# A2: Metric Spaces

Ben Green

# Contents

These notes cover the first ten lectures of A2: Metric Spaces and Complex Analysis, which deals with the theory of Metric Spaces. In preparing these notes I made considerable use of the previous notes for this section of the course, written by Kevin McGerty.

**Synopsis**

Basic definitions: metric spaces, isometries, continuous functions ($\varepsilon$-$\delta$ definition), homeomorphisms, open sets, closed sets. Examples of metric spaces, including metrics derived from a norm on a real vector space, particularly $\ell^1, \ell^2, \ell^\infty$-norms on $\mathbf{R}^n$, the sup norm on the bounded real-valued functions on a set, and on the bounded continuous real-valued functions on a metric space. The characterisation of continuity in terms of the pre-image of open sets or closed sets. The limit of a sequence of points in a metric space. A subset of a metric space inherits a metric. Discussion of open and closed sets in subspaces. The closure of a subset of a metric space.

Completeness (but not completion). Completeness of the space of bounded real-valued functions on a set, equipped with the norm, and the completeness of the space of bounded continuous real-valued functions on a metric space, equipped with the metric. Lipschitz maps and contractions. Contraction Mapping Theorem.

Connected metric spaces, path-connectedness. Closure of a connected space is connected, union of connected sets is connected if there is a non-empty intersection, continuous image of a connected space is connected. Path-connectedness implies connectedness. Connected open subset of a normed vector space is path-connected.

Definition of sequential compactness and proof of basic properties of sequentially compact sets. Preservation of sequential compactness under continuous maps, equivalence of continuity and uniform continuity for functions on a sequentially compact set. Equivalence of sequential compactness with being complete and totally bounded. The Arzelà-Ascoli theorem (proof non-examinable). Open cover definition of compactness. Heine-Borel (for the interval only) and proof that compactness implies sequential compactness (statement of the converse only).

*Important notes for 2021/22.* The synopsis above differs very slightly from the published one, in that I have replaced "compact" with "sequentially compact" on a few occasions. These notions turn out to be the same for metric spaces (as we will show) but in my opinion using the words interchangably is nonstandard and potentially confusing. I will fix the synopsis in future years.

I have omitted the number of lectures, since the division into, and timings of, video lectures will be quite different. This part of the course should be thought of as comprising roughly one third of A2: Metric Spaces and Complex Analysis, and would normally be scheduled for 10 lectures.

CHAPTER 1

# Metric spaces

## 1.1. The real numbers and the axiom of choice

*The real numbers.* I will assume familiarity with the real numbers $\mathbf{R}$ as discussed at some length in the Prelims course Analysis I. I will not repeat the long list of axioms for the real numbers here. The most important properties we shall need are

- Any non-empty, bounded subset $S \subseteq \mathbf{R}$ has a least upper bound $\sup(S)$, which is a real number $c$ such that $x \leqslant c$ for all $x \in S$, and such that if $c'$ is any other number with this property then $c' \geqslant c$;
- Similarly, any non-empty, bounded subset $S \subseteq \mathbf{R}$ has a greatest lower bound $\inf(S)$;
- (Bolzano-Weierstauss) Any bounded subsequence of the reals has a convergent subsequence;
- Any Cauchy sequence of real numbers converges.

It might be a good idea to remind yourself of the precise meaning of these statements now, though we shall be going over the last two points in a more general context later in the course.

The Prelims course Analysis I assumed that the real numbers exist. This is not, by any means, obvious! We will also assume they exist. For some comments on how they can actually be constructed, see Section 6.5 (which is non-examinable).

*The axiom of choice.* The following statement, used for example in the proof of Corollary 5.1.5, seems very uncontroversial: given nonempty subsets $S_1, S_2, \ldots$ of some set $X$, we may find a sequence $(x_n)_{n=1}^{\infty}$ with $x_n \in S_n$ for all $n$. One might have thought that this is the most trivial induction imaginable: pick $x_1 \in S_1$, then pick $x_2 \in S_2$, and so on. This does indeed show that there are $x_1, \ldots, x_N$ with $x_n \in S_n$ for $n = 1, \ldots, N$, but it does *not* show the infinitary statement about the existence of a sequence. In fact, the existence of a sequence $(x_n)_{n=1}^{\infty}$ with $x_n \in S_n$ for all $n$ has the status of a separate axiom of mathematics, called the *axiom of countable choice*.

You can learn much more about this and, more particularly, the axiom of choice itself in the course *B1.2: Set Theory*. However, the introduction of the Wikipedia page on the Axiom of Choice is a good read.

## 1.2. The definition of a metric space

One of the key definitions of Analysis I was that of the *continuity* of a function. Recall that if $f\colon \mathbf{R} \to \mathbf{R}$ is a function, we say that $f$ is continuous at $a \in \mathbf{R}$ if, for any $\epsilon > 0$, we can find a $\delta > 0$ such that if $|x - a| < \delta$ then $|f(x) - f(a)| < \epsilon$.

Stated somewhat more informally, this means that no matter how small an $\epsilon$ we are given, we can ensure $f(x)$ is within distance $\epsilon$ of $f(a)$ provided we demand $x$ is sufficiently close to – that is, within distance $\delta$ of – the point $a$.

Now consider what it is about real numbers that we need in order for this definition to make sense: Really we just need, for any pair of real numbers $x_1$ and $x_2$, a measure of the *distance* between them. (Note that we needed this notion of distance in the above definition of continuity for both the pairs $(x, a)$ and $(f(x), f(a))$.) Thus we should be able to talk about continuous functions $f\colon X \to X$ on any set $X$ provided it is equipped with a notion of distance. Even more generally, provided we have prescribed a notion of distance on two sets $X$ and $Y$, we should be able to say what it means for a function $f\colon X \to Y$ to be continuous. In order to make this precise, we will therefore need to give a mathematically rigorous definition of what a "notion of distance" on a set $X$ should be. This is the concept of a metric space.

Definition 1.2.1. Let $X$ be a set. Then a *distance function* on $X$ is a function $d\colon X \times X \to \mathbf{R}$ with the following properties:

(i) (positivity) $d(x, y) \geqslant 0$ and $d(x, y) = 0$ if and only if $x = y$;
(ii) (symmetry) $d(x, y) = d(y, x)$;
(iii) (triangle inequality) if $x, y, z \in X$ then we have $d(x, z) \leqslant d(x, y) + d(y, z)$.

The pair $(X, d)$ consisting of a set $X$ together with a distance function $d$ on it is called a *metric space*.

*Remark.* Often we will not be quite so formal, and will refer to $X$ (rather than the pair $(X, d)$) as a metric space. However, it is important to note that the same space $X$ can have many different distances on it, and in fact that different distances on the same space $X$ can have wildly differing properties.

Occasionally, we will be *more* formal, for instance when we have two metric spaces $(X, d_X)$ and $(Y, d_Y)$ and wish to make it clear which distance we are talking about.

The axioms that a distance function $d$ is required to satisfy are very basic, and one feels that any "reasonable" notion of distance ought to satisfy these properties. This, coupled with the fact that using just these axioms one can develop a satisfactory theory of continuity of functions – as well as many other things – is the point of the definition.

Before moving on, let us record one very simple but useful equivalent form of the triangle inequality, sometimes (but not by me) known as the reverse triangle inequality.

LEMMA 1.2.2. *Let $x, y, z$ be points in a metric space. Then we have $|d(x,y) - d(x,z)| \leqslant d(y,z)$.*

*Proof.* This is two inequalities in one, namely the inequality $d(x,y) - d(x,z) \leqslant d(y,z)$, and the inequality $d(x,z) - d(x,y) \leqslant d(y,z)$. Both are instances of (in fact, equivalent to) the triangle inequality. □

## 1.3. Some examples of metric spaces

In this section we look at some examples of metric spaces. A very basic example is that of the real numbers.

EXAMPLE 1.3.1. Take $X = \mathbf{R}$ and $d(x,y) = |x - y|$.

Let us generalise this to higher dimensions. In fact, there is no "obvious" generalisation. Here are several natural ones.

EXAMPLE 1.3.2. Take $X = \mathbf{R}^n$. Then each of the following functions define metrics on $X$.

$$d_1(v,w) = \sum_{i=1}^{n} |v_i - w_i|;$$

$$d_2(v,w) = \big(\sum_{i=1}^{n} (v_i - w_i)^2\big)^{1/2}$$

$$d_\infty(v,w) = \max_{i \in \{1,2,\ldots,n\}} |v_i - w_i|.$$

These are called the $\ell^1$- ("ell one"), $\ell^2$- (or Euclidean) and $\ell^\infty$-distances respectively. Of course, the Euclidean distance is the most familiar one.

The proof that each of $d_1, d_2, d_\infty$ defines a distance is mostly very routine, with the exception of proving that $d_2$, the Euclidean distance, satisfies the triangle inequality. To establish this, recall that the Euclidean norm $\|v\|_2$ of a vector $v = (v_1, \ldots, v_n) \in \mathbf{R}^n$ is

$$\|v\|_2 := \big(\sum_{i=1}^{n} |v_i|^2\big)^{1/2} = \langle v, v \rangle^{1/2},$$

where the inner product is given by

$$\langle v, w \rangle = \sum_{i=1}^{n} v_i w_i.$$

Then $d_2(v,w) = \|v-w\|_2$, and so the triangle inequality is the statement that

$$\|u-w\|_2 \leqslant \|u-v\|_2 + \|v-w\|_2.$$

This follows immediately by taking $x = u-v$ and $y = v-w$ in the following lemma.

LEMMA 1.3.3. *If $x, y \in \mathbf{R}^n$ then $\|x+y\|_2 \leqslant \|x\|_2 + \|y\|_2$.*

*Proof.* Since $\|v\|_2 \geqslant 0$ for all $v \in \mathbf{R}^n$ the desired inequality is equivalent to

$$\|x+y\|_2^2 \leqslant \|x\|_2^2 + 2\|x\|_2\|y\|_2 + \|y\|_2^2.$$

But since $\|x+y\|_2^2 = \langle x+y, x+y\rangle = \|x\|_2^2 + 2\langle x,y\rangle + \|y\|_2^2$, this inequality is immediate from the Cauchy-Schwarz inequality, that is to say the inequality $|\langle x,y\rangle| \leqslant \|x\|_2\|y\|_2$. □

The next example is rather a routine and trivial one. However, it behaves very differently to the Euclidean examples and can often provide counterexamples to over-optimistic conjectures based on geometric intuition.

EXAMPLE 1.3.4 (Discrete metric). Let $X$ be an arbitrary set. The *discrete* metric on a set $X$ is defined as follows:

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

The axioms for a distance function are easy to check.

Now we turn to some metrics which come up very naturally in diverse areas of mathematics. Our first example is critical in number theory, and also serves to show that metrics need not conform to one's most naïve understand of "distance".

EXAMPLE 1.3.5 (2-adic metric). Let $X = \mathbf{Z}$, and define $d(x,y)$ to be $2^{-m}$, where $2^m$ is the largest power of two dividing $x-y$. The triangle inequality holds in the following stronger form, known as the *ultrametric property*:

$$d(x,z) \leqslant \max(d(x,y), d(y,z)).$$

Indeed, this is just a rephrasing of the statement that if $2^m$ divides both $x-y$ and $y-z$, then $2^m$ divides $x-z$.

This metric is very unlike the usual distance. For example, $d(999, 1000) = 1$, whilst $d(0, 1000) = \frac{1}{8}$!

The role of 2 can be replaced by any other prime $p$, and the metric may also be extended in a natural way to the rationals $\mathbf{Q}$.

Metrics are also ubiquitous in graph theory:

EXAMPLE 1.3.6 (path metric). Let $G$ be a graph, that is to say a finite set of vertices $V$ joined by edges. Suppose that $G$ is connected, that is to say that

there is a path joining any pair of distinct vertices. Define a distance $d$ as follows: $d(v, v) = 0$, and $d(v, w)$ is the length of the shortest path from $v$ to $w$. Then $d$ is a metric on $V$, as can be easily checked.

They also come up in group theory:

EXAMPLE 1.3.7 (Word metric). Let $G$ be a group, and suppose that it is generated by elements $a, b$ and their inverses. Define a distance on $G$ as follows: $d(v, w)$ is the minimal $k$ such that $v = wg_1 \cdots g_k$, where $g_i \in \{a, b, a^{-1}, b^{-1}\}$ for all $i$.

When $G$ is finite, the word metric is a special case of the path metric – you may wish to think about why.

There are many metrics with a prominent position in computer science, for instance:

EXAMPLE 1.3.8 (Hamming distance). Let $X = \{0, 1\}^n$ (the boolean cube), the set of all strings of $n$ zeroes and ones. Define $d(x, y)$ to be the number of coordinates in which $x$ and $y$ differ.

*Remark.* In fact, one can if desired see $\{0, 1\}^n$ as a subset of $\mathbf{R}^n$, and in this case $d$ is the restriction of one of the metrics already considered in Example 1.3.2 (you may care to contemplate which one).

It hardly need be said that metrics are ubiquitous in geometry.

EXAMPLE 1.3.9 (Projective space). Consider the set $\mathbf{P}(\mathbf{R}^n)$ of one-dimensional subspaces of $\mathbf{R}^n$, that is to say lines through the origin). One way to define a distance on this set is to take, for lines $L_1, L_2$, the distance between $L_1$ and $L_2$ to be

$$d(L_1, L_2) = \sqrt{1 - \frac{|\langle v, w\rangle|^2}{\|v\|^2 \|w\|^2}},$$

where $v$ and $w$ are any non-zero vectors in $L_1$ and $L_2$ respectively. It is easy to see this is independent of the choice of vectors $v$ and $w$. The Cauchy-Schwarz inequality ensures that $d$ is well-defined, and moreover the criterion for equality in that inequality ensures positivity. The symmetry property is evident, while the triangle inequality is left as an exercise.

It is useful to think of the case when $n = 2$ here, that is, the case of lines through the origin in the plane $\mathbf{R}^2$. The distance between two such lines given by the above formula is then $\sin(\theta)$ where $\theta$ is the angle between the two lines (another exercise).

## 1.4. Norms

In Example 1.3.2, we looked at three examples of metrics on $\mathbf{R}^n$. They are all, as it turns out, induced from *norms*. This is an important notion which we now develop in its general context.

DEFINITION 1.4.1 (Norms). Let $V$ be any vector space (over the reals). A function $\|\cdot\| : V \to [0,\infty)$ is called a *norm* if the following are all true:

- $\|x\| = 0$ if and only if $x = 0$;
- $\|\lambda x\| = |\lambda|\|x\|$ for all $\lambda \in \mathbf{R}$, $x \in V$;
- $\|x + y\| \leqslant \|x\| + \|y\|$ whenever $x, y \in V$.

Given a norm, it is very easy to check that $d(x, y) := \|x - y\|$ defines a metric on $V$. Indeed, we have already seen that when $V = \mathbf{R}^n$, $\|\cdot\|_2$ is a norm (and so the name "Euclidean norm" is appropriate) and we defined $d_2(x, y) = \|x - y\|_2$.

As we mentioned, the other metrics in Example 1.3.2 also come from norms. Indeed, $d_1$ comes from the $\ell^1$-norm

$$\|x\|_1 := \sum_{i=1}^{n} |x_i|,$$

whilst $d_\infty$ comes from the $\ell^\infty$-norm

$$\|x\|_\infty := \max_{i=1,\dots,n} |x_i|.$$

As the notation suggests, these are special cases of a more general family of norms, the $\ell^p$-norms

$$\|x\|_p := \big(\sum_{i=1}^{n} |x_i|^p\big)^{1/p}.$$

It is true (but we shall not prove it in this course) that these do indeed define norms for $1 \leqslant p < \infty$. Moreover,

$$\lim_{p\to\infty} \|x\|_p = \|x\|_\infty,$$

which is how the $\ell^\infty$-norm comes to have its name.

The principle of turning norms into metrics is important enough that we state it as a lemma in its own right.

LEMMA 1.4.2. *Let $V$ be a vector space over the reals, and let $\|\cdot\|$ be a norm on it. Define $d : V \times V \to [0,\infty)$ by $d(x, y) := \|x - y\|$. Then $(V, d)$ is a metric space.*

It is important to note that the converse is very far from true. For instance, the discrete metric does not arise from a norm. All metrics arising from a norm have the *translation invariance* property $d(x + z, y + z) = d(x, y)$, as well as the scalar invariance $d(\lambda x, \lambda y) = |\lambda| d(x, y)$, neither of which are properties of arbitrary metrics. Conversely one can show that a metric with these two additional properties *does* come from a norm, an exercise we leave to the reader (*Hint*: the norm must arise as $\|v\| = d(v, 0)$).

We call a vector space endowed with a norm $\|\cdot\|$ a *normed space*. Whenever we talk about normed spaces it is understood that we are also thinking of them as metric spaces, with the metric being defined by $d(v, w) = \|v - w\|$.

Note that we do not assume that the underlying vector space $V$ is finite-dimensional. Here are some examples which are not finite-dimensional (whilst we do not *prove* that they are not finite-dimensional here, it is not hard to do so and we suggest this as an exercise).

EXAMPLE 1.4.3 ($\ell^p$ spaces). Let

$$\ell_1 = \{(x_n)_{n=1}^\infty : \sum_{n\geqslant 1} |x_n| < \infty\}$$

$$\ell_2 = \{(x_n)_{n=1}^\infty : \sum_{n\geqslant 1} x_n^2 < \infty\}$$

$$\ell_\infty = \{(x_n)_{n=1}^\infty : \sup_{n\in\mathbb{N}} |x_n| < \infty\}.$$

The sets $\ell_1, \ell_2, \ell_\infty$ are all real vector spaces, and moreover $\|(x_n)\|_1 = \sum_{n\geqslant 1} |x_n|$, $\|(x_n)\|_2 = \left(\sum_{n\geqslant 1} x_n^2\right)^{1/2}$, $\|(x_n)\|_\infty = \sup_{n\in\mathbf{N}} |x_n|$ define norms on $\ell_1, \ell_2$ and $\ell_\infty$ respectively. Note that $\ell_2$ is in fact an inner product space where

$$\langle (x_n), (y_n)\rangle = \sum_{n\geqslant 1} x_n y_n,$$

(the fact that the right-hand side converges if $(x_n)$ and $(y_n)$ are in $\ell_2$ follows from the Cauchy-Schwarz inequality).

