Texts and Readings in Mathematics 37

Terence Tao

Analysis I

Third Edition

HINDUSTAN BOOK AGENCY

Springer

Copyrighted material

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: 978-93-80250-64-9
© Hindustan Book Agency 2015.

ISSN 2366-8725 (electronic)
Texts and Readings in Mathematics
ISBN 978-981-10-1789-6 (eBook)
DOI 10.1007/978-981-10-1789-6

Library of Congress Control Number: 2016940817

© Springer Science+Business Media Singapore 2016 and Hindustan Book Agency 2015
This work is subject to copyright. All rights are reserved by the Publishers, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publishers, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publishers nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

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

Copyrighted material

Contents

Copyrighted material

Copyrighted material

Copyrighted material

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.

Copyrighted material

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

Copyrighted material

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

Copyrighted material

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

Copyrighted material

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

Copyrighted material

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 11.27 (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.

Copyrighted material

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

Copyrighted material

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”.

Copyrighted material

Chapter 1

Introduction

1.1 What is analysis?

This text is an honours-level undergraduate introduction to *real analysis*: the analysis of the real numbers, sequences and series of real numbers, and real-valued functions. This is related to, but is distinct from, *complex analysis*, which concerns the analysis of the complex numbers and complex functions, *harmonic analysis*, which concerns the analysis of harmonics (waves) such as sine waves, and how they synthesize other functions via the Fourier transform, *functional analysis*, which focuses much more heavily on functions (and how they form things like vector spaces), and so forth. *Analysis* is the rigorous study of such objects, with a focus on trying to pin down precisely and accurately the qualitative and quantitative behavior of these objects. Real analysis is the theoretical foundation which underlies *calculus*, which is the collection of computational algorithms which one uses to manipulate functions.

In this text we will be studying many objects which will be familiar to you from freshman calculus: numbers, sequences, series, limits, functions, definite integrals, derivatives, and so forth. You already have a great deal of experience of *computing* with these objects; however here we will be focused more on the underlying theory for these objects. We will be concerned with questions such as the following:

1. What is a real number? Is there a largest real number? After 0, what is the "next" real number (i.e., what is the smallest positive real number)? Can you cut a real number into pieces infinitely many times? Why does a number such as 2 have a square root, while a number such as -2 does not? If there are infinitely many

© Springer Science+Business Media Singapore 2016 and Hindustan Book Agency 2015

T. Tao, *Analysis I*, Texts and Readings in Mathematics 37,
DOI 10.1007/978-981-10-1789-6_1
Copyrighted material

reals and infinitely many rationals, how come there are "more" real numbers than rational numbers?

2. How do you take the limit of a sequence of real numbers? Which sequences have limits and which ones don't? If you can stop a sequence from escaping to infinity, does this mean that it must eventually settle down and converge? Can you add infinitely many real numbers together and still get a finite real number? Can you add infinitely many rational numbers together and end up with a non-rational number? If you rearrange the elements of an infinite sum, is the sum still the same?

3. What is a function? What does it mean for a function to be continuous? differentiable? integrable? bounded? Can you add infinitely many functions together? What about taking limits of sequences of functions? Can you differentiate an infinite series of functions? What about integrating? If a function $f(x)$ takes the value 3 when $x = 0$ and 5 when $x = 1$ (i.e., $f(0) = 3$ and $f(1) = 5$), does it have to take every intermediate value between 3 and 5 when x goes between 0 and 1? Why?

You may already know how to answer some of these questions from your calculus classes, but most likely these sorts of issues were only of secondary importance to those courses; the emphasis was on getting you to perform computations, such as computing the integral of $x \sin(x^2)$ from $x = 0$ to $x = 1$. But now that you are comfortable with these objects and already know how to do all the computations, we will go back to the theory and try to *really* understand what is going on.

1.2 Why do analysis?

It is a fair question to ask, "why bother?", when it comes to analysis. There is a certain philosophical satisfaction in knowing *why* things work, but a pragmatic person may argue that one only needs to know *how* things work to do real-life problems. The calculus training you receive in introductory classes is certainly adequate for you to begin solving many problems in physics, chemistry, biology, economics, computer science, finance, engineering, or whatever else you end up doing - and you can certainly use things like the chain rule, L'Hôpital's rule, or integration by parts without knowing why these rules work, or whether there are any exceptions to these rules. However, one can get into trouble if

Copyrighted material

one applies rules without knowing where they came from and what the limits of their applicability are. Let me give some examples in which several of these familiar rules, if applied blindly without knowledge of the underlying analysis, can lead to disaster.

