Texts and Readings in Mathematics 38

Terence Tao

Analysis II

Third Edition

HINDUSTAN BOOK AGENCY

Springer

Terence Tao
Department of Mathematics
University of California, Los Angeles
Los Angeles, CA
USA

This work is a co-publication with Hindustan Book Agency, New Delhi, licensed for sale in all countries in electronic form only. Sold and distributed in print across the world by Hindustan Book Agency, P-19 Green Park Extension, New Delhi 110016, India. ISBN: 978-93-80250-65-6

ISSN 2366-8725 (electronic)
Texts and Readings in Mathematics
ISBN 978-981-10-1804-6 (eBook)
DOI 10.1007/978-981-10-1804-6

Library of Congress Control Number: 2016940817

This Springer imprint is published by Springer Nature
The registered company is Springer Science+Business Media Singapore Pte Ltd.

Contents

Preface to the second and third editions

Since the publication of the first edition, many students and lecturers have communicated a number of minor typos and other corrections to me. There was also some demand for a hardcover edition of the texts. Because of this, the publishers and I have decided to incorporate the corrections and issue a hardcover second edition of the textbooks. The layout, page numbering, and indexing of the texts have also been changed; in particular the two volumes are now numbered and indexed separately. However, the chapter and exercise numbering, as well as the mathematical content, remains the same as the first edition, and so the two editions can be used more or less interchangeably for homework and study purposes.

The third edition contains a number of corrections that were reported for the second edition, together with a few new exercises, but is otherwise essentially the same text.

Preface to the first edition

This text originated from the lecture notes I gave teaching the honours undergraduate-level real analysis sequence at the University of California, Los Angeles, in 2003. Among the undergraduates here, real analysis was viewed as being one of the most difficult courses to learn, not only because of the abstract concepts being introduced for the first time (e.g., topology, limits, measurability, etc.), but also because of the level of rigour and proof demanded of the course. Because of this perception of difficulty, one was often faced with the difficult choice of either reducing the level of rigour in the course in order to make it easier, or to maintain strict standards and face the prospect of many undergraduates, even many of the bright and enthusiastic ones, struggling with the course material.

Faced with this dilemma, I tried a somewhat unusual approach to the subject. Typically, an introductory sequence in real analysis assumes that the students are already familiar with the real numbers, with mathematical induction, with elementary calculus, and with the basics of set theory, and then quickly launches into the heart of the subject, for instance the concept of a limit. Normally, students entering this sequence do indeed have a fair bit of exposure to these prerequisite topics, though in most cases the material is not covered in a thorough manner. For instance, very few students were able to actually *define* a real number, or even an integer, properly, even though they could visualize these numbers intuitively and manipulate them algebraically. This seemed to me to be a missed opportunity. Real analysis is one of the first subjects (together with linear algebra and abstract algebra) that a student encounters, in which one truly has to grapple with the subtleties of a truly rigorous mathematical proof. As such, the course offered an excellent chance to go back to the foundations of mathematics, and in particular the opportunity to do a proper and thorough construction of the real numbers.

Thus the course was structured as follows. In the first week, I described some well-known "paradoxes" in analysis, in which standard laws of the subject (e.g., interchange of limits and sums, or sums and integrals) were applied in a non-rigorous way to give nonsensical results such as $0 = 1$. This motivated the need to go back to the very beginning of the subject, even to the very definition of the natural numbers, and check all the foundations from scratch. For instance, one of the first homework assignments was to check (using only the Peano axioms) that addition was associative for natural numbers (i.e., that $(a + b) + c = a + (b + c)$ for all natural numbers a, b, c: see Exercise 2.2.1). Thus even in the first week, the students had to write rigorous proofs using mathematical induction. After we had derived all the basic properties of the natural numbers, we then moved on to the integers (initially defined as formal differences of natural numbers); once the students had verified all the basic properties of the integers, we moved on to the rationals (initially defined as formal quotients of integers); and then from there we moved on (via formal limits of Cauchy sequences) to the reals. Around the same time, we covered the basics of set theory, for instance demonstrating the uncountability of the reals. Only then (after about ten lectures) did we begin what one normally considers the heart of undergraduate real analysis - limits, continuity, differentiability, and so forth.

