MATH3611/5705 Higher Analysis

Ian Doust

July 30, 2022

Contents

I The entree 6

1 Cardinality 7

1.1 Introduction 7
1.2 TL;DR — the quick version! 8
1.3 1–1, onto and bijective maps 12
1.4 Infinite sets of different sizes 14
1.5 Countable sets 15
1.6 Problems 18

II The main course 20

2 Metric Spaces 21

2.1 Introduction and Motivation 21
2.2 Metric Spaces 22
2.3 Convergence 25
2.4 The topology of metric spaces 27
2.5 Closure and density 32
2.6 Indexed collections, unions and intersections 33
2.7 Continuity 36
2.8 Normed spaces 39
2.9 Boundedness 41
2.10 Completeness 42
2.11 Contraction Mappings 47
2.12 An application to Ordinary Differential Equations 50
2.13 Baire Category Theorem 53
2.14 Appendix: Classical Banach Spaces 56
2.15 Problems 60

3 Sequences and series of functions 68

3.1 Introduction 68
3.2 Types of Convergence 69
3.3 Bounded functions 70

3.4 Series of functions 72
3.5 Differentiating and integrating limits of sequences 73
3.6 Aside: limit superior limit inferior 75
3.7 Differentiating power series 76
3.8 Appendix: L_1 Convergence 78
3.9 Problems 79

4 Topological spaces **82**
4.1 Introduction 82
4.2 Convergence in Topological Spaces 85
4.3 Nets 87
4.4 Bases for Topologies 90
4.5 The weak topology 93
4.6 Pointwise convergence 96
4.7 Comparing topologies 97
4.8 Continuity 98
4.9 Homeomorphisms 100
4.10 Homotopy 101
4.11 Problems 102

5 Compactness and Connectedness **109**
5.1 Introduction 109
5.2 Covers and Compactness 110
5.3 Compactness and the relative topology 115
5.4 The finite intersection property 116
5.5 Compactness in metric spaces 118
5.6 Uniform Continuity 120
5.7 The Weierstrass Approximation Theorem 121
5.8 The Arzelà-Ascoli Theorem 125
5.9 Connectedness 127
5.10 Appendix: The product topology [Not examinable] 131
5.11 Appendix: Commutative C^* algebras 133
5.12 Problems 134

III Dessert **139**

6 L^p spaces — a brief introduction **140**
6.1 Introduction 140
6.2 Simple functions and measure properties 141
6.3 Sets of measure zero 143
6.4 σ-algbras and Lebesgue measure 144
6.5 Lebesgue integration 146

6.6 Convergence Theorems 149
6.7 L^p spaces 150
6.8 The case $p = \infty$ 152
6.9 General measure spaces 153
6.10 Nesting properties 154
6.11 Problems 154

7 Linear functionals and the weak topology **156**
7.1 Introduction 156
7.2 The dual space 156
7.3 Weak convergence 158
7.4 Duality 160
7.5 Aside: some probability theory 162
7.6 Problems 163

8 Linear operators **165**
8.1 Introduction 165
8.2 First thoughts 165
8.3 Continuity 166
8.4 Examples 168
8.5 The inverse problem 169
8.6 Structure Theorems 170
8.7 Adjoints and self-adjointness 172
8.8 Eigenvalues and the spectrum 174
8.9 $B(\mathcal{H})$ as an algebra 175
8.10 Structure Theorems 2 177

Preface

These notes are based on ones prepared when I taught the course MATH3611 Higher Analysis at UNSW at various years during the period 2000 – 2012. This was a roughly 38 lecture course aimed at introducing students to some of the main ideas of abstract analysis: metric and topological spaces, Hilbert and Banach spaces, and a little measure theory. Incoming students had typically done strong courses in multivariable calculus and linear algebra, but were not always very experienced in writing abstract analysis proofs. The course is a core part of the program for Pure Mathematics Honours students and is recommended to many other students working in Applied Mathematics or Statistics.

As lecture notes, these are a little more informal than a textbook. The contents of the lectures would often vary depending on what the students may have been looking at in their assignments. They contain a reasonable number of asides which may or may not have been mentioned in the lectures, often pointing to issues either outside the syllabus, or perhaps things which we would get back to later.

You will see some parts of the notes marked with this Bourbakian 'dangerous bend' symbol. This usually indicates either some point where students often miss some important point, or else it provides some details which go rather beyond the syllabus of this course.

The notes fall into three parts. The **Entree** is just a short introduction to the basics of cardinality. The main thing which will be needed throughout the course is the difference between countable and uncountable infinite sets, and this is covered in Section 1.2. For those who want to go just a little deeper, some more details are given in the rest of Chapter 1. This chapter is a good opportunity to make sure that you are comfortable with abstract proofs.

The **Main Course** looks at metric and topological spaces. The motivation for this area of mathematics is to see to what extent the results of calculus in one or several variables might be entended to more general settings. This includes results such as the Intermediate Value Theorem, and the Max-Min Theorem. The material in this part is covered quite carefully. Students are expected to follow the proofs and be able to produce careful arguments themselves. There are plenty of problems at the ends of the chapters, as well as exercises scattered through the text. It is important that you have a go at a good selection of these problems. Throughout this part of the course we have made an effort to go beyond just straight abstract definitions, theorems and proofs, and to indicate where these ideas might be used in other areas of mathematics such as differential equations, differential geometry, algebraic topology or probability theory, as well as in further work in real, functional or harmonic analysis.

Many students won't take any further courses in analysis, so the final part of the notes (**Dessert**: Chapters 6 to 8) is a rather more informal tour of some important topics that students might need to know about such as Lebesgue integration, weak convergence and duality, and the theory of linear operators. There are more or less no proofs in this section. In order to give students a useful flavour of the ideas, we have often concentrated on describing special cases rather than the most general theory. The notes contain much more than was ever lectured on this part of the course.

Finally, there will no doubt be typos and careless statements scattered through them. If in doubt — look up a proper textbook! (If students wanted a text for the course, we recommended Kolmogorov and Fomin's 'Introductory Real Analysis', referred to occasionally in the notes as K&F. This book has a few idiosyncrasies, but it covers pretty much all the topics in the course, and, importantly, is very cheap!)

Ian Doust

Part I

The entree

Chapter 1

Cardinality

You will all be happy saying that $\{2, 3, 5\}$ is a finite set, and that $\mathbb{R}$ is an infinite one. For most of your mathematical education this degree of distinction was sufficient.

It is not obvious that you can have different 'sizes' of infinite set. Of course there are several ways that you might want to think about one set being bigger than another, but the one that we are interested in here is the theory of 'cardinality' which was developed by Georg Cantor[1] in the latter part of the 1800s.

This material is formally more a part of set theory than analysis. It will however be important to at least understand the distinction between a small, or countable, infinite set and a large, or uncountable, one at several points of this course.

1.1 Introduction

Question 1: Suppose that $f : [0, 1] \to [0, 1]$ is increasing (ie if $x \leq y$ then $f(x) \leq f(y)$). Must f be continuous? If not, what can the set D_f of discontinuities of f look like?

It is easy to see that f could have infinitely many discontinuities. It is more of a challenge to construct examples where $D_f = A = [0, 1] \cap \mathbb{Q}$, or to show that it is impossible to find f such that $D_f = B = [0, 1] \setminus \mathbb{Q}$. What is the difference between A and B?

Later, we'll want to make sense of saying that f must be continuous at 'most' points in $[0, 1]$ (technically, 'almost everywhere').

Question 2: Consider the sequence $x_n = \sin n \in [-1, 1]$, $n = 1, 2, 3, \ldots$. This sequence is dense in $[-1, 1]$ in the sense that if you pick any $a \in [-1, 1]$ and any $\epsilon > 0$ then you can pick n so that $|a - x_n| < \epsilon$.

(a) Can you construct a dense sequence in $\mathbb{R}$. Or $\mathbb{R}^2$?

[1] According to Wikipedia, some Christian theologians saw Cantor's work as a challenge to the uniqueness of the absolute infinity in the nature of God! Kronecker's response to Cantor's work was to describe him as a 'corrupter of youth', so my apologies for what is to follow.

(b) Can you construct a dense sequence of **functions** in $C[0,1]$, the space of continuous functions on $[0,1]$? What about in $\mathcal{B}[0,1]$ the space of bounded functions on $[0,1]$.

The answer to (a) is yes in both cases — see the problems. For (b) you have first to decide how to measure the distance between functions. We'll see that for the standard distance function on these spaces, the answer is yes for $C[0,1]$, but no for $\mathcal{B}[0,1]$.

Related to these facts is that there is a reasonable version of an 'infinite basis' for the vector space $C[0,1]$, but there isn't one for $\mathcal{B}[0,1]$.

To understand what is going on with these questions we need to look at how one can compare infinite sets. It turns out that some infinite sets really are bigger in important ways than others. The technical term for the size of an infinite set is its **cardinality**.

1.2 TL;DR — the quick version!

This section contains the basics of what we will really need for this course, and is probably sufficient for everything that we will do. The remaining sections of the chapter look at cardinality in a little more detail (although without diving too deeply into the set theoretic issues that arise), and are there for those who might be interested in such things.

Definition 1.2.1. A set A is **countably infinite** if you can write it as a list[2],

$$A = a_0, a_1, a_2, \ldots. \tag{1.2.1}$$

You might reasonably wonder whether that definition is formal enough to be used. In writing (1.2.1) we mean that every element of A occurs in the list, and it occurs exactly once. More formally, it means that there is a one-to-one and onto map, that is a **bijection**, $f : \mathbb{N} \to A$ so that $a_n = f(n)$.

Example 1.2.2. Here are some lists, and hence countably infinite sets.

1. $\mathbb{Z}^+ = 1, 2, 3, \ldots.$
2. $\mathbb{Z} = 0, 1, -1, 2, -2, 3, -3, \ldots.$
3. $\mathbb{N} \times \mathbb{N} = (0,0), (1,0), (0,1), (2,0), (1,1), (0,2), (3,0), \ldots.$
4. The set of primes $= 2, 3, 5, 7, \ldots.$

[2]Later we'll usually write $A = \{a_0, a_1, a_2, \ldots\}$ but here I want you to notice that are differences between a set and a list. A list has a definite order to it, and here includes each element just once. For a set, the order is not important, and, as sets, $\{1,1,2,2,3,3,\ldots\} = \{1,2,3,\ldots\}$. It may perhaps be better to write a list as $[a_0, a_1, a_2, \ldots]$, although this does not really get around the technical point that (1.2.1) has objects from different categories on the two sides of the equal sign.

You might try to explicitly write down a formula for the bijections associated to those lists. The first is easy: $f(n) = n + 1$. But as you can see from the last one, sometimes it is much easier to convince people that the set can be written as a list than it is to write a nice formula for the n-th element of the list.

Definition 1.2.3. A set A is **countable** if it is either finite or countably infinite.

People are frequently very sloppy about distinguishing between countable and countably infinite. Usually this doesn't cause any confusion, but the distinction may be important in exam questions!

Proposition 1.2.4. *If A is a set and you can write a list which contains all the elements of A (perhaps repeating some), then A is countable.*

Proof. (Informal) Take the list and rub out any term which has appeared at an earlier point in the list. If this only leaves finitely many terms in the list then the set was finite. Otherwise you have an infinite list in which every element of A occurs exactly once and so A is countably infinite. ■

Example 1.2.5. Consider the following list

$$0, \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \frac{1}{5}, \dots.$$

With a little effort you could write down a formula for the n-th term in this list. It contains every element of $A = \mathbb{Q} \cap [0, 1)$. In fact it contains all the elements (except 0) infinitely many times! By the proposition, A is countable (and so clearly countably infinite).

Sometime you want to write a set as say $B = \{b_i\}_{i \in \mathbb{Z}}$ where you have labeled the elements by a different countable set to $\mathbb{N}$. The next (easy!) result says that if you can label the elements of a set using any countably infinite set, then the set is countably infinite.

Proposition 1.2.6. *If A is countably infinite and $g : A \to B$ is a bijection, then B is countably infinite.*

Proof. By definition, if A is countably infinite then there exists a bijection $f : \mathbb{N} \to A$. Then $g \circ f$ is a bijection from $\mathbb{N}$ to B and hence B is countably infinite. ■

One can use the idea in Example 1.2.2 (3) to prove the following.

Proposition 1.2.7. *Suppose that I is a countable set, and that for all $i \in I$, A_i is a countable set. Then $A = \bigcup_{i \in I} A_i$ is countable. In other words, a countable union of countable sets is countable.*

Proof. (Just for the case where everything is countably infinite!) Since I is countably infinite there is a bijection between I and $\mathbb{N}$. This means that we can just as easily relabel the collection as $\{A_i\}_{i\in\mathbb{N}}$. Having done that, write each $A_i = a_{i,0}, a_{i,1}, a_{i,2}, \dots$ as a list. Then

$$a_{0,0}, a_{1,0}, a_{0,1}, a_{2,0}, a_{1,1}, a_{0,2}, a_{3,0}, \dots$$

is a list which contains every element of A at least once and hence A is countable. ∎

Example 1.2.8. $\mathbb{Q} = \bigcup_{i\in\mathbb{Z}} \mathbb{Q}\cap[i, i+1)$ is countably infinite as, by the previous example, $A_i = \mathbb{Q} \cap [i, i+1)$ is countably infinite.

You might at this point be wondering whether every set is countable! The answer is no!

Example 1.2.9. (Cantor Diagonalization Argument!) Let $\mathcal{B}$ denote the set of infinite bit strings

$$\mathbf{b} = (b_0, b_1, b_2, \dots)$$

where $b_i \in \{0, 1\}$ for all i.

Suppose that you claim to have written $\mathcal{B}$ as a list $B = \mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2, \dots$. We'll show that that was impossible. Start by writing out each one of the bit strings:

$$\begin{aligned}
\mathbf{b}_0 &= b_{0,0}\, b_{0,1}\, b_{0,2}\, b_{0,3}\, \dots \\
\mathbf{b}_1 &= b_{1,0}\, b_{1,1}\, b_{1,2}\, b_{1,3}\, \dots \\
\mathbf{b}_0 &= b_{2,0}\, b_{2,1}\, b_{2,2}\, b_{2,3}\, \dots \\
\vdots\ & \ \vdots
\end{aligned}$$

Now define $\mathbf{b} = (b_0, b_1, b_2, \dots)$ where $b_j = 1 - b_{j,j}$, $j \in \mathbb{N}$. Notice that $\mathbf{b}$ is not equal to $\mathbf{b}_0$ as it differs in at least the 0th position: $b_0 \neq b_{0,0}$. And $\mathbf{b}$ can't be equal to $\mathbf{b}_1$ as it differs in the 1st position. And so on. So you can't possibly have written down $\mathcal{B}$ as a list — you must have left at least one element of $\mathcal{B}$ out. Thus $\mathcal{B}$ is not countable.

Definition 1.2.10. A set which is not countable is said to be **uncountable**.

Note that the set $\mathcal{B}$ (more or less) encodes several important examples. There is an easy bijection from $\mathcal{B}$ to the power set[3] of $\mathbb{N}$:

$$f(b_0, b_1, b_2, \dots) = \{i \in \mathbb{N} : b_i = 1\}.$$

For example $f(1, 0, 1, 0, 0, \dots) = \{0, 2\}$, while the inverse image of the set of primes is $\mathbf{b} = (0, 0, 1, 1, 0, 1, 0, 1, \dots)$.

Slightly less neatly, you could consider any bit string to be the binary expansion of a number in $[0, 1]$:

$$g(b_0, b_1, b_2, \dots) = \sum_{i=0}^{\infty} \frac{b_i}{2^{i+1}}.$$

[3]That is, the set of all subsets.

The small complication here is that the map g is not a bijection since $g(0, 1, 1, 1 \dots) = g(1, 0, 0, 0, \dots)$.

Exercise 1.2.11. Noting that decimal expansions (for some numbers) are not unique, adapt the Cantor Diagonalization Argument to prove that $[0, 1]$ is uncountable.

Useful is the following observation.

Proposition 1.2.12. *If A is a subset of a countable set, then it is countable. If A has an uncountable subset then it is uncountable.*

Thus, for example, $\mathbb{R}$ and $\mathbb{C}$ must be uncountable.

As we shall discuss in the next few sections, two sets A and B are said to have the same cardinality if there is a bijection $f : A \to B$, and in this case we write $|A| = |B|$. So A is countably infinite if $|A| = |\mathbb{N}|$. It is not obvious, but $|\mathcal{B}| = |[0, 1]| = |\mathbb{R}| = |\mathbb{C}|$.

Note that two finite sets have the same number of elements if and only if there is a bijection between the sets! In particular for finite sets you could never have $|A| = |B|$ with A a proper subset of B. As you can see from the examples, it is common to have an infinite set having the same cardinality as one of its proper subsets. Indeed, one can take this as a defining property of being an infinite set (see Definition 1.3.5).

You should think of countably infinite sets as being the smallest infinite sets. We'll write, for example,[4] $|\mathbb{N}| < |\mathbb{R}|$.

Proposition 1.2.13. *Every infinite set contains a countably infinite subset.*

Proof. (Slightly informal) Pick any element $a_0 \in A$, and let $S_0 = \{a_0\}$ and let $A_1 = A \setminus S_0$. As A is infinite, A_1 is nonempty, so choose $a_1 \in A_1$, let $S_1 = \{a_0, a_1\}$ and set $A_2 = A \setminus S_1$. Continue recursively in this way to produce a nested sequence of sets $S_n = \{a_0, a_1, \dots, a_n\}$ with $n \in \mathbb{N}$, and let $S = \cup_n S_n$. Then $S = \{a_0, a_1, a_2, \dots\}$ is a countably infinite subset of A. ■

There are 'bigger' infinite sets than $\mathbb{R}$. In fact there is a whole tower of sets of larger and larger cardinality. Whether there are sets which are intermediate in cardinality between $\mathbb{N}$ and $\mathbb{R}$ is known as the Continuum Hypothesis, something which is now known to be undecidable under the usual axioms of set theory! But for most of the later theory, the important distinction will be between countable and uncountable sets. For example, for Question 1 in the Introduction, the answer is that the set of discontinuities of an increasing function can be any countable set.

So read on if you want to see a bit more about cardinality. And if you don't, have a go at Problem 5 at the end of the chapter.

---

[4] The formal definition is given below!

1.3 1–1, onto and bijective maps

When you count you are actually matching each of the objects being counted with a numbers in a standardized list. That is object 1, that is object 2, and so forth, until every object has a matching number. What matters is not which object is object 1, but rather that there is some way of match all of the objects with, say, the numbers from 1 to 27. This idea of matching the elements of two sets is what is behind the theory of cardinality. We'll begin by giving a more set-theoretic definition of 'matching'.

Definition 1.3.1. Recall that a function $f : A \to B$ from a set A to a set B is

- **1–1** if whenever $x \neq y$, $f(x) \neq f(y)$ (usually written as $f(x) = f(y) \implies x = y$),
- **onto** if for all $b \in B$ there exists $x \in A$ such that $f(x) = b$,
- a **bijection** if it is 1–1 and onto.

So far you have mainly dealt with functions from $\mathbb{R} \to \mathbb{R}$ or $\mathbb{R}^n \to \mathbb{R}^m$. In this course the domains and codomains could be much more general. You need to be happy considering $f : C[0,1] \to M_{2,2}$

$$f(h) = \begin{pmatrix} h(0) & h(1) \\ h(\frac{1}{2}) & h(1) - h(0) \end{pmatrix}.$$

Is this function 1–1, onto, or a bijection?

Basic results from discrete mathematics tell us that if A and B each have finitely many elements, say $|A| = n$ and $|B| = m$, then

- if there exists a 1–1 map $f : A \to B$ then $|A| \leq |B|$;
- if there exists an onto map $f : A \to B$ then $|A| \geq |B|$;
- if there exists a bijection $f : A \to B$ then $|A| = |B|$.

Cantor in the 1880s saw that this gave a way of comparing the sizes of arbitrary sets. One has to tread somewhat carefully here since some of the results are counterintuitive for many people. A significant difficulty is in deciding precisely what a set is. (Is the set of all sets a set?) We are going to skip over this interesting question and assume that you can recognise sets (and functions) when you them.

Definition 1.3.2. Suppose that A and B are sets. We shall say that A **is equivalent to** B, written $A \sim B$, if there exists a bijection $f : A \to B$.

Exercise 1.3.3. Show that

- for all sets A, $A \sim A$,
- $A \sim B \iff B \sim A$,
- if $A \sim B$ and $B \sim C$, then $A \sim C$.

This means that $\sim$ acts as an **equivalence relation** on any set of sets. If A and B are finite then $A \sim B \iff |A| = |B|$, so we want to think of $A \sim B$ as saying that A and B are 'the same size'. Some things happen with infinite sets which can't happen with finite ones.

Example 1.3.4. (i) If A and B are finite and $A \subsetneq B$, then $A \not\sim B$. Consider however

$$
\begin{aligned}
A &= \mathbb{N} = \{0, 1, 2, 3, \dots\}, \\
B &= \mathbb{Z}^+ = \{1, 2, 3, 4, \dots\}.
\end{aligned}
$$

The map $f : A \to B$, $f(n) = n + 1$ is a bijection so $A \sim B$ even though $B \subsetneq A$.

(ii) Let $A = \mathbb{N}$, $C_1 = \{0, 2, 4, \dots\}$, $C_2 = \{1, 3, 5, \dots\}$. Then $A \sim C_1$ (via $f(n) = 2n$), $C_1 \sim C_2$ (via $f(n) = n + 1$) and so $C_1 \cup C_2 \sim C_1 \sim C_2$. Thus you can split an infinite set into two parts, each the same 'size' as the original???

Up until now we have used the terms 'finite set' and 'infinite set' rather intuitively. One possible definition is to say that a nonempty set A is finite if it is empty, or else equivalent to any set $\{1, 2, \dots, n\}$ where $n \in \mathbb{Z}^+$. Otherwise it is infinite. While this works, it does depend on you having constructed the positive integers as a set first. An alternative is to use a definition due to Dedekind which does not involve reference to some special set like $\mathbb{Z}^+$.

Definition 1.3.5. A is an **(Dedekind) infinite set** iff there exists $A_0 \subsetneq A$ with $A \sim A_0$. A set which is not infinite is said to be **finite**.

The downside of this definition is that it is now a bit of a challenge to show that if A is a nonempty finite set, then $A \sim \{1, 2, \dots, n\}$ for some $n \in \mathbb{Z}^+$ and so we can write $|A| = n$. Such a proof would require careful use of the ZFC axioms of set theory. You may assume that this is true in this course!

ZFC stands for Zermelo–Fraenkel with the Axiom of Choice. ZFC is the standard form of axiomatic set theory — if you ask most mathematicians of about the foundational basis of the mathematics that they are doing they will mumble something about ZFC, although very few actually can actually name the ZF axioms, and most don't really consider such foundational issues of a day-to-day basis.

The big issue that we will completely skate over is the question of "what is a set?" Under ZFC you can build up larger and more complicated sets using the constructions allowed by the axioms. For example one typically progresses by first constructing $\mathbb{N}$, then $\mathbb{Z}$, then $\mathbb{Q}$ and then $\mathbb{R}$ and $\mathbb{C}$. You can then construct sets of functions and so forth. But if you try to say something like "Let S be the set of all sets which don't contain themselves as a subset" you run into all sorts of paradoxical problems!

The Axiom of Choice (the 'C') is controversial and many mathematicians would prefer to avoid its use if at all possible. Over the past 50 years we have discovered that there are quite a few natural questions in mathematics which are not decidable within the ZFC axiom system. As we go through the course we will see a few points, such as the Hahn–Banach Theorem or existence of nonmeasurable sets, where the Axiom of Choice enters the story. Despite the surprising consequences of this axiom, most analysts keep it in their toolkit as it makes the overall theory rather neater. As far as I know, no bridges have fallen down due to a reliance on the Axiom of Choice. (Be warned that there are mathematicians with even more extreme views, such as not believing in the existence of the set of integers.)

1.4 Infinite sets of different sizes

Question 1.4.1. If A and B are infinite sets, is $A \sim B$?

If all infinite sets turned out to be equivalent, then this concept of similarity wouldn't really produce much of a theory of sizes for sets. There is, however, a rather clever way of seeing that that this isn't the case.

Let A be a set and let $B = \mathcal{P}(A) = \{U : U \subseteq A\}$ be the power set of A. Suppose that $f : A \to B$ is a bijection. Let $S = \{a \in A : a \notin f(a)\} \subseteq A$. Thus $S \in B$ and so as f is onto, there exists $a_0 \in A$ such that $S = f(a_0)$.

Now if $a_0 \in S = f(a_0)$, then $a_0 \notin f(a_0)$?! If $a_0 \notin S = f(a_0)$ then $a_0 \in f(a_0)$?! Thus no such bijection can exist.

This means that we **never** get $A \sim \mathcal{P}(A)$.

For finite sets, if $|A| = n$ then $|\mathcal{P}(A)| = 2^n$ which is generally rather bigger. This gives you the idea that there might be different sizes of infinite sets.

Definition 1.4.2. Let A and B be sets. We shall say that

1. A and B have the **same cardinality**, written $|A| = |B|$, if $A \sim B$. If A and B don't have the same cardinality we'll write $|A| \neq |B|$.

2. the **cardinality of A is less than or equal to the cardinality of B**, written $|A| \leq |B|$, if there exists an **onto** map $f : B \to A$.

3. If $|A| \leq |B|$ and $|A| \neq |B|$ then we'll write $|A| < |B|$ and say that the **cardinality of A is smaller than the cardinality of B**.

Remark 1.4.3. 1. K&F use the term **power** rather than cardinality. Their definition of $|A| \leq |B|$ is also slightly different (see problem sheet 1).

2. At this stage $|A|$ does not have any meaning by itself. There is however a well-developed theory of cardinal numbers, and their arithmetic which you might want to look into. We might in this course say things like 'the cardinality of X is $|\mathbb{R}|$', but that is really just shorthand for $|X| = |\mathbb{R}|$.

3. You can use all the usual inequality language in this setting. For example, if $|A| \leq |B|$ then you might say that the cardinality of B is greater than or equal to the cardinality of A, and perhaps even write $|B| \geq |A|$.

4. The empty set causes a bit of a problem in the above definitions. By decree, we write $\emptyset \sim \emptyset$ and $|\emptyset| \leq |A|$ for all sets A.

Example 1.4.4. (i) $|\mathbb{Z}| \leq |\mathbb{R}|$ via $f(x) = \begin{cases} x, & x \in \mathbb{Z}, \\ 0, & x \neq \mathbb{Z}. \end{cases}$

(ii) $|\{0, 2, 4, \dots\}| = |\mathbb{N}|$.

(iii) $|\mathbb{N}| \leq |\mathcal{P}(\mathbb{N})|$ via $f(A) = \begin{cases} n, & \text{if } A = \{n\}, \\ 0, & \text{otherwise.} \end{cases}$

Note that $\mathbb{N} \not\sim \mathcal{P}(\mathbb{N})$ so in this case $|\mathbb{N}| < |\mathcal{P}(\mathbb{N})|$.

Exercise 1.4.5. Prove that if $A \subseteq B$, then $|A| \leq |B|$.

The above notation makes the next (very important) result look obvious, but it is important to unpack just what it is saying.

Theorem 1.4.6. ***(Schröder-Bernstein)*** *(or Cantor-Bernstein)*
If $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$.

Decoding the notation, this says that if you can find onto maps in both directions, then you can find a bijection between the two sets. Proving this is not so straightforward. The Schröder-Bernstein is quite useful in practice because it often turns out to be easier to construct two onto maps than to construct an explicit bijection.

Example 1.4.7. Let $A = [0, 1)$ and $B = [0, 1]$. Finding a bijection here is slightly painful (try it!). However $f : A \to B$, $f(x) = \max\{2x, 1\}$ is clearly onto, as is $g : B \to A$, $g(y) = y \pmod 1$. Thus $|A| = |B|$.

The proof of Schröder-Bernstein is not examinable. You should have a think about how you might construct a bijection from two onto maps. Have a look at the proof in K&F and see if you can adapt it to our definitions (they use 1–1 maps rather than onto ones).

A natural question[5] at this stage is to ask whether $\leq$ actually provides a good way of comparing sets.

Question 1.4.8. Given sets A and B, must either $|A| \leq |B|$ or $|B| \leq |A|$?

1.5 Countable sets

We know lots of infinite sets: $\mathbb{N}$, $\mathbb{Q}$, $\mathbb{R}$, $[0, 1]$, ... The next thing to look at is how the cardinalities of these sets are related.

Definition 1.5.1. Let A be a set. We shall say that
(i) A is **countably infinite** if $|A| = |\mathbb{N}|$;
(ii) A is **countable** if A is finite or countably infinite;
(iii) A is **uncountable** if it is not countable.

[5]The truth of this statement (in ZF) actually turns out to be equivalent to the Axiom of Choice!

If A is countably infinite then there exists a bijection $a : \mathbb{N} \to A$. Such a map defines a **sequence**; we usually write a_n rather than $a(n)$. The **list**

$$a_0, a_1, a_2, \ldots$$

contains each element of A exactly once. Conversely, if you can write A as such a list, then it must be countably infinite. In practice, it is often much easier to convince someone that you can write A as a list than to write an explicit formula for a bijection!

Example 1.5.2. (i) $\mathbb{Z} = \{0, 1, -1, 2, -2, \ldots\}$ is countably infinite. In this case you could right down a formula for a suitable bijection, but it is clear that the list $0, 1, -1, 2, -2, 3, -3, \ldots$ includes all of $\mathbb{Z}$ (and nothing else).

(ii) $A = \mathbb{N} \times \mathbb{N} = \{(n, m) : n, m \in \mathbb{N}\}$ is countably infinite as

$$A = \{(0, 0), (1, 0), (0, 1), (2, 0), (1, 1), (0, 2), \ldots\}.$$

(Exercise: find a formula for the bijection.)

(iii) $A = [0, 1] \cap \mathbb{Q}$ is countably infinite as $A = \{0, 1, \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{3}{4}, \ldots\}$. (Exercise: What is the 1000000^{th} element in the list?)

(iv) $A = \mathbb{Q}$ is countably infinite. You could write a list as in (iii) or else note that $\mathbb{N} \subseteq \mathbb{Q}$ and so $|\mathbb{N}| \leq |\mathbb{Q}|$. Consider now $f : \mathbb{N} \times \mathbb{N} \to \mathbb{Q}$,

$$f(n, m) = \begin{cases} \frac{n}{m}, & \text{if } n, m \text{ coprime, } m \neq 0, \\ -\frac{n}{m}, & \text{if } \gcd(n, m) = 2 \text{ and } m \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

This is onto, so $|\mathbb{Q}| \leq |\mathbb{N} \times \mathbb{N}| = |\mathbb{N}|$ by (ii). By Schröder-Bernstein, $|\mathbb{Q}| = |\mathbb{N}|$.

Countably infinite sets are in some sense the smallest infinite sets.

Proposition 1.5.3. *If A is an infinite set then it contains a countably infinite subset $B \subseteq A$.*

Proof. First note that it is easy to show from the definitions that if A is infinite and $A \sim A_0$ then A_0 is also infinite.

Now, if A is infinite there exists a proper subset $A_1 \subsetneq A$ with $A \sim A_1$. Choose $x_1 \in A \setminus A_1$. Since A_1 is infinite, one can repeat this to find $A_2 \subsetneq A_1$ with $A_1 \sim A_2$ and choose $x_2 \in A_1 \setminus A_2$. One can recursively then construct a sequence of distinct elements $x_n \in A_{n-1} \setminus A_n$ which forms a countable subset of A. ■

Not all infinite sets are countable! We saw in Example 1.4.4 that $|\mathbb{N}| < |\mathcal{P}(\mathbb{N}|$ and hence that $\mathcal{P}(\mathbb{N})$ must be uncountable. Perhaps more interesting is that the set of real numbers is uncountable. It is actually easier to start withe the 'smaller' set $[0, 1)$.

Theorem 1.5.4. $[0, 1)$ *is uncountable.*

Proof. (A diagonalization argument) Suppose that $[0,1)$ is countable. The we could write this set as a list $[0,1) = \{x_1, x_2, x_3, \dots\}$. Write the decimal expansion[6] of each element $x_k = 0.d_{k1}d_{k2}d_{k3}\dots$. For $k = 1, 2, \dots$, let

$$c_k = \begin{cases} 1, & \text{if } d_{kk} \neq 1, \\ 2, & \text{if } d_{kk} = 1 \end{cases}$$

and let $x = 0.c_1c_2c_3\dots$. (Certainly this expansion does not have an infinite tail of 9's.) Clearly $x \in [0,1)$. But for each k, $x \neq x_k$ since x and x_k differ in the kth decimal place (and x has a unique allowable decimal expansion). Thus the list above doesn't contain x which is a contradiction. ■

Exercise 1.5.5. Prove that $|\mathbb{R}| = |[0,1)| = |\mathbb{R} \setminus \mathbb{Q}|$.

Countable sets have many special properties not possessed by general infinite sets. The hypothesis of countability appears regularly in theorems in analysis and elsewhere.

A powerful tool in showing that a set is countable is the fact that **'a countable union of countable sets is countable'**.

Theorem 1.5.6. *Suppose that A is a nonempty countable set and that $\{S_a\}_{a\in A}$ is a family of countable sets indexed by A. Then $\mathcal{S} = \cup_{a\in A} S_a$ is also countable.*

Heuristically this is pretty much just the same as showing that $\mathbb{N} \times \mathbb{N}$ is countable, by writing down a list of lists and then reordering them into a single list. A more formal proof is given here.

Proof. [Very abstract!] Since A is countable there exists an onto map $f : \mathbb{N} \to A$. For each $\alpha \in A$, the set S_α is countable and so there exists an onto map $g_\alpha : \mathbb{N} \to S_\alpha$. Define $F : \mathbb{N} \times \mathbb{N} \to \mathcal{S}$ by

$$F(n,m) = g_{f(n)}(m).$$

Then F is onto so $|\mathcal{S}| \leq |\mathbb{N} \times \mathbb{N}| = |\mathbb{N}|$ and hence $\mathcal{S}$ is countable. ■

Question 1.5.7. Is there a set $A \subseteq \mathbb{R}$ such that $|\mathbb{N}| < |A| < |\mathbb{R}|$?

This seemingly innocuous question (basically the **contimuum hypothesis**) turns out to be independent of the usual Zermelo-Fraenkel axioms of set theory!

It is possible to construct sets of larger and larger cardinality; $|\mathbb{R}| < |\mathcal{B}[0,1]|$ for example. For this course however, and indeed usually in analysis, the important distinction is whether an infinite set is countable or uncountable.

[6]... where, for definiteness, we never choose an expansion which ends with an infinite string of 9's.

1.6 Problems

Some of these are easy. Some are hard — and are there to perhaps motivate some of the concepts we'll look at later on to enable us to properly address the problems! Many of these problems would require that you read the more formal definitions in Sections 1.3 to 1.5. If you skipped here after reading Section 1.2, then you might just concentrate on Problems 5.

1. Find an explicit bijection $f : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$.

2. Prove carefully from the definition that if $|A| \leq |B|$ and $|B| \leq |C|$, then $|A| \leq |C|$. (This says that this relation is **transitive** on any set of sets.)

3. Prove that $|A| \leq |B|$ if and only if there exists a 1–1 map $g : A \to B$.

4. Let A be any infinite set and B any countable set. Prove that $|A| = |A \cup B|$.

5. Decide whether the following sets are countable or uncountable. Be clear about whether you are using the definition or a theorem.

   (a) $\mathbb{R} \times \mathbb{Q}$.

   (b) The set of all open intervals of $\mathbb{R}$ with rational endpoints.

   (c) $\mathbb{N}^{\mathbb{N}}$, the set of all sequences of natural numbers.

   (d) The set $\mathcal{M}$ of all matrices of any (finite) size which have integer entries. (Thus $(-9) \in \mathcal{M}$ and $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \in \mathcal{M}$.)

   (e) The set of all $x \in [0, 1]$ whose decimal expansion contains no 7s.

   (f) The set of all ways of writing 1 as a (finite) sum of rationals (such as $\frac{1}{2} + \frac{1}{2}$ or $\frac{1}{3} + \frac{3}{8} + \frac{7}{24}$).

   (g) The set of all polynomials with rational coefficients.

   (h) The set of convergent sequences of rational numbers.

6. A 2-D lattice walk is a (finite or infinite) list of points in $\mathbb{Z}^2$, starting at $(0, 0)$, and with each subsequent point distance 1 from the previous point. For example $(0, 0), (1, 0), (1, 1), (1, 2), (1, 1)$ is a 4 step lattice walk. Are the following sets countable or uncountable?

   (a) The set of all finite length lattice walks.

   (b) The set of all infinite length lattice walks.

   (c) The set of all infinite length lattice walks which are never more than distance 42 from the origin.

7. Suppose that $|A_n| = |\mathbb{R}|$, for $n = 1, 2, 3, \ldots$. Prove that $|\bigcup_{n=1}^{\infty} A_n| = |\mathbb{R}|$.

8. Let $\mathcal{R}[0,1]$ denote the set of all real-valued functions from $[0,1]$ to $\mathbb{R}$ and let $C[0,1]$ denote the set of continuous functions on $[0,1]$.

   (a) Prove that $|C[0,1]| = |\mathbb{R}|$.

   (b) Prove that $|\mathcal{R}[0,1]| > |\mathbb{R}|$.

9. Let S be a countable subset of $[0,1]$. Define an increasing function $f : [0,1] \to [0,1]$ whose set of discontinuities is S.

Part II

The main course

Chapter 2

Metric Spaces

2.1 Introduction and Motivation

If you are trying to numerically find $\sqrt{7}$ one might typically follow some sort of iterative procedure like Newton's method to find a sequence $\{x_k\}_{k=1}^{\infty}$ of real numbers which converge to a root of $f(x) = x^2 - 7$. You are using the fact that if $x_k \to x$ and $f(x_k) \to 0$ then, as f is continuous, $f(x) = 0$. What we will look at in this chapter who you can generalize these (and other) ideas to setting where the solution sought is not a number, but instead may be a vector, a matrix, a function or something else entirely.

Suppose for example, that you are trying to solve the initial value problem

$$\begin{cases} y''(x) + \sin(x)y'(x) - y(x)^2 = 0 \\ y(0) = 0, \qquad y'(0) = 1 \end{cases} \tag{2.1.1}$$

and you can find a sequence of functions y_k so that (2.1.1) is 'nearly satisfied'.

- Can you make sense of 'nearly satisfied'?
- Can you find an actual solution by finding $\lim\limits_{k\to\infty} y_k$ is some sense?

To answer these questions we'll need to be able to generalize our ideas of distance, continuity and convergence.

In second year you saw how to generalize these ideas to functions from $\mathbb{R}^n$ to $\mathbb{R}^m$. The (Euclidean) distance between two points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is given in the d_2^n metric as

$$d_2^n(\mathbf{x}, \mathbf{y}) = \left(\sum_{k=1}^{n} (x_k - y_k)^2 \right)^{1/2}$$

Convergence of a sequence of points $\{\mathbf{x}_k\}_{k=0}^{\infty} \subset \mathbb{R}^n$ can be defined in terms of this distance:

$$\mathbf{x}_k \to \mathbf{x} \iff d_2^n(\mathbf{x}_k, \mathbf{x}) \to 0 \quad \text{as } k \to \infty.$$

Similarly, you can express continuity of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ by saying that such a function is continuous if

$$d_2^m(f(\mathbf{x}_k), f(\mathbf{x})) \to 0 \quad \text{whenever} \quad d_2^n(\mathbf{x}_k, \mathbf{x}) \to 0.$$

Although we shall later generalize things even further[1] much can be done as soon as one has some decent way of measuring the **distance** between two objects.

2.2 Metric Spaces

In management speak, a metric is just some way of measuring something. In mathematics, we apply the term a little more narrowly. For us it will be something that tells us how close or distant two objects are. The objects in question could be numbers, vectors, functions, DNA sequences, facial images, student essays, ...

It turns out that most of the good properties of Euclidean distance in $\mathbb{R}^n$ just depend on a few simple conditions.

Definition 2.2.1. Let X be a nonempty[2] set. A function $d : X \times X \to [0, \infty)$ is a **metric** (or **distance function**) if it satisfies the following conditions[3].

(0) $d(x, y) \geq 0$ for all $x, y \in X$.

(i) $d(x, y) = 0 \iff x = y$.

(ii) (Symmetry) $d(x, y) = d(y, x)$ for all $x, y \in X$.

(iii) (Triangle Inequality) $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

The pair (X, d) is called a **metric space**.

There are **lots** of example of metric spaces, and indeed often lots of different useful metrics on any given set X! Your first task is to start collecting a decent mental library of standard examples with which you can use as test cases. Although in general I tend to imagine a general metric space as though it were a nice subset of $\mathbb{R}^2$ you need to know that metric spaces often look quite different to $(\mathbb{R}^n, d_2^n)$.

Example 2.2.2. Here are some standard metric spaces which we shall use as examples to illustrate things as we go along — learn these!

1. $X = \mathbb{R}$, with $d(x, y) = |x - y|$. This is the 'usual' metric on $\mathbb{R}$. If we don't say otherwise you should assume that $\mathbb{R}$ has this metric. (*Moral:* if you want to check whether some proposed theorem is true in metric spaces, first check whether it works in $\mathbb{R}$.)

[1] when we study topological spaces

[2] I may sometimes forget this — but I shouldn't!

[3] Condition (0) can be omitted as we have insisted the range of d is in $[0, \infty)$.

2. X = any nonempty set. The function $d : X \times X \to \{0, 1\}$

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

is always a metric, called the **discrete metric**[4].

3. $X = \mathbb{R}^n$. Suppose that $p > 0$. Define $d_p : X \times X \to [0, \infty)$ by

$$d_p(x,y) = \left(\sum_{k=1}^{n} |x_k - y_k|^p \right)^{1/p}.$$

Then $(\mathbb{R}^n, d_p)$ is a metric if and only if $p \geq 1$. (The triangle inequality here is called Minkowski's Inequality, which you may have seen in MATH2701.) What goes wrong if $0 < p < 1$?

4. $X = \mathbb{R}^n$ with metric

$$d_\infty(x,y) = \max_{1 \leq k \leq n} |x_k - y_k|$$

5. X = vertices of a weighted graph. If v_1, v_2 are two vertices we define $d : X \times X \to [0, \infty)$ as

$$d(v_1, v_2) = \text{Minimum weighted path from } v_1 \text{ to } v_2$$

then (X, d) is a metric space.

6. $X = C[0, 1]$. For $1 \leq p < \infty$ define $d_p : X \times X \to [0, \infty)$ by

$$d_p(f,g) = \left(\int_0^1 |f(t) - g(t)|^p \, \mathrm{d}t \right)^{1/p}.$$

You saw these metrics (especially the $p = 2$ case) in MATH2111 with L^2 convergence of Fourier series. Uniform convergence uses the metric

$$d_\infty(f,g) = \sup_{t \in [0,1]} |f(t) - g(t)|.$$

7. $X = \ell^\infty :=$ all bounded sequences (in $\mathbb{R}$ or $\mathbb{C}$ or $\ldots$). Define $d_\infty : X \times X \to [0, \infty)$ by

$$d_\infty(x,y) = \sup_{k \in \mathbb{N}} |x_k - y_k|$$

Then (X, d_∞) becomes a metric space.

[4] The discrete metric is in some sense at the opposite extreme in terms of behaviour to the usual distance on $\mathbb{R}$. As it is easy to work with, it is another good test case for any general proof you are trying to work out.

8. $X = \mathrm{M}_n(\mathbb{C})$. There are lots of interesting metrics on X! Writing $A = (a_{ij})_{i,j=1}^n$ we could have

$$\begin{aligned}
d(A,B) &= \sup_{\|x\|=1} \|Ax - Bx\|, \\
d(A,B) &= \sum_{i,j=1}^n |a_{ij} - b_{ij}|, \\
d(A,B) &= \left(\sum_{i,j=1}^n |a_{ij} - b_{ij}|^2 \right)^{1/2}, \quad \text{or} \\
d(A,B) &= \mathrm{tr}((A-B)^*(A-B))^{1/2}.
\end{aligned}$$

It is easier to check some of these than others!

9. Smooth surfaces in $\mathbb{R}^n$ such as spheres and toruses have a concept of geodesic (or shortest) distance between points which makes them into a metric space. This generalizes to the concept of a Riemannian manifold which you may study in the differential geometry course.

10. $X = \{z \in \mathbb{C} : |z| < 1\}$ with

$$d(z,w) = \cosh^{-1}\left(1 + \frac{2|z-w|^2}{(1-|z|)(1-|w|)}\right).$$

This provides something called the Poincaré disk model of hyperbolic geometry. (Checking the triangle inequality looks interesting!)

11. X =squares on a chess board, with

$$d(x,y) = \text{minimum number of knight moves to get from } x \text{ to } y.$$

Note that here you first need to prove that d is well-defined, that is that you can get from any square to any other square. Once you know this the metric space properties are easy. Determining the maximum distance in (X,d) is not so easy!

12. (For students who did MATH2701) Let $X = \mathbb{Q}$ and suppose that p is a prime. Let $|\cdot|_p$ denote the p-adic valuation on $\mathbb{Q}$. Then $d(x,y) = |x-y|_p$ is a metric on $\mathbb{Q}$. This metric is important in number theory.

13. Suppose that (X,d) is a metric space and that Y is a nonempty proper subset of X. The restriction of d to $Y \times Y \subseteq X \times X$ is a metric on Y. Thus we might have

    (a) $Y = [0,1]$ with the usual metric from $\mathbb{R}$, namely $d(x,y) = |x-y|$.

    (b) $Y = \mathbb{Z}$ with the usual metric from $\mathbb{R}$. (It will turn out that in terms of convergence and continuity etc, this is pretty much the same as using the discrete metric. Can you see why?)

(c) $Y = \{0,1\}^n$, or bit-strings of length n. The standard metric used in coding theory is the **Hamming metric**, $d((x_1,\dots,x_n),(y_1,\dots,y_n))$ is the number of places in which the entries differ. This is really just the restriction of the d_1 metric on $\mathbb{R}^n$ considered above.

(d) Important subsets of the space (ℓ^∞, d_∞) are the spaces c consisting of all convergent sequences, and c_0 which are all the sequences which converge to 0.

Example 2.2.3. It is perhaps also worth mentioning some 'nonexamples'. You should think about what is wrong in these examples.

1. Let X be the set of all integrable functions on $[0,1]$ with

$$d(f,g) = \int_0^1 |f(t) - g(t)|\,dt$$

2. Let X be the set of all continuous functions $f : \mathbb{R} \to \mathbb{R}$ with

$$d(f,g) = \sup_{t\in\mathbb{R}} |f(t) - g(t)|.$$

3. $X = \mathbb{R}$ with $d(x,y) = (x-y)^2$.

4. In many real-life situations, the problem is a lack of symmetry. For example, if one defines $d(a,b)$ to be the minimum time to travel from a to b, this may be different to $d(b,a)$.

Many of the example above have additional structure apart from their metric. In particular, many important examples are vector spaces (or subsets of vector spaces). Later on we will pay special attention to metrics which come from norms, which are a way of measuring the size of an element of a vector space.

In elementary calculus f is a function and x is a number. Once you move to abstract settings like metric spaces, you'll need to be much more flexible in your notation. In different examples, f (or even x) might be used for an element of a metric space, or a function defined on a metric space. If $f, g \dots$ are used for elements of the space (as in some of the function metric space examples above) we'll need to use something different for functions defined this metric space. Often we use capital letters such as $D(f) = f'' - 3f' + 2f$, and use the term operator, for a function which acts on functions.

2.3 Convergence

Analysis is largely about convergence, limits and continuity. All of these can we expressed in terms of a metric. Remarkably, for very many of the proofs that you saw in first and second year, all you need to do in order to extend the theorem to general metric spaces is to replace $|x-a|$ with $d(x,a)$ in the appropriate places.

We can start with the following definition.

Definition 2.3.1. Let (X, d) be a metric space. Suppose that $\{x_k\}_{k=1}^{\infty}$ is a sequence of elements of X and that $x \in X$. Then $\{x_k\}$ **converges** to x (in the d metric), written $x_k \to x$ as $k \to \infty$, or $\lim\limits_{k\to\infty} x_k = x$, if

$$d(x_k, x) \to 0 \text{ as } k \to \infty.$$

There is an ϵ-δ definition hiding in the last line. If we expand things out using the definition of convergence of a sequence in $\mathbb{R}$, this says that

> $x_k \to x$ as $k \to \infty$ in (X, d) if for all $\epsilon > 0$ there exists K such that for all $k \geq K$, $d(x_k, x) < \epsilon$.

You often need this in abstract proofs, but usually if you have a specific concrete sequence, then you can just find (or estimate) $d(x_k, x)$ and look at its convergence without resorting to epsilonics!

Example 2.3.2. Let $(X, d) = (\ell^\infty, d_\infty)$. For $k = 1, 2, 3, \ldots$ let

$$x_k = \left(\tfrac{1}{k}, \tfrac{1}{2k}, \tfrac{1}{3k}, \ldots\right)$$

and let $x = (0, 0, 0, \ldots)$. Then $d_\infty(x_k, x) = \frac{1}{k} \to 0$ so $x_k \to x$ in (ℓ^∞, d_∞).

If you write $x_k \to x$ you need to make sure your reader understands which metric you are using! In $X = C[0, 1]$ if

$$f_k(t) = \begin{cases} 1 - kt, & 0 \leq t \leq \frac{1}{k} \\ 0, & \frac{1}{k} < t \leq 1 \end{cases}$$

and $f \equiv 0$, then $f_k \to f$ in (X, d_1) but not in (X, d_∞) (see Example 6 above).

A property that we would like to have of limits is that they are always unique[5]!

Theorem 2.3.3. *Limits in a metric space are unique. (That is, if $x_k \to a$ and $x_k \to b$ then $a = b$.)*

Proof. (By contradiction) Suppose not. That is, there is a sequence of points $\{x_k\}_{k=1}^{\infty}$ in a metric space (X, d), and two distinct points $a, b \in X$ such that $x_k \to a$ and $x_k \to b$ as $k \to \infty$.

Since $a \neq b$, $\delta = d(a, b) > 0$ (by Condition (i) for a metric!). Now, as $x_k \to a$, $d(x_k, a) \to 0$. Here we need the ϵ–K definition we recalled above. Taking $\epsilon = \delta/2$ this means that there exists K_1 such that for all $k \geq K_1$, $d(x_k, a) < \delta/2$.

[5]While this result is important, the main reason for including it is to provide a simple proof which just depends on the properties of a metric.

Since $x_k \to b$, we can also say that there exists K_2 such that for all $k \geq K_2$, $d(x_k, b) < \delta/2$. Now choose k to be that larger of K_1 and K_2. From what we have just said

$$\begin{aligned}
\delta = d(a,b) &\leq d(a, x_k) + d(x_k, b) && \text{(triangle inequality)} \\
&= d(x_k, a) + d(x_k, b) && \text{(symmetry)} \\
&< \frac{\delta}{2} + \frac{\delta}{2} = \delta
\end{aligned}$$

which is impossible. ■

When we generalize convergence further to topological spaces, we'll see, somewhat unfortunately, that in that setting you can have crazy examples where every sequence converges to every point!

2.4 The topology of metric spaces

In one variable calculus, an important hypothesis on many theorems was that the function should be defined on some closed and bounded interval. Or perhaps on some open interval. You may have seen generalizations of these 'topological' notions in multivariable calculus or complex analysis. Our aim now is to entend these ideas to general metric spaces, where they are, if anything, even more important.

An example of the sort of theorem which we want to generalize is the Max–Min Theorem.

Theorem 2.4.1 (Max–Min Theorem)**.** *If f is continuous on the close interval $[a, b]$, then there exist $x_1, x_2 \in [a, b]$ such that*

$$f(x_1) \leq f(x) \leq f(x_2) \qquad \forall x_1, x_2 \in [a, b].$$

This theorem fails if we replace the closed interval with an open one say; take

$$f(x) = \frac{x-1}{x(x-2)}$$

on $(0, 2)$ say. One of our aims in this course is to understand what is about the interval $[a, b]$ which makes this theorem true.

The first fundamental notion is that of an ϵ-ball in a metric space.

Definition 2.4.2. Suppose (X, d) is a metric space, $x_0 \in X$ and $\epsilon > 0$. The **(open) ϵ-ball** centred at x_0 is the set

$$B(x_0, \epsilon) = \{x \in X \,:\, d(x_0, x) < \epsilon\}$$

Remark 2.4.3. 1. In $\mathbb{R}$, $B(x_0, \epsilon)$ is just the open interval $(x - \epsilon, x + \epsilon)$. In $\mathbb{R}^2$ with the usual Euclidean metric, $B(\mathbf{x}, \epsilon)$ is the open disk centred at $\mathbf{x}$ of radius ϵ, and so on.

2. Don't rely on $B(x_0, \epsilon)$ begin 'round'. Even in $\mathbb{R}^n$, if you use a different metric to the Euclidean one, you might get a rather different shape. For example in $\mathbb{R}^3$ with d_∞, then the open ϵ-balls are all cubes.

3. In 'big' metric spaces like $(C[0, 1], d_1)$ it can be hard to get any decent geometric feel for what an ϵ-ball looks like.

4. $B(x_0, \epsilon) \neq \emptyset$ since no matter how small ϵ is, $x_0 \in B(x_0, \epsilon)$. In $\mathbb{R}^n$ with the Euclidean metric, ϵ-balls always contain infinitely many points, but in some metric spaces (eg discrete metric spaces), you can easily get ϵ-balls with just this one element in them.

5. Despite the above remarks we usually draw pictures as if spaces are all like $\mathbb{R}^2$. You just need to develop some skill in not using more than the definitions give you!

Definition 2.4.4. Let (X, d) be a metric space and suppose that $Y \subseteq X$.

(i) A point $x_0 \in Y$ is an **interior point** of Y if there is some ϵ-ball $B(x_0, \epsilon)$ which lies completely inside Y.

(ii) A point $x_0 \in X$ is a **boundary point** of Y if every ϵ-ball $B(x_0, \epsilon)$ contains some points of Y and some points of the complement $X \setminus Y$.

Every element of Y is either an interior point of Y or a boundary point of Y — **it can't be both**! Interior points of Y have to be elements of Y, but boundary points might or might not belong to Y.

Note that the boundary of Y is exactly the same as the boundary of the complement $X \setminus Y$. This means that we can split up X into three **disjoint** sets:

$$X = \{\text{interior points of } Y\} \cup \{\text{boundary of } Y\} \cup \{\text{interior points of } X \setminus Y\}.$$

For simple sets in simple metric spaces it is easy to identify the boundary points (and hence the interior points). For less familiar spaces you need to go back to the definition and you may need to work hard.

Example 2.4.5. 1. Let $X = \mathbb{R}^2$ and let $Y = \mathbf{x}$ be a single point set. Then with the usual metric $\mathbf{x}$ is a boundary point of Y, but with the silly discrete metric, $\mathbf{x}$ is an interior point of Y since $B(\mathbf{x}, 1/2)$ only contains $\mathbf{x}$.

2. In an open interval $Y = (a, b) \subseteq \mathbb{R}$, every element of Y is an interior point (under the usual metric). If $Y = \mathbb{Q} \subseteq \mathbb{R}$, then every element of Y is a boundary point.

3. In a space like $(C[0,1], d_\infty)$ you can't use your geometric knowledge. Let[6] $Y = C^1[0,1] \subseteq X$. What are the interior and boundary points of Y?