The space $\ell^2$ is known as *Hilbert space* and it is of great importance in mathematics.

## 1.5. New metric spaces from old ones

*Subspaces.* Suppose that $(X, d)$ is a metric space and let $Y$ be a subset of $X$. Then the restriction of $d$ to $Y \times Y$ gives $Y$ a metric so that $(Y, d_{|Y\times Y})$ is a metric space. We call $Y$ equipped with this metric a *subspace*.

The word "subspace" is rather overused in mathematics. If $X = \mathbf{R}^n$, so that $X$ is a vector space, then $Y$ need not be a *vector* subspace – it is just a subset of $X$.

Let us give an example of a subspace of a metric space. If $X = \mathbf{R}$, we could take $Y = [0, 1]$, for instance, or $Y = \mathbf{Q}$ (the rationals) or $Y = \mathbf{Z}$ (the integers). (It would be perverse to *define* the usual metric on $\mathbf{Z}$ or on $\mathbf{Q}$ by restricting from $X = \mathbf{R}$. Indeed, the metric space $(X, d)$ with $X = \mathbf{Z}$ and $d(x, y) := |x - y|$ is a much more basic object than $\mathbf{R}$.)

*Product spaces.* If $(X, d_X)$ and $(Y, d_Y)$ are metric spaces, then it is natural to try to make $X \times Y$ into a metric space. One method is as follows: if $x_1, x_2 \in X$ and $y_1, y_2 \in Y$ then we set

$$d_{X\times Y}((x_1, y_1), (x_2, y_2)) = \sqrt{d_X(x_1, x_2)^2 + d_Y(y_1, y_2)^2}$$

The use of the square mean on the right, rather than the max or the sum, is appealing since then the product $\mathbf{R} \times \mathbf{R}$ becomes the space $\mathbf{R}^2$ with the Euclidean metric. However, either of those alternative definitions results in a metric which is equivalent, in the sense made precise in Section 3.3. See Sheet 1, Q4 for more details.

LEMMA 1.5.1. *With notation as above, $d_{X\times Y}$ gives a metric on $X \times Y$.*

*Proof.* Reflexivity and symmetry are obvious. Less clear is the triangle inequality. We need to prove that

$$\sqrt{d_X(x_1,x_3)^2 + d_Y(y_1,y_3)^2} + \sqrt{d_X(x_3,x_2)^2 + d_Y(y_3,y_2)^2} \geqslant \sqrt{d_X(x_1,x_2)^2 + d_Y(y_1,y_2)^2}. \tag{1.1}$$

To make this appear less frightening, write $a_1 = d_X(x_2,x_3)$, $a_2 = d_X(x_1,x_3)$, $a_3 = d_X(x_1,x_2)$ and similarly $b_1 = d_Y(y_2,y_3)$, $b_2 = d_Y(y_1,y_3)$ and $b_3 = d_Y(y_1,y_2)$. Thus we want to show

$$\sqrt{a_2^2 + b_2^2} + \sqrt{a_1^2 + b_1^2} \geqslant \sqrt{a_3^2 + b_3^2}. \tag{1.2}$$

To prove this, note that from the triangle inequality we have $a_1 + a_2 \geqslant a_3$, $b_1 + b_2 \geqslant b_3$. Squaring and adding gives

$$a_1^2 + b_1^2 + a_2^2 + b_2^2 + 2(a_1a_2 + b_1b_2) \geqslant a_3^2 + b_3^2.$$

By Cauchy-Schwarz,

$$a_1a_2 + b_1b_2 \leqslant \sqrt{a_1^2 + b_1^2}\sqrt{a_2^2 + b_2^2}.$$

Substituting this into the previous line gives precisely the square of (1.2), and (1.1) follows. □

## 1.6. Balls and boundedness

DEFINITION 1.6.1 (Balls). Let $X$ be a metric space. If $a \in X$ and $\varepsilon > 0$ then we define the *open ball of radius* $\varepsilon$ to be the set

$$B(a,\varepsilon) = \{x \in X : d(x,a) < \varepsilon\}.$$

Similarly we defined the *closed ball* of radius $\epsilon$ about $a$ to be the set

$$\overline{B}(a,\varepsilon) = \{x \in X : d(x,a) \leqslant \varepsilon\}.$$

Thus when $X = \mathbf{R}^3$ with the Euclidean metric we see that $B(0,1)$ really is what we understand geometrically as a ball (minus its boundary, the unit sphere), whilst $\overline{B}(0,1)$ contains the unit sphere and everything inside it.

We caution that this intuitive picture of the closed ball being the open ball "together with its boundary" is totally misleading in general. For instance, in the discrete metric on a set $X$, the open ball $B(a, 1)$ contains only the point $a$, whereas the closed ball $\overline{B}(a, 1)$ is the whole of $X$.

DEFINITION 1.6.2. Let $X$ be a metric space, and let $Y \subseteq X$. Then we say that $Y$ is *bounded* if $Y$ is contained in some open ball.

LEMMA 1.6.3. *Let $X$ be a metric space and let $Y \subseteq X$. Then the following are equivalent.*

(i) *$Y$ is bounded;*
(ii) *$Y$ is contained in some closed ball;*
(iii) *The set $\{d(y_1, y_2) : y_1, y_2 \in Y\}$ is a bounded subset of $\mathbf{R}$.*

*Proof.* That (i) implies (ii) is totally obvious. That (ii) implies (iii) follows immediately from the triangle inequality. Finally, suppose $Y$ satisfies (iii). Then there is some $K$ such that $d(y_1, y_2) \leqslant K$ whenever $y_1, y_2 \in Y$. If $Y$ is empty, it is certainly bounded. Otherwise, let $a \in Y$ be an arbitrary point. Then $Y$ is contained in $B(a, r)$ where $r = K + 1$. □

CHAPTER 2

# Limits and continuity

The main purpose of introducing the idea of a metric space is that many notions familiar over $\mathbf{R}$, such as those of limit and continuous function, can be extended to metric spaces, and theorems about them proven in that context.

## 2.1. Basic definitions and properties

DEFINITION 2.1.1 (Limit). Suppose that $(x_n)_{n=1}^{\infty}$ is a sequence of elements of a metric space $(X, d)$. Let $x \in X$. Then we say that $x_n \to x$, or that $\lim_{n\to\infty} x_n = x$, if the following is true. For every $\varepsilon > 0$, there is an $N$ such that $d(x_n, x) < \varepsilon$ for all $n \geqslant N$.

Let us bolster this definition with a couple of easy remarks. First, it is quite possible and indeed usual for a sequence $(x_n)_{n=1}^{\infty}$ to have no limit. Take, for instance, the sequence $(0, 1, 0, 1, 0, 1, \dots)$ in $\mathbf{R}$. Second, if the limit does exist then it is unique. To see this, suppose that $x_n \to a$ and $x_n \to b$, but that $a \neq b$. Let $\delta := d(a, b)$. Then, taking $\varepsilon = \delta/2$ in the definition of limit, we see that for $n$ sufficiently large we have $d(x_n, a), d(x_n, b) < \delta/2$. But then the triangle inequality yields

$$\delta = d(a, b) \leqslant d(x_n, a) + d(x_n, b) < \delta,$$

a contradiction.

DEFINITION 2.1.2 (Continuity). Let $(X, d_X)$ and $(Y, d_Y)$ be metric spaces. We say a function $f\colon X \to Y$ is continuous at $a \in X$ if for any $\epsilon > 0$ there is a $\delta > 0$ such that for any $x \in X$ with $d_X(a, x) < \delta$ we have $d_Y(f(x), f(a)) < \epsilon$.

We say $f$ is *continuous* if it is continuous at every $a \in X$.

Although we will come across it all that much in this course, it is important to note that the definition of *uniform* continuity may be extended to metric spaces as well. As for real functions, the idea is that "$\delta$ should depend only on $\varepsilon$".

DEFINITION 2.1.3 (Uniform continuity). Let $(X, d_X)$ and $(Y, d_Y)$ be metric spaces. We say a function $f\colon X \to Y$ is uniformly continuous if for any $\epsilon > 0$ there is a $\delta > 0$ such that for any $x, y \in X$ with $d_X(x, y) < \delta$ we have $d_Y(f(x), f(y)) < \epsilon$.

As for functions on the reals, one may also phrase the definition of continuity in terms of limits.

LEMMA 2.1.4. *Let $f : X \to Y$ be a function between metric spaces. Then $f$ is continuous at $a$ if and only if the following is true: for any sequence $(x_n)_{n=1}^{\infty}$ with $\lim_{n\to\infty} x_n = a$, we have $\lim_{n\to\infty} f(x_n) = f(a)$.*

*Proof.* Suppose first that $f$ is continuous at $a$. Then given $\epsilon > 0$ there is a $\delta > 0$ such that for all $x \in X$ with $d(x, a) < \delta$ we have $d(f(x), f(a)) < \epsilon$. Now if $(x_n)_{n=1}^{\infty}$ is a sequence with limit $a$ then, by the definition of limit, there is an $N > 0$ such that $d(a, x_n) < \delta$ for all $n \geqslant N$. But then for all $n \geqslant N$ we see that $d(f(a), f(x_n)) < \epsilon$, so indeed $\lim_{n\to\infty} f(x_n) = f(a)$ as required.

In the other direction, suppose $f$ is not continuous at $a$. Then there is an $\epsilon > 0$ such that for all $\delta > 0$ there is some $x \in X$ with $d(x, a) < \delta$ and $d(f(x), f(a)) \geqslant \epsilon$. Taking $\delta = 1/n$, we see that for each $n$ there is some $x_n \in X$ with $d(x_n, a) < 1/n$ and $d(f(x_n), f(a)) \geqslant \epsilon$. Therefore $\lim x_n = a$, but $\lim f(x_n) \neq f(a)$. □

## 2.2. Continuity of linear functions in normed spaces

An important source of metric spaces are the normed spaces. Recall Lemma 1.4.2: If $V$ is a vector space, and $\|\cdot\|$ a norm on it, we can define a metric $d : V \times V \to [0, \infty)$ by $d(x, y) := \|x - y\|$.

Suppose now that we have two normed spaces $V, W$, with norms $\|\cdot\|_V$ and $\|\cdot\|_W$ respectively; henceforth, we will drop the subscripts since it will always be clear which space we are working on. There is a pleasant criterion for when a *linear* map $f : V \to W$ is continuous.

LEMMA 2.2.1. *Let $f\colon V \to W$ be a linear map between normed vector spaces. Then $f$ is continuous if and only if $\{\|f(x)\| : \|x\| \leqslant 1\}$ is bounded.*

*Proof.* Suppose first that $f$ is continuous. In particular, it is continuous at $0 \in V$. Therefore, taking $\varepsilon = 1$ in the definition of continuity, there is some $\delta > 0$ such that $d(f(x), f(0)) < 1$ whenever $\|x\| < \delta$. Since $f(0) = 0$, this implies that $\|f(x)\| \leqslant 1$ for these $x$. Now suppose that $\|v\| = 1$. Then $\|\delta v/2\| = \delta/2 < \delta$, and so $\|f(\delta v/2)\| \leqslant 1$. Since $f$ is linear, $f(\delta v/2) = \delta f(v)/2$, and so $\|f(\delta v/2)\| = \delta\|f(v)\|/2$. It follows that $\|f(v)\| \leqslant 2/\delta$, and so indeed the set $\{\|f(x)\| : \|x\| \leqslant 1\}$ is bounded.

For the converse, suppose that $\|f(v)\| < M$ for all $v$ with $\|v\| \leqslant 1$. Let $\varepsilon > 0$, and set $\delta := \epsilon/M$. Then if $\|v - w\| < \delta$ we have

$$\|f(v) - f(w)\| = \|f(v - w)\| = \delta\|f(\delta^{-1}(v - w))\| < \delta M = \epsilon,$$

so that $f$ is in fact uniformly continuous on $V$. □

As a consequence of Lemma 2.2.1, one never really hears about "continuous linear functions", it being completely standard to refer to them as bounded instead.

## 2.3. Function spaces

A great deal of power comes from considering the set of all functions on a space satisfying some property, such as continuity, as a metric space in its own right. In this section we consider some important examples of such spaces.

We begin with the space of bounded real-valued functions on a set $X$. At this stage we assume nothing about $X$.

DEFINITION 2.3.1. If $X$ is any set we define $B(X)$ to be the space of functions $f : X \to \mathbf{R}$ for which $f(X) = \{f(x) : x \in X\}$ is bounded. If $f \in B(X)$, define $\|f\|_\infty = \sup_{x \in X} |f(x)|$.

LEMMA 2.3.2. *For any set $X$, $B(X)$ is a vector space, and $\|\cdot\|_\infty$ is a norm.*

We leave the proof as an easy exercise.

Now we turn to the space of continuous real-valued functions, $C(X)$. To make sense of what this means we now need $X$ to be a metric space.

DEFINITION 2.3.3. Let $X$ be a metric space. Then we write $C(X)$ for the space of all continuous functions $f : X \to \mathbf{R}$.

LEMMA 2.3.4. *The space $C(X)$ is a vector space over $\mathbf{R}$, with pointwise addition and multiplication by scalars.*

*Proof.* One must check that $C(X)$ is closed under addition and scalar multiplication. We do the case of addition; scalar multiplication is left as an (easy) exercise.

Suppose that $f, g \in C(X)$, and let $\varepsilon > 0$. Let $a \in X$.

Since $f$ is continuous at $a$, there is some $\delta_1$ such that $d(x, a) < \delta_1$ implies $|f(x) - f(a)| < \varepsilon/2$.

Since $g$ is continuous at $a$, there is some $\delta_2$ such that $d(x, a) < \delta_2$ implies $|g(x) - g(a)| < \varepsilon/2$.

Take $\delta = \min(\delta_1, \delta_2)$. Then, if $d(x, a) < \delta$ we have

$$\begin{aligned} |(f+g)(x) - (f+g)(a)| &= |f(x) + g(x) - f(a) - g(a)| \\ &\leqslant |f(x) - f(a)| + |g(x) - g(a)| \\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned}$$

Therefore $f + g$ is continuous at $a$. □

In general, we certainly do not have $B(X) \subseteq C(X)$, and unless $X$ is special we do not have $C(X) \subseteq B(X)$. We will discuss situations in which this *is* true later on; you

will already be familiar with a nontrivial example, namely that $C([0,1]) \subseteq B([0,1])$, that is to say all continuous functions on $[0,1]$ are bounded.

DEFINITION 2.3.5. Let $X$ be a metric space. Write $C_b(X) := C(X) \cap B(X)$ for the space of continuous, bounded functions on $X$. Since $C_b(X)$ is a subspace of $B(X)$, it inherits the norm $\|f\|_\infty = \sup_{x \in X} |f(x)|$, and we may define a metric $d_\infty$ on $C_b(X)$ in the usual way via $d_\infty(f,g) := \|f - g\|_\infty$

A useful exercise in checking your understanding of these definitions is Example 2.3.6 below. Here, let $X = [0,1]$. Then, as we just remarked, $C(X) = C_b(X)$. Instead of writing $C([0,1])$, it is conventional to write $C[0,1]$ for the vector space of continuous (and automatically bounded) functions on $[0,1]$.

EXAMPLE 2.3.6. Consider the space $C[0,1]$ together with the metric $d_\infty$ induced from the norm $\|\cdot\|_\infty$. Let $(f_n)_{n=1}^\infty$ be a sequence of elements (functions) of this space, and let $f$ be a further element. Then $f_n \to f$ in the metric $d_\infty$ if, and only if, $f_n$ converges to $f$ uniformly.

*Proof.* This is essentially a tautology, but it takes a little thought to unravel all the definitions. □

The norm $\|\cdot\|_\infty$ is by no means the only natural one on $C[0,1]$.

LEMMA 2.3.7. *For $f \in C[0,1]$, define*

$$\|f\|_1 = \int_0^1 |f(t)| dt.$$

*Then $\|\cdot\|_1$ is a norm on $C[0,1]$.*

*Remarks.* This norm is called the $L^1$- ("big ell one") norm. Note that, although we use the same notation as for the $\ell^1$-norm, this is quite a different object; in particular, the underlying vector space $C[0,1]$ is infinite-dimensional.

*Proof.* Most of what needs to be shown is very routine, at least given the results in the Prelims course Analysis III: integration. The fact that $\|f\|_1$ exists, behaves well with respect to the scalar multiplication and satisfies the triangle inequality all fall into this category.

One point deserves further comment. It needs to be shown that $\|f\|_1 = 0$ implies $f = 0$. Suppose not. Then there is some point $x \in [0,1]$ with $|f(x)| > 0$, let us say $|f(x)| = \varepsilon$. Since $f$ is continuous, there is some $\delta > 0$ such that if $|x - y| \leqslant \delta$ then $|f(y)| \geqslant \varepsilon/2$. The set of all $y \in [0,1]$ with $|x - y| \leqslant \delta$ is a subinterval $I \subset [0,1]$ with length at least $\min(1, \delta)$, and so

$$\int |f| \geqslant \int_I |f| \geqslant \frac{\varepsilon}{2} \min(1, \delta) > 0,$$

a contradiction. □

Note carefully that for the last part of the argument crucial use was made of the continuity of $f$. Indeed, the result is false without at least some assumption. Suppose one attempts to define $\|f\|_1$ for all bounded Riemann integrable functions. This is well-defined, and satisfies the scalar multiplication property and the triangle inequality, as required in the definition of norm. However, it is *not* a norm, since there are non-zero functions with $\|f\|_1 = 0$, for instance the function $f : [0,1] \to \mathbf{R}$ defined by $f(0) = 1$ and $f(x) = 0$ for $0 < x \leqslant 1$.