The response to this format was quite interesting. In the first few weeks, the students found the material very easy on a conceptual level, as we were dealing only with the basic properties of the standard number systems. But on an intellectual level it was very challenging, as one was analyzing these number systems from a foundational viewpoint, in order to rigorously derive the more advanced facts about these number systems from the more primitive ones. One student told me how difficult it was to explain to his friends in the non-honours real analysis sequence (a) why he was still learning how to show why all rational numbers are either positive, negative, or zero (Exercise 4.2.4), while the non-honours sequence was already distinguishing absolutely convergent and conditionally convergent series, and (b) why, despite this, he thought his homework was significantly harder than that of his friends. Another student commented to me, quite wryly, that while she could obviously *see* why one could always divide a natural number n into a positive integer q to give a quotient a and a remainder r less than q (Exercise 2.3.5), she still had, to her frustration, much difficulty in writing down a proof of this fact. (I told her that later in the course she would have to prove statements for which it would not be as obvious to see that the statements were true; she did not seem to be particularly consoled by this.) Nevertheless, these students greatly enjoyed the homework, as when they did perservere and obtain a rigorous proof of an intuitive fact, it solidified the link in their minds between the abstract manipulations of formal mathematics and their informal intuition of mathematics (and of the real world), often in a very satisfying way. By the time they were assigned the task of giving the infamous "epsilon and delta" proofs in real analysis, they had already had so much experience with formalizing intuition, and in discerning the subtleties of mathematical logic (such as the distinction between the "for all" quantifier and the "there exists" quantifier), that the transition to these proofs was fairly smooth, and we were able to cover material both thoroughly and rapidly. By the tenth week, we had caught up with the non-honours class, and the students were verifying the change of variables formula for Riemann-Stieltjes integrals, and showing that piecewise continuous functions were Riemann integrable. By the conclusion of the sequence in the twentieth week, we had covered (both in lecture and in homework) the convergence theory of Taylor and Fourier series, the inverse and implicit function theorem for continuously differentiable functions of several variables, and established the dominated convergence theorem for the Lebesgue integral.

In order to cover this much material, many of the key foundational results were left to the student to prove as homework; indeed, this was an essential aspect of the course, as it ensured the students truly appreciated the concepts as they were being introduced. This format has been retained in this text; the majority of the exercises consist of proving lemmas, propositions and theorems in the main text. Indeed, I would strongly recommend that one do as many of these exercises as possible - and this includes those exercises proving "obvious" statements - if one wishes to use this text to learn real analysis; this is not a subject whose subtleties are easily appreciated just from passive reading. Most of the chapter sections have a number of exercises, which are listed at the end of the section.

To the expert mathematician, the pace of this book may seem somewhat slow, especially in early chapters, as there is a heavy emphasis on rigour (except for those discussions explicitly marked "Informal"), and justifying many steps that would ordinarily be quickly passed over as being self-evident. The first few chapters develop (in painful detail) many of the "obvious" properties of the standard number systems, for instance that the sum of two positive real numbers is again positive (Exercise 5.4.1), or that given any two distinct real numbers, one can find rational number between them (Exercise 5.4.5). In these foundational chapters, there is also an emphasis on *non-circularity* - not using later, more advanced results to prove earlier, more primitive ones. In particular, the usual laws of algebra are not used until they are derived (and they have to be derived separately for the natural numbers, integers, rationals, and reals). The reason for this is that it allows the students to learn the art of abstract reasoning, deducing true facts from a limited set of assumptions, in the friendly and intuitive setting of number systems; the payoff for this practice comes later, when one has to utilize the same type of reasoning techniques to grapple with more advanced concepts (e.g., the Lebesgue integral).

The text here evolved from my lecture notes on the subject, and thus is very much oriented towards a pedagogical perspective; much of the key material is contained inside exercises, and in many cases I have chosen to give a lengthy and tedious, but instructive, proof instead of a slick abstract proof. In more advanced textbooks, the student will see shorter and more conceptually coherent treatments of this material, and with more emphasis on intuition than on rigour; however, I feel it is important to know how to do analysis rigorously and "by hand" first, in order to truly appreciate the more modern, intuitive and abstract approach to analysis that one uses at the graduate level and beyond.

The exposition in this book heavily emphasizes rigour and formalism; however this does not necessarily mean that lectures based on this book have to proceed the same way. Indeed, in my own teaching I have used the lecture time to present the intuition behind the concepts (drawing many informal pictures and giving examples), thus providing a complementary viewpoint to the formal presentation in the text. The exercises assigned as homework provide an essential bridge between the two, requiring the student to combine both intuition and formal understanding together in order to locate correct proofs for a problem. This I found to be the most difficult task for the students, as it requires the subject to be genuinely *learnt*, rather than merely memorized or vaguely absorbed. Nevertheless, the feedback I received from the students was that the homework, while very demanding for this reason, was also very rewarding, as it allowed them to connect the rather abstract manipulations of formal mathematics with their innate intuition on such basic concepts as numbers, sets, and functions. Of course, the aid of a good teaching assistant is invaluable in achieving this connection.

With regard to examinations for a course based on this text, I would recommend either an open-book, open-notes examination with problems similar to the exercises given in the text (but perhaps shorter, with no unusual trickery involved), or else a take-home examination that involves problems comparable to the more intricate exercises in the text. The subject matter is too vast to force the students to memorize the definitions and theorems, so I would not recommend a closed-book examination, or an examination based on regurgitating extracts from the book. (Indeed, in my own examinations I gave a supplemental sheet listing the key definitions and theorems which were relevant to the examination problems.) Making the examinations similar to the homework assigned in the course will also help motivate the students to work through and understand their homework problems as thoroughly as possible (as opposed to, say, using flash cards or other such devices to memorize material), which is good preparation not only for examinations but for doing mathematics in general.