Definition 2.4.6. Let (X, d) be a metric space and suppose that $Y \subseteq X$. Then Y is **open** (in (X, d)) if **every** element of Y is an interior point of Y.

Example 2.4.7.

1. In $\mathbb{R}$ open intervals (a, b) are open, but so are many other sets, such as $\emptyset$, $\mathbb{R}$, $(0,1) \cup (2,3) \cup (4,5) \cup \ldots$.

2. In a discrete metric space, every set is open since if $x \in Y \subseteq X$ then $B(x, 1/2) = \{x\} \subseteq Y$.

3. In any metric space there are at least two open sets: $\emptyset$ and X.

Many of you will have met many of these topological concepts in $\mathbb{C}$ or $\mathbb{R}^n$ is your second year courses. Our definitions here of course match those you have already seen. A challenge for students at this point is to not rely *too* heavily on the examples you met there to inform your intuition.

The 'in (X, d)' is important! Let Y be the set of all rationals in $(0,1)$. Considered as a subset of $X = \mathbb{R}$ with the usual metric, all the elements of Y are boundary points so the set is far from being open. But what if we take $X = \mathbb{Q}$ with the usual metric? Balls in this X can only contain elements of X, ie rationals. For example

$$B(1/2, 1/4) = \{x \in \mathbb{Q} \cap (0,1) \,:\, |x - 1/2| < 1/4\} = \{x \in \mathbb{Q} \cap (0,1) \,:\, 1/4 < x < 3/4\} \subseteq Y$$

so $1/2$, and similarly all other elements of Y, are interior points.

A couple of easy results. Check you can prove these (or ask in the tutorial!). (The second requires more work than the first.)

Proposition 2.4.8. *A set Y in (X, d) is open if and only if it contains none of its boundary points.*

Proposition 2.4.9. *Every ϵ-ball in a metric space is open.*

Theorem 2.4.10. *Let (X, d) be a metric space, suppose $\{x_k\}_{k=1}^\infty \subseteq X$ and that $x \in X$. Then the following are equivalent.*

(i) $x_k \to x$ *as* $k \to \infty$.

(ii) For all $\epsilon > 0$, there exists K such that for all $k \geq K$, $x_k \in B(x, \epsilon)$.

(iii) For every open set U containing x, there exists K such that for all $k \geq K$, $x_k \in U$.

[6] $C^1[0,1]$ is the set of all functions whose derivative is in $C[0,1]$, taking one-sided derivatives at the endpoints.

Proof. The first thing to note is that

$$d(x_k, x) < \epsilon \iff x_k \in B(x, \epsilon).$$

Saying $d(x_k, x) \to 0$ is just saying that for all $\epsilon > 0$, there exists K such that for all $k \geq K$, $|d(x_k, x)| = d(x_k, x) < \epsilon$. Thus (i) and (ii) are just different ways of writing exactly the same thing.

Since ϵ-balls are open (iii) must imply (ii). So the real work here is to show that (ii) implies (iii)

Suppose then that (ii) holds, and that U is an open set containing x. As U is open, x is an interior point, and hence there exists $\epsilon > 0$ such that $B(x, \epsilon) \subseteq U$. By (ii) there exists K such that for all $k \geq K$, $x_k \in B(x, \epsilon)$. But this implies that for all $k \geq K$, $x_k \in U$ and so we have established (iii). ■

Thus a sequence converges to x if and only if given any open set containing x, the sequence eventually enters and stays in that set.

The definition of closed sets is easy!

Definition 2.4.11. Let (X, d) be a metric space and suppose $Y \subseteq X$. Then Y is closed (in (X, d)) if $X \setminus Y$ is open.

Beware of thinking of closed as being the opposite of open! If a set isn't open it doesn't mean that it is closed (eg $(0, 1]$ in $\mathbb{R}$). In most useful metric spaces, most sets are neither. In discrete metric spaces all sets are both open and closed. In every space both $\emptyset$ and X are both open and closed.

The definition is often not the most useful way of identifying closed sets.

Definition 2.4.12. Let (X, d) be a metric space and suppose $Y \subseteq X$. A point $x \in X$ is a **limit point** of Y if every ϵ-ball $B(x, \epsilon)$ contains at least one point of Y different from x.

Note that a limit point might or might not be in Y. In something like $\mathbb{R}^n$ with the usual metric, every interior point of Y is a limit point, but this is not the case in general (eg in a discrete metric space).

The name 'limit point' is justified by the following result.

Proposition 2.4.13. *A point $x \in X$ is a* ***limit point*** *of $Y \subseteq X$ if and only if there is a sequence of points $\{x_k\}_{k=1}^{\infty} \subseteq Y \setminus \{x\}$ such that $x_k \to x$.*

Proof. ($\implies$) Suppose that x is a limit point of Y. For $k = 1, 2, \ldots$, the ball $B(x, 1/k)$ contains an element of Y different from x. Call this point x_k. Then $d(x_k, x) < 1/k \to 0$ and hence $x_k \to x$.

($\Leftarrow$) Suppose that there is a sequence of points $\{x_k\}_{k=1}^{\infty} \subseteq Y \setminus \{x\}$ such that $x_k \to x$. Fix $\epsilon > 0$. Then as $x_k \to x$ there exists K such that (in particular) $d(x_K, x) < \epsilon$. Thus x_K is an element of $B(x, \epsilon)$ and it is different from x. Hence x is a limit point of Y. ■

Exercise 2.4.14. One thing that you will need to get used to doing is negating the properties. Show that x is not a limit point of Y if and only if there exists some $\epsilon > 0$ such that the ball $B(x, \epsilon)$ contains no element of Y except possibly x itself.

Theorem 2.4.15. *Let (X, d) be a metric space and suppose $Y \subseteq X$. Then the following are equivalent.*

(i) Y is closed.

(ii) Y contains all its boundary points.

(iii) Y contains all its limit points.

Proof. [7] (i) $\Rightarrow$ (ii): Suppose that Y is closed. That is $X \setminus Y$ is open. Then every element of $X \setminus Y$ is an interior point of $X \setminus Y$. Thus every boundary point of $X \setminus Y$ is contained in $X \setminus (X \setminus Y) = Y$. But the boundaries of Y and $X \setminus Y$ are the same. Thus every boundary point of Y is contained in Y.

(ii) $\Rightarrow$ (iii): Suppose that Y contains all its boundary points. Let x be a limit point of Y. Either

(a) $x \in Y$ — which is what we want — or

(b) $x \notin Y$.

Suppose that $x \notin Y$ and that $\epsilon > 0$. Then the ball $B(x, \epsilon)$ contains a point of Y (other than x), and it also contains the point x which is not in Y. This means that x must be a boundary point, and so it must be in Y!?!

(iii) $\Rightarrow$ (i): Suppose that Y contains all its limit points. We need to show that $X \setminus Y$ is open. Suppose then that $z \in X \setminus Y$. As $z \notin Y$, z can't be a limit point of Y. Thus there is some $\epsilon > 0$ so that the ball $B(z, \epsilon)$ contains no point of Y other than possibly z. But we know that $z \notin Y$ and so $B(z, \epsilon) \subseteq X \setminus Y$. Thus z is an interior point of $X \setminus Y$ and so $X \setminus Y$ is open as required. ■

The reason that this theorem works is that, although in general boundary points are distinct from limit points, if x is an element of X which is not in Y, then x is a boundary point of Y if and only if it is a limit point of Y.

In practice we often apply the following result.

Theorem 2.4.16. *Let (X, d) be a metric space and suppose $Y \subseteq X$. Then Y is closed in (X, d) if whenever $\{x_k\}_{k=1}^{\infty}$ is a sequence of elements in Y which converges to an element $x \in X$, we have $x \in Y$.*

[7] Study this proof. It provides a good example of abstract arguments which just depend on unpacking the definitions and using the logic rules from discrete mathematics. If faced wth proving such a result, the first thing that you need to do is write out precisely what each of the conditions is defined as.

Proof. Suppose that Y contains the all the limits of convergent sequences of elements of Y. Let x be a limit point of Y. By Proposition 2.4.13 there exists a sequence $\{x_k\} \subseteq Y$ such that $x_k \to x$. As x must therefore lie in Y, we see that Y contains all its limit points and hence by Theorem 2.4.15 is closed. ■

Exercise 2.4.17. Prove the converse of this!

Example 2.4.18.

1. For $X = \mathbb{C}$ with the metric $d(z, w) = |z - w|$, the definitions that we have given here match with the ones that you would have seen in complex analysis.

2. Let $X = C[0, 1]$ and let

$$Y = \{f \in C[0, 1] \,:\, f(0) = 0\}.$$

Whether Y is open or closed depends on the metric we use.

First consider the metric $d_\infty(f, g) = \sup\limits_{t\in[0,1]} |f(t) - g(t)|$. Suppose that $\{f_k\}_{k=1}^\infty$ is a sequence of elements of Y and that $f_k \to f$ in the d_∞ metric. Then

$$|f(0)| = |f(0) - f_k(0)| \leq \sup_{t\in[0,1]} |f(t) - f_k(t)| = d_\infty(f, f_k) \to 0$$

and hence $f(0) = 0$. That is $f \in Y$. Thus Y contains the all the limits of convergent sequences of elements of Y and hence is closed.

On the other hand, suppose that $f \in Y$, and that $\epsilon > 0$. The ball $B(f, \epsilon)$ obviously contains an element of Y, namely f. Let $g(t) = f(t) + \epsilon/2$. Then $g \in C[0, 1]$, and $g \in B(f, \epsilon)$, but $g(0) \neq 0$ and so $B(f, \epsilon)$ contains an element of $X \setminus Y$. Thus every element of Y is a boundary point. In particular Y is not open in this metric.

3. Same sets with metric $d_1(f, g) = \displaystyle\int_0^1 |f(t) - g(t)|\, dt$. For $k = 1, 2, \ldots$ let

$$f_k(t) = \begin{cases} kt, & 0 \leq t \leq \frac{1}{k} \\ 1, & \frac{1}{k} < t \leq 1. \end{cases}$$

Then (check) $f_k \in Y$ and $d_1(f_k, f) \to 0$ where $f(t) = 1$ for all t. However $f \notin Y$ and hence Y is not closed. As an exercise, repeat the proof from (2) to show that Y is not closed in this metric either.

2.5 Closure and density

Sometimes in analysis it is easier to first prove something for objects that have some special form, such as

- rational numbers in $\mathbb{R}$;
- polynomials in $C[0,1]$;
- eventually zero sequences in ℓ^∞.

One can then sometimes extend the proof to all objects in the space by employing some sort of limiting argument. To do this, you need to know what things are limits of objects in your 'nice' set.

Definition 2.5.1. Let (X,d) be a metric space and let $Y \subseteq X$. The **closure** of Y (in (X,d)) is the set $\mathrm{cl}(Y)$ consisting of Y and all its limit points.

From the comment after Theorem 2.4.15 the closure of Y is also equal to the union of Y and the set of its boundary points. Note that the complement of $\mathrm{cl}(Y)$ is therefore the set of interior points of $X \setminus Y$, which is always an open set. Thus $\mathrm{cl}(Y)$ is always closed.

Proposition 2.5.2. $Y = \mathrm{cl}(Y)$ *if and only if* Y *is closed.*

Proof. Exercise! ■

Example 2.5.3. 1. The closure of $(0,1)$ in $\mathbb{R}$ is $[0,1]$. The closure of $\mathbb{Z}$ in $\mathbb{R}$ is $\mathbb{Z}$.

2. The closure of $\mathbb{Q}$ in $(\mathbb{Q}, |\cdot|)$ is $\mathbb{Q}$. In $(\mathbb{R}, |\cdot|)$ it is all of $\mathbb{R}$.

3. Let
$$c_{00} = \{(x_1, x_2, \dots) \in \ell^\infty \,:\, \exists K \text{ such that } x_k = 0 \text{ for } k \geq K\}.$$
Then the closure of c_{00} in ℓ^∞ is
$$c_0 = \{(x_1, x_2, \dots) \in \ell^\infty \,:\, \lim_{k\to\infty} x_k = 0\}.$$

Definition 2.5.4. Let (X,d) be a metric space and let $Y \subseteq X$. We say that Y is dense in (X,d) if $X = \mathrm{cl}(Y)$.

In other words, if Y is dense in X then every element of X can be approximated arbitrarily closely by an element of Y.

Thus $\mathbb{Q}$ is dense in $\mathbb{R}$ but c_{00} is not dense in ℓ^∞. Later we'll prove the Stone-Weierstrass Theorem which tells us that the polynomials are dense in $(C[0,1], d_\infty)$.

2.6 Indexed collections, unions and intersections

We'll assume that everyone is happy with taking the union $A \cup B$ and the intersection $A \cap B$ of two sets — or of finitely many sets. In analysis we sometimes need to use somewhat more complicated unions and intersections.

Suppose that for each α in some 'index set' $\mathcal{I}$ we are given a set S_α (all sitting inside some big 'universal set' X).

Example 2.6.1. 1. Take $\mathcal{I} = \mathbb{N}$. For $k \in \mathbb{N}$, let $S_k = (-1/k, k^2) \subseteq \mathbb{R}$.

2. Take $\mathcal{I} = C[0,1]$ and for $f \in \mathcal{I}$ let

$$S_f = \{g \in C[0,1] \, : \, |g(t)| \leq |f(t)| \text{ for all } t\}.$$

   Here the index set is uncountable and doesn't even have any natural way of ordering the elements.

3. For any collection (that is, set) of subsets of X, you can take that collection as your index set $\mathcal{I}$ and for $A \in \mathcal{I}$, set $S_A = A$.

   For example, fix a metric space (X, d) and a subset $Y \subseteq X$. Let[8]

$$\mathcal{I} = \{C \subseteq X \, : \, Y \subseteq C \text{ and } C \text{ is closed}\}.$$

   Note that whatever X and Y are, $\mathcal{I}$ is always nonempty since $X \in \mathcal{I}$!

   For $C \in \mathcal{I}$, let $S_C = C$.

We define general unions and intersections by

$$\bigcup_{\alpha \in \mathcal{I}} S_\alpha = \{x \in X \, : \, \text{there exists } \alpha_0 \in \mathcal{I} \text{ such that } x \in S_{\alpha_0}\},$$
$$\bigcap_{\alpha \in \mathcal{I}} S_\alpha = \{x \in X \, : \, x \in S_\alpha \text{ for all } \alpha \in \mathcal{I}\}.$$

Taking Example 1,

$$\bigcup_{k \in \mathbb{N}} S_k = (-1, \infty),$$
$$\bigcap_{k \in \mathbb{N}} S_k = [0, 1).$$

Example 2 requires a bit more thought. Every $f \in C[0,1]$ is in at at least one of these sets, namely S_f and so $\bigcup_{f \in C[0,1]} S_f = C[0,1]$. On the other hand, there is very little that lies in all the sets S_f: $\bigcap_{f \in C[0,1]} \{z\}$ where z is the function which is identically zero.

In Example 3, $\bigcap_{C \in \mathcal{I}} C$ is just the closure of Y. (Prove this!)

We will often show that sets are open or closed by expressing them as unions or intersections of simpler open and closed sets.

[8]If you want something even more concrete, let $X = \mathbb{R}$ and let $Y = \mathbb{Q} \cap (0, 1)$. In this case $\mathcal{I}$ is prety big!

Theorem 2.6.2. *Let (X, d) be a metric space.*

1. *If $\{S_\alpha\}_{\alpha\in\mathcal{I}}$ is an indexed collection of open sets then $\bigcup_{\alpha\in\mathcal{I}} S_\alpha$ is open.*

2. *If $\{S_k\}_{k=1}^n$ is a finite collection of open sets then $\bigcap_{k=1}^n S_k$ is open.*

Proof. [9] 1. Suppose that $x \in \bigcup_{\alpha\in\mathcal{I}} S_\alpha$. Then there exists some $\alpha_0 \in \mathcal{I}$ such that $x \in S_{\alpha_0}$. As S_{α_0} is open there exists $\epsilon > 0$ such that $B(x, \epsilon) \subseteq S_{\alpha_0}$. But then $B(x, \epsilon) \subseteq \bigcup_{\alpha\in\mathcal{I}} S_\alpha$ and hence x is an interior point of this union. Thus the union is open.

2. Suppose that $x \in \bigcap_{k=1}^n S_k$. Pick any $k \in \{1, 2, \dots, n\}$. Then $x \in S_k$ and S_k is open so there exists $\epsilon_k > 0$ such that $B(x, \epsilon_k) \subseteq S_k$. Let $\epsilon = \min\{\epsilon_1, \dots, \epsilon_n\}$. Then $\epsilon > 0$ and $B(x, \epsilon) \subseteq B(x, \epsilon_k) \subseteq S_k$ for each k. Thus $B(x, \epsilon) \subseteq \bigcap_{k=1}^n S_k$ and hence x is an interior point of the intersection. Thus the intersection is open. ■

The proof for (2) here would fail if we tried taking the intersection of infinitely many sets since it might not produce a positive ϵ.

Example 2.6.3. Let $S_k = (-1/k, 1/k) \subseteq \mathbb{R}$. Then $\bigcap_{k=1}^\infty S_k = \{0\}$ which is not open.

Exercise 2.6.4. Prove the corresponding result for closed sets:

1. If $\{S_\alpha\}_{\alpha\in\mathcal{I}}$ is an indexed collection of closed sets then $\bigcap_{\alpha\in\mathcal{I}} S_\alpha$ is closed.

2. If $\{S_k\}_{k=1}^n$ is a finite collection of closed sets then $\bigcup_{k=1}^n S_k$ is closed.

Every open set Y in a metric space (X, d) can be written as a union of ϵ-balls. Suppose that Y is open. Every $x \in Y$ is an interior point so there exists $\epsilon_x > 0$ such that $B(x, \epsilon_x) \subseteq Y$. Clearly then[10]

$$\bigcup_{x\in Y} B(x, \epsilon_x) \subseteq Y.$$

On the other hand if $x \in Y$ then x is in at least one of the balls: $x \in B(x, \epsilon_x)$ and hence $Y \subseteq \bigcup_{x\in Y} B(x, \epsilon_x)$.

Exercise 2.6.5. Prove that every open set $U \subseteq \mathbb{R}$ can be written as a **countable** union of **disjoint** open ϵ-balls. (Is this true in $\mathbb{R}^2$?)

[9] This is basically just abstract set-theoretic nonsense. Make sure you understand it!

[10] This idea of using the whole set as your index set gets used a lot — remember it!

2.7 Continuity

We want to be able to talk about continuous maps between metric spaces. For functions $f : \mathbb{R} \to \mathbb{R}$ you use a couple of equivalent tests for continuity.

1. f is continuous at x_0 if $\lim_{x \to x_0} f(x) = f(x_0)$.
2. f is continuous at x_0 if for all $\epsilon > 0$ there exists $\delta > 0$ such that if $|x - x_0| < \delta$ then $|f(x) - f(x_0)| < \epsilon$.
3. f is continuous at x_0 if whenever $x_n \to x_0$, we have $f(x_n) \to f(x_0)$.

Condition (2) is just expanding out explicitly what condition (1) means. Condition (3) is called sequential continuity, and one has to prove that it is equivalent to (1) and (2).

Either of conditions (2) or (3) could be easily generalized to the metric space setting. We need to check that they are still equivalent. We'll also see that many of the proofs turn out to be rather neater if we use another equivalent condition which involves open sets. This last one will be the version which we use when we get to topological spaces.

Definition 2.7.1. Let (X, d_X) and (Y, d_Y) be metric spaces and suppose that $f : X \to Y$.

1. We say that f is **continuous at a point** $x_0 \in X$ if for all $\epsilon > 0$ there exists $\delta > 0$ such that if $d_X(x, x_0) < \delta$ then $d_Y(f(x), f(x_0)) < \epsilon$.
2. We say that f is **continuous on** X if f is continuous at every point of X.

As in the real line, as much as possible you avoid getting buried in complicated ϵ-δ arguments, but sometimes there isn't much option.

Suppose that X is a set of functions, such as $C[0, 1]$. We'll usually use capital letters like T to denote functions (or operators) from X to X. If $f \in X$ then $T(f)$ is also a function so we can write things like $T(f)(t)$ for the value of that function at t. More usually we write Tf and $Tf(t)$ as all those parentheses get a bit confusing!

Example 2.7.2. The Volterra operator $V : C[0, 4] \to C[0, 4]$ (using the d_∞ metric). For $f \in C[0, 4]$, Vf is defined as the function

$$Vf(t) = \int_0^t f(u)\, du, \qquad t \in [0, 4].$$

Suppose that $f, f_0 \in C[0,4]$. Then, taking things slowly,

$$\begin{aligned}
d_\infty(Vf, Vf_0) &= \sup_{t\in[0,4]} |Vf(t) - Vf_0(t)| \\
&= \sup_{t\in[0,4]} \Big| \int_0^t f(u)\,du - \int_0^t f_0(u)\,du \Big| \\
&\le \sup_{t\in[0,4]} \int_0^t |f(u) - f_0(u)|\,du \\
&\le \int_0^4 |f(u) - f_0(u)|\,du \\
&\le 4 \sup_{u\in[0,4]} |f(u) - f_0(u)| \\
&= 4 d_\infty(f, f_0).
\end{aligned}$$

Suppose then that $\epsilon > 0$. If we choose $\delta = \epsilon/4$ then if $d_\infty(f, f_0) < \delta$ then $d_\infty(Vf, Vf_0) < 4\delta = \epsilon$ and hence V is continuous at f_0. As f_0 was arbitrary, V is continuous on all of $C[0,4]$.

Exercise 2.7.3. Let $X = \ell^\infty$ (with the usual d_∞ metric) and define $f : \ell^\infty \to \ell^\infty$ by

$$f(x_1, x_2, x_3, \dots) = (x_1 + x_2, x_3 + x_4, x_5 + x_6, \dots).$$

Prove that f is a continuous function from (X, d_∞) to (X, d_∞).

You may have seen a version of the following theorem in second year. It provides the condition we'll need to generalize the idea of continuity to topological spaces.

Theorem 2.7.4. *Let (X, d_X) and (Y, d_Y) be metric spaces and suppose that $f : X \to Y$. Then f is continuous on X if and only if $f^{-1}(U)$ is an open set in (X, d_X) for every open set U in (Y, d_Y).*

Proof. [11] ($\Rightarrow$) Suppose that f is continuous on X, and suppose that U is an open set in (Y, d_Y). Let $W = f^{-1}(U)$, and suppose that $x_0 \in W$. We need to show that x_0 is an interior point of W and hence that W is open.

Let $y_0 = f(x_0) \in U$. As U is open, y_0 is an interior point of U and hence there exists $\epsilon > 0$ so that $B(y_0, \epsilon) \subseteq U$. As f is continuous at x_0, there exists $\delta > 0$ such that if $d_X(x, x_0) < \delta$ then $d_Y(f(x), f(x_0)) = d_Y(f(x), y_0) < \epsilon$. Reinterpreting this: if $x \in B(x_0, \delta)$ then $f(x) \in B(y_0, \epsilon)$ and hence $f(x) \in U$, and hence $x \in f^{-1}(U) = W$. Thus $B(x_0, \delta) \subseteq W$ and hence x_0 is an interior point of W.

($\Leftarrow$) Suppose that the inverse image of every open set is open, and suppose that $x_0 \in X$. Suppose that $\epsilon > 0$. We need to show that there exists $\delta > 0$ such that if $d_X(x, x_0) < \delta$, then $d_Y(f(x), f(x_0)) < \epsilon$.

[11] The main idea here is just swapping between writing $b \in B(a, \epsilon)$ and the equivalent $d(b, a) < \epsilon$.

Let $U = B(f(x_0), \epsilon)$. All ϵ-balls are open so $W = f^{-1}(U)$ must be open. Now clearly $x_0 \in W$ so x_0 is an interior point of W. That is, there exists $\delta > 0$ so that $B(x_0, \delta) \subseteq W$. Reinterpreting this: if $d_X(x, x_0) < \delta$ then $f(x) \in U = B(d(x_0), \epsilon)$ and hence $d_Y(f(x), f(x_0)) < \epsilon$. ■

Continuous functions don't usually send open sets to open sets! Try $X = Y = \mathbb{R}$ and $f(x) = 1$ or $f(x) = \sin x$. See the problems.

Theorem 2.7.4 allows many nice 'ϵ-free' proofs of standard theorems.

Theorem 2.7.5. *The composition of continuous functions is continuous. That is, if (X, d_X), (Y, d_Y) and (Z, d_Z) are metric spaces and $f : X \to Y$ and $g : Y \to Z$ are continuous, then $g \circ f : X \to Z$ is continuous.*

Proof. Note that if $U \subseteq Z$, then $(g \circ f)^{-1}(U) = f^{-1}(g^{-1}(U))$. If U is open then $g^{-1}(U)$ is open as g is continuous and hence $f^{-1}(g^{-1}(U))$ is open as f is continuous. Thus $g \circ f$ is continuous. ■

Another very useful tool is 'sequential continuity'.

Definition 2.7.6. Let (X, d_X) and (Y, d_Y) be metric spaces and suppose that $f : X \to Y$. We say that f is **sequentially continuous at** $x_0 \in X$ if whenever $x_k \to x_0$ in (X, d_X), $f(x_k) \to f(x_0)$ in (Y, d_Y). We say f is **sequentially continuous on** X if it is sequentially continuous at every point in X.

In the metric space setting this turns out to equivalent to being continuous. In more general settings this won't be the case. The proof is left as an exercise.

Theorem 2.7.7. *Let (X, d_X) and (Y, d_Y) be metric spaces and suppose that $f : X \to Y$. Then f is sequentially continuous at $x_0 \in X$ if and only if f is continuous at x_0.*

Exercise 2.7.8. Use sequential continuity to give another proof of Theorem 2.7.5.

In multivariable calculus you may have seen that you could use any of the metrics d_p (from Example 2.2.2) to test the continuity of a function $f : \mathbb{R}^n \to \mathbb{R}^m$. If it was continuous using one, it was continuous using all of them. The point is that although the metrics are different, they produce the same collections of open and closed sets in these spaces. One consequence of Theorem 2.7.4 is that for many things, it is these collections of open and closed sets which really matter, not the actual metrics themselves. Of course you can find some metrics (like the discrete metric) which give a different collection of open sets, but for these spaces, pretty much all sensible metrics give the same 'topological structure' to the space. For infinite dimensional spaces, the situation is very different! In the next chapter we'll pursue these ideas further by concentrating on the topological structure and completely losing the metric!

2.8 Normed spaces

Much of what we are trying to do is to extend ideas from $\mathbb{R}^n$ to more general settings. Of course $\mathbb{R}^n$ has more structure than just a sense of distance. In particular it has the structure of a vector space. Fortunately many of the setting where you want to do analysis also have this structure. For example, the sets of continuous, differentiable or integrable functions are all closed under addition and scalar multiplication. The standard metrics on$\mathbb{R}^n$ are often rather special in that they combine nicely with the vector space structure. For example the metrics are translation invariant ($d(x+z, y+z) = d(x,y)$) and 'positively homogeneous' ($d(\lambda x, \lambda y) = \lambda d(x,y)$ if $\lambda \geq 0$).

Definition 2.8.1. Let V be a real[12] vector space. A function $\|\cdot\| : V \to \mathbb{R}$ is a **norm** if

(i) $\|v\| \geq 0$ for all $v \in V$ and $\|v\| = 0$ iff $v = 0$;

(ii) $\|\lambda v\| = |\lambda| \, \|v\|$ for all $\lambda \in \mathbb{R}$, $v \in V$;

(iii) $\|u + v\| \leq \|u\| + \|v\|$ for all $u, v \in V$.

The proof of the following observation is an easy exercise.

Proposition 2.8.2. *If $\|\cdot\|$ is any norm on a vector space V, then setting $d(u,v) = \|u - v\|$ defines a metric on V.*

Example 2.8.3.

1. $V = \mathbb{R}^n$, $\|\mathbf{x}\| = \left(\sum_{j=1}^n |x_j|^2\right)^{1/2}$.
2. $V = \ell^1 = \{\mathbf{x} = (x_1, x_2, \dots) : \sum_{k=1}^\infty |x_k| < \infty\}$ with $\|\mathbf{x}\| = \sum_{k=1}^\infty |x_k|$.
3. $V = C[0,1]$ (or $C^k[0,1]$), $\|f\|_\infty = \sup_{t \in [0,1]} |f(t)|$.
4. $V = C[0,1]$, $\|f\|_2 = \left(\int_0^1 |f(t)|^2 \, dt\right)^{\frac{1}{2}}$.

Exercise 2.8.4. Look through the examples of metric spaces we gave earlier and work out which ones are normed spaces. Which ones are subsets of normed spaces?

Some of these spaces come with even more of the structure of $\mathbb{R}^n$. Norms give us a measure of length and distance in our space, but no measure of angle — and in particular, no concept of orthogonality. In the abstract, this is supplied by an inner product, which is something which behaves like the dot product in $\mathbb{R}^n$.

Definition 2.8.5. Let V be a real vector space. An **inner product** on V is a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ such that

[12]This definition also applies to complex vector spaces, but we usually don't go so far as to allow vector spaces over finite fields! For the complex case, you need (ii) to hold for $\lambda \in \mathbb{C}$.

(i) $\langle v, v\rangle \geq 0$ for all $v \in V$ and $\langle v, v\rangle = 0$ iff $v = 0$;

(ii) $\langle \lambda u, v\rangle = \lambda\langle v, v\rangle$ for all $\lambda \in \mathbb{R}$, $u, v \in V$;

(iii) $\langle u, v\rangle = \langle v, u\rangle$ for all $u, v \in V$;

(iv) $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ for all $u, v, w \in V$.

Again it is an easy exercise to check that if $\langle \cdot, \cdot\rangle$ is an inner product, then setting $\|v\| = (\langle v, v\rangle)^{1/2}$ defines a norm, and hence a metric.

Complex inner product spaces need a little more care. The usual dot product on $\mathbb{C}^n$ has $\mathbf{x}\cdot\mathbf{y} = \sum_{k=1}^{n} x_k \overline{y}_k$, where the complex conjugate is there to make sure that we can define a norm as above: $\|\mathbf{x}\|^2 = \mathbf{x}\cdot\overline{\mathbf{x}} \geq 0$. For a complex inner product space you need to replace condition (iii) with $\langle u, v\rangle = \overline{\langle v, u\rangle}$, which means that $\langle u, \lambda v\rangle = \overline{\lambda}\langle v, v\rangle$.

Annoyingly, physicists do all this backwards. They put the complex conjugate on the first component rather than the second.

Example 2.8.6. 1. $\mathbb{R}^n$ and $\mathbb{C}^n$ with their usual dot products.

2. $V = C[0, 1]$ with the inner product

$$\langle f, g\rangle = \int_0^1 f(t)g(t)\, dt.$$

(or $\int_0^1 f(t)\,\overline{g(t)}\, dt$ in the complex case.) This gives example 4 above.

3. $V = \ell^2 = \{\mathbf{x} = (x_1, x_2, \dots) : \sum_{k=1}^{\infty} |x_k|^2 < \infty\}$ with $\langle \mathbf{x}, \mathbf{y}\rangle = \sum_{k=1}^{\infty} x_k y_k$.

Exercise 2.8.7. Vectors u and v is an inner product space V are said to be **orthogonal** if $\langle u, v\rangle = 0$. Prove that in an inner product space you get **Pythagoras' Theorem**: if u and v are orthogonal then $\|u + v\|^2 = \|u\|^2 + \|v\|^2$.

Exercise 2.8.8. 1. Show that in $C[0, 1]$ with the above inner product that $f(x) = \sin 2\pi x$ and $g(x) = \cos 2\pi x$ are orthogonal.

2. When are $\sin(\alpha x)$ and $\cos(\beta x)$ orthogonal? (Perhaps just some special cases will suffice!)

Having an inner product will allow us to really extend much of the $\mathbb{R}^n$ theory to infinite dimensional vector spaces. The theory of Fourier series, for example, is really about choosing a nice orthogonal basis for vector space of functions. Indeed this basis comprises functions which are eigenfunctions for $\frac{d^2}{dx^2}$, so that we are essentially diagonalizing that linear transformation in much the same way that one diagonalized self-adjoint matrices in second year. It turns out that you need one more technical property called completeness for your inner product space to make everything work nicely. We'll get to that shortly.

It is worth pausing at this point to just take stock of how all the different types of spaces we have met are related to one another. Figure 2.1 shows diagrammatically the types of spaces that we'll study at in this course. One of the challenges for us will be to make sure that any definition we give for one class of spaces is consistent with the definition we give in another.

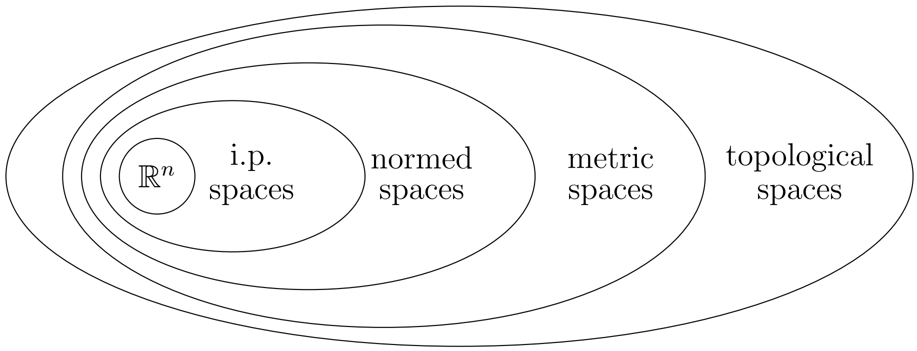


Figure 2.1: Different classes of spaces

A common problem that students have as they progress through the course is that they want to apply very abstract definitions and methods from the most general settings to the problems that they knew how to deal with in first year. Don't forget what you already know!

2.9 Boundedness

In $\mathbb{R}^n$ we saw that 'compact' sets have particularly nice properties. The definition in that setting was that a set is compact if it is closed and bounded. The definition of being bounded extends very easily to a general metric space.

Definition 2.9.1. Suppose that (X, d) is a metric space and that $Y \subseteq X$. We shall say that Y is **bounded** (in (X, d)) if there exists $x_0 \in X$ and $M < \infty$ such that $d(y, x_0) \leq M$ for all $y \in Y$.

Note that if $d(y, x_0) \leq K$ for all $y \in Y$ and x_1 is any other point in X then $d(y, x_1) \leq d(y, x_0) + d(x_0, x_1) \leq M + d(x_0, x_1)$. This means that if you are checking boundedness you can pick any convenient point in X. If you are in a normed space, the most convenient space is usually $x_0 = 0$, so you are just checking that $\|x\| \leq M$ for all $x \in Y$.

Thought of another way, a set is bounded if it sits entirely inside some (big) ϵ-ball around some point in the space.

Exercise 2.9.2. Let $X = C[0,1]$ with norm $\|\cdot\|_\infty$. Let

$$Y = \{f \in C[0,1] \,:\, \int_0^1 |f(t)|\, dt \le 1\}.$$

Prove that Y is not bounded.

It would be great if bounded sets in a general metric space had the same sorts of properties that they do in $\mathbb{R}^n$. For example, in $\mathbb{R}^n$, a bounded sequence always has a convergent subsequence. Unfortunately, such things fail in general metric spaces, and indeed even in normed spaces.

Example 2.9.3. Consider the normed space

$$\ell^2 = \Big\{\mathbf{x} = (x_1, x_2, \dots) \,:\, \|\mathbf{x}\|_2 = \Big(\sum_{j=1}^{\infty} |x_j|^2\Big)^{1/2}\Big\}.$$

For $k = 1, 2, \dots$, let $\mathbf{x}_k$ be the sequence with a one in the kth spot and zeros elsewhere. These are clearly in the closed unit ball $\{\mathbf{x} \,:\, \|\mathbf{x}\|_2 \le 1\}$ but they can't possibly have a convergent subsequence as each pair of vectors in the sequence $\{\mathbf{x}_k\}$ is $\sqrt{2}$ apart.

Nonetheless, we'll see many results where boundedness is an important hypothesis. We can do one easy result now.

Theorem 2.9.4. *Suppose that $\{x_k\}_{k=1}^\infty$ is a convergent sequence in a metric space (X, d). Then the set $\{x_k\}_{k=1}^\infty$ is bounded.*

Proof. Let x be the limit. By the definition of convergence, there exists K such that for all $k \ge K$, $d(x_k, x) < 1$. Let

$$M = \max\{d(x_1, x), \dots, d(x_{K-1}, x), 1\}.$$

Then $d(x_k, x) \le M$ for all k and hence the set is bounded. ■

2.10 Completeness

We don't do analysis with functions $f : \mathbb{Q} \to \mathbb{Q}$ because it has too many 'holes' in it! Let $f : \mathbb{Q} \to \mathbb{Q}$, $f(x) = x^2 - 2$. The iteration

$$a_0 = 2, \qquad a_{n+1} = \frac{a_n^2 + 2}{2a_n}, \qquad n \ge 0$$

produces a sequence of rationals for which $f(a_n) \to 0$, but the sequence doesn't converge in $(\mathbb{Q}, |\cdot|)$ to a root of the function.

In analysis we want sequences that look like they should converge to actually converge to something. Sometimes this involves filling in the holes, which is essentially what you do when you construct $\mathbb{R}$ from $\mathbb{Q}$. We first need to be more precise about what a hole is!

Example 2.10.1. Let $X = C[0,1]$ with the d_1 metric. For $k = 1, 2, \ldots$ let

$$f_k(t) = \begin{cases} 0, & 0 \le t \le \frac{2k-1}{4k} \\ 2\,k\left(t - \frac{1}{2}\right) + \frac{1}{2}, & \frac{2k-1}{4k} < t < \frac{2k+1}{4k} \\ 1, & \frac{2k+1}{4k} \le t \le 1. \end{cases}$$

Exactly calculating $d_1(f_k, f_\ell)$ is a bit of a challenge, but if $k < \ell$ then it is easy to see that $d_1(f_k, f_\ell) < 1/k$, so as the indices get large these functions are getting close together. With a little effort you can show that if $f_k \to f$ then $f(t)$ must be 0 on $[0, 1/2)$ and 1 on $(1/2, 1]$ which is impossible for a continuous function. Roughly speaking, this sequence is converging to a function which isn't in the original set, so we have a 'hole' where the limit should be.

This example gives an idea of how one might make the statement 'it looks like it should converge' more precise!

Definition 2.10.2. Let $\{x_k\}_{k=1}^{\infty}$ be a sequence in a metric space (X, d). We say that $\{x_k\}$ is a **Cauchy sequence** if for all $\epsilon > 0$ there exists K such that for all $k, \ell \ge K$, $d(x_k, x_\ell) < \epsilon$.

The sequence in Example 2.10.1 is therefore Cauchy. To confirm that we are on the right track we have the following result.

Theorem 2.10.3. *Every convergent sequence is Cauchy.*

Proof. Suppose that $\{x_k\}_{k=1}^{\infty}$ converges in (X, d) to $x \in X$. Fix any $\epsilon > 0$ By the definition of convergence, since $\epsilon/2 > 0$ there exists K such that for all $k \ge K$, $d(x_k, x) < \epsilon/2$. But this means that if $k, \ell \ge K$,

$$d(x_k, x_\ell) \le d(x_k, x) + d(x, x_\ell) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

and hence the sequence is Cauchy. ■

This theorem is really useful if you want to show that a sequence doesn't converge.

Example 2.10.4. Take the real sequence $x_k = \sin(k)$, $k = 1, 2, \ldots$. If you use the definition directly you need to show that for any $L \in \mathbb{R}$ there exists $\epsilon > 0$ such that for all K there exists $k \ge K$ such that $d(x_k, L) \ge \epsilon$. This is hard work!

Alternatively, to show that the sequence is not Cauchy we can just look at $d(x_k, x_\ell)$. Now consider the points e^{ik}, $k = 1, 2, \ldots$. The distance from $e^{i\pi/6}$ to $e^{i5\pi/6}$ around the circle is $\frac{2\pi}{3} > 1$. It is clear then that on each 'circuit' at least one of these points must land in the top bit of the circle. Said more formally, for all K, there exists $k \ge K$ such that $\sin k \ge \frac{1}{2}$. The same thing works with the bottom half of the circle, so for any K there exists $\ell \ge K$ such that $\sin \ell \le \frac{1}{2}$. In particular, no matter how big K is, there exists $k, \ell \ge K$ with $d(x_k, x_\ell) \ge 1$.

Knowing that $d(x_k, x_{k+1})$ gets small as k gets large is not enough for a sequence to be Cauchy. To show a sequence is Cauchy you need to consider all possible pairs of elements with big indices. To show that it is not Cauchy of course you only need to find some pairs of sequence elements with arbitrarily high indices which are not close together.

Exercise 2.10.5. Adapt the proof of Theorem 2.9.4 to show that a Cauchy sequence must be bounded.

Definition 2.10.6. A metric space (X, d) is said to be **complete** if every Cauchy sequence in X converges (to an element of X!).

From the examples above we can see that $(\mathbb{Q}, |\cdot|)$ and $(C[0,1], d_1)$ are not complete. On the other hand $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{R}^n$ with their usual metrics are complete. That $\mathbb{R}$ is complete is not trivial! Just how one does this depends very much on how you have defined the real numbers. Roughly speaking, the completeness of $\mathbb{R}$ is equivalent to the Least Upper Bound property of $\mathbb{R}$ that you saw in first year. In this course you may assume that $\mathbb{R}$ is complete.

Given any incomplete metric space (X, d) there is an construction which produces a complete space $(\hat{X}, \hat{d})$ which contains an exact dense copy of (X, d). This space, called the **completion** of (X, d), is essentially unique. (See §7.4 of Kolmogorov and Fomin.)

If you can find a complete space (Y, d) that (X, d) is sitting inside, then the completion of (X, d) is just the closure of X in Y. For example, $\mathbb{R}$ is complete, so the completions of $\mathbb{Q}$ and $(0, 1)$ are $\mathbb{R}$ and $[0, 1]$ respectively. (See Proposition 2.10.12 below.)

In general however, you don't have much of a handle on what the extra stuff ('the holes') that you have filled in are. For example, if (X, d) is a vector space of functions on some domain Ω, you can't necessarily identify the elements of $\hat{X}$ as functions on Ω. Later we will take a space like $(C[0,1], d_1)$ and 'fill in the holes' to make a complete space called $L^1[0,1]$, but there are some real technical difficulties here as the elements of $L^1[0,1]$ are not actually functions!

Fortunately, some of the bigger spaces we have met so far are already complete. One that you have already seen in second year is $(C[0,1], d_\infty)$. The completeness of this space is almost the statement that if a sequence of continuous functions converges uniformly (ie in d_∞ metric), then the limit is continuous.

Theorem 2.10.7. *$(C[0,1], d_\infty)$ is complete.*

We'll do the proof in two parts.

Lemma 2.10.8. *If $\{f_k\}_{k=1}^\infty$ is a Cauchy sequence in $(C[0,1], d_\infty)$ then there exists a function $f : [0,1] \to \mathbb{R}$ such that $f_k \to f$ uniformly.*

Proof. Fix $t_0 \in [0, 1]$. Note that for any k, ℓ

$$|f_k(t_0) - f_\ell(t_0)| \le d_\infty(f_k, f_\ell) = \sup_{t \in [0,1]} |f_k(t) - f_\ell(t)|.$$

Since the RHS can be made arbitrarily small by making k and ℓ large enough, so can the LHS and hence the real sequence $\{f_k(t_0)\}$ is Cauchy. But $\mathbb{R}$ is complete, so this sequence converges to some real number, which we shall call $f(t_0)$ (since it depends on t_0). This uniquely determines a function $f : [0,1] \to \mathbb{R}$, which is the pointwise limit of f_k.

Fix $\epsilon > 0$. There exists K such that if $k, \ell \geq K$, $d_\infty(f_k, f_\ell) < \epsilon/2$. Suppose now that $t \in [0,1]$. As $f_k(t) \to f(t)$, there exists K_t such that if $\ell \geq K_t$ then $|f(t) - f_\ell(t)| < \epsilon/2$. Choose $L_t = \max(K, K_t)$. Then for all $k \geq K$,

$$\begin{aligned} |f(t) - f_k(t)| &= |f(t) - f_{L_t}(t) + f_{L_t}(t) - f_k(t)| \\ &\leq |f(t) - f_{L_t}(t)| + |f_{L_t}(t) - f_k(t)| \\ &\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

This holds for all t, so $d_\infty(f_k, f) < \epsilon$ for all $k \geq K$, ie $f_k \to f$ uniformly. ■

Lemma 2.10.9. *If $\{f_k\}_{k=1}^\infty$ is a Cauchy sequence in $(C[0,1], d_\infty)$ then its limit f is in $C[0,1]$.*

Proof. Fix $t_0 \in [0,1]$ and suppose that $\epsilon > 0$. We need to show that there exists $\delta > 0$ such that if $|t - t_0| < \delta$ then $|f(t) - f(t_0)| < \epsilon$.

As $f_k \to f$ uniformly there exists K such that for all $k \geq K$,

$$d_\infty(f, f_k) = \sup_{t \in [0,1]} |f(t) - f_k(t)| < \frac{\epsilon}{3}. \tag{2.10.1}$$

Now f_K is continuous at t_0 so there exists $\delta > 0$ such that if $|t - t_0| < \delta$ then $|f_K(t) - f_K(t_0)| < \epsilon/3$. Thus if $|t - t_0| < \delta$ then (using (2.10.1) with $k = K$)

$$\begin{aligned} |f(t) - f(t_0)| &= |f(t) - f_K(t) + f_K(t) - f_K(t_0) + f_K(t_0) - f(t_0)| \\ &\leq |f(t) - f_K(t)| + |f_K(t) - f_K(t_0)| + |f_K(t_0) - f(t_0)| \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon \end{aligned}$$

as required. ■

An obvious question is what role $[0,1]$ plays in this proof. The short answer is 'very little'! If (X, d) is any metric space, then the set of continuous functions from X to $\mathbb{R}$, denoted $C(X)$ is always a vector space. Unfortunately (eg if $X = \mathbb{R}$ or $X = (0,1)$) functions in $C(X)$ need not be bounded. You can fix this by working in the normed space $C_b(X)$ of continuous bounded real-valued functions on X. This space is always complete, and the proof is essentially exactly the same as the one just given. Of course $C_b[0,1] = C[0,1]$. We'll see later that $C_b(X) = C(X)$ whenever (X, d) is a compact metric space.

'Functional analysis', which we shall touch on a little, but which you will mainly study in 4th year, combines a mixture of linear algebra and analysis. That is, it largely concerns linear transformations defined on infinite dimensional vector spaces. Here issues such as expansions in terms of a basis (think Fourier series), and the continuity of linear transformations (think differentiation) raise delicate analytical questions. The standard setting for this study is the class of complete normed vector spaces.

Definition 2.10.10. (i) A **Hilbert**[13] **space** is a complete inner product space.

(ii) A **Banach**[14] **space.** is a complete normed vector space.

What is not at all clear from this definition is how much bigger and wilder the class of Banach spaces is compared to the class of Hilbert spaces. Next year you will see that all Hilbert spaces 'of the same dimension' are isomorphic. Over the past 40 years it has become clear that the class of Banach spaces is a zoo!

Example 2.10.11. (i) The standard examples of Hilbert spaces are $\mathbb{R}^n$ and $\mathbb{C}^n$ with the usual dot products, and ℓ^2 with the 'infinite dot product', $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{k=1}^{\infty} x_k y_k$ (with a complex conjugate in the complex case!). An important example we'll look at later is $L^2[0,1]$ which is roughly the set of functions $f : [0,1] \to \mathbb{R}$ for which $\displaystyle\int_0^1 |f(t)|^2 \, dt$ is finite, with the inner product $\langle f, g \rangle = \displaystyle\int_0^1 f(t)g(t) \, dt$. Making sense of the 'roughly' here requires some work! The fact that ℓ^2 and $L^2[0,1]$ are really the same space in disguise (that is, they are isometrically isomorphic) is central to the theory of Fourier series.

(ii) Of course all Hilbert spaces are Banach spaces. The standard infinite-dimensional examples of non-Hilbertian Banach spaces are $(C[0,1], \|\cdot\|_\infty)$, $(\ell^p, \|\cdot\|_p)$ and $(c_0, \|\cdot\|_\infty)$. It was once hoped that these 'classical Banach spaces' might form the building blocks for a general theory of Banach spaces, it is now known that that the class of Banach spaces contains many strange and mysterious spaces! Nonetheless, these classical Banach spaces are very important in the general theory. You can read a little more about them in Section 2.14.

Showing that a normed vector space is complete is often non-trivial! One of the real nuisances of analysis is that the space $C[0,1]$ equipped with the inner product $\langle f, g \rangle = \int_0^1 f(t)g(t) \, dt$ is not complete. Filling in the 'holes' here is messy — we'll come back to this later!

Much of the time we can take a shortcut by noting that the space we want to use is already sitting inside a complete space.

[13]David Hilbert (1862-1943) was a German mathematician who was one of the most influential mathematicians of the early 20th century.

[14]Stefan Banach (1892 - 1945) was a Polish mathematician who was one of the founders of functional analysis

Proposition 2.10.12. *Let (X, d) be a complete metric space and suppose that $Y \subseteq X$. Then (Y, d) is complete if and only if Y is closed in X.*

Proof. $(\Rightarrow)$ Suppose that (Y, d) is complete. Suppose that $\{x_k\}$ is a sequence in Y that converges to a point $x \in X$. As $\{x_k\}$ converges it is Cauchy. As it is a Cauchy sequence in a complete metric space it must converge to an element of $Y \subseteq X$. Since limits are unique this means that $x \in Y$. By Theorem 2.4.16 this means that Y is closed.

$(\Leftarrow)$ Suppose that Y is closed in X. Suppose that $\{x_k\}$ is a Cauchy sequence in Y — and hence in X. As (X, d) is complete, this sequence must converge to an element $x \in X$. But Y is closed, and so it contains the limits of all sequences in Y. Therefore $x \in Y$, so $\{x_k\}$ converges in Y and hence Y is complete. ■

2.11 Contraction Mappings

Before we go further here is a nice application of these ideas.

Consider the complex polynomial

$$p(z) = z^7 + z^3 + 2z^2 - 8z + 3i.$$

Finding roots of p is not easy. Note however that if

$$f(z) = \frac{z^7 + z^3 + 2z^2 + 3i}{8}$$

then

$$p(z) = 0 \iff f(z) = z.$$

It turns out that solving this fixed point problem is easy numerically. Set $z_0 = 0$ and then calculate $z_k = f(z_{k-1})$ for $k = 1, 2, \ldots$. If you try this with Maple, you'll discover that the sequence $\{z_k\}$ quickly converges:

$$\begin{aligned}\{z_k\}_{k=0}^{\infty} = \{&0., 0.37500i, -0.035156 + 0.36828i, -0.031742 + 0.36236i,\\ &- 0.030955 + 0.36335i, -0.031178 + 0.36342i, -0.031175 + 0.36338i,\\ &- 0.031168 + 0.36338i, -0.031169 + 0.36338i, -0.031169 + 0.36338i,\\ &- 0.031169 + 0.36338i, \ldots\}.\end{aligned}$$

Some things to note:

- Try starting this at a point other than zero. Try a complex starting point. Sometimes it converges and sometimes not. Why?

- If $z_k \to z_0$ then, as f is continuous, $f(z_0) = f(\lim z_k) = \lim f(z_k) = \lim z_{k+1} = z_0$ and so z_0 is a fixed point.

- It is important that $\mathbb{C}$ is complete. If you tried this with complex rationals, the sequence would lie in that set, but the limit wouldn't.

We'll see that the things that make this work can actually be applied to solving more complicated equations like ODEs. The first property that f has is that it moves points closer together.

Definition 2.11.1. Let (X, d) be a metric space and suppose that $f : (X, d) \to (X, d)$. We say that f is a **contraction** if there exists a constant $c <$ such that for all $x, y \in X$

$$d(f(x), f(y)) \leq cd(x, y).$$

This is saying more than $d(f(x), f(y)) < d(x, y)$. Actually if you take $x = y$ this would always fail, so you might tweak things to require $d(f(x), f(y)) < d(x, y)$ whenever $x \neq y$.
The map $f : [1, \infty) \to [1, \infty)$

$$f(x) = x + 1/x$$

satisfies

$$|f(x) - f(y)| < |x - y|$$

but is not a contraction.

Theorem 2.11.2 (Contraction Mapping Theorem)**.** *Let (X, d) be a **complete** metric space. If $f : (X, d) \to (X, d)$ is a contraction then there exists a unique point $x_F \in X$ such that $f(x_F) = x_F$.*

Proof. Pick any $x_0 \in X$. Define $x_k = f(x_{k-1})$ for $k = 1, 2, 3, \ldots$. If $0 \leq k < \ell$ then

$$\begin{aligned} d(x_k, x_\ell) &= d(f(x_{k-1}), f(x_{\ell-1})) \\ &\leq c\, d(x_{k-1}, x_{\ell-1}) \\ &= c\, d(f(x_{k-1}), f(x_{\ell-1})) \\ &\;\;\vdots \\ &\leq c^k\, d(x_0, x_{\ell-k}). \end{aligned}$$

Now applying the triangle inequality and the above inequality

$$\begin{aligned} d(x_0, x_{\ell-k}) &\leq d(x_0, x_1) + d(x_1, x_2) + d(x_2, x_3) + \cdots + d(x_{\ell-k-1}, x_{\ell-k}) \\ &\leq d(x_0, x_1) + c\, d(x_0, x_1) + c^2\, d(x_0, x_1) + \cdots + c^{\ell-k-1} d(x_0, x_1) \\ &< d(x_0, x_1) \sum_{i=0}^{\infty} c^i \\ &= \frac{d(x_0, x_1)}{1 - c}. \end{aligned}$$

Thus

$$d(x_k, x_\ell) \leq \frac{c^k\, d(x_0, x_1)}{1 - c}.$$

Note that this bound does not depend on ℓ, so, since $c < 1$, we can make $d(x_k, x_\ell)$ as small as we like by insisting that the smaller index k is sufficiently large. Thus, the sequence $\{x_k\}_{k=1}^{\infty}$ is Cauchy. But (X, d) is complete so there exists $x_F \in X$ such that $d(x_k, x_F) \to 0$. To see that x_F is a fixed point, note that for any $k \geq 1$,

$$\begin{aligned} d(x_F, f(x_F)) &\leq d(x_F, x_k) + d(x_k, f(x_F)) \\ &= d(x_F, x_k) + d(f(x_{k-1}, f(x_F)) \\ &\leq d(x_F, x_k) + c\, d(x_{k-1}, x_F) \end{aligned}$$

which can be made arbitrarily small by making k large enough. Hence $d(x_F, f(x_F)) = 0$ or $f(x_F) = x_F$.

For uniqueness, suppose that $f(x) = x$ and $f(y) = y$ with $d(x, y) = \epsilon > 0$. Then

$$\epsilon = d(x, y) = d(f(x), f(y)) \leq c\, d(x, y) < \epsilon$$

which is clearly impossible. Hence the fixed point is unique. ■

It is important to note that the Contraction Mapping Theorem doesn't just tell us that a fixed point exists, the proof gives us a recipe for approximating this point. That is, if we write $f^1 = f$ and, for $k \geq 2$, $f^k = f \circ f^{k-1}$ for the **iterates** of f, then, given any starting point $x_0 \in X$, $f^k(x_0)$ converges to the fixed point x_F.

