Undergraduate Texts in Mathematics

UTM

Charles C. Pugh

Real Mathematical Analysis

Second Edition

Springer

Undergraduate Texts in Mathematics

Undergraduate Texts in Mathematics

Undergraduate Texts in Mathematics are generally aimed at third- and fourth-year undergraduate mathematics students at North American universities. These texts strive to provide students and teachers with new perspectives and novel approaches. The books include motivation that guides the reader to an appreciation of interrelations among different aspects of the subject. They feature examples that illustrate key concepts as well as exercises that strengthen understanding.

More information about this series at http://www.springer.com/series/666

Charles C. Pugh

Real Mathematical Analysis

Second Edition

Springer

Charles C. Pugh
Department of Mathematics
University of California
Berkeley, CA, USA

ISSN 0172-6056 ISSN 2197-5604 (electronic)
Undergraduate Texts in Mathematics
ISBN 978-3-319-17770-0 ISBN 978-3-319-17771-7 (eBook)
DOI 10.1007/978-3-319-17771-7

Library of Congress Control Number: 2015940438

Mathematics Subject Classification (2010): 26-xx

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

To Candida
and to the students who have encouraged me –
– especially A.W., D.H., and M.B.

Preface

Was plane geometry your favorite math course in high school? Did you like proving theorems? Are you sick of memorizing integrals? If so, real analysis could be your cup of tea. In contrast to calculus and elementary algebra, it involves neither formula manipulation nor applications to other fields of science. None. It is pure mathematics, and I hope it appeals to you, the budding pure mathematician.

This book is set out for college juniors and seniors who love math and who profit from pictures that illustrate the math. Rarely is a picture a proof, but I hope a good picture will cement your understanding of *why* something is true. Seeing is believing.

Chapter 1 gets you off the ground. The whole of analysis is built on the system of real numbers $\mathbb{R}$, and especially on its Least Upper Bound property. Unlike many analysis texts that assume $\mathbb{R}$ and its properties as axioms, Chapter 1 contains a natural construction of $\mathbb{R}$ and a natural proof of the LUB property. You will also see why some infinite sets are more infinite than others, and how to visualize things in four dimensions.

Chapter 2 is about metric spaces, especially subsets of the plane. This chapter contains many pictures you have never seen. ϵ and δ will become your friends. Most of the presentation uses sequences and limits, in contrast to open coverings. It may be less elegant but it's easier to begin with. You will get to know the Cantor set well.

Chapter 3 is about Freshman Calculus – differentiation, integration, L'Hôpital's Rule, and so on, for functions of a single variable – but this time you will find out why what you were taught before is actually true. In particular you will see that a bounded function is integrable if and only if it is continuous almost everywhere, and how this fact explains many other things about integrals.

Chapter 4 is about functions viewed en masse. You can treat a set of functions as a metric space. The "points" in the space aren't numbers or vectors – they are functions. What is the distance between two functions? What should it mean that a sequence of functions converges to a limit function? What happens to derivatives and integrals when your sequence of functions converges to a limit function? When can you approximate a bad function with a good one? What is the best kind of function? What does the typical continuous function look like? (Answer: "horrible.")

Chapter 5 is about Sophomore Calculus – functions of several variables, partial derivatives, multiple integrals, and so on. Again you will see why what you were taught before is actually true. You will revisit Lagrange multipliers (with a picture proof), the Implicit Function Theorem, etc. The main new topic for you will be differential forms. They are presented not as mysterious "multi-indexed expressions," but rather as things that assign numbers to smooth domains. A 1-form assigns to a smooth curve a number, a 2-form assigns to a surface a number, a 3-form assigns to a solid a number, and so on. Orientation (clockwise, counterclockwise, etc.) is important and lets you see why cowlicks are inevitable – the Hairy Ball Theorem. The culmination of the differential forms business is Stokes' Formula, which unifies what you know about div, grad, and curl. It also leads to a short and simple proof of the Brouwer Fixed Point Theorem – a fact usually considered too advanced for undergraduates.

Chapter 6 is about Lebesgue measure and integration. It is not about measure theory in the abstract, but rather about measure theory in the plane, where you can see it. Surely I am not the first person to have rediscovered J.C. Burkill's approach to the Lebesgue integral, but I hope you will come to value it as much as I do. After you understand a few nontrivial things about area in the plane, you are naturally led to define the integral as the area under the curve – the elementary picture you saw in high school calculus. Then the basic theorems of Lebesgue integration simply fall out from the picture. Included in the chapter is the subject of density points – points at which a set "clumps together." I consider density points central to Lebesgue measure theory.

At the end of each chapter are a great many exercises. Intentionally, there is no solution manual. You should expect to be confused and frustrated when you first try to solve the harder problems. Frustration is a good thing. It will strengthen you and it is the natural mental state of most mathematicians most of the time. Join the club! When you do solve a hard problem yourself or with a group of your friends, you will treasure it far more than something you pick up off the web. For encouragement, read Sam Young's story at http://legacyrlmoore.org/reference/young.html.

I have adopted Moe Hirsch's star system for the exercises. One star is hard, two stars is very hard, and a three-star exercise is a question to which I do not know the answer. Likewise, starred sections are more challenging.

Berkeley, California, USA — Charles Chapman Pugh

Contents

1

Real Numbers

1 Preliminaries

Before we discuss the system of real numbers it is best to make a few general remarks about mathematical outlook.

Language

By and large, mathematics is expressed in the language of set theory. Your first order of business is to get familiar with its vocabulary and grammar. A set is a collection of elements. The elements are members of the set and are said to belong to the set. For example, $\mathbb{N}$ denotes the set of **natural numbers**, 1, 2, 3, ... The members of $\mathbb{N}$ are whole numbers greater than or equal to 1. Is 10 a member of $\mathbb{N}$? Yes, 10 belongs to $\mathbb{N}$. Is 0 a member of $\mathbb{N}$? No. We write

$$x \in A \quad \text{and} \quad y \notin B$$

to indicate that the element x is a member of the set A and y is not a member of B. Thus, $6819 \in \mathbb{N}$ and $0 \notin \mathbb{N}$.

We try to write capital letters for sets and small letters for elements of sets. Other standard sets have standard names. The set of **integers** is denoted by $\mathbb{Z}$, which stands for the German word *Zahlen*. (An integer is a positive whole number, zero, or a negative whole number.) Is $\sqrt{2} \in \mathbb{Z}$? No, $\sqrt{2} \notin \mathbb{Z}$. How about -15? Yes, $-15 \in \mathbb{Z}$.

C.C. Pugh, *Real Mathematical Analysis*, Undergraduate Texts in Mathematics, DOI 10.1007/978-3-319-17771-7_1

The set of **rational numbers** is called $\mathbb{Q}$, which stands for "quotient." (A rational number is a fraction of integers, the denominator being nonzero.) Is $\sqrt{2}$ a member of $\mathbb{Q}$? No, $\sqrt{2}$ does not belong to $\mathbb{Q}$. Is π a member of $\mathbb{Q}$? No. Is 1.414 a member of $\mathbb{Q}$? Yes.

You should practice reading the notation "$\{x \in A :$" as "the set of x that belong to A such that." The **empty set** is the collection of no elements and is denoted by $\emptyset$. Is 0 a member of the empty set? No, $0 \notin \emptyset$.

A **singleton set** has exactly one member. It is denoted as $\{x\}$ where x is the member. Similarly if exactly two elements x and y belong to a set, the set is denoted as $\{x, y\}$.

If A and B are sets and each member of A also belongs to B then A is a subset of B and A is contained in B. We write†

$$A \subset B.$$

Is $\mathbb{N}$ a subset of $\mathbb{Z}$? Yes. Is it a subset of $\mathbb{Q}$? Yes. If A is a subset of B and B is a subset of C, does it follow that A is a subset of C? Yes. Is the empty set a subset of $\mathbb{N}$? Yes, $\emptyset \subset \mathbb{N}$. Is 1 a subset of $\mathbb{N}$? No, but the singleton set $\{1\}$ is a subset of $\mathbb{N}$. Two sets are equal if each member of one belongs to the other. Each is a subset of the other. This is how you prove two sets are equal: Show that each element of the first belongs to the second, and each element of the second belongs to the first.

The union of the sets A and B is the set $A \cup B$, each of whose elements belongs to either A, or to B, or to both A and to B. The intersection of A and B is the set $A \cap B$ each of whose elements belongs to both A and to B. If $A \cap B$ is the empty set then A and B are **disjoint**. The **symmetric difference** of A and B is the set $A \Delta B$ each of whose elements belongs to A but not to B, or belongs to B but not to A. The **difference** of A to B is the set $A \setminus B$ whose elements belong to A but not to B. See Figure 1.

A **class** is a collection of sets. The sets are members of the class. For example we could consider the class $\mathcal{E}$ of sets of even natural numbers. Is the set $\{2, 15\}$ a member of $\mathcal{E}$? No. How about the singleton set $\{6\}$? Yes. How about the empty set? Yes, each element of the empty set is even.

When is one class a subclass of another? When each member of the former belongs also to the latter. For example the class $\mathcal{T}$ of sets of positive integers divisible by 10

†When some mathematicians write $A \subset B$ they mean that A is a subset of B, but $A \neq B$. We do *not* adopt this convention. We accept $A \subset A$.

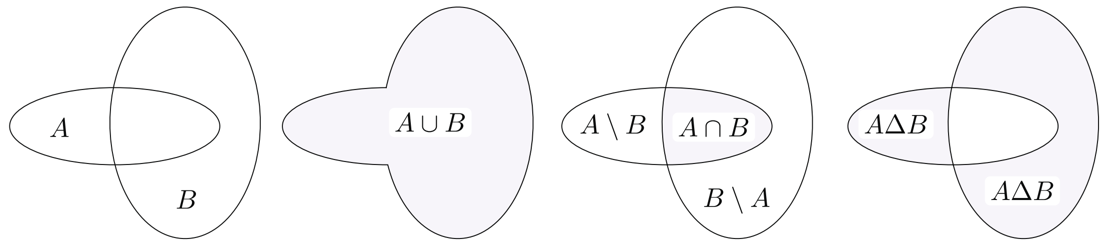

Figure 1 Venn diagrams of union, intersection, and differences

is a subclass of $\mathcal{E}$, the class of sets of even natural numbers, and we write $\mathcal{T} \subset \mathcal{E}$. Each set that belongs to the class $\mathcal{T}$ also belongs to the class $\mathcal{E}$. Consider another example. Let $\mathcal{S}$ be the class of singleton subsets of $\mathbb{N}$ and let $\mathcal{D}$ be the class of subsets of $\mathbb{N}$ each of which has exactly two elements. Thus $\{10\} \in \mathcal{S}$ and $\{2, 6\} \in \mathcal{D}$. Is $\mathcal{S}$ a subclass of $\mathcal{D}$? No. The members of $\mathcal{S}$ are singleton sets and they are not members of $\mathcal{D}$. Rather they are subsets of members of $\mathcal{D}$. Note the distinction, and think about it.

Here is an analogy. Each citizen is a member of his or her country – I am an element of the USA and Tony Blair is an element of the UK. Each country is a member of the United Nations. Are citizens members of the UN? No, countries are members of the UN.

In the same vein is the concept of an **equivalence relation** on a set S. It is a relation $s \sim s'$ that holds between some members $s, s' \in S$ and it satisfies three properties: For all $s, s', s'' \in S$

(a) $s \sim s$.
(b) $s \sim s'$ implies that $s' \sim s$.
(c) $s \sim s' \sim s''$ implies that $s \sim s''$.

Figure 2 on the next page shows how the equivalence relation breaks S into disjoint subsets called **equivalence classes**† defined by mutual equivalence: The equivalence class containing s consists of all elements $s' \in S$ equivalent to s and is denoted $[s]$. The element s is a **representative** of its equivalence class. Think again of citizens and countries. Say two citizens are equivalent if they are citizens of the same country. The world of equivalence relations is egalitarian: I represent my equivalence class USA just as much as does the president.

† The phrase "equivalence class" is standard and widespread, although it would be more consistent with the idea that a class is a collection of sets to refer instead to an "equivalence set."

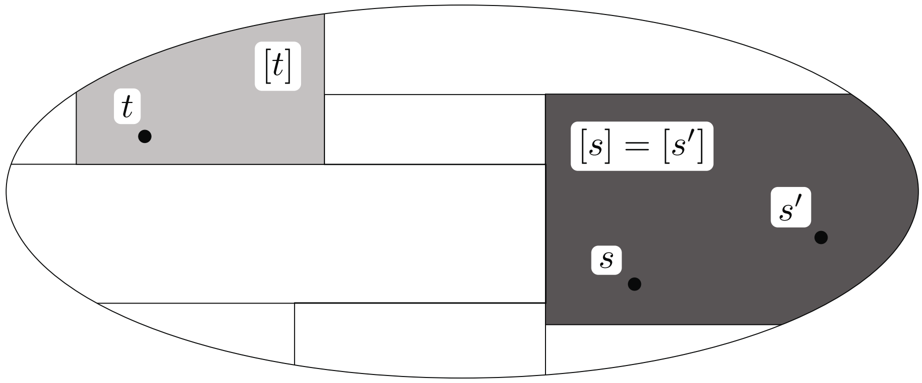

Figure 2 Equivalence classes and representatives

Truth

When is a mathematical statement accepted as true? Generally, mathematicians would answer "Only when it has a proof inside a familiar mathematical framework." A picture may be vital in getting you to believe a statement. An analogy with something you know to be true may help you understand it. An authoritative teacher may force you to parrot it. A formal proof, however, is the ultimate and only reason to accept a mathematical statement as true. A recent debate in Berkeley focused the issue for me. According to a math teacher from one of our local private high schools, his students found proofs in mathematics were of little value, especially compared to "convincing arguments." Besides, the mathematical statements were often seen as obviously true and in no need of formal proof anyway. I offer you a paraphrase of Bob Osserman's response.

> But a convincing argument is not a proof. A mathematician generally wants both, and certainly would be less likely to accept a convincing argument by itself than a formal proof by itself. Least of all would a mathematician accept the proposal that we should generally replace proofs with convincing arguments.
>
> There has been a tendency in recent years to take the notion of proof down from its pedestal. Critics point out that standards of rigor change from century to century. New gray areas appear all the time. Is a proof by computer an acceptable proof? Is a proof that is spread over many journals and thousands of pages, that is too long for any one person to master, a proof? And of course, venerable Euclid is full of flaws, some filled in by Hilbert, others possibly still lurking.

> Clearly it is worth examining closely and critically the most basic notion of mathematics, that of proof. On the other hand, it is important to bear in mind that all distinctions and niceties about what precisely constitutes a proof are mere quibbles compared to the enormous gap between any generally accepted version of a proof and the notion of a convincing argument. Compare Euclid, with all his flaws to the most eminent of the ancient exponents of the convincing argument – Aristotle. Much of Aristotle's reasoning was brilliant, and he certainly convinced most thoughtful people for over a thousand years. In some cases his analyses were exactly right, but in others, such as heavy objects falling faster than light ones, they turned out to be totally wrong. In contrast, there is not to my knowledge a single theorem stated in Euclid's *Elements* that in the course of two thousand years turned out to be false. That is quite an astonishing record, and an extraordinary validation of proof over convincing argument.