Some of the material in this textbook is somewhat peripheral to the main theme and may be omitted for reasons of time constraints. For instance, as set theory is not as fundamental to analysis as are the number systems, the chapters on set theory (Chapters 3, 8) can be covered more quickly and with substantially less rigour, or be given as reading assignments. The appendices on logic and the decimal system are intended as optional or supplemental reading and would probably not be covered in the main course lectures; the appendix on logic is particularly suitable for reading concurrently with the first few chapters. Also, Chapter 5 (on Fourier series) is not needed elsewhere in the text and can be omitted.

For reasons of length, this textbook has been split into two volumes. The first volume is slightly longer, but can be covered in about thirty lectures if the peripheral material is omitted or abridged. The second volume refers at times to the first, but can also be taught to students who have had a first course in analysis from other sources. It also takes about thirty lectures to cover.

I am deeply indebted to my students, who over the progression of the real analysis course corrected several errors in the lectures notes from which this text is derived, and gave other valuable feedback. I am also very grateful to the many anonymous referees who made several corrections and suggested many important improvements to the text.

I also thank Biswaranjan Behera, Tai-Danae Bradley, Brian, Eduardo Buscicchio, Carlos, EO, Florian, Gökhan Güçlü, Evangelos Georgiadis, Ulrich Groh, Bart Kleijngeld, Erik Koelink, Wang Kuyyang, Matthis Lehmkühler, Percy Li, Ming Li, Jason M., Manoranjan Majji, Geoff Mess, Pieter Naaijkens, Vineet Nair, Cristina Pereyra, David Radnell, Tim Reijnders, Pieter Roffelsen, Luke Rogers, Marc Schoolderman, Kent Van Vels, Daan Wanrooy, Yandong Xiao, Sam Xu, Luqing Ye, and the students of Math 401/501 and Math 402/502 at the University of New Mexico for corrections to the first and second editions.

Terence Tao

About the Author

Terence Tao, FAA FRS, is an Australian mathematician. His areas of interests are in harmonic analysis, partial differential equations, algebraic combinatorics, arithmetic combinatorics, geometric combinatorics, compressed sensing and analytic number theory. As of 2015, he holds the James and Carol Collins chair in mathematics at the University of California, Los Angeles. Professor Tao is a co-recipient of the 2006 Fields Medal and the 2014 Breakthrough Prize in Mathematics. He maintains a personal mathematics blog, which has been described by Timothy Gowers as “the undisputed king of all mathematics blogs”.

Chapter 1

Metric spaces

1.1 Definitions and examples

In Definition 6.1.5 we defined what it meant for a sequence $(x_n)_{n=m}^\infty$ of real numbers to converge to another real number x; indeed, this meant that for every $\varepsilon > 0$, there exists an $N \geq m$ such that $|x - x_n| \leq \varepsilon$ for all $n \geq N$. When this is the case, we write $\lim_{n\to\infty} x_n = x$.

Intuitively, when a sequence $(x_n)_{n=m}^\infty$ converges to a limit x, this means that somehow the elements x_n of that sequence will eventually be as close to x as one pleases. One way to phrase this more precisely is to introduce the *distance function* $d(x, y)$ between two real numbers by $d(x, y) := |x - y|$. (Thus for instance $d(3, 5) = 2$, $d(5, 3) = 2$, and $d(3, 3) = 0$.) Then we have

Lemma 1.1.1. *Let $(x_n)_{n=m}^\infty$ be a sequence of real numbers, and let x be another real number. Then $(x_n)_{n=m}^\infty$ converges to x if and only if $\lim_{n\to\infty} d(x_n, x) = 0$.*

Proof. See Exercise 1.1.1. □

One would now like to generalize this notion of convergence, so that one can take limits not just of sequences of real numbers, but also sequences of complex numbers, or sequences of vectors, or sequences of matrices, or sequences of functions, even sequences of sequences. One way to do this is to redefine the notion of convergence each time we deal with a new type of object. As you can guess, this will quickly get tedious. A more efficient way is to work *abstractly*, defining a very general class of spaces - which includes such standard spaces as the real numbers, complex numbers, vectors, etc. - and define the notion of convergence on this entire class of spaces at once. (A *space* is just the set

T. Tao, *Analysis II*, Texts and Readings in Mathematics 38,
DOI 10.1007/978-981-10-1804-6_1

of all objects of a certain type - the space of all real numbers, the space of all 3×3 matrices, etc. Mathematically, there is not much distinction between a space and a set, except that spaces tend to have much more structure than what a random set would have. For instance, the space of real numbers comes with operations such as addition and multiplication, while a general set would not.)

It turns out that there are two very useful classes of spaces which do the job. The first class is that of *metric spaces*, which we will study here. There is a more general class of spaces, called *topological spaces*, which is also very important, but we will only deal with this generalization briefly, in Section 2.5.