Example 2.11.3. Sometimes you need to choose X carefully to ensure that you have a contraction. Take the problem of finding a root of $p(z) = z^7 + z^3 + 2z^2 - 8z + 3i$, or equivalently, of finding a fixed point of

$$f(z) = \frac{z^7 + z^3 + 2z^2 + 3i}{8}$$

which we considered above. Let $X = \{z : |z| \leq r\}$. Then $(X, |\cdot|)$ is a complete metric space. We need to choose r so that

1. f maps X back into X, that is $|f(z)| \leq r$ if $|z| \leq r$;
2. f is a contraction on X, that is, there exists $c < 1$ so that if $z, w \in X$ then $|f(z) - f(w)| \leq c|z - w|$.

If $z \in X$ then

$$|f(z)| \leq \frac{1}{8}\left(|z|^7 + |z|^3 + 2|z|^2 + 3\right). \tag{2.11.1}$$

Note that for any k

$$z^k - w^k = (z - w)(z^{k-1} + z^{k-2}w + \cdots + zw^{k-2} + w^{k-1}$$

and so if $z, w \in X$

$$|z^k - w^k| \leq |z - w|kr^{k-1}.$$

If $z, w \in X$ then

$$d(f(z), f(w)) = |f(z) - f(w)| = \frac{1}{8}\left|(z^7 - w^7) + (z^3 - w^3) + 2(z^2 - w^2)\right|$$
$$\leq \frac{|z-w|}{8}\left(7r^6 + 3r^2 + 4r\right) \tag{2.11.2}$$

Solving for the 'optimal' value of r here is difficult, but for any given value of r it is easy use these bounds to verify whether (1) and (2) are satisfied.

Let's try $r = \frac{1}{2}$. If $|z| \leq \frac{1}{2}$ then by (2.11.1)

$$|f(z)| \leq \frac{1}{8} \cdot \frac{465}{128} = \frac{465}{1024} < \frac{1}{2}$$

so (1) is satisfied. If $|z|, |w| < \frac{1}{2}$ then by (2.11.2)

$$|f(z) - f(w)| \leq \frac{|z-w|}{8} \cdot \frac{183}{64} = \frac{183}{512}|z - w|$$

and so (2) is satisfied too. Thus we know that there is a unique fixed point for f inside $|z| \leq \frac{1}{2}$, and hence a unique root for p there.

If you make r smaller then you'll get faster convergence to the root, but you can't choose r too small or else f won't map the disk back into itself. If you make r larger, then you will get a larger region in which you know that there are no other solutions. Again, clearly you can't make r too large as this might cause either (1) or (2) to fail.

Often one has the problem of finding a fixed point of a function $f : X \to X$ where X does not come with any specified metric. Iterating f will produce a sequence $\{f^k(x_0)\}$, but you would need a metric to really make sense of whether this sequence is converging! It is a somewhat remarkable result due to Polish mathematician Czesław Bessaga in 1959 that if f, f^2, f^3,... each has a unique fixed point then, for any $c \in (0, 1)$ there exists a complete metric d on X such that f is contractive, and c is the contraction constant.

2.12 An application to Ordinary Differential Equations

Does the initial value problem

$$\begin{cases} y' = x \sin y + x^2(y + 3) \\ y(0) = 0 \end{cases}$$

have a unique solution?

It turns out that we can answer this by turning the problem into a fixed point problem in a suitable metric space and then applying the contraction mapping theorem. Importantly, the use of a contraction mapping gives an algorithm for approximating the solution by starting at any initial function and iterating the contraction mapping.

To simplify the notation, let $g(x,y) = x\sin y + x^2(y+3)$, so that the ODE is $y' = g(x,y)$. For suitable differentiable functions $y : (-a,a) \to \mathbb{R}$ define the operator T by

$$(Ty)(x) = \int_0^x g(t,y(t))\,dt.$$

Suppose that $Ty = y$, that is

$$\int_0^x g(t,y(t))\,dt = y(x), \qquad x \in (-a,a).$$

Differentiating gives the original ODE, with the appropriate initial value, so this is the mapping that we want a fixed point for. The problem now is to concoct a suitable metric space on which it is a contraction mapping.

More generally, consider the general initial value problem

$$(1) = \begin{cases} y' = g(x,y) \\ y(x_0) = y_0 \end{cases}$$

What we want is at least a small interval around x_0 on which a unique solution exists. That is, we'd like to have a small rectangle $G = (a,b)\times(c,d)$ containing (x_0,y_0) so that there is a unique curve in G which is the graph of a solution of (1). To construct our contraction mapping we need to impose some conditions on g:

(A) g is continuous and bounded on G, with $|g(x,y)| \le K$ for $(x,y) \in G$.

(B) g is Lipschitz in the y-direction. That is, there exists M such that for all $x \in (a,b)$ and all $y_1, y_2 \in (c,d)$

$$|g(x,y_1) - g(x,y_2)| \le M\,|y_1 - y_2|.$$

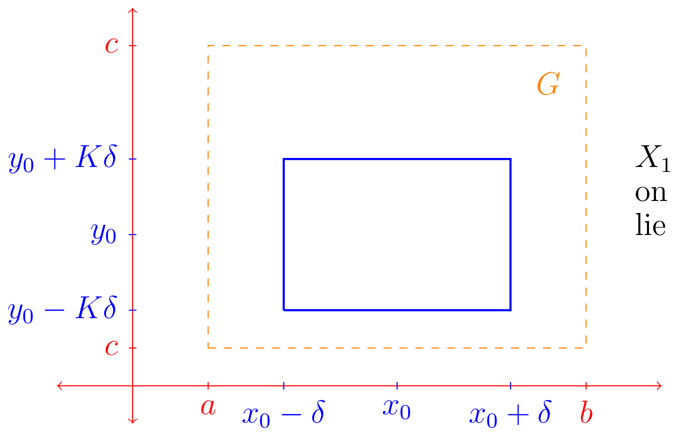
Theorem 2.12.1 (Picard)**.** *Let $G = (a,b)\times(c,d) \subseteq \mathbb{R}^2$ and suppose that $g : G \to \mathbb{R}$ satisfies (A) and (B) above. Then the ODE (1) above has a unique solution in some neighbourhood of $(x_0,y_0) \in G$.*

Proof. Choose $\delta > 0$ such that

- $a < x_0 - \delta < x_0 + \delta < b$,
- $c < y_0 - K\delta < y_0 + K\delta < d$,
- $M\delta < 1$.

Let $I = [x_0-\delta, x_0+\delta]$. Let X be the complete space $C(I)$ under the d_∞ metric. The metric space we'll work on is

$$X_1 = \{y \in X \,:\, y_0 - K\delta \le y(x) \le y_0 + K\delta\,,\ \forall x \in I\}.$$



X_1 is the space of continuous functions on $I = [x_0 - \delta, x_0 + \delta]$ whose graphs lie entirely in the blue rectangle.

Exercise: Show that this set is closed and hence (by Proposition 2.10.12) that X_1 is a complete metric space under the d_∞ metric.

Define the operator T on X_1 by

$$(Ty)(x) = y_0 + \int_{x_0}^{x} g(t, y(t))\, dt. \qquad x \in I.$$

Our aim now is to show that T maps X_1 back into X_1 and that T is a contraction mapping on this space. First, if $y \in X_1$ then, for any $x \in I$,

$$\begin{aligned} |(Ty)(x) - y_0| &= \Big| \int_{x_0}^{x} g(t, y(t))\, dt \Big| \\ &\le \int_{x_0}^{x} |g(t, y(t))|\, dt \\ &\le |x - x_0| K \le K\delta \end{aligned}$$

and hence $y_0 - K\delta \le y(x) \le y_0 + K\delta$. That is $Ty \in X_1$.

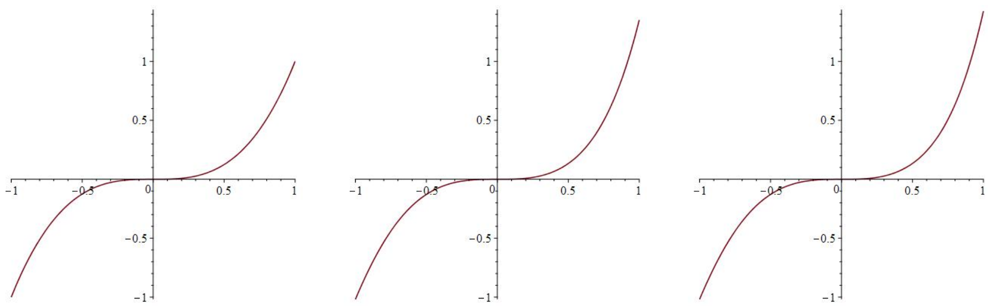
Furthermore, if $y_1, y_2 \in X_1$ then we have

$$\begin{aligned} d_\infty(T(y_1), T(y_2)) &= \sup_{x \in I} |(Ty_1)(x) - (Ty_2)(x)| \\ &= \sup_{x \in I} \Big| \int_{x_0}^{x} g(t, y_1(t)) - g(t, y_2(t))\, dt \Big| \\ &\le \sup_{x \in I} \int_{x_0}^{x} |g(t, y_1(t)) - g(t, y_2(t))|\, dt \\ &\le \sup_{x \in I} \int_{x_0}^{x} M|y_1(t) - y_2(t)|\, dt \qquad \text{(From (B))} \\ &\le (x - x_0) M \sup_{x \in I} |y_1(t) - y_2(t)| \\ &\le (M\delta)\, d_\infty(y_1(t), y_2(t)) \end{aligned}$$

As $M\delta < 1$ this means that T is a contraction mapping on X_1.

The Contraction Mapping Theorem now tells us that there is a unique function y_F in X_1 such that $Ty_F = y_F$ and hence a unique solution in X_1 to the IVP (1). ■

To use Picard's Theorem, one can start with $y_0(x) = y_0$ on $(x_0 - \delta, x_0 + \delta)$, and then finding the functions $y_n = T^n y_{n-1}$. In practice, you won't be able to give an explicit for y_n. Rather you will end up with a list of values at some discrete set of x values, which will be sufficient to do numerical integration. I did this with the problem at the start of this section, working on the interval $(-1, 1)$. The graphs of y_1, y_2, y_3 are below. For $n > 3$, the graphs appear to have stabilized, showing that this method does converge rather quickly.



2.13 Baire Category Theorem

One final general theorem about metric spaces is the Baire Category Theorem. This will be mainly important next year in Functional Analysis when you prove 'Big Theorems' like the Open Mapping Theorem and the Closed Graph Theorem.

Definition 2.13.1. Let (X, d) be a metric space and suppose that $Y \subseteq X$. We say that Y is nowhere dense if the complement of its closure, $X \setminus \text{cl}(Y)$ is dense.

Exercise 2.13.2. Prove that Y is nowhere dense if and only if $\text{cl}(Y)$ doesn't contain any open balls.

Example 2.13.3. (i) The integers are nowhere dense in $\mathbb{R}$.

(ii) $Y = (0, 1)$ is not dense in $\mathbb{R}$, but neither is it nowhere dense. The complement of the closure is $(-\infty, 0) \cup (1, \infty)$ which is not dense, or note that since $\text{cl}(Y)$ contains many open balls.

(iii) Indeed, a nowhere dense set cannot contain any nonempty open subset.

One aspect of the Baire Category Theorem is that it says that open dense sets need to be quite big. Note that in $\mathbb{R}$, an open set can always be written as a countable union of open intervals. Thus an open dense set must be just $\mathbb{R}$ with a countable number of elements removed. If you intersect countably many such sets you are again left with $\mathbb{R}$ less a countable number of points. The intersection of course might no longer be open!

Exercise 2.13.4. Give an example of a sequence $\{U_n\}_{n=1}^{\infty}$ of open dense subsets of $\mathbb{R}$ such that $\bigcap_{n=1}^{\infty} U_n$ is the irrational numbers.

The next theorem generalizes this.

Theorem 2.13.5. *Let (X, d) be a complete metric space, and suppose that $\{U_n\}_{n=1}^{\infty}$ is a countable collection of open dense subsets of X. Then $\bigcap_{n=1}^{\infty} U_n$ is not empty.*

Proof. First recall that the intersection of a dense set with any open set must be nonempty.

Suppose that $x_1 \in U_1$. As U_1 is open there exists a ball $B_1 = B(x_1, r_1) \subseteq U_1$. Now U_2 is dense, so there exists a point $x_2 \in B(x_1, r_1) \cap U_2$. Now $B(x_1, r_1) \cap U_2$ is open so there exists $B_2 = B(x_2, r_2) \subseteq B(x_1, r_1) \cap U_2$. Indeed for r_2 small enough $\mathrm{cl}(B_2) \subseteq B(x_1, r_1) \cap U_2$. Proceeding inductively we can produce a sequence of points and balls with

$$B_k = B(x_k, r_k) \subseteq \mathrm{cl}(B_k) \subseteq B_{k-1} \cap U_k.$$

Again, we can decide at each step to make sure that $r_k \leq r_{k-1}/2$ and hence that $r_k \to 0$. Note that

$$\mathrm{cl}(B_1) \supseteq \mathrm{cl}(B_2) \supseteq \mathrm{cl}(B_3) \supseteq \ldots$$

If $k, \ell \geq K$, then $x_k \in B_k \subseteq B_K$ and $x_\ell \in B_\ell \subseteq B_K$ and so $d(x_k, x_\ell) < 2r_K$. This ensures that $\{x_k\}$ is Cauchy and hence converges to some limit $x \in X$. The nesting of the balls means that for any K, $x_k \in \mathrm{cl}(B_K)$ for all $k \geq K$ and hence that $x = \lim_k x_k \in \mathrm{cl}(B_K)$, Hence

$$x \in \bigcap_K \mathrm{cl}(B_K) \subseteq \bigcap_k U_k.$$

The nesting of the balls means that the sequence $\{x_k\}$ is Cauchy and hence converges. ■

Theorem 2.13.6 (Baire Category Theorem)**.** *A complete metric space can not be decomposed as a countable union of nowhere dense sets.*

Proof. Let $\{Y_n\}_{n=1}^{\infty}$ is a countable collection of nowhere dense subsets of X. Then $U_n = X \setminus \mathrm{cl}(Y_n)$ forms a countable collection of open dense subsets, so there is an element of $\bigcap_{n=1}^{\infty} U_n$. But this is equivalent to saying that $x \notin \bigcup_{n=1}^{\infty} Y_n$. ■

Thus, for example, if $\mathbb{R} = \bigcup_{n=1}^{\infty} A_n$ then at least one of the sets A_n must 'bunch up' somewhere, in the sense that the closure of A_n must contain an open interval.

The power of the Baire Category Theorem is that it sometimes lets us prove uniform bounds from just knowing pointwise bounds.

Theorem 2.13.7. *Suppose that we have a (big) family $\mathcal{F}$ of continuous real-valued functions defined on a complete metric space X. Suppose that for each $x \in X$ the set*

$$V_x = \{f(x) : f \in \mathcal{F}\}$$

is bounded in $\mathbb{R}$, by say M_x. Then there is a nonempty open set $U \subseteq X$ and a constant M such that $|f(x)| \leq M$ for all $x \in U$ and $f \in \mathcal{F}$.

Proof. For $m = 1, 2, 3, \ldots$ and $f \in \mathcal{F}$, define

$$Y_{m,f} = \{x \in X : |f(x)| \leq m\}.$$

Let $Y_m = \bigcap_{f \in \mathcal{F}} Y_{m,f}$. The map $x \mapsto |f(x)|$ is continuous and so we can consider each set $Y_{m,f}$ as the inverse image under a continuous map of the closed set $[0, m]$, and hence deduce that each of these sets is closed.

The hypothesis on the boundedness of the sets V_x just says that $x \in Y_m$ for all $m \geq M_x$. In particular, each $x \in X$ is in at least Y_m and hence $\bigcap_{m=1}^{\infty} Y_m$ contains all elements of X.

By the Baire Category Theorem, the sets Y_m can't all be nowhere dense. Choose one, say Y_M which is not nowhere dense. By Exercise 2.13.2 $Y_M = \text{cl}(Y_M)$ must contain an open ball U. Thus, for all $x \in U$, $|f(x)| \leq M$ for all $x \in U$ and $f \in \mathcal{F}$. ■

2.14 Appendix: Classical Banach Spaces

Sequence spaces

The obvious infinite-dimensional version of $\mathbb{R}^n$ would be $\mathbb{R}^\infty$, the space of all real sequences. A typical element would look like

$$\mathbf{x} = (x_j)_{j=1}^\infty = (1, 1, 2, \tfrac{1}{2}, 3, \tfrac{1}{3}, 4, \tfrac{1}{4}, \dots).$$

The problem with $\mathbb{R}^\infty$ is that the natural Euclidean norm or modulus from $\mathbb{R}^n$,

$$\|(x_1, \dots, x_n)\|_2 = \left(\sum_{j=1}^n |x_j|^2\right)^{1/2}$$

doesn't always make sense in the infinite case (in particular for the vector $\mathbf{x} \in \mathbb{R}^\infty$ given above!). To do analysis we need to cut down to restricted sets of sequences on which particular norms do make sense. For example, we often work in

$$\ell^2 = \left\{\mathbf{x} \in \mathbb{R}^\infty \,:\, \|\mathbf{x}\|_2 = \Big(\sum_{j=1}^\infty |x_j|^2\Big)^{1/2} < \infty\right\}.$$

This space has many of the nice properties of $\mathbb{R}^n$. For example it is an inner product space where the inner product is given by the infinite dot product

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \cdot \mathbf{y} = \sum_{j=1}^\infty x_j y_j.$$

Importantly, but not obviously, the space $(\ell^2, \|\cdot\|_2)$ is complete and so this is a Hilbert space.

A sequence of elements in ℓ^2 is therefore a sequence $\{\mathbf{x}_k\}_{k=1}^\infty$ of sequences:

$$\begin{aligned} \mathbf{x}_1 &= (x_{11}, x_{12}, x_{13}, \dots) \\ \mathbf{x}_2 &= (x_{21}, x_{22}, x_{23}, \dots) \\ \vdots \quad & \quad \vdots \end{aligned}$$

It is actually easy to check that if $\mathbf{x}_k \to \mathbf{x}$ in ℓ^2 norm, then the sequence certainly converges 'elementwise'. That is, the first coordinates of the $\mathbf{x}_k$'s converge to the first coordinate of $\mathbf{x}$, and similarly for each of the other coordinates. This means that identifying the potential limit $\mathbf{x}$ is usually easy. You then have to check that $\|\mathbf{x}_k - \mathbf{x}\|_2 \to 0$ as $k \to \infty$.

Example 2.14.1. For $k = 1, 2, 3, \dots$ let

$$\mathbf{x}_k = \left(\frac{1}{k}, \dots, \frac{1}{k}, 0, 0, \dots\right)$$

where there are k nonzero entries. Noting that in each jth position, the entries go to zero, we let $\mathbf{x} = (0, 0, 0, \dots)$. Then

$$\|\mathbf{x}_k - \mathbf{x}\|_2 = \|\mathbf{x}_k\|_2 = \left(\sum_{j=1}^{k} \frac{1}{k^2}\right)^{1/2} = \frac{1}{\sqrt{k}} \to 0$$

and so $\mathbf{x}_k \to \mathbf{x}$ in $(\ell^2, \|\cdot\|_2)$.

The Hilbert space $(\ell^2, \|\cdot\|_2)$ is just one of a large number of sequence spaces. The most important class of these are the ℓ^p spaces. If $1 \leq p < \infty$ then

$$\ell^p = \Big\{\mathbf{x} \in \mathbb{R}^\infty \,:\, \|\mathbf{x}\|_p = \Big(\sum_{j=1}^{\infty} |x_j|^p\Big)^{1/p} < \infty\Big\}.$$

It requires quite a bit of work to check that $\|\cdot\|_p$ actually is a norm, and that $(\ell^p, \|\cdot\|_p)$ is a complete space. That $\|\cdot\|_p$ satisfies the triangle inequality is known as Minkowski's Theorem and we prove this in MATH2701. In any case, each of the spaces $(\ell^p, \|\cdot\|_p)$ is a Banach space.

If $0 < p < 1$ then $\|\cdot\|_p$ fails the triangle inequality and so is not a norm. These spaces still turn out to be quite useful however, and give an interesting examples of what are known as Fréchet spaces.

Another important sequence norm is

$$\|\mathbf{x}\|_\infty = \sup_j |x_j|.$$

This norm is usually applied to the spaces

$$\begin{aligned}
\ell^\infty &= \{\mathbf{x} \in \mathbb{R}^\infty \,:\, \|\mathbf{x}\|_\infty < \infty\}, \\
c &= \{\mathbf{x} \in \mathbb{R}^\infty \,:\, \lim_{j\to\infty} x_j \text{ exists}\}, \\
c_0 &= \{\mathbf{x} \in \mathbb{R}^\infty \,:\, \lim_{j\to\infty} x_j = 0\}, \\
c_{00} &= \{\mathbf{x} \in \mathbb{R}^\infty \,:\, \exists n \text{ such that } x_j = 0 \text{ for all } j > n\}.
\end{aligned}$$

Here it is rather easier to check that $\|\cdot\|_\infty$ is a norm and that the spaces $(\ell^\infty, \|\cdot\|_\infty)$ and $(c_0, \|\cdot\|_\infty)$ are complete and hence are Banach spaces. The space $(c_{00}, \|\cdot\|_\infty)$ is not complete, and indeed it is dense in c_0.

Note that if $1 < p < q < \infty$ then

$$c_{00} \subsetneq \ell^1 \subsetneq \ell^p \subsetneq \ell^q \subsetneq c_0 \subsetneq c \subsetneq \ell^\infty$$

and that

$$\|\mathbf{x}\|_1 \geq \|\mathbf{x}\|_p \geq \|\mathbf{x}\|_q \geq \|\mathbf{x}\|_\infty$$

where this makes sense. Indeed $\|\mathbf{x}\|_\infty = \lim_{p\to\infty} \|\mathbf{x}\|_p$.

Example 2.14.2. (i) $(1, \frac{1}{2}, \frac{1}{3}, \dots)$ is in ℓ^2 but not in ℓ^1.

(ii) $(1, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{3}}, \dots)$ is in ℓ^3 (or even ℓ^π) but not in ℓ^2.

(iii) $\left(\frac{1}{\ln 2}, \frac{1}{\ln 3}, \frac{1}{\ln 4}, \dots\right)$ is in c_0 but not in any ℓ^p spaces with $p < \infty$.

(iv) $(0, 1, 0, 1, 0, 1, \dots)$ is in ℓ^∞ but not in c.

Because of the above nesting, if $1 \le p < q \le \infty$ then $(\ell^p, \|\cdot\|_q)$ is a normed space, but it isn't complete. In fact ℓ^p is dense in $(\ell^q, \|\cdot\|_q)$ (unless $q = \infty$).

The above spaces are called the **classical sequence spaces**. There are many other natural norms and sequence spaces however. An example is

$$\|\mathbf{x}\|_{bv} = |x_1| + \sum_{j=1}^{\infty} |x_{j+1} - x_j|, \qquad bv = \{\mathbf{x} \in \mathbb{R}^\infty \,:\, \|\mathbf{x}\|_{bv} < \infty\}.$$

Exercise 2.14.3. (i) Show that c is not dense in $(\ell^\infty, \|\cdot\|_\infty)$ by showing that there is no element $\mathbf{y} \in c$ which is close to $\mathbf{x} = (0, 1, 0, 1, 0, 1, \dots) \in \ell^\infty$.

(ii) Show that c_0 is not dense in $(c, \|\cdot\|_\infty)$ by showing that there is no element $\mathbf{y} \in c_0$ which is close to $\mathbf{x} = (1, 1, 1, \dots) \in c$.

A concept that you may come across is that of separability. A metric space is said to be **separable** if there is a countable dense set in the space. Thus $\mathbb{R}$ is separable (in the usual metric) as $\mathbb{Q}$ is a countable dense subset.

For $k = 1, 2, 3, \dots$, let $\mathbf{e}_k$ be the sequence with a 1 in the kth position and zeros elsewhere. It is not too hard to show that the set of sequences which can be written as finite rational combinations of these 'basis vectors',

$$\mathbf{x} = \sum_{k=1}^{K} q_k \mathbf{e}_k, \qquad q_k \in \mathbb{Q},$$

is dense in c_0, and in ℓ^p for $1 \le p < \infty$, and hence these spaces are separable. This set is not dense in ℓ^∞. It is a little harder is to show that no countable dense subset of ℓ^∞ exists. So ℓ^∞ is really rather larger than the other spaces. One consequence of this will be that there is no hope of representing all linear transformations on ℓ^∞ as infinite matrices. We'll come back to this at the end of the course.

Function spaces

The classical Banach function spaces are the spaces $C[0,1]$ and $L^p[0,1]$ for $1 \le p \le \infty$. We have seen the space $C[0,1]$ quite a lot. The standard norm is

$$\|f\|_\infty = \sup_{t \in [0,1]} |f(t)|$$

and we showed that $(C[0,1], \|\cdot\|_\infty)$ is complete.

The L^p spaces are more complicated. The norm is easy enough to define. If $f : [0,1] \to \mathbb{R}$ then

$$\|f\|_p = \left(\int_0^1 |f(t)|^p \, dt \right)^{1/p}.$$

The problem is to work out a suitable vector space of functions to use in order that this gives a complete normed space. We'll come back to this later. For the moment just take $L^p[0,1]$ to be roughly all the functions for which $\|f\|_p$ is finite. The space $L^\infty[0,1]$ will contain all decent bounded functions. It will turn out that $L^2[0,1]$ is a Hilbert space and that $L^p[0,1]$ is a Banach space for $1 \le p \le \infty$.

What we will see is that if $1 < p < q < \infty$ then

$$C[0,1] \subsetneq L^\infty[0,1] \subsetneq L^q[0,1] \subsetneq L^p[0,1] \subsetneq L^1[0,1].$$

Note that this nesting goes the other way to the ℓ^p spaces! The norms have corresponding inequalities

$$\|f\|_1 \le \|f\|_p \le \|f\|_q \le \|f\|_\infty .$$

The problem alluded to above is that if you use the p-norm on $C[0,1]$ then the space is not complete. If you add too many functions in to fill in the holes, then you lose the property that $\|f - g\|_p = 0 \iff f = g$. The elements of L^p will turn out to be equivalence classes of functions rather than functions themselves. Most analysts will work with these spaces as if the elements actually are functions, but novices have to take care not to trip up on some of the issues this causes! We'll look at this more closely in Chapter 6.

You can define these spaces for functions defined on more general sets, so you get $C[a,b]$ or $L^2(\mathbb{R})$ or $L^1(\mathbb{R}^3)$. The nesting and inequality facts just listed depend very much on the functions having domain $[0,1]$. For a bounded interval $[a,b]$ you still get nesting but you need some factors in the inequalities. For unbounded domains, nothing works!

It is not obvious, but $L^2[0,1]$ is isometrically isomorphic (as a Hilbert space) to ℓ^2. For $p \ne 2$, this nice relationship does not hold! On the other hand, the Anderson–Kadec Theorem says that there is always a continuous bijection, with continuous inverse[15] between any two separable infinite dimensional Banach spaces — you just can't in general make this map linear! Note that as in the sequence space case, $L^p[0,1]$ is separable unless $p = \infty$.

Example 2.14.4. (i) $f(t) = t^{-1/2}$ is in $L^1[0,1]$ but not in $L^2[0,1]$.

(ii) $f(t) = t^{-1/4}$ is in $L^2[0,1]$ but not in $L^\infty[0,1]$.

Inside $C[0,1]$ there are lots of subspaces like $C^1[0,1]$. This is the space of functions $f : [0,1] \to \mathbb{R}$ such that $f' \in C[0,1]$. (There is a small issue here as to what you mean by $f'(0)$ and $f'(1)$. More formally, we want f to be differentiable on $(0,1)$ and for f' to have a finite right limit at 0 and a finite left limit at 1). Thus $f(t) = |t - \frac{1}{2}|$ or $f(t) = \sqrt{t}$ are in $C[0,1]$ but not in $C^1[0,1]$. You can extend this further to look at C^2

[15] In Chapter 4 we shall call such a map a homeomorphism.

where you want f, f' and f'' to all be in $C[0,1]$. Or more generally to $C^k[0,1]$ with $k \geq 1$. The spaces $C^k[0,1]$ all contain the polynomials $\mathcal{P}$ so we have

$$\mathcal{P} \subsetneq \cdots \subsetneq C^2[0,1] \subsetneq C^1[0,1] \subsetneq C[0,1].$$

Later we will prove that $\mathcal{P}$ is dense in $(C[0,1], \|\cdot\|_\infty)$, so all the $C^k[0,1]$ spaces must be dense too.

Sometimes it is useful to turn $C^k[0,1]$ into a Banach space by giving it a norm which makes it complete. The usual one to use is

$$\|f\|_{C^k} = \|f\|_\infty + \|f'\|_\infty + \cdots + \left\|f^{(k)}\right\|_\infty .$$

Hint: Many students think of elements of these function spaces in terms of some formula for $f(t)$. It is usually better to think of the graph of f. That is, if you want to think about a sequence of functions which have certain behaviour, first draw pictures of what the graphs need to look like, and worry about a formula for the functions later. In fact, in many cases, the graphs are much more convincing that the formulas. Conversely, if someone has defined a sequence of functions using formulas for $f_k(t)$, the best thing is to try to work out what the graphs of the first few look like.

Exercise 2.14.5. For $k = 1, 2, \ldots$ let

$$f_k(t) = \exp\left(k^3(t - 1/k)^2\right), \qquad t \in [0,1].$$

Is $\{f_k\}_{k=1}^\infty$ a bounded sequence in $(C[0,1], \|\cdot\|_\infty)$? If $\{f_k\}_{k=1}^\infty$ a bounded sequence in $(C^1[0,1], \|\cdot\|_{C^1})$? Does the sequence converge in $(C[0,1], \|\cdot\|_\infty)$?

2.15 Problems

1. Let M be a set. Prove that if $A, B \subseteq M$, then either $|A| \leq |B|$ or $|B| \leq |A|$ (or both).

2. Show that $d_1(f,g) = \displaystyle\int_0^1 |f(t) - g(t)|\, dt$ and $d_2(f,g) = \left(\int_0^1 |f(t) - g(t)|^2\, dt\right)^{1/2}$ are both metrics on the set $C[0,1]$.

3. Consider the metrics on the space $X = \mathbb{R}^n$:

$$\begin{aligned} d_p(\mathbf{x}, \mathbf{y}) &= \Big(\sum_{j=1}^n |x_j - y_j|^p\Big)^{1/p}, \quad (1 \leq p < \infty), \\ d_\infty(\mathbf{x}, \mathbf{y}) &= \max_{1 \leq j \leq n} |x_j - y_j|. \end{aligned}$$

Prove that these metrics are all **equivalent**. In other words, show that if $1 \leq p, q \leq \infty$, then given any sequence $\{\mathbf{x}_k\}_{k=1}^\infty \subseteq \mathbb{R}^n$ and any $\mathbf{x} \in \mathbb{R}^n$,

$$d_p(\mathbf{x}_k, \mathbf{x}) \to 0 \quad \Longleftrightarrow \quad d_q(\mathbf{x}_k, \mathbf{x}) \to 0.$$

4. Which of the following are metrics on $\mathbb{R}^n$?
   (a) $d(\mathbf{x},\mathbf{y}) = \Big| \, |\sum_{j=1}^n x_j| - |\sum_{j=1}^n y_j| \, \Big|$.
   (b) $d(\mathbf{x},\mathbf{y}) = \sum_{j=1}^n \sqrt{|x_j - y_j|}$.
   (c) $d(\mathbf{x},\mathbf{y}) = \sum_{j=1}^n (x_j - y_j)^2$.
5. Let $X = \mathbb{R}^2$. Fix positive integers n, m and define
$$d(\mathbf{x},\mathbf{y}) = \left(|x_1 - y_1|^n + |x_2 - y_2|^m\right)^{\frac{1}{n+m}}.$$
   Is d a metric on X? Does it depend on the values of n and m?
6. Find three interesting examples of metric spaces that were not presented in class.
7. Prove that if $f_k \to f$ in $(C[0,1], d_\infty)$ then $f_k \to f$ in $(C[0,1], d_1)$. (That the converse fails will be shown in lectures.)
8. Prove that convergence under the discrete metric is very boring. Prove that every subset of a discrete metric space is open.
9. Are there any metric spaces in which there are exactly 5 sets which are both open and closed? What about 6 open and closed sets?
10. Let (X, d) be a metric space.
    (a) Prove that every ball of the form $B(x,\epsilon) = \{y \in X : d(x,y) < \epsilon\}$ is open.
    (b) Is the set $\overline{B}(x,\epsilon) = \{y \in X : d(x,y) \leq \epsilon\}$ ever open?
11. Let $X = C[0,1]$ under the metric d_∞. Let
$$Y = \{f \in C[0,1] : f(0) \neq 0\}.$$
    (a) Prove that Y is open in $(C[0,1], d_\infty)$.
    (b) Is Y open in $(C[0,1], d_1)$?
12. Prove that every open set in $\mathbb{R}$ can be written as a **countable** union of open intervals.
13. Let $X = \mathbb{R}^2$ with the usual metric. For each of the following sets, identify the interior points, the boundary points, the limits points, and whether the set is open, closed or neither.
    (a) $Y_1 = \{(x,y) \in \mathbb{R}^2 : \sin x < y < \cos x\}$,
    (b) $Y_2 = \mathbb{Q} \times \mathbb{R}$,
    (c) $Y_3 = \left\{\left(\cos \frac{1}{t}, t\right) : t > 0\right\}$,

(d) $Y_4 = \bigcup_{k=1}^{\infty} B\big((\frac{1}{k}, 0), \frac{1}{3k^2}\big)$.

(Recall that $B(x, r)$ is the ball centred at x with radius r.)

14. Let $Y = \left\{\frac{1}{k} \,:\, k = 1, 2, 3, \ldots\right\}$. Let $d(x, y) = |x - y|$. For each of the following metric spaces (X_j, d), identify the interior points, the boundary points and the limits points of Y as a subset of (X_j, d) and decide whether Y is open, closed or neither in (X_j, d).

    (a) $X_1 = \mathbb{R}$,

    (b) $X_2 = (0, \infty)$,

    (c) $X_3 = \left\{\frac{n}{k} \,:\, 1 \le n \le k,\ k = 1, 2, 3, \ldots\right\}$.

15. Prove that a finite union of closed sets is closed. Prove that any intersection of closed sets is closed.

16. Give examples of a metric space (X, d) and an infinite family of closed sets $\{E_j\}_{j=1}^{\infty}$ in X such that

    (a) $\cup_{j=1}^{\infty} E_j$ is not closed,

    (b) $\cup_{j=1}^{\infty} E_j$ is closed

    (c) $\cup_{j=1}^{\infty} E_j$ is open and closed.

17. Let $X = [0, 1]$ with the usual metric. Let Z be the set of all closed subsets of $[0, 1]$. For $U, V \in Z$ let $\rho(U, V) = \inf\{|u - v| \,:\, u \in U,\ v \in V\}$. Is it possible to have $\rho(U, V) = 0$ with U and V disjoint? What happens if you replace $[0, 1]$ by $(0, 1)$? (If you don't understand inf, see over!)

18. (a) Suppose that $f : \mathbb{R} \to \mathbb{Z}$ (with their usual metrics). Can f be continuous?

    (b) Suppose that $f : \mathbb{Z} \to \mathbb{R}$. Can f be continuous? Must f be continuous?

    (c) If you look back you'll notice that our definition of continuity at a point is not quite the one you used in first year — that $\lim\limits_{x \to x_0} f(x)$ exists and equals $f(x_0)$. Why don't we use that definition in general metric spaces?

19. Let (X, d) and (X', d') be metric spaces and fix $x_0 \in X$. Prove that if $f : X \to X'$ is continuous at x_0 then f is sequentially continuous at x_0.

20. Let (X, d) be a metric space and fix $x_0 \in X$. Define $g : X \to \mathbb{R}$ by $g(x) = d(x, x_0)$. Must g be continuous on X?

21. (a) Suppose that $f : (X, d) \to (X', d')$. Prove that f is continuous on X if and only if $f^{-1}(B)$ is open in X for every open **ball** $B \subseteq X'$.
    (Is it true that $f^{-1}(B)$ is always an open ball?)

(b) Suppose that $f : (X, d) \to (X', d')$. If f is continuous, must it map open sets to open sets? If f does map open sets to open sets must f be continuous?

22. Prove that if a sequence in a metric space converges then it is Cauchy.

23. Let (X, d) be a complete metric space. Show that $Y \subseteq X$ is closed in X if and only if (Y, d) is complete.

24. Let $X = C[-1, 1]$. Suppose that $\{f_k\}_{k=1}^{\infty} \subseteq C[-1, 1]$ and $f \in C[-1, 1]$. In the next chapter we'll define that

    - $f_k \to f$ **pointwise** if, for all $t \in [-1, 1]$, $f_k(t) \to f(t)$ (in $\mathbb{R}$);
    - $f_k \to f$ **uniformly** if $d_\infty(f_k, f) \to 0$;
    - $f_k \to f$ **in d_1 metric** if $d_1(f_k, f) = \int_{-1}^{1} |f_k(t) - f(t)|\, dt \to 0$.

    (a) Let $f_k(t) = \dfrac{1 - kt^2}{1 + kt^2}$, $t \in [-1, 1]$. For each of the above three senses of convergence, determine whether $\{f_k\}$ has a limit. (Try using Maple to plot the curves.)

    (b) Repeat this with the sequence $g_k(t) = \dfrac{kt^3}{1 + kt^2}$.

25. Let (X, d) be a metric space and suppose that $Y \subseteq X$.

    (a) Prove that $\mathrm{cl}(Y) \cap \mathrm{cl}(X \setminus Y)$ equals the boundary of Y.

    (b) Prove that if $V \subseteq X$ is closed and $Y \subseteq V$, then $\mathrm{cl}(Y) \subseteq V$.

    (c) Prove that $\mathrm{cl}(Y) = \cap\{V : Y \subseteq V \text{ and } V \text{ is closed}\}$. (This is two lines!)

26. Let (X, d) be a metric space. A point $x \in X$ is **isolated** if there exists $\epsilon > 0$ such that $B(x, \epsilon) = \{x\}$.

    (a) Prove that $\{x\}$ is nowhere dense iff x is not isolated.

    (b) Prove that a complete metric space without isolated points must be uncountable.

    (c) Give an example of a countably infinite complete metric space.

27. Suppose that $a < b$, $c \leq d$ are real numbers. Let $X = C[a, b]$ with the usual supremum metric d_∞. Let

$$X_1 = \{f \in X : f(t) \in [c, d]\}.$$

    Prove that (X_1, d_∞) is a complete metric space.

28. Let $p, q : \mathbb{C} \to \mathbb{C}$ be defined by

$$\begin{aligned} p(z) &= z^7 + z^3 - 9z - i, \\ q(z) &= \frac{z^7 + z^3 - i}{9} \end{aligned}$$

(a) Prove that p has a zero at z_0 if and only if z_0 is a fixed point for q.

(b) Hence or otherwise show that p has at exactly one zero in the closed unit disk $D = \{z \in \mathbb{C} \,:\, |z| \leq 1\}$.

(c) Where are the other zeros?

29. Suppose that f is a contraction mapping on a complete metric space (X, d). It is all very well to know that $x_k = f^k(x_0)$ will converge to the fixed point x_F as $k \to \infty$, but it often just as important to know how good an approximation x_k is to x_F.

(a) Show that if c is the contraction constant of f, that $d(x_k, x_F) \leq c^k \operatorname{diam}(X) = c^k \sup\{d(x, y) \,:\, x, y \in X\}$.

(b) Apply this to Example 2.11.3 to determine how large k needs to be in order that $z_k = f^k(0)$ is within 10^{-10} of the root of p.

(c) Try running this on Maple or Matlab to see how sharp that estimate is.

30. Let $\emptyset \neq S \subseteq \mathbb{C}$. Let $C_S[0, 1]$ denote the set of all continuous functions from $[0, 1]$ to S. Applying the usual supremum norm

$$\|f\|_\infty = \sup_{x \in [0,1]} |f(x)|$$

makes $C_S[0, 1]$ into a metric space. Let

$$X = \{f \in C_S[0, 1] \,:\, f(0) = f(1)\}$$

denote the space of closed parameterized paths in S. Two paths $f, g \in X$ are said to be **homotopic** (written $f \sim g$) if there exists a continuous function $H : [0, 1] \to X$ such that $H(0) = f$ and $H(1) = g$.

(You should think, only slightly inaccurately, of f and g as specifying some closed loop of string lying inside S. These loops are homotopic if you can deform the f loop to the g loop in a continuous way with all of the intermediate loops staying inside S. If you were animating this with time going from 0 to 1, then you should think of $H(t)$ as giving the loop at time t. Note that $H(t)$ is a continuous function, so it makes sense to talk about $H(t)(x)$, which would be some point along the loop at time t.)

(a) Prove that if $f \sim g$ and $g \sim h$ then $f \sim h$.

(b) Let $S = \mathbb{C} \setminus \mathbb{Z}$. Let $f(x) = \pi e^{2\pi i x}$, $g(x) = \frac{\pi}{2} e^{2\pi i x}$ and $h(x) = \pi \cos^3(2\pi x) + 2i \sin^3(2\pi x)$. Prove that $f \sim h$, but that $f \not\sim g$.

31. Consider the initial value problem

$$\begin{cases} \dfrac{dy}{dx} = \exp(xy) \\ y(0) = 1. \end{cases}$$

(a) Verify that this IVP has a unique solution in a neighbourhood of $x = 0$.

(b) Following the notation of Section 2.12, find values of K, M and δ that will work for this case.

32. Let (X, d) be a complete metric space. The **diameter** of a set $A \subseteq X$ is the quantity $\operatorname{diam}(A) = \sup_{x,y \in A} d(x, y)$.

Prove that if $\{A_k\}$ is a sequence of nonempty **closed** subsets of X which satisfy

(a) $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$, and

(b) $\lim_{k\to\infty} \operatorname{diam}(A_k) = 0$,

then $\bigcap_{k=1}^{\infty} A_k$ is nonempty. (What happens if you drop the condition that the sets are closed?)

33. Let $X = M_{2,2}(\mathbb{R})$, the vector space of 2×2 real matrices. For $A \in M_{2,2}(\mathbb{R})$, define

$$\|A\|_{op} = \sup_{\mathbf{x} \in \mathbb{R}^2,\ \mathbf{x} \neq 0} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|},$$

where $\|\cdot\|$ on the right-hand side is just the usual norm on $\mathbb{R}^2$.

(a) Prove that $\|\cdot\|$ defines a norm on X.

(b) Calculate $\left\| \begin{pmatrix} 2 & -3 \\ 5 & 1 \end{pmatrix} \right\|_{op}$.

(c) Prove that for any $A \in M_{2,2}(\mathbb{R})$

$$\|A\|_{op} = \sup\{|\langle A\mathbf{x}, \mathbf{y}\rangle| \ : \ \|\mathbf{x}\| = \|\mathbf{y}\| = 1\}.$$

(d) Generalize! (Does it matter whether we use the usual norm on $\mathbb{R}^2$? Does this work with different size matrices?)

34. Let $(V, \|\cdot\|)$ be a normed vector space. Prove that the norm is a continuous function. That is show that the map $n : V \to \mathbb{R}$, $n(v) = \|v\|$ is continuous.

35. A set $C \subseteq \mathbb{R}^2$ is said to be **convex** if for all $\mathbf{x}, \mathbf{y} \in C$, the line segment joining $\mathbf{x}$ and $\mathbf{y}$,

$$S = \{\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \,:\, 0 \leq \lambda \leq 1\}$$

line entirely inside C. We say that C is **centrally symmetric** if $\mathbf{x} \in C \iff -\mathbf{x} \in C$.

   (a) Let $\|\cdot\|$ be **any** norm on $\mathbb{R}^2$. Let $C = \{\mathbf{x} \in \mathbb{R}^2 \,:\, \|\mathbf{x}\| \leq 1\}$ be the closed unit ball in $(\mathbb{R}, \|\cdot\|)$. Prove that C is closed, convex and centrally symmetric.

   (b) Suppose that C is a closed, bounded, convex and centrally symmetric subset of $\mathbb{R}^2$ and that 0 is an interior point of C. Prove that setting

$$\|\mathbf{x}\|_C = \inf\Big\{\lambda \,:\, \frac{1}{\lambda}\mathbf{x} \in C,\ \lambda > 0\Big\}$$

   defines a norm on $\mathbb{R}^2$.

36. Consider the vector space M_{22} of all 2×2 real matrices. For $A, B \in M_{22}$, define

$$\langle A, B\rangle := \operatorname{tr}(AB^T)$$

where tr is the trace of the matrix and B^T is the trace of B. Is $\langle\cdot,\cdot\rangle$ an inner product on M_{22}?

37. Let $(V, \|\cdot\|)$ be an infinite dimensional normed vector space. A sequence $\{\mathbf{e}_k\}_{k=1}^{\infty}$ of elements on V is called a (Schauder) **basis** for V if, for all $\mathbf{x} \in V$ there exists a unique sequence of scalars $\{a_k\}$ such that

$$\mathbf{x} = \sum_{k=1}^{\infty} a_k \mathbf{e}_k.$$

What that means is that $\|\mathbf{x} - \sum_{k=1}^{N} a_k\mathbf{e}_k\| \to 0$ as $N \to \infty$. The sequence $(a_1, a_2, \dots)$ is called the **coordinate vector** of $\mathbf{x}$ with respect to the basis. Check that the 'obvious standard basis' for ℓ^p is actually a basis if $1 \leq p < \infty$, but not for ℓ^∞.

38. Suppose that $(V, \|\cdot\|)$ be an infinite dimensional inner product vector space and that $\{\mathbf{e}_k\}_{k=1}^{\infty}$ is a Schauder basis for V. Suppose further that $\{\mathbf{e}_k\}$ is orthonormal. That is $\langle \mathbf{e}_i, \mathbf{e}_j\rangle = 0$ if $i \neq j$ (and equals 1 if $i = j$). Suppose that $\mathbf{x}$ has coordinate vector $(a_1, a_2, \dots)$ with respect to this basis. Show that $a_k = \langle \mathbf{x}, \mathbf{e}_k\rangle$ for all k. (Remark: in $\mathbb{R}^n$, finding coordinate vectors is in general quite painful requiring a lot of Gaussian elimination, so having an orthonormal basis saves a lot of work. In infinite dimensions, you can imagine that the Gaussian elimination is a killer!)

39. There are many important function spaces which are not 'classical Banach spaces'. One family which is very important in complex function theory are the **Hardy spaces** on the disk. These are complex spaces of analytic functions on the open unit disk $\mathbb{D} = \{z : |z| < 1\}$. For example $H^\infty(\mathbb{D})$ is the space of bounded analytic functions on $\mathbb{D}$ with $\|f\|_\infty = \sup_{z\in\mathbb{D}} |f(z)|$. The space $H^2(\mathbb{D})$ which consists of all the functions of the form $f(z) = \sum_{k=0}^\infty a_k z^k$ with $(a_k) \in \ell^2(\mathbb{N})$ is a Hilbert space whose norm is conveniently equal to $\|f\|_2 = \left(\sum_{k=0}^\infty |a_k|^2\right)^{1/2}$. (There is a whole scale of spaces $H^p(\mathbb{D})$, $1 \le p \le \infty$, whose definition is unfortunately not what you hope it is!) Prove that the set of (complex) polynomials is dense in $H^2(\mathbb{D})$.

40. A **Banach algebra** is a Banach space $(X, \|\cdot\|)$ which also has a multiplication, and for which one has $\|xy\| \le \|x\| \, \|y\|$ for all $x, y \in X$. (We say in this case that the norm is **submultiplicative**.)

    (a) Check that $(C[0,1], \|\cdot\|_\infty)$, with the usual multiplication of functions, and $(\ell^i nfty, \|\cdot\|_\infty)$, with elementwise multiplication of sequences, are both Banach algebras.

    (b) A more exotic (but important) example is

    $$\ell^1(\mathbb{Z}) = \{x = \{x_k\}_{k=\infty}^\infty : \sum_{k=-\infty}^\infty |x_k| < \infty\}.$$

    Here the multiplication is convolution:

    $$(x * y)_k = \sum_{i=-\infty}^\infty x_i y_{k-i}.$$

    Convince yourself that the norm is submultiplicative for this multiplication.

    (c) Another example is the $n \times n$ matrices with matrix multiplication and the operator norm (Question 33). Check this.

Chapter 3

Sequences and series of functions

3.1 Introduction

You already saw much of this is MATH2111, but now you **need** to know it! The sorts of questions that we want to address are things like

Question 3.1.1. Power series such as

$$f(z) = \sum_{k=0}^{\infty} \frac{z^k}{1+k^2}$$

are always infinitely differentiable inside their radius of convergence. Why?

Question 3.1.2. Weierstrass' Function is defined as

$$W(x) = \sum_{k=0}^{\infty} \frac{\sin(2^k x)}{2^k}.$$

In first year you could prove that this converges at each $x \in \mathbb{R}$. How can you prove that W is continuous? Is W differentiable?

Question 3.1.3. Suppose that $f : [-1, 1] \to \mathbb{R}$ can be written as a Taylor series

$$f(x) = \sum_{k=0}^{\infty} a_k x^k.$$

Then f can, in some sense, be though of as a limit of polynomials. Such a function must be infinitely differentiable. What about a function like $f(x) = |x|$ or the Weierstrass function? Can they be approximated by polynomials?

You saw in MATH2111 that although there were several sensible metrics you might use in $\mathbb{R}^n$, the ones that you use were all equivalent. That is, a sequence converges in one if and only if it converges in the other. (This is because all the metrics came from norms and all norms on $\mathbb{R}^n$ are equivalent.)

In infinite dimensional spaces, things are much more complicated, and you really need to be able to deal carefully with definitions with lots of quantifiers.

3.2 Types of Convergence

We have already mentioned pointwise and uniform convergence in the last chapter, so we should formally introduce them, and examine how general the ideas are.

Definition 3.2.1. Let A be any set. For $k \in \mathbb{Z}^+$, let $f_k : A \to \mathbb{R}$ and let $f : A \to \mathbb{R}$. We say that the sequence $\{f_k\}_{k=1}^{\infty}$ converges **pointwise** to f if for all $a \in A$, $f_k(a) \to f(a)$ in $\mathbb{R}$ as $k \to \infty$.

Some remarks on pointwise convergence.

- We don't require any structure at all on the domain of the functions. Any set A will do! Note that if $A = \{1, 2, \ldots, n\}$ then a function $f : A \to \mathbb{R}$ is really just an n-tuple $(f(1), \ldots, f(n)) \in \mathbb{R}^n$. If $A = \{1, 2, 3, \ldots\}$, then a function $f : A \to \mathbb{R}$ is just a real sequence.
- Note that limits in $\mathbb{R}$ are unique, so a sequence can have at most one pointwise limit function.
- You can replace the codomain set $\mathbb{R}$ in this definition by any metric space (X, d) (or later, any topological space) and now require that $f_k(a) \to f(a)$ in (X, d).

Example 3.2.2. Let $f_k : [-1, 1] \to \mathbb{R}$ be defined as $f_k(t) = \tan^{-1}(kt)$. Then

$$f_k(t) \to f(t) = \begin{cases} -\frac{\pi}{2}, & -1 \le t < 0, \\ 0, & t = 0, \\ \frac{\pi}{2}, & 0 < t \le 1. \end{cases}$$

As we noted earlier, we often deal with normed vector spaces of functions, and these come with a different sense of convergence.

In applications, such as: Solve

$$y' = x \sin y + x^2(y + 3),$$

you are not presented with the space to work in. Part of the skill in analysis is choosing an appropriate set of functions and an appropriate norm or metric to get everything to work.

Unless A only has finitely many elements, pointwise convergence does not correspond to norm (or indeed metric) convergence on any vector space of functions on A. This is partly why we shall need to generalize convergence from metric spaces to topological spaces.

Some common norms applied to functions include

$$\begin{aligned} \|f\|_p &= \left(\int_A |f(t)|^p \, dt \right)^{1/p}, \qquad 1 \le p < \infty \\ \|f\|_\infty &= \sup_{a \in A} |f(a)|, \\ \|f\|_{C^1[0,1]} &= \|f\|_\infty + \|f'\|_\infty, \\ \|f\|_{BV[0,1]} &= \|f\|_\infty + \|f'\|_1 \end{aligned}$$

There are of course conditions on the functions to which you can apply these norms! A fairly common procedure is to define your space to be all the functions for which the norm you have written down makes sense, and then to proceed to show that that is a vector space.

3.3 Bounded functions

Let's start with $\|\cdot\|_\infty$. If A is any nonempty set, let $\mathcal{B}(A)$ denote the vector space of all **bounded** functions on A. This is precisely the set of functions for which $\|\cdot\|_\infty$ is finite, and it is clearly a vector space.

As earlier we'll freely flip between describing this as a normed space and a metric space. We'll denote the corresponding metric as $d_\infty(f, g) = \|f - g\|_\infty$.

The normed vector space $(\mathcal{B}(A), \|\cdot\|_\infty)$ contains many useful and important subspaces. Depending on A, this might include subspaces such as the set of piecewise continuous functions, the set of polynomials, the set of continuous functions, the set of C^1 functions and so forth. (Note however that if $A = \mathbb{R}$, then these spaces are not subsets of $\mathcal{B}(A)$!)

Definition 3.3.1. A sequence of functions $\{f_k\}$ in $\mathcal{B}(A)$ converges **uniformly** to f (on A) if it converges to f in the d_∞ metric. That is, if $d_\infty(f, f_k) = \|f - f_k\|_\infty \to 0$.

Example 3.3.2. Let

$$f_k(x) = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^k}{k!}, \qquad x \in A.$$

Then f_k certainly converges pointwise (on $\mathbb{R}$ or even $\mathbb{C}$) to $f(x) = e^x$. Whether it converges uniformly depends on what you choose as the domain space A. If $A = \mathbb{R}$, then we aren't even in the setting where we can measure $d_\infty(f, f_k)$ as the functions aren't bounded. If A is a closed and bounded interval such as $[0, 1]$, then you could use the bounds on the error term in Taylor's Theorem to prove that $f_k \to f$ uniformly on $[0, 1]$. **Moral:** always specify the domain for uniform convergence!

An easy result from MATH2111 (which you should be able to easily prove for yourself).

Proposition 3.3.3. *If $f_k \to f$ uniformly on A, then $f_k \to f$ pointwise on A.*

What about the topological notions here? For ease of visualization let's take $A = [0, 1]$, f to be a continuous function on A, and $\epsilon > 0$. Sketch the graphs of f, and then $f + \epsilon$ and $f - \epsilon$. The latter two describe the 'ϵ-sausage' around (the graph of) f. The ball $B(f, \epsilon)$ is just the set of all functions g whose graph sits strictly inside that sausage. (See Figure 3.1.)