Here are some guidelines for writing a rigorous mathematical proof. See also Exercise 0.

1. Name each object that appears in your proof. (For instance, you might begin your proof with a phrase, "Consider a set X, and elements x, y that belong to X," etc.)
2. Draw a diagram that captures how these objects relate, and extract logical statements from it. Quantifiers precede the objects quantified; see below.
3. Become confident that the mathematical assertion you are trying to prove is really true before trying to write down a proof of it. If there a specific function involved – say $\sin x^\alpha$ – draw the graph of the function for a few values of α before starting any ϵ, δ analysis. Belief first and proof second.
4. Proceed step by step, each step depending on the hypotheses, previously proved theorems, or previous steps in your proof.
5. Check for "rigor": All cases have been considered, all details have been tied down, and circular reasoning has been avoided.
6. Before you sign off on the proof, check for counterexamples and any implicit assumptions you made that could invalidate your reasoning.

Logic

Among the most frequently used logical symbols in math are the quantifiers $\forall$ and $\exists$. Read them always as "for each" and "there exists." Avoid reading $\forall$ as "for all," which in English has a more inclusive connotation. Another common symbol is $\Rightarrow$. Read it as "implies."

The rules of correct mathematical grammar are simple: Quantifiers appear at the beginning of a sentence, they modify only what follows them in the sentence, and assertions occur at the end of the sentence. Here is an example.

(1) *For each integer n there is a prime number p which is greater than n.*

In symbols the sentence reads

$$\forall n \in \mathbb{Z} \quad \exists\, p \in P \quad \text{such that} \quad p > n,$$

where P denotes the set of prime numbers. (A **prime number** is a whole number greater than 1 whose only divisors in $\mathbb{N}$ are itself and 1.) In English, the same idea can be reexpressed as

(2) *Every integer is less than some prime number.*

or

(3) *A prime number can always be found which is bigger than any integer.*

These sentences are correct in English grammar, but disastrously WRONG when transcribed directly into mathematical grammar. They translate into disgusting mathematical gibberish:

(WRONG (2)) $\quad \forall n \in \mathbb{Z} \quad n < p \quad \exists\, p \in P$

(WRONG (3)) $\quad \exists\, p \in P \quad p > n \quad \forall n \in \mathbb{Z}.$

Moral Quantifiers first and assertions last. In stating a theorem, try to apply the same principle. Write the hypothesis first and the conclusion second. See Exercise 0.

The order in which quantifiers appear is also important. Contrast the next two sentences in which we switch the position of two quantified phrases.

(4) $\quad (\forall n \in \mathbb{N}) \quad (\forall m \in \mathbb{N}) \quad (\exists\, p \in P) \quad \text{such that} \quad (nm < p).$

(5) $\quad (\forall n \in \mathbb{N}) \quad (\exists\, p \in P) \quad \text{such that} \quad (\forall m \in \mathbb{N}) \quad (nm < p).$

(4) is a true statement but (5) is false. A quantifier modifies the part of a sentence that follows it but not the part that precedes it. This is another reason never to end with a quantifier.

Moral Quantifier order is crucial.

There is a point at which English and mathematical meaning diverge. It concerns the word "or." In mathematics "a or b" always means "a or b or both a and b," while in English it can mean "a or b but not both a and b." For example, Patrick Henry certainly would not have accepted both liberty and death in response to his cry of "Give me liberty or give me death." In mathematics, however, the sentence "17 is a prime or 23 is a prime" is correct even though both 17 and 23 are prime. Similarly, in mathematics $a \Rightarrow b$ means that if a is true then b is true but that b might also be true for reasons entirely unrelated to the truth of a. In English, $a \Rightarrow b$ is often confused with $b \Rightarrow a$.

Moral In mathematics "or" is inclusive. It means *and/or*. In mathematics $a \Rightarrow b$ is not the same as $b \Rightarrow a$.

It is often useful to form the negation or logical opposite of a mathematical sentence. The symbol $\sim$ is usually used for negation, despite the fact that the same symbol also indicates an equivalence relation. Mathematicians refer to this as an **abuse of notation**. Fighting a losing battle against abuse of notation, we write $\neg$ for negation. For example, if $m, n \in \mathbb{N}$ then $\neg(m < n)$ means it is not true that m is less than n. In other words

$$\neg(m < n) \quad \equiv \quad m \geq n.$$

(We use the symbol $\equiv$ to indicate that the two statements are equivalent.) Similarly, $\neg(x \in A)$ means it is not true that x belongs to A. In other words,

$$\neg(x \in A) \quad \equiv \quad x \notin A.$$

Double negation returns a statement to its original meaning. Slightly more interesting is the negation of "and" and "or." Just for now, let us use the symbols & for "and" and $\vee$ for "or." We claim

(6) $$\neg(a \,\&\, b) \quad \equiv \quad \neg a \ \vee \neg b.$$

(7) $$\neg(a \ \vee \ b) \quad \equiv \quad \neg a \,\&\, \neg b.$$

For if it is not the case that both a and b are true then at least one must be false. This proves (6), and (7) is similar. Implication also has such interpretations:

(8) $$a \Rightarrow b \quad \equiv \quad \neg a \Leftarrow \neg b \quad \equiv \quad \neg a \vee b.$$

(9) $$\neg(a \Rightarrow b) \quad \equiv \quad a \;\&\; \neg b.$$

What about the negation of a quantified sentence such as

$$\neg(\forall n \in \mathbb{N},\ \exists\, p \in P \text{ such that } n < p).$$

The rule is: change each $\forall$ to $\exists$ and vice versa, leaving the order the same, and negate the assertion. In this case the negation is

$$\exists\, n \in \mathbb{N}, \quad \forall p \in P, \quad n \geq p.$$

In English it reads "There exists a natural number n, and for all primes p we have $n \geq p$." The sentence has correct mathematical grammar but of course is false. To help translate from mathematics to readable English, a comma can be read as "and," "we have," or "such that."

All mathematical assertions take an implication form $a \Rightarrow b$. The hypothesis is a and the conclusion is b. If you are asked to prove $a \Rightarrow b$, there are several ways to proceed. First you may just see right away why a does imply b. Fine, if you are so lucky. Or you may be puzzled. Does a really imply b? Two routes are open to you. You may view the implication in its equivalent contrapositive form $\neg a \Leftarrow \neg b$ as in (8). Sometimes this will make things clearer. Or you may explore the possibility that a fails to imply b. If you can somehow deduce from the failure of a implying b a contradiction to a known fact (for instance, if you can deduce the existence of a planar right triangle with legs x, y but $x^2 + y^2 \neq h^2$, where h is the hypotenuse), then you have succeeded in making an **argument by contradiction**. Clearly (9) is pertinent here. It tells you what it means that a fails to imply b, namely that a is true and simultaneously b is false.

Euclid's proof that $\mathbb{N}$ contains infinitely many prime numbers is a classic example of this method. The hypothesis is that $\mathbb{N}$ is the set of natural numbers and that P is the set of prime numbers. The conclusion is that P is an infinite set. The proof of this fact begins with the phrase "Suppose not." It means to suppose, after all, that the set of prime numbers P is merely a finite set, and see where this leads you. It does not mean that we think P really is a finite set, and it is not a hypothesis of a theorem. Rather it just means that we will try to find out what awful consequences would follow from P being finite. In fact if P were[†] finite then it would consist of m numbers $p_1, \ldots, p_m$. Their product $N = 2 \cdot 3 \cdot 5 \cdot \cdots \cdot p_m$ would be evenly divisible (i.e., remainder 0 after division) by each p_i and therefore $N + 1$ would be evenly divisible by no prime (the remainder of p_i divided into $N + 1$ would always be 1), which would contradict the fact that every integer ≥ 2 can be factored as a product of primes. (The latter fact has nothing to do with P being finite or not.) Since the supposition that P is finite led to a contradiction of a known fact, prime factorization, the supposition was incorrect, and P is, after all, infinite.

Aficionados of logic will note our heavy use here of the "law of the excluded middle," to wit, that a mathematically meaningful statement is either true or false. The possibilities that it is neither true nor false, or that it is both true and false, are excluded.

Notation The symbol $\lightning$ indicates a contradiction. It is used when writing a proof in longhand.

Metaphor and Analogy

In high school English, you are taught that a metaphor is a figure of speech in which one idea or word is substituted for another to suggest a likeness or similarity. This can occur very simply as in "The ship plows the sea." Or it can be less direct, as in "His lawyers dropped the ball." What give a metaphor its power and pleasure are the secondary suggestions of similarity. Not only did the lawyers make a mistake, but it was their own fault, and, like an athlete who has dropped a ball, they could not follow through with their next legal action. A secondary implication is that their enterprise was just a game.

Often a metaphor associates something abstract to something concrete, as "Life is a journey." The preservation of inference from the concrete to the abstract in this metaphor suggests that like a journey, life has a beginning and an end, it progresses in one direction, it may have stops and detours, ups and downs, etc. The beauty of a metaphor is that hidden in a simple sentence like "Life is a journey" lurk a great many parallels, waiting to be uncovered by the thoughtful mind.

[†]In English grammar, the subjunctive mode indicates doubt, and I have written Euclid's proof in that form – "if P *were* finite" instead of "if P *is* finite," "each prime *would* divide N evenly," instead of "each prime *divides* N evenly," etc. At first it seems like a fine idea to write all arguments by contradiction in the subjunctive mode, clearly exhibiting their impermanence. Soon, however, the subjunctive and conditional language becomes ridiculously stilted and archaic. For consistency then, as much as possible, *use the present tense.*

Metaphorical thinking pervades mathematics to a remarkable degree. It is often reflected in the language mathematicians choose to define new concepts. In his construction of the system of real numbers, Dedekind could have referred to $A|B$ as a "type-2, order preserving equivalence class," or worse, whereas "cut" is the right metaphor. It corresponds closely to one's physical intuition about the real line. See Figure 3. In his book, *Where Mathematics Comes From*, George Lakoff gives a comprehensive view of metaphor in mathematics.

An analogy is a shallow form of metaphor. It just asserts that two things are similar. Although simple, analogies can be a great help in accepting abstract concepts. When you travel from home to school, at first you are closer to home, and then you are closer to school. Somewhere there is a halfway stage in your journey. You *know* this, long before you study mathematics. So when a curve connects two points in a metric space (Chapter 2), you should expect that as a point "travels along the curve," somewhere it will be equidistant between the curve's endpoints. Reasoning by analogy is also referred to as "intuitive reasoning."

Moral Try to translate what you know of the real world to guess what is true in mathematics.

Two Pieces of Advice

A colleague of mine regularly gives his students an excellent piece of advice. When you confront a general problem and do not see how to solve it, make some extra hypotheses, and try to solve it then. If the problem is posed in n dimensions, try it first in two dimensions. If the problem assumes that some function is continuous, does it get easier for a differentiable function? The idea is to reduce an abstract problem to its simplest concrete manifestation, rather like a metaphor in reverse. At the minimum, look for at least one instance in which you can solve the problem, and build from there.

Moral If you do not see how to solve a problem in complete generality, first solve it in some special cases.

Here is the second piece of advice. Buy a notebook. In it keep a diary of your own opinions about the mathematics you are learning. Draw a picture to illustrate every definition, concept, and theorem.

2 Cuts

We begin at the beginning and discuss $\mathbb{R}$ = the system of all real numbers from a somewhat theological point of view. The current mathematics teaching trend treats the real number system $\mathbb{R}$ as a given – it is defined axiomatically. Ten or so of its properties are listed, called axioms of a complete ordered field, and the game becomes to deduce its other properties from the axioms. This is something of a fraud, considering that the entire structure of analysis is built on the real number system. For what if a system satisfying the axioms failed to exist? Then one would be studying the empty set! However, you need not take the existence of the real numbers on faith alone – we will give a concise mathematical proof of it.

It is reasonable to accept all grammar school arithmetic facts about

The set $\mathbb{N}$ of natural numbers, $1, 2, 3, 4, \ldots$.
The set $\mathbb{Z}$ of integers, $0, 1, -1, -2, 2, \ldots$.
The set $\mathbb{Q}$ of rational numbers p/q where p, q are integers, $q \neq 0$.

For example, we will admit without question facts like $2 + 2 = 4$, and laws like $a+b = b+a$ for rational numbers a, b. All facts you know about arithmetic involving integers or rational numbers are fair to use in homework exercises too.† It is clear that $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q}$. Now $\mathbb{Z}$ improves $\mathbb{N}$ because it contains negatives and $\mathbb{Q}$ improves $\mathbb{Z}$ because it contains reciprocals. $\mathbb{Z}$ legalizes subtraction and $\mathbb{Q}$ legalizes division. Still, $\mathbb{Q}$ needs further improvement. It doesn't admit irrational roots such as $\sqrt{2}$ or transcendental numbers such as π. We aim to go a step beyond $\mathbb{Q}$, completing it to form $\mathbb{R}$ so that

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}.$$

As an example of the fact that $\mathbb{Q}$ is incomplete we have

1 Theorem *No number r in $\mathbb{Q}$ has square equal to 2; i.e., $\sqrt{2} \notin \mathbb{Q}$.*

Proof To prove that every $r = p/q$ has $r^2 \neq 2$ we show that $p^2 \neq 2q^2$. It is fair to assume that p and q have no common factors since we would have canceled them out beforehand.

Case 1. p is odd. Then p^2 is odd while $2q^2$ is not. Therefore $p^2 \neq 2q^2$.

† A subtler fact that you may find useful is the prime factorization theorem mentioned above. Any integer ≥ 2 can be factored into a product of prime numbers. For example, 120 is the product of primes $2 \cdot 2 \cdot 2 \cdot 3 \cdot 5$. Prime factorization is unique except for the order in which the factors appear. An easy consequence is that if a prime number p divides an integer k and if k is the product mn of integers then p divides m or it divides n. After all, by uniqueness, the prime factorization of k is just the product of the prime factorizations of m and n.

Case 2. p is even. Since p and q have no common factors, q is odd. Then p^2 is divisible by 4 while $2q^2$ is not. Therefore $p^2 \neq 2q^2$.

Since $p^2 \neq 2q^2$ for all integers p, there is no rational number $r = p/q$ whose square is 2. $\square$

The set $\mathbb{Q}$ of rational numbers is incomplete. It has "gaps," one of which occurs at $\sqrt{2}$. These gaps are really more like pinholes; they have zero width. Incompleteness is what is *wrong* with $\mathbb{Q}$. Our goal is to complete $\mathbb{Q}$ by filling in its gaps. An elegant method to arrive at this goal is **Dedekind cuts** in which one visualizes real numbers as places at which a line may be cut with scissors. See Figure 3.