Roughly speaking, a metric space is any space X which has a concept of *distance* $d(x, y)$ - and this distance should behave in a reasonable manner. More precisely, we have

Definition 1.1.2 (Metric spaces)**.** A *metric space* (X, d) is a space X of objects (called *points*), together with a *distance function* or *metric* $d : X \times X \to [0, +\infty)$, which associates to each pair x, y of points in X a non-negative real number $d(x, y) \geq 0$. Furthermore, the metric must satisfy the following four axioms:

(a) For any $x \in X$, we have $d(x, x) = 0$.

(b) (Positivity) For any *distinct* $x, y \in X$, we have $d(x, y) > 0$.

(c) (Symmetry) For any $x, y \in X$, we have $d(x, y) = d(y, x)$.

(d) (Triangle inequality) For any $x, y, z \in X$, we have $d(x, z) \leq d(x, y) + d(y, z)$.

In many cases it will be clear what the metric d is, and we shall abbreviate (X, d) as just X.

Remark 1.1.3. The conditions (a) and (b) can be rephrased as follows: for any $x, y \in X$ we have $d(x, y) = 0$ if and only if $x = y$. (Why is this equivalent to (a) and (b)?)

Example 1.1.4 (The real line)**.** Let $\mathbf{R}$ be the real numbers, and let $d : \mathbf{R} \times \mathbf{R} \to [0, \infty)$ be the metric $d(x, y) := |x - y|$ mentioned earlier. Then $(\mathbf{R}, d)$ is a metric space (Exercise 1.1.2). We refer to d as the *standard metric* on $\mathbf{R}$, and if we refer to $\mathbf{R}$ as a metric space, we assume that the metric is given by the standard metric d unless otherwise specified.

Example 1.1.5 (Induced metric spaces)**.** Let (X, d) be any metric space, and let Y be a subset of X. Then we can restrict the metric function $d : X \times X \to [0, +\infty)$ to the subset $Y \times Y$ of $X \times X$ to create a restricted metric function $d|_{Y\times Y} : Y \times Y \to [0, +\infty)$ of Y; this is known as the metric on Y *induced* by the metric d on X. The pair $(Y, d|_{Y\times Y})$ is a metric space (Exercise 1.1.4) and is known the *subspace* of (X, d) induced by Y. Thus for instance the metric on the real line in the previous example induces a metric space structure on any subset of the reals, such as the integers $\mathbf{Z}$, or an interval $[a, b]$, etc.

Example 1.1.6 (Euclidean spaces)**.** Let $n \geq 1$ be a natural number, and let $\mathbf{R}^n$ be the space of n-tuples of real numbers:

$$\mathbf{R}^n = \{(x_1, x_2, \ldots, x_n) : x_1, \ldots, x_n \in \mathbf{R}\}.$$

We define the *Euclidean metric* (also called the l^2 *metric*) $d_{l^2} : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$ by

$$\begin{aligned} d_{l^2}((x_1, \ldots, x_n), (y_1, \ldots, y_n)) &:= \sqrt{(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2} \\ &= \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{1/2}. \end{aligned}$$

Thus for instance, if $n = 2$, then $d_{l^2}((1, 6), (4, 2)) = \sqrt{3^2 + 4^2} = 5$. This metric corresponds to the geometric distance between the two points $(x_1, x_2, \ldots, x_n)$, $(y_1, y_2, \ldots, y_n)$ as given by Pythagoras' theorem. (We remark however that while geometry does give some very important examples of metric spaces, it is possible to have metric spaces which have no obvious geometry whatsoever. Some examples are given below.) The verification that $(\mathbf{R}^n, d)$ is indeed a metric space can be seen geometrically (for instance, the triangle inequality now asserts that the length of one side of a triangle is always less than or equal to the sum of the lengths of the other two sides), but can also be proven algebraically (see Exercise 1.1.6). We refer to $(\mathbf{R}^n, d_{l^2})$ as the *Euclidean space* of *dimension n*. Extending the convention from Example 1.1.4, if we refer to $\mathbf{R}^n$ as a metric space, we assume that the metric is given by the Euclidean metric unless otherwise specified.

Example 1.1.7 (Taxi-cab metric)**.** Again let $n \geq 1$, and let $\mathbf{R}^n$ be as before. But now we use a different metric d_{l^1}, the so-called *taxicab* *metric* (or l^1 *metric*), defined by

$$d_{l^1}((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)) := |x_1 - y_1| + \ldots + |x_n - y_n| = \sum_{i=1}^{n} |x_i - y_i|.$$

Thus for instance, if $n = 2$, then $d_{l^1}((1, 6), (4, 2)) = 3 + 4 = 7$. This metric is called the taxi-cab metric, because it models the distance a taxi-cab would have to traverse to get from one point to another if the cab was only allowed to move in cardinal directions (north, south, east, west) and not diagonally. As such it is always at least as large as the Euclidean metric, which measures distance "as the crow flies", as it were. We claim that the space $(\mathbf{R}^n, d_{l^1})$ is also a metric space (Exercise 1.1.7). The metrics are not quite the same, but we do have the inequalities

$$d_{l^2}(x, y) \leq d_{l^1}(x, y) \leq \sqrt{n} d_{l^2}(x, y) \tag{1.1}$$

for all x, y (see Exercise 1.1.8).