Thus another way of thinking about uniform convergence is that $f_k \to f$ uniformly if no matter how skinny an ϵ-sausage you draw around the graph of f,beyond some

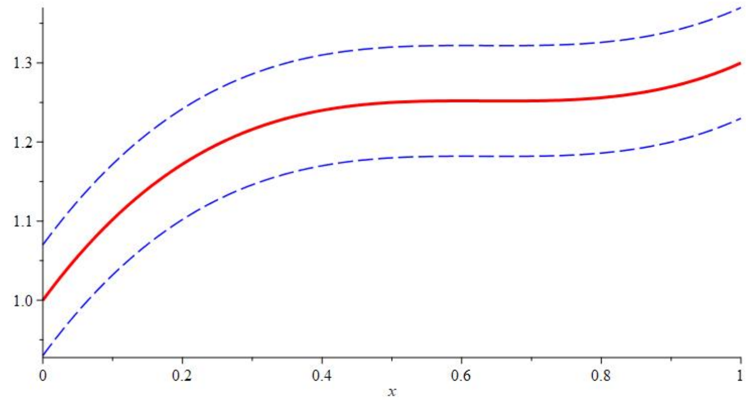


Figure 3.1: The ϵ-sausage. The red curve is the graph of f. The dashed lines are ϵ above and below this graph. If the graph of g lies within the dashed lines, then $d_\infty(f,g) < \epsilon$.

point in the sequence, the graphs of f_k are all inside that sausage. This can be useful if you care trying to determine whether a subset of $\mathcal{B}(A)$ is open or not.

What about completeness?

Theorem 3.3.4. $(\mathcal{B}(A), d_\infty)$ *is a complete metric space.*

This is much easier than the result that $(C[0,1], d_\infty)$ is complete!

Proof. Just repeat the proof of Lemma 2.10.8, which was the first part of proving that $C[0,1]$ is complete! ∎

Example 3.3.5. Let $\mathcal{P}[0,1]$ denote the set of all functions on $[0,1]$ which are polynomials. This is a vector subspace of $\mathcal{B}([0,1])$. But it isn't complete in d_∞. Let

$$f_k(x) = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^k}{k!}.$$

Then $\{f_k\}$ is a sequence in $\mathcal{P}[0,1]$ whose limit $f(x) = e^x$ isn't in $\mathcal{P}[0,1]$. Thus $\mathcal{P}[0,1]$ isn't closed in $\mathcal{B}([0,1])$ and hence it is not complete. We'll come back to this soon!

An important set of examples was given by taking $X = C[0,1]$ and

$$\|f\|_p = \left(\int_0^1 |f(t)|^p \, dt\right)^{1/p}$$

where $1 \le p < \infty$, particularly the case where $p = 2$, where we have the further structure of an inner product.

In MATH2111 we indicated that if f is a sufficiently nice (eg piecewise differentiable) element of $\mathcal{B}([-\pi,\pi])$, then the partial sums of its Fourier series

$$S_k = \frac{a_0}{2} + \sum_{n=1}^{k} a_n \cos nx + b_n \sin nx$$

converge to f in the d_2 metric, but not necessarily pointwise (or uniformly).

Unfortunately, it is easy to show that $(C[0,1], d_p)$ is not complete, and it turns out to be rather difficult to define a subspace of $\mathcal{B}([0,1])$ on which it is complete.

Of course d_p is not defined on all of $\mathcal{B}([0,1])$. You can try to define d_1 say on all Riemann integrable functions on $[0,1]$. Unfortunately this is not a metric as you can have $d_1(f,g) = 0$ with $f \neq g$. Just as serious is the problem that Riemann integration has poor limit properties.

3.4 Series of functions

Suppose that $(X, \|\cdot\|)$ is a vector space of functions on a set A, and that $\{f_k\}_{k=1}^{\infty}$ is a sequence of functions in X. The partial sum sequence is defined as

$$S_n = \sum_{k=1}^{n} f_k, \qquad n = 1, 2, \ldots.$$

Definition 3.4.1. We say that the series $\sum_{k=1}^{\infty} f_k$ converges pointwise/in $(X, \|\cdot\|)/\ldots$ if the partial sum sequence S_n converges pointwise/in $(X, \|\cdot\|)/\ldots$ as $n \to \infty$ to some function S in X.

The main examples of such series are

1. **Power series:** $S(x) = \sum_{k=0}^{\infty} a_k x^k$.
2. **Fourier series:** $S(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos kx + b_k \sin kx = \sum_{k=-\infty}^{\infty} c_k e^{ikx}$.

In both these cases the partial sum functions S_n are very nice, indeed infinitely differentiable. We want to know what we can tell about the limit S?

First an abstract result.

Theorem 3.4.2. *Suppose that $(X, \|\cdot\|)$ is a complete normed space (ie a Banach space), and that $\{x_k\}$ is a sequence in X. If $\sum_{k=1}^{\infty} \|x_k\|$ converges, then the series $\sum_{k=1}^{\infty} x_k$ converges in $(X, \|\cdot\|)$.*

Proof. Let $s_n = \sum_{k=1}^{n} \|x_k\|$. Since this sequence converges in $\mathbb{R}$ it is Cauchy. Let $S_n = \sum_{k=1}^{n} x_k$. Suppose that $m \leq n$. Then

$$\|S_n - S_m\| = \left\| \sum_{k=m+1}^{n} x_k \right\| \leq \sum_{k=m+1}^{n} \|x_k\| = |s_n - s_m|.$$

Since $\{s_n\}$ is Cauchy this implies that $\{S_n\}$ is too, and so, as $(X, \|\cdot\|)$ is complete, the series must converge. ■

Exercise 3.4.3. Give an example to show that this result may fail if $(X, \|\cdot\|)$ is not complete.

Corollary 3.4.4 (Weierstrass M-test 1). *Suppose that $\{f_k\}$ is a sequence of functions in $\mathcal{B}(A)$ and that there are constants M_k such that $\|f_k\|_\infty \le M_k$ for $k = 1, 2, \ldots$. If $\sum_{k=1}^\infty M_k$ converges then the series $\sum_{k=1}^\infty f_k$ converges uniformly on A to a bounded function f.*

Corollary 3.4.5 (Weierstrass M-test 2). *Suppose that $\{f_k\}$ is a sequence of functions in $C[0,1]$ and that there are constants M_k such that $\|f_k\|_\infty \le M_k$ for $k = 1, 2, \ldots$. If $\sum_{k=1}^\infty M_k$ converges then the series $\sum_{k=1}^\infty f_k$ converges uniformly to a function $f \in C[0,1]$.*

Remark 3.4.6. Again, the $C[0,1]$ in the last result could be replaced by $C_b(X)$.

Example 3.4.7. Weierstrass' function: $W(x) = \displaystyle\sum_{k=0}^\infty \frac{\sin 2^k x}{2^k}$. Here $f_k(x) = \dfrac{\sin 2^k x}{2^k}$, which we can consider as an element of $C_b(\mathbb{R})$. Then $\|f_k\|_\infty = 2^{-k}$ and $\sum 2^{-k}$ converges so by the Weierstrass M-test, W is a bounded continuous function on $\mathbb{R}$.

3.5 Differentiating and integrating limits of sequences

At its simplest we are interested in knowing when, if $g = \lim_{k\to\infty} g_k$, then $g' = \lim_{k\to\infty} g_k'$. If we write this out a little more explicitly, is

$$\frac{d}{dx}\lim_{k\to\infty} g_k(x) = \lim_{h\to 0}\lim_{k\to\infty} \frac{g_k(x+h) - g_k(x)}{h} = \lim_{k\to\infty}\lim_{h\to 0} \frac{g_k(x+h) - g_k(x)}{h} = \lim_{k\to\infty} g_k'(x)?$$

In general, doing limits in a different order can easily produce different answers:

$$1 = \lim_{m\to\infty}\lim_{n\to\infty} \frac{n}{n+m} \ne \lim_{n\to\infty}\lim_{m\to\infty} \frac{n}{n+m} = 0.$$

Under some conditions, however things do turn out OK here.

Any power series $S(x) = \sum_{k=0}^\infty a_k x^k = \sum f_k$ is differentiable with

$$S'(x) = \sum_{k=1}^\infty k\, a_k\, x^{k-1} = \sum f_k'(x)$$

inside the radius of convergence. That is, we can differentiate the sum 'term-by-term'. This clearly doesn't work so well with Weierstrass' function!

What we need is essentially uniform convergence of both the sequence of functions **and** the sequence of derivatives.

Theorem 3.5.1. *Suppose that $\{f_k\}_{k=1}^\infty \subseteq C[0,1]$.*

(i) If the sequence converges uniformly to $f : [0,1] \to \mathbb{R}$ then $\displaystyle\int_0^1 f_k(t)\,dt \to \int_0^1 f(t)\,dt.$

(ii) If the f_k are differentiable on $(0,1)$, $f_k \to f$ uniformly, $f_k' \in C_b(a,b)$ and $f_k' \to \phi$ uniformly. Then f' exists on $(0,1)$ and $f' = \phi$.

Proof.

(i) This is easy:

$$\begin{aligned}\Big|\int_0^1 f_k(t)\,dt - \int_0^1 f(t)\,dt\Big| &\leq \int_0^1 |f_k(t) - f(t)|\,dt \\ &\leq \|f_k - f\|_\infty \to 0\end{aligned}$$

as $k \to \infty$.

(ii) This is just a little bit harder. First note that $\phi \in C_b(0,1)$ since $C_b(0,1)$ is complete. As in (i), for any $x \in (0,1)$

$$\Big|\int_0^x f_k'(t)\,dt - \int_0^x \phi(t)\,dt\Big| \leq x\,\|f_k' - \phi\|_\infty \leq \|f_k' - \phi\|_\infty .$$

But of course $\int_0^x f_k'(t)\,dt = f_k(x) - f_k(0)$, or $f_k(x) = f_k(0) + \int_0^x f_k'(t)\,dt$. The first term converges to $f(0)$, so by the above, $f_k(x) \to f(0) + \int_0^x \phi(t)\,dt$ uniformly on $(0,1)$. Since uniform limits are unique,

$$f(x) = f(0) + \int_0^x \phi(t)\,dt, \qquad x \in (0,1).$$

Since ϕ is continuous, the Fundamental Theorem of Calculus implies that f is differentiable and $f'(x) = \phi(x)$. ■

Corollary 3.5.2. *(i') If $\{f_k\}_{k=1}^\infty \subseteq C[0,1]$ and $\sum f_k \to g$ uniformly in $C[0,1]$ then*

$$\int_0^1 g(t)\,dt = \int_0^1 \sum_{k=1}^\infty f_k(t)\,dt = \sum_{k=1}^\infty \int_0^1 f_k(t)\,dt$$

(ii') If $f_k : (0,1) \to \mathbb{R}$ are continuously differentiable, $\sum f_k$ converges uniformly and $\sum f_k'$ converges uniformly, then

$$\Big(\sum f_k\Big)' = \sum f_k'$$

Statements (i) and (i') are not true on unbounded intervals (look at the proof!). Let

$$f_k(x) = \frac{1}{\sqrt{2\pi}k}\exp(-x^2/2k^2).$$

Then $f_k \to 0$ uniformly on $\mathbb{R}$, but the $\displaystyle\int_{-\infty}^\infty f_k(t)\,dt = 1 \not\to \int_{-\infty}^\infty 0\,dt = 0.$

Part (ii) of the theorem is strong enough for most applications, but the hypotheses about f can be weakened substantially (see Rudin's Principles of Mathematical Analysis for example), or perhaps more accurately, be deduced from the facts about f'_k. Such theorems require rather more delicate proofs!

The proof for part (ii) clearly works if $[0, 1]$ is replaced by a closed interval $[a, b]$. For an unbounded interval, it is enough that one gets uniform convergence on a some closed interval around each point.

Exercise 3.5.3. Prove part (ii) where $[0, 1]$ is replaced by a disk $\{z \in \mathbb{C} : |z| \leq r\}$ in the complex plane. (Note that there is a Fundamental Theorem of Calculus for complex functions!) We'll need this soon!

3.6 Aside: limit superior limit inferior

In order to prove that you can differentiate power series term-by-term, you need to recall (or learn) a few things from the real-valued setting.

The first concepts are those of limit superior and limit inferior. In general a bounded real sequence $\{a_k\}$ need not converge, but it does always have convergent subsequences. (This is the *Bolzano-Weierstrass Theorem.*) The limit superior of $\{a_k\}$ can be defined as the largest number that any subsequence has as a limit. The limit inferior of $\{a_k\}$ is the smallest number that any subsequence has as a limit.

Example 3.6.1. Let $a_k = \frac{1}{k} - \sin k$. The supremum of this sequence is in fact $a_5 \approx 1.16$, and the sequence is greater than one infinitely often. Any convergent subsequence however must have limit $L \in [-1, 1]$, so $\limsup_k a_k = 1$ and $\liminf_k a_k = -1$.

The usual definition given is less useful, but perhaps easier to see that it exists! Suppose that $m \leq a_k \leq M$ for all k. For each $n \geq 1$ the set $\{a_k\}_{k \geq n}$ is a bounded set, and hence has a supremum

$$b_n = \sup_{k \geq n} a_k.$$

Furthermore as you remove more and more elements from the start of this list the supremums can only decrease, so

$$b_1 \geq b_2 \geq b_3 \cdots \geq m.$$

Thus $\{b_n\}$ is a monotone sequence which is bounded below and hence (via the Least Upper Bound Property of $\mathbb{R}$) it converges. The limit superior is then defined as

$$\limsup_k a_k = \lim_{n \to \infty} b_n.$$

The limit inferior is defined similarly.

Exercise 3.6.2. Show that the two definitions agree.

Theorem 3.6.3. *Suppose that $\{a_k\}$ is a bounded sequence. Then $\limsup_k a_k$ and $\liminf_k a_k$ both exist and*

$$\inf_k a_k \leq \liminf_k a_k \leq \limsup_k a_k \leq \sup_k a_k.$$

Exercise 3.6.4. Prove that $\{a_k\}$ converges to L if and only if $\limsup_k a_k = \liminf_k a_k = L$.

Exercise 3.6.5. Show that if $L = \limsup_k a_k$ then for all $M > L$ there exists K such that for all $k \geq K$, $a_k < M$.

The first place that these really appear is in the proper versions of the ratio and root tests. The versions that you saw in first year involved limits, but you actually need the professional grade root test.

Theorem 3.6.6. *The (real or complex) series $\sum w_k$ converges (absolutely) if*

1. *(Ratio Test)* $\limsup_k \left|\dfrac{w_{k+1}}{w_k}\right| < 1$, *or*

2. *(Root Test)* $\limsup_k \sqrt[k]{|w_k|} < 1$.

Proof. (Just 2.) Let $\alpha = \limsup_k \sqrt[k]{|w_k|} < 1$. Choose $\beta \in (\alpha, 1)$. By Exercise 3.6.5 there exists K such that for all $k \geq K$, $\sqrt[k]{|w_k|} < \beta$, and hence $|w_k| < \beta^k$. But $\sum_{k=K}^{\infty} \beta^k$ converges and hence by the comparison test $\sum w_k$ converges absolutely.

(The proof of 1. is very similar) ■

In first year we gave the radius of convergence of the power series $S(z) = \sum_{k=0}^{\infty} a_k z^k$ as

$$R = \lim_{k\to\infty} \left|\frac{a_k}{a_{k+1}}\right|.$$

This isn't much use if the limit doesn't exist. A more general result is that

$$\frac{1}{R} = \limsup_k \sqrt[k]{|a_k|}. \tag{3.6.1}$$

which follows easily from the root test.

3.7 Differentiating power series

To show that power series can be differentiated term-by-term we want to apply Theorem 3.5.2. One needs to be a little careful however since power series typically do not converge uniformly on their disk of convergence. (We'll work here with power series defined for a complex variable, even if ultimately we may only be interested in what happens on some part of the real line.)

Example 3.7.1. Let $S(z) = \dfrac{1}{1-z} = \displaystyle\sum_{k=0}^{\infty} z^k$. The radius of convergence is 1, and on $D = \{z \in \mathbb{C} : |z| < 1\}$, S is not even bounded. The partial sum functions $S_n(z) = \sum_{k=0}^{n} z^k$ are at least in $C_b(D)$, but they don't form a bounded sequence.

The idea, which is very important in much of analysis is to work with uniform convergence on bigger and bigger subsets of the domains of the functions.

First an observation which follows from the formula for the radius of convergence (3.6.1) above.

Lemma 3.7.2. *The radius of convergence of the power series* $S_1(z) = \sum_{k=0}^{\infty} a_k\, z^k$ *and* $S_2(z) = \sum_{k=0}^{\infty} a_k\, z^{k+1} = zS_1(z)$ *is the same.*

Theorem 3.7.3. *Suppose that* $S(z) = \displaystyle\sum_{k=0}^{\infty} a_k z^k$ *is a power series which converges pointwise for* $|z| < R$ *in* $\mathbb{C}$. *Then on this disk,* S *is differentiable with*

$$S'(z) = \sum_{k=1}^{\infty} k\, a_k\, z^{k-1}.$$

Proof. For $n = 0, 1, 2, \ldots$, let $S_n(z) = \sum_{k=0}^{n} a_k z^k$.

Suppose that $r < R$ and consider the disk $D_r = \{z : |z| \le r\}$. For $k = 0, 1, 2, \ldots$, let $M_k = |a_k| r^k$. Note that by the radius of convergence formula (3.6.1)

$$\limsup_k \sqrt[k]{|M_k|} = \limsup_k \sqrt[k]{|a_k|}\, r < \limsup_k \sqrt[k]{|a_k|}\, R = 1$$

and so $\sum_k M_k$ converges by Theorem 3.6.6.

Let $f_k : D_r \to \mathbb{C}$, $f_k(z) = a_k z^k$. Then (on D_r), $\|f_k\|_\infty \le M_k$, so by the Weierstrass M-test (Version 2), the series for S converges uniformly on D. That is $S_n \to S$ uniformly on D_r.

Note that as $\sqrt[k]{k} \to 1$ as $k \to \infty$,

$$\limsup_k \sqrt[k]{|k a_k|} = \limsup_k \sqrt[k]{k}\, \sqrt[k]{|a_k|} = \limsup_k \sqrt[k]{|a_k|}$$

and so the series $\sum k\, a_k\, z^{k-1}$ has the same radius of convergence as the series for S. So the same proof shows that the series $\displaystyle\sum_{k=1}^{\infty} k\, a_k\, z^{k-1}$ also converges uniformly on D_r. That is S_n' converges uniformly to $\phi(z) = \displaystyle\sum_{k=1}^{\infty} k\, a_k\, z^{k-1}$ on D_r

Corollary 3.5.2 now allows us to deduce that on the interior of D_r, S is differentiable, and its derivative is given by ϕ.

As this holds on all closed disks inside $|z| < R$, it holds on all of the big open disk! ■

3.8 Appendix: L_1 Convergence

We initially defined the metric

$$d_1(f,g) = \int_0^1 |f(t) - g(t)|\, dt \tag{3.8.1}$$

on $C[0,1]$. Unfortunately $(C[0,1], d_1)$ is not complete, and in any case, (3.8.1) makes sense for many noncontinuous functions.

Question 3.8.1. What is the largest set of functions on $[0,1]$ for which the metric d_1 makes sense?

The evaluation of $d_1(f,g)$ clearly relies on the integrability of f and g. The problem is that if

$$f(t) = 0 \quad \forall t \in [0,1] \qquad g(t) = \begin{cases} 0 & t \neq \frac{1}{2} \\ 1 & t = \frac{1}{2} \end{cases}$$

then $d_1(f,g) = 0$ but $f \neq g$. So it seems that d_1 doesn't really behave like a metric anymore.

Even worse! List the rationals in $[0,1]$ as

$$\mathbb{Q} \cap [0,1] = \{r_1, r_2, \dots\}$$

For $k \geq 1$, let

$$f_k(t) = \begin{cases} 1 & t \in \{r_1, r_2, \dots, r_k\} \\ 0 & \text{otherwise} \end{cases} \qquad f(t) = \begin{cases} 1 & t \in \mathbb{Q} \cap [0,1] \\ 0 & \text{otherwise} \end{cases}$$

Then clearly $f_k \to f$ pointwise, but $d_1(f_k, f_l) = 0$ for all k, l (so $\{f_k\}$ is in some sense Cauchy). But f is not Riemann integrable, so $d_1(f_k, f)$ doesn't really make sense!

Fixing these problems is complicated! The second problem is circumvented by using Lebesgue instead of Riemann integration. The function f in the above example **is** Lebesgue integrable. We'll discuss how to deal with these issues in Chapter 6.

To fix the first problem we "declare" f and g to be "equal" if $d_1(f,g) = 0$. That is, the elements of the new metric space are not functions, but **equivalence classes of functions** that are equal 'almost everywhere'. We'll need to make this more precise soon too.

Once one does this, you can construct a **complete** normed vector space $L_1[0,1]$ with[1]

$$\|f\|_1 = \int_0^1 |f|\, d\mu$$

Moreover, similar ideas are used with d_p to make spaces $L^p[0,1]$ for $1 \leq p < \infty$.

[1] One should, but rarely do, write $\|[f]\|_1$ where $[f]$ stands for the equivalence class of f.

3.9 Problems

1. For $k = 1, 2, \dots$, define $f_k : [0,1] \to \mathbb{R}$ by

$$f_k(x) = \begin{cases} 4k^2 x, & 0 \le x \le \frac{1}{2k}, \\ 4k(1-kx), & \frac{1}{2k} < x \le \frac{1}{k}, \\ 0, & \frac{1}{k} < x \le 1. \end{cases}$$

   (a) Show that $\{f_k\}$ has a pointwise limit, f, but that $f_k \not\to f$ uniformly.

   (b) Does $\displaystyle\int_0^1 f_k(x)\,dx \to \int_0^1 f(x)\,dx$?

2. For the sequences $\{f_k\}_{k=1}^{\infty}$ below, answer the following. Does $\{f_k\}$ converge? In what sense? Is the limit integrable? Differentiable? Is the integral/derivative of the limit equal to the limit of the integrals/derivatives of the functions in the sequence?

   (a) $f_k(x) = \sqrt{k}\, x^k(1-x)$, $x \in \mathbb{R}$.

   (b) $f_k(x) = e^{-(x-k)^2}$, $x \in \mathbb{R}$.

   (c) $f_k(x) = \left(1 + \frac{x}{k}\right)^k$, $x \in [0,1]$.

3. Define $S(x) = \displaystyle\sum_{k=1}^{\infty} \frac{x \sin kx}{k^4}$, $x \in \mathbb{R}$.

   (a) Show that S converges pointwise on $\mathbb{R}$.

   (b) S is not bounded on $\mathbb{R}$, but show that it converges uniformly on any compact interval of the real line.

   (c) Can S be differentiated term by term?

4. For $k = 1, 2, \dots$, define $g_k : \mathbb{R} \to \mathbb{R}$ by $g_k(x) = \dfrac{\sin(2^k x)}{2^k}$.

   (a) Verify that $\sum_{k=1}^{\infty} g_k$ converges uniformly to say g, and that g is continuous.

   (b) Does $\sum_{k=1}^{\infty} g_k'$ converge?

   (c) (Harder) Prove that g is not differentiable anywhere.

5. Let $\mathbb{T} = [0, 2\pi)$. As one usually does with angles on the circle, there is a natural 'mod 2π' projection M from the real line onto $\mathbb{T}$. (So, for example, $M(5\pi) = \pi$.) If $a < b$ are real numbers we'll let $I_{a,b} = \{M(x) \in \mathbb{T} : a \le x \le b\}$.

   For $k = 1, 2, \dots$, let $t_k = \sum_{m=1}^{k} \frac{1}{m}$ so that $t_k \to \infty$ as $k \to \infty$, but $t_{k+1} - t_k \to 0$. Define $f_k : \mathbb{T} \to \mathbb{R}$ be the characteristic function of $I_{t_k, t_{k+1}}$.

   (a) Does f_k converge pointwise?

(b) Prove that $f_k \to 0$ in the $L^1(\mathbb{T})$ metric.

6. For $k = 1, 2, \ldots$, let $f_k(x) = \frac{k}{\sqrt{\pi}} e^{-(kx)^2}$, $x \in \mathbb{R}$. Physicists are wont to write things like $\delta(x) = \lim\limits_{k \to \infty} f_k(x)$. Does this sequence actually converge in any sense? (We'll see more about this later!)

7. Given a set $S \subseteq \mathbb{R}$ and $a, b \in \mathbb{R}$, we write $aS + b$ for the set $\{ax + b : x \in S\}$. Let $C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. For $n = 2, 3, 4, \ldots$, let

$$C_n = \frac{1}{3} C_{n-1} \cup \left(\frac{1}{3} C_{n-1} + \frac{2}{3} \right).$$

The **Cantor set** is the set $C = \bigcap\limits_{n=1}^{\infty} C_n$.

(a) For $n \geq 1$, let $\chi_n : [0, 1] \to \mathbb{R}$ be the characteristic function of C_n. What is the value of $\displaystyle\int_0^1 \chi_n(t)\, dt$? Define $f_n : [0, 1] \to \mathbb{R}$ by

$$f_n(x) = \frac{\int_0^x \chi_n(t)\, dt}{\int_0^1 \chi_n(t)\, dt}.$$

(b) Prove that $\{f_n\}$ converges uniformly to a continuous function f. What is the range of f?

(c) Where is f differentiable? Find the derivative of f at those points.

8. Check out the following web sites

`http://www.cut-the-knot.org/do_you_know/hilbert.shtml`
`http://www.math.uwaterloo.ca/`
`~wgilbert/Research/HilbertCurve/HilbertCurve.html`
`http://www.dcs.napier.ac.uk/~andrew/hilbert.html`
`http://mathworld.wolfram.com/HilbertCurve.html`

for descriptions of space-filling curves. These sites actually give a Cauchy sequence of **continuous** functions $f_k : [0, 1] \to [0, 1] \times [0, 1]$.

(a) Convince yourself that the sequences given actually are Cauchy! (The metric here is $d(f, g) = \sup_{t \in [0,1]} ||f(t) - g(t)||$.)

(b) Prove that $X = \{f : [0, 1] \to [0, 1] \times [0, 1] : f \text{ is continuous}\}$ is a complete metric space.

(c) Deduce that the sequences given must have a limit function $f \in X$.

(d) Choose $(x, y) \in [0, 1] \times [0, 1]$. Describe how one finds $t \in [0, 1]$ such that $f(t) = (x, y)$.

(e) Thinking of the Jordan Curve Theorem from complex analysis, does the curve $\{f(t) : t \in [0,1]\}$ have an inside and an outside?

9. (To be done using MAPLE) Suppose that $f : [0,1] \to \mathbb{R}$. For $n \in \mathbb{N}$ let

$$p_n(x) = \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

By choosing a range of functions f, investigate the relationship between f and the sequence p_n.

Chapter 4

Topological spaces

4.1 Introduction

When dealing with functions many different limiting processes arise, eg

- $f_k \to f$ pointwise.
- $f_k \to f$ uniformly.
- $\displaystyle\int_0^1 |f_k(t) - f(t)|\, dt \to 0$.
- for all $g \in C[0,1]$, $\displaystyle\int_0^1 f_k(t)g(t)\, dt \to \int_0^1 f(t)g(t)\, dt$.
- on every closed disk inside the open unit disk, $f_k \to f$ uniformly.

Can we describe all of these in terms of convergence in some suitable metric space (or better yet, a normed space)?

The answer is unfortunately NO. In order to provide a theory that covers more or less all reasonable ideas of convergence, the theory of Point Set Topology was developed in the first half of the 20th century. This theory is based not on the idea of distance, but on the idea of an open set. A great many of our metric space concepts can be expressed just in terms of open sets:

- a sequence $\{x_k\}$ converges to x if and only if it eventually sits in any open set containing x.
- a function $f : X \to Y$ is continuous if and only if the inverse image of every open set is open.

The idea is to extract the properties of open sets in metric spaces that make this all work.

Definition 4.1.1. Let X be (nonempty) set. A family τ of subsets of X is a **topology** on X if

1. $\emptyset, X \in \tau$.

2. If $U_1, \ldots, U_n \in \tau$ then $\bigcap_{i=1,\ldots,n} U_i \in \tau$.

3. If $\{U_\alpha\}_{\alpha\in A} \subseteq \tau$ then $\bigcup_{\alpha\in A} U_\alpha \in \tau$.

The elements of τ are called the **open** (or τ-open) sets in X. The pair (X, τ) is called a **topological space**

It is probably not at all obvious why the above conditions give you anything interesting, but it is at least reassuring that the things that we called open sets earlier in the course are still open!

Example 4.1.2. Let (X, d) be a metric space, then let

$$\tau = \{U \subseteq X \ : \ U \text{ is open in } d\ \}$$

Then τ defines a topology on X, since we know that the open sets in (X, d) have the above properties. We will often describe this as the '**metric topology on** X'.

Example 4.1.3. It is easy to give examples of very trivial topologies on a set X

- **Indiscrete Topology**: $\tau_1 = \{\emptyset, X\}$.
- **Discrete Topology**: $\tau_2 = \mathcal{P}(X)$.

Clearly τ_1, τ_2 both define topologies on X. We will see shortly that these topologies aren't very interesting at all. Note that the discrete topology is really just the one that comes from the discrete metric on X. The indiscrete topology however can't come from a metric (except in the trivial case where X has just one element). Why?

Example 4.1.4. Slightly more interesting is the **Cofinite Topology**: Let X be a nonempty set.

$$\tau_3 = \{\emptyset\} \cup \{U \subseteq X \ : \ X \setminus U \text{ is finite}\}.$$

Exercise 4.1.5. Prove that this is a topology!

It is clear from these examples that the theory of topological spaces will generalize the theory of metric spaces. What is not clear is

1. whether it gives a useful generalization,

2. whether it captures all the interesting examples of convergence listed earlier.

Convincing you that the answer to these questions is 'Yes' will take some time because it is pretty much the case that for all the useful topologies for analysis it is quite hard to describe what all elements of τ look like.

In practice, knowing what the open sets look like turns out to be not so important. We usually have a test for convergence, such as pointwise convergence, which doesn't involve eventually getting inside open sets. What *is* important is knowing that there is an (nice) underlying topology which corresponds to your sense of convergence, and hence that you can apply the theorems that hold in all (nice) topological spaces.

Example 4.1.6. Let $X = \mathbb{R}^2$ and consider the metrics d_1 and d_2 on X. Note that a set $U \subseteq \mathbb{R}^2$ is open with respect to one of these metrics if and only if it is open with respect to the other. Thus the corresponding topologies are the same. This is picking out the fact that a sequence converges in one metric if and only if it converges in the other. In terms of convergence and continuity, the topology is what really matters, not the metric!

Many of our metric space notions can be extended to the topological space setting.

Definition 4.1.7. Let (X, τ) be a topological space.

1. $Y \subseteq X$ is **closed** if $X \setminus Y$ is open.
2. $x \in Y \subseteq X$ is an **interior point** of Y if there exists $U \in \tau$ such that $x \in U$ and $U \subseteq Y$.
3. A **neighbourhood** of a point $x \in X$ is any open set containing x.
4. $x \in X$ is a **boundary point** of $Y \subseteq X$ if every neighbourhood of x intersects both Y and $X \setminus Y$.
5. Suppose $Y \subseteq X$. A point $x \in X$ is a **limit point** of Y if every neighbourhood of x contains an element of Y which is different from x.
6. The **closure** of a set $Y \subseteq X$ is the set
$$\mathrm{cl}(Y) = Y \cup \{\text{limit points of } Y\}.$$

Note that there one should check that we haven't changed the definitions as we move from metric spaces to topological spaces. For example the metric space definition of an interior point requires that a ball $B(x, \epsilon)$ sits entirely inside Y. That certainly implies that x is an interior point using the above topological definition. Conversely if some open set U containing x sits entirely inside Y, then as U is open in the metric space, there is a small ball $B(x, \epsilon)$ sitting inside U, and hence inside Y. In Section 4.4 we'll come back and look at something that plays the role of a ball in a topological space.

It is worthwhile checking that a few of our metric space facts still hold true in topological spaces.

Theorem 4.1.8. *Let (X, τ) be a topological space and suppose that $Y \subseteq X$. Then* $\mathrm{cl}(Y)$ *is the smallest closed set containing Y. That is, if Z is a closed set containing Y, then* $\mathrm{cl}(Y) \subseteq Z$.

Proof. Suppose that Z is a closed set containing Y and that x is a limit point of Y. If $x \notin Z$ then $X \setminus Z$ is an open set containing x which does not intersect Y, which contradicts that x is a limit point. Therefore $x \in Z$ and hence $\mathrm{cl}(Y) \subseteq Z$.

For the proof that $\mathrm{cl}(Y)$ is itself closed, see the problems. ■

Theorem 4.1.9. *Let (X, τ) be a topological space and $Y \subseteq X$. Then Y is open if and only if every element of Y is an interior point.*

Proof. ($\Rightarrow$) Suppose Y is open in the τ-topology and $y \in Y$, then we note simply that Y is an open set containing y and $Y \subseteq Y$. Hence y is an interior point of Y.

($\Leftarrow$) Suppose every $y \in Y$ is an interior point of Y. Then for each $y \in Y$,there exists $U_y \subseteq U$ with $y \in U_y$ and clearly

$$Y = \bigcup_{y \in Y} U_y$$

That is, Y can be written as a union of open sets and so Y is open. ■

4.2 Convergence in Topological Spaces

We can now proceed to make a definition of convergence, and show that it is equivalent to the definition of convergence in a metric space. However, we will also highlight in this section that many of the nice convergence properties of sequences in a metric space do not hold for the general topological space.

Definition 4.2.1. Suppose that $\{x_k\}_{k=1}^{\infty}$ is a sequence in a topological space (X, τ). Then we say that $\{x_k\}_{k=1}^{\infty}$ **converges to** $x \in X$ **in the τ-topology** if for any open neighbourhood G of x, there exists K such that for all $k \geq K$, $x_k \in G$. We'll write $x_k \xrightarrow{\tau} x$ if we want to specify the topology being used.

This was not quite the definition we had for metric spaces but we showed (Theorem 2.4.10) that this is equivalent to the original definition (that is, that $d(x_k, x) \to 0$).

For more exotic topologies, convergence can be rather counterintuitive.

Example 4.2.2. Let (X, τ) be a space with the indiscrete topology $\tau = \{\emptyset, X\}$. Suppose that $\{x_k\}$ is any sequence in X and x any element of X. Then there is only one open set containing x, namely X and the sequence is clearly always in this set, so $x_k \xrightarrow{\tau} x$. That is every sequence converges to every element — yuck!

One thing that this example shows is that there are topological spaces where the open sets can't have come from some metric on the space.

Example 4.2.3. Let $X = \mathbb{Z}$. Two standard metrics on X are the usual metric $d(x, y) = |x-y|$ and the discrete metric, $\delta(x, y) = 1$ unless $x = y$. Although these metrics are very different, convergence in the metric spaces (X, d) and (X, δ) is identical: $\{x_k\}$ converges in $(X, d) \iff \{x_k\}$ converges in $(X, \delta) \iff$ the sequence is eventually constant. This is picking up that the topologies generated by the two metrics are the same, in this case the discrete topology $\tau = \mathcal{P}(\mathbb{Z})$. This illustrates that it is the topology that really matters, not the underlying metric!

Example 4.2.4. Let $X = \mathbb{R}$ and for $n \in \mathbb{Z}$, define $I_n = (n, n+1]$.
Define $\tau = \{\emptyset\} \cup \{\cup_{j \in J} I_j \,:\, \emptyset \neq J \subseteq \mathbb{Z}\}$. It is a routine exercise to verify that τ defines a topology on X.

Let us now explore some of (X, τ)'s properties.

Let $x_k = \frac{1}{k}$, $k \in \mathbb{Z}^+$ and set $x = \frac{1}{\pi}$. If G is a neighbourhood of x then $(0, 1] \subseteq G$, and so $x_k \in G$ for all $k \in \mathbb{Z}^+$. Thus, $x_k \xrightarrow{\tau} x$. In fact $x_k \to y$ for all $y \in (0, 1]$ under the τ-topology. Even worse, $x_k \overset{\tau}{\not\to} 0$ under the τ-topology since $(-1, 0]$ is an open set containing 0 and the sequence never even enters it.

The reason why the some of the above examples are so poorly behaved is due to the fact that there aren't really enough open sets defined by τ.

Recall that when we proved that limits are unique in a metric space (X, d) , we used the fact that if $x, y \in X$ with $x \neq y$ and $\epsilon = d(x, y)/2 > 0$. Then $B(x, \epsilon/2)$ and $B(y, \epsilon/2)$ are disjoint open sets around x and y and so a sequence can't eventually end up in both of them.

Definition 4.2.5. A topological space (X, τ) is said to be **Hausdorff** if for all $x, y \in X$ with $x \neq y$ there exist disjoint open neighbourhoods of x and y respectively. Hausdorff space are sometimes called T_2-spaces or 'topological spaces satisfying the Second Axiom of Separability[1]'.

Theorem 4.2.6. *In a Hausdorff topological space (X, τ), limits of sequences are unique. That is, if $x_k \to x$ and $x_k \to y$ then $x = y$.*

Exercise 4.2.7. Adapt the metric space proof!

Theorem 4.2.8. *Let (X, τ) be a Hausdorff topological space and suppose that $Y \subseteq X$. Then x is a limit point of Y iff every neighbourhood of x contains infinitely many elements of Y.*

[1] You may take from this that there is more than one axiom of separability. There are in fact lots of these. Note only are there T_0, T_1 and T_3 spaces, there are also $T_{1\frac{1}{2}}$ and $T_{2\frac{1}{2}}$ spaces, and many others. If you are interested, you can look these up on Wikipedia.

Proof. ($\Leftarrow$) This is clear.

($\Rightarrow$) Suppose then that x is a limit point of Y and that U is a neighbourhood of x. Then there must exist $u_1 \in Y \cap U$ with $u_1 \neq x$. As (X, τ) is Hausdorff, there exists neighbourhoods V_1 of u_1 and W_1 of x such that $V_1 \cap W_1 = \emptyset$. Let $U_1 = U \cap W_1$. Then U_1 is a neighbourhood of x and hence there exists an element $u_2 \in Y \cap U_1$. Note that $u_2 \in Y \cap U$ too and that $u_2 \neq u_1$. One can proceed recursively to construct a sequence of elements $\{u_k\}_{k=1}^{\infty} \subseteq Y \cap U$. ■

Remark 4.2.9. 1. Note that this implies that you only get limit points in topological spaces with infinitely many elements!

2. All useful topologies' in analysis are Hausdorff, however algebraists[2] use things like the Zariski topology which is usually not Hausdorff.

In a metric space, $x \in \text{cl}(Y)$ if and only if there is a sequence $\{x_k\}_{k=1}^{\infty} \subseteq Y$ which converges to x. Does this still hold in topological spaces?

Example 4.2.10. Define a topology τ on $\mathbb{R}$ by saying that $U \in \tau$ if

$$U = \emptyset \quad \text{or} \quad \mathbb{R} \setminus U \text{ is countable.}$$

This is called the **co-countable** topology on $\mathbb{R}$. As with many topologies it is slightly easier to say what a closed sets is: a set C is closed in $(\mathbb{R}, \tau)$ if $C = \mathbb{R}$, or if C is countable.

Let $Y = [0, 1]$, then $\text{cl}(Y) = \mathbb{R}$ since this is the only closed set containing Y . Now, suppose $x \in \text{cl}(Y) = \mathbb{R}$ and suppose $\{x_k\} \subseteq \mathbb{R}$. Then the set $C = \{x_k\}_{k=1}^{\infty} \setminus \{x\}$ is closed so $U = \mathbb{R} \setminus C$ is open. Since U contains x, it is neighbourhood of x.

Note that no element of $\{x_k\}_{k=1}^{\infty}$ can be in U unless it equals x so this sequence isn't eventually inside U unless x_k is eventually equal to x.

Note in particular that although $2 \in \text{cl}(Y)$, there does not exist a sequence $\{x_k\}_{k=1}^{\infty} \subseteq Y$ such that $x_k \to 2$.

The above highlights that this is not a very nice topology, it is clearly not Hausdorff. However, you can find examples of Hausdorff spaces with the same property: there exists $x \in \text{cl}(Y)$ such that no sequence $\{x_k\} \subseteq Y$ converges to x.

4.3 Nets

One fix to this problem is to deal not with sequences, but with 'generalized sequences'. A sequence is a collection of elements $\{x_k\}$ indexed by k in the ordered set $\mathbb{N}$. With 'generalized sequences' we work with a collection of elements $\{x_\alpha\}_{\alpha \in A}$ which are indexed by more general 'partially ordered' sets A. A partial order is something which is like the usual $\leq$ relation, but for which you can't necessarily compare every pair of elements in your set.

[2] Algebraists, like physicists, need to be approached with extreme caution!

Definition 4.3.1. Let A be a set. A **partial order** on A is a relation $\preceq$ on A satisfying

1. $a \preceq a, \quad \forall a \in A$.
2. If $a \preceq b$ and $b \preceq c$ then $a \preceq c$.
3. If $a \preceq b$ and $b \preceq a$ then $a = b$.

As usual $a \prec b$ means $a \preceq b$ with $a \neq b$.

A pretty useless partial order would be one where the only comparison statements one can make are that $a \preceq a$ for all a. A total order would be one where we could compare any two elements. For our purposes we can get by with a bit less.

Definition 4.3.2. A **directed set** is a partially ordered set $(A, \preceq)$ that satisfies

4. If $a, b \in A$ then there exists $c \in A$ such that $a \preceq c$ and $b \preceq c$.

Example 4.3.3. 1. $(\mathbb{N}, \leq)$ and $(\mathbb{R}, \leq)$ are totally ordered.

2. $(\mathbb{N} \times \mathbb{N}, \preceq)$, where $(n_1, m_1) \preceq (n_2, m_2)$ if $n_1 < n_2$ or if $n_1 = n_2$ and $m_1 \leq m_2$. This is also totally ordered.

3. Let V be a vector space and let A denote the set of all vector subspaces of V. Say $V_1 \preceq V_2$ if $V_1 \subseteq V_2$. This is not totally ordered, but given V_1 and V_2, you always have $V_1 \preceq V$ and $V_2 \preceq V$, so $(A, \preceq)$ is a directed set.

4. Let $\mathbb{P} = \{$ All partitions of $[0,1]$ $\}$. We now define a partial ordering $\preceq$ on $\mathbb{P}$. Recall that a partition of $[0,1]$ is a finite subset of the form $\Lambda = \{\lambda_j\}_{j=0}^n$ where

$$0 = \lambda_0 < \lambda_1 < \cdots < \lambda_n = 1$$

Let $\Gamma = \{\gamma_j\}_{j=1}^m \in \mathbb{P}$ then we say that $\Lambda \preceq \Gamma$ if $\Lambda \subseteq \Gamma$. Note that $(\mathbb{P}, \preceq)$ defines a directed set, since if $\Gamma, \Lambda \in \mathbb{P}$, then if $S = \Gamma \cup \Lambda$, we clearly have $\Gamma \preceq S$ and $\Lambda \preceq S$.

5. Let X be a nonempty set and let $A = \mathcal{P}(X)$ be the power set of X. Define $\preceq$ on A by saying that $U \preceq V$ if $V \subseteq U$ ('reverse inclusion'). Again this is a directed set since if $U, V \in A$ the set $S = U \cap V$ satisfies $U \preceq S$ and $V \preceq S$.

Definition 4.3.4. A **net** (or generalised sequence) in X is a family of elements $\{x_\alpha\}_{\alpha \in A}$ of X indexed by a directed set $(A, \preceq)$.

Example 4.3.5. 1. Every sequence is a net, using the directed set $(\mathbb{N}, \leq)$.

2. Let $f : [0,1] \to \mathbb{R}$. For a partition $\Lambda \in \mathbb{P}$ as in a previous example define

$$S_\Lambda = \sum_{j=1}^n f(\lambda_j)(\lambda_j - \lambda_{j-1})$$

Then $\{S_\Lambda\}_{\Lambda \in \mathbb{P}}$ is a net in $\mathbb{R}$ indexed by $(\mathbb{P}, \preceq)$.

Definition 4.3.6. Let (X, τ) be a topological space, and suppose that $\{x_\alpha\}_{\alpha \in A}$ is a net in X indexed by $(A, \preceq)$. We say that $\{x_\alpha\}$ **converges** to $x \in X$ if given any neighbourhood G of x, there exists $\alpha_0 \in A$ such that for all $\alpha \succeq \alpha_0$, $x_\alpha \in G$. We will write this as

$$x_\alpha \xrightarrow{\tau} x \qquad \text{or} \qquad \tau\text{-}\lim_{\alpha \in A} x_\alpha = x$$

Example 4.3.7. Consider $\mathbb{R}$ with its usual topology. We'll take our directed set as $A = \mathbb{R} \setminus \{0\}$ with the partial ordering $a \prec b$ if $|b| < |a|$. Consider the net $\{a\}_{a \in A}$.

I claim that this net converges to 0 in the usual topology. Suppose then that G is a neighbourhood of 0. As G is open, there exists $\epsilon > 0$ such that $(-\epsilon, \epsilon) \subseteq G$. Let $a_0 = \epsilon/2 \in A$ and suppose that $a \succeq a_0$. Then $|a| \leq \epsilon/2$ and so $a \in (-\epsilon, \epsilon) \subseteq G$. That is, for all $a \succeq a_0$, $a \in G$. Thus $\{a\}_{a \in A}$ converges to 0 in the usual topology in $\mathbb{R}$.

Example 4.3.8. If $\{S_\Lambda\}_{\Lambda \in \mathbb{P}}$ converges in $\mathbb{R}$ then we say that f is Riemann integrable.

If we use nets, then we can recover our characterization of the closure of a set.

Theorem 4.3.9. *Let (X, τ) be a topological space and suppose that $Y \subseteq X$. Then $x \in \mathrm{cl}(Y)$ if and only if there exists a net $\{x_\alpha\}_{\alpha \in A} \subseteq Y$ such that $x_\alpha \xrightarrow{\tau} x$.*

Proof. ($\Rightarrow$) Suppose $x \in \mathrm{cl}(Y)$. If $x \in Y$ then let $x_n = x$ for all $n \in \mathbb{N}$ and then certainly $x_n \to x$.

If $x \notin Y$ then x is a limit point of Y. The challenge now is to cook up a suitable net that converges to x, and for this we need a suitable index set.

Let $\mathcal{N}_x$ denote the set of all neighbourhoods of x. (Note that $\mathcal{N}_x$ is necessarily nonempty!) Define an order $\preceq$ on $\mathcal{N}_x$ by reverse inclusion:

$$U \preceq V \iff V \subseteq U$$

As we noted earlier, this partial order makes $(\mathcal{N}_x, \preceq)$ a directed set.

For $U \in \mathcal{N}_x$, as x is a limit point of Y it follows that $U \cap Y \neq \emptyset$. By the Axiom of Choice we can choose one element from each set $U \cap Y$; let us call it y_U. Thus, $\{y_U\}_{U \in \mathcal{N}_x}$ is a net in (X, τ).

Now, suppose V is a neighbourhood of x. Choose $U_0 = V$. Then if $U \succeq U_0$ we have $y_U \in U \subseteq V$. Thus $y_U \xrightarrow{\tau} x$.

($\Leftarrow$) Let $\{x_\alpha\}_{\alpha \in A} \subseteq Y$ be a net and suppose that $x_\alpha \xrightarrow{\tau} x$, but that $x \notin \mathrm{cl}(Y)$. Now $\mathrm{cl}(Y)$ is closed, so $U = X \setminus \mathrm{cl}(Y)$ is an open neighbourhood of x which doesn't intersect Y. As $x_\alpha \xrightarrow{\tau} x$, there exists $\alpha_a \in A$ such that for all $\alpha \succeq \alpha_0$, $x_\alpha \in U$. But if $x_\alpha \in U$, then $x_\alpha \notin Y$ which is impossible. Thus $x \in \mathrm{cl}(Y)$. ■

It is tempting to think that 'OK, everything is pretty much the same with nets as sequences; you just replace $\leq$ by $\preceq$.' But there are some important ways in which they differ. For example, every convergent sequence is bounded, but not every convergent net is. So one needs to take a little care working with nets.

Working with nets is not the only way to extend things to general topological spaces. There is a parallel theory of **filters**. Like a topology, a filter is a collection of subsets of the space X which has certain properties. These have the advantage of not requiring one to condider possibly strange index sets, but the disadvantage that they look less like sequences. If you want to see more about this you might consult R. G. Bartle, Nets and Filters In Topology, *American Mathematical Monthly*, Vol. 62 (1955), 551–557.

Given a topological space (X, τ), it is not always easy to tell whether it is possible to find some metric d on X which generates the original topology τ. Knowing that such a metric exists is helpful even if you can't actually say what it is, since in that case we would know that the space was Hausdorff and that we can reach any point in the closure of any set using a sequence rather than a net. There are many general theorems about when a topological space is **metrizable**. We'll shortly look at the weak and pointwise convergence topologies. These are both Hausdorff, but not metrizable, although we won't directly prove this last property.

4.4 Bases for Topologies

So far all the non-metric topologies that we have shown are rather artificial, and don't really give very interesting notion of convergence. In practice, what analysts do is write down what they want convergence to mean, and then show that this does indeed correspond to convergence in some topological space. What is very difficult in practice is to say 'the open sets in this topology are the ones which have the property that ...'.

Remember that in metric spaces, open sets are actually defined in terms of a collection of simple open sets, namely open balls, which were easier to describe. As we saw, every open set in a metric space can be written as a union of open balls

$$U = \bigcup_{x \in U} B(x, \epsilon_x).$$

Perhaps more importantly, although the the open balls do not themselves form a topology, they are enough to determine convergence, since a sequence eventually gets inside any open set around x if and only if it eventually gets in side any open ball centred at x.

It turns out that this is a useful idea more generally. Instead of describing exactly what sets are in a topology τ we'll instead describe a simpler collection $\mathcal{G} \subseteq \tau$ of subsets of X which is nonetheless large enough to completely determine τ-convergence. That is, $x_\alpha \xrightarrow{\tau} x \iff x_\alpha$ eventually ends up in every set $G \in \mathcal{G}$ which contains x. The trick is to choose $\mathcal{G}$ the right size. If you make $\mathcal{G}$ too big (eg $\mathcal{G} = \tau$) then it can be hard to describe the elements, and checking convergence is harder. If you make $\mathcal{G}$ too small (eg $\mathcal{G} = \{\emptyset, X\}$) then you'll lose the $\Leftarrow$ implication. What we want is for $\mathcal{G}$ to be just big enough to generate all the open sets of the topology.

The first step is to say what it means for $\mathcal{G}$ to be big enough to generate all of τ in a certain sense.

Definition 4.4.1. Let (X, τ) be a topological space. A family $\mathcal{G} \subseteq \tau$ is a **(topological) base** for (X, τ) if every open set $U \in \tau$ can be written as some arbitrary union of elements of $\mathcal{G}$.

Note that we allow an empty union giving the empty set! Many authors use the word 'basis' rather than base, but as I might want to talk about a basis for a Banach space, I'd rather not overburden that term.

Example 4.4.2.

1. The set of all open balls in a metric space is a topological base for the topology generated by the metric.
2. In the discrete topology $\tau = \mathcal{P}(X)$, the singleton sets form a topological base.

It turns out that there is an easy test for whether a collection of sets $\mathcal{G}$ is a topological base for *some* topology on X.

Theorem 4.4.3. *Let X be a nonempty set and let $\mathcal{G} \subseteq \mathcal{P}(X)$ satisfy*

1. $\emptyset \in \mathcal{G}$.
2. *Every element of X is in at least one $G \in \mathcal{G}$, that is,* $X = \bigcup_{G \in \mathcal{G}} G$.
3. *If $G_1, G_2 \in \mathcal{G}$ and $x \in G_1 \cap G_2$ then there exists $G \in \mathcal{G}$ such that $x \in G \subseteq G_1 \cap G_2$.*

Let $\tau_{\mathcal{G}}$ be the set of all subsets of X which can be written as unions of elements in $\mathcal{G}$. Then

1. $(X, \tau_{\mathcal{G}})$ *is a topological space*
2. $\mathcal{G}$ *is a base for* $(X, \tau_{\mathcal{G}})$

Proof.

1. From conditions 1 and 2 in the theorem it is clear that $\emptyset, X \in \tau_{\mathcal{G}}$.
2. Let $U_1, \ldots, U_n \in \tau_{\mathcal{G}}$ and that $U = \bigcap_{k=1}^n U_k$. (Assume, to avoid the trivial case, that $n \geq 2$.) For $k = 1, \ldots, n$ we have now that U_k can be written as a union of elements of $\mathcal{G}$, say
$$U_k = \bigcup_{\beta \in B_k} G_\beta \qquad G_\beta \in \mathcal{G}.$$
Suppose then that $x \in \bigcap_{k=1}^n U_k$. Since x is in U_1 and U_2, there exist $\beta_1 \in B_1$ and $\beta_2 \in B_2$ so that $x \in G_{\beta_1} \cap G_{\beta_2}$. From condition 3 then there exists $G_2 \in \mathcal{G}$ with $x \in G_2 \subseteq G_{\beta_1} \cap G_{\beta_2}$.
We now proceed recursively: $x \in G_{\beta_3}$ for some $\beta_3 \in B_3$ so there exists $G_3 \in \mathcal{G}$ such that $x \in G_{\beta_3} \subseteq G_{\beta_3} \cap G_2$, and so on until we have that
$$x \in G_x \subseteq G_{\beta_n} \cap G_{n-1} \subseteq \bigcap_{k=1}^n G_{\beta_k} \subseteq U$$

with $G_x \in \mathcal{G}$. We can now write $U = \bigcup_{x \in U} G_x$ as a union of elements of $\mathcal{G}$ and hence $\bigcap_{k=1}^n U_k \in \tau_{\mathcal{G}}$.

3. Suppose that $\{U_\alpha\}_{\alpha \in A} \subseteq \tau_{\mathcal{G}}$. For all $\alpha \in A$ we have that
$$U_\alpha = \bigcup_{\beta \in B_\alpha} G_{\alpha,\beta} \qquad G_{\alpha,\beta} \in \mathcal{G}$$
Hence
$$\bigcup_{\alpha \in A} U_\alpha = \bigcup_{\alpha \in A} \bigcup_{\beta \in B_\alpha} G_{\alpha,\beta} \in \tau_{\mathcal{G}}$$

Thus, $\tau_{\mathcal{G}}$ defines a topology on X. That $\mathcal{G}$ is a base for $\tau_{\mathcal{G}}$ is immediate from its definition. ■

Example 4.4.4. A topology will generally have lots of bases. Let τ denote the usual open sets on $\mathbb{R}$. We saw that the open intervals (plus $\emptyset$) form a useful base for τ. Choosing $\mathcal{G} = \tau$ works too, but is less useful.

Sometimes it helps to make $\mathcal{G}$ rather small. The collection
$$\mathcal{G}_{\mathbb{Q}} = \{B(x,r) \,:\, x \in \mathbb{Q}, r \in \mathbb{Q}^+\} \cup \{\emptyset\}$$
is a base for τ. What makes this interesting is that $\mathcal{G}_{\mathbb{Q}}$ is a countable collection of sets, whereas the base of all open balls is not. The existence of a countable base will turn out to be important — you can't always do this.

The next proposition confirms that convergence is always determined by any base for the topology.

Proposition 4.4.5. *Let $\mathcal{G}$ be a base for a topological space (X, τ). A net $\{x_\alpha\}_{\alpha \in A}$ converges to x if and only if*

($*$) *for all $U \in \mathcal{G}$ containing x there exists α_0 such that for all $\alpha \succeq \alpha_0$ we have $x_\alpha \in U$.*

Proof. ($\Rightarrow$) Trivial.

($\Leftarrow$) Suppose that ($*$) holds and V is a neighbourhood of x. As $\mathcal{G}$ is a base for τ then there is a collection $\{G_\beta\}_{\beta \in B} \subseteq \mathcal{G}$ such that $V = \bigcup_{\beta \in B} G_\beta$. As $x \in V$, there exists β_0 such that $x \in G_{\beta_0} \subseteq V$. By ($*$), there exists α_0 such that for all $\alpha \succeq \alpha_0$ we have $x_\alpha \in G_{\beta_0}$.