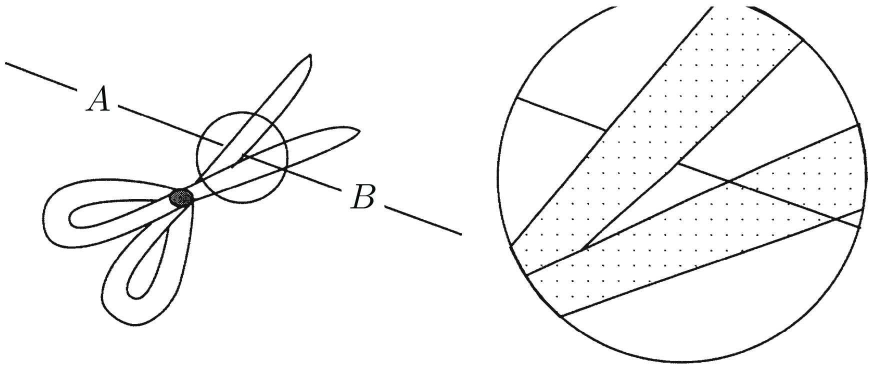

Figure 3 A Dedekind cut

Definition A **cut** in $\mathbb{Q}$ is a pair of subsets A, B of $\mathbb{Q}$ such that

(a) $A \cup B = \mathbb{Q}$, $A \neq \emptyset$, $B \neq \emptyset$, $A \cap B = \emptyset$.
(b) If $a \in A$ and $b \in B$ then $a < b$.
(c) A contains no largest element.

A is the left-hand part of the cut and B is the right-hand part. We denote the cut as $x = A|B$. Making a semantic leap, we now answer the question "what is a real number?"

Definition A **real number** is a cut in $\mathbb{Q}$.

$\mathbb{R}$ is the class† of all real numbers $x = A|B$. We will show that in a natural way $\mathbb{R}$ is a complete ordered field containing $\mathbb{Q}$. Before spelling out what this means, here are two examples of cuts.

†The word "class" is used instead of the word "set" to emphasize that for now the members of $\mathbb{R}$ are set-pairs $A|B$, and not the numbers that belong to A or B. The notation $A|B$ could be shortened to A since B is just the rest of $\mathbb{Q}$. We write $A|B$, however, as a mnemonic device. It *looks* like a cut.

(i) $A|B = \{r \in \mathbb{Q} : r < 1\} \mid \{r \in \mathbb{Q} : r \geq 1\}$.
(ii) $A|B = \{r \in \mathbb{Q} : r \leq 0 \text{ or } r^2 < 2\} \mid \{r \in \mathbb{Q} : r > 0 \text{ and } r^2 \geq 2\}$.

It is convenient to say that $A|B$ is a **rational cut** if it is like the cut in (i): For some fixed rational number c, A is the set of all rationals $< c$ while B is the rest of $\mathbb{Q}$. The B-set of a rational cut contains a smallest element c, and conversely, if $A|B$ is a cut in $\mathbb{Q}$ and B contains a smallest element c then $A|B$ is the rational cut at c. We write c^* for the rational cut at c. This lets us think of $\mathbb{Q} \subset \mathbb{R}$ by identifying c with c^*. It is like thinking of $\mathbb{Z}$ as a subset of $\mathbb{Q}$ since the integer n in $\mathbb{Z}$ can be thought of as the fraction $n/1$ in $\mathbb{Q}$. In the same way the rational number c in $\mathbb{Q}$ can be thought of as the cut at c. It is just a different way of looking at c. It is in this sense that we write

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}.$$

There is an order relation $x \leq y$ on cuts that fairly cries out for attention.

Definition If $x = A|B$ and $y = C|D$ are cuts such that $A \subset C$ then x is **less than or equal** to y and we write $x \leq y$. If $A \subset C$ and $A \neq C$ then x is **less than** y and we write $x < y$.

The property distinguishing $\mathbb{R}$ from $\mathbb{Q}$ and which is at the bottom of every significant theorem about $\mathbb{R}$ involves upper bounds and least upper bounds or, equivalently, lower bounds and greatest lower bounds.

$M \in \mathbb{R}$ is an **upper bound** for a set $S \subset \mathbb{R}$ if each $s \in S$ satisfies

$$s \leq M.$$

We also say that the set S is **bounded above** by M. An upper bound for S that is less than all other upper bounds for S is a **least upper bound** for S. The least upper bound for S is denoted l.u.b.(S). For example,

3 is an upper bound for the set of negative integers.
-1 is the least upper bound for the set of negative integers.
1 is the least upper bound for the set of rational numbers $1 - 1/n$ with $n \in \mathbb{N}$.
-100 is an upper bound for the empty set.

A least upper bound for S may or may not belong to S. This is why you should say "least upper bound *for S*" rather than "least upper bound *of S*."

2 Theorem *The set $\mathbb{R}$, constructed by means of Dedekind cuts, is **complete**[†] in the sense that it satisfies the*

***Least Upper Bound Property**: If S is a nonempty subset of $\mathbb{R}$ and is bounded above then in $\mathbb{R}$ there exists a least upper bound for S.*

Proof Easy! Let $\mathcal{C} \subset \mathbb{R}$ be any nonempty collection of cuts which is bounded above, say by the cut $X|Y$. Define

$$C = \{a \in \mathbb{Q} : \text{ for some cut} A|B \in \mathcal{C} \text{ we have } a \in A\} \text{ and } D = \text{the rest of } \mathbb{Q}.$$

It is easy to see that $z = C|D$ is a cut. Clearly, it is an upper bound for $\mathcal{C}$ since the A for every element of $\mathcal{C}$ is contained in C. Let $z' = C'|D'$ be any upper bound for $\mathcal{C}$. By the assumption that $A|B \leq C'|D'$ for all $A|B \in \mathcal{C}$, we see that the A for every member of $\mathcal{C}$ is contained in C'. Hence $C \subset C'$, so $z \leq z'$. That is, among all upper bounds for $\mathcal{C}$, z is least. $\square$

The simplicity of this proof is what makes cuts good. We go from $\mathbb{Q}$ to $\mathbb{R}$ by pure thought. To be more complete, as it were, we describe the natural arithmetic of cuts. Let cuts $x = A|B$ and $y = C|D$ be given. How do we add them? subtract them? ... Generally the answer is to do the corresponding operation to the elements comprising the two halves of the cuts, being careful about negative numbers. The sum of x and y is $x + y = E|F$ where

$$\begin{aligned} E &= \{r \in \mathbb{Q} : \text{ for some } a \in A \text{ and for some } c \in C \text{ we have } r = a + c\} \\ F &= \text{the rest of } \mathbb{Q}. \end{aligned}$$

It is easy to see that $E|F$ is a cut in $\mathbb{Q}$ and that it doesn't depend on the order in which x and y appear. That is, cut addition is well defined and $x + y = y + x$. The zero cut is 0^* and $0^* + x = x$ for all $x \in \mathbb{R}$. The additive inverse of $x = A|B$ is $-x = C|D$ where

$$\begin{aligned} C &= \{r \in \mathbb{Q} : \text{ for some } b \in B, \text{ not the smallest element of } B, \ r = -b\} \\ D &= \text{the rest of } \mathbb{Q}. \end{aligned}$$

Then $(-x) + x = 0^*$. Correspondingly, the difference of cuts is $x - y = x + (-y)$. Another property of cut addition is **associativity**:

$$(x + y) + z = x + (y + z).$$

[†]There is another, related, sense in which $\mathbb{R}$ is complete. See Theorem 5 below.

This follows from the corresponding property of $\mathbb{Q}$.

Multiplication is trickier to define. It helps to first say that the cut $x = A|B$ is **positive** if $0^* < x$ or **negative** if $x < 0^*$. Since 0 lies in A or B, a cut is either positive, negative, or zero. If $x = A|B$ and $y = C|D$ are positive cuts then their product is $x \cdot y = E|F$ where

$$E = \{r \in \mathbb{Q} : r \leq 0 \text{ or } \exists\, a \in A \text{ and } \exists\, c \in C \text{ such that } a > 0,\, c > 0, \text{ and } r = ac\}$$

and F is the rest of $\mathbb{Q}$. If x is positive and y is negative then we define the product to be $-(x \cdot (-y))$. Since x and $-y$ are both positive cuts this makes sense and is a negative cut. Similarly, if x is negative and y is positive then by definition their product is the negative cut $-((-x) \cdot y)$, while if x and y are both negative then their product is the positive cut $(-x) \cdot (-y)$. Finally, if x or y is the zero cut 0^* we define $x \cdot y$ to be 0^*. (This makes five cases in the definition.)

Verifying the arithmetic properties for multiplication is tedious, to say the least, and somehow nothing seems to be gained by writing out every detail. (To pursue cut arithmetic further you could read Landau's classically boring book, *Foundations of Analysis.*) To get the flavor of it, let's check the commutativity of multiplication: $x \cdot y = y \cdot x$ for cuts $x = A|B$, $y = C|D$. If x, y are positive then

$$\{ac : a \in A,\ c \in C,\ a > 0,\ c > 0\} \ = \ \{ca : c \in C,\ a \in A,\ c > 0,\ a > 0\}$$

implies that $x \cdot y = y \cdot x$. If x is positive and y is negative then

$$x \cdot y \ = \ -(x \cdot (-y)) \ = \ -((-y) \cdot x) \ = \ y \cdot x.$$

The second equality holds because we have already checked commutativity for positive cuts. The remaining three cases are checked similarly. There are twenty seven cases to check for associativity and twenty seven more for distributivity. All are simple and we omit their proofs. The real point is that cut arithmetic can be defined and it satisfies the same field properties that $\mathbb{Q}$ does:

The operation of cut addition is
well defined, natural, commutative, associative, and
has inverses with respect to the neutral element 0^*.

The operation of cut multiplication
is well defined, natural, commutative, associative,
distributive over cut addition, and has inverses of
nonzero elements with respect to the neutral element 1^*.

By definition, a **field** is a system consisting of a set of elements and two operations, addition and multiplication, that have the preceding algebraic properties – commutativity, associativity, etc. Besides just existing, cut arithmetic is consistent with $\mathbb{Q}$ arithmetic in the sense that if $c, r \in \mathbb{Q}$ then

$$c^* + r^* = (c+r)^* \quad \text{and} \quad c^* \cdot r^* = (cr)^*.$$

By definition, this is what we mean when we say that $\mathbb{Q}$ is a **subfield** of $\mathbb{R}$. The cut order enjoys the additional properties of

transitivity $x < y < z$ implies $x < z$.
trichotomy Either $x < y$, $y < x$, or $x = y$, but only one of the three things is true.
translation $x < y$ implies $x + z < y + z$.

By definition, this is what we mean when we say that $\mathbb{R}$ is an **ordered field**. Besides, the product of positive cuts is positive and cut order is consistent with $\mathbb{Q}$ order: $c^* < r^*$ if and only if $c < r$ in $\mathbb{Q}$. By definition, this is what we mean when we say that $\mathbb{Q}$ is an ordered subfield of $\mathbb{R}$. To summarize

3 Theorem *The set $\mathbb{R}$ of all cuts in $\mathbb{Q}$ is a complete ordered field that contains $\mathbb{Q}$ as an ordered subfield.*

The **magnitude** or absolute value of $x \in \mathbb{R}$ is

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0. \end{cases}$$

Thus, $x \leq |x|$. A basic, constantly used fact about magnitude is the following.

4 Triangle Inequality *For all $x, y \in \mathbb{R}$ we have $|x + y| \leq |x| + |y|$.*

Proof The translation and transitivity properties of the order relation imply that adding y and $-y$ to the inequalities $x \leq |x|$ and $-x \leq |x|$ gives

$$\begin{aligned} x + y &\leq |x| + y \leq |x| + |y| \\ -x - y &\leq |x| - y \leq |x| + |y|. \end{aligned}$$

Since

$$|x + y| = \begin{cases} x + y & \text{if } x + y \geq 0 \\ -x - y & \text{if } x + y \leq 0 \end{cases}$$

and both $x + y$ and $-x - y$ are less than or equal to $|x| + |y|$, we infer that $|x + y| \leq |x| + |y|$ as asserted. $\square$

Next, suppose we try the same cut construction in $\mathbb{R}$ that we did in $\mathbb{Q}$. Are there gaps in $\mathbb{R}$ that can be detected by cutting $\mathbb{R}$ with scissors? The natural definition of a cut in $\mathbb{R}$ is a division $\mathcal{A}|\mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are disjoint, nonempty subcollections of $\mathbb{R}$ with $\mathcal{A} \cup \mathcal{B} = \mathbb{R}$, and $a < b$ for all $a \in \mathcal{A}$ and $b \in \mathcal{B}$. Further, $\mathcal{A}$ contains no largest element. Each $b \in \mathcal{B}$ is an upper bound for $\mathcal{A}$. Therefore $y = \text{l.u.b.}(\mathcal{A})$ exists and $a \le y \le b$ for all $a \in \mathcal{A}$ and $b \in \mathcal{B}$. By trichotomy,

$$\mathcal{A}|\mathcal{B} = \{x \in \mathbb{R} : x < y\} \mid \{x \in \mathbb{R} : x \ge y\}.$$

In other words, $\mathbb{R}$ has no gaps. Every cut in $\mathbb{R}$ occurs exactly *at* a real number.

Allied to the existence of $\mathbb{R}$ is its uniqueness. Any complete ordered field $\mathbb{F}$ containing $\mathbb{Q}$ as an ordered subfield corresponds to $\mathbb{R}$ in a way preserving all the ordered field structure. To see this, take any $\varphi \in \mathbb{F}$ and associate to it the cut $A|B$ where

$$A = \{r \in \mathbb{Q} : r < \varphi \text{ in } \mathbb{F}\} \qquad B = \text{ the rest of } \mathbb{Q}.$$

This correspondence makes $\mathbb{F}$ equivalent to $\mathbb{R}$.

Upshot The real number system $\mathbb{R}$ exists and it satisfies the properties of a complete ordered field. The properties are not assumed as axioms, but are proved by logically analyzing the Dedekind construction of $\mathbb{R}$. Having gone through all this cut rigmarole, we must remark that it is a rare working mathematician who actually thinks of $\mathbb{R}$ as a complete ordered field or as the set of all cuts in $\mathbb{Q}$. Rather, he or she thinks of $\mathbb{R}$ as points on the x-axis, just as in calculus. You too should picture $\mathbb{R}$ this way, the only benefit of the cut derivation being that you should now unhesitatingly accept the least upper bound property of $\mathbb{R}$ as a true fact.

Note $\pm\infty$ are not real numbers, since $\mathbb{Q}|\emptyset$ and $\emptyset|\mathbb{Q}$ are not cuts. Although some mathematicians think of $\mathbb{R}$ together with $-\infty$ and $+\infty$ as an "extended real number system," it is simpler to leave well enough alone and just deal with $\mathbb{R}$ itself. Nevertheless, it is convenient to write expressions like "$x \to \infty$" to indicate that a real variable x grows larger and larger without bound.