Example 1.2.1 (Division by zero)**.** This is a very familiar one to you: the cancellation law $ac = bc \implies a = b$ does not work when $c = 0$. For instance, the identity $1 \times 0 = 2 \times 0$ is true, but if one blindly cancels the 0 then one obtains $1 = 2$, which is false. In this case it was obvious that one was dividing by zero; but in other cases it can be more hidden.

Example 1.2.2 (Divergent series)**.** You have probably seen geometric series such as the infinite sum

$$S = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \ldots.$$

You have probably seen the following trick to sum this series: if we call the above sum S, then if we multiply both sides by 2, we obtain

$$2S = 2 + 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 2 + S$$

and hence $S = 2$, so the series sums to 2. However, if you apply the same trick to the series

$$S = 1 + 2 + 4 + 8 + 16 + \ldots$$

one gets nonsensical results:

$$2S = 2 + 4 + 8 + 16 + \ldots = S - 1 \implies S = -1.$$

So the same reasoning that shows that $1 + \frac{1}{2} + \frac{1}{4} + \ldots = 2$ also gives that $1 + 2 + 4 + 8 + \ldots = -1$. Why is it that we trust the first equation but not the second? A similar example arises with the series

$$S = 1 - 1 + 1 - 1 + 1 - 1 + \ldots;$$

we can write

$$S = 1 - (1 - 1 + 1 - 1 + \ldots) = 1 - S$$

and hence that $S = 1/2$; or instead we can write

$$S = (1 - 1) + (1 - 1) + (1 - 1) + \ldots = 0 + 0 + \ldots$$

Copyrighted material

and hence that $S = 0$; or instead we can write

$$S = 1 + (-1 + 1) + (-1 + 1) + \ldots = 1 + 0 + 0 + \ldots$$

and hence that $S = 1$. Which one is correct? (See Exercise 7.2.1 for an answer.)

Example 1.2.3 (Divergent sequences)**.** Here is a slight variation of the previous example. Let x be a real number, and let L be the limit

$$L = \lim_{n\to\infty} x^n.$$

Changing variables $n = m + 1$, we have

$$L = \lim_{m+1\to\infty} x^{m+1} = \lim_{m+1\to\infty} x \times x^m = x \lim_{m+1\to\infty} x^m.$$

But if $m + 1 \to \infty$, then $m \to \infty$, thus

$$\lim_{m+1\to\infty} x^m = \lim_{m\to\infty} x^m = \lim_{n\to\infty} x^n = L,$$

and thus

$$xL = L.$$

At this point we could cancel the L's and conclude that $x = 1$ for an arbitrary real number x, which is absurd. But since we are already aware of the division by zero problem, we could be a little smarter and conclude instead that either $x = 1$, or $L = 0$. In particular we seem to have shown that

$$\lim_{n\to\infty} x^n = 0 \text{ for all } x \neq 1.$$

But this conclusion is absurd if we apply it to certain values of x, for instance by specializing to the case $x = 2$ we could conclude that the sequence $1, 2, 4, 8, \ldots$ converges to zero, and by specializing to the case $x = -1$ we conclude that the sequence $1, -1, 1, -1, \ldots$ also converges to zero. These conclusions appear to be absurd; what is the problem with the above argument? (See Exercise 6.3.4 for an answer.)