We should also say that there is nothing special about the interval $[0,1]$; everything we have said works, with essentially identical proofs, for any closed interval $[a,b] \subset \mathbf{R}$ with $a < b$.

As with the $\ell^p$ norms on $\mathbf{R}^n$, one may also define norms $\|f\|_p = (\int_0^1 |f(t)|^p dt)^{1/p}$ on $C[0,1]$ for any $p \in [1,\infty)$. These are called $L^p$-norms, and the case $p = 2$ is particularly important. We will not discuss it further here.

CHAPTER 3

# Isometries, homeomorphisms and equivalence

One learns as mathematician that, when one studies a type of structure, one should also study maps which preserve that structure. In this chapter we will look at various such notions applicable to metric spaces.

## 3.1. Isometries

Maps which genuinely preserve the distance function are called isometries.

DEFINITION 3.1.1. Let $(X, d_X)$ and $(Y, d_Y)$ be metric spaces. A function $f\colon X \to Y$ between metric spaces $(X, d_X)$ and $(Y, d_Y)$ is said to be an *isometry* if

$$d_Y(f(x), f(y)) = d_X(x, y) \text{ for all } x, y \in X. \tag{3.1}$$

*Remarks.* An isometry is automatically injective, but not automatically surjective. For instance, the right-shift map on $\ell^2$ defined by $f((x_1, x_2, x_3, \dots)) = (0, x_1, x_2, \dots)$ is an isometry, but it is not surjective.

If an isometry is surjective as well, we call it a *bijective isometry*. Some authors use the word "isometry" to mean "bijective isometry", but we have refrained from doing this so that we are consistent with Prelims course M4: Geometry. There, isometries in the case $X = Y = \mathbf{R}^n$ were discussed in considerable detail. It was shown that they all have the form $f(x) = Ax + b$ for some orthogonal matrix $A$; in particular, they are automatically surjective.

For any metric space $X$ the set of all bijective isometries from $X$ to itself is a group under composition, denoted $\mathrm{Isom}(X)$.

## 3.2. Homeomorphisms

The notion of isometry is rather rigid. A weaker notion is that of a homeomorphism.

DEFINITION 3.2.1. Let $f\colon X \to Y$ be a continuous function between metric spaces $X$ and $Y$. We say that $f$ is a *homeomorphism* if it is continuous, a bijection, and if its inverse $f^{-1} : Y \to X$ is also continuous.

*Remark.* Note that it is possible for a map $f : X \to Y$ to be both continuous and a bijection, but for its inverse to fail to be continuous (so in this case $f$ is

not a homeomorphism). For instance, consider the spaces $X = [0,1) \cup [2,3]$ and $Y = [0,2]$. Then the function $f\colon X \to Y$ given by

$$f(x) = \begin{cases} x, & \text{if } x \in [0,1) \\ x-1, & \text{if } x \in [2,3] \end{cases}$$

is a bijection and is clearly continuous. However, its inverse $g\colon Y \to X$ is not continuous at 1 – the one-sided limits of $g$ as $x$ tends to 1 from above and below are 1 and 2 respectively.

The following examples illustrate the extent to which homeomorphisms are less rigid than isometries.

EXAMPLE 3.2.2. The closed disk $\bar{B}(0,1)$ of radius 1 in $\mathbf{R}^2$ is homeomorphic to the square $[-1,1] \times [-1,1]$. The easiest way to see this is to inscribe the disk in the square and stretch the disk radially out to the square. One can write explicit formulas for this in the four quarters of the disk given by the lines $x \pm y = 0$ to check this does indeed give a homeomorphism.

EXAMPLE 3.2.3. The open interval $(-1,1)$ is homeomorphic to $\mathbf{R}$: an explicit homeomorphism is given by $f(x) = x/(1-|x|)$, which has inverse $g(x) = x/(1+|x|)$. It follows (using translation and scaling maps) that any open interval is homeomorphic to $\mathbf{R}$. Similarly, the function $h(x) = 1/x$ shows that $(0,1)$ and $(1,\infty)$ are homeomorphic, and from this one can see that the spaces $\mathbf{R}$, $(a,b)$, $(-\infty,a)$ and $(a,\infty)$ are all homeomorphic for any $a,b \in \mathbf{R}$ with $a < b$.

EXAMPLE 3.2.4. A coffee cup (reusable, with a handle) is homeomorphic to a doughnut.

## 3.3. *Equivalent metrics

One space $X$ can certainly support wildly different metrics. For instance, the 2-adic metric on $\mathbf{Q}$ is very different to the standard Euclidean metric. However, there is a useful notion of two metrics $d_1, d_2$ on the same space being equivalent.

DEFINITION 3.3.1 (Equivalent metrics). Let $X$ be a set, and let $d, d'$ be two metrics on $X$. Then we say that the metrics $d, d'$ are equivalent if the identity map $\iota : (X,d) \to (X,d')$ is a homeomorphism.

An easy exercise in the definitions show that this is equivalent to the following property: every open ball $B(x,\varepsilon)$ with respect to the $d$-metric contains an open ball $B'(x,\varepsilon')$ in the $d'$-metric, and vice versa.

If two metrics $d, d'$ are equivalent then, for example, the notions of limit coincide in the two metric spaces $(X,d)$ and $(X,d')$. We leave the detailed proof as an exercise.

PROPOSITION 3.3.2. *The metrics $d_1, d_2, d_\infty$ on $\mathbf{R}^n$ are equivalent.*

*Proof.* In fact, we will show that these metrics are *strongly* equivalent. Two metrics $d, d'$ on a space $X$ are strongly equivalent if there is a constant $C$ such that

$$d(x,y) \leqslant Cd'(x,y) \text{ and } d'(x,y) \leqslant Cd(x,y)$$

for all $x \neq y$. We leave it as an easy exercise to show that strongly equivalent metrics are indeed equivalent (the converse is not true).

The three metrics under consideration all come from norms, and it is enough to find some constant $C$ such that

$$\|x\| \leqslant C\|x\|' \tag{3.2}$$

for each pair $\|\cdot\|, \|\cdot\|'$ of these norms. Four such inequalities are obvious, namely

$$\|x\|_\infty \leqslant \|x\|_1 \leqslant n\|x\|_\infty$$

and

$$\|x\|_\infty \leqslant \|x\|_2 \leqslant n^{1/2}\|x\|_\infty.$$

The remaining two inequalities follow from these two, or one could use the Cauchy-Schwarz inequality to get better constants. □

CHAPTER 4

# Open and closed sets

## 4.1. Basic definitions

The definition of open set, which we give now, is one of the most important in the course.

DEFINITION 4.1.1 (Open sets). If $X$ is a metric space then we say a subset $U \subseteq X$ is *open* (or *open in* $X$) if for each $y \in U$ there is some $\delta > 0$ such that the open ball $B(y, \delta)$ is contained in $U$.

To check you have understood the definition, convince yourself of the following facts:

- The interval $(0, 1)$ is open in $\mathbf{R}$, but $[0, 1]$ is not;
- The rationals $\mathbf{Q}$ are not open in $\mathbf{R}$;
- If $X$ is a set with the discrete metric, every set is open.

Note carefully that the notion of open set is a relative notion, depending on $U$ being contained in $X$. Thus, while $[0, 1]$ is not open in $\mathbf{R}$, it is an open set considered as a subspace of itself.

The first basic result about open sets is that open balls $B(a, \varepsilon)$ are open. Note that this is not a tautology – at this point "open ball" is just the name we gave to the sets $B(a, \varepsilon)$, and the fact that they are indeed open in the sense of Definition 4.1.1 needs to be proven.

LEMMA 4.1.2. *Every open ball in an metric space is an open set.*

*Proof.* Let the ball be $B(a, \varepsilon)$. Let $x \in B(a, \varepsilon)$. Then $d(x, a) < \varepsilon$, so there is $\varepsilon' > 0$ so that $d(x, a) < \varepsilon - \varepsilon'$. We claim that the open ball $B(x, \varepsilon')$ is contained in $B(a, \varepsilon)$. To see this, suppose that $z \in B(x, \varepsilon')$. Then $d(z, x) < \varepsilon'$ and so by the triangle inequality $d(z, a) \leqslant d(z, x) + d(x, a) < \varepsilon' + (\varepsilon - \varepsilon') = \varepsilon$. □

The complement of an open set is a closed set.

DEFINITION 4.1.3 (Closed sets). If $X$ is a metric space, then a subset $F \subseteq X$ is said to be a *closed* subset of $X$ if and only if its complement $F^c = X \backslash F$ is an open subset.

It is important to note that the property of being closed is *not* the property of not being open! In a metric space, it is possible for a subset to be open, closed, both or neither: In $\mathbf{R}$ the set $\mathbf{R}$ is open and closed, the set $(0,1)$ is open and not closed, the set $[0,1]$ is closed and not open while the set $(0,1]$ is neither.

Just as open balls are open sets, so closed balls are closed sets, and this is also a fact requiring proof.

LEMMA 4.1.4. *Every closed ball in a metric space is a closed set. In particular, singleton sets are closed.*

*Proof.* Let the ball be $\bar{B}(a,\varepsilon)$. We will show that the complement $\bar{B}(a,\varepsilon)^c$ is open. Let $x \in \bar{B}(a,\varepsilon)^c$. Then $d(x,a) > \varepsilon$, so there is $\varepsilon' > 0$ so that $d(x,a) > \varepsilon + \varepsilon'$. We claim that the open ball $B(x,\varepsilon')$ is contained in $\bar{B}(a,\varepsilon)^c$. To see this, suppose that $z \in B(x,\varepsilon')$. Then $d(z,x) < \varepsilon'$ and so by the triangle inequality $d(z,a) \geqslant d(x,a) - d(z,x) > (\varepsilon + \varepsilon') - \varepsilon' = \varepsilon$.

The second statement – that singleton sets are closed – follows from the observation that $\{a\} = \bar{B}(a,0)$. □

## 4.2. Basic properties of open sets

LEMMA 4.2.1. *Let $X$ be metric space. Then we have*

(i) *The subsets $X$ and $\emptyset$ are open.*
(ii) *For any indexing set $I$ and $\{U_i : i \in I\}$ a collection of open sets, the set $\bigcup_{i\in I} U_i$ is an open set.*
(iii) *If $I$ is finite and $\{U_i : i \in I\}$ are open sets then $\bigcap_{i\in I} U_i$ is open.*

*Proof.* The first claim is trivial. For the second claim, if $x \in \bigcup_{i\in I} U_i$ then there is some $i \in I$ with $x \in U_i$. Since $U_i$ is open, some open ball $B(x,\varepsilon)$ is contained in $U_i$ and hence in $\bigcup_{i\in I} U_i$.

Finally, for claim (iii), suppose that $I$ is finite and that $x \in \bigcap_{i\in I} U_i$. For each $i \in I$, we have $x \in U_i$, and so some ball $B(x,\varepsilon_i)$ is contained in $U_i$. Set $\varepsilon := \min_{i\in I} \varepsilon_i$; then $\varepsilon > 0$ (here it is, of course, crucial that $I$ be finite), and $B(x,\varepsilon) \subseteq B(x,\varepsilon_i) \subseteq U_i$ for all $i$. Therefore $B(x,\varepsilon) \subseteq \bigcap_{i\in I} U_i$. □

*Remarks.* (i) is in fact a special case of (ii) and (iii), taking $I$ to be the empty set.

It is extremely important to note that, whilst the indexing set $I$ in (ii) can be arbitrary, the indexing set in (iii) must be finite. In general, an arbitrary intersection of open sets is not open; for instance, the intervals $U_i = (-1/i, 1/i)$ are all open in $\mathbf{R}$, but their intersection $\bigcap_{i=1}^{\infty} U_i$ is just the singleton $\{0\}$, which is not an open set.

A result equivalent to Lemma 4.2.1 may be formulated in terms of closed sets, simply by taking complements and applying de Morgan's laws. We simply state the outcome.

LEMMA 4.2.2. *Let $X$ be a metric space and let $\{F_i : i \in I\}$ be a collection of closed subsets.*

(i) *The subsets $X$ and $\emptyset$ are closed.*
(ii) *The intersection $\bigcap_{i\in I} F_i$ is a closed subset.*
(iii) *If $I$ is finite then $\bigcup_{i\in I} F_i$ is closed.*

If $X$ is a metric space, the collection of all open sets in $X$ is calld the *topology* of $X$.

## 4.3. Continuity in terms of open sets

An interesting and important fact is that continuity of a function may be expressed in an $\varepsilon$-$\delta$-free manner using open sets. To formulate a precise statement, we introduce the concept of a *neighbourhood*.

DEFINITION 4.3.1 (Neighbourhood). Let $X$ be a metric space, and let $Z \subseteq X$. Let $z \in Z$. We say $Z$ is a *neighbourhood* of $z$ if some open ball about $z$ is contained in $Z$: that is, if there is some $\delta > 0$ such that $B(z, \delta) \subseteq Z$.

Thus any open set containing $z$ is a neighbourhood of $z$. However, the converse is not true: there is no requirement that a neighbourhood itself be an open set.

Here is the formulation of continuity in terms of neighbourhoods.

PROPOSITION 4.3.2. *Let $X, Y$ be metric spaces and let $f : X \to Y$ be a map. If $a \in X$ then $f$ is continuous at $a$ if and only if for every neighbourhood $N \subseteq Y$ of $f(a)$, the preimage $f^{-1}(N)$ is a neighbourhood of $a \in X$.*

*Proof.* Before embarking on the proof, let us make sure that the statement is clear. Here, $f^{-1}(N)$ denotes the preimage of $N$ under $f$, the set of all points in $X$ which map, under $f$, to $N$. Note that we are *not* asserting that $f$ is invertible, and $f^{-1}$ is not a function; a given point in $Y$ may have several preimages.

We now turn to the proof. There is little more to it than working through the definitions, but there have been sufficiently many of them that it is worth doing properly. Let $d_X, d_Y$ be the metrics on $X, Y$ respectively.

Suppose first that $f$ is continuous at $a$. Let $N$ be a neighbourhood of $f(a)$. By the definition of neighbourhood, $N$ contains some open ball $B(f(a), \varepsilon)$. By the definition of continuity, there is some $\delta > 0$ such that, if $d_X(a, x) < \delta$, then $d_Y(f(a), f(x)) < \varepsilon$. Equivalently, if $x \in B(a, \delta)$, then $f(x) \in B(f(a), \varepsilon)$. Put another way, $f^{-1}(B(f(a), \varepsilon)) \supseteq B(a, \delta)$. Since $N$ contains $B(f(a), \varepsilon)$, $f^{-1}(N)$

contains $f^{-1}(B(f(a),\varepsilon)$. Therefore $f^{-1}(N) \supseteq B(a,\delta)$. We have found an open ball about $a$ which is contained in $f^{-1}(N)$, which is precisely what it means for $f^{-1}(N)$ to be a neighbourhood of $a$.

In the other direction, suppose that $f$ satisfies the neighbourhood preimages property. We show that $f$ is continuous at $a$. Let $\varepsilon > 0$, and consider the open ball $B(f(a),\varepsilon)$. This is an open set containing $f(a)$, and hence it is a neighbourhood of $f(a)$. By assumption, its preimage $f^{-1}(B(f(a),\varepsilon))$ is a neighbourhood of $a$, which means that it contains some open ball $B(a,\delta)$. Thus $B(a,\delta) \subseteq f^{-1}(B(f(a),\varepsilon))$, or in other words if $x \in B(a,\delta)$ then $f(x) \in B(f(a),\varepsilon)$. This is what it means for $f$ to be continuous at $a$. □

One can prove a very closely related characterisation of what it means for a function to be continuous at *every* point. I personally find it easier to prove this afresh, rather than deduce it from Proposition 4.3.2.

PROPOSITION 4.3.3. *Let $X, Y$ be metric spaces and let $f : X \to Y$ be a map. Then $f$ is continuous on all of $X$ if and only if for each open subset $U$ of $Y$, its preimage $f^{-1}(U)$ is open in $X$.*

*Proof.* Suppose first that $f$ is continuous at every point. and let $U \subseteq Y$ be open; we want to show that $f^{-1}(U)$ is open. Let $a \in f^{-1}(U)$ be arbitrary. Then $f(a) \in U$, and so, since $U$ is open, some ball $B(f(a),\varepsilon)$ also lies in $U$. By the definition of continuity, there is some $\delta > 0$ such that if $x \in B(a,\delta)$ then $f(x) \in B(f(a),\varepsilon)$, and therefore $f^{-1}(B(f(a),\varepsilon)) \supseteq B(a,\delta)$. Therefore $f^{-1}(U)$ contains $B(a,\delta)$, which means that $f^{-1}(U)$ is open.

Now suppose that $f$ satisfies the open sets preimages property, and let $a \in X$. The ball $B(f(a),\varepsilon)$ is open, and so by assumption the preimage $f^{-1}(B(f(a),\varepsilon))$ is open. Since $a$ lies in this set, it follows from the definition of open that there is some $\delta > 0$ such that $B(a,\delta) \subseteq f^{-1}(B(f(a),\varepsilon))$, whence $f(B(a,\delta)) \subseteq B(f(a),\varepsilon)$. This is what it means for $f$ to be continuous at $a$. □

By taking complements, one can show the following version of Proposition 4.3.3 for closed sets: $f : X \to Y$ is continuous if and only if for each *closed* subset $V$ of $Y$, its preimage $f^{-1}(V)$ is a *closed* subset of $X$.

Finally, it is important to take note of what Proposition 4.3.3 does *not* say, namely that a continuous function maps open sets to open sets. This is obvious since, for example, constant functions are continuous. Less obvious is the fact that it still fails even under the assumption that $f$ is injective. For instance, the natural map $f : [0,1) \to \mathbf{R}/\mathbf{Z}$ is continuous. The set $[0,1/2)$ is open in $[0,1)$, but its image is not open in $\mathbf{R}/\mathbf{Z}$.