If S is a nonempty subset of $\mathbb{R}$ then its **supremum** is its least upper bound when S is bounded above and is said to be $+\infty$ otherwise; its **infimum** is its greatest lower bound when S is bounded below and is said to be $-\infty$ otherwise. (In Exercise 19 you are asked to invent the notion of greatest lower bound.) By definition the supremum of the empty set is $-\infty$. This is reasonable, considering that every real number, no matter how negative, is an upper bound for $\emptyset$, and the least upper bound should be as far leftward as possible, namely $-\infty$. Similarly, the infimum of the empty set is $+\infty$. We write $\sup S$ and $\inf S$ for the supremum and infimum of S.

Cauchy sequences

As mentioned above there is a second sense in which $\mathbb{R}$ is complete. It involves the concept of convergent sequences. Let $a_1, a_2, a_3, a_4, \ldots = (a_n),\ n \in \mathbb{N}$, be a sequence of real numbers. The sequence (a_n) **converges to the limit** $b \in \mathbb{R}$ as $n \to \infty$ provided that for each $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n \geq N$ we have

$$|a_n - b| < \epsilon.$$

The statistician's language is evocative here. Think of $n = 1, 2, \ldots$ as a sequence of times and say that the sequence (a_n) converges to b provided that *eventually* all its terms nearly equal b. In symbols,

$$\forall\, \epsilon > 0\ \exists\, N \in \mathbb{N} \text{ such that } n \geq N \ \Rightarrow\ |a_n - b| < \epsilon.$$

If the limit b exists it is not hard to see (Exercise 20) that it is unique, and we write

$$\lim_{n\to\infty} a_n = b \ \text{ or } \ a_n \to b.$$

Suppose that $\lim_{n\to\infty} a_n = b$. Since all the numbers a_n are eventually near b they are all near each other; i.e., every convergent sequence obeys a **Cauchy condition**:

$$\forall\, \epsilon > 0\ \ \exists\, N \in \mathbb{N} \text{ such that if } n, k \geq N \text{ then } |a_n - a_k| < \epsilon.$$

The converse of this fact is a fundamental property of $\mathbb{R}$.

5 Theorem $\mathbb{R}$ *is* ***complete*** *with respect to Cauchy sequences in the sense that if (a_n) is a sequence of real numbers which obeys a Cauchy condition then it converges to a limit in $\mathbb{R}$.*

Proof First we show that (a_n) is bounded. Taking $\epsilon = 1$ in the Cauchy condition implies there is an N such that for all $n, k \geq N$ we have $|a_n - a_k| < 1$. Take K large enough that $-K \ \leq\ a_1, \ldots, a_N \ \leq\ K$. Set $M = K + 1$. Then for all n we have

$$-M \ <\ a_n \ <\ M,$$

which shows that the sequence is bounded.

Define a set X as

$$X = \{x \in \mathbb{R} : \exists \text{ infinitely many } n \text{ such that } a_n \geq x\}.$$

$-M \in X$ since for *all* n we have $a_n > -M$, while $M \notin X$ since *no* x_n is $\geq M$. Thus X is a nonempty subset of $\mathbb{R}$ which is bounded above by M. The least upper bound property applies to X and we have $b = \text{l.u.b.}\, X$ with $-M \leq b \leq M$.

We claim that a_n converges to b as $n \to \infty$. Given $\epsilon > 0$ we must show there is an N such that for all $n \geq N$ we have $|a_n - b| < \epsilon$. Since (a_n) is Cauchy and $\epsilon/2$ is positive there does exist an N such that if $n, k \geq N$ then

$$|a_n - a_k| < \frac{\epsilon}{2}.$$

Since $b - \epsilon/2$ is less than b it is not an upper bound for X, so there is $x \in X$ with $b - \epsilon/2 \leq x$. For infinitely many n we have $a_n \geq x$. Since $b + \epsilon/2 > b$ it does not belong to X, and therefore for only finitely many n do we have $a_n > b + \epsilon/2$. Thus, for infinitely many n we have

$$b - \frac{\epsilon}{2} \;\leq\; x \;\leq\; a_n \;\leq\; b + \frac{\epsilon}{2}.$$

Since there are infinitely many of these n there are infinitely many that are $\geq N$. Pick one, say a_{n_0} with $n_0 \geq N$ and $b - \epsilon/2 \leq a_{n_0} \leq b + \epsilon/2$. Then for *all* $n \geq N$ we have

$$|a_n - b| \;\leq\; |a_n - a_{n_0}| + |a_{n_0} - b| \;<\; \frac{\epsilon}{2} + \frac{\epsilon}{2} \;=\; \epsilon$$

which completes the verification that (a_n) converges. See Figure 4. $\square$

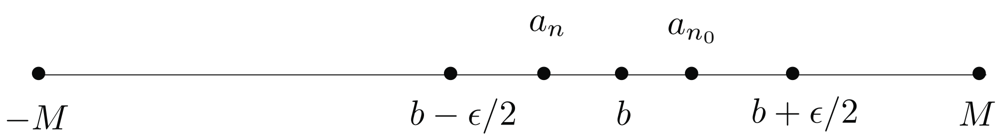

Figure 4 For all $n \geq N$ we have $|a_n - b| < \epsilon$.

Restating Theorem 5 gives the

6 Cauchy Convergence Criterion *A sequence (a_n) in $\mathbb{R}$ converges if and only if*

$$\forall \epsilon > 0 \;\; \exists\, N \in \mathbb{N} \text{ such that } n, k \geq N \;\Rightarrow\; |a_n - a_k| < \epsilon.$$

Further description of $\mathbb{R}$

The elements of $\mathbb{R} \setminus \mathbb{Q}$ are irrational numbers. If x is irrational and r is rational then $y = x + r$ is irrational. For if y is rational then so is $y - r = x$, the difference of rationals being rational. Similarly, if $r \neq 0$ then rx is irrational. It follows that the reciprocal of an irrational number is irrational. From these observations we will show that the rational and irrational numbers are thoroughly mixed up with each other.

Let $a < b$ be given in $\mathbb{R}$. Define the intervals (a, b) and $[a, b]$ as

$$\begin{aligned}(a, b) &= \{x \in \mathbb{R} : a < x < b\}\\ [a, b] &= \{x \in \mathbb{R} : a \leq x \leq b\}.\end{aligned}$$

7 Theorem *Every interval (a, b), no matter how small, contains both rational and irrational numbers. In fact it contains infinitely many rational numbers and infinitely many irrational numbers.*

Proof Think of a, b as cuts $a = A|A'$, $b = B|B'$. The fact that $a < b$ implies the set $B \setminus A$ is a nonempty set of rational numbers. Choose a rational $r \in B \setminus A$. Since B has no largest element, there is a rational s with $a < r < s < b$. Now consider the transformation

$$T : t \mapsto r + (s - r)t.$$

It sends the interval $[0, 1]$ to the interval $[r, s]$. Since r and $s - r$ are rational, T sends rationals to rationals and irrationals to irrationals. Clearly $[0, 1]$ contains infinitely many rationals, say $1/n$ with $n \in \mathbb{N}$, so $[r, s]$ contains infinitely many rationals. Also $[0, 1]$ contains infinitely many irrationals, say $1/n\sqrt{2}$ with $n \in \mathbb{N}$, so $[r, s]$ contains infinitely many irrationals. Since $[r, s]$ contains infinitely many rationals and infinitely many irrationals, the same is true of the larger interval (a, b). □

Theorem 7 expresses the fact that between any two rational numbers lies an irrational number, and between any two irrational numbers lies a rational number. This is a fact worth thinking about for it seems implausible at first. Spend some time trying to picture the situation, especially in light of the following related facts:

(a) There is no first (i.e., smallest) rational number in the interval $(0, 1)$.
(b) There is no first irrational number in the interval $(0, 1)$.
(c) There are strictly more irrational numbers in the interval $(0, 1)$ (in the cardinality sense explained in Section 4) than there are rational numbers.

The transformation in the proof of Theorem 7 shows that the real line is like rubber: stretch it out and it never breaks.

A somewhat obscure and trivial fact about $\mathbb{R}$ is its Archimedean property: for each $x \in \mathbb{R}$ there is an integer n that is greater than x. In other words, there exist arbitrarily large integers. The Archimedean property is true for $\mathbb{Q}$ since $p/q \le |p|$. It follows that it is true for $\mathbb{R}$. Given $x = A|B$, just choose a rational number $r \in B$ and an integer $n > r$. Then $n > x$. An equivalent way to state the Archimedean property is that there exist arbitrarily small reciprocals of integers.

Mildly interesting is the existence of ordered fields for which the Archimedean property fails. One example is the field $\mathbb{R}(x)$ of rational functions with real coefficients. Each such function is of the form

$$R(x) = \frac{p(x)}{q(x)}$$

where p and q are polynomials with real coefficients and q is not the zero polynomial. (It does not matter that $q(x) = 0$ at a finite number of points.) Addition and multiplication are defined in the usual fashion of high school algebra, and it is easy to see that $\mathbb{R}(x)$ is a field. The order relation on $\mathbb{R}(x)$ is also easy to define. If $R(x) > 0$ for all sufficiently large x then we say that R is positive in $\mathbb{R}(x)$, and if $R - S$ is positive then we write $S < R$. Since a nonzero rational function vanishes (has value zero) at only finitely many $x \in \mathbb{R}$, we get trichotomy: either $R = S$, $R < S$, or $S < R$. (To be rigorous, we need to prove that the values of a rational function do not change sign for x large enough.) The other order properties are equally easy to check, and $\mathbb{R}(x)$ is an ordered field.

Is $\mathbb{R}(x)$ Archimedean? That is, given $R \in \mathbb{R}(x)$, does there exist a natural number $n \in \mathbb{R}(x)$ such that $R < n$? (A number n is the rational function whose numerator is the constant polynomial $p(x) = n$, a polynomial of degree zero, and whose denominator is the constant polynomial $q(x) = 1$.) The answer is "no." Take $R(x) = x/1$. The numerator is x and the denominator is 1. Clearly we have $n < x$, not the opposite, so $\mathbb{R}(x)$ fails to be Archimedean.

The same remarks hold for any positive rational function $R = p(x)/q(x)$ where the degree of p exceeds the degree of q. In $\mathbb{R}(x)$, R is never less than a natural number. (You might ask yourself: exactly which rational functions are less than n?)

The ϵ-principle

Finally let us note a nearly trivial principle that turns out to be invaluable in deriving inequalities and equalities in $\mathbb{R}$.

8 Theorem (**ϵ-principle**) *If a, b are real numbers and if for each $\epsilon > 0$ we have $a \leq b + \epsilon$ then $a \leq b$. If x, y are real numbers and for each $\epsilon > 0$ we have $|x - y| \leq \epsilon$ then $x = y$.*

Proof Trichotomy implies that either $a \leq b$ or $a > b$. In the latter case we can choose ϵ with $0 < \epsilon < a - b$ and get the absurdity

$$\epsilon < a - b \leq \epsilon.$$

Hence $a \leq b$. Similarly, if $x \neq y$ then choosing ϵ with $0 < \epsilon < |x - y|$ gives the contradiction $\epsilon < |x - y| \leq \epsilon$. Hence $x = y$. See also Exercise 12. $\square$

3 Euclidean Space

Given sets A and B, the **Cartesian product** of A and B is the set $A \times B$ of all ordered pairs (a, b) such that $a \in A$ and $b \in B$. (The name comes from Descartes who pioneered the idea of the xy-coordinate system in geometry.) See Figure 5.

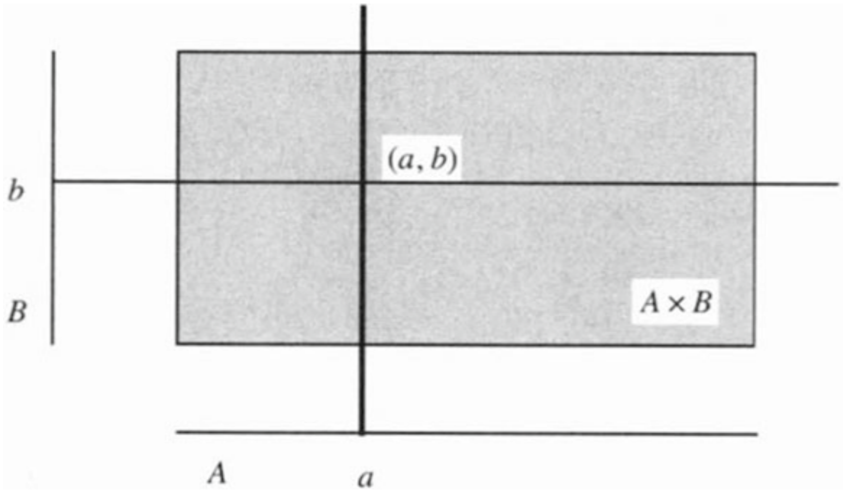

Figure 5 The Cartesian product $A \times B$

The Cartesian product of $\mathbb{R}$ with itself m times is denoted $\mathbb{R}^m$. Elements of $\mathbb{R}^m$ are vectors, ordered m-tuples of real numbers $(x_1, \ldots, x_m)$. In this terminology real numbers are called scalars and $\mathbb{R}$ is called the scalar field. When vectors are added, subtracted, and multiplied by scalars according to the rules

$$\begin{aligned} (x_1, \ldots, x_m) + (y_1, \ldots, y_m) &= (x_1 + y_1, \ldots, x_m + y_m) \\ (x_1, \ldots, x_m) - (y_1, \ldots, y_m) &= (x_1 - y_1, \ldots, x_m - y_m) \\ c(x_1, \ldots, x_m) &= (cx_1, \ldots, cx_m) \end{aligned}$$

then these operations obey the natural laws of linear algebra: commutativity, associativity, etc. There is another operation defined on $\mathbb{R}^m$, the **dot product** (also called the scalar product or inner product). The dot product of $x = (x_1, \ldots, x_m)$ and $y = (y_1, \ldots, y_m)$ is

$$\langle x, y \rangle = x_1 y_1 + \ldots + x_m y_m.$$

Remember: the dot product of two vectors is a scalar, not a vector. The dot product operation is bilinear, symmetric, and positive definite; i.e., for any vectors $x, y, z \in \mathbb{R}^m$

and any $c \in \mathbb{R}$ we have

$$\begin{aligned}\langle x,\, y + cz\rangle &= \langle x, y\rangle + c\langle x, z\rangle \\ \langle x, y\rangle &= \langle y, x\rangle \\ \langle x, x\rangle &\geq 0 \text{ and } \langle x, x\rangle = 0 \text{ if and only if } x \text{ is the zero vector.}\end{aligned}$$

The **length** or **magnitude** of a vector $x \in \mathbb{R}^m$ is defined to be

$$|x| = \sqrt{\langle x, x\rangle} = \sqrt{x_1^2 + \ldots + x_m^2}.$$

See Exercise 16 which legalizes taking roots. Expressed in coordinate-free language, the basic fact about the dot product is the

9 Cauchy-Schwarz Inequality *For all $x, y \in \mathbb{R}^m$ we have $\langle x, y\rangle \leq |x||y|$.*

Proof Tricky! For any vectors x, y consider the new vector $w = x + ty$, where $t \in \mathbb{R}$ is a varying scalar. Then

$$Q(t) = \langle w, w\rangle = \langle x + ty,\ x + ty\rangle$$

is a real-valued function of t. In fact, $Q(t) \geq 0$ since the dot product of any vector with itself is nonnegative. The bilinearity properties of the dot product imply that

$$Q(t) = \langle x, x\rangle + 2t\langle x, y\rangle + t^2\langle y, y\rangle = c + bt + at^2$$

is a quadratic function of t. Nonnegative quadratic functions of $t \in \mathbb{R}$ have nonpositive discriminants, $b^2 - 4ac \leq 0$. For if $b^2 - 4ac > 0$ then $Q(t)$ has two real roots, between which $Q(t)$ is negative. See Figure 6.