Example 1.2.4 (Limiting values of functions)**.** Start with the expression $\lim_{x\to\infty} \sin(x)$, make the change of variable $x = y + \pi$ and recall that $\sin(y + \pi) = -\sin(y)$ to obtain

$$\lim_{x\to\infty} \sin(x) = \lim_{y+\pi\to\infty} \sin(y + \pi) = \lim_{y\to\infty} (-\sin(y)) = -\lim_{y\to\infty} \sin(y).$$

Copyrighted material

Since $\lim_{x\to\infty}\sin(x) = \lim_{y\to\infty}\sin(y)$ we thus have

$$\lim_{x\to\infty}\sin(x) = -\lim_{x\to\infty}\sin(x)$$

and hence

$$\lim_{x\to\infty}\sin(x) = 0.$$

If we then make the change of variables $x = \pi/2 + z$ and recall that $\sin(\pi/2 + z) = \cos(z)$ we conclude that

$$\lim_{x\to\infty}\cos(x) = 0.$$

Squaring both of these limits and adding we see that

$$\lim_{x\to\infty}(\sin^2(x) + \cos^2(x)) = 0^2 + 0^2 = 0.$$

On the other hand, we have $\sin^2(x) + \cos^2(x) = 1$ for all x. Thus we have shown that $1 = 0$! What is the difficulty here?

Example 1.2.5 (Interchanging sums)**.** Consider the following fact of arithmetic. Consider any matrix of numbers, e.g.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

and compute the sums of all the rows and the sums of all the columns, and then total all the row sums and total all the column sums. In both cases you will get the same number - the total sum of all the entries in the matrix:

$$\begin{matrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} & \begin{matrix} 6 \\ 15 \\ 24 \end{matrix} \\ \begin{matrix} 12 & 15 & 18 \end{matrix} & 45 \end{matrix}$$

To put it another way, if you want to add all the entries in an $m \times n$ matrix together, it doesn't matter whether you sum the rows first or sum the columns first, you end up with the same answer. (Before the invention of computers, accountants and book-keepers would use this fact to guard against making errors when balancing their books.) In

Copyrighted material

series notation, this fact would be expressed as

$$\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} = \sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij},$$

if a_{ij} denoted the entry in the i^{th} row and j^{th} column of the matrix.

Now one might think that this rule should extend easily to infinite series:

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} a_{ij} = \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} a_{ij}.$$

Indeed, if you use infinite series a lot in your work, you will find yourself having to switch summations like this fairly often. Another way of saying this fact is that in an infinite matrix, the sum of the row-totals should equal the sum of the column-totals. However, despite the reasonableness of this statement, it is actually false! Here is a counterexample:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \dots \\ -1 & 1 & 0 & 0 & \dots \\ 0 & -1 & 1 & 0 & \dots \\ 0 & 0 & -1 & 1 & \dots \\ 0 & 0 & 0 & -1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

If you sum up all the rows, and then add up all the row totals, you get 1; but if you sum up all the columns, and add up all the column totals, you get 0! So, does this mean that summations for infinite series should not be swapped, and that any argument using such a swapping should be distrusted? (See Theorem 8.2.2 for an answer.)

Example 1.2.6 (Interchanging integrals)**.** The interchanging of integrals is a trick which occurs in mathematics just as commonly as the interchanging of sums. Suppose one wants to compute the volume under a surface $z = f(x, y)$ (let us ignore the limits of integration for the moment). One can do it by slicing parallel to the x-axis: for each fixed value of y, we can compute an area $\int f(x, y)\ dx$, and then we integrate the area in the y variable to obtain the volume

$$V = \int \int f(x, y) dx dy.$$

Copyrighted material

Or we could slice parallel to the y-axis for each fixed x and compute an area $\int f(x,y)\ dy$, and then integrate in the x-axis to obtain

$$V = \int \int f(x,y) dy dx.$$

This seems to suggest that one should always be able to swap integral signs:

$$\int \int f(x,y)\ dx dy = \int \int f(x,y)\ dy dx.$$

And indeed, people swap integral signs all the time, because sometimes one variable is easier to integrate in first than the other. However, just as infinite sums sometimes cannot be swapped, integrals are also sometimes dangerous to swap. An example is with the integrand $e^{-xy} - xye^{-xy}$. Suppose we believe that we can swap the integrals:

$$\int_0^\infty \int_0^1 (e^{-xy} - xye^{-xy})\ dy\ dx = \int_0^1 \int_0^\infty (e^{-xy} - xye^{-xy})\ dx\ dy. \quad (1.1)$$

Since

$$\int_0^1 (e^{-xy} - xye^{-xy})\ dy = ye^{-xy}|_{y=0}^{y=1} = e^{-x},$$

the left-hand side of (1.1) is $\int_0^\infty e^{-x}\ dx = -e^{-x}|_0^\infty = 1$. But since

$$\int_0^\infty (e^{-xy} - xye^{-xy})\ dx = xe^{-xy}|_{x=0}^{x=\infty} = 0,$$

the right-hand side of (1.1) is $\int_0^1 0\ dx = 0$. Clearly $1 \neq 0$, so there is an error somewhere; but you won't find one anywhere except in the step where we interchanged the integrals. So how do we know when to trust the interchange of integrals? (See Theorem 11.50.1 for a partial answer.)