Remark 1.1.8. The taxi-cab metric is useful in several places, for instance in the theory of error correcting codes. A string of n binary digits can be thought of as an element of $\mathbf{R}^n$, for instance the binary string 10010 can be thought of as the point $(1, 0, 0, 1, 0)$ in $\mathbf{R}^5$. The taxi-cab distance between two binary strings is then the number of bits in the two strings which do not match, for instance $d_{l^1}(10010, 10101) = 3$. The goal of error-correcting codes is to encode each piece of information (e.g., a letter of the alphabet) as a binary string in such a way that all the binary strings are as far away in the taxicab metric from each other as possible; this minimizes the chance that any distortion of the bits due to random noise can accidentally change one of the coded binary strings to another, and also maximizes the chance that any such distortion can be detected and correctly repaired.

Example 1.1.9 (Sup norm metric)**.** Again let $n \geq 1$, and let $\mathbf{R}^n$ be as before. But now we use a different metric d_{l^∞}, the so-called *sup norm metric* (or l^∞ *metric*), defined by

$$d_{l^\infty}((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)) := \sup\{|x_i - y_i| : 1 \leq i \leq n\}.$$

Thus for instance, if $n = 2$, then $d_{l^\infty}((1, 6), (4, 2)) = \sup(3, 4) = 4$. The space $(\mathbf{R}^n, d_{l^\infty})$ is also a metric space (Exercise 1.1.9), and is related to

Similarly the sequence converges to $(0,0)$ in the sup norm metric d_{l^∞} (why?). However, the sequence $(x^{(n)})_{n=1}^\infty$ does *not* converge to $(0,0)$ in the discrete metric d_{disc}, since

$$\lim_{n\to\infty} d_{\text{disc}}(x^{(n)}, (0,0)) = \lim_{n\to\infty} 1 = 1 \neq 0.$$

Thus the convergence of a sequence can depend on what metric one uses[1].

In the case of the above four metrics - Euclidean, taxi-cab, sup norm, and discrete - it is in fact rather easy to test for convergence.

Proposition 1.1.18 (Equivalence of l^1, l^2, l^∞)**.** *Let* $\mathbf{R}^n$ *be a Euclidean space, and let* $(x^{(k)})_{k=m}^\infty$ *be a sequence of points in* $\mathbf{R}^n$*. We write* $x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)})$*, i.e., for* $j = 1, 2, \ldots, n$*,* $x_j^{(k)} \in \mathbf{R}$ *is the* j^{th} *co-ordinate of* $x^{(k)} \in \mathbf{R}^n$*. Let* $x = (x_1, \ldots, x_n)$ *be a point in* $\mathbf{R}^n$*. Then the following four statements are equivalent:*

- *(a)* $(x^{(k)})_{k=m}^\infty$ *converges to* x *with respect to the Euclidean metric* d_{l^2}*.*
- *(b)* $(x^{(k)})_{k=m}^\infty$ *converges to* x *with respect to the taxi-cab metric* d_{l^1}*.*
- *(c)* $(x^{(k)})_{k=m}^\infty$ *converges to* x *with respect to the sup norm metric* d_{l^∞}*.*
- *(d) For every* $1 \leq j \leq n$*, the sequence* $(x_j^{(k)})_{k=m}^\infty$ *converges to* x_j*. (Notice that this is a sequence of real numbers, not of points in* $\mathbf{R}^n$*.)*

Proof. See Exercise 1.1.12. □

In other words, a sequence converges in the Euclidean, taxi-cab, or sup norm metric if and only if each of its components converges individually. Because of the equivalence of (a), (b) and (c), we say that the Euclidean, taxicab, and sup norm metrics on $\mathbf{R}^n$ are *equivalent*. (There are infinite-dimensional analogues of the Euclidean, taxicab, and sup norm metrics which are *not* equivalent, see for instance Exercise 1.1.15.)

---

[1] For a somewhat whimsical real-life example, one can give a city an "automobile metric", with $d(x,y)$ defined as the time it takes for a car to drive from x to y, or a "pedestrian metric", where $d(x,y)$ is the time it takes to walk on foot from x to y. (Let us assume for sake of argument that these metrics are symmetric, though this is not always the case in real life.) One can easily imagine examples where two points are close in one metric but not another.

For the discrete metric, convergence is much rarer: the sequence must be eventually constant in order to converge.

Proposition 1.1.19 (Convergence in the discrete metric)**.** *Let X be any set, and let d_{disc} be the discrete metric on X. Let $(x^{(n)})_{n=m}^\infty$ be a sequence of points in X, and let x be a point in X. Then $(x^{(n)})_{n=m}^\infty$ converges to x with respect to the discrete metric d_{disc} if and only if there exists an $N \geq m$ such that $x^{(n)} = x$ for all $n \geq N$.*

Proof. See Exercise 1.1.13. □

We now prove a basic fact about converging sequences; they can only converge to at most one point at a time.