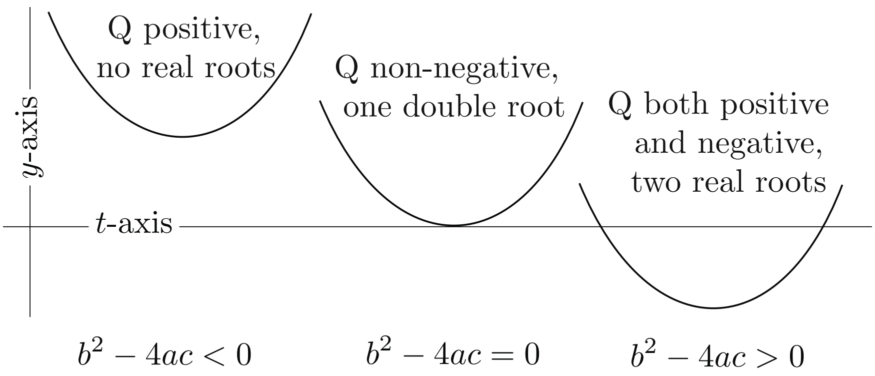

Figure 6 Quadratic graphs

But $b^2 - 4ac \le 0$ means that $4\langle x, y\rangle^2 - 4\langle x, x\rangle\langle y, y\rangle \le 0$, i.e.,

$$\langle x, y\rangle^2 \le \langle x, x\rangle\langle y, y\rangle.$$

Taking the square root of both sides gives $\langle x, y\rangle \le \sqrt{\langle x, x\rangle}\sqrt{\langle y, y\rangle} = |x||y|$. (We use Exercise 17 here and below without further mention.) □

The Cauchy-Schwarz inequality implies easily the **Triangle Inequality for vectors**: For all $x, y \in \mathbb{R}^m$ we have

$$|x + y| \le |x| + |y|.$$

For $|x + y|^2 = \langle x + y,\ x + y\rangle = \langle x, x\rangle + 2\langle x, y\rangle + \langle y, y\rangle$. By Cauchy-Schwarz, $2\langle x, y\rangle \le 2|x||y|$. Thus,

$$|x + y|^2 \le |x|^2 + 2|x||y| + |y|^2 = (|x| + |y|)^2.$$

Taking the square root of both sides gives the result.

The **Euclidean distance** between vectors $x, y \in \mathbb{R}^m$ is defined as the length of their difference,

$$|x - y| = \sqrt{\langle x - y,\ x - y\rangle} = \sqrt{(x_1 - y_1)^2 + \ldots + (x_m - y_m)^2}.$$

From the Triangle Inequality for vectors follows the **Triangle Inequality for distance**. For all $x, y, z \in \mathbb{R}^m$ we have

$$|x - z| \le |x - y| + |y - z|.$$

To prove it, think of $x - z$ as the vector sum $(x - y) + (y - z)$ and apply the Triangle Inequality for vectors. See Figure 7.

Geometric intuition in Euclidean space can carry you a long way in real analysis, especially in being able to forecast whether a given statement is true or not. Your geometric intuition will grow with experience and contemplation. We begin with some vocabulary.

In real analysis, vectors in $\mathbb{R}^m$ are referred to as points in $\mathbb{R}^m$. The j^{th} coordinate of the point $(x_1, \ldots, x_m)$ is the number x_j appearing in the j^{th} position. The j^{th} coordinate axis is the set of points $x \in \mathbb{R}^m$ whose k^{th} coordinates are zero for all $k \ne j$. The origin of $\mathbb{R}^m$ is the zero vector, $(0, \ldots, 0)$. The **first orthant** of $\mathbb{R}^m$ is the set of points $x \in \mathbb{R}^m$ all of whose coordinates are nonnegative. When $m = 2$, the first orthant is the first quadrant. The **integer lattice** is the set $\mathbb{Z}^m \subset \mathbb{R}^m$ of

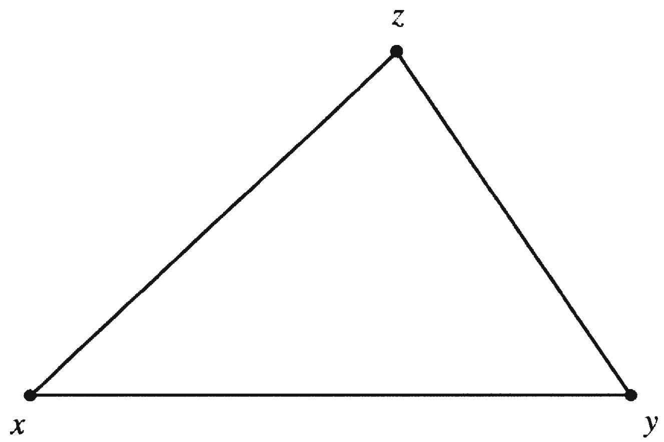

Figure 7 How the Triangle Inequality gets its name

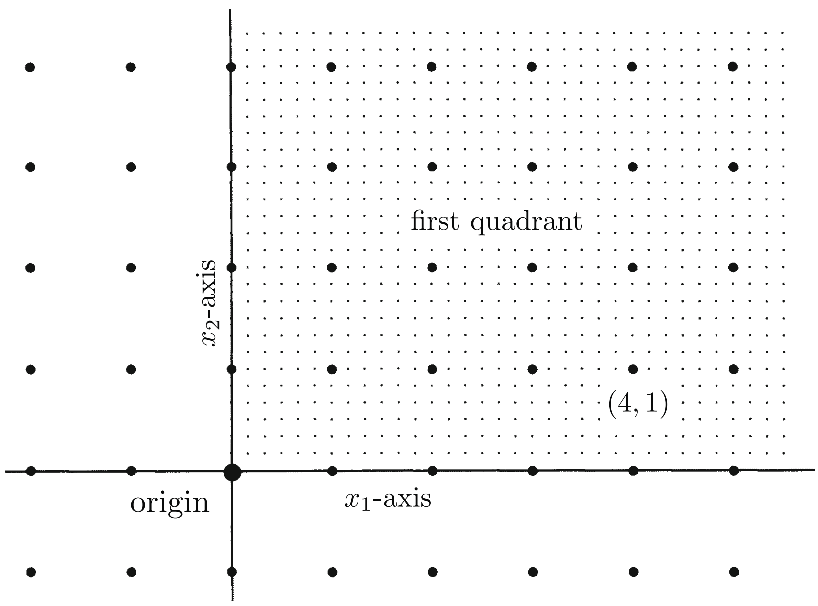

Figure 8 The integer lattice and first quadrant

ordered m-tuples of integers. The integer lattice is also called the **integer grid**. See Figure 8.

A **box** is a Cartesian product of intervals

$$[a_1, b_1] \times \cdots \times [a_m, b_m]$$

in $\mathbb{R}^m$. (A box is also called a **rectangular parallelepiped**.) The **unit cube** in $\mathbb{R}^m$ is the box $[0,1]^m = [0,1] \times \cdots \times [0,1]$. See Figure 9.

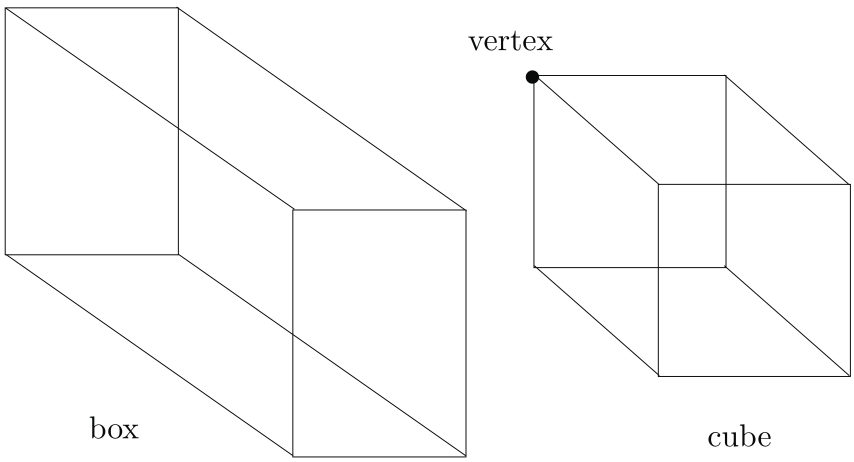

Figure 9 A box and a cube

The **unit ball** and **unit sphere** in $\mathbb{R}^m$ are the sets

$$\begin{aligned} B^m &= \{x \in \mathbb{R}^m : |x| \leq 1\} \\ S^{m-1} &= \{x \in \mathbb{R}^m : |x| = 1\}. \end{aligned}$$

The reason for the exponent $m-1$ is that the sphere is $(m-1)$-dimensional as an object in its own right although it does *live* in m-space. In 3-space, the surface of a ball is a two-dimensional film, the 2-sphere S^2. See Figure 10.

A set $E \subset \mathbb{R}^m$ is **convex** if for each pair of points $x, y \in E$, the straight line segment between x and y is also contained in E. The unit ball is an example of a convex set. To see this, take any two points in B^m and draw the segment between them. It "obviously" lies in B^m. See Figure 11.

To give a mathematical proof, it is useful to describe the line segment between x and y with a formula. The straight line determined by distinct points $x, y \in \mathbb{R}^m$ is the set of all linear combinations $sx + ty$ where $s + t = 1$, and the line segment is the set of these linear combinations where s and t are ≤ 1. Such linear combinations

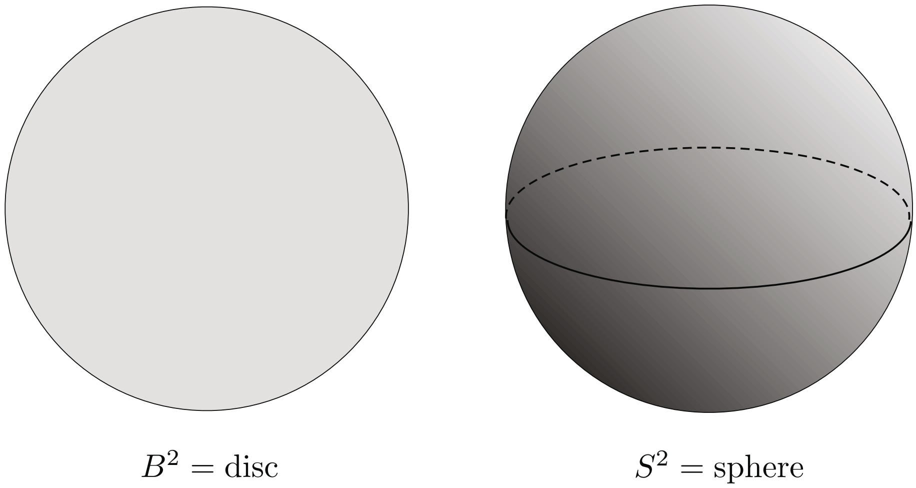

Figure 10 A 2-disc B^2 with its boundary circle, and a 2-sphere S^2 with its equator

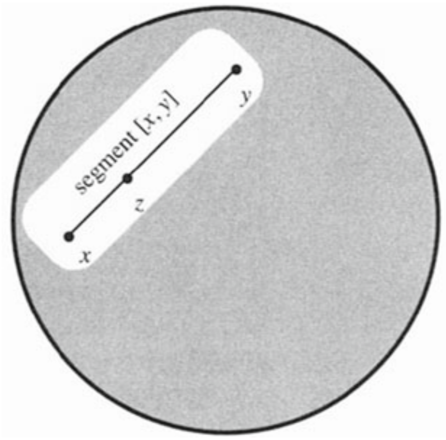

Figure 11 Convexity of the ball

$sx + ty$ with $s + t = 1$ and $0 \le s, t \le 1$ are called **convex combinations**. The line segment is denoted as $[x, y]$. (This notation is consistent with the interval notation $[a, b]$. See Exercise 27.) Now if $x, y \in B^m$ and $sx + ty = z$ is a convex combination of x and y then, using the Cauchy-Schwarz Inequality and the fact that $2st \ge 0$, we get

$$\begin{aligned} \langle z, z\rangle &= s^2\langle x, x\rangle + 2st\langle x, y\rangle + t^2\langle y, y\rangle \\ &\le s^2|x|^2 + 2st|x||y| + t^2|y|^2 \\ &\le s^2 + 2st + t^2 = (s+t)^2 = 1. \end{aligned}$$

Taking the square root of both sides gives $|z| \le 1$, which proves convexity of the ball.

Inner product spaces

An **inner product** on a vector space V is an operation $\langle\ ,\ \rangle$ on pairs of vectors in V that satisfies the same conditions that the dot product in Euclidean space does: Namely, bilinearity, symmetry, and positive definiteness. A vector space equipped with an inner product is an **inner product space**. The discriminant proof of the Cauchy-Schwarz Inequality is valid for any inner product defined on any real vector space, even if the space is infinite-dimensional and the standard coordinate proof would make no sense. For the discriminant proof uses only the inner product properties, and not the particular definition of the dot product in Euclidean space.

$\mathbb{R}^m$ has dimension m because it has a basis $e_1, \ldots, e_m$. Other vector spaces are more general. For example, let $C([a,b],\mathbb{R})$ denote the set of all of continuous real-valued functions defined on the interval $[a,b]$. (See Section 6 or your old calculus book for the definition of continuity.) It is a vector space in a natural way, the sum of continuous functions being continuous and the scalar multiple of a continuous function being continuous. The vector space $C([a,b],\mathbb{R})$, however, has no finite basis. It is infinite-dimensional. Even so, there is a natural inner product,

$$\langle f, g\rangle = \int_a^b f(x)g(x)\,dx.$$

Cauchy-Schwarz applies to this inner product, just as to any inner product, and we infer a general integral inequality valid for any two continuous functions,

$$\int_a^b f(x)g(x)\,dx \quad \leq \quad \sqrt{\int_a^b f(x)^2\,dx}\,\sqrt{\int_a^b g(x)^2\,dx}.$$

It would be challenging to prove such an inequality from scratch, would it not? See also the first paragraph of the next chapter.