Clearly then, for all $\alpha \succeq \alpha_0$ we have $x_\alpha \in V$. That is $x_\alpha \xrightarrow{\tau} x$. ■

In practice, we define most topologies via Theorem 4.4.3. That is, we describe a base $\mathcal{G}$ satisfying the conditions of Theorem 4.4.3 and then say that τ is the corresponding topology generated by that base. A more detailed description of a general open set in τ is rarely needed.

Exercise 4.4.6. Suppose that $\mathcal{G}$ is a base for a topology τ. Prove that (X, τ) is Hausdorff if and only if for all distinct $x, y \in X$, there exist disjoint G_x and G_y in $\mathcal{G}$ with $x \in G_x$ and $y \in G_y$. (Thus we don't need to know the full topology to test whether the space if Hausdorff.)

4.5 The weak topology

We are now in a position to introduce some topologies which are useful to analysts, which are not generated by a metric. In the next section we'll give a base for the topology of pointwise convergence. First though we'll look at something known as the weak topology. The norm topology on ℓ^2 lacks some of the nice properties that the norm topology has on $\mathbb{R}^n$. For example, a bounded sequence in ℓ^2 may fail to have any convergent subsequences. It will turn out that this topology recovers some of these good properties.

Let X be an inner product space with inner product $\langle \cdot, \cdot \rangle$. In many applications one discovers not that $\|x_n - x\| \to 0$ but that $\langle x_n, y \rangle \to \langle x, y \rangle$ for all $y \in X$. This is called **weak convergence**.

Why 'weak convergence'? Because it is weaker than the usual convergence. Suppose, for example, that $X = \ell^2$. Note that if $\|\mathbf{x}_n - \mathbf{x}\| \to 0$ in ℓ^2, then, by Cauchy-Schwarz

$$|\langle \mathbf{x}_n, \mathbf{y} \rangle - \langle \mathbf{x}, \mathbf{y} \rangle| = |\langle \mathbf{x}_n - \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}_n - \mathbf{x}\|_2 \|\mathbf{y}\|_2 \to 0$$

and so $\mathbf{x}_n \to \mathbf{x}$ weakly. The reverse implication fails however. Let $\mathbf{e}_n$ be the sequence with a one in the nth spot and zeros elsewhere. The sequence $\{\mathbf{e}_n\}_{n=1}^{\infty}$ is certainly not Cauchy so it can't converge in norm. On the other hand, if $\mathbf{y} = (y_1, y_2, \dots) \in \ell^2$ then

$$\langle \mathbf{e}_n, \mathbf{y} \rangle = \sum_{k=1}^{\infty} e_{n,k} y_k = y_n \to 0, \qquad \text{as } n \to \infty.$$

(Why?) Thus if $\mathbf{x} = (0, 0, \dots)$ then $\langle \mathbf{e}_n, \mathbf{y} \rangle \to \langle \mathbf{x}, \mathbf{y} \rangle$ and so $\mathbf{e}_n \to \mathbf{x}$ weakly.

Question 4.5.1. Does this convergence behave nicely? Are limits unique? Does it come from a metric? Or a topology? A Hausdorff topology?

We will define the topology by giving a suitable base.

Definition 4.5.2. Let X be an inner product space (real or complex) with inner product $\langle \cdot, \cdot \rangle$. Suppose that $\mathbf{x}, \mathbf{u} \in X$ and that $r > 0$. The set

$$S(\mathbf{x}, \mathbf{u}, r) = \{\mathbf{v} \in X \,:\, |\langle \mathbf{v} - \mathbf{x}, \mathbf{u} \rangle| < r\}$$

is called a **slice** centred at $\mathbf{x}$, orthogonal to $\mathbf{u}$.

Example 4.5.3. Let $X = \mathbb{R}^2$ with the usual dot product. For fixed $\mathbf{x}, \mathbf{u}$, the equation

$$\langle \mathbf{v} - \mathbf{x}, \mathbf{u} \rangle = 0 \qquad \text{or} \qquad (\mathbf{v} - \mathbf{x}) \cdot \mathbf{u} = 0 \tag{4.5.1}$$

is just the point-normal form of the line through $\mathbf{x}$ with normal $\mathbf{u}$.

For $c \in \mathbb{R}$, the equations $(\mathbf{v} - \mathbf{x}) \cdot \mathbf{u} = c$ describe a family of parallel lines, all orthogonal to $\mathbf{u}$. A slice

$$S(\mathbf{x}, \mathbf{u}, r) = \{\mathbf{v} \in X \,:\, |\langle \mathbf{v} - \mathbf{x}, \mathbf{u} \rangle| < r\}$$

is just an open strip with the line (4.5.1) down the middle. You can check that the width of the strip is $2r / \|\mathbf{u}\|$.

Example 4.5.4. Let $X = \mathbb{R}^3$. Now the equation $\langle \mathbf{v} - \mathbf{x}, \mathbf{u} \rangle = c$ describes a plane which is orthogonal to $\mathbf{u}$ and a slice set is the open region between to parallel planes.

In $\mathbb{R}^n$ (or indeed in any inner product space) a slice set is the open region between two parallel **hyperplanes**. In ℓ^2 this can be a little harder to imagine.

Let $\mathbf{u} = (1, 0, 0, \dots) \in \ell^2$ and let $\mathbf{x} = (1, \frac{1}{2}, \frac{1}{3}, \dots) \in \ell^2$. Suppose that $\mathbf{v} = (v_1, v_2, v_3, \dots) \in \ell^2$ satisfies

$$|\langle \mathbf{v} - \mathbf{x}, \mathbf{u} \rangle| < r.$$

Expanding what this means

$$|\langle (v_1 - 1, v_2 - \frac{1}{2}, v_3 - \frac{1}{3}, \dots) \, , \, (1, 0, 0, \dots) \rangle| = |(v_1 - 1) \cdot 1| = |v_1 - 1| < r.$$

The sets $\{\mathbf{v} : v_1 = 1\}$ and $\{\mathbf{v} : v_1 = 2\}$ are hyperplanes in ℓ^2, and the slice $S(\mathbf{x}, \mathbf{u}, r)$ is the region between these hyperplanes.

The set of slices does not form a base for a topology, even in $\mathbb{R}^2$. To get a base you need to take intersections of slices.

In $\mathbb{R}^2$, how can two slices $S(\mathbf{x}_1, \mathbf{u}_1, r_1)$ and $S(\mathbf{x}_2, \mathbf{u}_2, r_2)$ intersect? If $\mathbf{u}_1$ and $\mathbf{u}_2$ are parallel, the slices might be disjoint, or else their intersection is a slice of the form $S(\mathbf{x}', \mathbf{u}_1, r')$. If $\mathbf{u}_1$ and $\mathbf{u}_2$ are not parallel then the slices must intersect in a more interesting way to give an open parallelogram. Intersecting with a third slice could produce a hexagon or several other shapes.

Exercise 4.5.5. Let $X = \mathbb{R}^2$ and let[3]

$$\mathcal{G} = \{\emptyset\} \cup \{S(\mathbf{x}_1, \mathbf{u}_1, r_1) \cap S(\mathbf{x}_2, \mathbf{u}_2, r_2) \, : \, \mathbf{x}_1, \mathbf{x}_2, \mathbf{u}_1, \mathbf{u}_2 \in \mathbb{R}^2, \ r_1, r_2 > 0\}.$$

Show that $\tau_{\mathcal{G}}$ is just the usual topology on $\mathbb{R}^2$.

In $\mathbb{R}^3$, the intersection of two slices is never a bounded set, but the intersection of three may produce a parallelepiped. In this space, the set of all intersections of three slices is a base for the usual topology on $\mathbb{R}^3$ (but intersections of just two slices is not a base — why?).

Things are different in infinite dimensions! There the idea is not to pick some fixed finite number of slices to intersect, but to allow intersections of an arbitrary, but finite, number of slices.

Definition 4.5.6. Given an inner product space, we define the collection of sets

$$\mathcal{G}_w = \{\emptyset\} \cup \{ \text{ finite intersections of slices } \}.$$

It is easy to check that this satisfies the hypotheses of Theorem 4.4.3, and hence that it is the base for a topology τ_w, called the **weak topology**.

[3]It is worth noting that explicitly including the empty set here is not really necessary since two slices could be disjoint. Also, $\mathcal{G}$ includes all the slices themselves as we could take the two slices we are intersecting to be the same!

As we saw above, in $\mathbb{R}^n$, the weak topology is just the usual topology. In ℓ^2, the intersection of a finite number of slices never produces a bounded set. (Think of what we saw in Example 4.5.4 above. The slice bounds what can happen in the first coordinate, but there are infinitely many coordinates on which we have not restriction.) This means that any union of sets in $\mathcal{G}_w$ must be unbounded and in particular, it can't be a ball.

It remains to check that convergence in the weak topology is actually weak convergence!

Proposition 4.5.7. *Let X be an inner product space, let $\{x_\alpha\}_{\alpha \in A}$ be a net in X and suppose that $x \in X$. Then $x_\alpha \to x$ in the weak topology if and only if $\langle x_\alpha, y\rangle \to \langle x, y\rangle$ for all $y \in X$.*

Proof. ($\Rightarrow$) Suppose that $x_\alpha \to x$ in the weak topology. Fix $y \in X$ and suppose that $\epsilon > 0$. Consider the slice

$$V = S(x, y, \epsilon) = \{v \in X \,:\, |\langle v - x, y\rangle| < \epsilon\}.$$

As $V \in \tau_w$ there exists $\alpha_0 \in A$ such that for all $\alpha \succeq \alpha_0$, $x_\alpha \in V$. But this just says that for $\alpha \succeq \alpha_0$,

$$|\langle x_\alpha, y\rangle - \langle x, y\rangle| < \epsilon$$

which tells us that $\langle x_\alpha, y\rangle \to \langle x, y\rangle$.

($\Leftarrow$) Suppose that $\langle x_\alpha, y\rangle \to \langle x, y\rangle$ for all $y \in X$. Suppose that $V \in \mathcal{G}_w$, that is

$$V = \bigcap_{k=1}^{n} S(w_k, u_k, r_k),$$

is a basic open set containing x. Fix $k \in \{1, 2, \dots, n\}$. As $x \in V$, $|\langle x - w_k, u_k\rangle| < r_k$. Let $\epsilon_k = r_k - |\langle x - w_k, u_k\rangle| > 0$. Now (taking $y = u_k$), we know that there exists $\alpha_k \in A$ such that for all $\alpha \succeq \alpha_k$, $|\langle x_\alpha, u_k\rangle - \langle x, u_k\rangle| < \epsilon_k$. Using the triangle inequality, it follows that, for $\alpha \succeq \alpha_k$,

$$|\langle x_\alpha, u_k\rangle - \langle w_k, u_k\rangle| \le |\langle x_\alpha, u_k\rangle - \langle x, u_k\rangle| + |\langle x, u_k\rangle - \langle w_k, u_k\rangle| < r_k$$

and so $x_\alpha \in S(w_k, u_k, r_k)$.

Since A is a directed set, we can find $\alpha_0 \in A$ which is larger than each of $\alpha_1, \dots, \alpha_n$. If $\alpha \succeq \alpha_0$, then $\alpha \succeq \alpha_k$ for each k and hence by the last paragraph $x_\alpha \in V$.

Since the net $\{x_\alpha\}$ eventually gets and remains inside any basic open set containing x, it follows from Proposition 4.4.5 that $x_\alpha \to x$ in the weak topology ■

Note that given any two distinct points x, y in an inner product space X, for small enough r, the slices $S(x, x - y, r)$ and $S(y, x - y, r)$ are disjoint. This implies that the weak topology is Hausdorff and hence that limits are unique in the weak topology.

We'll see soon that the weak topology has many nice properties not possessed by the norm topology.

Exercise 4.5.8. Fill in the details of the above claims about what happens in $\mathbb{R}^n$: Prove that the distinction between the weak and norm topologies occurs only in infinite dimensions. That is, show that if if X is a finite dimensional inner product space and $x_n \to x$ weakly, then $x_n \to x$ in norm.

We proved above that the weak topology is different to the usual norm topology on ℓ^2. What we haven't proven is that there is no metric d on ℓ^2 which generates τ_w. If one can fine a metric d on a topological space (X, τ) so that τ is just the topology generated by d, then we say that (X, τ) is a **metrizable** topological space. Even if you can't say what the metric is, this can be a good thing to know. For example, you would know that every point in the closure of a set was the limit of a sequence in the set, rather than a net. We won't prove it, but the weak topology on ℓ^2 is *not* metrizable.

4.6 Pointwise convergence

Another standard type of convergence is pointwise convergence of a net of functions on a set Ω: the net $\{f_\alpha\}_{\alpha \in A}$ converges pointwise to f if for all $\omega \in \Omega$, $f_\alpha(\omega) \to f(\omega)$. This also comes from a topology!

For notational convenience let's do this in $X = \mathcal{B}(\Omega)$, but you could do this for other sets of functions (such as $C[0,1]$).

For a fixed $f \in X$, finite subset $F \subseteq \Omega$ and $\epsilon > 0$, the set

$$G(f, F, \epsilon) = \{g \in X \, : \, |f(x) - g(x)| < \epsilon \quad \forall x \in F\}$$

is called a **gate set**. This is because it consists of all functions[4] g whose graph passes through each 'gate' of width 2ϵ around $f(x_k)$ at x_k.

Let

$$\mathcal{G}_G = \{\emptyset\} \cup \{G(f, F, \epsilon) \, : \, f, F, \epsilon \text{ as above.}\}$$

Exercise 4.6.1. Prove that $\mathcal{G}_G$ satisfies the hypotheses of Theorem 4.4.3 and hence is the base for a topology.

We might call the topology generated by $\mathcal{G}_G$ the **gate set topology**, although this name is not standard.

Note that this topology is Hausdorff: if $f_1 \neq f_2$ then there is some point $x \in \Omega$ such that $f_1(x) \neq f_2(x)$. Let $\epsilon = |f_1(x) - f_2(x)|/2$. Then $f_1 \in G(f_1, \{x\}, \epsilon)$ and $f_2 \in G(f_2, \{x\}, \epsilon)$, but the two gate sets are disjoint.

Proposition 4.6.2. *Convergence in the gate set topology is pointwise convergence.*

The proof is almost identical to the corresponding proof for the weak topology. You should have a go at adapting it.

[4]Compare this to an ϵ-ball around f in the d_∞ norm. There an element needs to have a graph which passes through gates at **every** point in Ω.

4.7 Comparing topologies

There are typically many topologies that can be used in a given setting. At one end there is the tiny indiscrete topology $\{\emptyset, X\}$ and at the other end the huge discrete topology $\mathcal{P}(X)$. The useful ones lie somewhere in the middle!

Definition 4.7.1. Suppose that τ_1 and τ_2 are topologies on a set X. We say that τ_1 is **weaker** than τ_2 if $\tau_1 \subseteq \tau_2$. That is, if every τ_1-open set is also τ_2-open.

Thus, in a metric space, the indiscrete topology is weaker than the metric space topology is weaker than the discrete topology. Of course, it is not automatic that given two topologies, one is always a subset of the other. However, most of the standard topologies that we use can be compared in this way.

Proposition 4.7.2. *Suppose that τ_1 and τ_2 are topologies on a set X. Then τ_1 is weaker than τ_2 then,*

$(*)$ $x_\alpha \to x$ *in* $\tau_2 \implies x_\alpha \to x$ *in* τ_1.

Proof. Suppose that τ_1 is weaker than τ_2. If $x_\alpha \to x$ in τ_2, then 'eventually' x_α gets inside every τ_2-open set. But every τ_1-open set is τ_2-open so this implies that $x_\alpha \to x$ in τ_1. ■

What is true is that given two topologies τ_1 and τ_2, you can form a new topology

$$\tau = \tau_1 \cap \tau_2$$

which is clearly weaker than either.

Exercise 4.7.3. Prove that if τ_α, $\alpha \in A$ is any collection of topologies on a set X, that

$$\tau = \bigcap_{\alpha \in A} \tau_\alpha$$

is a topology on X.

A collection $\mathcal{T}$ of subsets of X clearly need not be a topology, nor even a base for a topology. It is however always a subset of at least one topology, namely the discrete topology $\mathcal{P}(X)$. It makes sense therefore to look at the collection of all topologies that contain $\mathcal{T}$ and then take the intersection of that collection. This results in what is called the weakest (or smallest) topology containing $\mathcal{T}$ or the topology generated by $\mathcal{T}$.

Example 4.7.4. Let $X = \mathbb{R}^2$. Then

$$\begin{aligned} \tau_1 &= \{U \times \mathbb{R} \,:\, U \text{ is open in } \mathbb{R}\}, \\ \tau_2 &= \{\mathbb{R} \times U \,:\, U \text{ is open in } \mathbb{R}\} \end{aligned}$$

are both topologies on X. Neither is contained in the other. Their intersection is just the indiscrete topology. Their union is not a topology, nor a base for a topology. The weakest topology which contains both τ_1 and τ_2, is the usual metric topology on X.

The trick in analysis is often to choose a topology with just the right properties. Too weak and you become non-Hausdorff and limits are not unique. Too strong and hardly anything converges.

4.8 Continuity

Definition 4.8.1. Suppose $(X, \tau_X), (Y, \tau_Y)$ are topological spaces, and $f : (X, \tau_X) \to (Y, \tau_Y)$. We say that f is **continuous on** X if $f^{-1}(U) \in \tau_X$ for all $U \in \tau_Y$.

This of course matches up with what happens in metric spaces!

Continuity then depends crucially on the two topologies. If τ_X is big and τ_Y is small, then it is easy for functions to be continuous. If τ_X is small and τ_Y is big, then it is hard for functions to be continuous.

Example 4.8.2.

1. If $\tau_X = \mathcal{P}(X)$, then every function is continuous, whatever topology you apply to Y!
2. If $X = Y = \mathbb{R}$, $\tau_X = \{\emptyset, \mathbb{R}\}$ and $\tau_Y = \mathcal{P}(\mathbb{R})$ then then only constant functions are continuous.

In many situations, you have a collection of functions $\mathcal{F}$ and you construct a topology specifically to make sure that all the functions in $\mathcal{F}$ are continuous!

Example 4.8.3. Suppose that $\mathcal{F}$ is a nonempty collection of functions mapping $\Omega \to \mathbb{R}$. Let

$$\mathcal{T} = \{V \subseteq \Omega \,:\, V = f^{-1}(U) \text{ for some open } U \subseteq \mathbb{R} \text{ and } f \in \mathcal{F}\}.$$

A common construction to let τ be the topology generated by $\mathcal{T}$, that is, the weakest topology which makes all the functions in $\mathcal{F}$ continuous. As noted above, there is always at least one topology $\mathcal{P}(X)$ which makes them continuous.

Example 4.8.4. More specifically, let $\mathcal{F}$ denote the set of all functions $\phi : \ell^2 \to \mathbb{R}$ of the form

$$\phi(\mathbf{x}) = \phi_{\mathbf{y}}(x) := \langle \mathbf{x}, \mathbf{y} \rangle$$

for some $\mathbf{y} \in \ell^2$. It is clear that all the functions in $\mathcal{F}$ are (norm) continuous and linear. Less obvious is that every continuous linear map from $(\ell^2, \|\cdot\|_2)$ to $\mathbb{R}$ is of this form. The weak topology constructed in Section 4.5 is the weakest topology which makes all the elements of $\mathcal{F}$ continuous.

Exercise 4.8.5. Let $\Omega = \mathbb{R}$ and let $\mathcal{F}$ be the set of continuous (with the usual topology!) even functions on $\mathbb{R}$. Is there a weaker topology you can impose on the domain Ω under which the even functions are still all continuous?

Question 4.8.6. What about continuity at a point?

Definition 4.8.7. Let (X, τ) be a topological space and suppose that $x \in X$. The set of all neighbourhoods[5] of x is denoted $\mathcal{N}_x$.

[5] Remember: neighbourhood = open set containing x

Definition 4.8.8. Suppose $f : (X, \tau_X) \to (Y, \tau_Y)$ and that $x \in X$. Then we say that f is **continuous at** $x \in X$ if for all $U \in \mathcal{N}_{f(x)} \subseteq \tau_Y$ we have that there exists $V \in \tau_X$ such that $V \subseteq f^{-1}(U)$.

This is perhaps better expressed in words: *f is continuous at x if the inverse image of every neighbourhood of $f(x)$ contains some neighbourhood of x.*

Example 4.8.9. Let $f : \mathbb{R} \to \mathbb{R}$ be defined as

$$f(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0. \end{cases}$$

Let's look at a few different values of x:

$x = 2$: Let U be an open set containing $f(2)$, then if $0 \notin U$ we have $f^{-1}(U) = (0, \infty)$ and if $0 \in U$, $f^{-1}(U) = \mathbb{R}$. We can see that in either case $f^{-1}(U)$ is an open set containing 2. Hence f is continuous at 2

$x = 0$: Let $U = (-\frac{1}{2}, \frac{1}{2})$ which is an open neighbourhood of $f(0) = 0$. Then $f^{-1}(U) = (-\infty, 0]$, and there is no open set contained in $(-\infty, 0]$ that contains 0. Hence f is discontinuous at 2

Exercise 4.8.10. Prove that $f : (X, \tau_X) \to (Y, \tau_Y)$ is continuous on X if and only if it is continuous at each element of X.

For many of our standard topologies we don't describe all the open sets, but rather a base for the topology. Fortunately that is all you need to worry about for continuity (at least on the image side).

Proposition 4.8.11. *Suppose $f : (X, \tau_X) \to (Y, \tau_Y)$ and that $\mathcal{G}$ is a base for τ_Y. Then f is continuous $\iff$ $f^{-1}(G) \in \tau_X$ for all $G \in \mathcal{G}$.*

Proof.

($\Rightarrow$) Trivial.

($\Leftarrow$) Suppose that $f^{-1}(G) \in \tau_X$ for all $G \in \mathcal{G}$. If $U \in \tau_Y$, then since $\mathcal{G}$ is a base for τ_Y, U can be written in the form $U = \bigcup_{\alpha in A} G_\alpha$ with $G_\alpha \in \mathcal{G}$ for all $\alpha \in A$. Then

$$f^{-1}(U) = f^{-1}\left(\bigcup_{\alpha} G_\alpha\right) = \bigcup_{\alpha \in A} f^{-1}(G_\alpha) \in \tau_X$$

■

4.9 Homeomorphisms

At school you may have heard people talk about topology and they showed you pictures of Möbius strips and Klein bottles. That doesn't look very much like the (point set) topology we have been looking at here! In general topology is the study of properties of surfaces (or other spaces) that that are preserved under 'continuous deformations' such as stretching and bending, but not tearing or gluing. We can now make this more precise.

Definition 4.9.1. Let $(X, \tau_X), (Y, \tau_Y)$ be topological spaces. If $f : X \to Y$ is a bijection, then $f^{-1} : Y \to X$. If both f and f^{-1} are continuous, then we say that

1. f is a **homeomorphism**
2. (X, τ_X) and (Y, τ_Y) are **homeomorphic** topological spaces.

Note that the definition ensures that the homeomorphism maps τ_X to τ_Y bijectively.

From a topological point of view, homeomorphic spaces are regarded as being indistinguishable since

$$\begin{aligned} U \in \tau_X \quad &\iff \quad f(U) \in \tau_Y \\ f^{-1}(V) \in \tau_X \quad &\iff \quad V \in \tau_Y \\ x_\alpha \xrightarrow{\tau_X} x \quad &\iff \quad f(x_\alpha) \xrightarrow{\tau_Y} f(x). \end{aligned}$$

The previous paragraph is perhaps a little dishonest. One might better say that a property is topological if it is invariant under homeomorphisms.

Example 4.9.2. $(0, 1)$ and $\mathbb{R}$ are homeomorphic via the map $f : (0, 1) \to \mathbb{R}$,

$$f(x) = \tan\left[\pi\left(x - \frac{1}{2}\right)\right].$$

(Note that the sequence $\{\frac{1}{k}\}_{k=2}^{\infty}$ is Cauchy in $(0, 1)$, but its image under f is not Cauchy in $\mathbb{R}$. So being Cauchy is not a topological property!)

Example 4.9.3. $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ and $\mathbb{C}$ are homeomorphic via $f : \mathbb{D} \to \mathbb{C}$

$$f(z) = \frac{z}{1 - |z|}.$$

Example 4.9.4. $\mathbb{T}$ and $I = [0, 1]$ are not homeomorphic.

This is despite the facts that

- there exists a bijection $f : \mathbb{T} \to I$ since they have the same cardinality.
- there exists a continuous onto map $f : \mathbb{T} \to I$, eg $z \mapsto \frac{1}{2}(\operatorname{Re} z + 1)$.

- there exists a continuous onto map $f : I \to \mathbb{T}$, eg $x \mapsto e^{10ix}$.

Proof: Suppose $f : \mathbb{T} \to [0,1]$ is a homeomorphism.

Let $h : [0,1] \to \mathbb{T}$, $h(x) = (f^{-1}(0))^{1-x}\,(f^{-1}(1))^x$. The image of h is the arc of the circle going from $f^{-1}(0)$ to $f^{-1}(1)$ (and is certainly not the whole circle!).

Now $f \circ h : I \to I$ is such that $f \circ h(0) = f(f^{-1}(0)) = 0$, and $f \circ h(1) = f(f^{-1}(1)) = 1$. Clearly f, h are continuous so the Intermediate Value Theorem implies that $f \circ h$ is onto.

Pick a $z \in \mathbb{T}$ which is not in the arc $h(I)$ and let $a = f(z) \in [0,1]$. As $f \circ h$ is onto, there exists $x \in [0,1]$ such that $f \circ h(x) = a$. Thus, if $w = f(x) \in h(I)$ then $f(w) = a = f(z)$ contradicting the fact that f is one-to-one.

Exercise 4.9.5. (Easy!) Prove that being homeomorphic is an equivalence relation on any family of topological spaces. In particular, if (X, τ_X) is homeomorphic to (Y, τ_Y), and (Y, τ_Y) is homeomorphic to (Z, τ_Z) then (X, τ_X) is homeomorphic to (Z, τ_Z).

In the next chapter we will look at some of these 'topological properties', that is, properties that are preserved under homeomorphisms. These include things like compactness and connectedness. It also includes things like 'the number of holes' in a space, but that is rather more complicated to make precise!

4.10 Homotopy

A closed curve in $\mathbb{C}$ can be considered as the image of a continuous function $\gamma : \mathbb{T} \to \mathbb{C}$. In complex analysis you have things like the Cauchy integral formula

$$f(a) = \frac{1}{2\pi i} \int_\gamma \frac{f(z)}{z-a}\, dz$$

for a suitable curve which loops once around the point a in an anticlockwise direction. We often talk there about deforming one curve γ to another γ' without changing the value of the integral.

It is worth rephrasing all this in our current language. To make things simpler, let's take $a = 0$. A closed curve around 0 is then a continuous function from $\mathbb{T}$ into the metric space $Y = (\mathbb{C}^*, |\cdot|)$ where $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$; that is, it is an element of $Z = C(\mathbb{T}, \mathbb{C}^*)$. Note that Z is a normed (and hence metric) space under the supremum norm.

Saying that you can 'deform' one curve γ_0 to another curve γ_1 is saying that there is a continuous map $H : Z \to Z$ with $H(0) = \gamma_0$ and $H_1 = \gamma_1$. The map H is called a **homotopy**, and we say that γ_1 and γ_2 are homotopic. Being homotopic is an equivalence relation on Z.

It turns out that there are several interesting **continuous** functions from Z to $\mathbb{Z}$. An example is

$$I(\gamma) = \frac{1}{2\pi i} \int_\gamma \frac{1}{z}\, dz$$

which gives the winding number of γ around 0. Clearly $I \circ H : Z \to \mathbb{Z}$ is continuous, so if γ_1 and γ_2 are homotopic, then (why?) they must have the same winding number.

One can do the same thing with $\mathbb{C}^*$ replaced by other topological spaces — think the sphere or the torus. Any two closed curves on a sphere are homotopic, but there are lots of nonhomotopic curves on a torus.

It is a fundamental concept in Algebraic Topology that you can make a group out of (a subset of) these curves. These homotopy groups are a topological invariant. For example the first homotopy group of the sphere in $\mathbb{R}^3$ is trivial $\{0\}$, while that for the torus is isomorphic to $\mathbb{Z}^2$. This means that the sphere and the torus can not be homeomorphic.

4.11 Problems

1. Let $X = (0, 1)$. Define
$$\tau = \{\emptyset\} \cup \{(0, t) \,:\, 0 < t \leq 1\}.$$
   (a) Prove that (X, τ) is a topological space.
   (b) Prove that a sequence $\{x_k\}_{k=1}^{\infty} \subseteq X$ converges to $\frac{1}{2}$ in (X, τ) if and only if $\limsup_k x_k \leq \dfrac{1}{2}$.
   (c) Given an example of a nonconvergent sequence $\{y_k\}_{k=1}^{\infty} \subseteq X$ in (X, τ).

2. Which of the following are topologies? What topologies are Hausdorff?
   (a) $X = \mathbb{R}$.
$$\tau = \{\emptyset, \mathbb{R}\} \cup \{(-\infty, a) \,:\, a \in \mathbb{R}\}.$$
   (b) $X = \mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$ is a fixed function.
$$\tau = \{f^{-1}(U) \,:\, U \text{ is an open set in } \mathbb{R} \text{ in the usual topology}\}.$$
   (c) $X = D = \{z \in \mathbb{C} \,:\, |z| \leq 1\}$.
$$\tau = \{U \cap D \,:\, U \text{ is an open set in } \mathbb{C} \text{ in the usual topology}\}.$$
   (d) $X = C[0, 1]$. For $r > 0$ let $B_r = \{f \in C[0, 1] \,:\, \|f\|_\infty < r\}$.
$$\tau = \{\emptyset, X\} \cup \{B_r \,:\, r > 0\}.$$
   (e) X be any set.
$$\tau = \{\emptyset\} \cup \{U \subseteq X \,:\, X \setminus U \text{ is finite}\}.$$

3. Consider the following nets. Do they converge in $(\mathbb{R}, |\cdot|)$?

   (a) $A = \mathbb{R}$ with the usual (partial) order $\leq$. The net is $\left\{\dfrac{1}{1+x^2}\right\}_{x\in A}$.

   (b) $A = \mathbb{Z}^+ \times \mathbb{Z}^+$ with partial order

   $$(n_1, m_1) \preceq (n_2, m_2) \iff n_1 \leq n_2 \text{ and } m_1 \leq m_2.$$

   The net is $x_{n,m} = \dfrac{nm}{n^2+m^2}$, $(n, m) \in A$.

   (c) The same net but with order

   $$(n_1, m_1) \trianglelefteq (n_2, m_2) \iff n_1 < n_2 \text{ or } \big(n_1 = n_2 \text{ and } m_1 \leq m_2\big).$$

4. Suppose that $\mathcal{G}_1$ and $\mathcal{G}_2$ are bases for topologies τ_1 and τ_2 on X. A $\mathcal{G}_1$ basic neighbourhood of a point $x \in X$ is any set $U \in \mathcal{G}_1$ which contains x. Suppose that for every $x \in X$ every $\mathcal{G}_1$ basic neighbourhood of x contains a $\mathcal{G}_2$ basic neighbourhood of x, and conversely that every $\mathcal{G}_2$ basic neighbourhood contains a $\mathcal{G}_1$ basic neighbourhood. Prove that $\tau_1 = \tau_2$. Prove the converse.

5. Let $X = \mathbb{R}^2$. Recall that a slice set in X is one of the form

   $$S(\mathbf{x}, \mathbf{u}, r) = \{\mathbf{v} \in X \,:\, |\langle \mathbf{v} - \mathbf{x}, \mathbf{u}\rangle| < r\}$$

   where $\mathbf{x}, \mathbf{u} \in X$ and $r > 0$.

   (a) Verify that

   $$\mathcal{G} = \{\emptyset\} \cup \{\text{ finite intersections of slices }\}.$$

   is a base for a topology τ on X.

   (b) Prove that τ is just the usual topology!

   (c) Explain, using the previous question, why when you do this on ℓ^2, you don't get the usual norm topology.

6. Prove that the gate sets form the base for a topology (see Exercise 4.6.1).

7. True or false? Suppose that τ_1 and τ_2 are two topologies on X with the property that whenever $x_\alpha \to x$ in the τ_1 topology then $x_\alpha \to x$ in the τ_2 topology. Then $\tau_2 \subseteq \tau_1$.

8. Let (X, d) be a metric space. Prove that $x_k \to x$ in (X, d) if and only if for every open set U containing x, there exists K such that for all $k \geq K$, $x_k \in U$.

9. Check that the definitions of closed, interior point, neighbourhood etc, given in this chapter are consistent with the metric space definitions we gave earlier.

10. Silly example. Let $X = \{a, b, c\}$. What are all the possible topologies on X? How many are Hausdorff?

11. Let X be a nonempty set. Let

$$\tau = \{\emptyset\} \cup \{U \subseteq X \,:\, X \setminus U \text{ is finite}\}.$$

   (a) Prove that τ is a topology on X (called the **cofinite** topology).

   (b) Consider $\mathbb{R}$ with the confinite topology and let $U = \mathbb{Z}$.

      i. Is U open, closed or neither?

      ii. Identify the interior points, boundary points, and limit points of U.

      iii. Identify the closure of U.

      iv. Consider the sequence $\{\frac{1}{k}\}_{k=1}^{\infty}$. Does this sequence converge in the cofinite topology on $\mathbb{R}$? If so, what is the limit?

12. Let X be a nonempty set, and let d_1 and d_2 be metrics on X. Recall that d_1 and d_2 are equivalent metrics if $x_k \to x$ in (X, d_1) iff $x_k \to x$ in (X, d_2). Given any metric d, the **topology generated by** d is defined to be the set τ_d of all sets which are open in the metric space (X, d). Prove that if d_1 and d_2 are equivalent metrics, then $\tau_{d_1} = \tau_{d_2}$.

13. Give examples of two **distinct** Hausdorff topologies on $\mathbb{R}^2$.

14. Consider the inner product space $\mathbb{R}^n$ with the usual dot product. Without considering all this topology machinery, check that in this case, a sequence $\{\mathbf{x}_k\}$ converges to $\mathbf{x}$ in norm if and only if it converges weakly to $\mathbf{x}$ (that is, if $\mathbf{x}_k \cdot \mathbf{y} \to \mathbf{x} \cdot \mathbf{y}$ for all $\mathbf{y} \in \mathbb{R}^n$). So in some inner product spaces, the weak topology is just the same as the norm topology.

15. Two standard infinite dimensional normed vector spaces are c_0 and ℓ^1:

$$c_0 = \{(x_1, x_2, \dots) \,:\, \lim_{n\to\infty} x_n = 0\}$$

under the norm $\|(x_1, x_2, \dots)\|_\infty = \sup_n |x_n|$, and

$$\ell^1 = \{(x_1, x_2, \dots) \,:\, \sum_{n=1}^{\infty} |x_n| < \infty\}$$

under the norm $\|(x_1, x_2, \dots)\|_1 = \sum_{n=1}^{\infty} |x_n|$.

   (a) Prove that if $\mathbf{x} = (x_n)_{n=1}^{\infty} \in c_0$ and $\mathbf{y} = (y_n)_{n=1}^{\infty} \in \ell^1$, then $\langle \mathbf{x}, \mathbf{y} \rangle := \sum_{n=1}^{\infty} x_n y_n$ converges, and that in fact

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \le \|\mathbf{x}\|_\infty \|\mathbf{y}\|_1 .$$

(b) Let $\{\mathbf{x}_k\}_{k=1}^{\infty}$ be a sequence in c_0. (That is, $\{\mathbf{x}_k\}_{k=1}^{\infty}$ is a sequence of sequences!) We say that $\{\mathbf{x}_k\}$ converges **weakly** to $\mathbf{x} \in c_0$ if, for **every** $\mathbf{y} \in \ell^1$,

$$\langle \mathbf{x}_k, \mathbf{y} \rangle \to \langle \mathbf{x}, \mathbf{y} \rangle \qquad \text{(that is, } \langle \mathbf{x}_k - \mathbf{x}, \mathbf{y} \rangle \to 0\text{)}.$$

Suppose that $\{\mathbf{x}_k\}$ converges to $\mathbf{x}$ in the metric that comes from the above norm on c_0 (that is, $\|\mathbf{x}_k - \mathbf{x}\|_\infty \to 0$). Prove that $\{\mathbf{x}_k\}$ converges weakly to $\mathbf{x}$ as well.

(c) Give an example of a sequence of elements in c_0 which converges weakly, but not in the norm metric.

16. Let $A = \mathbf{Z}^+ \times \mathbf{Z}^+$. Define the relation $\trianglelefteq$ on $A \times A$ by

$$(m, n) \trianglelefteq (m', n') \qquad \iff \qquad (m < m') \text{ or } (m = m' \text{ and } n \leq n').$$

(a) Prove that $(A, \trianglelefteq)$ is a directed set.

(b) For $(m, n) \in A$ let $x_{m,n} = \dfrac{m^2}{m^2 + n^2}$ and let $y_{m,n} = \dfrac{m^2 + n^2 + n}{m^2 + n^2}$. Discuss (with proof) the convergence (or otherwise) of the nets $\{x_{m,n}\}_{(m,n)\in A}$ and $\{y_{m,n}\}_{(m,n)\in A}$.

17. Which of the following families are bases for topologies on $\mathbb{R}^2$?

(a) $\mathcal{G}_1 = \{\emptyset\} \cup \{(a, b) \times (c, d) \,:\, a < b \text{ and } c < d\}$.

(b) $\mathcal{G}_2 = \{\emptyset\} \cup \{E(x_0, y_0, a, b) \,:\, x_0, y_0 \in \mathbb{R} \text{ and } a, b > 0\}$ where

$$E(x_0, y_0, a, b) = \{(x, y) \,:\, \frac{(x - x_0)^2}{a^2} + \frac{(y - y_0)^2}{b^2} < 1\}.$$

(c) $\mathcal{G}_3 = \{\emptyset\} \cup \{\overline{B(\mathbf{x}_0, r)} \,:\, \mathbf{x}_0 \in \mathbb{R}^2,\ r > 0\}$ where

$$\overline{B(\mathbf{x}_0, r)} = \{\mathbf{x} \,:\, \|\mathbf{x} - \mathbf{x}_0\| \leq r\}.$$

18. A topological space (X, τ) is said to be a **first countable** if for all $x \in X$, there exists a **countable** family of open neighbourhoods of x, say $\mathcal{F}$, such that given any open neighbourhood U of x, there exists $F \in \mathcal{F}$ such that $F \subseteq U$. Prove that if (X, τ) is first countable and x is a limit point of a set $Y \subseteq X$, then there exists a **sequence** $\{x_j\}_{j=1}^{\infty} \subseteq Y$ such that $x_j \to x$.

19. A topological space (X, τ) is said to be a **second countable** if (X, τ) has a **countable** base $\mathcal{G}$. A space with a countable dense subset is said to be **separable**.

(a) Prove that if (X, τ) is second countable, then X is separable.

(b) Prove that if a metric space (X, d) is separable, then (X, d) is second countable.

(c) Prove that ℓ^2 is separable.

(d) Prove that the space of all bounded sequences ℓ^∞ with the norm

$$\|(x_1, x_2, \dots)\|_\infty = \sup_n |x_n|$$

is not separable.

20. Let X be a set and suppose that τ_1 and τ_2 are topologies on X. If $\tau_1 \subseteq \tau_2$ then we say that τ_1 is **weaker** than τ_2, or that τ_2 is **stronger** than τ_1.

   (a) Is $\tau = \tau_1 \cap \tau_2$ a topology on X?

   (b) What if you had more than two topologies?

   (c) Suppose that $\sigma \subseteq \mathcal{P}(X)$. Explain why it makes sense to talk about the 'weakest' topology on X that contains σ.

21. Let $X = M_n(\mathbb{R})$ with the operator norm $\|T\|_{op} = \sup_{\|\mathbf{x}\|=1} \|T\mathbf{x}\|$. We say that a sequence $\{T_n\} \subseteq X$ converges to $T \in X$ in the **strong operator topology** if $T_n\mathbf{x} \to T\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$. (Remember that these matrices are really just linear transformations $T : \mathbb{R}^n \to \mathbb{R}^n$, that is, functions defined on $\mathbb{R}^n$. Thought of this way, SOT convergence is just pointwise convergence, so it definitely comes from a topology!) We say that $\{T_n\} \subseteq X$ converges to $T \in X$ in the **weak operator topology** if $\langle T_n\mathbf{x}, \mathbf{y}\rangle \to \langle T\mathbf{x}, \mathbf{y}\rangle$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Prove that

$$T_n \xrightarrow{\text{norm}} T \quad \Longrightarrow \quad T_n \xrightarrow{\text{SOT}} T \quad \Longrightarrow \quad T_n \xrightarrow{\text{WOT}} T.$$

[Actually, in this setting, these senses of convergence are all equivalent. Later, we'll replace $\mathbb{R}^n$ with ℓ^2 and the $n \times n$ matrices with the continuous linear maps from ℓ^2 to ℓ^2. The proof you give here will still work, but in that setting you end up with three distinct, and useful topologies on this set of linear maps]

22. Let X be a set and let (Y, τ_Y) be a topological space. Let $\mathcal{F}$ be a nonempty family of functions from $X \to Y$.

   (a) Is there a topology on X which makes all the functions in $\mathcal{F}$ continuous? Show in fact that there a weakest topology $\tau_{\mathcal{F}}$ on X which makes all the functions in $\mathcal{F}$ continuous?

   (b) Let $X = Y = \mathbb{R}$ with Y having its usual topology. Let $\mathcal{F} = \{x \mapsto x^2 + a : a \in \mathbb{R}\}$. What is $\tau_{\mathcal{F}}$?

   (c) [Extracurricular!] Let $X = \ell^2$ and $Y = \mathbb{R}$ with its usual topology. Let $\mathcal{F} = \{\mathbf{x} \mapsto \langle \mathbf{x}, \mathbf{y}\rangle : \mathbf{y} \in \ell^2\}$. Show that $\tau_{\mathcal{F}}$ is the weak topology on ℓ^2.

23. (a) Check that the 'topological' definitions of continuity agree with the 'metric space' definitions.

(b) Define $f : \mathbb{Z} \to \mathbb{C}$ by $f(n) = n + n^2 i$. If $\mathbb{Z}$ and $\mathbb{C}$ have their usual (metric space) topologies, is f continuous? Generalize.

(c) Suppose that $f : (X, \tau_X) \to (Y, \tau_Y)$. Prove that f is continuous at $x \in X$ iff $f(x_\alpha) \to f(x)$ in (Y, τ_Y) whenever the net $x_\alpha \to x$ in (X, τ_X).

24. Some quick checks! Which of the following functions are continuous?

(a) $f : M_n(\mathbb{R}) \to \mathbb{R}$, $f(A) = \det(A)$. (Pick you favourite norm on $M_n(\mathbb{R})$.)

(b) $f : \{A \in M_n(\mathbb{R}) \,:\, A \text{ is invertible}\} \to M_n(\mathbb{R})$, $f(A) = A^{-1}$.

(c) $f : (C[0,1], \|\cdot\|_\infty) \to \mathbb{R}$, $f(g) = g(0)$.

(d) $f : (C[0,1], \|\cdot\|_1) \to \mathbb{R}$, $f(g) = g(0)$.

(e) $f : (\ell^2, \|\cdot\|_2) \to \mathbb{R}$, $f(\mathbf{x}) = \sum_{k=1}^\infty x_k^2$.

(f) $f : (\ell^2, \text{weak}) \to \mathbb{R}$, $f(\mathbf{x}) = \sum_{k=1}^\infty x_k^2$.

(g) $f : (C[0,1], \text{pointwise}) \to \mathbb{R}$, $f(g) = \int_0^1 g(t)\, dt$.

25. Let $X = \ell^1$. Given a net $\{\mathbf{x}_\alpha\}_\alpha \in \ell^1$, we say that:

- $\mathbf{x}_\alpha \to \mathbf{x}$ in **norm** if $\|\mathbf{x}_\alpha - \mathbf{x}\|_1 \to 0$;
- $\mathbf{x}_\alpha \to \mathbf{x}$ in the **weak topology** if, for all $\mathbf{y} \in \ell^\infty$, $\langle \mathbf{x}_\alpha, \mathbf{y}\rangle \to \langle \mathbf{x}, \mathbf{y}\rangle$;
- $\mathbf{x}_\alpha \to \mathbf{x}$ in the **weak-$*$ topology** if, for all $\mathbf{y} \in c_0$, $\langle \mathbf{x}_\alpha, \mathbf{y}\rangle \to \langle \mathbf{x}, \mathbf{y}\rangle$.

Here, as usual $\langle \mathbf{x}, \mathbf{y}\rangle = \sum_{k=1}^\infty x_k y_k$. All three senses of convergence come from a topology.

(a) It is easy to check that norm convergence implies weak convergence implies weak-$*$ convergence — do it!

(b) Find a sequence (of sequences!) which converges in the weak-$*$ topology but not in the weak topology.

(c) Look up (or try it yourself if you are brave) Schur's Theorem for ℓ^1. This says that for **sequences** in ℓ^1, weak convergence implies norm convergence. This doesn't work for nets though, and therefore the norm and weak topologies are distinct. (This is another reason why when we get to topological spaces you need to consider nets and not just sequences. We'll mention more about weak and weak-$*$ topologies later.)

26. The space $C(\mathbb{T}; \mathbb{C})$ consists of the continuous functions from the unit circle $\mathbb{T}$ to to complex plane. This is a complex Banach space under the norm $\|f\|_\infty = \sup_{z\in\mathbb{T}} |f(z)|$. Let $Z = \{f \in C(\mathbb{T}; \mathbb{C}) \,:\, 0 \notin f(\mathbb{T})\}$ be the set of such functions which are never zero. Informally, you can think of every $f \in X$ as a continuous path in the complex plane; as z moves around the circle, $f(z)$ traces out a continuous path in the plane. Attached to every such path is an integer $I(f)$ which measures the winding number of the path (clockwise) around the origin. So, for example, $f(z) = z^3$ has $I(f) = 3$. Why is the map $I : Z \to \mathbb{Z}$ continuous?

27. (a) $\mathrm{GL}(2,\mathbb{R})$ is the set of invertible 2×2 real matrices. Pick your favourite metric to apply to this set. Find a continuous map $H : [0,1] \to \mathrm{GL}(2,\mathbb{R})$ with $H(0) = I$ (the identity matrix), and $H(1) = -I$.

    (b) Now try this with $\mathrm{GL}(2,\mathbb{R})$ replaced by $\mathrm{GL}(2,\mathbb{Z})$ the set of 2×2 matrices with integer entries and determinant ± 1.

Chapter 5

Compactness and Connectedness

5.1 Introduction

One of the most important theorems of elementary calculus is the Max-Min Theorem. A slightly more sophisticated statement of this than what you saw in first year is that if $f : [a, b] \to \mathbb{R}$ is continuous then the range of f is also a closed and bounded interval[1] $[c, d]$. We also saw in first year that you need the domain to be both closed and bounded — it is easy to find examples which map unbounded closed intervals to non-closed sets. or non-closed bounded intervals to unbounded sets. (It is perhaps worth noting here that being open is also typically not preserved under a continuous map!)

The multivariate version of this is the Max-Min Theorem that if $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous and $K \subseteq \mathbb{R}^n$ is closed and bounded then $f(K)$ is closed and bounded. This is pretty useful since closed and bounded subsets of $\mathbb{R}^n$ are particularly nice. For example, we have the Bolzano–Weierstrass Theorem:

Theorem 5.1.1. *Suppose that $K \subseteq \mathbb{R}^n$. Then the following conditions are equivalent:*

- *K is closed and bounded.*
- *Every sequence $\{x_k\} \subseteq K$ contains a subsequence which converges to an element of K.*

It is natural to hope that all this might extend to all metric spaces, but unfortunately it doesn't, at least not in the obvious way.

Example 5.1.2. 1. Let $X = (0, 1)$ with its usual topology. Then $Y = (0, 1)$ is a closed subset of X, and Y is bounded. But the sequence $\{\frac{1}{k}\} \subseteq Y$ does not have any subsequence which converges in Y, and the function $f(x) = \dfrac{2x - 1}{x(1 - x)}$ is unbounded on Y.

[1]To be strictly accurate we are also including the intermediate value theorem here!

2. Let $X = \mathbb{Z}$ with the discrete metric. Then every subset of X (including in particular X itself) is closed and bounded, but obviously there are sequence in X which don't converge, and continuous functions from X to $\mathbb{R}$ which are not bounded.

Even in general normed spaces, closed and bounded sets need not behave as they do in $\mathbb{R}^n$.

It turns out that there is a topological condition, called **compactness**, which does have the properties that we are looking for even in general topological spaces. The definition however is quite complicated, and it is not at all obvious that in the case of $\mathbb{R}^n$, it is equivalent to being closed and bounded!

5.2 Covers and Compactness

Let's start by looking at one one distinction between $(0, 1)$ and $[0, 1]$.

Example 5.2.1. Let $X = (0, 1)$ with its usual topology. We can write X as a union of open sets in many ways:

$$\begin{aligned}
X &= \left(0, \frac{2}{3}\right) \cup \left(\frac{1}{3}, 1\right) \\
&= \left(\frac{1}{3}, 1\right) \cup \left(\frac{1}{4}, \frac{1}{2}\right) \cup \left(\frac{1}{5}, \frac{1}{3}\right) \cup \cdots = \bigcup_{k=1}^{\infty} \left(\frac{1}{k+2}, \frac{1}{k}\right) \\
&= \left(0, \frac{2}{3}\right) \cup \left(\frac{1}{3}, 1\right) \cup \left(\frac{1}{4}, \frac{1}{2}\right) \cup \left(\frac{1}{5}, \frac{1}{3}\right) \cup \cdots = \left(0, \frac{2}{3}\right) \cup \bigcup_{k=1}^{\infty} \left(\frac{1}{k+2}, \frac{1}{k}\right).
\end{aligned}$$

The first writes X as a union of only two open sets. The second way uses infinitely many — and they are all needed! The third uses infinitely many, but actually we could manage with just the first two sets in the list.

Example 5.2.2. We could try the same thing with $X = [0, 1]$. That is, how can you write $[0, 1]$ as a union of open sets? Remember that $[0, \frac{1}{2})$ is an open subset of X.

If you sit down and try this, you'll find that every time you write $[0, 1] = \bigcup_{k=1}^{\infty} U_k$ where each U_k is an open subset of $[0, 1]$, then it is actually like the third example above, in that you really only needed finitely many of the sets in your list. This isn't because you aren't clever enough — we'll see later why it is impossible to come up with a collection of infinitely many open sets which cover $[0, 1]$ and which are all needed.

Surprisingly, the properties of these 'open coverings' of a set capture just the behaviour that we are looking for.

Definition 5.2.3. Let (X, τ) be a topological space. A family of open sets $\{U_\alpha\}_{\alpha \in A} \subseteq \tau$ is called an **open cover** for $Y \subseteq X$ if $Y \subseteq \cup_{\alpha \in A} U_\alpha$.

The families $\{U_\alpha\}_{\alpha \in A}$ in this definition could be finite, countable or uncountable. The sets U_α need not be contained inside Y as in our example — they can hang over the edges. Some could even be completely disjoint from Y, although such elements of the cover would be redundant.

Example 5.2.4. We saw some examples above. Others are

1. $\{(x-1, x+1)\}_{x\in\mathbb{R}}$ is an open cover for $\mathbb{R}$.
2. $\{(-1, 2)\}$ is an open cover for $[0, 1]$.
3. $\{\{x\}\}_{x\in X}$ is an open cover in any discrete topological space.
4. $\{B(\pm\mathbf{e}_k, \sqrt{2})\}_{k=1}^{\infty}$ is an open cover of the unit ball $B(\mathbf{0}, 1)$ in ℓ^2.

Definition 5.2.5. A **subcover** of an open cover $\{U_\alpha\}_{\alpha\in A}$ is a subcollection $\mathcal{C}' = \{U_\alpha\}_{\alpha\in A'}$ such that $A' \subseteq A$ and $\mathcal{C}'$ is still a cover.

Example 5.2.6. $\{(x-1, x+1)\}_{x\in\mathbb{N}}$ is a subcover of the cover $\{(x-1, x+1)\}_{x\in\mathbb{R}}$ of $\mathbb{R}$.

Definition 5.2.7. Let (X, τ) be a topological space. A subset $Y \subseteq X$ is **compact** if given any open cover of Y there exists a finite subcover.

This is perhaps better understood as saying that a set W is **not compact** if there is some infinite open cover $\{U_\alpha\}_{\alpha\in A}$ of W for which no finite subcollection covers W.

Example 5.2.8. 1. $(0, 1)$ is not compact as the second cover we gave has no finite subcover.

2. $\mathbb{R}$ is not compact, since there is no subcover of $\{(n-1, n+1)\}_{n\in\mathbb{N}}$ that covers $\mathbb{R}$.
3. The unit ball in ℓ^2 is not compact, using the open cover given in Example 5.2.4.
4. Any infinite discrete topological space is not compact.

Compactness is not the easiest property to check, as you have to say something about *every possible* way of covering subset of a topological space with open sets. The definition is definitely NOT saying that you need to be able to cover Y with finitely many open sets. (On the other hand, as these examples show, proving noncompactness is usually easy!)

Let's begin with an easy fact.

Proposition 5.2.9. *Every finite subset of a topological space (X, τ) is compact.*

Proof. Let $Y \subseteq X$ be finite, say

$$Y = \{y_1, y_2, \ldots, y_n\}$$

If $\{U_\alpha\}_{\alpha\in A} \subseteq \tau$ is any open cover of Y, then it follows that there must at least one open set containing each element. It follows that (by picking one set that contains y_1, one that contains y_2, etc.) we can construct a new family of sets $\{U_\alpha\}_{\alpha\in A'}$, with at most n elements. ■

Example 5.2.10. Let $Y = \{0\} \cup \{1, \frac{1}{2}, \frac{1}{3}, \dots\} \subseteq \mathbb{R}$ with the usual metric topology.

Claim: Y is compact.

Let $\{U_\alpha\}_{\alpha \in A}$ be an open cover of Y. Then there exists α_0 such that $0 \in U_{\alpha_0}$. As U_{α_0} is open in the metric topology it follows that there exists $\delta > 0$ such that $(-\delta, \delta) \subseteq U_{\alpha_0}$. Then there exists $K \in \mathbb{Z}^+$ such that $\frac{1}{K} < \delta$ so

$$\left\{0, \frac{1}{K}, \frac{1}{K+1}, \dots\right\} \subseteq B(0, \delta) \subseteq U_{\alpha_0}$$

Clearly, $Y \setminus \{0, \frac{1}{K}, \frac{1}{K+1}, \dots\}$ is a finite set. As in the previous example, we can choose U_{α_j} so that $\frac{1}{j} \in U_{\alpha_j}$ for $j = 1, \dots, K-1$. and so $\{U_{\alpha_j}\}_{j=0}^{K-1}$ is a finite subcover. Thus, Y is compact.

You might at this stage complain that it is not at all clear that using the term 'compact' is justified as it isn't at all obvious that this new concept matches up with our use in $\mathbb{R}^n$.

The first and main step is to prove the following.

Theorem 5.2.11. $[0, 1]$ *is a compact subset of* $\mathbb{R}$.