Proposition 1.1.20 (Uniqueness of limits)**.** *Let (X, d) be a metric space, and let $(x^{(n)})_{n=m}^\infty$ be a sequence in X. Suppose that there are two points $x, x' \in X$ such that $(x^{(n)})_{n=m}^\infty$ converges to x with respect to d, and $(x^{(n)})_{n=m}^\infty$ also converges to x' with respect to d. Then we have $x = x'$.*

Proof. See Exercise 1.1.14. □

Because of the above Proposition, it is safe to introduce the following notation: if $(x^{(n)})_{n=m}^\infty$ converges to x in the metric d, then we write $d - \lim_{n\to\infty} x^{(n)} = x$, or simply $\lim_{n\to\infty} x^{(n)} = x$ when there is no confusion as to what d is. For instance, in the example of $(\frac{1}{n}, \frac{1}{n})$, we have

$$d_{l^2} - \lim_{n\to\infty} \left(\frac{1}{n}, \frac{1}{n}\right) = d_{l^1} - \lim_{n\to\infty} \left(\frac{1}{n}, \frac{1}{n}\right) = (0, 0),$$

but $d_{\mathrm{disc}} - \lim_{n\to\infty}(\frac{1}{n}, \frac{1}{n})$ is undefined. Thus the meaning of $d - \lim_{n\to\infty} x^{(n)}$ can depend on what d is; however Proposition 1.1.20 assures us that once d is fixed, there can be at most one value of $d - \lim_{n\to\infty} x^{(n)}$. (Of course, it is still possible that this limit does not exist; some sequences are not convergent.) Note that by Lemma 1.1.1, this definition of limit generalizes the notion of limit in Definition 6.1.8.

Remark 1.1.21. It is possible for a sequence to converge to one point using one metric, and another point using a different metric, although such examples are usually quite artificial. For instance, let $X := [0, 1]$, the closed interval from 0 to 1. Using the usual metric d, we have $d - \lim_{n\to\infty} \frac{1}{n} = 0$. But now suppose we "swap" the points 0 and 1 in the following manner. Let $f : [0, 1] \to [0, 1]$ be the function defined by $f(0) := 1$, $f(1) := 0$, and $f(x) := x$ for all $x \in (0, 1)$, and then define

$d'(x, y) := d(f(x), f(y))$. Then (X, d') is still a metric space (why?), but now $d' - \lim_{n\to\infty} \frac{1}{n} = 1$. Thus changing the metric on a space can greatly affect the nature of convergence (also called the *topology*) on that space; see Section 2.5 for a further discussion of topology.

— Exercises —

Exercise 1.1.1. Prove Lemma 1.1.1.

Exercise 1.1.2. Show that the real line with the metric $d(x, y) := |x-y|$ is indeed a metric space. (Hint: you may wish to review your proof of Proposition 4.3.3.)

Exercise 1.1.3. Let X be a set, and let $d : X \times X \to [0, \infty)$ be a function.

(a) Give an example of a pair (X, d) which obeys axioms (bcd) of Definition 1.1.2, but not (a). (Hint: modify the discrete metric.)

(b) Give an example of a pair (X, d) which obeys axioms (acd) of Definition 1.1.2, but not (b).

(c) Give an example of a pair (X, d) which obeys axioms (abd) of Definition 1.1.2, but not (c).

(d) Give an example of a pair (X, d) which obeys axioms (abc) of Definition 1.1.2, but not (d). (Hint: try examples where X is a finite set.)

Exercise 1.1.4. Show that the pair $(Y, d|_{Y\times Y})$ defined in Example 1.1.5 is indeed a metric space.

Exercise 1.1.5. Let $n \geq 1$, and let $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ be real numbers. Verify the identity

$$\left(\sum_{i=1}^n a_i b_i\right)^2 + \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n (a_i b_j - a_j b_i)^2 = \left(\sum_{i=1}^n a_i^2\right)\left(\sum_{j=1}^n b_j^2\right),$$

and conclude the *Cauchy-Schwarz inequality*

$$\left|\sum_{i=1}^n a_i b_i\right| \leq \left(\sum_{i=1}^n a_i^2\right)^{1/2}\left(\sum_{j=1}^n b_j^2\right)^{1/2}. \tag{1.3}$$

Then use the Cauchy-Schwarz inequality to prove the *triangle inequality*

$$\left(\sum_{i=1}^n (a_i + b_i)^2\right)^{1/2} \leq \left(\sum_{i=1}^n a_i^2\right)^{1/2} + \left(\sum_{j=1}^n b_j^2\right)^{1/2}.$$

Exercise 1.1.6. Show that $(\mathbf{R}^n, d_{l^2})$ in Example 1.1.6 is indeed a metric space. (Hint: use Exercise 1.1.5.)

Exercise 1.1.7. Show that the pair $(\mathbf{R}^n, d_{l^1})$ in Example 1.1.7 is indeed a metric space.