A **norm** on a vector space V is any function $|\quad| : V \to \mathbb{R}$ with the three properties of vector length: Namely, if $v, w \in V$ and $\lambda \in \mathbb{R}$ then

$$\begin{aligned} &|v| \geq 0 \text{ and } |v| = 0 \text{ if and only if } v = 0, \\ &|\lambda v| = |\lambda|\,|v|, \\ &|v + w| \leq |v| + |w|. \end{aligned}$$

An inner product $\langle\ ,\ \rangle$ defines a norm as $|v| = \sqrt{\langle v, v\rangle}$, but not all norms come from inner products. The unit sphere $\{v \in V : \langle v, v\rangle = 1\}$ for every inner product is smooth (has no corners) while for the norm

$$|v|_{\max} = \max\{|v_1|, |v_2|\}$$

defined on $v = (v_1, v_2) \in \mathbb{R}^2$, the unit sphere is the perimeter of the square $\{(v_1, v_2) \in \mathbb{R}^2 : |v_1| \leq 1 \text{ and } |v_2| \leq 1\}$. It has corners and so it does not arise from an inner product. See Exercises 46, 47, and the Manhattan metric on page 76.

The simplest Euclidean space beyond $\mathbb{R}$ is the plane $\mathbb{R}^2$. Its xy-coordinates can be used to define a multiplication,

$$(x, y) \bullet (x', y') = (xx' - yy',\ xy' + x'y).$$

The point $(1, 0)$ corresponds to the multiplicative unit element 1, while the point $(0, 1)$ corresponds to $i = \sqrt{-1}$, which converts the plane to the field $\mathbb{C}$ of complex numbers. Complex analysis is the study of functions of a complex variable, i.e., functions $f(z)$ where z and $f(z)$ lie in $\mathbb{C}$. Complex analysis is the good twin and real analysis the evil one: beautiful formulas and elegant theorems seem to blossom spontaneously in the complex domain, while toil and pathology rule the reals. Nevertheless, complex analysis relies more on real analysis than the other way around.

4 Cardinality

Let A and B be sets. A **function** $f : A \to B$ is a rule or mechanism which, when presented with any element $a \in A$, produces an element $b = f(a)$ of B. It need not be defined by a formula. Think of a function as a device into which you feed elements of A and out of which pour elements of B. See Figure 12. We also call f a **mapping**

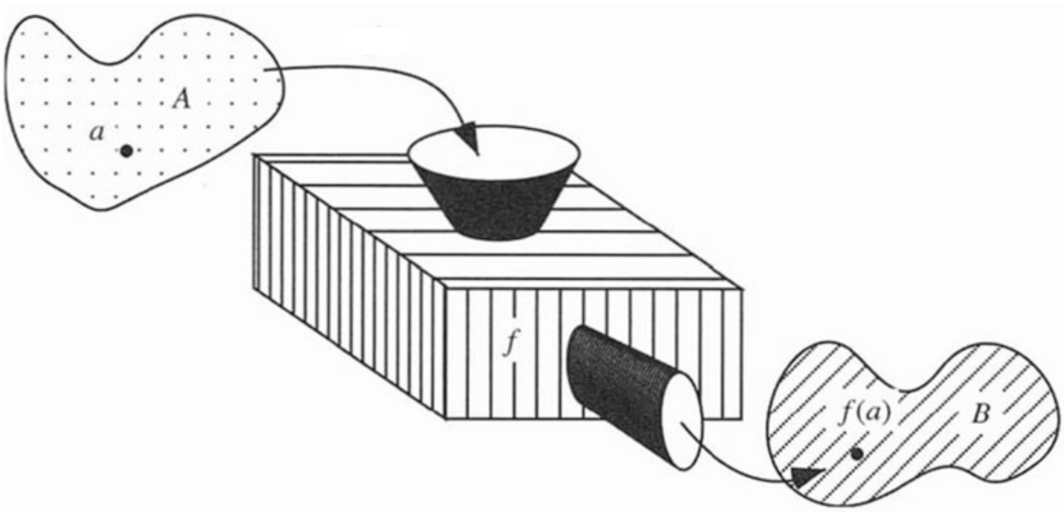

Figure 12 The function f as a machine

or a **map** or a **transformation**. The set A is the **domain** of the function and B is its **target**, also called its **codomain**. The **range** or **image** of f is the subset of the target

$$\{b \in B : \text{ there exists at least one element } a \in A \text{ with } f(a) = b\}.$$

See Figure 13.

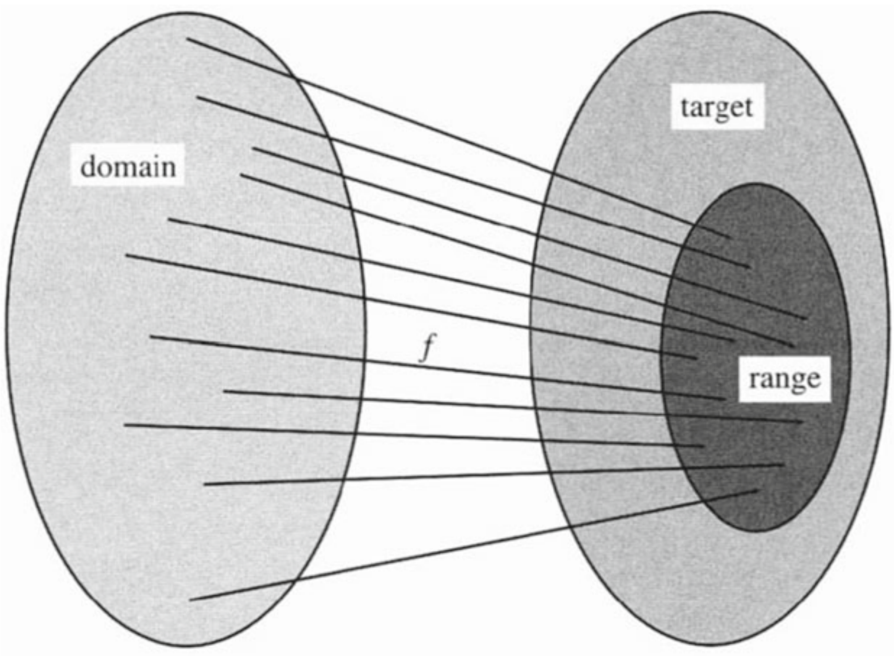

Figure 13 The domain, target, and range of a function

Try to write f instead of $f(x)$ to denote a function. The function is the device which when confronted with input x produces output $f(x)$. The function is the device, not the output.

Think also of a function dynamically. At time zero all the elements of A are sitting peacefully in A. Then the function applies itself to them and throws them into B. At time one all the elements that were formerly in A are now transferred into B. Each $a \in A$ gets sent to some element $f(a) \in B$.

A mapping $f : A \to B$ is an **injection** (or is **one-to-one**) if for each pair of distinct elements $a, a' \in A$, the elements $f(a), f(a')$ are distinct in B. That is,

$$a \neq a' \quad \Rightarrow \quad f(a) \neq f(a').$$

The mapping f is a **surjection** (or is **onto**) if for each $b \in B$ there is at least one $a \in A$ such that $f(a) = b$. That is, the range of f is B.

A mapping is a **bijection** if it is both injective and surjective. It is one-to-one and onto. If $f : A \to B$ is a bijection then the inverse map $f^{-1} : B \to A$ is a bijection where $f^{-1}(b)$ is by definition the unique element $a \in A$ such that $f(a) = b$.

The **identity map** of any set to itself is the bijection that takes each $a \in A$ and sends it to itself, $\mathrm{id}(a) = a$.

If $f : A \to B$ and $g : B \to C$ then the **composite** $g \circ f : A \to C$ is the function that sends $a \in A$ to $g(f(a)) \in C$. If f and g are injective then so is $g \circ f$, while if f and g are surjective then so is $g \circ f$,

$$\begin{tikzcd}
& B \arrow[dr, "g"] & \\
A \arrow[ur, "f"] \arrow[rr, "g \circ f"'] & & C.
\end{tikzcd}$$

In particular the composite of bijections is a bijection. If there is a bijection from A onto B then A and B are said to have **equal cardinality**,[†] and we write $A \sim B$. The relation $\sim$ is an equivalence relation. That is,

(a) $A \sim A$.
(b) $A \sim B$ implies $B \sim A$.
(c) $A \sim B \sim C$ implies $A \sim C$.

(a) follows from the fact that the identity map bijects A to itself. (b) follows from the fact that the inverse of a bijection $A \to B$ is a bijection $B \to A$. (c) follows from the fact that the composite of bijections f and g is a bijection $g \circ f$.

A set S is

finite if it is empty or for some $n \in \mathbb{N}$ we have $S \sim \{1, \ldots, n\}$.
infinite if it is not finite.
denumerable if $S \sim \mathbb{N}$.
countable if it is finite or denumerable.
uncountable if it is not countable.

[†] The word "cardinal" indicates the number of elements in the set. The cardinal numbers are $0, 1, 2, \ldots$ The first infinite cardinal number is **aleph null**, $\aleph_0$. One says the $\mathbb{N}$ has $\aleph_0$ elements. A mystery of math is the **Continuum Hypothesis** which states that $\mathbb{R}$ has cardinality $\aleph_1$, the second infinite cardinal. Equivalently, if $\mathbb{N} \subset S \subset \mathbb{R}$, the Continuum Hypothesis asserts that $S \sim \mathbb{N}$ or $S \sim \mathbb{R}$. No intermediate cardinalities exist. You can pursue this issue in Paul Cohen's book, *Set Theory and the Continuum Hypothesis*.

We also write $\operatorname{card} A = \operatorname{card} B$ and $\#A = \#B$ when A, B have equal cardinality.

If S is denumerable then there is a bijection $f : \mathbb{N} \to S$, and this gives a way to list the elements of S as $s_1 = f(1)$, $s_2 = f(2)$, $s_3 = f(3)$, etc. Conversely, if a set S is presented as an infinite list (without repetition) $S = \{s_1, s_2, s_3, \ldots\}$, then it is denumerable: Define $f(k) = s_k$ for all $k \in \mathbb{N}$. In brief, denumerable = listable.

Let's begin with a truly remarkable cardinality result, that although $\mathbb{N}$ and $\mathbb{R}$ are both infinite, $\mathbb{R}$ is more infinite than $\mathbb{N}$. Namely,

10 Theorem $\mathbb{R}$ *is uncountable.*

Proof There are other proofs of the uncountability of $\mathbb{R}$, but none so beautiful as this one. It is due to Cantor. I assume that you accept the fact that each real number x has a decimal expansion, $x = N.x_1x_2x_3\ldots$, and it is uniquely determined by x if one agrees never to terminate the expansion with an infinite string of 9s. (See also Exercise 18.) We want to prove that $\mathbb{R}$ is uncountable. Suppose it is not uncountable. Then it is countable and, being infinite, it must be denumerable. Accordingly let $f : \mathbb{N} \to \mathbb{R}$ be a bijection. Using f, we list the elements of $\mathbb{R}$ along with their decimal expansions as an array, and consider the digits x_{ii} that occur along the diagonal in this array. See Figure 14.

$$
\begin{array}{ccccccccccc}
f(1) & = & N_1 & x_{11} & x_{12} & x_{13} & x_{14} & x_{15} & x_{16} & x_{17} & \\
f(2) & = & N_2 & x_{21} & x_{22} & x_{23} & x_{24} & x_{25} & x_{26} & x_{27} & \\
f(3) & = & N_3 & x_{31} & x_{32} & x_{33} & x_{34} & x_{35} & x_{36} & x_{37} & \\
f(4) & = & N_4 & x_{41} & x_{42} & x_{43} & x_{44} & x_{45} & x_{46} & x_{47} & \\
f(5) & = & N_5 & x_{51} & x_{52} & x_{53} & x_{54} & x_{55} & x_{56} & x_{57} & \\
f(6) & = & N_6 & x_{61} & x_{62} & x_{63} & x_{64} & x_{65} & x_{66} & x_{67} & \\
f(7) & = & N_7 & x_{71} & x_{72} & x_{73} & x_{74} & x_{75} & x_{76} & x_{77} & \\
\vdots & & & & & & & & & & \ddots
\end{array}
$$

Figure 14 Cantor's diagonal method

For each i, choose a digit y_i such that $y_i \neq x_{ii}$ and $y_i \neq 9$. Where is the number $y = 0.y_1y_2y_3\ldots$? Is it $f(1)$? No, because the first digit in the decimal expansion of

$f(1)$ is x_{11} and $y_1 \neq x_{11}$. Is it $f(2)$? No, because the second digit in the decimal expansion of $f(2)$ is x_{22} and $y_2 \neq x_{22}$. Is it $f(k)$? No, because the k^{th} digit in the decimal expansion of $f(k)$ is x_{kk} and $y_k \neq x_{kk}$. Nowhere in the list do we find y. Nowhere! Therefore the list could not account for every real number, and $\mathbb{R}$ must have been uncountable. □

11 Corollary *$[a, b]$ and (a, b) are uncountable.*

Proof There are bijections from (a, b) onto $(-1, 1)$ onto the unit semicircle onto $\mathbb{R}$ shown in Figure 15. The composite f bijects (a, b) onto $\mathbb{R}$, so (a, b) is uncountable.

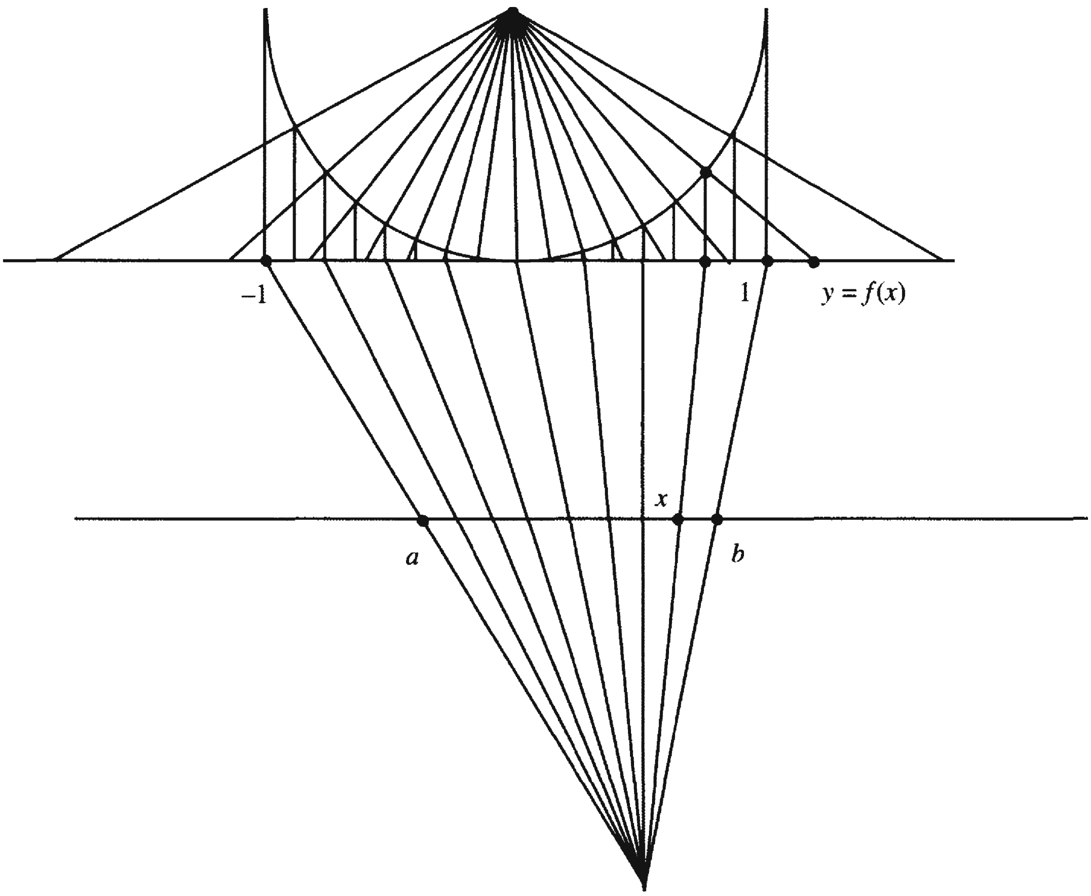

Figure 15 Equicardinality of (a, b), $(-1, 1)$, and $\mathbb{R}$

Since $[a, b]$ contains (a, b), it too is uncountable. □

The remaining results in this section are of a more positive flavor.