The proof needs to use something more than the proof in the previous example. The reason is that the proof in Example 5.2.10 is also valid if we consider Y to be a subset of the metric space $(\mathbb{Q}, |\cdot|)$.

It is **not** true that $Y = \mathbb{Q} \cap [0, 1]$ is a compact subset of $\mathbb{Q}$. The sets $U_n = \{x \in \mathbb{Q} : |x - 1/\pi| > 1/n\}$ form an open cover for Y which has no finite subcover. To prove the theorem we will need to bring in the completeness of $\mathbb{R}$. This is done via the Least Upper Bound Axiom.

Proof. Let $\{U_\alpha\}_{\alpha \in A}$ be an open cover of $[0, 1]$. Define $Z \subseteq [0, 1]$ by the condition that $x \in Z$ if $[0, x]$ can be covered by finitely many U_α.

Since 0 is in some U_α is is clear that $0 \in Z$ and hence, in particular, that $Z \neq \emptyset$. As Z is a nonempty, bounded subset of $\mathbb{R}$ then (by the Least Upper Bound Axiom of $\mathbb{R}$) $c = \sup Z$ exists.

Claim 1: $c \in Z$.

Proof of Claim 1: As $c \in [0, 1]$, there exists $\alpha_0 \in A$ such that $c \in U_{\alpha_0}$. Note that if $c = 0$, then $c \in Z$. If $c > 0$, then there exists $\delta > 0$ such that $B(c, \delta) \subseteq U_{\alpha_0}$ (since U_{α_0} is open). Also, by the definition of the supremum, there exists $x \in Z$ with $c - \delta < x \leq c$. As $x \in Z$, it follows that $[0, x] \subseteq \displaystyle\bigcup_{i=1}^{n} U_{\alpha_i}$ Thus, $[0, c] \subseteq \displaystyle\bigcup_{i=0}^{n} U_{\alpha_i}$ Hence $c \in Z$.

Claim: $c = 1$.

Proof of Claim: Suppose that $c < 1$, then $c \in U_{\alpha_0}$ for some $\alpha_0 \in A$. Again there exists $\delta > 0$ with $B(c, \delta) \subseteq U_{\alpha_0}$. As $c \in Z$, $[0, c] \subseteq \bigcup_{i=1}^{n} U_{\alpha_i}$.

Now, choose $c' \in [0, 1]$ with $c < c' < c + \delta$. Then $[0, c'] \subseteq \bigcup_{i=0}^{n} U_{\alpha_i}$ too, so $c' \in Z$. However, this contradicts the fact that $c = \sup Z$. Hence $c = 1$. ■

The same proof would clearly prove that any closed and bounded interval $[a, b]$ is also compact in $\mathbb{R}$. If fact we can do more.

Corollary 5.2.12 (Heine-Borel Version 1)**.** *A set $Y \subseteq \mathbb{R}$ is compact if and only if it is closed and bounded.*

Proof. Suppose first that $Y \subseteq \mathbb{R}$ is closed and bounded. Then $Y \subseteq [a, b]$ for some closed interval $[a, b]$. Suppose that $\mathcal{C} = \{U_\alpha\}_{\alpha \in A}$ is an open cover for Y. Let $V = [a, b] \setminus Y$, which is open. Then $\{V\} \cup \mathcal{C}$ is an open cover of $[a, b]$ and hence has a finite subcover $\mathcal{C}'$. The finitely many sets in $\mathcal{C}'$ certainly cover Y. The set V may or may not be in $\mathcal{C}'$. If it is, then $C'' = \mathcal{C}' \setminus \{V\}$ is a finite subcover of $\mathcal{C}$ and it must still cover Y as V certainly contains no elements of Y. If V is not in $\mathcal{C}'$ then $\mathcal{C}'$ is a finite subcover of $\mathcal{C}$.

For the converse, suppose that Y is not bounded. Then $\mathcal{C} = \{(-n, n)\}_{n=1}^{\infty}$ is an open cover for Y, but any subcollection only covers a bounded set so it can't cover Y. Thus any compact set must be bounded.

Finally, suppose that Y is not closed. Then there exists a limit point x of Y which is not in Y. For $n = 1, 2, \ldots$, let $U_n = \mathbb{R} \setminus [x - \frac{1}{n}, x + \frac{1}{n}]$. Then $\bigcup_{n=1}^{\infty} U_n = \mathbb{R} \setminus \{x\}$ and so $\mathcal{C} = \{U_n\}$ is an open cover of Y. But any finite subcollection of $\mathcal{C}$ must miss some open ball around x and hence not contain some elements of Y. Thus $\mathcal{C}$ has no finite subcover. Therefore any compact set must be closed. ■

The proof of the second half of the corollary doesn't really depend at all on the fact that we are in $\mathbb{R}$. If you replace the intervals with balls, then the proof would work in any **metric** space. We therefore have:

Corollary 5.2.13. *Suppose that (X, d) is metric space. Then every compact set is closed and bounded.*

Example 5.2.14. Compact sets need not be closed! Let X have the indiscrete topology $\{\emptyset, X\}$. Suppose that Y is any nonempty proper subset of X. Then Y is not closed, but any open cover of Y must contain at least one copy of X and so has a finite subcover of the form $\{X\}$. Thus, every subset of an indiscrete topological space is compact, although most are not closed. (If we stay away from the 'Dark Side', this doesn't happen. In Hausdorff spaces we'll see that compact sets must be closed).

Example 5.2.15. The closed unit ball in ℓ^2 is closed and bounded, but as we saw, it is easy to construct an open cover which has no finite subcover, and so this set is not compact. We'll see later that in normed spaces you need to strengthen the boundedness condition to something called total boundedness.

The first part of the proof of Theorem 5.2.12 really shows the following.

Theorem 5.2.16. *Any closed subset of a compact set is compact.*

Proof. Suppose that $Y \subseteq K \subseteq X$, that Y is closed and that K is compact. Suppose that $\mathcal{C} = \{U_\alpha\}$ is an open cover of Y. Let

$$\mathcal{C}' = \mathcal{C} \cup \{X \setminus Y\}.$$

Then $\mathcal{C}'$ covers all of X and certainly covers K so there is a finite subcover of $\mathcal{C}'$ which covers K. It is possible that this finite subcover includes the added open set $X \setminus Y$, but since no element of Y is in that set, the remaining (finite) number of open sets in the subcover must cover all of Y. Therefore Y is compact. ■

The Max-Min Theorem can be restated as saying that if $f : \mathbb{R} \to \mathbb{R}$ is continuous then the image of a closed and bounded subset $K \subseteq \mathbb{R}$ is always a closed and bounded subset of $\mathbb{R}$. The generalization to general spaces is not too hard.

Theorem 5.2.17. *Suppose that (X, τ_X) and (Y, τ_Y) are topological spaces and that $f : X \to Y$ is continuous. If $K \subseteq X$ is compact in (X, τ) then $f(K)$ is compact in (Y, τ_Y).*

Proof. Suppose that $\{U_\alpha\}$ is an open cover for $f(K)$. Since each U_α is open, the sets $V_\alpha = f^{-1}(U_\alpha)$ are open in X, and so $\{V_\alpha\}$ forms an open cover for K. As K is compact, we can find a finite subcover

$$K \subseteq \bigcup_{j=1}^{n} V_{\alpha_j}$$

and going back to Y, this says that

$$f(K) \subseteq \bigcup_{j=1}^{n} U_{\alpha_j}.$$

Thus our original open cover for $f(K)$ has a finite subcover and hence $f(K)$ is compact. ■

Note in particular that this tells us that compactness is a topological invariant.

Corollary 5.2.18. *Suppose that (X_1, τ_1) and (X_2, τ_2) are homeomorphic topological spaces. Then X_1 is compact if and only if X_2 is compact.*

Example 5.2.19. Thus while the open unit disk $\mathbb{D}$ is homeomorphic to the whole complex plane $\mathbb{C}$, the closed disk (which is compact!) can not be homeomorphic to any unbounded subset of the plane (as they are all noncompact).

Corollary 5.2.20. *Suppose that (X, τ) is a topological space, that $K \subseteq X$ is compact, and that $f : K \to \mathbb{R}$ is continuous. Then there exists $x_{max} \in K$ such that $f(x) \leq f(x_{max})$ for all $x \in K$.*

Proof. $f(K)$ must be compact, and hence is a closed and bounded subset of $\mathbb{R}$. A closed and bounded subset of $\mathbb{R}$ must have a largest element, say $y_{max} = f(x_{max})$. ■

What we often want is to use compactness to deduce the existence of a convergent sequence.

Definition 5.2.21. A topological space (X, τ) is **sequentially compact** if every sequence in X has a convergent subsequence (to an element of X).

The Bolzano-Weierstrass Theorem says that every closed and bounded subset of $\mathbb{R}^n$ is sequentially compact.

In general, a topological space can be sequentially compact without being compact — and vice versa! (see Royden's *Real Analysis* for an example). All is not lost however. If we restrict to metric spaces, then compactness and sequential compactness are equivalent (see Theorem 5.5.1 below.) In a general topological space the problem is that, as we saw in Chapter 4, sequences don't fully describe convergence in the space. The general statement is that a space is compact if and only if every net has a subnet which converges to an element of the space. Subnets are much harder to work with than subsequences so this is not quite as useful as you might hope.

The story so far: Compactness is a nice property for a set K to have in terms of the functions on K being not just bounded, but always attaining maximum and minimum values. In $\mathbb{R}^n$, compact sets are quite easy to recognise, but in some many spaces this is not the case. Over the next few sections we'll see some further characterizations of compact sets, and some further properties that they possess. We'll see for example, that compactness plays an important role in proving the Weierstrass Approximation theorem which says that the polynomials are dense in $C[0,1]$. We'll also see a characterization of compact subsets of $C[0,1]$.

5.3 Compactness and the relative topology

Is $[0,1]$ open? Certainly not as a subset of $X = \mathbb{R}$. But it is open as a subset of the metric space $X = [0,1]$ or of the space $X = [0,1] \cup [3,\infty)$. Note that what we are changing here is not the metric so much as the containing metric space X. *Being closed is dependent on what space you are considered to be a subset of.*

An important property of compactness is that it depends only on the topology and is independent of the containing space. This needs to be made precise.

If Y is a subset of a metric space (X,d) then it is obvious that Y is also a metric space by just restricting the metric d to only considering elements of Y. Usually one doesn't make a distinction, but for the moment, let's define $d_Y(y_1,y_2) = d(y_1,y_2)$. A sequence $\{y_k\}$ in Y converges to $y \in Y$ in (Y,d_Y) if and only if it converges to y in (X,d). (Although of course a sequence of elements of Y might converge in (X,d) without converging in (Y,d_Y).)

The open balls in (Y,d) look like

$$B_Y(y,r) = \{u \in Y : d_Y(u,y) < r\} = \{u \in X : d(u,y) < r\} \cap Y = B_X(y,r) \cap Y.$$

That is, they are all formed by looking at the intersection of a ball in X with the set Y. This turns out to be how to produce a suitable topology on a subset.

Theorem 5.3.1. *Suppose that (X,τ) is a topological space and that Y is a nonempty subset of X. Then*

$$\tau_Y = \{U \cap Y : U \in \tau\}$$

is a topology on Y.

Proof. This is basically just set theory. For example, if $\{V_\alpha = U_\alpha \cap Y\}$ is a family of elements of τ_Y then

$$\bigcup_\alpha V_\alpha = \bigcup_\alpha (U_\alpha \cap Y) = \left(\bigcup_\alpha U_\alpha)\right) \cap Y$$

which is again an element of τ_Y. ■

Definition 5.3.2. The topology τ_Y is called the relative topology on Y induced by the topology τ on X.

As we noted in the metric space examples above, a set might be open or closed in the relative topology on Y, but not in the original topology on X. On the other hand, if $U \subseteq Y$ is open in X, then it is clearly also open in (Y, τ_Y).

Exercise 5.3.3. Prove that a net $\{y_\alpha\}_{\alpha\in A} \subseteq Y$ converges in the relative topology to an element $y \in Y$ if and only if it converges in (X, τ).

Theorem 5.3.4. *Suppose that (X, τ) is a topological space and that Y is a nonempty subset of X. A subset K of Y is compact in (Y, τ_Y) if and only if it is compact in (X, τ).*

Proof. ($\Rightarrow$) Suppose that K is compact in (Y, τ_Y) and that $\{U_\alpha\}$ is an open cover of K in (X, τ). Then $\{U_\alpha \cap Y\}$ is an open cover of K in (Y, τ_Y) and hence has a finite subcover $\{U_{\alpha_j} \cap Y\}_{j=1}^n$. But clearly then $K \subseteq \bigcup_{j=1}^n U_{\alpha_j}$ so $\{U_{\alpha_j}\}_{j=1}^n$ is a finite subcover in (X, τ).

($\Leftarrow$) Suppose that K is compact in (X, τ) and that $\{U_\alpha \cap Y\}$ is an open cover of K in (Y, τ_Y). Then $\{U_\alpha$ is an open cover of K in (X, τ) and hence has a finite subcover $\{U_{\alpha_j}\}_{j=1}^n$ and again this restricts to a finite subcover of the original cover. ■

Because of this theorem, we'll often just consider compact topological spaces, rather than compact subsets of topological spaces as this usually simplifies the statement of the results.

5.4 The finite intersection property

Sometimes it is easier to work with closed sets rather than open ones. Suppose that $\{U_\alpha\}$ is an open cover for $Y \subseteq X$. Then the sets $F_\alpha = Y \setminus U_\alpha = Y \setminus (U_\alpha \cap Y)$ are all closed in the relative topology on Y and

$$\bigcap_\alpha F_\alpha = \emptyset.$$

If $\{U_\alpha\}_{\alpha\in A}$ has no finite subcover then no intersection of finitely many of the sets F_α is empty.

Definition 5.4.1. A family of sets $\{F_\alpha\}_{\alpha\in A}$ has the **finite intersection property** if every finite subfamily has a non-empty intersection.

Example 5.4.2. 1. $\{(0, \frac{1}{k})\}_{k\in\mathbf{Z}^+}$ in $\mathbb{R}$ has the finite intersection property. (But this family has empty intersection.)

2. $\{(\alpha, \alpha+7)\}_{\alpha\in\mathbb{R}}$ in $\mathbb{R}$ doesn't have the finite intersection property.

Theorem 5.4.3. *Let (X, τ) be topological space. Then X is compact if and only if every family of closed sets $\{F_\alpha\}_{\alpha\in A}$ having the finite intersection property has a non-empty intersection, $\bigcap_{\alpha\in A} F_\alpha \neq \emptyset$.*

Proof. ($\Rightarrow$) Suppose that X is compact and assume that there exists a family of closed subsets $\{F_\alpha\}_{\alpha\in A}$ in X having the finite intersection property with an empty intersection. Now,

$$X = X \setminus \emptyset = X \setminus \bigcap_{\alpha\in A} F_\alpha = \bigcup_{\alpha\in A}(X \setminus F_\alpha). \qquad \text{(De Morgan's Laws)}$$

But the above implies that $\{U_\alpha = X \setminus F_\alpha\}_{\alpha\in A}$ is an open cover for X. Since X is compact it follows that there is a finite subcover, $\{U_{\alpha_j}\}_{j=1}^n$. Then

$$\emptyset = X \setminus \bigcup_{j=1}^{n}(X \setminus F_{\alpha_j}) = \bigcap_{j=1} F_{\alpha_j}$$

which clearly contradicts the finite intersection property. ($\Leftarrow$) (via the contrapositive) Let $\{U_\alpha\}_{\alpha\in A}$ be an open cover for X, set $F_\alpha = X \setminus U_\alpha$ for $\alpha \in A$. So $\{F_\alpha\}_{\alpha\in A}$ is a family of closed subsets of X. Now,

$$\bigcap_{\alpha\in A} F_\alpha = \bigcap_{\alpha\in A}(X \setminus U_\alpha) = X \setminus \bigcup_{\alpha\in A} U_\alpha = X \setminus X = \emptyset$$

Thus, $\{F_\alpha\}_{\alpha\in A}$ cannot have the finite intersection property and so there exists $A' \subseteq A$ with $|A'| \in \mathcal{N}$ such that

$$\bigcap_{\alpha\in A'} F_\alpha = \emptyset$$

Hence

$$X = X \setminus \emptyset = X \setminus \bigcap_{\alpha\in A'} F_\alpha = \bigcup_{\alpha\in A'}(X \setminus F_\alpha) = \bigcup_{\alpha\in A'} U_\alpha$$

That is, $\{U_\alpha\}_{\alpha\in A'}$ is a finite subcover.

■

5.5 Compactness in metric spaces

Compactness in metric spaces (and more generally in Hausdorff spaces) is somewhat less complicated than the situation in a general topological space.

First let us recover the Bolzano-Weierstrass property.

Theorem 5.5.1. *Let (X, d) be a metric space. Then X is compact if and only if it is sequentially compact.*

Proof. Suppose first that X is compact, and that $\{x_k\}_{k=1}^{\infty}$ is a sequence in X. For $n = 1, 2, \ldots$, let $F_n = \mathrm{cl}(\{x_k\}_{k=n}^{\infty})$. It is easy to see that the family $\{F_n\}_{n=1}^{\infty}$ is a sequence of closed sets which has the finite intersection property. By Theorem 5.4.3, this family has a nonempty intersection. That is, there exists

$$x \in \bigcap_{n=1}^{\infty} F_n.$$

We will now define a subsequence of $\{x_k\}$ recursively. To start, as $x \in \mathrm{cl}(\{x_k\}_{k=1}^{\infty})$ there exists k_1 such that x_{k_1} sits in the open ball around x of radius 1. Suppose now that $m \geq 1$ and that we can chosen $k_1 < k_2 < \cdots < k_m$. As $x \in \mathrm{cl}(\{x_k\}_{k=k_m+1}^{\infty})$, there exists $k_{m+1} > k_m$ such that $x_{k_{m+1}}$ sits in the open ball around x of radius $\frac{1}{m+1}$.

Thus $\{x_{k_n}\}_{n=1}^{\infty}$ is a subsequence of $\{x_k\}$, and as $d(x_{k_n}, x) < \frac{1}{n}$, this subsequence converges to x. That is, our sequence has a convergent subsequence.

We leave the converse as an exercise. ■

We saw that in a metric space that a compact set must be closed and bounded, but that the converse does not hold. The correct converse requires a related concept called total boundedness.

Exercise 5.5.2. Note that being closed and bounded does not imply compactness even in a low dimensional setting. If $X = \mathbb{Q}$ (with the usual metric), then $Y = \mathbb{Q} \cap [0, 1]$ is a closed and bounded subset of X, but it is not compact. The problem here is that the set has holes, so what we really want is completeness rather than being closed.

Definition 5.5.3. Let (X, d) be a metric space. Fix $\epsilon > 0$, and let $Y \subseteq X$. We call $C \subseteq X$ an **ϵ-net** for Y if

$$Y \subseteq \bigcup_{x \in C} B(x, \epsilon)$$

Example 5.5.4. $\{0, 1/2, 1\}$ is a 1/3-net for $[0, 1]$ in $\mathbb{R}$

Definition 5.5.5. Let $Y \subseteq X$. We say that Y is **totally bounded** if for all $\epsilon > 0$, there exists a finite ϵ-net for Y.

Example 5.5.6. Let $Y \subseteq \mathbb{R}^2$ be bounded. We can certainly put a rectangle around Y, say $Y \subseteq [a_1, b_1] \times [a_2, b_2]$. Fix $\epsilon > 0$. Let C be any lattice of points spaced at most $\epsilon/\sqrt{2}$ apart in each of the x and y . Then C is a finite ϵ-net for the rectangle, and hence Y. Thus Y is totally bounded.

Example 5.5.7. It is clear that for each $n \in \mathbb{Z}^+$ any bounded subset of $\mathbb{R}^n$ will also be totally bounded. We do this by choosing for each fixed $\epsilon > 0$ a lattice of points that are at most $\epsilon/\sqrt{n}$ apart.

Indeed, it is easy to show (try it!) that any totally bounded set is bounded. We have therefore shown the following.

Theorem 5.5.8. *A set $Y \subseteq \mathbb{R}^n$ is bounded iff it is totally bounded.*

The above construction doesn't work in the infinite dimensional case (eg ℓ^2) and indeed, the unit ball in ℓ^2 is not totally bounded.

The main theorem in this section is the following.

Theorem 5.5.9. *A metric space is compact if and only if it is complete and totally bounded.*

Proof. Suppose that X is a compact metric space. Suppose that $\epsilon > 0$. Then $\{B(x, \epsilon) : x \in X\}$ is an open cover of X and hence it has a finite subcover. Thus there is a finite ϵ-net for X, and so X is totally bounded.

Suppose that $\{x_k\}_{k=1}^\infty$ is a Cauchy sequence in X. By Theorem 5.5.1 this sequence has a convergent subsequence with limit $x \in X$. Since the sequence is Cauchy, one can show that in fact the whole sequence must converge to x (exercise!). Therefore X is complete.

For the converse, suppose that X is a complete and totally bounded metric space. Suppose that $\{x_k\}_{k=1}^\infty$ is any sequence in X. Since X is totally bounded we can cover X with a finite number of balls of radius 1. At least one of these balls, say B_1, must contain an infinite subsequence of the original sequence.

Now cover B_1 with a finite number of balls of radius $\frac{1}{2}$. (We can cover all of X so we can certainly cover B_1 in such a way!) At least one of these balls, call it B_2, must contain infinitely many elements of the subsequence, that is a subsubsequence of the original one. Indeed we can continue on to produce a sequence of balls B_n of radius $\frac{1}{n}$ so that at each stage $B_1 \cap \cdots \cap B_n$ contains infinitely many elements of the original sequence. We can therefore recursively choose $x_{k_n} \in B_1 \cap \cdots \cap B_n$ with $k_1 < k_2 < k_3 \ldots$.

If $n, m \geq N$, then $x_{k_n}, x_{k_m} \in B_N$, and so $d(x_{k_n}, x_{k_m}) < \frac{2}{N}$. Thus, the subsequence $\{x_{k_n}\}_{n=1}^\infty$ is Cauchy, and hence it converges as X is complete. Since every sequence has a convergent subsequence, again we use Theorem 5.5.1 to deduce that X is compact. ■

In a complete metric space a subset is closed if and only if it is complete, so we have the following.

Corollary 5.5.10. *A subset of a complete metric space is compact if and only if it is closed and totally bounded.*

Corollary 5.5.11 (Heine–Borel Theorem)**.** *A subset of $\mathbb{R}^n$ is compact if and only if it is closed and and bounded.*

5.6 Uniform Continuity

Our next aim is to prove that every continuous function on $[0,1]$ can be estimated uniformly by a polynomial. For this we need the following.

Definition 5.6.1 (Uniform Continuity)**.** Let (X,d) be a metric space and let $f: X \to \mathbb{R}$. We say that f is **uniformly continuous** if for all $\epsilon > 0$ there exists $\delta > 0$ such that

$$|f(x) - f(y)| < \epsilon \quad \text{whenever} \quad d(x,y) < \delta.$$

Remark 5.6.2. For continuity at x we are allowed to choose δ such that it is depends on x_1 and ϵ. For f to be continuous at all x, you are allowed to choose a different δ at each x. For uniform continuity, δ is only a function of ϵ.

For the masochists: f is continuous on X if

$$\forall x \in X \ \forall \epsilon > 0 \ \exists \delta > 0 \ \forall y \in B(x,\delta) \ |f(x) - f(y)| < \epsilon$$

whereas f is uniformly continuous if

$$\forall \epsilon > 0 \ \exists \delta > 0 \ \forall x \in X \ \forall y \in B(x,\delta) \ |f(x) - f(y)| < \epsilon.$$

Does that really help?

Example 5.6.3. $X = (0,1)$, $f(x) = \frac{1}{x} \in C(X)$ but f is not uniformly continuous. Prove this! Nor is $g: \mathbb{R} \to \mathbb{R}$, $g(x) = x^2$.

Example 5.6.4. $X = \mathbb{R}$, $f(x) = \sin(x)$ and f is uniformly continuous since for every ϵ, if $\delta = \epsilon$ then $|f(x) - f(y)| \le |x-y| < \epsilon$ if $|x-y| < \delta$.

One of the important things that compactness gives you is uniform continuity.

Theorem 5.6.5. *Suppose that (X,d) is a compact metric space and $f \in C(X)$. Then f is uniformly continuous.*

Proof. Suppose not. Then there exists $\epsilon > 0$ such that for all $n \in \mathbb{Z}^+$ there exists $x_n, x_n' \in X$ such that

$$d(x_n, x_n') < \frac{1}{n} \quad \text{but} \quad |f(x_n) - f(x_n')| \ge \epsilon \tag{5.6.1}$$

As X is compact the sequence $\{x_n\}_{n \in \mathbf{Z}^+}$ has a convergent subsequence $\{x_{n_k}\}_{k \in \mathbf{Z}^+}$ with limit $x \in X$. Note that

$$d(x_{n_k}', x) \le d(x_{n_k}', x_{n_k}) + d(x_{n_k}, x) < \frac{1}{n_k} + d(x_{n_k}, x) \to 0$$

That is $x_{n_k}' \to x$ as $k \to \infty$. As f is continuous $f(x_{n_k}) \to f(x)$ and $f(x_{n_k}') \to f(x)$ so that $|f(x_{n_k}) - f(x_{n_k}')| \to 0$ which contradicts (5.6.1). ■

5.7 The Weierstrass Approximation Theorem

Theorem 5.7.1 (Weierstrass Approximation Theorem)**.** *Let $f \in C[0,1]$ and let $\epsilon > 0$. Then there exists a polynomial $p \in C[0,1]$ such that $\|f - p\|_\infty < \epsilon$.*

In other words, the polynomials are a dense subset of $C[0,1]$ under the uniform norm. The proof is not just an existence proof. It gives a formula of how to construct a good approximating polynomial to f.

Definition 5.7.2. Fix $f \in C[0,1]$. For $n \in \mathbb{N}$, the n**th Bernstein polynomial for** f is defined to be

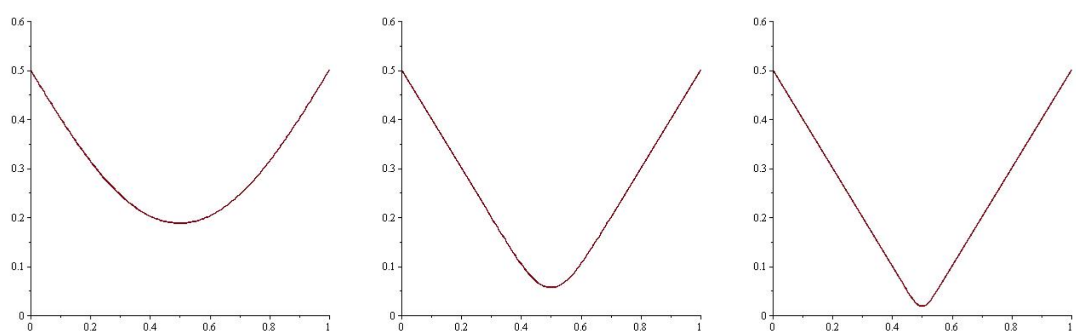
$$p_{f,n}(x) = \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

The proof of the Weierstrass Approximation Theorem that we will use will show that the Bernstein polynomials for f converge uniformly to f.

Example 5.7.3. Let $f(x) = |x - \frac{1}{2}|$. As this function is not differentiable in $[0,1]$ you can't try approximating this by a Taylor series. The first 4 Bernstein polynomials for f are:

$$\begin{aligned} p_{f,1} &= \frac{1}{2} \\ p_{f,2} = p_{f,3} &= \frac{1}{2} - x + x^2 \\ p_{f,3} &= \frac{1}{2} - x + 2x^3 - x^4. \end{aligned}$$

The graphs for $n = 5$, 50 and 500 are:



Proof. We need a few identities for the proof. From the binomial theorem

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \qquad (1)$$

Differentiate (1) with respect to x and multiply by x then we arrive at

$$nx(x+y)^{n-1} = \sum_{k=0}^{n} k\binom{n}{k}x^k y^{n-k} \qquad (2)$$

Differentiate (1) twice with respect to x then multiply by x^2

$$n(n-1)x^2(x+y)^{n-2} = \sum_{k=0}^{n} k(k-1)\binom{n}{k}x^k y^{n-k} \qquad (3)$$

Set $r_k(x) = \binom{n}{k}x^k(1-x)^{n-k}$ and putting $y = 1-x$ into $(1), (2), (3)$ implies that[2]

$$\begin{aligned} 1 &= \sum_{k=0}^{n} r_k(x) \\ nx &= \sum_{k=0}^{n} k r_k(x) \\ n(n-1)x^2 &= \sum_{k=0}^{n} k(k-1)r_k(x). \end{aligned}$$

Thus,

$$\begin{aligned} \sum_{k=0}^{n}(k-nx)^2 r_k(x) &= \sum_{k=0}^{n}(k^2 - 2knx + n^2x^2)r_k(x) \\ &= \sum_{k=0}^{n} k^2 r_k(x) - 2nx\sum_{k=0}^{n} k r_k(x) + n^2x^2\sum_{k=0}^{n} r_k(x) \\ &= (nx + n(n-1)x^2) - 2nx(nx) + n^2x^2 \\ &= nx(1-x). \end{aligned}$$

Choose M such that $|f(x)| \leq M$ on $[0,1]$. Since f is uniformly continuous there is for $\epsilon > 0$ a $\delta > 0$ such that

$$|f(x) - f(y)| < \frac{\epsilon}{2} \quad \text{whenever } |x-y| < \delta.$$

[2]The first of these says that at each x value, $\{r_k(x)\}_{k=0}^{n}$ is a probability distribution. Thus the Berstein polynomial $p_{f,n}(x)$ is actually giving the expected value of the random variable where value $f(\frac{k}{n})$ occurs with probability $r_k(x)$.

Now,

$$\begin{aligned}
|f(x) - p_{f,n}(x)| &= \Big| f(x) - \sum_{k=0}^{n} f\left(\frac{k}{n}\right) r_k(x) \Big| \\
&= \Big| \sum_{k=0}^{n} \left[f(x) - f\left(\frac{k}{n}\right)\right] r_k(x) \Big| \qquad \left(\text{Since } \sum_{k=0}^{n} r_k = 1\right) \\
&\leq \Big| \sum_{k\in K_1} \left[f(x) - f\left(\frac{k}{n}\right)\right] r_k(x) \Big| + \Big| \sum_{k\in K_2} \left[f(x) - f\left(\frac{k}{n}\right)\right] r_k(x) \Big| \\
&= S_1 + S_2,
\end{aligned}$$

where $K_1 = \{k : |k - nx| < \delta n\}$ and $K_2 = \{0, \ldots, n\} - K_1$
If $k \in K_1$, that is $|x - k/n| < \delta$ then clearly $|f(x) - f(k/n)| < \frac{\epsilon}{2}$ by uniform continuity.
Hence,

$$S_1 \leq \sum_{k\in K_1} \Big| f(x) - f\left(\frac{k}{n}\right) \Big| r_k(x) < \frac{\epsilon}{2}$$

Note that $r_k(x) \geq 0$ for all x, k. Moreover,

$$\begin{aligned}
S_2 &\leq \sum_{k\in K_2} \Big| f(x) - f\left(\frac{k}{n}\right) \Big| r_k(x) \\
&\leq \sum_{k\in K_2} 2M r_k(x) \\
&\leq \sum_{k\in K_2} 2M \frac{(k-nx)^2}{n^2\delta^2} r_k(x) \qquad (k \in K_2 \iff |k - nx| \geq \delta n) \\
&\leq \frac{2M}{n^2\delta^2} \sum_{k\in K_2} (k-nx)^2 r_k(x) \\
&= \frac{2M}{n^2\delta^2} nx(1-x) \\
&\leq \frac{M}{2n\delta^2} \qquad (\text{as } 0 \leq x(1-x) \leq 1/4\ )
\end{aligned}$$

Thus if n is large enough this is less than $\epsilon/2$ also. Hence,

$$|f(x) - p_{f,n}(x)| < S_1 + S_2 < \epsilon \qquad \text{for all } x \in [0, 1].$$

Thus $p_{f,n} \to f$ uniformly. ■

Remark 5.7.4. The interval $[0, 1]$ in the above proof can changed to $[a, b]$ with a simple adaption of the Bernstein polynomials.

Example 5.7.5. Let $\overline{\mathbb{D}} = \{z \in \mathbb{C} : |z| \leq 1\}$. One might hope that the the proof can be adapted to show that the polynomials of a complex variable are uniformly dense in $C(\overline{\mathbb{D}})$. This is however not the case!

In particular, the function $f(z) = \overline{z}$ cannot be approximated uniformly by polynomials on $\mathbb{D}$. Suppose that for all $\epsilon > 0$ there exists a complex polynomial $p_\epsilon : \overline{\mathbb{D}} \to \mathbb{C}$ such that $\|f - p_\epsilon\|_\infty < \epsilon$.

Since the convergence is uniform this implies that

$$\lim_{\epsilon \to 0^+} \int_{\mathbb{T}} p_\epsilon = \int_{\mathbb{T}} f.$$

By parametrizing $\mathbb{T}$ you can easily calculate that

$$\int_{\mathbb{T}} f = \int_0^1 f(e^{2\pi i t})(2\pi i e^{2\pi i t})\, dt = \int_0^1 2\pi i\, dt = 2\pi i$$

But by the Cauchy-Goursat theorem (since every polynomial is analytic on and inside $\mathbb{T}$)

$$\int_{\mathbb{T}} p_\epsilon = 0,$$

which is a contradiction.

The full Stone-Weierstrass Theorem generalizes Weierstrass' Theorem to more general algebras.

Theorem 5.7.6. *Let $X \subseteq \mathbb{C}$ be a compact Hausdorff space and suppose that $\mathcal{B} \subseteq C(X; \mathbb{C})$ satisfies*

1. *$\mathcal{B}$ is an algebra.*
2. *$\mathcal{B}$ contains the constant functions.*
3. *$\mathcal{B}$ separates points (ie. if $x_1 \neq x_2 \in X$ then there exists $f \in \mathcal{B}$ such that $f(x_1) \neq f(x_2)$).*
4. *If $f \in \mathcal{B}$, then $\overline{f} \in \mathcal{B}$.*

Then $\mathcal{B}$ is dense in $C(X; \mathbb{C})$.

We'll omit the proof. If you are just dealing with algebras of real-valued functions, then you can omit the condition that $\mathcal{B}$ be conjugate closed.

5.8 The Arzelà-Ascoli Theorem

The Heine-Borel Theorem tells us how to recognise a compact subset of $\mathbb{R}^n$, but how do you recognise a compact subset of $C[0,1]$ or some other infinite dimensional Banach space? The closed unit ball is compact in $\mathbb{R}^n$, but not in any infinite dimensional Banach space.

Example 5.8.1. For $k = 1, 2, 3, \ldots$, define $f_k \in C[0,1]$ by

$$f_k(x) = \begin{cases} 2kx, & 0 \le x \le \frac{1}{2k}, \\ 2 - 2kx, & \frac{1}{2k} < x \le \frac{1}{k}, \\ 0, & \frac{1}{k} < x \le 1. \end{cases}$$

For each k, $\|f_k\|_\infty = 1$, so this is a sequence in the closed unit ball of $C[0,1]$ (which is a closed and bounded set!). But this sequence has no convergent subsequence. To see this you might either show that no subsequence is Cauchy, or else note that the pointwise limit of this sequence is $f \equiv 0$. If any subsequence of $\{f_k\}$ converged in norm, it would have to converge to this pointwise limit, but $\|f_k - f\|_\infty = 1$ for all k and so no subsequence can possibly converge to f.

What this example shows is that the closed unit ball is a subset of $C[0,1]$ which is closed and bounded, but, as it contains a sequence with no convergent subsequence, it is not compact.

In general, in a Banach space a set is compact if and only if it is closed and totally bounded, but this is often not the easiest test to apply. The Arzelà-Ascoli Theorem gives a condition which is relatively easy to use. It involves a property known as equicontinuity.

Definition 5.8.2. Suppose $A \subseteq C[0,1]$. Then A is **equicontinuous** if for all $\epsilon > 0$, there exists $\delta > 0$ such that for all $f \in A$ and all x, y with $|x - y| < \delta$,

$$|f(x) - f(y)| < \epsilon.$$

One might say that the functions in A need to be uniformly uniformly continuous! That is, the way that δ depends on ϵ needs to not just be independent of the points $x, yin[0,1]$, but also the function f.

Again for the masochists: A is equicontinuous if

$$\forall \epsilon > 0 \ \exists \delta > 0 \ \forall f \in A \ \forall x \in [0,1] \ \forall y \in B(x,\delta), \ |f(x) - f(y)| < \epsilon.$$

Most people's brains struggle to deal with more than 2 quantifiers, so you can see why this concept is a challenge.

If you look back at Example reftents you'll see that the sequence $\{f_k\}$ is not equicontinuous.

The easiest test for equicontinuity is if the functions in A have uniformly bounded derivatives.

Example 5.8.3. Let

$$A = \{f \in C[0,1] \,:\, f(0) = 0,\ f \text{ is diff'ble}, |f'(x)| \leq 2 \text{ for all } x \in [0.1]\}.$$

Then A is equicontinuous.

Proof. Let $\epsilon > 0$. Let $\delta = \epsilon/2$. Let $f \in A$. Suppose $x, y \in [0,1]$ and $|x - y| < \delta = \epsilon/2$. By the Mean-Value Theorem there exists a c such that $|f(x) - f(y)| = |f'(c)||x - y| < 2 \times \epsilon/2 = \epsilon$. Then A is equicontinuous. ■

Theorem 5.8.4 (Arzelà-Ascoli)**.** *A set $\mathcal{B} \subseteq C[0,1]$ is compact (in the norm topology) if it is closed, bounded and equicontinuous.*

Rather than prove this, we'll prove a related result which is often good enough.

Corollary 5.8.5. *If $\{f_k\}_{k=1}^{\infty} \subseteq C[0,1]$ is bounded and equicontinuous, then it contains a convergent subsequence.*

Proof. [Not examinable] Enumerate $\mathbb{Q} \cap [0,1] = \{x_1, x_2, \ldots\}$
Now observe what happens at x_1 for $\{f_n(x_1)\}_{n=1}^{\infty}$. Since $\{f_n(x_1)\}_{n=1}^{\infty}$ is a bounded sequence in $\mathbb{R}$, it has a convergent subsequence. Denote the corresponding subsequence of functions by

$$f_{11}, f_{12}, f_{13}, \ldots \subseteq \{f_1, f_2, \ldots\}$$

Now $\{f_{11}(x_2), f_{12}(x_2), \ldots\}$ is also a bounded sequence of reals and hence it has a convergent subsequence. Denote the functions by

$$f_{21}, f_{22}, f_{23}, \ldots \subseteq \{f_{11}, f_{12}, \ldots\} \subseteq \{f_1, f_2, \ldots\}$$

Continue in this way to form (sub)^k sequences.

$$\begin{array}{cccc} f_{11} & f_{12} & f_{13} & \cdots \\ f_{21} & f_{22} & f_{23} & \cdots \\ f_{31} & f_{32} & f_{33} & \cdots \\ \vdots & \vdots & \vdots & \cdots \\ f_{n1} & f_{n2} & f_{n3} & \ddots \end{array}$$

Define the sequence $\{g_n\}_{n=1}^{\infty}$ by setting $g_n = f_{nn}$. Note that $\{g_n\}_{n=1}^{\infty}$ is a subsequence of $\{f_{11}, f_{12}, \ldots\}$ so $\{g_n(x_1)\}_{n=1}^{\infty}$ converges. Also $\{g_n(x)\}_{n=2}^{\infty}$ is a subsequence of $\{f_{21}, f_{22}, \ldots\}$ so $\{g_n(x_2)\}_{n=1}^{\infty}$ converges too. Indeed $\{g_n\}_{n=m}^{\infty}$ is a subsequence of $\{f_{m1}, f_{m2}, \ldots\}$ so $\{g_n(x_m)\}_{n=1}^{\infty}$ converges for all $m = 1, 2, \ldots$. That is g_n converges pointwise on $\mathbb{Q} \cap [0,1]$.

Now fix $\epsilon > 0$. Choose $\delta > 0$ such that for all n

$$|f_n(x) - f_n(y)| < \epsilon \quad \text{whenever} \quad |x - y| < \delta \quad x, y \in [0,1].$$

Choose rationals $y_1, \dots, y_k \in [0,1]$ such that for all $x \in [0,1]$ there exists i such that $|x - y_i| < \delta$. Now $\{g_n(y_1)\}_{n=1}^{\infty}$, $\{g_n(y_2)\}_{n=1}^{\infty} \dots \{g_n(y_k)\}_{n=1}^{\infty}$ all converge so there exists N such that for all $n, m \geq N$ $|g_n(y_i) - g_m(y_i)| < \epsilon$ for $i = 1, 2, \dots, k$.

Suppose thta $x \in [0,1]$. Choose y_i such that $|x - y_i| < \delta$. Then $|g_n(x) - g_n(y_i)| < \epsilon$ (by equicontinuity). Then we have

$$\begin{aligned} |g_n(x) - g_m(x)| &\leq |g_n(x) - g_n(y_i)| + |g_n(y_i)| + |g_n(y_i) - g_m(y_i)| + |g_m(y_i) - g_m(x)| \\ &< \epsilon + \epsilon + \epsilon \end{aligned}$$

for all $n, m \geq N$. It follows then that for $n, m \geq N$

$$\|g_n - g_m\|_\infty = \sup_x |g_n(x) - g_m(x)| \leq 3\epsilon.$$

This implies that $\{g_n\}$ is Cauchy and as $C[0,1]$ is complete it must converge. ■

We saw that the set A in Example 5.8.3 is equicontinuous. Also, if $f \in A$ and $x \in [0,1]$ then, using the fact that $f(x) = f(0) + \int_0^x f'(t)\,dt$ we can deduce that $|f(x)| \leq 2x \leq 2$, and so $\|f\|)\infty \leq 2$. Thus A is a bounded subset of $C[0,1]$. By the corollary then, any sequence in A would have a convergent subsequence. Note that A itself is not closed and so it is not compact. This means that the limit of the convergent subsequence might not lie in A, but rather in the closure of A.

Remark 5.8.6. The Arzela-Ascoli Theorem actually works for any set $\mathcal{B} \subseteq C(X; \mathbb{R}^n)$ with X a compact metric space. That is we can change the definition of equicontinuity by changing $|f_n(x) - f_n(y)|$ to $d(f_n(x), f_n(y))$ as the metric.

Exercise 5.8.7. Let $\{f_n\} \subseteq C[0,1]$ by uniformly bounded and let $F_n(x) = \int_0^x f_n(t)\,dt$ with $x \in [0,1]$. Prove $\{F_n\}$ has a uniform convergent subsequence.

5.9 Connectedness

We saw that compactness is a topological invariant. Another important topological invariant is connectedness. The most intuitive version of connectedness is that you can draw a line between any two points of the set without leaving the set. This turns out to not be the best definition in a general topological space, but, once we say it a bit more carefully, it is still often the most useful!

Definition 5.9.1. Suppose that (X, τ) is a topological space. We say that X is **path-connected** if for all $x, y \in X$ there exists a continuous $f : [0,1] \to X$ with $f(0) = x$ and $f(1) = y$.

We can also talk about path-connected subsets of a topological space (using the relative topology on the subset).

Example 5.9.2. $X = X_1 \cap X_2 \subseteq \mathbb{C}$, $X_1 = \mathbb{D}$ and $X_2 = \{z \in \mathbb{C} : |z-3| \leq 1\}$ with the usual metric topology. Suppose $f : [0,1] \to X$ is continuous with $f(0) = 0$ and $f(1) = 3$. Let $g : X \to \mathbb{R}$ with $g(z) = \Re(z)$. Then $g \circ f : [0,1] \to \mathbb{R}$ is contiuous with $g \circ f(0) = 0$ and $g \circ f(1) = 3$. So by the Intermediate Value Theorem there exists $c \in (0,1)$ with $(g \circ f)(c) = \frac{3}{2}$. But the range of $(g \circ f) \subseteq (-1,1) \cup [2,4]$ which is a contradiction, so no such f can exist. Thus f is not path-connected.

Example 5.9.3. Let X be the set of all 2×2 matrices with determinant 1. This is a subset of the set of all 2×2 matrices which is a normed space under several equivalent norms. If you wish, you can take

$$\left\| \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right\| = |a| + |b| + |c| + |d|.$$

Is X path-connected? If I give you two matrices in X, how can you find a suitable function f parametrizing a continuous path in X from one to the other?

Example 5.9.4. Let

$$X = \{(x,y) \,:\, x \neq 0,\ y = \sin(\tfrac{1}{x})\} \cup \{(0,0)\}$$

(with the usual metric topology in the plane). Is X path-connected or not? The sort of proof used in the first example won't work here.

A more 'straight topological' notion is that of connectedness.

Definition 5.9.5. Suppose that (X, τ) is a topological space. We say that X is **connected** if it is impossible to write $X = U \cup V$ where $U, V \in \tau$, $U, V \neq \emptyset$ and $U \cap V = \emptyset$.

Actually, this is perhaps more appropriately thought of as telling you what a disconnected space is! That is, a topological space is **disconnected** if there exist disjoint nonempty open sets U and V such that $X = U \cup V$. Note that this implies that both U and V are also closed!

The examples above are all really subsets X of some bigger standard space (Y, τ) where we have given X the relative topology inherited from Y. In practice it is usually easier to just worry about which sets are open in (Y, τ).

Definition 5.9.6. Suppose that (Y, τ) is a topological space and that $X \subseteq Y$. We shall say that X is a **disconnected subset** of Y is there exist open sets $U, V \in \tau$ such that

1. $X \subseteq U \cup V$,
2. $X \cap U \neq \emptyset$ and $X \cap V \neq \emptyset$, and
3. $X \cap U \cap V = \emptyset$.

If X is not disconnected, then we say that X is a **connected subset** of Y.

Exercise 5.9.7. Suppose that (Y, τ) is a topological space and that $X \subseteq Y$ inherits the relative topology τ_X. Prove that (X, τ_X) is a connected space if and only if X is a connected subset of (Y, τ).

Example 5.9.8. The first example above is disconnected as X_1 and X_2 are both open subsets in the metric topology on X. Alternatively, you could take $U = \{z : |z| < 1.2\}$ and $V = \{z : |z - 3| \le 1.2\}$ which are obviously disjoint open sets in the 'big space' $Y = \mathbb{C}$ and $X \subseteq U \cup V$. Note that this is an easier proof than the proof that it is not path-connected.

Example 5.9.9. Let $Y = \{a, b, c\}$ with topology $\{\emptyset, \{a\}, \{a, b\}, \{a, c\}, \{a, b, c\}\}$, and let $X = \{b, c\}$. Then the relative topology on X is $\{\emptyset, \{b\}, \{c\}, \{b, c\}\}$. Writing $X = \{b\} \cup \{c\}$ shows that X is a disconnected space in the relative topology. Note there that you can't find disjoint open sets U and V in Y such that $X \subseteq U \cup V$.

This example shows that it is sufficient in order to prove that $X \subseteq Y$ is disconnected to find disjoint open sets U, V in Y whose union contains X, it is not necessary. The issue is that if one has written $X = U_X \cup V_X$ with U_X, V_X nonempty, disjoint and open in the relative topology, one can always find open sets U, V such that $U_X = U \cap X$ and $V_X = V \cap X$, but you can't always choose U, V disjoint.

Theorem 5.9.10. *If (X, τ) is path-connected then it is connected.*

Proof. Suppose that X is path-connected but that it is not connected. That is, suppose $X = U \cup V$, $U, V \in \tau$, $U, V \neq \emptyset$, $U \cap V = \emptyset$. Choose $u \in U$ and $v \in V$. As X is path-connected there exists a continuous function $f : [0, 1] \to X$ with $f(0) = u$ and $f(1) = v$. Then $f^{-1}(U)$ and $f^{-1}(V)$ are non-empty open subsets of $[0, 1]$ and $[0, 1]$ equals the disjoint union of open sets $f^{-1}(U)$ and $f^{-1}(V)$ which is impossible (check that!). ■

Exercise 5.9.11. Prove that (X, τ) is connected if and only if the only sets which are both open and closed ('clopen') are X and the empty set[3].

The main theorem that we want to prove is the following.

Theorem 5.9.12. *Let (X, τ_X) and (Y, τ_Y) be topological spaces and suppose that $f : X \to Y$ is continuous. Then if X is connected, so is $f(X)$ (in the relative topology). (More generally, if $X_0 \subseteq X$ is connected, then $f(X_0) \subseteq Y$ is connected.)*

Proof. We'll prove the contrapositive. Suppose then that $f(X)$ is not connected, that is, there exist open sets U, V in $f(X)$ in the relative topology so that $f(X) = U \cup V$, $U, V \neq \emptyset$ and $U \cap V = \emptyset$. As U and V are open we have that $U = U_0 \cap f(X)$, $V = V_0 \cap f(X)$ for some $U_0, V_0 \in \tau_Y$.

Since f is continuous, $f^{-1}(U_0) = f^{-1}(U) \in \tau_X$ and $f^{-1}(V_0) = f^{-1}(V) \in \tau_X$. Then $X = f^{-1}(U) \cup f^{-1}(V)$ and $f^{-1}(U) \cap f^{-1}(V) = \emptyset$ with $f^{-1}(U), f^{-1}(V) \neq \emptyset$. Therefore, X is not connected. ■

[3] Under no circumstances should you google 'Hitler learns topology'.

Exercise 5.9.13. Prove that the image of a path-connected set under a continuous map is path-connected.

The proof we just did, and the proof in the exercise illustrate one difference between the two concepts of connectedness. For path-connectedness it is usually more straightforward to show that something is path-connected, rather than to show that something is not path-connected. That is because for the latter you need to show that something is impossible, which is often harder than just exhibiting some suitable paths. For connectedness, the harder sort of problem is often showing that a set is connected, as that is what involves showing that a certain way of splitting up the space is impossible. In practice, one often resorts to proving that a set is connected by instead showing that it has the stronger property of being path-connected.

In any case, from what we have done, both connectedness and path-connectedness are topological invariants. That is, if (X, τ_X) and (Y, τ_Y) are homeomorphic topological spaces, then they are either both connected or else they are both disconnected.

Note that these results should be considered as generalizations of the Intermediate Value Theorem. A connected subset of $\mathbb{R}$ is just an interval. If $f : [a, b] \to \mathbb{R}$ is continuous, Theorem 5.9.12 tells us that the range $R = f([a, b])$ must be a connected. That is, R must be an interval. Since $f(a)$ and $f(b)$ both lie in R, that means that every point in between $f(a)$ and $f(b)$ must also lie in R, which is precisely the conclusion of the Intermediate Value Theorem.

Example 5.9.14. Consider the set X in Example 5.9.4. Suppose that $X = U \cup V$ has been written as the union of two nonempty disjoint open sets (in the metric topology of X). We may as well assume that $(0, 0) \in U$. And we may as well assume that V contains a point $\mathbf{y} = (t, \sin \frac{1}{t})$ with $t > 0$. As U is open, this means that there exists $\epsilon > 0$ such that $B((0, 0), \epsilon) \subseteq U$. (Here we are considering the ball containing just the elements of X that are distance less than ϵ from the origin, not the whole ϵ-disk in the plane!) Choose k so that $2k\pi > \frac{1}{\epsilon}$. Thus if $s = \frac{1}{2k\pi}$, the point $(s, \sin \frac{1}{s}) = (s, 0)$ lies in $B((0, 0), \epsilon)$. Now just consider the subset $\hat{X} = \{(x, \sin \frac{1}{x}) : s \le x \le t\} \subseteq X$. Then $\hat{U} = U \cap \hat{X}$ and $\hat{V} = V \cap \hat{X}$ are nonempty disjoint open subsets of $\hat{X}$ whose union is $\hat{X}$, and so $\hat{X}$ is disconnected. But $\hat{X}$ is homeomorphic to the connected set $[s, t]$ via the obvious homeomorphism $f : [a, b] \to \hat{X}$, $f(x) = (x, \sin \frac{1}{x})$ which is impossible. This X must be disconnected.

Example 5.9.15. Let $\mathcal{H} = \ell^2$ and suppose that $T : \mathcal{H} \to \mathcal{H}$ is a continuous linear transformation (or operator). In MATH2601 you proved that the kernel of T is a subspace of $\mathcal{H}$. This might be finite dimensional or it might be infinite dimensional. The adjoint operator T^* is the (unique) continuous linear operator which satisfies

$$\langle Tx, y \rangle = \langle x, T^* y \rangle \qquad x, y \in \mathcal{H}.$$

The kernel of T^* turns out to be the subspace of vectors orthogonal to everything in the range of T. If both T and T^* have finite dimensional kernels then we call T a Fredholm operator. The set of all Fredholm operators is denoted $\mathcal{F}$. We define the Fredholm index to be map from $\mathcal{F}$ to $\mathbb{Z}$,

$$\text{index}(T) = \dim(\ker(T)) - \dim(\ker(T^*)).$$

If $\mathcal{H}$ were a finite dimensional Hilbert space, the index would be zero for every operator T, but in infinite dimensions you can get any integer value. For example.

$$T(x_1, x_2, x_3, \dots) = (0, x_1, x_2, x_3, \dots)$$

has index -1, while its adjoint

$$S(x_1, x_2, x_3, \dots) = (x_2, x_3, x_4, \dots)$$

has index 1. The index give some sort of measure of how non-invertible a matrix is.

The set $\mathcal{F}$ is a subset of the Banach space of all continuous linear operators which has the norm $\|T\| = \sup\limits_{\|x\|=1} \|Tx\|$. Note that the map $\mathcal{F} \to \mathbb{Z}$, $T \mapsto \dim(\ker(T))$ is definitely not continuous. To see this consider

$$T_\epsilon(x_1, x_2, x_3, \dots) = (\epsilon x_1, \epsilon x_2, \dots, \epsilon x_k, x_{k+1}, x_{k+2}, \dots)$$

and let $\epsilon \to 0$.

It turns out that index : $\mathcal{F} \to \mathbb{Z}$ **is** continuous however (and onto). This tells us quite a bit about the topological structure of $\mathcal{F}$. It must split up into lots of different unconnected components on which the index is constant!

5.10 Appendix: The product topology [Not examinable]

Suppose that (X_1, τ_1) and (X_2, τ_2) are two topological spaces. Their Cartesian product is the set of all ordered pairs

$$X_1 \times X_2 = \{(x_1, x_2) \,:\, x_1 \in X_1,\; X_2 \in X_2\}.$$

You would like a topology on this set.

If you think about $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ it is obvious that the Cartesian product of two open sets is open. For example $(0,1) \times (0,1)$ is the open unit square in the first quadrant. However a typical open subset (think of the open disk) of $\mathbb{R}^2$ cannot be expressed as a Cartesian product of two one dimensional open sets.

Theorem 5.10.1. *Suppose that (X_1, τ_1) and (X_2, τ_2) are two topological spaces. Then*

$$\mathcal{G} = \{U_1 \times U_2 \,:\, U_1 \in \tau_1,\; U_2 \in \tau_2\}$$

is a base for a topology on $X_1 \times X_2$.

Proof. We'll use Theorem 4.3.3. It is clear that $\mathcal{G}$ contains the empty set and that every element of $X_1 \times X_2$ is in at least one element of $\mathcal{G}$ (namely $X_1 \times X_2$!).