Example 1.2.7 (Interchanging limits)**.** Suppose we start with the plausible looking statement

$$\lim_{x\to 0} \lim_{y\to 0} \frac{x^2}{x^2+y^2} = \lim_{y\to 0} \lim_{x\to 0} \frac{x^2}{x^2+y^2}. \quad (1.2)$$

But we have

$$\lim_{y\to 0} \frac{x^2}{x^2+y^2} = \frac{x^2}{x^2+0^2} = 1,$$

Copyrighted material

so the left-hand side of (1.2) is 1; on the other hand, we have

$$\lim_{x \to 0} \frac{x^2}{x^2 + y^2} = \frac{0^2}{0^2 + y^2} = 0,$$

so the right-hand side of (1.2) is 0. Since 1 is clearly not equal to zero, this suggests that interchange of limits is untrustworthy. But are there any other circumstances in which the interchange of limits is legitimate? (See Exercise 11.9.9 for a partial answer.)

Example 1.2.8 (Interchanging limits, again)**.** Consider the plausible looking statement

$$\lim_{x \to 1^-} \lim_{n \to \infty} x^n = \lim_{n \to \infty} \lim_{x \to 1^-} x^n$$

where the notation $x \to 1^-$ means that x is approaching 1 from the left. When x is to the left of 1, then $\lim_{n \to \infty} x^n = 0$, and hence the left-hand side is zero. But we also have $\lim_{x \to 1^-} x^n = 1$ for all n, and so the right-hand side limit is 1. Does this demonstrate that this type of limit interchange is always untrustworthy? (See Proposition 11.15.3 for an answer.)

Example 1.2.9 (Interchanging limits and integrals)**.** For any real number y, we have

$$\int_{-\infty}^{\infty} \frac{1}{1 + (x - y)^2}\, dx = \arctan(x - y)|_{x=-\infty}^{\infty} = \frac{\pi}{2} - \left(-\frac{\pi}{2}\right) = \pi.$$

Taking limits as $y \to \infty$, we should obtain

$$\int_{-\infty}^{\infty} \lim_{y \to \infty} \frac{1}{1 + (x - y)^2}\, dx = \lim_{y \to \infty} \int_{-\infty}^{\infty} \frac{1}{1 + (x - y)^2}\, dx = \pi.$$

But for every x, we have $\lim_{y \to \infty} \frac{1}{1+(x-y)^2} = 0$. So we seem to have concluded that $0 = \pi$. What was the problem with the above argument? Should one abandon the (very useful) technique of interchanging limits and integrals? (See Theorem 11.18.1 for a partial answer.)

Example 1.2.10 (Interchanging limits and derivatives)**.** Observe that if $\varepsilon > 0$, then

$$\frac{d}{dx}\left(\frac{x^3}{\varepsilon^2 + x^2}\right) = \frac{3x^2(\varepsilon^2 + x^2) - 2x^4}{(\varepsilon^2 + x^2)^2}$$

Copyrighted material

and in particular that

$$\frac{d}{dx}\left(\frac{x^3}{\varepsilon^2+x^2}\right)|_{x=0} = 0.$$

Taking limits as $\varepsilon \to 0$, one might then expect that

$$\frac{d}{dx}\left(\frac{x^3}{0+x^2}\right)|_{x=0} = 0.$$

But the right-hand side is $\frac{d}{dx}x = 1$. Does this mean that it is always illegitimate to interchange limits and derivatives? (See Theorem 11.19.1 for an answer.)

