$$t^n{\cdot}(x, y) = (x + n, y + n\alpha) \quad \text{for all } (x, y) \in \mathbb{R}^2 \text{ and } n \in \mathbb{Z}.$$
This action commutes with the action of $\mathbb{Z}^2$, and thus descends to an action of G on $\mathbb{R}^2/\mathbb{Z}^2 = \mathbb{T}^2$. Prove the following:

(a) The action of G on $\mathbb{T}^2$ is free.
(b) For all $p \in \mathbb{T}^2$, the G-orbit $G{\cdot}p = \{g{\cdot}p \mid g \in G\}$ is dense in $\mathbb{T}^2$.
(c) The quotient $\mathbb{T}^2/G$ is not Hausdorff. □

EXERCISE J.7. Let $X = \mathbb{R}^2 \setminus 0$ and let $G = \mathbb{Z}$ with generator $t = 1$. Define an action of G on X as follows:
$$t^n{\cdot}(x, y) = (2^n x, 2^{-n} y) \quad \text{for all } (x, y) \in X \text{ and } n \in \mathbb{Z}.$$
Prove the following:

(a) This is a covering space action.
(b) Letting $q\colon X \to X/G$ be the quotient, the points $q(1, 0)$ and $q(0, 1)$ do not have disjoint open neighborhoods. □

EXERCISE J.8. Let $X = \mathbb{R}^2 \setminus 0$ and let $G = \mathbb{Z}$ with generator $t = 1$. Define an action of G on X as follows:

$$t^n \cdot (x, y) = (2^n x, 2^{-n} y) \quad \text{for all } (x, y) \in X \text{ and } n \in \mathbb{Z}.$$

Let $q\colon X \to X/G$ be the quotient. Prove the following:

(a) The action of G on X is a covering space action.

(b) The points $q(1, 0)$ and $q(0, 1)$ then do *not* have disjoint open neighborhoods, so X/G is not Hausdorff. □

EXERCISE J.9. Let $(M, \mathfrak{d})$ be a second countable metric space and let G be a group acting on M by isometries. Assume that the action of G on M is proper, so M/G is Hausdorff (Lemma J.6.2).

(a) Prove that M/G is regular.

(b) Prove that M/G is second countable and metrizable. □

Bibliography

[1] M. Kapovich, A note on properly discontinuous actions, São Paulo J. Math. Sci. 18 (2024), no. 2, 807–836. (Cited on page 233.)

Essays on miscellaneous topics

ESSAY K

Some set-theoretic technicalities (to be written)