## 4.4. *Topological spaces

In this section we offer a very brief taster of the course A5: Topology by discussing the notion of a topological space. One may of course observe that Proposition 4.3.3 allows one to define the notion of a continuous function without explicitly mentioning the metric $X$ or concepts equivalent to it such as the notion of an open ball of radius $\delta$. Of course, those notions are embedded within the definition of the notion of an open set, so this comment is a little misleading.

In the concept of a topological space, the open sets are to the fore. Thus a topological space is a set $X$ together with a collection of sets $U$ (which we call the open sets) satisfying certain properties. The properties we require are precisely those which we *proved* in Lemma 4.2.1, namely

(i) The subsets $X$ and $\emptyset$ are open.
(ii) For any indexing set $I$ and $\{U_i; i \in I\}$ a collection of open sets, the set $\bigcup_{i\in I} U_i$ is an open set.
(iii) If $I$ is finite and $\{U_i : i \in I\}$ are open sets then $\bigcap_{i\in I} U_i$ is open in $X$.

Note that we have not said anything about the "geometry" of the open sets, or anything about them containing balls - indeed there are no such notions, because $X$ is equipped with no structure.

Lemma 4.2.1 may then be phrased as follows.

LEMMA 4.4.1 (Lemma 4.2.1). *Let $X$ be a metric space together with the open sets as defined in Definition 4.1.1. Then $X$ is a topological space, with the same collection of open sets.*

The concept of a topological space is considerably more general than that of a metric space, and there are certainly topological spaces which do not have the structure of a metric space (are not *metrizable*). However, as a consequence of Proposition 4.3.3 we may still formulate the notion of a continuous function between two topological spaces, in such a way that when restricted to metric spaces it coincides with the usual definition.

DEFINITION 4.4.2. Suppose that $X$ and $Y$ are two topological spaces. Then we say that $f : X \to Y$ is continuous if and only if, for every open set $U \subseteq Y$, the inverse $f^{-1}(U)$ is open in $X$.

Let us emphasise that in the generality of topological spaces, there is no equivalent form of this definition in terms of $\varepsilon$s and $\delta$s.

## 4.5. Subspaces

If $(X, d)$ is a metric space, then as we noted in Section 1.5, any subset $Y \subseteq X$ is automatically also a metric space since the distance function $d\colon X \times X \to \mathbf{R}_{\geqslant 0}$

restricts to a distance function on $Y$. We will use the letter $d$ for both metrics, but it is important to distinguish the balls in $Y$ from the balls in $X$, because these are quite different objects.

We will write

$$B_Y(y,r) = \{z \in Y : d(z,y) < r\}$$

for the open ball about $y$ of radius $r$ in $Y$ and

$$B_X(y,r) = \{x \in X : d(x,y) < r\}$$

for the open ball of radius $r$ about $y$ in $X$.

Note that $B_Y(y,r) = Y \cap B_X(y,r)$.

Similarly, the notions of a set being open in $X$ and of being open in $Y$ are quite different.

By way of an example, consider $X = \mathbf{R}^2$ and $Y = \mathbf{R} \times \{0\}$, that is to say $Y$ is the $x$-axis. The the ball $B_X(0,1)$ is simply the open unit disc of radius 1, whilst the ball $B_Y(0,1)$ is the open unit line segment $(-1,1)$. Note carefully that $B_Y(0,1)$, whilst it is open as a subset of $Y$, does not look remotely like an open subset of $X$.

The following lemma clarifies the relationship between open sets in $X$ and open sets in the subspace $Y$.

LEMMA 4.5.1. *Let $X$ be a metric space and suppose that $Y \subseteq X$. Then a subset $U \subseteq Y$ is an open subset of $Y$ if and only if there is an open subset $V$ of $X$ such that $U = Y \cap V$. Similarly a subset $Z \subseteq Y$ is a closed subset of $Y$ if and only if there is a closed subset $F$ of $X$ such that $Z = F \cap Y$.*

*Proof.* Suppose first that $U = Y \cap V$, where $V$ is open in $X$. We will show that $U$ is open in $Y$. Let $y \in U$. Then, since $V$ is open, there is some $\varepsilon > 0$ such that $B_X(y,\varepsilon) \subseteq V$. Therefore

$$B_Y(y,\varepsilon) = Y \cap B_X(y,\varepsilon) \subseteq V \cap Y = U.$$

We have shown that some open ball (in $Y$) about $y$ is contained in $U$, and therefore $U$ is open.

In the other direction, suppose that $U$ is an open subset of $Y$. Then for each $y \in U$ we may pick an open ball $B_Y(y,\varepsilon_y)$ contained in $U$. We have $\bigcup_{y\in U} B_Y(y,\varepsilon_y) = U$. Now define $V = \bigcup_{y\in U} B_X(y,\varepsilon_y)$. Then $V$, being a union of open balls in $X$, is open. Moreover

$$Y \cap V = Y \cap \bigcup_{y\in Y} B_X(y,\varepsilon_y) = \bigcup_{y\in Y} (Y \cap B_X(y,\varepsilon_y)) = \bigcup_{y\in Y} B_Y(y,\varepsilon_y) = U.$$

The corresponding result for closed sets follows by taking complements – we leave the detailed verification as an exercise. □

The concept of being open in a subspace is a bit confusing when you first meet it, so let us give an example. Let $X = \mathbf{R}$, and $Y = (0,1] \cup [2,3]$. Set $U = (0,1]$. Then $U$ is not open as a subset of $X$ – for instance, no open ball $B_X(1,\varepsilon)$ is contained in $U$. However, $U$ *is* open as a subset of $Y$. For example, the ball $B_Y(1,\frac{1}{2})$, which consists of all points of $Y$ at distance less than $\frac{1}{2}$ from 1, is the set $(\frac{1}{2},1]$, and this is contained in $U$.

**Remark.* (For those who have read Section 4.4). Lemma 4.5.1 shows that the topology on $X$ determines the topology on the subspace $Y \subseteq X$ without reference to the metric. If $X$ is just a topological space (not necessarily a metric space) and if $Y \subseteq X$ is a subset, this shows how to provide $Y$ with the structure of a topological space, by declaring the open sets in $Y$ to be the intersections of $Y$ with open sets in $X$.

CHAPTER 5

# Interiors, closures, limit points

In this chapter we explore some further concepts in the basic theory of metric spaces.

## 5.1. Interiors and closures

DEFINITION 5.1.1. Let $X$ be a metric space, and let $S \subset X$. The *interior* $\text{int}(S)$ of $S$ is defined to be the union of all open subsets of $X$ contained in $S$. The closure $\bar{S}$ is defined to be the intersection of all closed subsets of $X$ containing $S$. The set $\bar{S} \backslash \text{int}(S)$ is known as the *boundary* of $S$ and denoted $\partial S$. A set $S \subseteq X$ is said to be *dense* if $\overline{S} = X$.

It is very important to note that, in Definition 5.1.1, the notion of open and closed is here being taken in the metric space $X$, *not* in the subspace metric on $S$, which would result in trivial definitions.

Since an arbitrary union of open sets is open (Lemma 4.2.1), $\text{int}(S)$ is itself an open set, and it is clearly the unique largest open subset of $X$ contained in $S$. If $S$ is itself open then evidently $S = \text{int}(S)$.

Since an arbitrary intersection of closed sets is closed, $\bar{S}$ is the unique smallest closed subset of $X$ containing $S$. If $S$ is itself closed then evidently $S = \overline{S}$.

If $x \in \text{int}(S)$ we say that $x$ is an *interior point* of $S$. One can also phrase this in terms of neighbourhoods: the interior of $S$ is the set of all points in $S$ for which $S$ is a neighbourhood.

EXAMPLE 5.1.2. If $S = [a, b]$ is a closed interval in $\mathbf{R}$ then its interior is just the open interval $(a, b)$. If we take $S = \mathbb{Q} \subset \mathbf{R}$ then $\text{int}(\mathbb{Q}) = \emptyset$.

EXAMPLE 5.1.3. The rationals $\mathbb{Q}$ are a dense subset of $\mathbf{R}$, as is the set $\{\frac{a}{2^n} : a \in \mathbb{Z}, n \in \mathbb{N}\}$.

Let us give a couple of simple characterisations of the closure of a set.

LEMMA 5.1.4. *Let $X$ be a metric space, and let $S \subseteq X$ be a subset. Then $a \in \bar{S}$ if and only if the following is true: every open ball $B(a, \varepsilon)$ contains a point of $S$.*

*Proof.* Suppose $a \in \bar{S}$. If $B(a,\varepsilon)$ does not meet $S$, then $B(a,\varepsilon)^c$ is a closed set containing $S$. Therefore $B(a,\varepsilon)^c$ contains $\bar{S}$, and hence it contains $a$, which is obviously nonsense.

Conversely, suppose that every ball $B(a,\varepsilon)$ meets $S$. If $a \notin \bar{S}$ then, since $\bar{S}^c$ is open, there is a ball $B(a,\varepsilon)$ contained in $\bar{S}^c$, and hence in $S^c$, contrary to assumption. □

*Remark.* A particular consequence of this is that $S \subseteq X$ is dense if and only if it meets every open set in $X$.

COROLLARY 5.1.5. *Let $X$ be a metric space, and let $S \subseteq X$ be a subset. Let $a \in X$. Then $a$ lies in the closure $\bar{S}$ if and only if there is a sequence $(x_n)_{n=1}^{\infty}$ of elements of $S$ with $\lim_{n\to\infty} x_n = a$. In particular, $S$ is closed if and only if the limit of every convergent sequence $(x_n)_{n=1}^{\infty}$ of elements of $S$ lies in $S$.*

*Proof.* We use Lemma 5.1.4. Suppose that $a \in \bar{S}$. Then by Lemma 5.1.4 every ball $B(a,1/n)$ contains a point of $S$, so we may pick a sequence $(x_n)_{n=1}^{\infty}$ with $x_n \in B(a,1/n) \cap S$. Clearly $\lim_{n\to\infty} x_n = a$.

Conversely, suppose $\lim_{n\to\infty} x_n = a$, where $x_n \in S$. If $a \notin \bar{S}$ then by Lemma 5.1.4 there must be some ball $B(a,\varepsilon)$ not meeting $S$. But if $n$ is large enough then $d(x_n,a) < \varepsilon$, and so $x_n \in S \cap B(a,\varepsilon)$, contradiction. □

We conclude with a cautionary example which has often confused people when they first meet it.

EXAMPLE 5.1.6. In general, it need *not* be the case that $\bar{B}(a,\varepsilon)$ is the closure of $B(a,\varepsilon)$. Since we have seen that $\bar{B}(a,\varepsilon)$ is closed, it is always true that $\overline{B(a,\varepsilon)} \subseteq \bar{B}(a,\varepsilon)$, but the containment can be proper. Indeed, take any set $X$ with at least two elements equipped with the discrete metric. Then if $x \in X$ we have $B(x,1) = \overline{B(x,1)} = \{x\}$, but $\bar{B}(x,1)$ is the whole space $X$.

## 5.2. Limit points

This section introduces the notion of *limit points* (also known in some places as cluster points or accumulation points). The notion is a well-studied one, introduced here for cultural reference, but we will not come across it in subsequent chapters of the course.

DEFINITION 5.2.1. If $X$ is a metric space and $S \subseteq X$ is any subset, then we say a point $a \in X$ is a *limit point* of $S$ if any open ball about $a$ contains a point of $S$ other than $a$ itself.

We will write $L(S)$ for the set of limit points of $S$; I am not sure that there is any completely standard notation for this. Note that we do not necessarily have

$S \subseteq L(S)$, that is to say it is quite possible for a point $a \in S$ not to be a limit point of $S$. This occurs if there is some ball $B(a, \varepsilon)$ such that $B(a, \varepsilon) \cap S = \{a\}$, and in this case we say that $a$ is an *isolated point* of $S$.

EXAMPLE 5.2.2. Take $X = \mathbf{R}$ and $S = (0, 1] \cup \{2\}$. Then $L(S) = [0, 1]$. Note in particular that 0 does not lie in $S$, but is a limit point; by contrast, 2 does lie in $S$, but it is not a limit point, so it is an isolated point.

LEMMA 5.2.3. *Let $S$ be a subset of a metric space $X$. Then $L(S)$ is a closed subset of $X$.*

*Proof.* We need to show that the complement $L(S)^c$ is open. Suppose $a \in L(S)^c$. Then there is a ball $B(a, \varepsilon)$ whose intersection with $S$ is either empty or $\{a\}$.

We claim that $B(a, \varepsilon/2) \subseteq L(S)^c$. Let $b \in B(a, \varepsilon/2)$. If $b = a$, then clearly $b \in L(S)^c$. If $b \neq a$, there is some ball about $b$ which is contained in $B(a, \varepsilon)$, but does not contain $a$: the ball $B(b, \delta)$ where $\delta = \min(\varepsilon/2, d(a, b))$ has this property. This ball meets $S$ in the empty set, and so $b \in L(S)^c$ in this case too. □

PROPOSITION 5.2.4. *Let $S$ be a subset of a metric space $X$. Let $L(S)$ be its set of limit points, and $\bar{S}$ its closure. Then $\bar{S} = S \cup L(S)$.*

*Proof.* We first show the containment $S \cup L(S) \subseteq \bar{S}$. Obviously $S \subseteq \bar{S}$, so we need only show that $L(S) \subseteq \bar{S}$. Suppose $a \in \bar{S}^c$. Since $\bar{S}^c$ is open, there is some ball $B(a, \varepsilon)$ which lies in $\bar{S}^c$, and hence also in $S^c$, and therefore $a$ cannot be a limit point of $S$. This concludes the proof of this direction.

Now we look at the opposite containment $\bar{S} \subseteq S \cup L(S)$. If $a \in \bar{S}$, we saw in Lemma 5.1.5 that there is a sequence $(x_n)_{n=1}^{\infty}$ of elements of $S$ with $x_n \to a$. If $x_n = a$ for some $n$ then we are done, since this implies that $a \in S$. Suppose, then, that $x_n \neq a$ for all $n$. Let $\varepsilon > 0$. Then all the $x_n$, for $n$ sufficiently large in terms of $\varepsilon$, are elements of $B(a, \varepsilon) \setminus \{a\}$, and they all lie in $S$. It follows that $a$ is a limit point of $S$, and so we are done in this case also. □

COROLLARY 5.2.5. *Let $S$ be a subset of a metric space $X$. Then $S$ is closed if and only if it contains all its limit points.*

*Proof.* We already remarked, in Section 5.1, that $S$ is closed if and only if $S = \bar{S}$. The corollary is immediate from this and Proposition 5.2.4. □

CHAPTER 6

# Completeness

Students may wish to remind themselves of the Prelims course M2: Analysis I, which covered some of the topics of this section in the specific case of the real numbers. Much of the theory in a general metric space is a natural generalisation of what was done there.

## 6.1. Basic definitions and examples

DEFINITION 6.1.1. Let $(x_n)_{n=1}^{\infty}$ be a sequence in some metric space $X$. Then we say that this sequence is

- *Bounded* if the set $\{x_n : n \geqslant 1\}$ is bounded in the sense of Definition 1.6.2, that is to say if all the $x_n$ lie in some ball $B(a, R)$;
- *Cauchy* if the $x_n$ become arbitrarily close together as $n \to \infty$, in the following sense: for every $\varepsilon > 0$, there is some $N$ such that $d(x_n, x_m) < \varepsilon$ whenever $n, m \geqslant N$.
- *Convergent* if there is some $a \in X$ such that $\lim_{n\to\infty} x_n = a$.

If a sequence $\sigma$ has any one of these properties, then any subsequence of $\sigma$ also has the property. We leave the proof of this as an exercise.

The relation between the above concepts is as follows.

PROPOSITION 6.1.2. *A convergent sequence is Cauchy. A Cauchy sequence is bounded. Neither of the reverse implications holds in general.*

*Proof.* We begin by showing that the reverse implications do not hold, since the examples we will give serve to illustrate the concepts. Take $X = (0, 1]$. Then the sequence $x_n = 1/n$ is Cauchy, but not convergent. The sequence in which $x_n = 1$ for $n$ odd and $x_n = 1/2$ for $n$ even is bounded, but it is not Cauchy since there is no $N$ such that $d(x_n, x_{n+1}) < 1/2$ for all $n \geqslant N$.

Now we show the two main implications. Suppose that $(x_n)_{n=1}^{\infty}$ is convergent, and that $\lim_{n\to\infty} x_n = a$. Let $\varepsilon > 0$. By the definition of limit, there is some $N$ such that, if $n \geqslant N$, $d(x_n, a) < \varepsilon/2$. Now suppose that $m, n \geqslant N$. Then

$$d(x_m, x_n) \leqslant d(x_n, a) + d(x_m, a) < \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

and so $(x_n)_{n=1}^{\infty}$ is Cauchy.

Now suppose that $(x_n)_{n=1}^{\infty}$ is Cauchy. Taking $\varepsilon = 1$ in the definition, we see that there is some $N$ such that $d(x_m, x_n) < 1$ whenever $m, n \geqslant N$. In particular, all points of the sequence except (possibly) $x_1, \ldots, x_{N-1}$ lie in $B(x_N, 1)$. It follows that *all* points of the sequence lie in $B(x_N, R)$, where $R$ is the largest value of the set $\{1, d(x_N, x_1), \ldots, d(x_N, x_{N-1})\}$, and so $(x_n)_{n=1}^{\infty}$ is bounded. □

Now we turn to the key definition of the chapter.

Definition 6.1.3 (Completeness). A metric space is said to be *complete* if every Cauchy sequence converges.

One of the main results of the Prelims course was that $\mathbf{R}$ is complete, and it is easy to deduce from this that $\mathbf{R}^n$ is complete also (since a sequence in $\mathbf{R}^n$ converges if and only if each of its coordinates converge).