Example 1.2.11 (Interchanging derivatives)**.** Let[1] $f(x,y)$ be the function $f(x,y) := \frac{xy^3}{x^2+y^2}$. A common maneuvre in analysis is to interchange two partial derivatives, thus one expects

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) = \frac{\partial^2 f}{\partial y \partial x}(0,0).$$

But from the quotient rule we have

$$\frac{\partial f}{\partial y}(x,y) = \frac{3xy^2}{x^2+y^2} - \frac{2xy^4}{(x^2+y^2)^2}$$

and in particular

$$\frac{\partial f}{\partial y}(x,0) = \frac{0}{x^2} - \frac{0}{x^4} = 0.$$

Thus

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) = 0.$$

On the other hand, from the quotient rule again we have

$$\frac{\partial f}{\partial x}(x,y) = \frac{y^3}{x^2+y^2} - \frac{2x^2y^3}{(x^2+y^2)^2}$$

and hence

$$\frac{\partial f}{\partial x}(0,y) = \frac{y^3}{y^2} - \frac{0}{y^4} = y.$$

[1] One might object that this function is not defined at $(x,y) = (0,0)$, but if we set $f(0,0) := (0,0)$ then this function becomes continuous and differentiable for all (x,y), and in fact both partial derivatives $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$ are also continuous and differentiable for all (x,y)!

Copyrighted material

Thus

$$\frac{\partial^2 f}{\partial y \partial x}(0,0) = 1.$$

Since $1 \neq 0$, we thus seem to have shown that interchange of derivatives is untrustworthy. But are there any other circumstances in which the interchange of derivatives is legitimate? (See Theorem 11.37.4 and Exercise 11.37.1 for some answers.)

Example 1.2.12 (L'Hôpital's rule)**.** We are all familiar with the beautifully simple L'Hôpital's rule

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = \lim_{x \to x_0} \frac{f'(x)}{g'(x)},$$

but one can still get led to incorrect conclusions if one applies it incorrectly. For instance, applying it to $f(x) := x$, $g(x) := 1 + x$, and $x_0 := 0$ we would obtain

$$\lim_{x \to 0} \frac{x}{1+x} = \lim_{x \to 0} \frac{1}{1} = 1,$$

but this is the incorrect answer, since $\lim_{x \to 0} \frac{x}{1+x} = \frac{0}{1+0} = 0$. Of course, all that is going on here is that L'Hôpital's rule is only applicable when both $f(x)$ and $g(x)$ go to zero as $x \to x_0$, a condition which was violated in the above example. But even when $f(x)$ and $g(x)$ do go to zero as $x \to x_0$ there is still a possibility for an incorrect conclusion. For instance, consider the limit

$$\lim_{x \to 0} \frac{x^2 \sin(x^{-4})}{x}.$$

Both numerator and denominator go to zero as $x \to 0$, so it seems pretty safe to apply L'Hôpital's rule, to obtain

$$\begin{aligned} \lim_{x \to 0} \frac{x^2 \sin(x^{-4})}{x} &= \lim_{x \to 0} \frac{2x \sin(x^{-4}) - 4x^{-3} \cos(x^{-4})}{1} \\ &= \lim_{x \to 0} 2x \sin(x^{-4}) - \lim_{x \to 0} 4x^{-3} \cos(x^{-4}). \end{aligned}$$

The first limit converges to zero by the squeeze test (since the function $2x \sin(x^{-4})$ is bounded above by $2|x|$ and below by $-2|x|$, both of which go to zero at 0). But the second limit is divergent (because x^{-3} goes to infinity as $x \to 0$, and $\cos(x^{-4})$ does not go to zero). So the limit $\lim_{x \to 0} \frac{2x \sin(x^{-4}) - 4x^{-2} \cos(x^{-4})}{1}$ diverges. One might then conclude using L'Hôpital's rule that $\lim_{x \to 0} \frac{x^2 \sin(x^{-4})}{x}$ also diverges; however we can

Copyrighted material

Chapter 2

Starting at the beginning: the natural numbers

In this text, we will review the material you have learnt in high school and in elementary calculus classes, but as rigorously as possible. To do so we will have to begin at the very basics - indeed, we will go back to the concept of *numbers* and what their properties are. Of course, you have dealt with numbers for over ten years and you know how to manipulate the rules of algebra to simplify any expression involving numbers, but we will now turn to a more fundamental issue, which is: *why* do the rules of algebra work at all? For instance, why is it true that $a(b+c)$ is equal to $ab+ac$ for any three numbers a, b, c? This is not an arbitrary choice of rule; it can be proven from more primitive, and more fundamental, properties of the number system. This will teach you a new skill - how to prove complicated properties from simpler ones. You will find that even though a statement may be "obvious", it may not be easy to prove; the material here will give you plenty of practice in doing so, and in the process will lead you to think about *why* an obvious statement really is obvious. One skill in particular that you will pick up here is the use of *mathematical induction*, which is a basic tool in proving things in many areas of mathematics.