Exercise 1.1.8. Prove the two inequalities in (1.1). (For the first inequality, square both sides. For the second inequality, use Exercise (1.1.5).

Exercise 1.1.9. Show that the pair $(\mathbf{R}^n, d_{l^\infty})$ in Example 1.1.9 is indeed a metric space.

Exercise 1.1.10. Prove the two inequalities in (1.2).

Exercise 1.1.11. Show that the discrete metric $(\mathbf{R}^n, d_{\text{disc}})$ in Example 1.1.11 is indeed a metric space.

Exercise 1.1.12. Prove Proposition 1.1.18.

Exercise 1.1.13. Prove Proposition 1.1.19.

Exercise 1.1.14. Prove Proposition 1.1.20. (Hint: modify the proof of Proposition 6.1.7.)

Exercise 1.1.15. Let

$$X := \left\{ (a_n)_{n=0}^\infty : \sum_{n=0}^\infty |a_n| < \infty \right\}$$

be the space of absolutely convergent sequences. Define the l^1 and l^∞ metrics on this space by

$$d_{l^1}((a_n)_{n=0}^\infty, (b_n)_{n=0}^\infty) := \sum_{n=0}^\infty |a_n - b_n|;$$
$$d_{l^\infty}((a_n)_{n=0}^\infty, (b_n)_{n=0}^\infty) := \sup_{n \in \mathbf{N}} |a_n - b_n|.$$

Show that these are both metrics on X, but show that there exist sequences $x^{(1)}, x^{(2)}, \ldots$ of elements of X (i.e., sequences of sequences) which are convergent with respect to the d_{l^∞} metric but not with respect to the d_{l^1} metric. Conversely, show that any sequence which converges in the d_{l^1} metric automatically converges in the d_{l^∞} metric.

Exercise 1.1.16. Let $(x_n)_{n=1}^\infty$ and $(y_n)_{n=1}^\infty$ be two sequences in a metric space (X, d). Suppose that $(x_n)_{n=1}^\infty$ converges to a point $x \in X$, and $(y_n)_{n=1}^\infty$ converges to a point $y \in X$. Show that $\lim_{n\to\infty} d(x_n, y_n) = d(x, y)$. (Hint: use the triangle inequality several times.)

1.2 Some point-set topology of metric spaces

Having defined the operation of convergence on metric spaces, we now define a couple other related notions, including that of open set, closed set, interior, exterior, boundary, and adherent point. The study of such notions is known as *point-set topology*, which we shall return to in Section 2.5.

We first need the notion of a *metric ball*, or more simply a *ball*.

Definition 1.2.1 (Balls)**.** Let (X, d) be a metric space, let x_0 be a point in X, and let $r > 0$. We define the *ball* $B_{(X,d)}(x_0, r)$ in X, centered at x_0, and with radius r, in the metric d, to be the set

$$B_{(X,d)}(x_0, r) := \{x \in X : d(x, x_0) < r\}.$$

When it is clear what the metric space (X, d) is, we shall abbreviate $B_{(X,d)}(x_0, r)$ as just $B(x_0, r)$.

Example 1.2.2. In $\mathbf{R}^2$ with the Euclidean metric d_{l^2}, the ball $B_{(\mathbf{R}^2, d_{l^2})}((0,0), 1)$ is the open disc

$$B_{(\mathbf{R}^2, d_{l^2})}((0,0), 1) = \{(x, y) \in \mathbf{R}^2 : x^2 + y^2 < 1\}.$$

However, if one uses the taxi-cab metric d_{l^1} instead, then we obtain a diamond:

$$B_{(\mathbf{R}^2, d_{l^1})}((0,0), 1) = \{(x, y) \in \mathbf{R}^2 : |x| + |y| < 1\}.$$

If we use the discrete metric, the ball is now reduced to a single point:

$$B_{(\mathbf{R}^2, d_{\text{disc}})}((0,0), 1) = \{(0,0)\},$$

although if one increases the radius to be larger than 1, then the ball now encompasses all of $\mathbf{R}^2$. (Why?)

Example 1.2.3. In $\mathbf{R}$ with the usual metric d, the open interval $(3, 7)$ is also the metric ball $B_{(\mathbf{R},d)}(5, 2)$.

Remark 1.2.4. Note that the smaller the radius r, the smaller the ball $B(x_0, r)$. However, $B(x_0, r)$ always contains at least one point, namely the center x_0, as long as r stays positive, thanks to Definition 1.1.2(a). (We don't consider balls of zero radius or negative radius since they are rather boring, being just the empty set.)

Using metric balls, one can now take a set E in a metric space X, and classify three types of points in X: interior, exterior, and boundary points of E.

Definition 1.2.5 (Interior, exterior, boundary)**.** Let (X, d) be a metric space, let E be a subset of X, and let x_0 be a point in X. We say that x_0 is an *interior point of* E if there exists a radius $r > 0$ such that $B(x_0, r) \subseteq E$. We say that x_0 is an *exterior point of* E if there exists a

From the above two remarks we see that the notions of being open and being closed are *not* negations of each other; there are sets that are both open and closed, and there are sets which are neither open nor closed. Thus, if one knew for instance that E was not an open set, it would be erroneous to conclude from this that E was a closed set, and similarly with the rôles of open and closed reversed. The correct relationship between open and closed sets is given by Proposition 1.2.15(e) below.