On the other hand, we observed above that $(0, 1]$ is not complete. For much the same reason, $(0, 1)$ is not complete. Note, however, that $(0, 1)$ is homeomorphic to $\mathbf{R}$, as we showed earlier. Therefore the notion of completeness is not (necessarily) preserved under homeomorphisms.

Let $V$ be a normed vector space with norm $\|\cdot\|$. As previously discussed, we can define a metric on $V$ by $d(v, w) = \|v - w\|$. We say that $V$ is complete if, when endowed with the structure of a metric space in this way, it is complete. That is, when we talk about completeness of normed spaces we implicitly assume that the obvious metric has been put on $V$, without necessarily mentioning it explicitly.

## 6.2. First properties of complete metric spaces

In this section we collect a couple of basic properties of complete metric spaces.

Lemma 6.2.1. *A subspace of a complete metric space is complete if and only if it is closed.*

*Proof.* Let $X$ be a complete metric space and suppose that $Y \subseteq X$. Suppose first that $Y$ is closed; we will show that it is complete. Let $(y_n)_{n=1}^{\infty}$ be a Cauchy sequence in $Y$. Then it is also a Cauchy sequence in $X$. Since $X$ is complete, it converges, say $\lim_{n\to\infty} y_n = a$. By Corollary 5.1.5, $a \in Y$.

In the other direction, suppose that $Y$ is complete. Let $(y_n)_{n=1}^{\infty}$ be a sequence of elements of $Y$ with $\lim_{n\to\infty} y_n = a$. Then $(y_n)_{n=1}^{\infty}$ is certainly a Cauchy sequence, and so by completeness it has a subsequence which converges to an element of $Y$. Since this subsequence must also converge to $a$, it follows that $a \in Y$. By Corollary 5.1.5, $Y$ is closed. □

The next lemma is sometimes known as Cantor's intersection theorem.

LEMMA 6.2.2. *Let $X$ be a complete metric space and suppose that $S_1 \supseteq S_2 \supseteq \ldots$ form a nested sequence of non-empty closed sets in $X$ with the property that $\operatorname{diam}(S_n) \to 0$ as $n \to \infty$. Then $\bigcap_{n=1}^{\infty} S_n$ contains a unique point $a$.*

*Proof.* For each $n$, pick $x_n \in S_n$. We claim that $(x_n)_{n=1}^{\infty}$ is Cauchy. To see this, let $\varepsilon > 0$, and suppose that $N$ is large enough that $\operatorname{diam}(S_N) < \varepsilon$. If $n, m \geqslant N$ then, since the $S_i$ are nested, $x_n, x_m \in S_N$. By the definition of diameter, $d(x_n, x_m) \leqslant \operatorname{diam}(S_N) < \varepsilon$.

Since $X$ is complete, we have $\lim_{n\to\infty} x_n = a$ for some $a$. For each $i$, the nesting property of the sets $S_i$ implies that we have $x_n \in S_i$ for all $i \leqslant n$. Therefore, since $S_i$ is closed, Corollary 5.1.5 tells us that $a \in S_i$. Since this is true for all $i$, we have $a \in \bigcap_{i=1}^{\infty} S_i$.

To show that $a$ is unique, suppose that $b \in \bigcap_{i=1}^{\infty} S_i$. Then $d(a, b) \leqslant \operatorname{diam}(S_i)$ for all $i$. Since $\operatorname{diam}(S_i) \to 0$, we have $d(a, b) = 0$ and so $a = b$. □

What if we drop the condition $\operatorname{diam}(S_i) \to 0$? We certainly could not expect $a$ to be unique since, for instance, we could take all the $S_i$ to be the whole space $X$. Somewhat surprisingly at first sight, the intersection $\bigcap_{i=1}^{\infty} S_i$ may even be empty. For instance, take $S_i = [i, \infty) \subset \mathbf{R}$.

## 6.3. Completeness of function spaces

In this section we show that two natural spaces of functions give rise to complete metric spaces.

For the first result, recall that if $X$ is a set then $B(X)$ denotes the normed vector space of bounded functions $f : X \to \mathbf{R}$, with norm $\|f\|_\infty = \sup_{x\in X} |f(x)|$.

THEOREM 6.3.1. *Let $X$ be any set. Then $B(X)$ is complete.*

*Proof.* Let $(f_n)_{n=1}^{\infty}$ be a Cauchy sequence in $B(X)$. Then for each $x$ the sequence $(f_n(x))_{n=1}^{\infty}$ is a Cauchy sequence of real numbers (convincing yourself of this is a good exercise to check you have understood the definitions). Since $\mathbf{R}$ is complete, each such sequence has a limit, and we write $f(x)$ for this limit. That is, $\lim_{n\to\infty} f_n(x) = f(x)$.

We claim that $f$ is a bounded function. To see this, take $\varepsilon = 1$ in the definition of Cauchy sequence. This gives an $N$ such that, if $n, m \geqslant N$, $\sup_x |f_n(x) - f_m(x)| \leqslant 1$. In particular, $|f_N(x) - f_n(x)| \leqslant 1$ for all $n \geqslant N$ and for all $x \in X$. Taking the limit as $n \to \infty$, it follows that $|f_N(x) - f(x)| \leqslant 1$ for all $x$. Since $f_N$ is a bounded function, so is $f$.

Finally, we need to show that $f_n \to f$ in the norm $\|\cdot\|_\infty$ (at the moment we have only shown pointwise convergence). The argument is a simple modification of the preceding one. Let $\varepsilon > 0$, and let $N$ be such that, if $n, m \geqslant N$, $|f_n(x) - f_m(x)| \leqslant \varepsilon$

for all $x \in X$. For each fixed $n \geqslant N$ and $x \in X$, we may let $m \to \infty$, obtaining that $|f_n(x) - f(x)| \leqslant \varepsilon$. That is, for all $n \geqslant N$ we have $\|f_n - f\|_\infty \leqslant \varepsilon$. It follows that $f_n \to f$ in the $\|\cdot\|_\infty$-norm. □

For the second result, recall that if $X$ is a metric space then $C_b(X)$ denotes the normed vector space of bounded continuous functions $f : X \to \mathbf{R}$, again with norm $\|f\|_\infty = \sup_{x \in X} |f(x)|$.

THEOREM 6.3.2. *Let $X$ be a metric space. Then $C_b(X)$ is complete.*

*Proof.* We have shown in Theorem 6.3.1 that $B(X)$ is complete, so by Lemma 6.2.1 it is enough to show that $C_b(X)$ is a closed subset of $B(X)$.

By Corollary 5.1.5, it suffices to show that if $(f_n)_{n=1}^\infty$ is a sequence of elements of $C_b(X)$ converging in the $\|\cdot\|_\infty$-norm to some $f \in B(X)$, then $f \in C_b(X)$, or in other words $f$ is continuous.

Let $a \in X$, and let $\varepsilon > 0$. Since $f_n \to f$ in the $\|\cdot\|_\infty$-norm, there is some $n$ such that $\|f_n - f\|_\infty \leqslant \varepsilon/3$. Since $f_n$ is continuous, there is a $\delta > 0$ such that $|f_n(x) - f_n(a)| < \varepsilon/3$ for all $x \in B(a, \delta)$. But then for $x \in B(a, \delta)$ we have

$$\begin{aligned} |f(x) - f(a)| &\leqslant |f(x) - f_n(x)| + |f_n(x) - f_n(a)| + |f_n(a) - f(a)| \\ &< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon. \end{aligned}$$

It follows that $f$ is continuous at $a$, and since $a$ was arbitrary, $f$ is a continuous function on $X$. □

*Remark.* You may have the impression that you have seen something like this argument before, and indeed that is the case. In Prelims: Analysis II you saw that a uniform limit of continuous functions on $\mathbf{R}$ is continuous, and our task here was essentially the same, but in the setting of a general metric space.

## 6.4. The contraction mapping theorem

The final topic of this section is a classic theorem about fixed points of certain maps from a metric space to itself. We will discuss the result for its own intrinsic interest, but it has important applications to the solutions of differential equations, as you will see in the course A1 : Differential Equations.

Let us begin with a couple of definitions.

DEFINITION 6.4.1. Let $(X, d_X)$ and $(Y, d_Y)$ be metric spaces and suppose that $f\colon X \to Y$. We say that $f$ is a *Lipschitz* map (or is *Lipschitz continuous*) if there is a constant $K \geqslant 0$ such that

$$d_Y(f(x), f(y)) \leqslant K d_X(x, y).$$

If $Y = X$ and $K \in [0, 1)$ then we say that $f$ is a *contraction mapping* (or simply a *contraction*).

An easy exercise is to check that every Lipschitz map is continuous, and to give an example of a continuous map between metric spaces which is *not* Lipschitz.

It is very important to note, in the definition of a contraction, that it says something stronger than that $d(f(x), f(y)) < d(x, y)$, namely that there is a constant $K < 1$ such that $d(f(x), f(y)) \leqslant Kd(x, y)$ for all $x, y$.

THEOREM 6.4.2 (Contraction mapping theorem). *Let $X$ be a nonempty complete metric space and suppose that $f\colon X \to X$ is a contraction. Then $f$ has a unique fixed point, that is, there is a unique $x \in X$ such that $f(x) = x$.*

*Proof.* We begin by showing that there cannot be two fixed points. Suppose that $f(x_1) = x_1$ and that $f(x_2) = x_2$. Then we have

$$d(x_1, x_2) = d(f(x_1), f(x_2)) \leqslant Kd(x_1, x_2).$$

Since $d(x_1, x_2) \geqslant 0$ and $K < 1$, we are forced to conclude that $d(x_1, x_2) = 0$ and hence that $x_1 = x_2$.

Now we show that there is a fixed point. The proof is constructive (and may be used in practical situations to find fixed points numerically). The idea is as follows. Pick an arbitrary $x_0 \in X$, and form the sequence of iterates $x_1 := f(x_0)$, $x_2 := f(x_1)$, and so on. We claim that (no matter which $x_0$ we started with) the sequence $(x_n)_{n=1}^{\infty}$ converges to some limit $x$, and that $f(x) = x$.

To show that $(x_n)_{n=1}^{\infty}$ converges, it suffices to show that it is Cauchy, since $X$ is complete. To do this, first observe that by repeated use of the contraction property and the definition of the sequence $(x_n)_{n=1}^{\infty}$ we have

$$d(x_n, x_{n-1}) \leqslant Kd(x_{n-1}, x_{n-2}) \leqslant K^2 d(x_{n-2}, x_{n-3}) \leqslant \ldots \leqslant K^{n-1} d(x_0, x_1)$$

(you could prove this formally by induction if you wanted). Therefore if $n > m$ we have

$$\begin{aligned} d(x_n, x_m) &\leqslant d(x_n, x_{n-1}) + \cdots + d(x_{m+1}, x_m) \\ &\leqslant (K^{n-1} + K^{n-2} + \cdots + K^m) d(x_0, x_1) \\ &\leqslant K^m (1 + K + K^2 + \ldots) d(x_0, x_1) = CK^m, \end{aligned}$$

where $C = d(x_0, x_1)/(1 - K)$ (by summing the geometric series).

It follows that if $n, m \geqslant N$ then $d(x_m, x_n) \leqslant CK^N$.

Since $K < 1$, for any $\varepsilon > 0$ there is some $N$ such that $CK^N < \varepsilon$, and therefore $(x_n)_{n=1}^{\infty}$ is indeed a Cauchy sequence.

Since $X$ is complete, $x_n \to x$ for some $x \in X$. To complete the proof we must show that $f(x) = x$. This is quite straightforward. Indeed, since $f$ is continuous

we have

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

This finishes the proof. □

*Remarks.* Over the years, many people have lost a mark in exam questions for forgetting that $X$ must be non-empty.

Let us conclude by giving a couple of examples to show that the hypotheses of the theorem are necessary. First, we observe that the weaker condition that $d(f(x), f(y)) < d(x, y)$ for all $x \neq y$ is *not* sufficient. For instance, it may be checked that the function $f\colon [1, \infty) \to [1, \infty)$ defined by $f(x) = x + 1/x$ has this property, but it obviously has no fixed points.

More obviously, the requirement that $X$ is complete is important. For instance, if we define $f\colon (0, 1) \to (0, 1)$ by $f(x) = x/2$ then clearly $f$ is a contraction, but $f$ has no fixed points in $(0, 1)$.

## 6.5. *Completions

In this section we mention the notion of the *completion* of a metric space. This is explicitly declared to be non-examinable in the schedules.

The idea is that one may take an arbitrary metric space $X$ and "add in the limits of all Cauchy sequences" to get a new, complete, space $\tilde{X}$.

The space $\tilde{X}$ consists of all Cauchy sequence $(x_n)_{n=1}^{\infty}$ in $X$, modulo a natural notion of equivalence: two Cauchy sequences $(x_n)_{n=1}^{\infty}$ and $(x'_n)_{n=1}^{\infty}$ are said to be equivalent if and only if $x_n - x'_n \to 0$ as $n \to \infty$. It is not too hard to check that this does give an equivalence relation. Write $[(x_n)_{n=1}^{\infty}]$ for the equivalence class of $(x_n)_{n=1}^{\infty}$.

To give $\tilde{X}$ the structure of a metric space, we define

$$\tilde{d}([(x_n)_{n=1}^{\infty}], [(x'_n)_{n=1}^{\infty}]) = \lim_{n\to\infty} d(x_n, x'_n).$$

There is a lot to be checked here: that the limit even exists and that it does not depend on which representatives $(x_n)_{n=1}^{\infty}, (x'_n)_{n=1}^{\infty}$ one takes in a given equivalence class, and finally that it is a distance.

Once this has been established, there are a number of other natural statements to be proven about $\tilde{X}$, which include the following:

- $\tilde{X}$ is a complete metric space;
- There is a natural map $\iota : X \to \tilde{X}$ given by $\iota(x) = [(x, x, x, \dots, )]$. It is continuous, injective and has dense image.

Example 6.5.1. The completion of $\mathbf{Q}$ with respect to the usual metric is (isometrically equivalent to) the real numbers $\mathbf{R}$.

As we noted at the start of the course, you have not been shown a proof that the real numbers $\mathbf{R}$ exist. Can we *define* them as the completion of $\mathbf{Q}$ with respect to the usual metric? Not exactly, because we have used the notion of $\mathbf{R}$ throughout our development of the theory of metric spaces, so there is some circularity. In particular, the definition of the metric $\tilde{d}$ makes crucial use of $\mathbf{R}$ in defining the limit $\lim_{n\to\infty} d(x_n, x'_n)$.

That said, arguably the most natural construction of $\mathbf{R}$ *is* as equivalence classes of Cauchy sequences in $\mathbf{Q}$ – but this needs to be done separately from the general definition of completion, and indeed should really be done before even beginning a discussion of metric spaces, which involve $\mathbf{R}$ in their definition.

If you are interested in how to construct the reals (which I think you should be!) then the Wikipedia page on the subject is a good place to start.

EXAMPLE 6.5.2. A very important example of a completion in number theory is the construction of the $p$-adics $\mathbf{Q}_p$, which is the completion of $\mathbf{Q}$ with respect to the $p$-adic metric, which we briefly discussed in Example 1.3.5.

CHAPTER 7

# Connectedness and path-connectedness

In this section we try to understand what makes a space "connected". We will consider two natural approaches to this question, and show that for reasonably nice spaces the two notions in fact coincide.

In particular, the two notions of connectedness coincide for open subsets of the complex plane, which will be our main subject of interest in the second half of the course.

## 7.1. Connectedness

The concept of *connectedness* formulates the intuitive idea of a space which cannot be split into two "separated" pieces.

DEFINITION 7.1.1. We say that a metric space is *disconnected* if we can write it as the disjoint union of two nonempty open sets. We say that a space is *connected* if it is not disconnected.

If $X$ is written as a disjoint union of two nonempty open sets $U$ and $V$ then we say that these sets *disconnect* $X$.

If $X = [0,1] \cup [2,3] \subset \mathbf{R}$ then we have seen that both $[0,1]$ and $[2,3]$ are open in $X$. Since $X$ is their disjoint union, $X$ is disconnected.

It is a little harder to give a nontrivial example of a connected space. Later on, we will show that all intervals in $\mathbf{R}$ are connected.

The following lemma gives some equivalent ways to formulate the concept of connected space.

LEMMA 7.1.2. *Let $X$ be a metric space. Then the following are equivalent.*

(i) *$X$ is connected.*

(ii) *If $f\colon X \to \{0,1\}$ is a continuous function then $f$ is constant.*

(iii) *The only subsets of $X$ which are both open and closed are $X$ and $\emptyset$.*

(*Here the set $\{0,1\}$ is viewed as a metric space via its embedding in $\mathbf{R}$, or equivalently with the discrete metric.*)

*Proof.* (i) $\Rightarrow$ (ii): Let $X$ be connected, and let $f\colon X \to \{0,1\}$ be a continuous function. The singleton sets $\{0\}$ and $\{1\}$ are both open in $\{0,1\}$ and so both $f^{-1}(0)$

and $f^{-1}(1)$ are open subsets of $X$.They are clearly disjoint, and their union is $X$. Therefore one of them must be empty, which means that $f$ is constant.

(ii) $\Rightarrow$ (iii): Suppose that $A \subseteq X$ is both open and closed. Then $A^c$ is open (and closed), and so the function $f : X \to \{0, 1\}$ defined by $f(x) = 1$ for $x \in A$ and $f(x) = 0$ for $x \in A^c$ (that is, the characteristic function of $A$) is continuous. Assuming (ii), it must be constant. If it takes the constant value 1, then $A = X$. If it takes the constant value 0, then $A = \emptyset$.