So in the first few chapters we will re-acquaint you with various number systems that are used in real analysis. In increasing order of sophistication, they are the *natural numbers* $\mathbf{N}$; the *integers* $\mathbf{Z}$; the *rationals* $\mathbf{Q}$, and the *real numbers* $\mathbf{R}$. (There are other number systems such as the *complex numbers* $\mathbf{C}$, but we will not study them until Section 11.26.) The natural numbers $\{0, 1, 2, \ldots\}$ are the most primitive of the number systems, but they are used to build the integers, which in turn are used to build the rationals. Furthermore, the rationals are used to build the real numbers, which are in turn used to build the complex numbers. Thus to begin at the very beginning, we must look at the

© Springer Science+Business Media Singapore 2016 and Hindustan Book Agency 2015

T. Tao, *Analysis I*, Texts and Readings in Mathematics 37,
DOI 10.1007/978-981-10-1789-6_2
Copyrighted material

natural numbers. We will consider the following question: how does one actually *define* the natural numbers? (This is a very different question from how to *use* the natural numbers, which is something you of course know how to do very well. It's like the difference between knowing how to use, say, a computer, versus knowing how to *build* that computer.)

This question is more difficult to answer than it looks. The basic problem is that you have used the natural numbers for so long that they are embedded deeply into your mathematical thinking, and you can make various implicit assumptions about these numbers (e.g., that $a + b$ is always equal to $b + a$) without even aware that you are doing so; it is difficult to let go and try to inspect this number system as if it is the first time you have seen it. So in what follows I will have to ask you to perform a rather difficult task: try to set aside, for the moment, everything you know about the natural numbers; forget that you know how to count, to add, to multiply, to manipulate the rules of algebra, etc. We will try to introduce these concepts one at a time and identify explicitly what our assumptions are as we go along - and not allow ourselves to use more "advanced" tricks such as the rules of algebra until we have actually proven them. This may seem like an irritating constraint, especially as we will spend a lot of time proving statements which are "obvious", but it is necessary to do this suspension of known facts to avoid *circularity* (e.g., using an advanced fact to prove a more elementary fact, and then later using the elementary fact to prove the advanced fact). Also, this exercise will be an excellent way to affirm the foundations of your mathematical knowledge. Furthermore, practicing your proofs and abstract thinking here will be invaluable when we move on to more advanced concepts, such as real numbers, functions, sequences and series, differentials and integrals, and so forth. In short, the results here may seem trivial, but the journey is much more important than the destination, for now. (Once the number systems are constructed properly, we can resume using the laws of algebra etc. without having to rederive them each time.)

We will also forget that we know the decimal system, which of course is an extremely convenient way to manipulate numbers, but it is not something which is fundamental to what numbers are. (For instance, one could use an octal or binary system instead of the decimal system, or even the Roman numeral system, and still get exactly the same set of numbers.) Besides, if one tries to fully explain what the decimal number system is, it isn't as natural as you might think. Why is 00423 the same number as 423, but 32400 isn't the same number as 324? Why

Copyrighted material

is 123.4444... a real number, while ...444.321 is not? And why do we have to carry of digits when adding or multiplying? Why is 0.999... the same number as 1? What is the smallest positive real number? Isn't it just 0.00...001? So to set aside these problems, we will not try to assume any knowledge of the decimal system, though we will of course still refer to numbers by their familiar names such as 1,2,3, etc. instead of using other notation such as I,II,III or 0++, (0++)++, ((0++)++)++ (see below) so as not to be needlessly artificial. For completeness, we review the decimal system in an Appendix (§B).