Suppose then that $G = U_1 \times U_2$, $H = V_1 \times V_2 \in \mathcal{G}$ and that $(x_1, x_2) \in G \cap H$. Then $x_1 \in U_1 \cap V_1 \in \tau_1$ and $x_2 \in U_2 \cap V_2 \in \tau_2$ so if $J = (U_1 \cap V_1) \times (U_2 \cap V_2) \in \mathcal{G}$ then $(x_1, x_2) \in J \subseteq G \cap H$. ■

Definition 5.10.2. The topology generated by $\mathcal{G}$ is called the **product topology** on $X_1 \times X_2$.

Exercise 5.10.3. Prove that in the product topology on $\mathbb{R} \times \mathbb{R}$ (with the usual metric topology on $\mathbb{R}$) is just the usual metric topology on $\mathbb{R}^2$.

It should be clear that it would be easy to extend this to a Cartesian product of finitely many spaces $X_1 \times X_2 \times \cdots \times X_n$. What is rather less trivial is to deal with a more complicated Cartesian product. These more complicated objects will appear in the proofs of some of the important compactness results in 4th year.

Definition 5.10.4. Suppose that A is a nonempty index set and for each $\alpha \in A$, X_α is a set. Then the **product space** $\prod_{\alpha \in A} X_\alpha$ is defined to be

$$\prod_{\alpha \in A} X_\alpha = \{(x_\alpha)_{\alpha \in A} \,:\, x_\alpha \in X_\alpha \text{ for all } \alpha \in A\}.$$

Example 5.10.5. Let $A = \mathbb{R}$ and for each $a \in \mathbb{R}$ let $X_a = \mathbb{C}$. Then

$$\prod_{\alpha \in \mathbb{R}} X_\alpha = \{(x_\alpha)_{\alpha \in \mathbb{R}} \,:\, x_\alpha \in \mathbb{C} \text{ for all } \alpha \in \mathbb{R}\}.$$

which is just a fancy way of describing all the functions from $\mathbb{R}$ to $\mathbb{C}$.

Given a product space $\prod_{\alpha \in A} X_\alpha$ and an element $\beta \in A$, you can define the projection map $\pi_\beta : \prod_{\alpha \in A} X_\alpha \to X_\beta$ by $\pi_\beta((x_\alpha)_{\alpha \in A}) = x_\beta$, which just picks out the 'βth' component.

Suppose now that all the spaces are topological spaces (X_α, τ_α). In general, the product topology on $\prod_{\alpha \in A} X_\alpha$ can be defined as the weakest topology that makes all the projection maps π_β, $\beta \in A$, continuous. Alternatively, a base for this topology is given by

$$\mathcal{G} = \{\prod_{\alpha \in A} U_\alpha \,:\, U_\alpha \in \tau_\alpha \text{ for all } \alpha \in A \text{ and } U_\alpha = X_\alpha \text{ except for finitely many } \alpha\}.$$

Exercise 5.10.6. Prove that if the example above, the product topology on $\prod_{a \in \mathbb{R}} \mathbb{C}$ is just the topology of pointwise convergence on the set of functions from $\mathbb{R}$ to $\mathbb{C}$.

Theorem 5.10.7 (Tychonoff). *Suppose that (X_α, τ_α), $\alpha \in A$ is a family of compact topological spaces. Then the product space $\prod_{\alpha \in A} X_\alpha$ is compact in the product topology.*

Many of the central theorems of functional analysis use Tychonoff's Theorem, often with each X_α being a relatively simple space, but with A being rather big! For example, in the proof that the closed unit ball in ℓ^2 is compact in the weak topology, one starts by considering the compact space

$$X = \prod_{\mathbf{x}\in\ell^2} \{z \in \mathbb{C} : |z| \leq \|\mathbf{x}\|\}$$

which is a big Cartesian product of closed disks in the plane.

This full version of Tychonoff's Theorem is in fact equivalent to the Axiom of Choice. We'll omit the proof.

Corollary 5.10.8 (Heine-Borel). *A set $K \subseteq \mathbb{R}^n$ is compact if and only if it is closed and bounded.*

Proof. We just need to prove the converse to Corollary 5.2.13. It follows from Tychonoff's Theorem and Theorem 5.2.11 that any 'box' $B = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$ in $\mathbb{R}^n$ is compact.

Suppose then $K \subseteq \mathbb{R}^n$ is closed and bounded. Since is it bounded it is sitting inside a box B of the above form. Therefore K is a closed subset of a compact set B in a Hausdorrf space and hence is compact. ■

Exercise 5.10.9. Using Tychonoff's Theorem is the proof of the corollary is a bit of overkill, so here is an alternative proof of the fact that the box B in that proof is compact.

Suppose that $\{\mathbf{x}_k\}_{k=1}^\infty$ is a sequence in B. Write each element as $\mathbf{x}_k = (x_{k1}, \ldots, x_{kn})$. Now $\{x_{k1}\}_{k=1}^\infty$ is a sequence in $[a_1, b_1]$ and hence it has a convergent subsequence. Take the corresponding subsequence of $\{\mathbf{x}_k\}$. Now repeat this exercise to find a sub-subsequence for which the second, as well as the first, components converge. Do this n times to get a sub-…-subsequence for which all the components converge. Check that the limit of this subsequence is indeed in B. It will follow that B is sequentially compact and hence compact.

5.11 Appendix: Commutative C^* algebras

One of the major theorems of Functional Analysis is the Spectral Theorem, which generalizes the fact that normal matrices can be diagonalized. The setting for the Spectral Theorem is the class of C^*-algebras.

Examples of C^*-algebras include $\mathbb{C}$, $\mathrm{M}_n(\mathbb{C})$, ℓ^∞, c_0, $C[0, 1]$, …. More formally C^*-algebras are both

- Complete normed spaces, ie Banach spaces.

- Algebras — complex vector spaces with multiplication and an 'involution'; that is, an operation * satisfying

$$(a^*)^* = a \quad (ab)^* = b^*a^* \quad (\lambda a)^* = \bar{\lambda}a^* \quad \|a^*a\| = \|a\|^2$$

  for all a, b and $\lambda \in \mathbb{C}$.

Question: What can a **commutative** C^*-algebra $\mathcal{A}$ look like?

Answer[4]**:** Turns out it must be isometrically isomorphic to $C(K)$, where K is a compact Hausdorff space.

The set $\mathcal{A}^*$ of continuous linear transformations $T : \mathcal{A} \to \mathbb{C}$ is itself a Banach space. This space can be given the topology of pointwise convergence on $\mathcal{A}$, which is weaker than its usual norm topology. The set Σ of nonzero C^*-algebra homomorphisms from $\mathcal{A} \to \mathbb{C}$ is a subset of $\mathcal{A}^*$. In this new topology it is closed and bounded and that turns out to be enough to make it compact in this topology. By some magic, we discover that $\mathcal{A}$ is isometrically isomorphic to $C(\Sigma)$.

What is more magical is that often this scary looking set Σ, which is a subset of some infinite dimensional Banach space with a crazy topology is actually homeomorphic to a compact subset of $\mathbb{C}$ with its usual topology.

The simplest case: Let $N \in M_n(\mathbb{C})$ be a normal matrix, and let

$$\mathcal{A}_N = \{p(N) : p \text{ is a complex polynomial}\}.$$

Then $\mathcal{A}_N$ is a commutative C^*-subalgebra of $M_n(\mathbb{C})$. The set Σ in this case formally consists of a set of multiplicative maps from $\mathcal{A}_N$ to $\mathbb{C}$, but in fact is homeomorphic to the set of eigenvalues of N. In the case of a normal operator on ℓ^2, Σ is homeomorphic to what is known as the spectrum of N,

$$\sigma(N) = \{\lambda \in \mathbb{C} : (N - \lambda I) \text{ is not invertible}\}.$$

5.12 Problems

1. Prove that in a Hausdorff topological space, every compact set is closed.

2. Let $X = \mathbb{Q} \cap [0, 1]$ with the usual metric topology. Prove directly from the definition that X is not compact.

3. Can a set be open and compact?

4. Give an example of a noncompact topological space in which a subset is compact if and only if it is finite.

[4]Strictly speaking you need $\mathcal{A}$ to have a multiplicative identity here.

5. Let $X = \mathbb{R}$ with the left-ray topology

$$\tau = \{\emptyset, \mathbb{R}\} \cup \{(-\infty, a) \,:\, a \in \mathbb{R}\}.$$

What are the compact subsets of X?

6. Let $X = \ell^2$ with the usual ℓ^2 norm. Define

$$Y = \Big\{(x_1, x_2, \dots) \in \ell^2 \,:\, |x_k| \le \frac{1}{k^2}\Big\}.$$

Is Y closed? Is Y bounded? Is Y compact?

7. Which of the following sets Y are compact in the given space (X, τ)?

   (a) $Y = \mathbb{Z}$ in $(\mathbb{Q}, |\cdot|)$.

   (b) $Y = \mathbb{Z}$ in $\mathbb{R}$ with the co-countable topology.

   (c) $Y = \{(\sin 3t, \cos 2t, \sin 5t) \in \mathbb{R}^3 \,:\, t \in [0, 2\pi]\}$ in $(\mathbb{R}, |\cdot|)$.

   (d) $Y = \{e^{ik} \,:\, k \in \mathbb{Z}\}$ in $(\mathbb{C}, |\cdot|)$.

   (e) $Y = \{A \in M_{44}(\mathbb{R}) \,:\, \det(A) = 1\}$ in $(M_{44}(\mathbb{R}), \|\cdot\|_{op})$.

   (f) $Y = \{f \in C^1[0,1] \,:\, f(0) = 0,\ \|f'\|_\infty \le 3\}$ in $(C[0,1], \|\cdot\|_\infty)$.

   (g) [Harder] $Y = \{\mathbf{x} \in \ell^1 \,:\, \|\mathbf{x}\|_1 \le 1\}$ in ℓ^1 with the topology $\mathbf{x}_\alpha \to \mathbf{x}$ if $\langle \mathbf{x}_\alpha, \mathbf{y}\rangle \to \langle \mathbf{x}, \mathbf{y}\rangle$ for all $\mathbf{y} \in c_0$.

8. Supppose that Y is a subset of a topological space (X, τ). Prove that a net $\{y_\alpha\}_{\alpha \in A} \subseteq Y$ converges in the relative topology to an element $y \in Y$ if and only if it converges in (X, τ).

9. Give an example of a family $\{F_\alpha\}_{\alpha \in A}$ of subsets of $\mathbb{R}$ with the finite intersection property which has empty intersection. Do this in the open unit disk in the plane.

10. Let Y be the unit n-cube $Y = [0,1]^n \subseteq \mathbb{R}^n$. Give a good estimate for the smallest ϵ-net that you need to cover Y. (You'll need a formula that depends on n and ϵ!)

11. Complete the proof of Theorem 5.5.1.

12. Prove that if $\{x_k\}$ is a Cauchy sequence in a metric space and some subsequence of $\{x_k\}$ converges to a point x, then the original sequence converges to x too.

13. Give an example of a closed and bounded subset of $C[0,1]$ which is not compact. Prove your answer!

14. Given two disjoint subsets A, B of a metric space (X, d) we can define

$$\operatorname{dist}(A, B) = \inf\{d(a, b) \,:\, a \in A,\ b \in B\}.$$

(a) Give an example of two disjoint closed subsets of $\mathbb{R}^2$ which are distance zero apart.

(b) Can you find two disjoint closed connected subsets of $\mathbb{R}$ which are distance zero apart?

(c) Prove that if A and B are compact, then $d(A, B) > 0$. (What if only one of the sets is compact and the other is closed?)

15. There are fairly general conditions which allow you to embed a non-compact topological space X inside a compact one K_X in a way that its original topology is the same as the relative topology it inherits from K_X. For $X = \mathbb{R}$, you can do this by adding just a single point ∞ for which basic open neighbourhoods are of the form $(-\infty, x) \cup (y, \infty) \cup \{\infty\}$. Fill in the details! Give a familiar space which is homeomorphic to K_X for this example.

16. Let $X = \mathbb{Z}^+$ and let $K_X = X \cup \{\infty\}$. Let $K = \{0, 1, \frac{1}{2}, \frac{1}{3}, \dots\}$ which is compact in the usual topology. The map $\psi : K \to K_X$

$$\Psi(x) = \begin{cases} k, & \text{if } x = \frac{1}{k},\ k \in \mathbb{Z}, \\ \infty, & \text{if } x = 0 \end{cases}$$

is a bijection. Use this to impose a topology on K_X: $\tau = \{\Psi(U) \,:\, \text{U is open in } K\}$. Check that this topology makes K_X compact, and that the relative topology on X is just the usual one on $\mathbb{Z}^+$. Check that $C(K_X)$ is really nothing but the sequence space c.

17. Let (X, d) be a finite metric space with $X = \{x_1, \dots, x_n\}$. (Assume that $n \geq 2$.) Define sets

$$H = \{\mathbf{v} \in \mathbb{R}^m \,:\, \textstyle\sum_{i=1}^n v_i = 1\}, \qquad H^+ = \{\mathbf{v} \in H \,:\, v_i \geq 0 \text{ for all } i\}.$$

Quantities that appear in metric geometry are

$$I(\mathbf{v}) = \sum_{i=1}^n \sum_{j=1}^n v_i v_j d(x_i, x_j), \qquad \mathbf{v} \in H,$$

and

$$M(X, d) = \sup_{\mathbf{v} \in H} I(\mathbf{v}), \qquad M^+(X, d) = \sup_{\mathbf{v} \in H^+} I(\mathbf{v}).$$

(a) Prove that $M(X, d)$ and $M^+(X, d)$ must be positive.

(b) Must these two quantities be finite?

(c) Explain why there must exist $\mathbf{v}_0 \in H^+$ such that $M^+(X, d) = I(\mathbf{v}_0)$.

18. Let $f(x) = \sin \pi x$. How big does n need to be so that the nth Bernstein polynomial for f, $p_{f,n}$ on $[0,1]$ satisfies $\|f - p_{f,n}\|_\infty < \frac{1}{100}$. How big does n have to be before the nth Taylor polynomial for f

$$t_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k$$

satisfies $\|f - t_n\|_\infty < \frac{1}{100}$?

19. Get Maple to calculate the Bernstein polynomials $p_{f,n}$ for the function $f(x) = |x - \frac{1}{2}|$ on $[0,1]$. Plot the derivatives of these. What do you see?

20. (For those who know some complex analysis) Let $\overline{\mathbb{D}}$ denote the closed unit disk in $\mathbb{C}$. One might hope that the complex polynomials are dense in $(C(\overline{\mathbb{D}}), \|\cdot\|_\infty)$. Unfortunately this is not the case. Let $f(z) = \bar{z}$ and suppose that there is a polynomial p so that $\|f - p\|_\infty < \frac{1}{100}$. Let Γ be the path $\theta \mapsto e^{i\theta}$, $0 \le \theta \le 2\pi$.

    (a) Use the M-L lemma of contour integrals to show that

    $$\left| \int_\Gamma f(z) - p(z)\, dz \right| \le \frac{\pi}{50}.$$

    (b) Now calculate $\int_\Gamma f(z)\, dz$ (via the definition of contour integration) and $\displaystyle\int_\Gamma p(z)\, dz$ (via the Cauchy-Goursat Theorem) and arrive at a contradiction. (Thus f cannot be approximated by polynomials on the disk.)

21. Consider the following subsets of $\mathbb{C}$, whose descriptions are given in polar coordinates. (Take $r \ge 0$ in this question.)

$$\begin{aligned} X_1 &= \{(r,\theta) \,:\, r = 1\} \\ X_2 &= \{(r,\theta) \,:\, r < 1\} \\ X_3 &= \{(r,\theta) \,:\, 0 < \theta < \pi,\ r > 0\} \\ X_4 &= \{(r,\theta) \,:\, r = \cos 2\theta\}. \end{aligned}$$

    Give each set the usual topology inherited from $\mathbb{C}$. Which, if any, of these sets are homeomorphic?

22. Consider the set $\mathcal{A}$ of letters A–Z written in a sans serif font:

    A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z.

    Define an equivalence relation on $\mathcal{A}$ by saying that letter x is equivalent to letter y if their images in the above list are homeomorphic (as subsets of the plane). How many equivalence classes are there?

23. Let Y be the subset of $C[0,1]$ consisting of functions which don't vanish.

    (a) Prove that if $f \in Y$ then $\inf_{x\in[0,1]} |f(x)| > 0$, and hence that $\frac{1}{f} \in C[0,1]$. (Indeed Y is precisely the invertible elements in the algebra $C[0,1]$.)

    (b) Is Y connected? Does this depend on whether you use real or complex scalars?

    (c) Now try this with $C[0,1]$ replaced by the sequence space c (which is an algebra under elementwise multiplication). What are the invertible elements of c? Is the set of invertible elements connected?

24. Prove that at any given moment there are always two antipodal points on the earth's surface which are at exactly the same temperature. (What assumptions did you need to make?)

25. Prove that in the product topology on $\mathbb{R} \times \mathbb{R}$ (with the usual metric topology on $\mathbb{R}$) is just the usual metric topology on $\mathbb{R}^2$.

26. Prove that the product of two Hausdorff spaces is always Hausdorff under the product topology.

Part III

Dessert

Chapter 6

L^p spaces — a brief introduction

The remaining chapters are a very brief overview of some important topics which will be treated much more carefully in some later courses. But not everyone will do any more analysis courses, so this is a chance to at least get some feel for some of the objects which you are likely to come across as you continue your studies.

In these sections there are very few proofs, and we will often just concentrate on some of the main special cases rather than trying to be too general.

6.1 Introduction

So far, we have noted that if $1 \le p < \infty$ then

$$\|f\|_p = \int_0^1 |f(t)|^p \, dt$$

defines a norm on $C[0,1]$. Unlike the $\|\cdot\|_\infty$ norm however, this doesn't give a complete metric space.

You can try to fix this by adding in some extra functions, but once you allow discontinuous functions in your space, you need to deal with the problem that the above formula can give that $\|f\|_p = 0$ for a nonzero function f.

We'd like especially to have a nice space of functions for which $\|\cdot\|_2$ is a complete norm, as this would provide a Hilbert space.

The spaces that we will define are called the L^p spaces. One could define them very abstractly as the 'completions' of $C[0,1]$ under the above norms, but this would leave the very big question as to exactly what the additional new elements are.

Instead we will explain how to arrive at these spaces via Lebesgue integration. This is a generalization of the Riemann integration that you have been using up until now. Its main advantage is not so much that you can integrate (many) more functions[1], but that it behaves much better when it comes to taking limits.

[1] More accurately, many more functions are Lebesgue integrable than are Riemann integrable. However, our main tool for actually computing an integral remains the Fundamental Theorem of Calculus. Lebesgue integration will not magically tell you how to find $\int_0^1 \tan^2(e^{t\cos t})\, dt$.

In this chapter we'll explain the main ideas behind Lebesgue integration and the construction of L^p spaces, essentially omitting all of the proofs. The details can be studied in MATH5825. The L^p spaces are so important in analysis however that everyone should at least have a decent idea of what they are and what their main properties are. The main tool in Lebesgue's theory is that of a measure, and Measure Theory forms the basis of modern probability theory.

I'll label things as 'Fact' when I have no intention of explaining why they are true!

6.2 Simple functions and measure properties

For Riemann integration you essentially approximate a function $f : [0,1] \to \mathbb{R}$ by **step functions**. These are functions of the form

$$\phi(x) = \sum_{j=1}^{n} c_j \chi_{[a_{j-1}, a_j)}(x)$$

where $0 = a_0 < a_1 < \cdots < a_n = 1$ is some partition of $[0,1]$ and χ_S is the characteristic function of S:

$$\chi_S(x) = \begin{cases} 1, & \text{if } x \in S, \\ 0, & \text{otherwise.} \end{cases}$$

The main change in Lebesgue integration is that instead of only looking at linear combinations of characteristic functions of intervals, we allow linear combinations of characteristic functions of '**measurable**' sets.

Roughly speaking, the measure of a set is a generalization of 'length'. We'd **like** to be able to measure every subset $A \subseteq [0,1]$ with a number $0 \le \mu(A) \le 1$ which satisfies

1. $\mu([a,b]) = \mu((a,b)) = \mu([a,b)) = \mu((a,b]) = b - a$ for any $0 \le a < b \le 1$.
2. if $A \subseteq B$ then $\mu(A) \le \mu(B)$. (Monotonicity)
3. $\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2)$ if A_1, A_2 are disjoint. (Additivity)
4. if $B = \{x + t \pmod 1 : x \in A\}$ is the wrapped translation[2] of A by t then $\mu(A) = \mu(B)$. (Translation invariance)
5. and ???

Thus if $A = [0, \frac{1}{4}] \cup [\frac{1}{2}, 1]$ then you want $\mu(A) = \frac{3}{4}$.

A second way of thinking about μ is that $\mu(A)$ should be the probability that a 'randomly chosen' element of $[0,1]$ lies in A. The difficulty comes in making formal sense of 'randomly chosen'. Nevertheless, measure theory lies at the heart of probability theory, and thinking of things in this way can give you good intuition.

[2]'Rotation' is perhaps a better word here as we are thinking of $[0,1)$ as being joined up at the endpoints to form a circle.

The area bounded by the characteristic function of a set should just be the 'length' of the base times the height, which would just come to be $\mu(A) \times 1 = \mu(A)$. That is, we want to define

$$\int_0^1 \chi_A(x)\,dx = \mu(A)$$

for any set A. Characteristic functions can be pretty messy; think for example of $A = \mathbb{Q} \cap [0,1]$ for which χ_A is certainly not Riemann integrable.

More generally we could consider combinations of such characteristic functions.

Definition 6.2.1. A **simple function** on $[0,1]$ is a function $\phi : [0,1] \to \mathbb{R}$ of the form

$$\phi(x) = \sum_{j=1}^{n} c_j\, \chi_{A_j}(x)$$

where $\{A_j\}_{j=1}^{n}$ is a finite collection of subsets of $[0,1]$ and $\{c_j\}_{j=1}^{n}$ is a set of scalars.

The definition doesn't insist that the sets $\{A_j\}$ are disjoint. In some proofs it is helpful to be able to assume this. Check that we could add that in to the definition without any problem!

Since we want integration to be linear, the integral of such as function would need to be

$$\int_0^1 \phi(x)\,dx = \int_0^1 \sum_{j=1}^{n} c_j\,\chi_{A_j}(x)\,dx = \sum_{j=1}^{n} c_j \int_0^1 \chi_{A_j}(x)\,dx = \sum_{j=1}^{n} c_j\,\mu(A_j).$$

For a general $f : [0,1] \to \mathbb{R}$ the idea is to find simple functions ϕ_n converging to f and then take

$$\int_0^1 f(x)\,dx = \lim_{n\to\infty} \int_0^1 \phi_n(x)\,dx.$$

There are three big challenges in doing this:

- defining $\mu(A)$ for messy sets A,
- deciding which functions can be approximated by simple functions (and what topology to use!), and
- proving the convergence theorems, so that it is OK to say that $\displaystyle\int_0^1 \lim_{n\to\infty} \phi_n(x)\,dx = \lim_{n\to\infty} \int_0^1 \phi_n(x)\,dx$.

With Riemann integration we get that if $f_n \to f$ uniformly then $\int_0^1 f_n(x)\,dx \to \int_0^1 f(x)\,dx$. There are easy examples (even in $C[0,1]$) where this fails if f_n only converges to f pointwise. The limit theorems that we'll be looking to prove require a version of convergence somewhere between these. To prove these it turns out that you need a bit more than additivity of μ. You actually want countable additivity:

3′. (Countable additivity) $\mu(\bigcup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \mu(A_j)$ for any countable collection of disjoint subsets $\{A_j\}$.

The countability of the collection $\{A_j\}$ is vital. You can make sense of an infinite sum of non-negative numbers over an uncountable set. The real problem is that you could never hope to have 'uncountable additivity'. The translation invariance tells you that $\mu(\{x\})$ must be independent of x. But there is no value you could choose so that $\mu\big(\bigcup_{x\in[0,1]}\{x\}\big) = \mu([0,1]) = \sum_{x\in[0,1]} \mu(\{x\})$.

I wrote the countable collection $\{A_j\}$ as a sequence which provides an order for the infinite sum. Remember that in general the value of an infinite sum $\sum_{j=1}^{\infty} a_j = \lim_{N\to\infty} \sum_{j=1}^{N} a_j$ depends on the ordering of the terms. If the sum were only conditionally convergent you might get a different answer if you rearranged the terms. In general a countable collection may not come with a definitive ordering, so it is worthwhile noting that if, as is the cases here, all the terms in a convergent series are nonnegative, then the series must converge unconditionally, and it doesn't matter what order the sum is taken in.

6.3 Sets of measure zero

A consequence of the above properties for μ is that if $B \subseteq A_1 \cup A_2 = (A_1 \setminus A_2) \cup (A_2 \setminus A_1) \cup (A_1 \cap A_2)$ then

$$\begin{aligned}
0 \le \mu(B) &\le \mu(A_1 \setminus A_2) + \mu(A_2 \setminus A_1) + \mu(A_1 \cap A_2) \\
&\le (\mu(A_1 \setminus A_2) + \mu(A_1 \cap A_2)) + (\mu(A_2 \setminus A_1) + \mu(A_1 \cap A_2)) \\
&= \mu(A_1) + \mu(A_2).
\end{aligned}$$

More generally, if we have countable additivity, and if $B \subseteq \bigcup_{j=1}^{n} A_j$ then we should have that $0 \le \mu(B) \le \sum_{j=1}^{\infty} \mu(A_j)$.

Example 6.3.1. Let $B = \mathbb{Q} \cap [0,1]$ which we can write as a list $B = \{r_1, r_2, r_2, \dots\}$. Fix $\epsilon > 0$. Let I_1 be any small interval containing r_1 of length $\frac{\epsilon}{2}$, let I_2 be any small interval containing r_2 of length $\frac{\epsilon}{4}$, and so on. Then $B \subseteq \bigcup_{j=1}^{n} I_j$ and $\sum_{j=1}^{\infty} \text{length}(I_j) = \sum_{j=1}^{\infty} \frac{\epsilon}{2^j} = \epsilon$. (It is interesting to ask yourself just which points are not in $\bigcup_{j=1}^{n} I_j$.)

As this is true for any $\epsilon > 0$, the properties we want of a measure force that $\mu(B) = 0$.

Exercise 6.3.2. More directly, but less helpfully for what is to come, it is not hard to show that if x is any point in $[0,1]$ then $\mu(\{x\})$ must be zero. In the notation of the previous example $B = \bigcup_{i=1}^{\infty} \{r_i\}_{i=1}^{\infty}$ and so the countable additivity would tell us that $\mu(B)$ must be zero too.

Definition 6.3.3. Suppose that $B \subseteq [0,1]$. We say that B has **measure zero** if for all $\epsilon > 0$ there exists a countable collection of intervals I_j such that $B \subseteq \bigcup_{j=1}^{n} I_j$ and $\sum_{j=1}^{\infty} \text{length}(I_j) \le \epsilon$.

In fact we have just proved the following

Theorem 6.3.4. *Every countable subset of* $[0, 1]$ *has measure zero.*

Is every set of measure zero countable? The answer is No. The standard example of an uncountable set of measure zero is the Cantor middle third set. This is the set C of all numbers $x \in [0, 1]$ which have a ternary (base 3) expansion

$$x = 0.t_1t_2t_3\dots$$

with no t_j equal to 2. There is an easy bijection between C and the set of all countably infinite bit strings so C is uncountable. The heuristic argument that this has measure zero is that the probability that you randomly choose an infinite string of elements of the set $\{0, 1, 2\}$ and never choose a 2 is surely zero. You should try to give a proof from the definition.

The integral of the characteristic function of a set of measure zero will be zero, so these are the sort of functions that will cause us problems in defining a norm via something like $\|f\| = \int_0^1 |f(t)|\, dt$. One way[3] to fix this problem is to essentially declare that if $\|f - g\| = 0$, then we will regard f and g are being the same point in our vector space!

Definition 6.3.5. If something happens for all x except for a set of measure zero, we shall say that if happens **almost everywhere** (a.e.).

Thus two functions f, g are equal almost everywhere if $f(x) = g(x)$ except on a set of measure zero.

6.4 σ-algbras and Lebesgue measure

It is an unfortunate consequence of the Axiom of Choice that it is impossible to measure all subsets on $[0, 1]$ in a sensible way that satisfies all the above requirements, and this is one of the main source of complication in the theory of Lebesgue integration. Subsets of $[0, 1]$ for which a sensible concept of 'generalized length' can be defined are called 'measurable'.

Note for example that if A is measurable then by (3) (or (3')) above, you should have

$$\mu([0, 1]) = 1 = \mu(A) + \mu([0, 1] \setminus A), \quad \text{or} \quad \mu([0, 1] \setminus A) = 1 - \mu(A).$$

That is, $[0, 1] \setminus A$ should be measurable too. More generally (3') says that countable unions of measurable sets should be measurable. Since

$$\bigcap_{j=1}^{\infty} A_j = [0, 1] \setminus (\bigcup_{j=1}^{\infty}([0, 1] \setminus A_j))$$

[3]Another option is to just live with the fact that you have a seminorm rather than a norm. Probabilists tend to choose this option as they like to work with actual functions. Analysts tend to prefer to deal with actual norms rather than seminorms, and so they live with the fact that the objects in their space are not quite functions. We'll come back to this in Section 6.7.

we get that intersections of measurable sets should be measurable too. This means that we need to be trying to define μ on a collection of sets called a σ-algebra.

Definition 6.4.1. A collection $\mathcal{A}$ of subsets of $[0,1]$ is called a **σ-algebra** if

1. $\emptyset$, $[0,1] \in \mathcal{A}$.
2. if $A \in \mathcal{A}$ then $([0,1] \setminus A) \in \mathcal{A}$
3. $\mathcal{A}$ is closed under countable unions and countable intersections.

This looks a little like the definition of a topology. The main difference is that a σ-algebra is closed under taking complements.

Example 6.4.2. (i) A big σ-algebra is $\mathcal{P}([0,1])$.
(ii) A little one is $\{\emptyset, [0, \frac{1}{2}), [\frac{1}{2}, 1], [0,1]\}$.

Question 6.4.3. What we want is the largest σ-algebra on which the measure can be defined.

In formally constructing Lebesgue measure μ, one can start, as we did, by defining $\mu(A)$ for intervals. You can then extend the definition to countable unions and intersections of intervals. And then to countable unions and intersections of those sets, and .... At each stage you can use the above properties to define $\mu(A)$ for more and more complicated sets, and importantly, this doesn't ever run into a problem. Unfortunately, this procedure doesn't terminate after a finite number of steps; at each stage you get more sets.

Fact 6.4.4. *There is a smallest σ-algebra $\mathcal{B}$, called the **Borel σ-algebra**, that contains all the open intervals in $[0,1]$. The elements of this σ-algebra are called **Borel sets**.*

Note that since every open subset of $[0,1]$ is a countable union of open intervals, every open set is Borel. And since σ-algebras are closed under taking complements, all closed sets are Borel too. It must contain all the singleton sets. For example $\{\frac{1}{2}\}$ is the complement of $[0, \frac{1}{2}) \cup (\frac{1}{2}, 1]$ and $\mu(\{\frac{1}{2}\}) = 1 - \frac{1}{2} - \frac{1}{2} = 0$. By (3') then every countable set S is in $\mathcal{B}$ with $\mu(S) = 0$.

Fact 6.4.5. *The Borel σ-algebra is strictly smaller that $\mathcal{P}([0,1])$.*

The Borel sets are however a large enough class that they suffice for most purposes. Trying to come up with a non-Borel set is a bit of a challenge; pretty much anything that you can write down 'naturally' turns out to be Borel.

The problem with the Borel σ-algebra is that it is possible to find a Borel set A with $\mu(A) = 0$ and a non-Borel set B such that $B \subseteq A$. If we can cover A with lots of very small intervals we can surely also cover B in this way so B is also a set of measure zero. Adding in these extra funny sets turns out to be important for proving the convergence theorems that we are looking for.

Fact 6.4.6. *There is a smallest σ-algebra $\mathcal{L}$, called the **Lebesgue σ-algebra**, that contains all the Borel sets and all the sets of measure zero in* $[0,1]$.

Again you don't get all the subsets of $[0,1]$, but it is again quite hard to construct a set which is not in $\mathcal{L}$.

Fact 6.4.7. *There is a unique map $\mu : \mathcal{L} \to [0,1]$, called **Lebesgue measure**, which satisfies properties (1), (2), (3′) and (4) from Section 6.2. The sets in $\mathcal{L}$ are said to be **Lebesgue measurable**.*

In fact things are not quite as mysterious as they may appear.

Fact 6.4.8. *For any set $A \in \mathcal{L}$,*

$$\mu(A) = \inf \sum_{j=1}^{\infty} \text{length}(I_j)$$

where the infimum is taken over all collections of open intervals such that $A \subseteq \bigcup_{j=1}^{\infty} I_j$.

The quantity on the right-hand side of the about equation exists for any set $A \subseteq [0,1]$, not just the measurable ones, so you could try defining μ on all sets by this formula. This gives what is called 'outer measure'; it behave quite like a measure, but it doesn't satisfy condition (3').

More generally one can work on an arbitrary **measure space**, that is a set X with some σ-algebra $\mathcal{A}$. A general measure on $(X, \mathcal{A})$ is a function $m : \mathcal{A} \to [0, \infty]$ which is countably additive and for which $m(\emptyset) = 0$. This greater generality is important, for example, for applications in probability theory.

6.5 Lebesgue integration

Definition 6.5.1. A **simple measurable function** is a function $\phi : [0,1] \to \mathbb{R}$ of the form

$$\phi(x) = \sum_{j=1}^{n} c_j \chi_{A_j}(x)$$

where $\{A_j\}_{j=1}^n$ is a finite collection of Lebesgue measurable subsets of $[0,1]$ and $\{c_j\}_{j=1}^n$ is a set of scalars.

Note that such a function only take on finitely many values. If we insist that the A_j are disjoint then if U is any subset of $\mathbb{R}$, then $f^{-1}(U)$ must be some union of the A_j and hence is certain in $\mathcal{L}$.

Definition 6.5.2. The **Lebesgue integral** of a simple measurable function $\phi(x) = \sum_{j=1}^n c_j \chi_{A_j}(x)$ is defined to be

$$\int_0^1 \phi(x)\, dx = \sum_{j=1}^{n} c_j \mu(A_j).$$

Remark 6.5.3. You'll see a range of different notation here for the integral:

$$\int \phi \, d\mu, \qquad \int_{[0,1]} \phi(x) \, d\mu(x), \qquad \int_0^1 \phi(x) \, \mu(dx), \qquad \dots$$

We'll suppress the μ as we won't be using any other measures.

It is easy to check that the set of simple measurable functions is a vector space and that the map $\phi \mapsto \int_0^1 \phi(x) \, dx$ is a linear transformation. (One should also check that the integral is in fact well-defined!) It is also an order-preserving map: if $0 \leq \phi \leq \psi$ on $[0, 1]$ then $0 \leq \int_0^1 \phi(x) \, dx \leq \int_0^1 \psi(x) \, dx$.

The general idea now is that if $f : [0, 1] \to \mathbb{R}$ can be written as a 'limit' of simple measurable functions ϕ_n then we set $\int_0^1 f(x) \, dx = \lim_{n\to\infty} \int_0^1 \phi_n(x) \, dx$.

The concept that we need to proceed is that of a measurable function.

Definition 6.5.4. A function $f : [0, 1] \to \mathbb{R}$ is **Lebesgue measurable** if for each $\alpha \in \mathbb{R}$, the set $f^{-1}((-\infty, \alpha))$ is a Lebesgue measurable subset of $\mathbb{R}$.

Actually, it makes no difference whether you use $(-\infty, \alpha)$ or $(-\infty, \alpha]$ or any other suitably large class of intervals in the definition. We will also extend the definition to allow functions like $f(x) = \frac{1}{x}$ which are infinite at certain points. Clearly simple measurable functions are indeed measurable, so we have at least some measurable functions.

Exercise 6.5.5. Use the open set characterization of continuity to show that every continuous function on $[0, 1]$ is Lebesgue measurable.

In practice one rarely uses the definition to show that a function is Lebesgue measurable. This is of course also the situation for Riemann integration. Rather, you collect a useful collection of conditions which are sufficient for measurability. For example, any piecewise continuous function is Lebesgue measurable.

Fact 6.5.6. *The set of Lebesgue measurable functions is an algebra (that is, a vector space that is closed under multiplication too).*

What is just as important, is that this set is closed under taking suitable limits.

Fact 6.5.7. *Suppose that* $\{f_k\}_{k=1}^{\infty}$ *is a sequence of Lebesgue measurable functions on* $[0, 1]$*. Then the following functions are also Lebesgue measurable:*

1. $g_1(x) = \sup_k f_k(x)$.

2. $g_2(x) = \inf_k f_k(x)$.

3. $g_3(x) = \lim_{k\to\infty} f_k(x)$, *if the sequence converges pointwise.*

4. $g_4(x) = \limsup_k f_k(x)$.

Much easier is:

Fact 6.5.8. *If f is measurable and $f = g$ almost everywhere, then g is measurable.*

Perhaps more surprising is the following, which depends on the Weierstrass Approximation Theorem.

Fact 6.5.9. *Suppose that $f : [0,1] \to \mathbb{R} \cup \{-\infty, \infty\}$ is measurable, and that $f(x)$ is finite almost everywhere. Then for all $\epsilon > 0$ there exist a step function ϕ and a continuous function g such that*

$$|f(x) - \phi(x)| < \epsilon \qquad \text{and} \qquad |f(x) - g(x)| < \epsilon$$

for all x in $[0,1]$ except a set of measure less than ϵ.

Suppose that $f : [0,1] \to \mathbb{R}$ is bounded and measurable and that n is a positive integer. For $j \in \mathbb{Z}$ let

$$A_j = f^{-1}\Big(\Big[\frac{j}{n}, \frac{j+1}{n}\Big)\Big).$$

As f is bounded only finitely many of these sets are nonempty, say $a \le j \le b$. Then

$$\phi_n = \sum_{j=a}^{b} \frac{j}{n}\, \chi_{A_j}$$

is a simple measurable function such that $\|f - \phi_n\|_\infty < \frac{1}{n}$. We define

$$\int_0^1 f(x)\, dx = \lim_{n \to \infty} \int_0^1 \phi_n(x)\, dx.$$

There is of course some work to be done to check that the limit must exist.

If $0 \le f$ then we define

$$\int_0^1 f(x)\, dx = \sup \int_0^1 h(x)\, dx$$

where the supremum is over all bounded measurable functions with $0 \le h \le f$. We say that f is **integrable** if this supremum is finite.

If f is a general measurable function, we define

$$f^+(x) = \begin{cases} f(x), & \text{if } f(x) \ge 0, \\ 0, & \text{otherwise,} \end{cases} \qquad f^-(x) = \begin{cases} -f(x), & \text{if } f(x) \le 0, \\ 0, & \text{otherwise.} \end{cases}$$

so that $f = f^+ - f^-$ with $f^+, f^- \ge 0$. Then f is **integrable** if both f^+ and f^- are and we set

$$\int_0^1 f(x)\, dx = \int_0^1 f^+(x)\, dx - \int_0^1 f^-(x)\, dx.$$

It is important that what we are doing really is an extension of what you did at school.

Fact 6.5.10. *If f is Riemann integrable then it is Lebesgue integrable too and you get the same answer!*

You almost never use the definition of Riemann integration to actually calculate $\int_0^1 f(x)\,dx$. Rather you find an antiderivative and then use the Fundamental Theorem of Calculus.

6.6 Convergence Theorems

The advantage of Lebesgue integration is that it gives us an additional tool for calculating integrals — taking limits. That is, if $f = \lim_n f_n$ and you can calculate $\int_0^1 f_n(x)\,dx$ for each n, you'd like to be able to say that f is Lebesgue integrable with $\int_0^1 f(x)\,dx$ being equal to $\lim_n \int_0^1 f_n(x)\,dx$.

Riemann integration lacks good convergence theorems!

Example 6.6.1. As before write $B = \mathbb{Q} \cap [0,1] = \{r_1, r_2, \dots\}$. For $k = 1, 2, 3, \dots$ let

$$f_k(x) = \begin{cases} 1, & \text{if } x \in \{r_1, \dots, r_k\}, \\ 0, & \text{otherwise.} \end{cases}$$

The each function f_k is Riemann integrable. The sequence converges pointwise to the characteristic function of B, but this limit is not Riemann integrable. However χ_B is Lebesgue integrable and $\int_0^1 f_k(x)\,dx \to \int_0^1 \chi_B(x)\,dx$.

More generally:

Theorem 6.6.2 (Monotone Convergence Theorem)**.** *Suppose that $\{f_k\}$ is a sequence of measurable functions on $[0,1]$ such that*

(i) $0 \le f_1 \le f_2 \le \dots$.

(ii) $f_k \to f$ *pointwise.*

(iii) $\displaystyle\lim_{k\to\infty} \int_0^1 f_k(x)\,dx < \infty$.

Then f is Lebesgue integrable and $\displaystyle\int_0^1 f(x)\,dx = \lim_{k\to\infty} \int_0^1 f_k(x)\,dx.$

Exercise 6.6.3. Give an example to show that this fails if the sequence is not monotone.

If you don't have a monotone sequence then you may be able to get by by just knowing that your functions are not bouncing around too much.

Theorem 6.6.4 (Dominated Convergence Theorem)**.** *Suppose that $\{f_k\}$ is a sequence of integrable functions on $[0,1]$ such that* *(i)* $f_k \to f$ *pointwise.*

(ii) *there is an integrable function* g *such that* $|f_k(x)| \leq |g(x)|$ *for all* k *and all* $x \in [0,1]$.

(iii) $\displaystyle\lim_{k\to\infty} \int_0^1 f_k(x)\,dx < \infty$.

Then f *is Lebesgue integrable and* $\displaystyle\int_0^1 f(x)\,dx = \lim_{k\to\infty} \int_0^1 f_k(x)\,dx$.

6.7 L^p spaces

Suppose that $1 \leq p < \infty$. Define

$$\mathcal{L}^p[0,1] = \left\{ f : [0,1] \to \mathbb{R} \,:\, f \text{ is measurable and } \int_0^1 |f(x)|^p\,dx < \infty \right\}.$$

This set contains, for example, all continuous functions on $[0,1]$ and all simple measurable functions.

Question 6.7.1. Is $\mathcal{L}^p[0,1]$ a vector space?

The answer is Yes, but this is not exactly easy! For $f \in \mathcal{L}^p[0,1]$ define

$$N_p(f) = \left(\int_0^1 |f(x)|^p\,dx \right)^{1/p}.$$

This is not a norm since $N_p(f) = 0$ does not imply $f \equiv 0$. It does however have the other properties of a norm, which makes it what is known as a **seminorm**. In particular it satisfies the triangle inequality.

Theorem 6.7.2 (Minkowski's Inequality)**.** *If* $f, g \in \mathcal{L}^p[0,1]$ *then* $f + g \in \mathcal{L}^p[0,1]$ *and*

$$N_p(f+g) \leq N_p(f) + N_p(g).$$

Minkowski's Inequality depends on another important result:

Theorem 6.7.3 (Hölder's Inequality)**.** *Suppose that* $1 < p, q < \infty$ *and* $\frac{1}{p} + \frac{1}{q} = 1$. *If* $f \in \mathcal{L}^p[0,1]$ *and* $g \in \mathcal{L}^q[0,1]$ *then* $fg \in \mathcal{L}^1[0,1]$ *and*

$$N_1(fg) \leq N_p(f) \cdot N_q(g).$$

Note that if $p = q = 2$ this is essentially just Cauchy-Schwarz. The theorem also holds when $p = 1$ and $q = \infty$ where $N_\infty(g)$ is interpreted as the usual supremum norm. The proofs are now essentially in MATH2701!

The standard way to turn a norm into a seminorm is to form a quotient space, by declaring all the elements where $N(f-g)=0$ to be equal. More formally, we say that $f,g \in \mathcal{L}^p[0,1]$ are equivalent, denoted $f \sim g$, if $f(x)=g(x)$ almost everywhere. It is easy to check that $\sim$ is an equivalence relation, and that

$$f \sim g \iff N_p(f-g)=0.$$

The equivalence class of f is defined as

$$[f]=\{g \in \mathcal{L}^p[0,1] \,:\, f \sim g\}.$$

Formally, $L^p[0,1]$ is the normed vector space of all equivalence classes of functions in $\mathcal{L}^p[0,1]$ under the norm

$$\|[f]\|_p=\left(\int_0^1 |f(x)|^p\,dx\right)^p$$

and vector space operations

$$[f]+[g]=[f+g], \qquad \lambda[f]=[\lambda f].$$

Fact 6.7.4. *These definitions don't depend on which elements of the equivalence class you choose!*

In practice, we usually work as if the elements of $L^p[0,1]$ are actually functions and not equivalence classes of functions. So people will write 'if $f \in L^2[0,1]$ then it is in $L^1[0,1]$' and so forth. We'll certainly write $\|f\|_p$ rather than $\|[f]\|_p$ or $N_p(f)$. On the other hand, writing

$$S=\{f \in L^p[0,1] \,:\, f(0)=0\}$$

is not allowed since the condition here depends on which representative of the equivalence class you take.

The way that you will usually see Minkowski and Hölder is thus that

$$\|f+g\|_p \leq \|f\|_p+\|g\|_p \qquad \text{and that} \qquad \|fg\|_1 \leq \|f\|_p\,\|g\|_q\,.$$

Note also that you can consider $(C[0,1],\|\cdot\|_p)$ as a subset of all these spaces since each continuous f belongs to a different equivalence class.

The important fact about $L^p[0,1]$ is:

Theorem 6.7.5 (Riesz-Fischer Theorem)**.** *For $1 \leq p < \infty$, $L^p[0,1]$ is complete, and is therefore a Banach space.*

The proof uses the Dominated Convergence Theorem that we saw earlier.

If $1 \leq p < \infty$ then $C[0,1]$ forms a dense subset of $L^p[0,1]$ in each of these norms, so you can consider $L^p[0,1]$ as the completion of $C[0,1]$ under the p-norm.

Because we are dealing with equivalence classes, we are not really concerned if a function is not defined on some set of measure zero. Many authors explicitly allow $f(x)=\infty$ (or $-\infty$) on a set of measure zero. Thus we will happily write that $f(x)=x^{-1/2}$ is in $L^1[0,1]$ but not in $L^2[0,1]$.

6.8 The case $p = \infty$

The case $p = \infty$ is slightly different. Of course $C[0,1]$ *is* complete under the $\|\cdot\|_\infty$ norm, so we are not taking a completion at all here.

Heuristically, we want $L^\infty[0,1]$ to be the space of all bounded measurable functions. To get consistency with the other L^p spaces we want to actually work with equivalence classes of measurable functions, and we now run into the problem that the norm

$$\|[f]\|_\infty = \sup_{x\in[0,1]} |f(x)|$$

isn't well-defined! Even worse, you can have $f \sim 0$ and f not even being bounded.

The fix to this is to use the essential supremum. This is perhaps best explained via an example. Let

$$f(x) = \begin{cases} 1 - (2x-1)^2, & \text{if } x \neq \frac{1}{2} \\ 3, & \text{otherwise.} \end{cases}$$

Then $\sup_x |f(x)| = 3$. On the other hand, if you change f on a small (ie measure zero) set, you can make the function have supremum equal to 1 — but not anything smaller. The essential supremum of f then is the smallest value of the supremum of any function which is equivalent to f. This clearly depends only on the equivalence class of f. If $f \in C[0,1]$ then the essential supremum of f is the same as the supremum of f.

We therefore define $L^\infty[0,1]$ to be the vector space of eqivalence classes of measurable functions f for which $\|f\|_\infty = \operatorname*{ess\text{-}sup}_{x\in[0,1]} |f(x)|$ is finite.

With this definition, we get Minkowski and Hölder working when $p = \infty$ too.

Theorem 6.8.1. *$L^\infty[0,1]$ is a Banach space under the essential supremum norm.*

Actually, $L^\infty[0,1]$ has even more structure. If $f, g \in L^\infty[0,1]$ then $fg \in L^\infty[0,1]$ with $\|fg\|_\infty \le \|f\|_\infty \|g\|_\infty$. This means that $L^\infty[0,1]$ is a (commutative) **Banach algebra**. Even better, the complex space has an involution $f^* = \overline{f}$ which is something which behaves like the complex conjugate in $\mathbb{C}$. This involution satisfies the condition that $\|f^*f\|_\infty = \|f\|_\infty^2$, which is known as the C^*-condition. This makes $L^\infty[0,1]$ into a special type of Banach algebra called a **C^*-algebra** (and indeed even a von Neumann algebra!). The study of commutative C^*-algebras lies at the heart of proving the **Spectral Theorem**, which is the infinite dimensional version of the theorem that a self-adjoint matrix can be diagonalized. The other L^p spaces don't enjoy this multiplicative structure.

The space $L^\infty[0,1]$ turns out to be a rather difficult beast from many viewpoints. One problem is that it is very 'big', in a certain sense. More precisely, it is not too hard to prove that the set of polynomials with rational coefficients is a countable dense subset of $L^p[0,1]$ for $1 \le p < \infty$, and consequently that these spaces are 'separable'. On the other hand, you can use a diagonalization argument to show that $L^\infty[0,1]$ is not separable.

This is a little counterintuitive, since $L^\infty[0,1]$ is a subset of all the other L^p spaces. Indeed, they all have cardinality equal to that of $\mathbb{R}$. The issue here is really that it is hard to get convergence in $L^\infty[0,1]$.

More subtle is that the unit ball of $L^\infty[0,1]$ has lots of flat faces. This geometric fact will turn out to make the weak topology on this Banach space both hard to work with and lacking good convergence properties.

6.9 General measure spaces

In this chapter we have concentrated on integration with respect to Lebesgue measure on $[0,1]$. One can do all this in much greater generality! Given a nonempty set X and a σ-algebra $\mathcal{A}$ of subsets of X, a (countably additive scalar) measure is a map $\mu : \mathcal{A} \to [0,\infty]$ such that

1. $\mu(\emptyset) = 0$, and
2. if $\{A_k\}_{k=1}^\infty$ is a sequence of pairwise disjoint elements of $\mathcal{A}$ then

$$\mu\Big(\cup_{k=1}^\infty A_k = \sum_{k=1}^\infty \mu(A_k).$$

If $\mu(X) = 1$ then $(X, \mathcal{A}, \mu)$ is called a probability space. The general integration theory is not too different in the general case. In many ways the main complication that arises in the previous sections is not the integration theory, but rather showing that you can construct a suitable σ-algebra and measure on $[0,1]$. But for integration one proceeds in the same way. For a simple measurable function $\phi = \sum_{j=1}^n c_j \chi_{A_j}$ one defines $\displaystyle\int_X \phi\, d\mu = \sum_{j=1}^n c_j \mu(A_j)$. For a more general measurable function, one writes it as a limit $f = \lim_n \phi_n$ of simple functions and defines $\displaystyle\int_X f\, d\mu = \lim_{n\to\infty} \int_X \phi_n\, d\mu$. There is of course quite a bit of technical work to do to check that everything works as it should.

A few important cases are:

Example 6.9.1. $X = \mathbb{R}$. One can generate the Borel and Lebesgue σ-algebras and Lebesgue measure here. The main problem is that you have lots of sets of infinite measure so not all simple measurable functions are integrable. And of course even the bounded continuous functions need not sit in $L^p(\mathbb{R})$. We do however get nice Banach spaces of (equivalence classes of) functions.

Example 6.9.2. In probability theory you want $\mu(A)$ to be the probability that a randomly chose element of the sample space lies in the "event" A. Suppose that $p(x)$ is a probability density function on $\mathbb{R}$, so $p(x) \geq 0$ and $\int_{-\infty}^\infty p(x)\, dx = 1$. Then for a Lebesgue measurable set A, setting $\mu(A) = \int_{-\infty}^\infty \chi_A(x) p(x)\, dx$ defines a measure on $\mathbb{R}$.

Example 6.9.3. Let $X = \mathbb{Z}^+$ and let $\mathcal{A} = \mathcal{P}(X)$ be the power set σ-algebra. Counting measure μ_c on a subset $A \subseteq X$ is just the number of elements of X. Here $\int_{\mathbb{Z}} f(n)\, d\mu(n)$

is just a fancy way of writing $\sum_{n=1}^{\infty} f(n)$! Things are rather easier here as there are no nonempty sets of measure zero to worry about.

The L^p spaces here just turn out to be the usual ℓ^p spaces.

6.10 Nesting properties

Recall that if $1 \leq p < q < \infty$ then $\ell^1 \subset \ell^p \subset \ell^q \subset c_0 \subset \ell^\infty$. On $[0,1]$ (or indeed any space of total measure one), the L^p spaces nest in the other direction:

$$C[0,1] \subset L^\infty[0,1] \subset L^q[0,1] \subset L^p[0,1] \subset L^1[0,1].$$

This follows from the fact that the map $p \to \|f\|_p$ is increasing. This is not especially easy to see — try it. (It *is* easy to check this for a characterisic function!)

Exercise 6.10.1. (i) Write down a function which is in $L^2[0,1]$ but which is not in $L^1[0,1]$.

(ii) Now, for general $p < q$, find a function which is in $L^q[0,1]$ but not in $L^p[0,1]$.

Even for Lebesgue measure on different subsets of $\mathbb{R}$, this are more complicated. A vital ingredient in the proof that $p \to \|f\|_p$ is increasing is that $\mu([0,1]) = 1$. If you change from $[0,1]$ to a different finite interval then the L^p spaces still nest as above, but with factors (depending on the length of the interval) appearing in the inequalities comparing the different L^p norms. To keep things tidy, people often work with scaled versions of Lebesgue measure. For example, in Fourier analysis, one typically works on $[0,2\pi]$ with the measure $\frac{\mu}{2\pi}$ to give the whole interval measure one again.

If you are working on $\mathbb{R}$, then there is no general nesting properties of the L^p spaces. The one useful fact is that there is a very nicely behaved space of functions which is dense in $L^p(\mathbb{R})$ for all $p < \infty$. This is the Schwartz space $\mathcal{S}$ of infinitely differentiable functions, all of whose derivatives are rapidly decreasing. (This includes, in particular, functions whose support is contained in a finite interval.) The Fourier Transform

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi}\, dx, \qquad \xi \in \mathbb{R}$$

works in a nice way on $\mathcal{S}$ and this helps you extend it to the L^p spaces. This is studied in the 4th year course on Harmonic Analysis.

6.11 Problems

1. Thinking probabilistically, what should the measures of the following subsets of $[0,1]$ be?

   (a) The set S_1 of all x which have two different decimal expansions (such as $0.4\overline{9} = 0.5$).

(b) The set S_2 of all x for which the first 0 in the decimal expansion of x occurs in an odd indexed position (so $0.0123 \in S_2$ and $0.123406 \in S_2$ but $0.10203 \notin S_2$).

(c) The set S_3 of all x for which any 7 in the decimal expansion of x occurs after the first 8.

2. Look up the construction of a nonmeasurable subset of $[0, 1]$. Note that this depends rather critically on the Axiom of Choice. (In 1970, Solovay showed that there are models of ZF in which every subset of $[0, 1]$ is Lebesgue measurable!)

3. Consider the map $J : C[0, 1] \to L^{[}0, 1]$ (where $1 \leq p \leq \infty$) defined by $J(f) = [f]$.

   (a) Mentally check that this makes sense!

   (b) Show that J is one-to-one.

   (c) Show that J is continuous and linear.