(iii) $\Rightarrow$ (i): Suppose that $X = U \cup V$ with $U, V$ open and disjoint. Then $U^c = V$ is open, so $U$ is also closed. Thus $U$ is both open and closed, and hence (assuming (iii)) is either $X$ or $\emptyset$. Similarly for $V$. Hence there is no way to disconnect $f(X)$. □

Frequently one has a metric space $X$ and a subset $Y$ of it whose connectedness or otherwise one wishes to ascertain. To this end, it is useful to record the following lemma.

LEMMA 7.1.3. *Let $X$ be a metric space, and let $Y \subseteq X$ be a subset, considered as a metric space with the metric induced from $X$. Then $Y$ is connected if and only if the following is true. If $U, V$ are open subsets of $X$, and $U \cap V \cap Y = \emptyset$, then whenever $Y \subseteq U \cup V$, either $Y \subseteq U$ or $Y \subseteq V$.*

*Proof.* The key point here is to recall that the open sets in $Y$ are precisely the sets of the form $U \cap Y$, where $U$ is open in $X$. This was proven in Lemma 4.5.1. Take a pair $U \cap Y$, $V \cap Y$ of such open sets. They disconnect $Y$ if and only if

(i) They are disjoint, thus $U \cap V \cap Y = \emptyset$;
(ii) They cover $Y$, which is equivalent to $Y \subseteq U \cup V$;
(iii) Neither is empty.

Thus $Y$ is connected if and only if (i) and (ii) imply that one of $U \cap Y$, $V \cap Y$ is empty or equivalently that $Y \subseteq V$ or $Y \subseteq U$. □

We now turn to some basic properties of the notion of connectedness. These broadly conform with one's intuition about how connected sets should behave, but of course proof is required in each case.

LEMMA 7.1.4 (Sunflower lemma). *Let $X$ be a metric space. Let $\{A_i : i \in I\}$ be a collection of connected subsets of $X$ such that $\bigcap_{i\in I} A_i \neq \emptyset$. Then $\bigcup_{i\in I} A_i$ is connected.*

*Proof.* We use the alternative characterisation of connectedness given in Lemma 7.1.2 (ii). Suppose that $f\colon \bigcup_{i\in I} A_i \to \{0, 1\}$ is continuous. We must show that $f$ is constant. Pick $x_0 \in \bigcap_{i\in I} A_i$. Then if $x \in \bigcup_{i\in I} A_i$ there is some $i$ for which

$x \in A_i$. But then the restriction of $f$ to $A_i$ is constant since $A_i$ is connected, so that $f(x) = f(x_0)$ as $x, x_0 \in A_i$. But since $x$ was arbitrary, it follows that $f$ is constant as required. □

LEMMA 7.1.5 (Connectedness and closures). *Let $X$ be a metric space. If $A \subseteq X$ is connected then if $B$ is such that $A \subseteq B \subseteq \bar{A}$, the set $B$ is also connected.*

*Proof.* We use the criterion for a subspace to be connected from Lemma 7.1.3. Suppose that $B \subseteq U \cup V$ where $U$ and $V$ are open in $X$ and $U \cap V \cap B = \emptyset$. Then certainly $A \subseteq U \cup V$ and $A \cap U \cap V = \emptyset$. Hence, since $A$ is connected, either $A \subseteq U$ or $A \subseteq V$. Without loss of generality, $A \subseteq U$, and since $A \cap U \cap V = \emptyset$ this means that $A \subseteq V^c$. However, $V^c$ is closed and so taking closures we obtain $\bar{A} \subseteq \bar{V^c} = V^c$. In particular $B \subseteq V^c$ and so, since $B \subseteq U \cup V$, we must have $B \subseteq U$. We have verified the criterion (Lemma 7.1.3) for a subspace to be connected. □

LEMMA 7.1.6 (Connected image of a connected set). *Let $X$ be a connected metric space, and let $f : X \to Y$ be continuous. Then $f(X)$ is connected.*

*Proof.* We may as well suppose that $f$ is surjective (otherwise replace $Y$ by $f(X)$). Suppose that $U$ and $V$ are disjoint open subsets of $Y$ with $U \cup V = Y$. Then $f^{-1}(U)$ and $f^{-1}(V)$ are disjoint open subsets of $X$ with $f^{-1}(U) \cup f^{-1}(V) = X$. Since $X$ is connected one of them, say $f^{-1}(U)$, is empty. Therefore $U$ is empty.

It follows that there is no way to disconnect $X$. □

A simple corollary is that (unlike completeness) the property of connectedness is preserved under homeomorphisms.

*Connected components.* A consequence of the Sunflower Lemma is that, for each $x \in X$, there is a unique maximal connected subset of $X$ containing $x$, which contains all other such sets (take the union of all connected subsets of $X$ containing $x$). This is called the *connected component* of $X$ containing $x$.

PROPOSITION 7.1.7 (Connected components). *The connected components of a metric space partition the space. A space is connected if and only if it has a unique connected component.*

*Proof.* Let $X$ be the space, and for $x \in X$ write $\Gamma(x)$ for the connected component containing $x$. Suppose that $\Gamma(x)$ and $\Gamma(y)$ are not disjoint, say $a \in \Gamma(x) \cap \Gamma(y)$. We wish to show that they coincide, which is what it means for them to partition the space. By the Sunflower Lemma, $\Gamma(x) \cup \Gamma(y)$ is connected. By the definition of connected component, $\Gamma(x)$ must contain this set, which of course means that $\Gamma(y) \subseteq \Gamma(x)$. Similarly $\Gamma(x) \subseteq \Gamma(y)$, and so $\Gamma(x) = \Gamma(y)$.

The second statement is obvious. □

## 7.2. *Connected subsets of R

In this section we classify the connected subsets of $\mathbf{R}$, showing that they are precisely the intervals. For the purposes of this section, the word *interval* includes half-infinite or infinite intervals, and intervals can be open or closed at either end. Thus the sets we are talking about are

$$[a, \infty), (a, \infty), (-\infty, a) \text{ and } (-\infty, a],$$

together with all bounded intervals

$$(a, b), (a, b], [a, b) \text{ and } [a, b] \text{ for } a, b \in \mathbf{R} \text{ with } a \leqslant b.$$

Note that singleton sets $\{a\}$ are intervals, as is the empty set.

THEOREM 7.2.1. *A subset of* $\mathbf{R}$ *is connected if and only if it is an interval.*

*Proof.* We will prove this theorem, which is not by any means trivial, in two stages. To this end, let us make a definition. Let $E \subseteq \mathbf{R}$ be a subset of the real line. We say that $E$ has the *interval property* if, whenever $x < y$ both lie in $E$, we have $[x, y] \subseteq E$.

The theorem is a consequence of the following two assertions:

(1) A subset of the real line $\mathbf{R}$ is connected if and only if it has the interval property;

(2) A subset of the real line has the interval property if and only if it is itself an interval.

We begin with (1). Suppose that $E$ is connected and that $x, y \in E$. By symmetry we may assume that $x < y$. If $[x, y]$ is not entirely contained in $E$, we may find $c \in (x, y)$ such that $c \notin E$. Take $U = (-\infty, c)$ and $V = (c, \infty)$. Clearly $U \cap V \cap E = \emptyset$, $E \subseteq U \cup V$, but we do not have $E \subseteq U$ (since $y \in E \setminus U$) or $E \subseteq V$ (since $x \in E \setminus V$). By Lemma 7.1.3, $E$ is not connected.

In the other direction, suppose that $E$ has the interval property. We wil show that $E$ is connected using Lemma 7.1.3. Suppose $E \subseteq U \cup V$ where $U$ and $V$ are open subsets of $\mathbf{R}$ with $E \cap U \cap V = \emptyset$, but that we do not have $E \subseteq U$ or $E \subseteq V$. Thus $E \cap U$ and $E \cap V$ are both non-empty. Let $x \in E \cap U$ and $y \in E \cap V$. Since $E \cap U \cap V = \emptyset$, $x$ and $y$ are distinct, and we may assume without loss of generality that $x < y$. Since $E$ has the interval property, $[x, y]$ is entirely contained in $E$.

Now define $S = \{z \in [x, y] : z \in U\}$. Then $S$ is non-empty and bounded and so $c = \sup(S)$ exists. Clearly $c \in [x, y]$. Since $[x, y] \subseteq E \subseteq U \cup V$, we have either $c \in U$ or $c \in V$.

If $c \in U$ then $c \neq y$ and so, since $U$ is open, there is some interval $[c, c+\varepsilon)$ contained in $U$ and also in $[x, y]$. This means that $[c, c+\varepsilon) \subseteq S$, which contradicts the fact that $c = \sup(S)$ (for instance, $c + \varepsilon/2$ lies in $S$ and is bigger than $c$).

If $c \in V$ then $c \neq x$ and so, since $V$ is open, there is some interval $(c-\varepsilon, c]$ contained in $V$ and also in $[x, y]$. In particular, $[c-\varepsilon/2, c]$ is disjoint from $S$, which contradicts the fact that $c = \sup(S)$ (for instance, $c - \varepsilon/2$ as an upper bound for $S$, and is smaller than $c$).

These two contradictions show that we were wrong to assume that neither $E \subseteq U$ or $E \subseteq V$. Therefore $E$ is connected. This concludes the proof of statement (1).

We turn now to statement (2). Here it is convenient to abuse some standard notation. In particular, we let $\inf(E)$ take the value $-\infty$ (if $E$ is not bounded from below) and $\sup(E)$ take the value $\infty$ (if $E$ is not bounded from above). Also, to save splitting into large numbers of cases, we allow ourselves to write $[-\infty, a]$, when really we mean the half-line $(-\infty, a]$. With these abuses of notation in place, suppose that $E$ has the interval property. Write $c = \inf(E)$ and $C = \sup(E)$ (where, as just discussed, these can take the values $-\infty$ and $\infty$ respectively. We claim that

$$(c, C) \subseteq E \subseteq [c, C], \tag{7.1}$$

which is easily seen to imply that $E$ is one of the sets listed at the start of the section. The right-hand inclusion is immediate from the definition of inf and sup. To show the left-hand inclusion, suppose that $z \in (c, C)$. Then there is some $x \in E$ with $c \leqslant x < z$, or else $z$ would be a lower bound for $E$, larger than $c$. Similarly, there is some $y \in E$ with $z < y \leqslant C$. By the interval property, $[x, y] \subseteq E$. But $z \in [x, y]$, and so $z \in E$. This concludes the proof. □

To finish this section, let us remark that the intermediate value theorem is an almost immediate consequence of Theorem 7.2.1 and Lemma 7.1.6. Indeed, suppose $f : [a, b] \to \mathbf{R}$ is continuous. Then, since $[a, b]$ is connected, $f([a, b])$ is connected. Therefore this latter set is an interval and in particular it contains every $c$ with between $f(a)$ and $f(b)$.

## 7.3. Path-connectedness

We now turn to a different, but equally intuitive, notion of what it means for a set to be connected: that one should be able to "continuously move" from any point to another. Here is the precise definition.

Definition 7.3.1 (Path connectedness). Let $X$ be a metric space. Then we say that $X$ is path-connected if the following is true: for any $a, b \in X$ there is a continuous map $\gamma : [0, 1] \to X$ with $\gamma(0) = a$ and $\gamma(1) = b$.

A continuous map $\gamma : [0,1] \to X$ is called a *path*. To develop the basic theory of path-connectedness, we introduce a couple of simple operations on paths.

Given two paths $\gamma_1, \gamma_2$ in $X$ such that $\gamma_1(1) = \gamma_2(0)$ we can form the *concatenation* $\gamma_1 \star \gamma_2$ of the two paths to be the path

$$\gamma_1 \star \gamma_2(t) = \begin{cases} \gamma_1(2t), & 0 \leqslant t \leqslant 1/2 \\ \gamma_2(2t-1), & 1/2 \leqslant t \leqslant 1 \end{cases}$$

We leave it as an easy exercise to show carefully that $\gamma_1 \star \gamma_2$ is continuous, and hence really is a path.

If $\gamma\colon [0,1] \to X$ is a path, then the *opposite* path $\gamma^-$ is defined by $\gamma^-(t) = \gamma(1-t)$.

LEMMA 7.3.2. *Let $X$ be a metric space. Define a relation $\sim$ on $X$ as follows: $a \sim b$ if and only if there is a path $\gamma : [0,1] \to X$ with $\gamma(0) = a$ and $\gamma(1) = b$. Then $\sim$ is an equivalence relation.*

*Proof.* To show that $a \sim a$, use the path $\gamma$ which takes the constant value $a$. To show that $a \sim b$ implies $b \sim a$, take a path $\gamma$ from $a$ to $b$ and consider its opposite path $\gamma^-$. Finally, to show transitivity, use the join of two paths. □

The equivalence classes into which this relation partitions $X$ are called the *path-components* of $X$.

## 7.4. Connectedness and path-connectedness

In the final part of this chapter, we explore the link between connectedness and path-connectedness. The key points to be covered are as follows:

- Path-connectedness implies connectedness;
- Connectedness does not imply path-connectedness in general, but it does in normed vector spaces.

THEOREM 7.4.1. *A path-connected metric space is connected.*

*Proof.* Suppose that $X$ is path-connected, and let $f : X \to \{0,1\}$. We claim that $f$ is constant, which is enough to establish connectedness of $X$ by Lemma 7.1.2 (ii). Let $a, b \in X$, Since $X$ is path-connected, there is a path $\gamma\colon [0,1] \to X$ such that $\gamma(0) = a$ and $\gamma(1) = b$. Consider the composition $f \circ \gamma$. This is a continuous function from $[0,1]$ to $\{0,1\}$ and hence, since $[0,1]$ is connected, it is constant. Therefore $f(a) = (f \circ \gamma)(0) = (f \circ \gamma)(1) = f(b)$. Since $a$ and $b$ were arbitrary, this implies that $f$ is indeed constant. □

THEOREM 7.4.2. *A connected open subset of a normed space is path-connected.*

*Proof.* Write $X$ for the connected open set. The key observation is that any path-component of $X$ is open. To see this, suppose that $P$ is a path-component of $X$, and let $a \in P$. Since $X$ is open, there is a ball $B(a, \varepsilon)$ contained in $X$. Let $b$ be a point in this ball. We can now write down an explicit path $\gamma$ between $a$ and $b$, namely $\gamma(t) = (1-t)a + tb$. This is easily seen to be continuous, and its image is contained in $B(a, \varepsilon)$ since

$$\|\gamma(t) - a\| = t\|a - b\| \leqslant \|a - b\| = d(a, b) < \varepsilon$$

for all $t$. Therefore $b$ lies in the same path-component $P$.

With this observation in place, the theorem follows easily. Indeed, the path-components partition $X$, and so if there was more than one of them we could write $X$ as a disjoint union of non-empty open sets, contrary to the assumption that $X$ is connected. □

THEOREM 7.4.3. *There is a connected subset of $\mathbf{R}^2$ which is not path-connected.*

*Proof.* *There is a classic example, known as the *Topologist's sine-curve*. This is the set $A \subseteq \mathbf{R}^2$ given by

$$\{(0, y) : -1 \leqslant y \leqslant 1\} \cup \{(x, \sin(1/x) : x \in (0, 1]\}.$$

Why is $A$ connected? It is quite easy to convince oneself that $A = \bar{E}$, where $E = \{(x, \sin(1/x) : x \in (0, 1]\}$. However, $E$ is connected, being the image of the connected set $(0, 1]$ under a continuous map, and so the connectedness of $A$ is immediate from Lemma 7.1.5.

Why is $A$ not path-connected? It is "intuitively clear" that there is no path $\gamma : [0, 1] \to A$ with $\gamma(0) = (0, 0)$ and $\gamma(1) = (1, \sin(1))$, but we must prove this. Suppose we have such a path $\gamma$. Write $\ell$ for the vertical line $\{0\} \times [-1, 1]$, thus $A = E \cup \ell$. Since $\ell$ is closed in $A$, $\gamma^{-1}(\ell)$ is closed, and in particular contains its supremum $t$. Thus $\gamma(t) \in \ell$, whilst $\gamma(u) \in E$ for all $u > t$.

Let $p_Y : \mathbf{R}^2 \to \mathbf{R}$ be projection onto the $y$-coordinate, i.e. $p_Y(x, y) = y$. Since $p_Y$ is continuous, so is the composition $p_Y \circ \gamma : [0, 1] \to \mathbf{R}$. Thus there is some $\delta > 0$ such that

$$|p_Y(\gamma(u_1)) - p_Y(\gamma(u_2))| \leqslant 1 \text{ for all } u_1, u_2 \in [t, t + \delta]. \tag{7.2}$$

Now let $p_X$ be projection onto the $x$-coordinate, i.e. $p_X(x, y) = x$. The composition $p_X \circ \gamma$ is continuous, and so by the intermediate value theorem and the fact that $p_X(\gamma(t + \delta)) > 0$, $(p_X \circ \gamma)[t, t + \delta]$ contains some interval $[0, c]$, $c > 0$.

However, as $x$ ranges over $(0, c]$, $\sin(1/x)$ takes all values in $[-1, 1]$ (infinitely often), so there are $u_1, u_2 \in [t, t + \delta]$ such that $p_Y(\gamma(u_1)) = 1$, $p_Y(\gamma(u_2)) = -1$. This contradicts (7.2). □

CHAPTER 8

# Sequential compactness

In this chapter (and in Chapter 8) we will be talking a lot about sequences and subsequences, so let us be clear about what these concepts are. If $X$ is some space, let $\sigma = (x_n)_{n=1}^{\infty} = (x_1, x_2, \dots)$ be a sequence of elements of $X$. Any sequence of the form $\sigma' = (x_{n_k})_{k=1}^{\infty}$, where $n_1 < n_2 < n_3 < \dots$, is called a subsequence of $\sigma$. For instance, $(x_1, x_4, x_9, x_{16}, \dots)$ is a subsequence of $(x_1, x_2, x_3, x_4, \dots)$.