2.1 The Peano axioms

We now present one standard way to define the natural numbers, in terms of the *Peano axioms*, which were first laid out by Guiseppe Peano (1858–1932). This is not the only way to define the natural numbers. For instance, another approach is to talk about the cardinality of finite sets, for instance one could take a set of five elements and define 5 to be the number of elements in that set. We shall discuss this alternate approach in Section 3.6. However, we shall stick with the Peano axiomatic approach for now.

How are we to define what the natural numbers are? Informally, we could say

Definition 2.1.1. (Informal) A *natural number* is any element of the set

$$\mathbf{N} := \{0, 1, 2, 3, 4, \ldots\},$$

which is the set of all the numbers created by starting with 0 and then counting forward indefinitely. We call $\mathbf{N}$ the *set of natural numbers*.

Remark 2.1.2. In some texts the natural numbers start at 1 instead of 0, but this is a matter of notational convention more than anything else. In this text we shall refer to the set $\{1, 2, 3, \ldots\}$ as the *positive integers* $\mathbf{Z}^+$ rather than the natural numbers. Natural numbers are sometimes also known as *whole numbers*.

In a sense, this definition solves the problem of what the natural numbers are: a natural number is any element of the set[1] $\mathbf{N}$. However,

[1]Strictly speaking, there is another problem with this informal definition: we have not yet defined what a "set" is, or what "element of" is. Thus for the rest of this chapter we shall avoid mention of sets and their elements as much as possible, except in informal discussion.

Copyrighted material

it is not really that satisfactory, because it begs the question of what **N** is. This definition of "start at 0 and count indefinitely" seems like an intuitive enough definition of **N**, but it is not entirely acceptable, because it leaves many questions unanswered. For instance: how do we know we can keep counting indefinitely, without cycling back to 0? Also, how do you perform operations such as addition, multiplication, or exponentiation?

We can answer the latter question first: we can define complicated operations in terms of simpler operations. Exponentiation is nothing more than repeated multiplication: 5^3 is nothing more than three fives multiplied together. Multiplication is nothing more than repeated addition; 5×3 is nothing more than three fives added together. (Subtraction and division will not be covered here, because they are not operations which are well-suited to the natural numbers; they will have to wait for the integers and rationals, respectively.) And addition? It is nothing more than the repeated operation of *counting forward*, or *incrementing*. If you add three to five, what you are doing is incrementing five three times. On the other hand, incrementing seems to be a fundamental operation, not reducible to any simpler operation; indeed, it is the first operation one learns on numbers, even before learning to add.

Thus, to define the natural numbers, we will use two fundamental concepts: the zero number 0, and the increment operation. In deference to modern computer languages, we will use n++ to denote the increment or *successor* of n, thus for instance $3{+}{+} = 4$, $(3{+}{+}){+}{+} = 5$, etc. This is a slightly different usage from that in computer languages such as C, where n++ actually *redefines* the value of n to be its successor; however in mathematics we try not to define a variable more than once in any given setting, as it can often lead to confusion; many of the statements which were true for the old value of the variable can now become false, and vice versa.

So, it seems like we want to say that **N** consists of 0 and everything which can be obtained from 0 by incrementing: **N** should consist of the objects

$$0, 0{+}{+}, (0{+}{+}){+}{+}, ((0{+}{+}){+}{+}){+}{+}, \text{etc.}$$

If we start writing down what this means about the natural numbers, we thus see that we should have the following axioms concerning 0 and the increment operation ++:

Axiom 2.1. *0 is a natural number.*

Copyrighted material

Axiom 2.2. *If n is a natural number, then $n{+}{+}$ is also a natural number.*

Thus for instance, from Axiom 2.1 and two applications of Axiom 2.2, we see that $(0{+}{+}){+}{+}$ is a natural number. Of course, this notation will begin to get unwieldy, so we adopt a convention to write these numbers in more familiar notation:

Definition 2.1.3. We define 1 to be the number $0{+}{+}$, 2 to be the number $(0{+}{+}){+}{+}$, 3 to be the number $((0{+}{+}){+}{+}){+}{+}$, etc. (In other words, $1 := 0{+}{+}$, $2 := 1{+}{+}$, $3 := 2{+}{+}$, etc. In this text I use "$x := y$" to denote the statement that x is *defined* to equal y.)