Now we list some more properties of open and closed sets.

Proposition 1.2.15 (Basic properties of open and closed sets)**.** *Let (X, d) be a metric space.*

(*a*) *Let E be a subset of X. Then E is open if and only if $E = \mathrm{int}(E)$. In other words, E is open if and only if for every $x \in E$, there exists an $r > 0$ such that $B(x, r) \subseteq E$.*

(*b*) *Let E be a subset of X. Then E is closed if and only if E contains all its adherent points. In other words, E is closed if and only if for every convergent sequence $(x_n)_{n=m}^{\infty}$ in E, the limit $\lim_{n\to\infty} x_n$ of that sequence also lies in E.*

(*c*) *For any $x_0 \in X$ and $r > 0$, then the ball $B(x_0, r)$ is an open set. The set $\{x \in X : d(x, x_0) \leq r\}$ is a closed set. (This set is sometimes called the* closed ball *of radius r centered at x_0.)*

(*d*) *Any singleton set $\{x_0\}$, where $x_0 \in X$, is automatically closed.*

(*e*) *If E is a subset of X, then E is open if and only if the complement $X \backslash E := \{x \in X : x \notin E\}$ is closed.*

(*f*) *If $E_1, \ldots, E_n$ are a finite collection of open sets in X, then $E_1 \cap E_2 \cap \ldots \cap E_n$ is also open. If $F_1, \ldots, F_n$ is a finite collection of closed sets in X, then $F_1 \cup F_2 \cup \ldots \cup F_n$ is also closed.*

(*g*) *If $\{E_\alpha\}_{\alpha\in I}$ is a collection of open sets in X (where the index set I could be finite, countable, or uncountable), then the union $\bigcup_{\alpha\in I} E_\alpha := \{x \in X : x \in E_\alpha \text{ for some } \alpha \in I\}$ is also open. If $\{F_\alpha\}_{\alpha\in I}$ is a collection of closed sets in X, then the intersection $\bigcap_{\alpha\in I} F_\alpha := \{x \in X : x \in F_\alpha \text{ for all } \alpha \in I\}$ is also closed.*

(*h*) *If E is any subset of X, then $\mathrm{int}(E)$ is the largest open set which is contained in E; in other words, $\mathrm{int}(E)$ is open, and given any* *other open set $V \subseteq E$, we have $V \subseteq \text{int}(E)$. Similarly $\overline{E}$ is the smallest closed set which contains E; in other words, $\overline{E}$ is closed, and given any other closed set $K \supset E$, $K \supset \overline{E}$.*

Proof. See Exercise 1.2.3. □

— Exercises —

Exercise 1.2.1. Verify the claims in Example 1.2.8.

Exercise 1.2.2. Prove Proposition 1.2.10. (Hint: for some of the implications one will need the axiom of choice, as in Lemma 8.4.5.)

Exercise 1.2.3. Prove Proposition 1.2.15. (Hint: you can use earlier parts of the proposition to prove later ones.)

Exercise 1.2.4. Let (X, d) be a metric space, x_0 be a point in X, and $r > 0$. Let B be the open ball $B := B(x_0, r) = \{x \in X : d(x, x_0) < r\}$, and let C be the closed ball $C := \{x \in X : d(x, x_0) \leq r\}$.

(a) Show that $\overline{B} \subseteq C$.

(b) Give an example of a metric space (X, d), a point x_0, and a radius $r > 0$ such that $\overline{B}$ is *not* equal to C.

1.3 Relative topology

When we defined notions such as open and closed sets, we mentioned that such concepts depended on the choice of metric one uses. For instance, on the real line $\mathbf{R}$, if one uses the usual metric $d(x, y) = |x-y|$, then the set $\{1\}$ is not open, however if instead one uses the discrete metric d_{disc}, then $\{1\}$ is now an open set (why?).

However, it is not just the choice of metric which determines what is open and what is not - it is also the choice of *ambient space* X. Here are some examples.

Example 1.3.1. Consider the plane $\mathbf{R}^2$ with the Euclidean metric d_{l^2}. Inside the plane, we can find the x-axis $X := \{(x, 0) : x \in \mathbf{R}\}$. The metric d_{l^2} can be restricted to X, creating a subspace $(X, d_{l^2}|_{X \times X})$ of $(\mathbf{R}^2, d_{l^2})$. (This subspace is essentially the same as the real line $(\mathbf{R}, d)$ with the usual metric; the precise way of stating this is that $(X, d_{l^2}|_{X \times X})$ is *isometric* to $(\mathbf{R}, d)$. We will not pursue this concept further in this text, however.) Now consider the set

$$E := \{(x, 0) : -1 < x < 1\}$$