## 8.1. Definitions

In this chapter we study metric spaces which satisfy the metric-space analogue of the Bolzano-Weierstrass property. Recall what the Bolzano-Weierstrass property of $\mathbf{R}$ is: any bounded sequence has a convergent subsequence. More precisely, if $(x_n)_{n=1}^{\infty}$ is a sequence of elements in some closed bounded interval $[a, b]$, there is a subsequence of the $x_n$ which converges to some $c \in [a, b]$.

There is an obvious way to generalise this notion to subsets of metric spaces, and the resulting notion is called sequential compactness.

Definition 8.1.1 (Sequential compactness). Let $X$ be a metric space. Then $X$ is said to be *sequentially compact* if any sequence of elements in $X$ has a convergent subsequence.

*Important remark.* Sometimes, you will see the notion of sequential compactness called simply "compactness". Indeed, the schedules for the course seem slightly confused on this point. To me, compactness is defined in terms of open covers, as in Chapter 9. It is then a nontrivial theorem that the notions of sequential compactness and compactness are the same. In my view it is very important to make the distinction, since both notions have obvious extensions to the context of topological spaces, but in this generality they are *not* the same and in fact neither notion implies the other. Examples showing this are beyond the scope of this course, but if you are interested, see here for discussion and further references.

Example 8.1.2. The closed interval $[0, 1]$ is sequentially compact, by the Bolzano-Weierstrass theorem.

The open interval $(0, 1)$ is not sequentially compact. For instance, the sequence $x_n = 1/n$ has no convergent subsequence in this space.

The set of rational numbers in $[0,1]$ is not sequentially compact – for instance, the sequence $0.1, 0.14, 0.141, 0.1415, \ldots$ consisting of decimal approximations to $\pi - 3$ has no convergent subsequence.

Finally, the real line $\mathbf{R}$ is not sequentially compact. For instance, the sequence $x_n = n$ has no convergent subsequence in this space.

## 8.2. Closure and boundedness properties

In this section we prove a couple of basic lemmas about sequentially compact spaces.

LEMMA 8.2.1. *A sequentially compact subspace of a metric space is closed and bounded.*

*Proof.* Let $X$ be the space and $Y$ the sequentially compact subspace.

Suppose first that $Y$ is not closed. Then $\bar{Y} \setminus Y$ is nonempty. Let $a$ be a point in this set. By Lemma 5.1.5, there is a sequence $(y_n)_{n=1}^{\infty}$ of elements of $Y$ with $\lim_{n\to\infty} y_n = a$. Then any subsequence of $(y_n)_{n=1}^{\infty}$ converges to $a$ and hence, by the uniqueness of limits, does not converge to an element of $Y$. Therefore $Y$ cannot be sequentially compact.

Suppose next that $Y$ is not bounded. Pick an arbitrary point $y_0 \in Y$, and pick a sequence $(y_n)_{n=1}^{\infty}$ such that $d(y_0, y_n) \geqslant n$ for all $n$. Suppose there is a subsequence $(y_{n_k})_{k=1}^{\infty}$ converging to $b$. Then for $k$ sufficiently large we have $d(y_{n_k}, b) < 1$, which implies that

$$d(y_0, b) \geqslant d(y_0, y_{n_k}) - d(y_{n_k}, b) \geqslant n_k - 1.$$

Since $n_k \to \infty$ as $k \to \infty$, whilst $d(y_0, b)$ is a fixed finite quantity, this is a contradiction. □

The converse is not true – for instance, take $X = Y = (0,1)$ (noting that $Y$ is closed *as a subset of* $X$).

LEMMA 8.2.2. *A closed subset of a sequentially compact metric space is sequentially compact.*

*Proof.* Let $X$ be the space and $Y$ the closed subspace. Consider a sequence $(y_n)_{n=1}^{\infty}$ of elements of $Y$. It is also a sequence of elements of $X$ and so, by sequential compactness of $X$, has a subsequence converging to $a$. However, $Y$ is closed, so the limit of any convergent sequence of elements of $Y$ lies in $Y$. In particular, $a \in Y$. □

The following is perhaps not quite so basic (though it is not hard). We include it because it will be needed later on, in the complex analysis part of the course.

LEMMA 8.2.3. *Let $X$ be a metric space. Suppose that $K$ is a sequentially compact subset of $X$, and that $U$ is an open subset of $X$ containing $K$. Then there is some $\varepsilon > 0$ such that the "$\varepsilon$-thickening" $\bigcup_{z\in K} B(z,\varepsilon)$ of $K$ is contained in $U$.*

*Proof.* Suppose this is not true for $\varepsilon = 1/n$. Then there is $x_n \in K$ such that $B(x_n, 1/n)$ is not contained in $U$, so there is $y_n \in U^c$ with $d(x_n, y_n) < 1/n$. Since $K$ is sequentially compact, there is a subsequence $(x_{n_k})_{k=1}^{\infty}$ which converges to some point $p \in K$.

But then it follows $(y_{n_k})$ also converges to $p$. Since $U^c$ is closed and all the $y_{n_k}$ lie in $U^c$, it follows that $p \in U^c$. But this is a contradiction, since $K \subseteq U$. □

## 8.3. Continuous functions on sequentially compact spaces

Sequential compactness has some nice properties with respect to continuous maps.

LEMMA 8.3.1. *The image of a sequentially compact metric space under a continuous map is sequentially compact.*

*Proof.* Let $X$ be sequentially compact, and suppose that $f : X \to Y$ is continuous. Let $\sigma = (f(x_n))_{n=1}^{\infty}$ be a sequence of elements of $f(X)$. The sequence $(x_n)$ contains a convergent subsequence $(x_{n_k})$ say, with $x_{n_k} \to a$ as $k \to \infty$ for some $a \in X$. But then, since $f$ is continuous, we have $f(x_{n_k}) \to f(a)$, and so $\sigma' = (f(x_{n_k}))_{k=1}^{\infty}$ is a convergent subsequence of $\sigma$. □

As a consequence of Lemma 8.2.1, we see that continuous function $f$ from a sequentially compact metric space $X$ to $\mathbf{R}$ has closed and bounded image, so in particular $f$ is bounded and attains its bounds.

Another consequence of Lemma 8.3.1 is the if $X$ and $Y$ are homeomorphic metric spaces and if $X$ is sequentially compact, then so is $Y$.

PROPOSITION 8.3.2. *A continuous function from a sequentially compact metric space to $\mathbf{R}$ is uniformly continuous.*

*Proof.* Let $X$ be a sequentially compact metric space, and suppose that $f : X \to \mathbf{R}$ is continuous but not uniformly continuous. Then there exists some $\epsilon > 0$ such that for each $n \in \mathbf{N}$ we may find $a_n, b_n \in X$ such that $d(a_n, b_n) < 1/n$ but $d(f(a_n), f(b_n)) \geqslant \varepsilon$. Since $X$ is sequentially compact, $(a_n)_{n=1}^{\infty}$ has a subsequence, $(a_{n_k})_{k=1}^{\infty}$ converging to some point $\ell$. Consider the corresponding sequence $(b_{n_k})_{k=1}^{\infty}$. Since $d(a_{n_k}, b_{n_k}) \leqslant 1/n_k \to 0$, it follows that $b_{n_k}$ also converges to $\ell$ as $k \to \infty$.

Relabelling (to avoid double subscripts) we may now assume we have sequences $(a_n)_{n=1}^{\infty}$, $(b_n)_{n=1}^{\infty}$ with $\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = \ell$ and $d(f(a_n), f(b_n)) \geqslant \varepsilon$ for all $n$.

Now $f$ is continuous at $\ell$, so there is a $\delta > 0$ such that for all $x \in X$ with $d(\ell, x) < \delta$, we have $d(f(\ell), f(x)) < \varepsilon/2$. If $n$ is sufficiently large, we have $d(\ell, a_n), d(\ell, b_n) < \delta$ and hence

$$\varepsilon \leqslant d(f(a_n), f(b_n)) \leqslant d(f(a_n), f(\ell)) + d(f(\ell), f(b_n)) < \epsilon/2 + \epsilon/2 < \epsilon,$$

which is a contradiction.

We were therefore wrong to assume that $f$ is not uniformly continuous. □

## 8.4. Product spaces

Recall that if $(X, d_X)$ and $(Y, d_Y)$ are metric spaces then their Cartesian product $X \times Y$ can be equipped with a metric $d_{X\times Y}$ by setting

$$d_{X\times Y}((x_1, y_1), (x_2, y_2)) = \sqrt{d_X(x_1, x_2)^2 + d_Y(y_1, y_2)^2}.$$

The main result of this section, Proposition 8.4.2 below, is that the product of two sequentially compact spaces is compact. Before proving this, we note an important lemma.

LEMMA 8.4.1. *Let $X$ and $Y$ be metric spaces. A sequence $((x_n, y_n))_{n=1}^{\infty}$ in $X \times Y$ converges if and only if $(x_n)_{n=1}^{\infty}$ converges in $X$ and $(y_n)_{n=1}^{\infty}$ converges in $Y$.*

*Proof.* The projection maps $p_X \colon X \times Y \to X$ and $p_Y \colon X \times Y \to Y$ are continuous. In fact it is easy to see that they are Lipschitz continuous with Lipschitz constant 1. It follows that if $\lim_{n\to\infty}(x_n, y_n) = (a, b)$ then

$$\lim_{n\to\infty} x_n = \lim_{n\to\infty} p_X(x_n, y_n) = p_X(a, b) = a,$$

and similarly $\lim_{n\to\infty} y_n = b$.

Conversely, if $x_n \to a$ and $y_n \to b$ then

$$d_{X\times Y}((x_n, y_n), (a, b)) = \sqrt{d_X(x_n, a)^2 + d_Y(y_n, b)^2} \to 0$$

as $n \to \infty$ and so $(x_n, y_n) \to (a, b)$ as $n \to \infty$, as required. □

Now we prove that the product of two sequentially compact spaces is compact, with apologies for using a rather unpleasant triple subscript notation in the argument.

PROPOSITION 8.4.2. *The product of two sequentially compact metric spaces is sequentially compact.*

*Proof.* Let $((x_n, y_n))_{n=1}^{\infty}$ be a sequence in $X \times Y$. As $X$ is sequentially compact, the sequence $\sigma = (x_n)_{n=1}^{\infty}$ in $X$ has a convergent subsequence $\sigma' = (x_{n_k})_{k=1}^{\infty}$, with $x_{n_k} \to a$ as $k \to \infty$. Now consider the sequence $(y_{n_k})_{k=1}^{\infty}$ in $Y$. Since $Y$

is sequentially compact this in turn has a convergent subsequence $(y_{n_{k_r}})_{r=1}^{\infty}$, say $y_{n_{k_r}} \to b$ as $r \to \infty$. Let $\sigma''$ be the corresponding subsequence of $x$s, that is to say $\sigma'' = (x_{n_{k_r}})_{r=1}^{\infty}$. Then $\sigma''$ is a subsequence of $\sigma'$, and so it converges to $a$.

By the previous Lemma it follows that $(x_{n_{k_r}}, y_{n_{k_r}}) \to (a, b)$ as $r \to \infty$, and so we have exhibited a convergent subsequence of $((x_n, y_n))_{n=1}^{\infty}$ . Therefore $X \times Y$ is sequentially compact. □

A corollary of this is the following result, which is often called the Bolzano-Weierstrass theorem (being a generalisation of the version on $\mathbf{R}$).

COROLLARY 8.4.3 (Bolzano-Weierstrass). *Any closed and bounded subset of* $\mathbf{R}^n$ *is sequentially compact.*

*Proof.* Let $X \subseteq \mathbf{R}^n$ be the set. Since $X$ is bounded, it is contained in some cube $[-M, M]^n$. The Bolzano-Weierstrass theorem on $\mathbf{R}$ implies that $[-M, M]$ is sequentially compact, and therefore by Proposition 8.4.2, $[-M, M]^n$ is sequentially compact. Since $X$ is closed, it is sequentially compact by Lemma 8.2.2. □

## 8.5. Sequentially compact equals complete and totally bounded

As a warm-up to the main business of this section, we prove the following.

PROPOSITION 8.5.1. *A sequentially compact metric space is complete and bounded. The converse is not true in general.*

*Proof.* Suppose that $X$ is sequentially compact. We have already shown that $X$ is bounded in Lemma 8.2.1. Let us now show that $X$ is complete. Suppose that $(x_n)_{n=1}^{\infty}$ is a Cauchy sequence in $X$. Since $X$ is sequentially compact, $(x_n)_{n=1}^{\infty}$ has a convergent subsequence $(x_{n_k})_{k=1}^{\infty}$. Suppose that $\lim_{k\to\infty} x_{n_k} = a$. We claim that in fact $\lim_{n\to\infty} x_n = a$.

Let $\epsilon > 0$. Then, since $(x_n)_{n=1}^{\infty}$ is Cauchy, there is some $N$ such that for all $n, m \geqslant N$ we have $d(x_n, x_m) < \epsilon/2$. Since $\lim_{k\to\infty} x_{n_k} = a$, we may find a $k$ such that $n_k \geqslant N$ and $d(x_{n_k}, a) < \epsilon/2$. But then if $n \geqslant N$ we have

$$d(x_n, a) \leqslant d(x_n, x_{n_K}) + d(x_{n_K}, a) < \epsilon/2 + \epsilon/2 = \epsilon,$$

as required.

To show that the converse is not true in general, consider the following example. Take $C_b(\mathbf{R})$ to be the normed space of continuous bounded functions on the real line equipped as usual with the $\|\cdot\|_\infty$-norm and the associated metric. Define a function $\phi : \mathbf{R} \to \mathbf{R}$ by

$$\phi(t) = \begin{cases} 2t + 1, & -1/2 \leqslant t \leqslant 0; \\ 1 - 2t, & 0 \leqslant t \leqslant 1/2 \end{cases}$$

and $\phi(t) = 0$ for $t \notin [-1/2, 1/2]$. For each $n \in \mathbf{N}$ set $f_n(t) = \phi(t+n)$ (we might call this sequence of functions a "moving bump"). All of the functions $f_n$ lie in $\bar{B}(0,1)$ (that is, have sup norm bounded by 1). However, if $n \neq m$ then $f_n(n) = 1$, whilst $f_m(n) = 0$, so $\|f_n - f_m\|_\infty = 1$. Thus the sequence $(f_n)_{n=1}^\infty$ has no Cauchy subsequence, and hence certainly no convergent subsequence. □

*Remark.* The Bolzano-Weierstass theorem (Corollary 8.4.3) asserts that the converse *is* true for subsets of $\mathbf{R}^n$.

It turns out that there is a stronger notion of boundedness called *total boundedness* which – together with completeness - implies sequential compactness and in fact is equivalent to it.

DEFINITION 8.5.2. A metric space is said to be *totally bounded* if, for any $\varepsilon > 0$, it may be covered by finitely many open balls of radius $\varepsilon$.

Here is one of the more substantial theorems of the course.

THEOREM 8.5.3. *A metric space is sequentially compact if and only if it is complete and totally bounded.*

*Proof.* Suppose first that we have a space $X$ which is sequentially compact. We have already shown in Proposition 8.5.1 that $X$ is complete. Let us now show that it is totally bounded. Suppose $X$ is *not* totally bounded, and let $\varepsilon$ be such that there is no way to cover $X$ by finitely many open balls of radius $\varepsilon$.

Using a greedy algorithm, we select an infinite sequence $(x_n)_{n=1}^\infty$ of elements of $X$ which are separated by at least $\varepsilon$, that is to say $d(x_i, x_j) \geqslant \varepsilon$ whenever $i \neq j$.

To do this, suppose that $x_1, \ldots, x_n$ have already been selected. By assumption, the balls $B(x_i, \varepsilon)$ do not cover $X$, and so we may select a point $x_{n+1} \in X$ which does not lie in any of these balls, and therefore $d(x_i, x_{n+1}) \geqslant \varepsilon$ for $i = 1, \ldots, n$.

It is clear that such a sequence has no convergent subsequence, and so we were wrong to assume that $X$ is not totally bounded.

We turn now to the more substantial direction of the theorem, which is to show that a complete and totally bounded metric space $X$ is sequentially compact. Let $\sigma$ be a sequence of elements of $X$. We will use the total boundedness assumption for balls of radii $1, \frac{1}{2}, \frac{1}{4}, \ldots$. Thus, for each nonnegative integer $m$ there is a finite collection of open balls $B_1^{(m)}, \ldots, B_{k_m}^{(m)}$ of radius $2^{-m}$ which cover $X$.

Start with the balls $B_1^{(0)}, \cdots, B_{k_0}^{(0)}$ of radius 1. One of these balls contains infinitely many elements of the sequence $\sigma$. Write $B_0$ for the ball with this property, and let $\sigma^{(0)}$ be the infinite subsequence of $\sigma$ of elements contained in this ball.

Now look at the balls $B_1^{(1)}, \ldots, B_{k_1}^{(1)}$ of radius $\frac{1}{2}$. One of *these* balls contains infinitely many elements of the new subsequence $\sigma^{(0)}$. Write $B_1$ for such a ball, and let $\sigma^{(1)}$ be the finite subsequence of $\sigma^{(0)}$ of elements contained in it.

Continue in the obvious fashion, producing new subsequences $\sigma^{(2)}, \sigma^{(3)}, \ldots$ with $\sigma^{(r)}$ contained in $B_r$ and a subsequence of $\sigma^{(r-1)}$.

Now consider the sequence $\sigma^*$ obtained by a diagonal argument: the $i$th element of $\sigma^*$ is taken to be the $i$th element of $\sigma^{(i)}$. Clearly $\sigma^*$ is a subsequence of $\sigma$ and, if we write $\sigma^* = (x_n)_{n=1}^{\infty}$, we have $x_n \in B_r$ for all $n \geqslant r$.

It is now clear that $\sigma^*$ is a Cauchy sequence. Indeed, given $\varepsilon > 0$, let $N$ be such that $2^{-N} < \varepsilon/2$. If $n, m \geqslant N$ then $x_n, x_m$ both lie in $B_N$, which is a ball of radius $2^{-N}$, and hence $d(x_n, x_m) < \varepsilon/2 + \varepsilon/2 = \varepsilon$.