Thus for instance, we have

Proposition 2.1.4. *3 is a natural number.*

Proof. By Axiom 2.1, 0 is a natural number. By Axiom 2.2, $0{+}{+} = 1$ is a natural number. By Axiom 2.2 again, $1{+}{+} = 2$ is a natural number. By Axiom 2.2 again, $2{+}{+} = 3$ is a natural number. □

It may seem that this is enough to describe the natural numbers. However, we have not pinned down completely the behavior of $\mathbf{N}$:

Example 2.1.5. Consider a number system which consists of the numbers $0, 1, 2, 3$, in which the increment operation wraps back from 3 to 0. More precisely $0{+}{+}$ is equal to 1, $1{+}{+}$ is equal to 2, $2{+}{+}$ is equal to 3, but $3{+}{+}$ is equal to 0 (and also equal to 4, by definition of 4). This type of thing actually happens in real life, when one uses a computer to try to store a natural number: if one starts at 0 and performs the increment operation repeatedly, eventually the computer will overflow its memory and the number will wrap around back to 0 (though this may take quite a large number of incrementation operations, for instance a two-byte representation of an integer will wrap around only after $65,536$ increments). Note that this type of number system obeys Axiom 2.1 and Axiom 2.2, even though it clearly does not correspond to what we intuitively believe the natural numbers to be like.

To prevent this sort of "wrap-around issue" we will impose another axiom:

Axiom 2.3. *0 is not the successor of any natural number; i.e., we have $n{+}{+} \neq 0$ for every natural number n.*

Copyrighted material

Axioms 2.1-2.5 are known as the *Peano axioms* for the natural numbers. They are all very plausible, and so we shall make

Assumption 2.6. (*Informal*) *There exists a number system* $\mathbf{N}$, *whose elements we will call* natural numbers, *for which Axioms 2.1-2.5 are true.*

We will make this assumption a bit more precise once we have laid down our notation for sets and functions in the next chapter.

Remark 2.1.12. We will refer to this number system $\mathbf{N}$ as *the* natural number system. One could of course consider the possibility that there is more than one natural number system, e.g., we could have the Hindu-Arabic number system $\{0, 1, 2, 3, \ldots\}$ and the Roman number system $\{O, I, II, III, IV, V, VI, \ldots\}$, and if we really wanted to be annoying we could view these number systems as different. But these number systems are clearly equivalent (the technical term is *isomorphic*), because one can create a one-to-one correspondence $0 \leftrightarrow O$, $1 \leftrightarrow I$, $2 \leftrightarrow II$, etc. which maps the zero of the Hindu-Arabic system with the zero of the Roman system, and which is preserved by the increment operation (e.g., if 2 corresponds to II, then $2{+}{+}$ will correspond to $II{+}{+}$). For a more precise statement of this type of equivalence, see Exercise 3.5.13. Since all versions of the natural number system are equivalent, there is no point in having distinct natural number systems, and we will just use a single natural number system to do mathematics.

We will not prove Assumption 2.6 (though we will eventually include it in our axioms for set theory, see Axiom 3.7), and it will be the only assumption we will ever make about our numbers. A remarkable accomplishment of modern analysis is that just by starting from these five very primitive axioms, and some additional axioms from set theory, we can build all the other number systems, create functions, and do all the algebra and calculus that we are used to.

Remark 2.1.13. (Informal) One interesting feature about the natural numbers is that while each individual natural number is finite, the *set* of natural numbers is infinite; i.e., $\mathbf{N}$ is infinite but consists of individually finite elements. (The whole is greater than any of its parts.) There are no infinite natural numbers; one can even prove this using Axiom 2.5, provided one is comfortable with the notions of finite and infinite. (Clearly 0 is finite. Also, if n is finite, then clearly $n{+}{+}$ is also finite. Hence by Axiom 2.5, all natural numbers are finite.) So the natural

Copyrighted material

Index

© Springer Science+Business Media Singapore 2016 and Hindustan Book Agency 2015
T. Tao, *Analysis I*, Texts and Readings in Mathematics 37,
DOI 10.1007/978-981-10-1789-6

Copyrighted material

Copyrighted material

Copyrighted material

Copyrighted material

Copyrighted material

Copyrighted material

Copyrighted material