12 Theorem *Each infinite set S contains a denumerable subset.*

Proof Since S is infinite it is nonempty and contains an element s_1. Since S is infinite the set $S \setminus \{s_1\} = \{s \in S : s \neq s_1\}$ is nonempty and there exists $s_2 \in S \setminus \{s_1\}$. Since S is an infinite set, $S \setminus \{s_1, s_2\} = \{s \in S : s \neq s_1, s_2\}$ is nonempty and there exists $s_3 \in S \setminus \{s_1, s_2\}$. Continuing this way gives a list (s_n) of distinct elements of S. The set of these elements forms a denumerable subset of S. □

13 Theorem *An infinite subset A of a denumerable set B is denumerable.*

Proof There exists a bijection $f : \mathbb{N} \to B$. Each element of A appears exactly once in the list $f(1), f(2), f(3), \ldots$ of B. Define $g(k)$ to be the k^{th} element of A appearing in the list. Since A is infinite, $g(k)$ is defined for all $k \in \mathbb{N}$. Thus $g : \mathbb{N} \to A$ is a bijection and A is denumerable. □

14 Corollary *The sets of even integers and of prime integers are denumerable.*

Proof They are infinite subsets of $\mathbb{N}$ which is denumerable. □

15 Theorem $\mathbb{N} \times \mathbb{N}$ *is denumerable.*

Proof Think of $\mathbb{N} \times \mathbb{N}$ as an $\infty \times \infty$ matrix and walk along the successive counter-diagonals. See Figure 16. This gives a list

$$(1,1), (2,1), (1,2), (3,1), (2,2), (1,3), (4,1), (3,2), (2,3), (1,4), (5,1), \ldots$$

of $\mathbb{N} \times \mathbb{N}$ and proves that $\mathbb{N} \times \mathbb{N}$ is denumerable. □

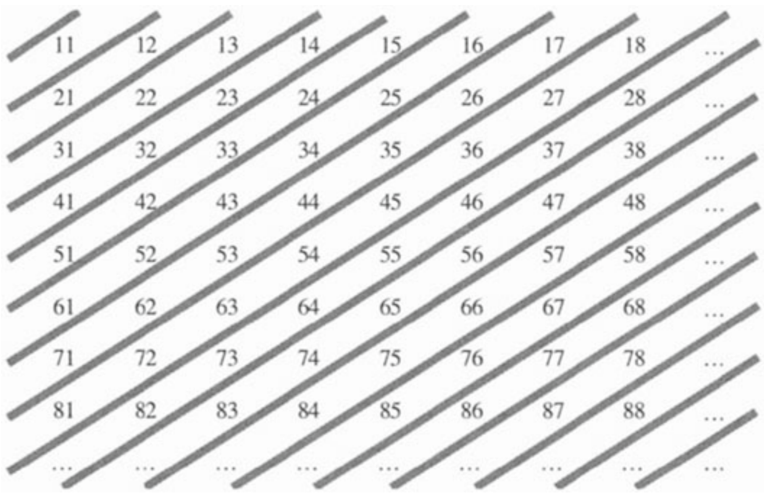

Figure 16 Counter-diagonals in an $\infty \times \infty$ matrix

16 Corollary *The Cartesian product of denumerable sets A and B is denumerable.*

Proof $\mathbb{N} \sim \mathbb{N} \times \mathbb{N} \sim A \times B$. □

17 Theorem *If $f : \mathbb{N} \to B$ is a surjection and B is infinite then B is denumerable.*

Proof For each $b \in B$, the set $\{k \in \mathbb{N} : f(k) = b\}$ is nonempty and hence contains a smallest element; say $h(b) = k$ is the smallest integer that is sent to b by f. Clearly, if $b, b' \in B$ and $b \neq b'$ then $h(b) \neq h(b')$. That is, $h : B \to \mathbb{N}$ is an injection which bijects B to $hB \subset \mathbb{N}$. Since B is infinite, so is hB. By Theorem 13, hB is denumerable and therefore so is B. □

18 Corollary *The denumerable union of denumerable sets is denumerable.*

Proof Suppose that $A_1, A_2, \ldots$ is a sequence of denumerable sets. List the elements of A_i as $a_{i1}, a_{i2}, \ldots$ and define

$$\begin{aligned} f : \mathbb{N} \times \mathbb{N} &\to A = \bigcup A_i \\ (i, j) &\mapsto a_{ij} \end{aligned}$$

Clearly f is a surjection. According to Theorem 15, there is a bijection $g : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$. The composite $f \circ g$ is a surjection $\mathbb{N} \to A$. Since A is infinite, Theorem 17 implies it is denumerable. □

19 Corollary $\mathbb{Q}$ *is denumerable.*

Proof $\mathbb{Q}$ is the denumerable union of the denumerable sets $A_q = \{p/q : p \in \mathbb{Z}\}$ as q ranges over $\mathbb{N}$. □

20 Corollary *For each $m \in \mathbb{N}$ the set $\mathbb{Q}^m$ is denumerable.*

Proof Apply the induction principle. If $m = 1$ then the previous corollary states that $\mathbb{Q}^1$ is denumerable. Knowing inductively that $\mathbb{Q}^{m-1}$ is denumerable and $\mathbb{Q}^m = \mathbb{Q}^{m-1} \times \mathbb{Q}$, the result follows from Corollary 16. □

Combination laws for countable sets are similar to those for denumerable sets. As is easily checked,

> Every subset of a countable set is countable.
> A countable set that contains a denumerable subset is denumerable.
> The Cartesian product of finitely many countable sets is countable.
> The countable union of countable sets is countable.

5* Comparing Cardinalities

The following result gives a way to conclude that two sets have the same cardinality. Roughly speaking the condition is that $\operatorname{card} A \leq \operatorname{card} B$ and $\operatorname{card} B \leq \operatorname{card} A$.

21 Schroeder-Bernstein Theorem *If A, B are sets and $f : A \to B$, $g : B \to A$ are injections then there exists a bijection $h : A \to B$.*

Proof-sketch Consider the dynamic Venn diagram, Figure 17. The disc labeled gfA

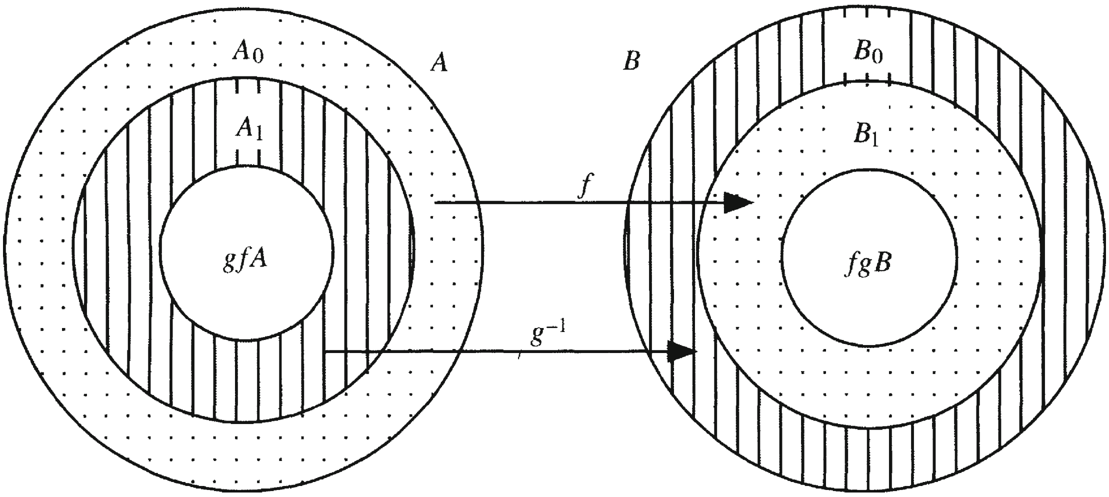

Figure 17 Pictorial proof of the Schroeder-Bernstein Theorem

is the image of A under the map $g \circ f$. It is a subset of A. The ring between A and gfA divides into two subrings. A_0 is the set of points in A that do not lie in the image of g, while A_1 is the set points in the image of g that do not lie in gfA. Similarly, B_0 is the set of points in B that do not lie in fA, while B_1 is the set of points in fA that do not lie in fgB. There is a natural bijection h from the pair of rings $A_0 \cup A_1 = A \setminus gfA$ to the pair of rings $B_0 \cup B_1 = B \setminus fgB$. It equals f on the outer ring $A_0 = A \setminus gB$ and it is g^{-1} on the inner ring $A_1 = gB \setminus gfA$. (The map g^{-1} is not defined on all of A, but it is defined on the set gB.) In this notation, h sends A_0 onto B_1 and sends A_1 onto B_0. It switches the indices. Repeat this on the next pair of rings for A and B. That is, look at gfA instead of A and fgB instead of B. The next two rings in A, B are

$$\begin{aligned} A_2 &= gfA \setminus gfgB & \quad A_3 &= gfgB \setminus gfgfA \\ B_2 &= fgB \setminus fgfA & \quad B_3 &= fgfA \setminus fgfgB. \end{aligned}$$

Send A_2 to B_3 by f and A_3 to B_2 by g^{-1}. The rings A_i are disjoint, and so are the rings B_i, so repetition gives a bijection

$$\phi : \bigsqcup A_i \quad \to \quad \bigsqcup B_i,$$

($\bigsqcup$ indicates disjoint union) defined by

$$\phi(x) = \begin{cases} f(x) & \text{if } x \in A_i \text{ and } i \text{ is even} \\ g^{-1}(x) & \text{if } x \in A_i \text{ and } i \text{ is odd.} \end{cases}$$

Let $A_* = A \setminus (\bigcup A_i)$ and $B_* = B \setminus (\bigcup B_i)$ be the rest of A and B. Then f bijects A_* to B_* and ϕ extends to a bijection $h : A \to B$ defined by

$$h(x) = \begin{cases} \phi(x) & \text{if } x \in \bigsqcup A_i \\ f(x) & \text{if } x \in A_*. \end{cases}$$

□

A supplementary aid in understanding the Schroeder Bernstein proof is the following crossed ladder diagram, Figure 18.

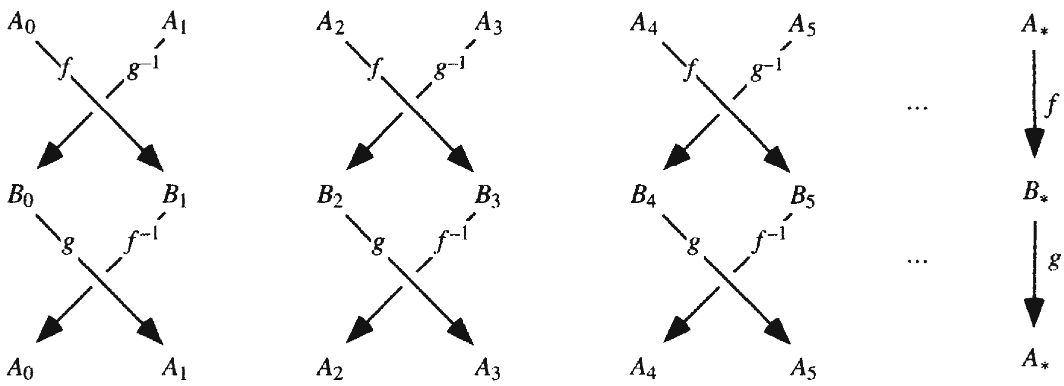

Figure 18 Diagramatic proof of the Schroeder-Bernstein Theorem

Exercise 36 asks you to show directly that $(a, b) \sim [a, b]$. This makes sense since $(a, b) \subset [a, b] \subset \mathbb{R}$ and $(a, b) \sim \mathbb{R}$ should certainly imply $(a, b) \sim [a, b] \sim \mathbb{R}$. The Schroeder-Bernstein theorem gives a quick indirect solution to the exercise. The inclusion map $i : (a, b) \hookrightarrow [a, b]$ sending x to x injects (a, b) into $[a, b]$, while the function $j(x) = x/2 + (a + b)/4$ injects $[a, b]$ into (a, b). The existence of the two injections implies by the Schroeder-Bernstein Theorem that there is a bijection $(a, b) \sim [a, b]$.

6* The Skeleton of Calculus

The behavior of a continuous function defined on an interval $[a, b]$ is at the root of all calculus theory. Using solely the Least Upper Bound Property of the real numbers we rigorously derive the basic properties of such functions. The function $f : [a, b] \to \mathbb{R}$ is **continuous** if for each $\epsilon > 0$ and each $x \in [a, b]$ there is a $\delta > 0$ such that

$$t \in [a, b] \text{ and } |t - x| < \delta \quad \Rightarrow \quad |f(t) - f(x)| < \epsilon.$$

See Figure 19.