Finally, since $X$ is complete the sequence $\sigma^*$ converges. We have shown that $\sigma$, which was an arbitrary sequence in $X$, has a convergent subsequence, and therefore $X$ is sequentially compact. □

*Remark.* Observe that the argument in fact shows that any sequence in a totally bounded metric space has a subsequence which is Cauchy. We only used completeness right at the end.

## 8.6. The Arzelà-Ascoli theorem

Let $X$ be a sequentially compact metric space. We have shown (Lemma 8.3.1) that any continuous function $f : X \to \mathbf{R}$ is bounded, and so the space $C_b(X)$ of bounded, real-valued continuous functions on $X$ is equal to $C(X)$, the space of continuous real-valued functions on $X$. We have seen that this is a normed space, with the sup norm $\|f\|_\infty := \sup_{x \in X} |f(x)|$, and moreover it is complete (Theorem 6.3.2).

The space $C(X)$ is never sequentially compact itself. To see this, consider the sequence $(f_n)_{n=1}^{\infty}$ in which $f_n$ is the constant function $n$; this sequence clearly has no convergent subsequence.

There are other, less trivial, examples, for instance when $X = [0, 1]$. Consider the sequence $(f_n)_{n=1}^{\infty}$ of continuous functions on $[0, 1]$ defined as follows: $f_n(x)$ is zero outside of the interval $(\frac{1}{n+1}, \frac{1}{n})$, but takes the value 1 at the midpoint $t_n := \frac{1}{2}(\frac{1}{n} + \frac{1}{n+1})$ of this interval, and is piecewise linear elsewhere. The $f_n$ are all continuous, but clearly $d(f_m, f_n) = 1$ whenever $m \neq n$, since $f_m(t_m) = 1$ whilst $f_n(t_m) = 0$. Thus this sequence has no convergent subsequence.

The issue here is that the functions $f_n$, whilst continuous, become "less and less continuous" as $n \to \infty$; the gradient of the piecewise linear sequences tends to infinity.

Whilst the whole space $C(X)$ is not sequentially compact, interesting subsets of it may be. Roughly speaking, the two types of example we have just mentioned are the only obstruction to sequential compactness, an idea made precise by the Arzelà-Ascoli theorem.

The property of uniform boundedness rules out trivial examples like the sequence $f_n = n$.

DEFINITION 8.6.1 (Uniformly bounded). Let $X$ be a sequentially compact metric space. Let $\mathcal{F} \subseteq C(X)$. Then we say that $\mathcal{F}$ is *uniformly bounded* if it is bounded as a subset of $C(X)$ with the $\|\cdot\|_\infty$ norm. That is, there is some $M$ such that $|f(x)| \leqslant M$ for all $x \in X$ and for all $f \in \mathcal{F}$.

The property of equicontinuity rules out "less and less continuous examples" like the one we described.

DEFINITION 8.6.2 (Equicontinuity). Let $X$ be a sequentially compact metric space. Let $\mathcal{F} \subseteq C(X)$. Suppose that, in the $\varepsilon$-$\delta$ definition of continuity, $\delta$ can be chosen independently of $f \in \mathcal{F}$. That is, for every $\varepsilon > 0$ there is $\delta > 0$ such that whenever $d(x, y) < \delta$ we have $d(f(x), f(y)) < \varepsilon$ for all $f \in \mathcal{F}$. Then we say that the family $\mathcal{F}$ is *equicontinuous*.

THEOREM 8.6.3 (Arzelà-Ascoli). *Let $X$ be a sequentially compact metric space. Let $\mathcal{F} \subseteq C(X)$ be an equicontinuous and uniformly bounded set of functions. Then any sequence of elements of $\mathcal{F}$ has a convergent subsequence. In particular, if $\mathcal{F}$ is closed then it is sequentially compact.*

*Proof.* (Non-examinable). $\bar{\mathcal{F}}$, being a closed subset of the complete metric space $C(X)$, is complete. Therefore, by Theorem 8.5.3, it is enough to show that $\bar{\mathcal{F}}$ is totally bounded.

*Claim.* It is enough to show that $\mathcal{F}$ is totally bounded. *Proof.* Let $\varepsilon > 0$, and suppose there is some collection of balls $B(f_i, \varepsilon/2)$, which cover $\mathcal{F}$. If $g \in \overline{\mathcal{F}}$, there is some $f \in \mathcal{F}$ with $d(f, g) < \varepsilon/2$. Suppose that $f \in B(f_i, \varepsilon/2)$. Then $d(g, f_i) \leqslant d(g, f) + d(f, f_i) < \varepsilon$. Therefore the balls $B(f_i, \varepsilon)$ cover $\bar{\mathcal{F}}$. (Note that the same argument works in any metric space.) This proves the claim.

It remains to show that $\mathcal{F}$ is totally bounded. Let $\varepsilon > 0$. Since $\mathcal{F}$ is uniformly bounded, there is some $M$ such that $|f(x)| \leqslant M$ for all $x \in X$ and $f \in \mathcal{F}$.

Since $\mathcal{F}$ is equicontinuous we know that there is a $\delta > 0$ such that if $x, y \in X$ are such that $d(x, y) < \delta$ then $|f(x) - f(y)| < \epsilon/4$.

Since $X$ is sequentially compact, it is totally bounded, so there is some finite collection of balls $B(x_i, \delta)$, $i = 1, 2, \ldots, k$, which covers $X$.

Divide $[-M, M]$ into $K$ intervals, all of length less than $\varepsilon/4$, and label these intervals $I_1, \ldots, I_K$. For each function $\alpha : \{1, \ldots, k\} \to \{1, \ldots, K\}$, there may or may not be a function $f \in \mathcal{F}$ such that $f(x_i) \in I_{\alpha(i)}$ for $i = 1, \ldots, k$. If there is, pick one and call it $f_\alpha$; otherwise, choose $f_\alpha$ arbitrarily.

We claim that the balls $B(f_\alpha, \varepsilon)$ cover $\mathcal{F}$. Since there are only finitely many (in fact, $K^k$) functions $\alpha$, this establishes the total boundedness of $\mathcal{F}$.

It remains to prove this claim. Let $f \in \mathcal{F}$ be arbitrary, and let $\alpha : \{1, \dots, k\} \to \{1, \dots, K\}$ be the function such that $f(x_i) \in I_{\alpha(i)}$ for all $i$. Consider the function $f_\alpha$, which has the same property by definition: $f_\alpha(x_i) \in I_{\alpha(i)}$ for all $i$. In particular,

$$|f(x_i) - f_\alpha(x_i)| < \varepsilon/2 \qquad \text{for all } i. \tag{8.1}$$

Now let $x \in X$ be arbitrary. By the choice of the $x_i$, there is some $i$ such that $d(x, x_i) < \delta$. From the definition of $\delta$, it follows that

$$|f(x) - f(x_i)| < \varepsilon/4, \tag{8.2}$$

and also that

$$|f_\alpha(x) - f_\alpha(x_i)| < \varepsilon/4 \tag{8.3}$$

Combining (8.1), (8.2) and (8.3) using the triangle inequality gives

$$|f(x) - f_\alpha(x)| < \varepsilon.$$

Since $x$ was arbitrary, it follows that $\|f - f_\alpha\|_\infty < \varepsilon$, or in other words that $f \in B(f_\alpha, \varepsilon)$. This confirms the claim, and completes the proof of the Arzelà-Ascoli theorem.

□

CHAPTER 9

# Compactness

## 9.1. Open covers and the definition of compactness

In this final chapter of the metric spaces part of the course, we come to one of the most powerful and important notions in all of mathematics: compactness.

Let us start by giving the definition.

DEFINITION 9.1.1. Let $X$ be a metric space and $\mathcal{U} = \{U_i : i \in I\}$ a collection of open subsets of $X$. We say that $\mathcal{U}$ is an *open cover* of $X$ if $X = \bigcup_{i \in I} U_i$. If $J \subseteq I$ is a subset such that $X = \bigcup_{i \in J} U_i$ then we say that $\{U_i : i \in J\}$ is a *subcover* of $\mathcal{U}$ and if $|J| < \infty$ then we say that it is a *finite subcover*.

DEFINITION 9.1.2 (Compactness). A metric space is said to be compact if every open cover has a finite subcover.

EXAMPLE 9.1.3. The real line $\mathbf{R}$ is not compact. For instance, the open cover $\bigcup_{n \in \mathbf{N}} (-n, n)$ has no finite subcover.

*Motivation.* It is quite hard to motivate the definition of compactness when one first sees it. Indeed, von Neumann's famous quote "... *in mathematics you don't understand things. You just get used to them*" is quite apposite. Nonetheless, a couple of comments are in order. First of all, it turns out that compactness and sequential compactness are the same concept in metric spaces. We prove this in Sections 9.2 and 9.4 below (with the second of these being non-examinable). Second, the notion of compactness looks rather natural in the context of topological spaces, since it talks about open sets in a very basic way. Whilst the notion of sequential compactness can also be formulated in topological spaces, it is somehow less basic and, in this more general situation, *not* equivalent to compactness. We already remarked on this point and suggested further reading in the last chapter.

*Subspaces.* Sometimes, we will have a metric space $X$ and a subspace $Y \subseteq X$, and we wish to talk about whether $Y$ is compact. In this context, by convention an open cover $\mathcal{U}$ of $Y$ is a collection $\{U_i : i \in I\}$ of open subsets *of* $X$, such that $Y \subseteq \bigcup_{i \in I} U_i$. A subcollection $\{U_i : i \in J\}$ is called a subcover if $Y \subseteq \bigcup_{i \in J} U_i$.

Then $Y$ is compact if and only if every open cover has a finite subcover. The reason this notion is the same as the previous one (which was "internal to $Y$",

making no reference to open sets in $X$) is Lemma 4.5.1, which says that open sets in $Y$ are the same thing as open sets in $X$ intersected with $Y$.

It should be said that this abuse of nomenclature of using the phrase "open cover" in two slightly different ways can be a touch confusing when you first see it. I recall being confused about this point myself when I was an undergraduate.

## 9.2. Compactness implies sequential compactness

PROPOSITION 9.2.1. *A compact metric space is sequentially compact.*

We isolate a lemma from the proof.

LEMMA 9.2.2. *Suppose that $X$ is a compact metric space and that we have a nested sequence $S_1 \supseteq S_2 \supseteq S_3 \supseteq \cdots$ of nonempty, closed subsets of $X$. Then the intersection $\bigcap_{n=1}^{\infty} S_n$ is nonempty.*

*Remark.* You might be interested in comparing this with Lemma 6.2.2, where the same conclusion was reached assuming that $X$ is complete and that the diameters of $S_i$ tend to 0.

*Proof.* Suppose the intersection is empty. Then the complements $S_i^c$ (which are open sets) are an open cover of $X$. By compactness, there is a finite subcover. In particular, for some $n$ the sets $S_1^c, \ldots, S_n^c$ cover $X$. However, we have $S_1^c \subseteq S_2^c \subseteq \cdots \subseteq S_n^c$, and therefore $S_n^c$ covers (is equal to) $X$. But this is a contradiction, since $S_n$ is nonempty. □

*Proof.* (Proof of Proposition 9.2.1.) Let $X$ be the space in question, and suppose that $(x_n)_{n=1}^{\infty}$ is a sequence of elements of $X$. We wish to find a convergent subsequence of this sequence.

For each natural number $n$, set $A_n := \{x_n, x_{n+1}, x_{n+2}, \ldots\}$. Obviously, $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$, and so $\bar{A}_1 \supseteq \bar{A}_2 \supseteq \bar{A}_3 \supseteq \cdots$. Applying Lemma 9.2.2, we see that $\bigcap_{n=1}^{\infty} \bar{A}_n$ is nonempty.

Let $a$ be a point in this intersection. We inductively construct a subsequence $(x_{n_k})_{k=1}^{\infty}$ such that $d(x_{n_k}, a) < 1/k$ for all $k$; it is then clear that this subsequence converges (to $a$) and the proof will be complete. Suppose that $n_1, \ldots, n_k$ have already been constructed. Now $a$ lies in $\bar{A}_{n_k+1}$, that is to say the closure of the set $\{x_{n_k+1}, x_{n_k+2}, \ldots\}$. In particular, there is some element of this sequence at distance less than $1/(k+1)$ from $a$, and we can take this to be our $x_{n_{k+1}}$. □

## 9.3. The Heine-Borel theorem

In Section 9.4, we will prove that any sequentially compact metric space is compact. However, that section is not examinable. The special case of a closed interval $[a, b]$ *is* examinable, and this is called the Heine-Borel theorem.

PROPOSITION 9.3.1 (Heine-Borel). *The interval $[a, b]$ is compact.*

*Proof.* Let $\mathcal{U} = \{U_i : i \in I\}$ be an open cover of $[a, b]$ (the $U_i$ are open in $\mathbf{R}$).

Define $S$ to be the set of all $x \in [a, b]$ for which $[a, x]$ is covered by some finite subcollection of the $U_i$.

Certainly $S \neq \emptyset$, since $a \in S$. $S$ is bounded above by $b$. Therefore it has a supremum $c = \sup(S)$, and $c \in [a, b]$. In fact $c > a$: if $a \in U_j$ then $U_j$ contains some interval $[a - \eta, a + \eta]$, $\eta > 0$, so $a + \eta \in S$.

Assume that $c < b$. Since $\mathcal{U}$ is an open cover of $[a, b]$, $c$ lies in some set $U_j$. Since $U_j$ is open, some open interval $[c - \varepsilon, c + \varepsilon]$, $\varepsilon > 0$, is contained in $U_j$. Assume $\varepsilon$ is chosen so small that $c - \varepsilon > a$ and $c + \varepsilon < b$.

Now $c - \varepsilon$ is contained in $S$, or else it would be an upper bound for $S$, smaller than $c$. Therefore $[a, c - \varepsilon]$ is covered by finitely many sets from $\mathcal{U}$. These sets, together with $U_j$, then give a covering of $[a, c + \varepsilon]$ by a finite subcollection of $\mathcal{U}$. This contradicts the fact that $c$ is an upper bound for $S$.

We are forced to conclude that $c = b$. Now if $b \in U_i$ then $U_i$ contains some interval $[b - \kappa, b + \kappa]$, $\kappa > 0$. Since $c = \sup(S) = b$, $b - \kappa \in S$, and so $[a, b - \kappa]$ is covered by a finite subcollection of $\mathcal{U}$. This subcollection, together with $U_i$, gives a finite subcover of $[a, b]$. □

## 9.4. Sequential compactness implies compactness

The converse of Proposition 9.2.1 is also true.

PROPOSITION 9.4.1. *A sequentially compact metric space is compact.*

As a consequence of this, Proposition 9.2.1 and Theorem 8.5.3, we have the following substantial and important theorem.

THEOREM 9.4.2. *Let $X$ be a metric space. Then the following are equivalent:*

(i) *$X$ is compact;*
(ii) *$X$ is sequentially compact;*
(iii) *$X$ is complete and totally bounded.*

We turn now to the proof of Proposition 9.4.1, which is nonexaminable.

*Proof.* (Proof of Proposition 9.4.1, non-examinable.) Let $X$ be a sequentially compact metric space. By (the easy direction of) Proposition 8.5.3, $X$ is complete

and totally bounded. For $m = 1, 2, 3, \ldots$, fix some collection of balls $B_1^{(m)}, \ldots, B_{k_m}^{(m)}$ of radius $2^{-m}$ which cover $X$.

Suppose we have an open cover of $X$ by sets $U_i$, $i \in I$, which has no finite subcover. Then one of the balls $B_j^{(1)}$ is not covered by finitely many of the $U_i$; let us write $B_1$ for this ball.

Now consider the balls $B_j^{(2)}$ which intersect $B_1$. One of these is not covered by finitely many of the $U_i$ (otherwise $B_1$ would be). Write $B_2$ for this ball.

Now consider the balls $B_j^{(3)}$ which intersect $B_2$, and so on.

Continuing in this fashion, we obtain a sequence $B_1, B_2, \ldots$ of open balls, with $B_m$ having radius $2^{-m}$, $B_m \cap B_{m+1} \neq \emptyset$ for all $m$, and with none of the $B_j$ covered by finitely many of the $U_i$. Let $x_m$ be the centre of $B_m$. Then, since $B_m$ and $B_{m+1}$ intersect in some point $t$, we have

$$d(x_m, x_{m+1}) \leqslant d(x_m, t) + d(x_{m+1}, t) < 2^{-m} + 2^{-(m+1)} < 2 \cdot 2^{-m}.$$

By the triangle inequality and summing the geometric series, it follows that for any $n \geqslant m$ we have

$$\begin{aligned} d(x_m, x_n) &\leqslant d(x_m, x_{m+1}) + \cdots + d(x_{n-1}, x_n) \\ &< 2(2^{-m} + 2^{-(m+1)} + \cdots) = 4 \cdot 2^{-m}. \end{aligned}$$

Therefore $(x_n)_{n=1}^{\infty}$ is a Cauchy sequence. Since $X$ is complete, we have $\lim_{n\to\infty} x_n = x$ for some $x \in X$. Since the sets $U_i$ cover $X$, one of them must contain $x$. Let us suppose $U_1$ contains $x$. Then, since $U_1$ is open, some ball $B(x, \varepsilon)$ is contained in $U_1$.

Choose $n$ large enough that $d(x_n, x) < \varepsilon/2$, and also that $2^{-n} < \varepsilon/2$. Recalling that $B_n$ is the ball of radius $2^{-n}$ centred on $x_n$, it follows that $B_n \subseteq B(x, \varepsilon)$. But then $B_n \subseteq U_1$, contrary to the assumption that $B_n$ is not covered by finitely many of the $U_i$.

We were wrong to assume the existence of an open cover of $X$ with no finite subcover, and so $X$ is indeed compact. $\square$