   (d) How are $\|f\|_\infty$ and $\|[f]\|_p$ related?

   (When we say $C[0, 1] \subseteq L^p[0, 1]$ are actually talking about the image of $C[0, 1]$ under J.)

4. Suppose that $1 \leq p < \infty$. Given $f \in L^p[0, 1]$, can you always find $g \in L^q[0, 1]$ so that you get equality in Hölder's inequality? (Does it make any difference if you assume that f and g are continuous?)

5. The definition of a general measure in Section 6.9 only has two conditions. Explain why these conditions imply the simpler condition that $\mu(A \cup B) = \mu(A) + \mu(B)$ if A and B are disjoint.

6. Following Example **??**, check that if Z is a continuous random variable with probability density function $p : \mathbb{R} \to [0, 1]$ then $\mathbb{E}(Z) = \int_{\mathbb{R}} x \, d\mu(x)$.

7. As in Example 6.9.3, let $X = \mathbb{N}^+$ with $\mathcal{A} = \mathcal{P}(X)$ the power set σ-algebra. Let $\{p_k\}_{k=0}^\infty$ be a probability distribution for some random variable Z. Of course a sequence is just a fancy way of writing a function $p : \mathbb{N} \to \mathbb{R}$. Define $\mu : \mathcal{A} \to [0, 1]$ by $\mu(A) = \sum_{k \in A} p_k$, $A \subseteq X$.

   (a) Check that μ is a countably additive measure in the sense of Section 6.9.

   (b) Check that the formula for the mean of Z is just $\mathbb{E}(Z) = \int_X x \, d\mu(x)$.

The last two exercises show that the distinct formulas you learnt for discrete and continuous random variables in first year can all be combined if you write everything using measure theory. Measure theory also allows you to deal with random variable, such as the amount of rainfall on a given day, which are neither discrete nor continuous.

Chapter 7

Linear functionals and the weak topology

7.1 Introduction

We looked earlier at the weak topology in a Hilbert space. This turns our to generalize to Banach spaces in a nice way.

Recall that $x_n \to x$ weakly in a Hilbert space $\mathcal{H}$ if

$$\langle x_n, y \rangle \to \langle x, y \rangle, \qquad \text{for all } y \in \mathcal{H}.$$

The idea is to note that for each $y \in \mathcal{H}$ the map $\phi_y : \mathcal{H} \to \mathbb{R}$

$$\phi_y(x) = \langle x, y \rangle$$

is a continuous linear transformation. Indeed, every continuous linear transformation from $\mathcal{H} \to \mathbb{R}$ is of this form! This is one of the Riesz Representations Theorems.

For a Banach space X, the idea is to ignore the lack of an inner product and to just look at the continuous linear transformations from $X \to \mathbb{R}$.

I'll do everything here with real scalars, but you could just as easily use complex ones. The main extra complication is that in a complex Hilbert space $y \mapsto \langle x, y \rangle$ is not linear in y but conjugate linear.

7.2 The dual space

Suppose that $(X, \|\cdot\|)$ is a Banach space. The **dual space** to X, denoted X^*, is the set of all continuous linear maps from X to $\mathbb{R}$. The elements of X^* are called continuous **linear functionals** on X. These always form a vector space. Even better, we can make X^* into a Banach space.

Theorem 7.2.1. *Suppose that X is a Banach space and that $\phi : \mathbf{X} \to \mathbb{R}$ is a linear transformation. Then ϕ is continuous if and only if $\sup_{\|x\|=1} |\phi(x)|$ is finite.*

Proof. See exercises. ■

We can turn X^* into a normed space by setting

$$\|\phi\|_* = \sup_{\|x\|=1} |\phi(x)| = \sup_{x\neq 0} \frac{|\phi(x)|}{\|x\|}.$$

Exercise 7.2.2. Prove that $|\phi(x)| \leq \|x\| \, \|\phi\|_*$ and that $\|\phi\|_*$ is the smallest number K so that $|\phi(x)| \leq K \|x\|$ for all $x \in X$.

Fact 7.2.3. $(X^*, \|\cdot\|_*)$ *is always complete and hence is a Banach space.*

Example 7.2.4. Suppose that $X = C[0,1]$ with the sup norm. There are lots of continuous linear functionals:

$$\begin{aligned}
\phi(f) &= f(0) && \|\phi\|_* = 1 \\
\phi(f) &= \int_0^1 f(t)\,dt && \|\phi\|_* = 1 \\
\phi(f) &= f(0) + \int_0^1 f(t)e^t\,dt && \|\phi\|_* \leq e?
\end{aligned}$$

Example 7.2.5. Let $X = \ell^1$. Let $y = (y_1, y_2, \dots) \in \ell^\infty$. Define

$$\phi_y(x_1, x_2, \dots) = \sum_{k=1}^{\infty} x_k y_k.$$

Then ϕ is clearly linear and

$$|\phi_y(x)| \leq \sum_{k=1}^{\infty} |x_k y_k| \leq \|y\|_\infty \sum_{k=1}^{\infty} |x_k| = \|x\|_1 \|y\|_\infty .$$

Thus every element of ℓ^∞ produces a (distinct) element of X^*. Furthermore $\|\phi_y\|_* = \|y\|_\infty$.

Theorem 7.2.6 (Riesz Representation Theorem)**.** *Let $\mathcal{H}$ be a (real) Hilbert space. The $(\mathcal{H}^*, \|\cdot\|_*)$ is* ***isometrically*** *isomorphic to $(H, \|\cdot\|)$ under the map $y \leftrightarrow \phi_y$, where $\phi_y(x) = \langle x, y\rangle$.*

The main point here is not just that each element of $\mathcal{H}$ defines a continuous linear functional, and not just that this gives all of them, but that the norms and all the other vector space structure matches. We often therefore blur the distinction between the element y and the linear functional it determines, and write something like $\mathcal{H}^* = \mathcal{H}$ (when perhaps we should say that these spaces are isometrically isomorphic).

In general the dual space doesn't look like X at all! In fact having X^* isometrically isomorphic to X characterizes Hilbert space.

You proved the Riesz Representation Theorem for $H = \mathbb{R}^n$ in first year! This just saying that a linear map $\phi : \mathbb{R}^n \to \mathbb{R}$ is given by $\phi(\mathbf{x}) = A\mathbf{x}$ for a $1 \times n$ matrix, ie by taking the dot product with an element of $\mathbb{R}^n$. All such maps are easily seen to be continuous, and indeed, in first and second year, we barely mention the continuity of linear maps.

A reasonable question is whether continuity comes for free in general for a continuous map. The answer is no, although this is a little delicate. If we just look at linear functionals on (possible incomplete) normed spaces then it is easy to give one which is discontinuous. Let c_{00} be the space of sequences which are eventually zero, with the usual $\|\cdot\|_\infty$ norm. Define $\phi(x_1, x_2, \dots) = \sum_{k=1}^\infty x_k$. For $n = 1, 2, \dots,$ let $\mathbf{x}_n = (\frac{1}{n}, \dots, \frac{1}{n}, 0, 0, \dots)$ where there are n nonzero entries. Then $\mathbf{x}_n \to 0$ in c_{00} but $\phi(\mathbf{x}_n) \not\to \phi(0)$.

If you want to do this on Banach spaces, then you need to work much harder. Indeed, the examples of discontinuous linear functionals on infinite dimensional Banach spaces seem to all be noncontructive and require the Axiom of Choice. For those who like working in exotic axiom systems, there are alternatives to ZFC in which all linear functionals on Banach spaces are continuous!

The Closed Graph Theorem, which you will meet in Functional Analysis, says that an **everywhere-defined** 'closed' linear map on a Banach space must be continuous. Pretty much any formula that you write down will generate a closed map, so what this means in practice, is that your proposed linear functional is either bounded or else it isn't defined somewhere. Consider for example the map $\phi(x_1, x_2, \dots) = \sum_{k=1}^\infty x_k$ considered now on ℓ^2. This is linear, but not bounded, and it is easy to pick vectors in ℓ^2 where the formula blows up.

7.3 Weak convergence

Definition 7.3.1. Suppose that $\{x_\alpha\}$ is a net in a Banach space X. We say that x_α **converges weakly** to $x \in X$ if for all $\phi \in X^*$,

$$\phi(x_\alpha) \to \phi(x).$$

It follows from Exercise 7.2.2 that if x_α converges to x in norm then it certainly converges weakly. We have seen that in an infinite dimensional Hilbert space that the converse may be false.

In a Hilbert space the weak topology is always Hausdorff. This is equivalent to asking that given any $x_1 \neq x_2 \in \mathcal{H}$ you can always find $y \in \mathcal{H}$ such that

$$\langle x_1, y\rangle \neq \langle x_2, y\rangle.$$

(Take $y = x_1 - x_2$!)

To ensure that this version of weak convergence is useful we want to know two things:

1. Are there always enough linear functionals to distinguish elements of X, and therefore ensure that the corresponding topology is Hausdorff?

2. How can you effectively test weak convergence unless you know what the elements of the dual space look like?

The first question gets resolved by one of the cornerstone results of functional analysis, the Hahn-Banach Theorem. This theorem has many versions and consequences.

Theorem 7.3.2 (Hahn–Banach Theorem — Version 1)**.** *Suppose that X is a Banach space and that $x \in X$. Then there exists a nonzero $\phi \in X^*$ such that $\phi(x) = \|x\| \, \|\phi\|_*$.*

Corollary 7.3.3. *Suppose that $x_1 \neq x_2 \in X$. Then there exists $\phi \in X^*$ such that $\phi(x_1) \neq \phi(x_2)$.*

Proof. Let $x = x_1 - x_2$. By the Hahn-Banach Theorem there exists a nonzero $\phi \in X^*$ such that $\phi(x) = \phi(x_1) - \phi(x_2) = \|x\| \, \|\phi\|_* \neq 0$ and so $\phi(x_1) \neq \phi(x_2)$. ■

Corollary 7.3.4. *Suppose that $\{x_\alpha\}$ converges to both x and x'. Then $x = x'$.*

The theorem above also gives a certain symmetry between X and X^*.

Corollary 7.3.5. *For all $x \in X$, $\|x\| = \sup_{\|\phi\|_*=1} |\phi(x)|$.*

It follows from the Hahn-Banach theorem that weak convergence comes from a Hausdorff topology. One can define a base for the topology by just mimicing the construction from Hilbert space. In practice however, one rarely cares just what the open sets look like, just that a topology exists. You are usually just interested in checking convergence and continuity and you can do these directly from Definition 7.3.1.

The second question above is a little more complicated. In general it is hard to say what the elements of X^* look like. However, for pretty much all of the 'standard' Banach spaces we have seen, it is possible to say explcitly what form a continuous linear functional on X must take.

In Example 7.2.5 we showed that if $X = \ell^1$, then every element y of ℓ^∞ determines a continuous linear functional ϕ_y (with matching norm) by taking a sort of dot product. It is actually not too hard to show that there are not any others. In a case like this we'll say that $(\ell^1)^*$ **is** ℓ^∞, again blurring the distinction between $y \in \ell^\infty$ and $\phi_y \in (\ell^1)^*$. (To be strictly correct, one should say that the map $J : \ell^\infty \to (\ell^1)^*$, $J(y) = \phi_y$ is an isometric isomorphism of Banach spaces. Note in particular that it is not only 1–1, onto and norm-preserving, but it also respects that vector space structure, since $\phi_{y+z} = \phi_y + \phi z$ and so forth.) To even further match what we do in Hilbert space we'll write $\langle x, y \rangle$ for $\phi_y(x)$. As we'll see shortly, this symmetry in the notation is quite convenient.

Something similar happens for almost all the other classical Banach spaces; we can say just what the continuous linear functionals look like and this determines an isometric isomorphism between the dual space and some space which we already know. This is a minor miracle! The classical reference Dunford and Schwartz has a 6 page table listing spaces and their dual spaces! As above, the general practice is to stay that the dual space **is** Y rather than that it is isometrically isomorphic to Y.

In the table below, which gives the duals of the main example spaces we've met, $1 < p < \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$.

X	X^*	Form of functional
c_0	ℓ^1	$\langle x, y\rangle = \sum_k^{\infty} x_k y_k$
ℓ^1	ℓ^∞	$\langle x, y\rangle = \sum_k^{\infty} x_k y_k$
ℓ^p	ℓ^q	$\langle x, y\rangle = \sum_k^{\infty} x_k y_k$
$L^1[0,1]$	$L^\infty[0,1]$	$\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$
$L^p[0,1]$	$L^q[0,1]$	$\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$

Table of dual spaces

The dual of $C[0,1]$ turns out to be a space of measures on $[0,1]$ which is a little more complicated, but is well understood. The dual of ℓ^∞ and $L^\infty[0,1]$ are worse. There are certainly spaces for which there is not nice description of X^*.

The effect of these tables is that you can actually do calculations to test whether a net converges weakly or not.

Example 7.3.6. Let e_k be the sequence with a 1 in the kth position and zeros elsewehere. What happens to the sequence $\{e_k\}$ in ℓ^p. It certainly doesn't converge in norm in any of these spaces. If $y \in \ell^q$ with $q < \infty$ then

$$\langle e_k, y\rangle = y_k \to 0 = \langle 0, y\rangle$$

since every sequence in ℓ^q has limit zero, Thus e_k converges weakly to the zero sequence. On the other hand, if $p = 1$ and $q = \infty$, then y_k need not converge at all and so this sequence doesn't converge weakly in ℓ^1. (Actually for sequences, any **sequence** that converges weakly in ℓ^1 automatically converges in the ℓ^1 norm!)

Exercise 7.3.7. For $k = 1, 2, \ldots$, let $f_k(t) = e^{2\pi ikt}$ in $L^p[0,1]$. It is easy to check that $\|f_k\|_p = 1$ for any p. To check whether f_k converges weakly to $f = 0$ in L^p, you need to fix $g \in L^q[0,1]$ and check whether

$$\langle f_k, g\rangle = \int_0^1 f_k(t)g(t)\,dt = \int_0^1 g(t)e^{2\pi ikt}\,dt \to 0.$$

This is of course asking about the decay of the Fourier coefficients on an L^q function.

7.4 Duality

A feature of the table in the last section is the symmetry in the formulae for the linear functionals. If we write down $\int_0^1 f(x)g(x)\,dx$ we might consider this as the formula for

$\phi_f(g)$ or for $\phi_g(f)$. If $1 < p < \infty$, then all the elements of ℓ^q determine linear functionals on ℓ^p — and all the elements of ℓ^p determine linear functionals on ℓ^q.

More generally[1], given any x in a Banach space X, you can define a linear functional $\hat{x}$ on the dual space X^* by setting

$$\hat{x}(\phi) = \phi(x).$$

Clearly $|\hat{x}(\phi)| = |\phi(x)| \leq \|x\| \, \|\phi\|_*$ so this functional is certainly continuous. Thus every element of X determines a continuous linear functional on X^*. The Hahn-Banach Theorem tells us that in fact the norm of $\hat{x}$ as an element of $X^{**} := (X^*)^*$ is precisely $\|x\|$.

In other words, there is always an isometric copy of X sitting inside X^{**}. As you can see from the table, sometimes this copy is in fact all of X^{**} and sometimes is it just a subspace of X^{**}. For example, if $X = c_0$ then X^{**} is ℓ^∞ which is actually quite a bit bigger than c_0.

Definition 7.4.1. If the copy of X in X^{**} is all of X^{**} then we say that X is **reflexive**.

Thus ℓ^p and $L^p[0,1]$ are reflexive when $1 < p < \infty$ and are not reflexive when $p = 1$ or $p = \infty$. The space $C[0,1]$ is not reflexive. It is clear that if X is reflexive, then so is X^*.

You can keep taking duals of duals of duals, to form X^{***}, X^{****}, ... If X is reflexive then these obviously just alternate between two spaces. If X is not reflexive then you get two never-stabilizing towers of spaces $X \underset{\neq}{\subset} X^{**} \underset{\neq}{\subset} X^{****} \underset{\neq}{\subset} \dots$ and $X^* \underset{\neq}{\subset} X^{***} \underset{\neq}{\subset} \dots$.

Why do you care?

It turns out that the weak topology is different if X is reflexive compared to when it is not.

Theorem 7.4.2. *Let X be a reflexive Banach space. Then the closed unit ball in X is compact in the weak topology on X.*

What this means is that in a reflexive space, any bounded sequence must have at least a weak limit point.

If you have a nonreflexive space then you need to use a different topology.

Definition 7.4.3. Let X be a Banach space with dual X^*. A net $\{\phi_\alpha\}_{\alpha \in A}$ converges to $\phi \in X^*$ in the **weak-$*$ topology** if for all $x \in X$,

$$\langle x, \phi_\alpha \rangle \to \langle x, \phi \rangle.$$

This topology is the same as the weak topology in a reflexive space, but otherwise it is weaker.

[1]Even more generally, if X is any vector space of functions on a set Ω, then given any $\omega \in \Omega$, the map $\phi_\omega(f) = f(\omega)$ is linear. In this way, not only do the functions in X act on the points in Ω, but the points in Ω act on the functions in X. Said another way; every point is a function and every function is a point!

Theorem 7.4.4 (Alaoglu's Theorem)**.** *The closed unit ball in X^* is compact in the weak-$*$ topology.*

The real problem children are the spaces which are not the dual of another space. It is not so easy to see, but c_0 and $L^1[0,1]$ are not the duals of any Banach spaces. This acts as an obstruction to many general Banach space proofs, and indeed there are many theorems of the form that 'nice property holds in X' if and only if X doesn't contain a subspace isomorphic to c_0.

If we are dealing with complex Hilbert spaces, the inner product is conjugate linear in the second variable. For example in ℓ^2

$$\langle x, y\rangle = \sum_{k=1}^{\infty} x_k \bar{y}_k.$$

This is to make sure that $\|x\|^2 = \langle x, x\rangle$. In Banach spaces we usually don't apply the complex conjugate in these sorts of formulas since we'd like the map $\phi_x(y) = \langle x, y\rangle$ to be linear rather than conjugate linear. This can lead to a small degree of ambiguity when someone writes $\langle x, y\rangle$; is the author a Hilbert space person or a Banach space person?

7.5 Aside: some probability theory

In first year you looked separately at discrete and continuous random variables. But some random variables, such as the amount of rainfall on Sydney on a given day, don't fall into either category. Measure theory provides a language to bring everything together.

Let X be a random variable taking values in $\mathbb{R}$. For a set $A \subseteq \mathbb{R}$, let

$$\mu(A) = P(X \in A).$$

As noted in Chapter 6, we want this to be a measure on $\mathbb{R}$. For a discrete random variable with probability distribution $\{p_k\}_{k=0}^{\infty}$, this means that

$$\mu(A) = \sum_{k \in A} p_k.$$

For a continuous random variable, we have a probability density function f, for which

$$P(a \le X \le b) = \int_a^b f(x)\,dx.$$

For measurable sets A one can then write

$$\mu(A) = \int_A f(x)\,dx = \int_{\mathbb{R}} \chi_A(x) f(x)\,dx.$$

As we saw, you run into problems if you try to extend this to every subset of $\mathbb{R}$.

In general probability theory one has a set of possible outcomes Ω, a σ-algebra $\mathcal{A}$ of measurable subsets of Ω, and a measure $\mu : \mathcal{A} \to [0, 1]$ so that

$$P(X \in A) = \int_A 1\, d\mu = \int_\Omega \chi_A \, d\mu, \qquad A \in \mathcal{A}.$$

An important issue then is whether one can apply all our analysis theory to study things like convergence of measures. These are the sorts of questions that arise in things like the Central Limit Theorem.

The simplest case is perhaps where $\Omega = \mathbb{N}$, and $\mathcal{A}$ consists of all the subsets of $\mathbb{N}$. (That is, discrete random variables!) Each measure corresponds to a probability distribution. Since $\sum_{k=0}^\infty p_k = 1$, each of these can be considered as a point $p = (p_0, p_1, p_2, \dots)$ in the unit sphere of the sequence space $\ell^1(\mathbb{N})$. Going back the other way, one might consider any sequence $\{q_k\} \in \ell^1(\mathbb{N})$ as a 'signed measure' on the set $\mathbb{N}$. Importantly, the extra structure of $\ell^1(\mathbb{N})$ provides a host of geometric and topological structure to the set $\mathcal{P}$ of probability measures on $\mathbb{N}$. For example $\mathcal{P}$ is a convex set in $\ell^1(\mathbb{N})$. We can look at norm, weak and weak-$*$ convergence. In the right topologies $\mathcal{P}$ is also compact, and this turns out to be extremely important.

The same ideas work rather more generally, but with many more technicalities. That is, one can, in reasonable generality[2], embed the space of probability measures into the unit sphere of a space $M(\Omega)$ of signed measures. The important aspect of this is that this space is the dual space of a space $\mathcal{C}$ of continuous functions on Ω. For example, if Ω is a compact topological space, then the dual of $C(\Omega)$ is the space of Radon signed measures acting on the Borel subsets of Ω. This provides an important weak-$*$ convergence in the space of measures:

$$\mu_n \to \mu \iff \int_\Omega f\, d\mu_n \to \int_\Omega f\, d\mu, \qquad \text{for all } f \in C(\Omega).$$

Confusingly, probabilists call this weak convergence of measures! In any case, the ideas of duality play an important role in the defining useful topologies on spaces of probability measures. This is pursued in much more detail in the course Measure, Integration and Probability.

7.6 Problems

1. Prove Theorem 7.2.1.

2. Which of the following are continuous linear functionals on the given Banach spaces? For those which are, find $\|\phi\|_*$.

[2] If you look this up you run across terms such as regular countably additive Borel measures on a locally compact Hausdorff space, but for the present discussion, the details are not very important.

(a) $X = c$, $\phi(\mathbf{x}) = \lim_{n\to\infty} x_n$.

(b) $X = C[0,1]$, $\phi(f) = \sum_{k=1}^{\infty} \frac{f(\frac{1}{k})}{2^k}$.

(c) $X = C[0,1]$, $\phi(f) = \int_0^1 f(t)g(t)\,dt$ where g is some fixed element of $C[0,1]$.

(d) $X = L^\infty[0,1]$, $\phi(f) = f(0)$.

(e) $X = \ell^3$, $\phi(\mathbf{x}) = \sum_{k=1}^{\infty} \frac{x_k}{k}$.

3. The map $\phi(f) = f(0)$ is a continuous linear functional on $C[0,1]$. Is there any function g such that $\phi(f) = \int_0^1 f(t)g(t)\,dt$ for all $f \in C[0,1]$?

4. For the following sequences determine whether they converge in norm, or weakly, or (if this makes sense) weak-$*$.

(a) $\mathbf{x}_k = (1, 1, \dots, 1, 0, 0, \dots) \in c_0$ where there are k 1's.

(b) $\mathbf{x}_k = \left(\frac{1}{k}, \dots, \frac{1}{k}, 0, \dots\right) \in \ell^1$ where are k nonzero entries.

(c) $f_k = k\chi_{[0,1/k]} \in L^1[0,1]$.

(d) $f_k(x) = \sin(2\pi kx) \in L^2[0,1]$.

Chapter 8

Linear operators

8.1 Introduction

Many of the standard operations of analysis take in a function and spit out a different one: differentiation, integration, composition, ... Many of these operations are in fact linear and so one can try to study them via the linear algebra of linear transformations on vectors spaces.

Unfortunately function spaces are almost always infinite dimensional and the well-understood picture of finite-dimensional linear algebra you studied in second year suddenly replaced by a real zoo!

In finite dimensions you more or less ignore any topological questions when doing linear algebra. All norms on $\mathbb{R}^n$ are equivalent, and all linear transformations $T : \mathbb{R}^n \to \mathbb{R}^n$ are continuous with respect to any of these norms. And all n-dimensional real spaces are isomorphic to $\mathbb{R}^n$. Nothing like this holds in infinite dimensions!

Operator Theory combines the theory that comes from algebra with the theory that comes from analysis. In this chapter we'll give a brief overview of some of the main concepts in this theory. There won't be any proofs, and some of the concepts may not be fully explained, but it hopefully will provide some flavour of what you might expect from some of the later courses in Functional Analysis and Operator Algebras.

In this chapter an operator will generally mean a continuous linear transformation between normed vector spaces. If someone says that they are an operator theorist, this is probably what they study, but one should be aware that there are theories which cover non-continuous linear operators (which arise frequently in differential equations) and non-linear operators.

8.2 First thoughts

The best behaved infinite dimensional space is the Hilbert space ℓ^2, so that is a good place to start. And indeed operator theory on ℓ^2 has a much richer and well-developed theory than the theory of operators on say Banach spaces.

It is not too hard to convince yourself that any linear transformation from ℓ^2 to ℓ^2 must have a representation in the form of an infinite matrix:

$$Ax = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots \\ a_{21} & a_{22} & a_{23} & \dots \\ a_{31} & a_{32} & a_{33} & \\ \vdots & \vdots & & \ddots \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{\infty} a_{1j} x_j \\ \sum_{j=1}^{\infty} a_{2j} x_j \\ \vdots \end{pmatrix}$$

As in finite dimensions, the jth column is just the image of the jth standard basis vector. The problem is that not all infinite matrices determine a linear transformation from ℓ^2 to ℓ^2. For a start, you need that each row is in ℓ^2 or else the sums on the right-hand side might not converge. But even if the output sequence is well-defined, it need not be in ℓ^2. (For example, if each row in the matrix were the same, then you would get a constant sequence as the output, and that won't be in ℓ^2.)

There are sufficient conditions for an infinite matrix to define an ℓ^2 operator, and also necessary conditions, but annoyingly there is no reasonable necessary and sufficient conditions which you can use. Even if you have a matrix representation, most of the tools that you used with finite matrices, such as Gaussian elimination, obviously no longer work. Nonetheless, it is *sometimes* useful to write down the matrix for an operator to get some idea of what it is doing.

Another drawback is that using a matrix requires that the space has a (Schauder) basis. The concept of a basis for an infinite dimensional normed space is much more complicated than that for a finite dimensional space. The space ℓ^2 has an 'obvious' basis, but it is less clear what one should use in $C[0,1]$ say. The fact that finite linear combinations of the functions $\{1, x, x^2, x^3, \dots\}$ are dense in $C[0,1]$, but these functions do not form a basis should give you some idea that something more involved is going on. The space $C[0,1]$ does have a basis, but, for example, ℓ^∞ does not.

8.3 Continuity

Just as much of a problem is that not all linear transformations are continuous. This of course depends on the topologies of the input and output spaces, but we certainly saw examples where $f_n \to f$ (in some sense) but $f_n' \not\to f'$.

Fortunately, there is a relatively easy test for continuity of linear transformations between normed spaces. Continuity here means that if $\|x_n - x\| \to 0$ then $\|Tx_n - Tx\| \to 0$.

Definition 8.3.1. Suppose that $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are normed spaces and that $T : X \to Y$ is a linear transformation. Then T is a **bounded operator** if

$$N_T = \sup_{\|x\|_X = 1} \|Tx\|_Y = \sup_{x \neq 0} \frac{\|Tx\|_Y}{\|x\|_X}$$

is finite.

The most important result of this section is the following.

Theorem 8.3.2. *An operator $T : X \to Y$ is continuous everywhere $\iff$ it is continuous at 0 $\iff$ it is bounded.*

Because of this, analysts will use the words bounded and continuous for linear operators interchangeably.

Theorem 8.3.3. *The set of all bounded operators from X to Y, denoted $B(X, Y)$ is a vector space.*

Theorem 8.3.4. *The quantity N_T is a norm on $B(X, Y)$. If Y is complete, then $B(X, Y)$ is a Banach space under this norm.*

We usually write $\|T\|_{op}$ or just $\|T\|$ rather than N_T. Note that $\|Tx\| \leq \|T\|_{op} \|x\|$ for all $x \in X$, and $\|T\|_{op}$ is the smallest number that makes this inequality true.

The most important case is when $X = Y$ in which case we write $B(X)$ rather than $B(X, X)$. In this case we can also compose the linear transformations. Remember that in finite dimensions, linear transformation composition corresponds exactly to matrix multiplication of the associated matrices.

Fact 8.3.5. *$B(X)$ is a Banach algebra with composition as multiplication. In particular*

$$\|ST\|_{op} \leq \|S\|_{op} \|T\|_{op} .$$

The inequality is easy as for any $x \in X$,

$$\|STx\| = \|S(Tx)\| \leq \|S\|_{op} \|Tx\| \leq \|S\|_{op} \|T\|_{op} \|x\| .$$

Actually calculating $\|T\|_{op}$ is in general a real challenge, even on a finite dimensional Banach space, where the linear transformations have matrix representations.

Exercise 8.3.6. Find the operator norm of

$$A = \begin{pmatrix} 3 & -4 \\ 2 & 1 \end{pmatrix}$$

acting on $(\mathbb{R}^2, \|\cdot\|_2)$.

As noted above, for infinite matrices acting on ℓ^2, there is no useful necessary and sufficient condition for identifying whether an infinite matrix corresponds to a continuous linear transformation. On the other hand, for many linear operators it is not too hard to determine whether they are bounded or not.

Some authors prefer to use $L(X, Y)$ rather that $B(X, Y)$. Others use $L(X, Y)$ for all the linear maps from X to Y (not just the continuous ones). The more categorically minded may write $\mathrm{End}(X)$.

The operator norm $\|\cdot\|_{op}$ imposes a topology on $B(X)$ and so we can talk about the convergence of a sequence of operators. We won't pursue this any further, but as with any Banach space, there are a number of different topologies which prove useful when looking at $B(X)$. As it is a space of functions (on X), one can look at pointwise convergence too: $T_n \to T$ pointwise if for all $x \in X$, $T_n x \to Tx$ (in X norm). This is usually known as the **strong operator topology** on $B(X)$. There is also a **weak operator topology** for which $T_n \to T$ if for all $x \in X$, $T_n x \to Tx$ in the weak topology on X. The list does not stop here!

8.4 Examples

Example 8.4.1. Define the right shift operator $S : \ell^p \to \ell^p$ by

$$S(x_1, x_2, \dots) = (0, x_1, x_2, \dots).$$

It is easy to see that $\|Sx\| = \|x\|$ and so $\|S\|_{op} = 1$. Indeed S is an isometry, or norm preserving operator. Note that S is one-to-one, but not onto, something that can't happen in finite dimensions.

Example 8.4.2. Let $X = L^p[0,1]$ and suppose that $h \in L^\infty[0,1]$. Define the multiplication operator M_h by $M_h f = hf$. Then

$$\begin{aligned}
\|M_h f\|_p &= \left(\int_0^1 |h(x)\, f(x)|^p \, dx \right)^{1/p} \\
&\le \left(\int_0^1 \|h\|_\infty^p \, |f(x)|^p \, dx \right)^{1/p} \\
&= \|h\|_\infty \left(\int_0^1 |f(x)|^p \, dx \right)^{1/p} \\
&= \|h\|_\infty \, \|f\|_p \, .
\end{aligned}$$

Thus M_h is bounded and $\|M_h\|_{op} \le \|h\|_\infty$. (Prove that in fact we have equality.)

Example 8.4.3. Let $X = C[0,1]$ and suppose that $k(x,y)$ is a 'suitable' function of two variables. The integral operator T_k is defined by

$$T_k f(x) = \int_0^1 k(x,y) f(y) \, dy, \qquad x \in [0,1].$$

Now you have to worry about whether $\|T_k f\|_\infty < \|f\|_\infty$ and whether $T_k f$ is actually continuous.

Example 8.4.4. Let $X = C(\mathbb{D})$ under the supremum norm. Suppose that $\phi : \mathbb{D} \to \mathbb{D}$ is continuous. The composition operator $C_\phi : X \to X$ is defined by

$$C_\phi f(z) = f(\phi(z)), \qquad z \in \mathbb{D}.$$

Here it is obvious that C_ϕ is a bounded operator and that $\|C_\phi\|_{op} = 1$ — why?

Example 8.4.5. Let $X = L^2[0, 2\pi]$ and $Y = \ell^2(\mathbb{Z})$, now with complex scalars. For $f \in L^2[0, 2\pi]$ and $n \in \mathbb{Z}$ let

$$c_n = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-inx} \, dx.$$

and let $F(f) = (\dots, c_{-1}, c_0, c_1, c_2, \dots)$. Then Parseval's identity says that $\|F(f)\|_Y = \sqrt{2\pi}\, \|f\|_X$, so $F \in B(X, Y)$. Indeed, if you are willing to scale the operator (or the norm in X), you can make F into an isometry of between these two Hilbert spaces.

You should now go back and see what happens in the above examples if you change the space X on which the operator is acting.

8.5 The inverse problem

In linear algebra the main problem is to solve $Ax = b$ where you know A and b and you want to find x. You typically do this via Gaussian Elimination. In infinite dimensions, even if you had a matrix representation, this would not be easy to do!

Example 8.5.1. Define $T : \ell^2 \to \ell^2$ by

$$T(x_1, x_2, x_3, \dots) = \big(x_1, \frac{x_2}{2}, \frac{x_2}{3}, \dots\big).$$

This is a bounded operator with $\|T\|_{op} = 1$. (This is easy!)

Can we solve $Tx = b$ for every $b \in \ell^2$. You an write down a formula for the inverse: if $Tx = b$ then

$$(x_1, x_2, x_3, \dots) = T^{-1}b = (b_1, 2b_2, 3b_3, \dots).$$

If $b = (1, \frac{1}{2}, \frac{1}{3}, \dots)$ then the 'solution' x isn't in ℓ^2! The inverse operator T^{-1} is linear, but it isn't bounded[1].

Definition 8.5.2. Suppose that $T \in B(X)$. We say that T is invertible if there exists an operator $S \in B(X)$ such that $ST = TS = I$, where I is the identity linear operator on X.

You may have noticed that we will sometimes talk about an 'inverse operator' even when T is not invertible. This doesn't seem to cause problems in practice, but be careful.

Proposition 8.5.3. *If such an operator S exists it is unique.*

[1] We are being a little loose here in writing T^{-1} as we haven't given a domain and codomain for the operator. If you want the domain to be all of ℓ^2, then the codomain can't be ℓ^2. Alternatively, you might restrict the domain of T^{-1} to be the range of T which is strictly smaller than ℓ^2.

Proof. Suppose $S_1T = TS_1 = I = S_2T = TS_2$. Then

$$S_1 = S_1I = S_1(TS_2) = (S_1T)S_2 = IS_2 = S_2.$$

$\blacksquare$

In finite dimensions for a linear transformation $T : V \to V$ the following conditions are equivalent:

1. T is invertible.
2. T is one-to-one.
3. T is onto.

In infinite dimensions these things are distinct as the example above shows. In particular, unlike in finite dimensions, you really need to check that ST and TS are both the identity.

Exercise 8.5.4. Let S be the right shift operator from Example 8.4.1. Show that S has a left inverse but not a right inverse. That is, there exists $U \in B(\ell^p)$ such that $US = I$ on ℓ^p, but no $V \in B(\ell^p)$ such that $SV = I$.

Exercise 8.5.5. Prove that if T has both a left inverse and a right inverse then these are equal.

The big question is whether there is any replacement for the Gaussian Elimination algorithm which will allow you to find the inverse operator. The answer unfortunately is 'No'. Sometimes, such as for the multiplication operators in Example 8.4.2, it is quite easy to write down the inverse and give conditions under which it is bounded. But for most operators, it is a really challenging (and perhaps impossible!) problem.

8.6 Structure Theorems

The main structure theorem for matrices is the Jordan Canonical Form. There is no such general result for operators, even those on Hilbert space!

By way of notation, suppose that $\mathcal{H}_1$ and $\mathcal{H}_2$ are (closed!) subspaces of a Hilbert space $\mathcal{H}$. We'll write $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$ if every vector $x \in \mathcal{H}$ can be written uniquely as $x = x_1 + x_2$ with $x_1 \in \mathcal{H}_1$ and $x_2 \in \mathcal{H}_2$. In this case there are projections P_1 and P_2 with ranges $\mathcal{H}_1$ and $\mathcal{H}_2$, such that $P_1 + P_2 = I$, $P_1P_2 = P_2P_1 = 0$. We say that $\mathcal{H}$ is the direct sum of $\mathcal{H}_1$ and $\mathcal{H}_2$. If $T \in B(\mathcal{H})$ then we can write it in operator matrix form

$$T = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \quad \text{on } \mathcal{H}_1 \oplus \mathcal{H}_2$$

where $T_{ij} \in B(\mathcal{H}_j, \mathcal{H}_i)$. Of course, you can do this with more than 2 summands, getting large operator matrices.

If T is linear transformation on an n dimensional complex vector space $\mathcal{H}$, then there exists a basis $\{x_1, x_2, \ldots, x_n\}$ of eigenvectors and generalized eigenvectors such that the matrix A for T is upper triangular. Given $1 \leq k < n$, if $\mathcal{H}_1 = \operatorname{span}\{x_1, \ldots, x_k\}$ and $\mathcal{H}_2 = \operatorname{span}\{x_{k+1}, \ldots, x_n\}$ then $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$. The Jordan form theorem says then that

$$T = \begin{pmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{pmatrix}$$

with respect to this decomposition, i.e. $T_{21} = 0$. That is, we get an upper triangular matrix form. This is equivalent to the statement that $\mathcal{H}_1$ is an **invariant subspace** for T. That is, $Tx \in \mathcal{H}_1$ for all $x \in \mathcal{H}_1$ (or, more succinctly, $T(\mathcal{H}_1) \subseteq \mathcal{H}_1$). Algebraically, this is the same as asking that $P_1 T P_1 = T P_1$.

For self-adjoint operators we also get that $T_{12} = 0$, so T has a diagonal operator matrix form. Algebraically, this is equivalent to saying that $P_1 T = T P_1$.

A first step then in finding an analogue of the Jordan form theorem would be to show that every operator T on an infinite dimensional Hilbert space has an invariant subspace. Of course the 'trivial' subspaces $\{0\}$ and $\mathcal{H}$ are always invariant, but they don't really count! The Invariant Subspace Problem asks whether every bounded linear operator on ℓ^2 has a nontrivial (closed) invariant subspace. This problem has been open for about 100 years. As of 2022 it is still unsolved, despite a huge amount of work having gone into research in the area.

The very first structure theorem that you meet as an undergraduate is the theorem that self-adjoint or normal matrices can be diagonalized. It turns out that this theorem **can** be generalized. The result, known as the Spectral Theorem, is one of the cornerstones of operator theory. The first step is to rewrite the matrix diagonalization theorem in a slightly different way.

Recall that if v is a nonzero vector in an inner product space $\mathcal{H}$, then the orthogonal projection onto the span of v is given by

$$P_v x = \left(\frac{\langle x, v \rangle}{\|v\|^2} \right) v.$$

It is easy to check that $P_v^2 = P_v$, and that for all $x, y \in \mathcal{H}$

$$\langle P_v x, y \rangle = \langle x, P_v y \rangle.$$

The diagonalization theorem for self-adjoint (or normal) matrices says that, if you rewrite the matrix with respect to a basis of eigenvectors the matrix, then you get a diagonal matrix:

$$\begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & & \ddots & \\ 0 & \ldots & 0 & \lambda_n \end{pmatrix} = \sum_{j=1}^{n} \lambda_j P_j$$

where P_j is the matrix with a 1 in the (j, j)th position and zeros elsewhere. This is just the matrix corresponding to the orthogonal projection onto the jth eigenvector!

It would be great if, for some reasonable class of operators, you could do this in infinite dimensions by writing

$$T = \sum_{j=1}^{\infty} \lambda_j P_j$$

where you now have an infinite sequence of eigenvalues and eigenvectors $Tv_j = \lambda_j v_j$ and P_j is again the orthogonal projection onto the span of v_j.

It turns out that this is the case, but to make sense of it we'll need to give some more definitions. First we need to say what self-adjoint and normal mean for operators (on a Hilbert space). Then we'll need to see why we need a more general concept than that of an eigenvalue to do things in infinite dimensions.

8.7 Adjoints and self-adjointness

Let's specialize now to $\mathcal{H} = \ell^2$ (or any other separable infinite dimensional Hilbert space), and at this stage it is best to use complex scalars and the inner product

$$\langle x, y \rangle = \sum_{j=1}^{\infty} x_j \, \overline{y}_j.$$

One fact that you will have seen is that the adjoint A^* of an $n \times n$ matrix A, defined as the conjugate transpose, satisfies

$$\langle Ax, y \rangle = \langle x, A^* y \rangle \qquad \text{for all } x, y \in \mathbb{C}^n. \tag{8.7.1}$$

We say that A is self-adjoint if $A = A^*$ and that it is normal if $A^*A = AA^*$. The identity (8.7.1) allows one to extend this concept to any inner product space.

Fact 8.7.1. *For every $T \in B(\mathcal{H})$ there exists a unique operator $T^* \in B(\mathcal{H})$, called the **adjoint** of T, such that for all $x, y \in \mathcal{H}$,*

$$\langle Tx, y \rangle = \langle x, T^* y \rangle.$$

If T is described by an infinite matrix, then the matrix for T^* is again the conjugate transpose of that for T. The properties of the inner product make it easy to check that $(T^*)^* = T$ in general.

It is not always easy to find T^*, but it is easy to check whether $T = T^*$!

Definition 8.7.2. We say that $T \in B(\mathcal{H})$ is **self-adjoint** if $T = T^*$, and **normal** if $T^*T = TT^*$.

This all relies critically on the existence of an inner product. In a Banach space X, if $T \in B(X)$ then you can find a unique operator T^* which now acts on the dual space X^* such that

$$\langle Tx, y \rangle = \langle x, T^* y \rangle, \qquad x \in X, y \in X^*,$$

but you can't ask for $T = T^*$ as they act on different spaces! Note that for Hilbert spaces $(\lambda T)^* = \bar{\lambda} T^*$ whereas for Banach spaces $(\lambda T)^* = \lambda T^*$. So again you need to be careful to understand whether an author is considering say $L^2[0,1]$ as just one space in the scale of Banach spaces $L^p[0,1]$, or whether they are really using the Hilbert space structure as well.

Exercise 8.7.3. Prove that the orthogonal projection P_v onto a nonzero vector $v \in \mathcal{H}$ is always a continuous linear operator of norm 1 and that it is self-adjoint.

Some of the most important operators are **idempotent**, that is $T^2 = T$. Given any closed subspace $\mathcal{K}$ of a Hilbert space $\mathcal{H}$ there is a 'projection' of $\mathcal{H}$ onto $\mathcal{K}$. Indeed there are many. For example, in $\mathbb{R}^2$ then matrices

$$P_c = \begin{pmatrix} 1 & c \\ 0 & 0 \end{pmatrix}$$

are all idempotents with range $\mathcal{K} = \{(x, 0) : x \in \mathbb{R}\}$. The complementary projection $Q_c = I - P_c$ has range $\operatorname{span}\{(-c, 1)\}$, which is the same as the kernel of P_c. The orthogonal projection onto $\mathcal{K}$ is the one where the range is orthogonal to the kernel — in this case where $c = 0$. In general there is a nice theorem which helps to identify the orthogonal projection.

Fact 8.7.4. *Suppose that $P \in B(\mathcal{H})$. Any two of the following conditions implies the third.*

1. $P^2 = P$.
2. $P = P^*$.
3. $\|P\|_{op} = 1$.

If P satisfies these conditions, then P is the orthogonal projection onto the range of P.

Exercise 8.7.5. Let S be the right shift operator from Example 8.4.1. Find the adjoint of S and prove that S is not self-adjoint (or even normal). (What are the matrices for S and S^*?)

Example 8.7.6. On $L^2[0,1]$ the inner product is

$$\langle f, g \rangle = \int_0^1 f(x)\, \overline{g(x)}\, dx$$

It is easy to check that if the multiplication operator M_h from Example 8.4.2 is always normal, and it is self-adjoint if and only if h is real-valued.

Exercise 8.7.7. Suppose that $T \in B(\mathcal{H})$ is normal. Show that there are unique commuting self-adjoint operators A and B such that $T = A + iB$.

8.8 Eigenvalues and the spectrum

As in linear algebra, we say that $\lambda \in \mathbb{C}$ is an **eigenvalue** for $T \in B(\mathcal{H})$ if there exists a nonzero '**eigenvector**' $v \in \mathcal{H}$ such that $Tv = \lambda v$.

In finite dimensions you can find these by finding the roots of the characteristic polynomial, but there isn't any obvious way to generalize that to infinite dimensions.

Example 8.8.1. Let $T(x_1, x_2, \dots) = (x_1, \frac{x_2}{2}, \frac{x_3}{3}, \dots)$. Then all the numbers of the form $\frac{1}{j}$ are eigenvalues as $Te_j = \frac{1}{j}e_j$ for the jth 'basis' vector e_j.

Example 8.8.2. Let $S(x_1, x_2, \dots) = (0, x_1, x_2, \dots)$ be the right shift operator we looked at earlier. Suppose that $Sx = \lambda x$. Then comparing components

$$\begin{aligned} 0 &= \lambda x_1 \\ x_1 &= \lambda x_2 \\ x_2 &= \lambda x_3 \end{aligned}$$

and so forth. If $\lambda \neq 0$ then we get that $x_1 = x_2 = \cdots = 0$ and so $x = 0$, so λ is not an eigenvalue. If $\lambda = 0$ then the same conclusion holds! Therefore S is an operator with no eigenvalues and eigenvectors, something that can't happen in a finite dimensional space.

It turns out that eigenvalues aren't quite the right thing to look for in infinite dimensions. Note that if $Tv = \lambda v$ then $(T - \lambda I)v = 0$. As $(T - \lambda I)0 = 0$ this tells me that $T - \lambda I$ is not one-to-one and hence it can't possible be invertible. In finite dimensions λ is an eigenvalue if and only if $T - \lambda I$ is not invertible, but this is not true in infinite dimensions.

Example 8.8.3. For the right shift, take $\lambda = 0$. This is not an eigenvalue, but $S - 0I = S$ is not invertible.

It turns out that what we really need is the non-invertibility condition.

Definition 8.8.4. Suppose that X is a Banach space and that $T \in B(X)$. Then the **spectrum** of T is the set

$$\sigma(T) = \{\lambda \in \mathbb{C} : (T - \lambda I) \text{ is not invertible}\}.$$

The **resolvent** of T is the complement of the spectrum: $\rho(T) = \mathbb{C} \setminus \sigma(T)$.

So in finite dimensions, $\sigma(T)$ is just the set of eigenvalues of T.

Fact 8.8.5. *For any $T \in B(X)$, the spectrum $\sigma(T)$ is always a non-empty compact subset of $\mathbb{C}$.*

In ℓ^2 you can get any nonempty compact subset of $\mathbb{C}$ as the spectrum, but there are some Banach spaces where this is not the case. Note, in particular, that the spectrum of an operator on ℓ^2 can be uncountable[2].

Exercise 8.8.6. Show that for the shift operator $\sigma(S) = \{z : |z| \leq 1\}$.

Exercise 8.8.7. Find the spectrum of the multiplication operator M_h from Exercise 8.4.2 (acting on ℓ^2).

At this point it is probably not clear why one would want to calculate $\sigma(T)$. And in general it is pretty hard to find $\sigma(T)$. But it does turn out to give quite a lot of information about the behaviour of the operator.

The diagonalization theorem tells you that if A is a self-adjoint matrix you can sensibly form $f(A)$ for any function f which is defined on the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A. If A is not self-adjoint, then you might have seen that you may also need certain derivatives of f at the eigenvalues. For example, it is easy to check that, at least for polynomials,

$$f\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} f(2) & f'(2) \\ 0 & f(2) \end{pmatrix}.$$

What happens for operators on infinite dimensional spaces is that $f(T)$ should typically only depend on the values (and derivatives) of f on $\sigma(T)$. It is an important theorem that you can always make some good sense of $f(T)$ if f is analytic on an open neighbourhood of the spectrum.

8.9 $B(\mathcal{H})$ as an algebra

We won't pursue this idea too deeply, but an important ingredient in operator theory is the fact that $B(\mathcal{H})$ is a C^*-algebra. You can think of this as something like a higher dimensional version of $\mathbb{C}$. In $B(\mathcal{H})$ you not only have the vector space operations, but also multiplication (via composition) and adjoints, which are an analogue of complex conjugation.

A good deal of complex analysis actually works for $B(\mathcal{H})$ valued functions, with more or less the same proofs. In particular you can have power series in $B(\mathcal{H})$. For example, suppose that $\lambda \neq 0$. Then

$$\frac{1}{z - \lambda} = -\sum_{k=0}^{\infty} \frac{z^k}{\lambda^{k+1}}$$

which converges if $|z| < |\lambda|$. This suggests that for $T \in B(\mathcal{H})$ we might consider the series

$$U = -\sum_{k=0}^{\infty} \frac{T^k}{\lambda^{k+1}}.$$

[2]This should make you wonder how you could write $T = \sum \lambda_i P_i$ if you had to do an uncountable sum?

Then

$$(T - \lambda I)U = -T \sum_{k=0}^{\infty} \frac{T^k}{\lambda^{k+1}} + \lambda \sum_{k=0}^{\infty} \frac{T^k}{\lambda^{k+1}} = \sum_{k=0}^{\infty} \frac{T^k}{\lambda^k} - \sum_{k=0}^{\infty} \frac{T^{k+1}}{\lambda^{k+1}} = I$$

(and similarly $U(T - \lambda I) = I$). So, as long as the series converges we can find an inverse to $T - \lambda I$. It is easy to check that a sufficient condition for this is that $\lambda > \|T\|_{op}$, from which we can deduce that $T - \lambda I$ is invertible if $|\lambda| > \|T\|_{op}$. That is,

$$\sigma(T) \subseteq \{z \,:\, |z| \le \|T\|\}.$$

In fact, an even better bound is known: the smallest disk centered at 0 which contains $\sigma(T)$ has radius

$$r = \lim_{k \to \infty} \left\|T^k\right\|_{op}^{1/k}.$$

A consequence of this[3] is that if $T^k = 0$ for some k, then $\sigma(T) = \{0\}$.

Fact 8.9.1. *If $T \in B(\mathcal{H})$ is self-adjoint then $\sigma(T) \subseteq \mathbb{R}$.*

You can use power series to define functions of operators

$$e^T = I + T + \frac{T^2}{2!} + \frac{T^3}{3!} + \dots$$
$$\sin(T) = T - \frac{T^3}{3!} + \frac{T^5}{5!} - \dots$$

and so on. Perhaps more surprisingly you can use the Cauchy integral formula to define functions of operators

$$f(T) = \frac{1}{2\pi i} \int_{\gamma} f(z)(zI - T)^{-1}\, dz$$

for suitable analytic functions and suitable curves γ around the spectrum. (And these methods all give the same answer!)

$B(\mathcal{H})$ is an example of a C^*-algebra. That is, it is both a Banach space and a ring, and it has an involution $T \mapsto T^*$. And of course all these structures need to fit together properly.

It is a consequence of the Weierstrass Approximation Theorem that given a nonempty compact set $\sigma \subseteq \mathbb{C}$, the complex polynomials $p(z, \overline{z})$ form a dense subalgebra of $C(\sigma)$. If T and T^* commute (that is, T is normal), then $p(T, T^*)$ is well-defined. Remarkably, in this case

$$\|p(T, T^*)\|_{op} = \sup_{z \in \sigma(T)} |p(z, \overline{z})| = \|p\|_{C(\sigma(T))}.$$

A consequence of this is that there is a natural way of defining $f(T)$ for any $f \in C(\sigma)$, and that the map $f \mapsto f(T)$ is an isometric $*$-isomorphism the (commutative) C^*-algebra $C(\sigma(T))$ to the C^*-algebra generated by T, $C^*(T)$. (Again, this emphasizes why you would want to know what $\sigma(T)$ is!)

[3] You should know this fact for matrices!

8.10 Structure Theorems 2

The big version of the Spectral Theorem for self-adjoint operators needs to deal with the fact that $\sigma(T)$ might be uncountable. One way of doing this is to replace the sum $\sum_{j=1}^n \lambda_j P_j$ with some sort of integral $\int_{\sigma(T)} \lambda \, dP(\lambda)$. This is quite technical, but very powerful!

Fortunately, there is a class of self-adjoint operators for which something simpler can be done.

Definition 8.10.1. An operator $T \in B(\mathcal{H})$ is **finite rank** if the range of T is finite dimensional.

Suppose that T is of finite rank, and that P is the orthogonal projection onto the range of T. Then $P\mathcal{H}$ is an invariant subspace for T, so that as an operator matrix

$$T = \begin{pmatrix} T_{11} & T_{12} \\ 0 & 0 \end{pmatrix} \qquad \text{on } P\mathcal{H} \oplus (I - P)\mathcal{H}.$$

Here T_{11} is just a linear transformation on the finite dimensional space $P\mathcal{H}$, so all the finite dimensional theory can be applied. In particular, T_{11} must have eigenvalues, which will also be eigenvalues of T. Indeed, one has that $\sigma(T) = \sigma(T_{11}) \cup \{0\}$, and, if T is self-adjoint, then can be written as a finite linear combination of the projections onto the eigenspaces, $T = \sum_{j=1}^n \lambda_j P_j$.

It turns out that things still work quite well when T is 'nearly finite rank'. By this we mean that there is a sequence of finite rank operators $\{T_n\}$ such that

$$\lim_{n\to\infty} \|T - T_n\|_{op} = 0.$$

On a Hilbert space, this turns out to be equivalent to a condition called compactness.

Definition 8.10.2. An operator $T \in B(\mathcal{H})$ is said to be **compact** if the closure of the image of the unit ball $B \subseteq \mathcal{H}$ has compact closure in $\mathcal{H}$. That is, if $\mathrm{cl}(T(B))$ is norm compact.

Many operators that arise in solving differential and integral equations turn out to be compact. On $\mathcal{H}$, the easiest way to check for compactness is the following.

Fact 8.10.3. *An operator $T \in B(\mathcal{H})$ is compact if and only if it is nearly finite rank. (But note that on some Banach spaces, there are compact operators which are not nearly finite rank!)*

Example 8.10.4. The diagonal operator $T(x_1, x_2, \dots) = (x_1, x_2/2, x_3/3, \dots)$ is compact. You can just take $T_n(x_1, x_2, \dots) = (x_1, x_2/2, \dots, x_n/n, 0, 0, \dots)$ as the approximating sequence. On the other hand, the identity is not compact on any infinite dimensional space.

Fact 8.10.5. *For the algebraically minded: the set of compact operators in $B(\mathcal{H})$ forms a closed two-sided ideal in $B(\mathcal{H})$.*

Compact operators have a spectral theory which is pretty close to that for matrices.

Fact 8.10.6 (Fredholm Alternative)**.** *Suppose that $T \in B(\mathcal{H})$ is compact. Then $\sigma(T)$ contains 0 and a (possibly empty or finite) countable set of eigenvalues $\{\lambda_j\}$ whose only possible limit point is 0. The eigenvalues each have finite dimensional eigenspaces.*

Fact 8.10.7 (Spectral Theorem for Compact Self-adjoint Operators)**.** *Suppose that $T \in B(\mathcal{H})$ is a compact self-adjoint operator. Then*

$$T = \sum_{j=1}^{\infty} \lambda_j P_j$$

where the numbers $\{\lambda_j\}$ are the nonzero eigenvalues of T and the operators P_j are the orthogonal projections onto the corresponding finite-dimensional eigenspaces.

The details will be given in Functional Analysis! The more general Spectral Theorem for normal operators will have to wait until Banach and Operator Algebras.

The Spectral Theorem for Compact Self-adjoint Operators has some important applications in differential equations. While the differential operators one is interested in are typically unbounded, they have inverses which are compact and this allows you to use the theorem to say something about the solution sets.