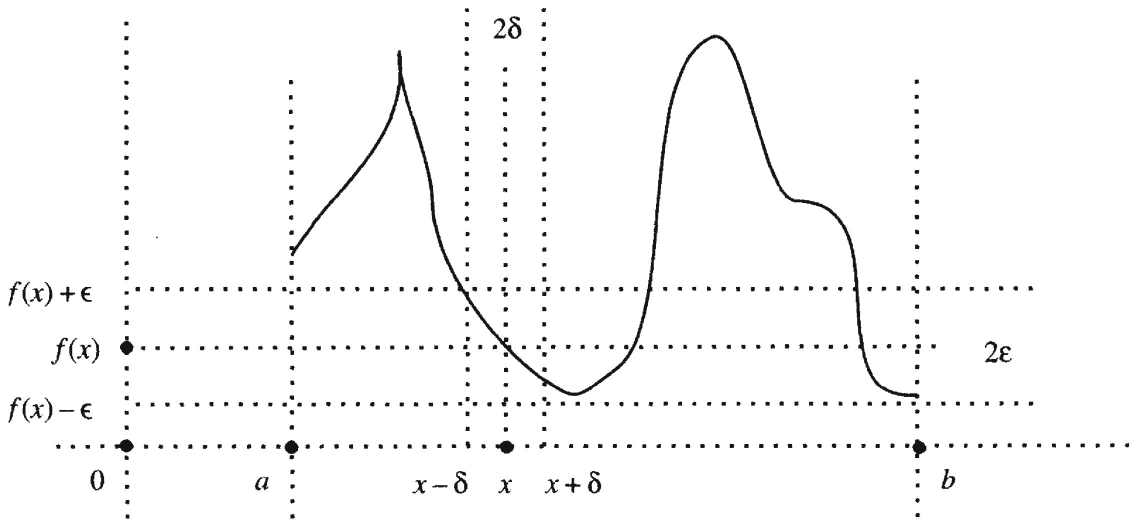

Figure 19 The graph of a continuous function of a real variable

Continuous functions are found everywhere in analysis and topology. Theorems 22, 23, and 24 present their simplest properties. Later we generalize these results to functions that are neither real valued nor dependent on a real variable. Although it is possible to give a combined proof of Theorems 22 and 23 I prefer to highlight the Least Upper Bound Property and keep them separate.

22 Theorem *The values of a continuous function defined on an interval $[a, b]$ form a bounded subset of $\mathbb{R}$. That is, there exist $m, M \in \mathbb{R}$ such that for all $x \in [a, b]$ we have $m \leq f(x) \leq M$.*

Proof For $x \in [a, b]$, let V_x be the value set of $f(t)$ as t varies from a to x,

$$V_x = \{y \in \mathbb{R} : \text{ for some } t \in [a, x] \text{ we have } y = f(t)\}.$$

Set

$$X = \{x \in [a, b] : V_x \text{ is a bounded subset of } \mathbb{R}\}.$$

We must prove that $b \in X$. Clearly $a \in X$ and b is an upper bound for X. Since X is nonempty and bounded above, there exists in $\mathbb{R}$ a least upper bound $c \leq b$ for X. Take $\epsilon = 1$ in the definition of continuity at c. There exists a $\delta > 0$ such that $|x - c| < \delta$ implies $|f(x) - f(c)| < 1$. Since c is the least upper bound for X, there exists $x \in X$ in the interval $[c - \delta, c]$. (Otherwise $c - \delta$ is a smaller upper bound for X.) Now as t varies from a to c, the value $f(t)$ varies first in the bounded set V_x and then in the bounded set $J = (f(c) - 1, f(c) + 1)$. See Figure 20.

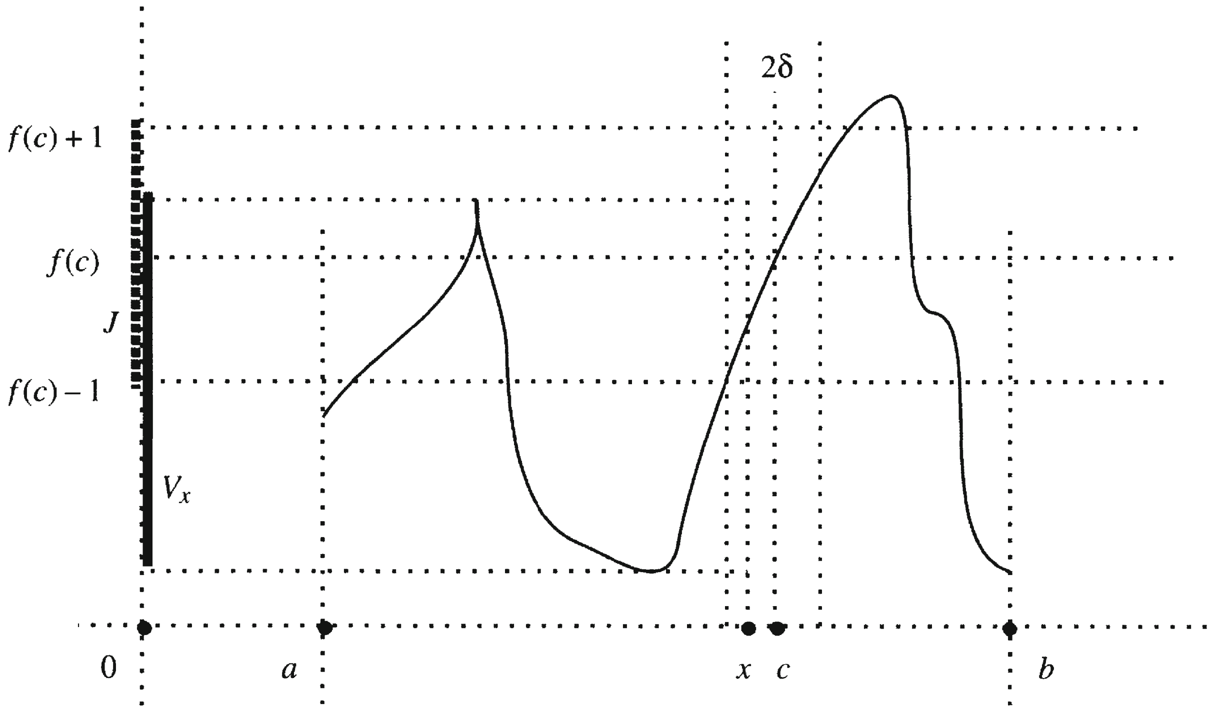

Figure 20 The value set V_x and the interval J

The union of two bounded sets is a bounded set and it follows that V_c is bounded, so $c \in X$. Besides, if $c < b$ then $f(t)$ continues to vary in the bounded set J for $t > c$, contrary to the fact that c is an upper bound for X. Thus, $c = b$, $b \in X$, and the values of f form a bounded subset of $\mathbb{R}$. □

23 Theorem *A continuous function f defined on an interval $[a, b]$ takes on absolute minimum and absolute maximum values: For some $x_0, x_1 \in [a, b]$ and for all $x \in [a, b]$ we have*

$$f(x_0) \;\leq\; f(x) \;\leq\; f(x_1).$$

Proof Let $M = \text{l.u.b.}\, f(t)$ as t varies in $[a, b]$. By Theorem 22 M exists. Consider the set $X = \{x \in [a, b] : \text{l.u.b.} V_x < M\}$ where, as above, V_x is the set of values of $f(t)$ as t varies on $[a, x]$.

Case 1. $f(a) = M$. Then f takes on a maximum at a and the theorem is proved.

Case 2. $f(a) < M$. Then $X \neq \emptyset$ and we can consider the least upper bound of X, say c. If $f(c) < M$, we choose $\epsilon > 0$ with $\epsilon < M - f(c)$. By continuity at c, there exists a $\delta > 0$ such that $|t - c| < \delta$ implies $|f(t) - f(c)| < \epsilon$. Thus, $\text{l.u.b.}V_c < M$. If $c < b$ this implies there exist points t to the right of c at which $\text{l.u.b.}V_t < M$, contrary to the fact that c is an upper bound of such points. Therefore, $c = b$, which implies that $M < M$, a contradiction. Having arrived at a contradiction from the supposition that $f(c) < M$, we duly conclude that $f(c) = M$, so f assumes a maximum at c. The situation with minima is similar. □

24 Intermediate Value Theorem *A continuous function defined on an interval $[a, b]$ takes on (or "achieves," "assumes," or "attains") all intermediate values: That is, if $f(a) = \alpha$, $f(b) = \beta$, and γ is given, $\alpha \leq \gamma \leq \beta$, then there is some $c \in [a, b]$ such that $f(c) = \gamma$. The same conclusion holds if $\beta \leq \gamma \leq \alpha$.*

The theorem is pictorially obvious. A continuous function has a graph that is a curve without break points. Such a graph can not jump from one height to another. It must pass through all intermediate heights.

Proof Set $X = \{x \in [a, b] : \text{l.u.b.}V_x \leq \gamma\}$ and $c = \text{l.u.b.}X$. Now c exists because X is nonempty (it contains a) and it is bounded above (by b). We claim that $f(c) = \gamma$, as shown in Figure 21.

To prove it we just eliminate the other two possibilities which are $f(c) < \gamma$ and $f(c) > \gamma$, by showing that each leads to a contradiction. Suppose that $f(c) < \gamma$ and take $\epsilon = \gamma - f(c)$. Continuity at c gives $\delta > 0$ such that $|t - c| < \delta$ implies $|f(t) - f(c)| < \epsilon$. That is,

$$t \in (c - \delta,\ c + \delta) \quad \Rightarrow \quad f(t) \ < \ \gamma,$$

so $c + \delta/2 \in X$, contrary to c being an upper bound of X.

Suppose that $f(c) > \gamma$ and take $\epsilon = f(c) - \gamma$. Continuity at c gives $\delta > 0$ such that $|t - c| < \delta$ implies $|f(t) - f(c)| < \epsilon$. That is,

$$t \in (c - \delta,\ c + \delta) \quad \Rightarrow \quad f(t) \ > \ \gamma,$$

so $c - \delta/2$ is an upper bound for X, contrary to c being the least upper bound for X. Since $f(c)$ is neither $< \gamma$ nor $> \gamma$ we get $f(c) = \gamma$. □

A combination of Theorems 22, 23, 24, and Exercise 43 could well be called the

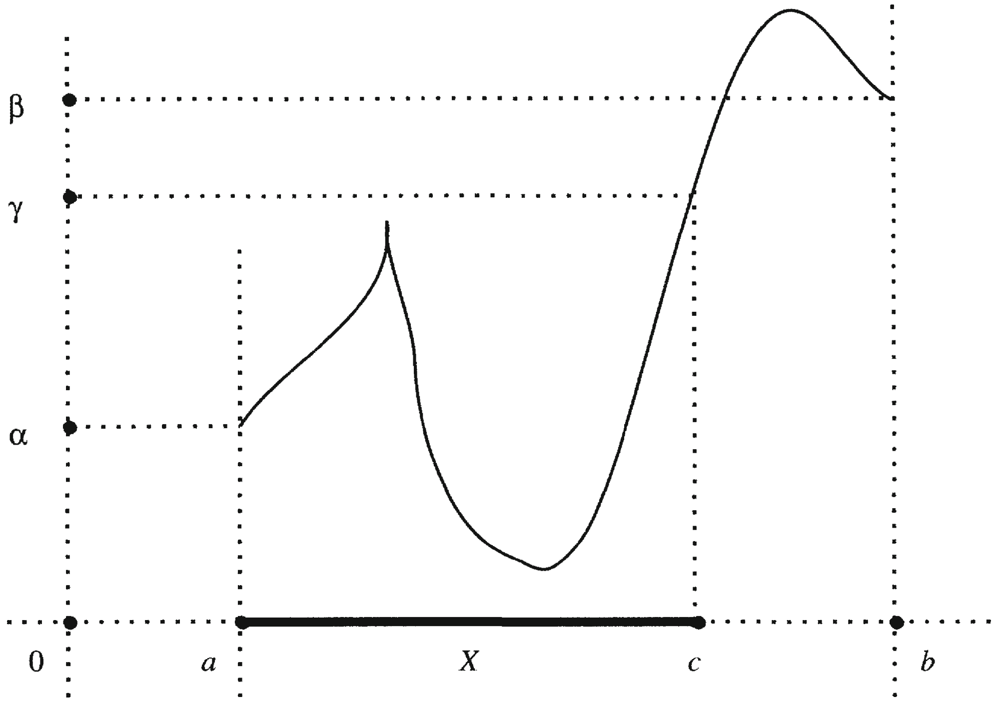

Figure 21 $x \in X$ implies that $f(x) \leq \gamma$.

25 Fundamental Theorem of Continuous Functions *Every continuous real valued function of a real variable $x \in [a, b]$ is bounded, achieves minimum, intermediate, and maximum values, and is uniformly continuous.*

7* Visualizing the Fourth Dimension

A lot of real analysis takes place in $\mathbb{R}^m$ but the full m-dimensionality is rarely important. Rather, most analysis facts which are true when $m = 1, 2, 3$ remain true for $m \geq 4$. Still, I suspect you would be happier if you could visualize $\mathbb{R}^4$, $\mathbb{R}^5$, etc. Here is how to do it.

It is often said that time is *the* fourth dimension and that $\mathbb{R}^4$ should be thought of as $xyzt$-space where a point has position (x, y, z) in 3-space at time t. This is only one possible way to think of *a* fourth dimension. Instead, you can think of color as a fourth dimension. Imagine our usual 3-space with its xyz-coordinates in which points are colorless. Then imagine that you can give color to points ("paint" them) with shades of red indicating positive fourth coordinate and blue indicating negative fourth coordinate. This gives $xyzc$-coordinates. Points with equal xyz-coordinates but different colors are different points.

How is this useful? We have not used time as a coordinate, reserving it to describe motion in 4-space. Figure 22 shows two circles – the unit circle C in the horizontal xy-plane and the circle V with radius 1 and center $(1, 0, 0)$ in the vertical xz-plane. They are linked. No continuous motion can unlink them in 3-space without one

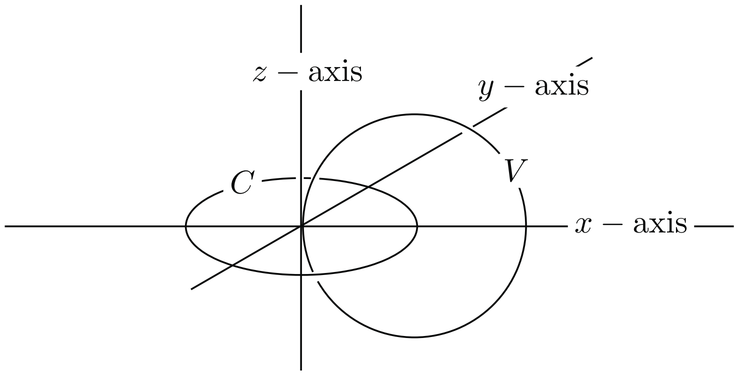

Figure 22 C and V are linked circles.

crossing the other. However, in Figure 23 you can *watch* them unlink in 4-space as follows.

Just gradually give redness to C while dragging it leftward parallel to the x-axis, until it is to the left of V. (Leave V always fixed.) Then diminish the redness of C until it becomes colorless. It ends up to the left of V and no longer links it. In formulas we can let

$$C(t) = \{(x, y, z, c) \in \mathbb{R}^4 : (x + 2t)^2 + y^2 = 1,\ z = 0,\ \text{and } c(t) = t(t - 1)\}$$

for $0 \le t \le 1$. See Figure 23.

The moving circle $C(t)$ never touches the stationary circle V. In particular, at time $t = 1/2$ we have $C(t) \cap V = \emptyset$. For $(-1, 0, 0, 1/4) \neq (-1, 0, 0, 0)$.

Other parameters can be used for higher dimensions. For example we could use pressure, temperature, chemical concentration, monetary value, etc. In theoretical mechanics one uses six parameters for a moving particle – three coordinates of position and three more for momentum.

Moral Choosing a new parameter as the fourth dimension (color instead of time) lets one visualize 4-space and observe motion there.

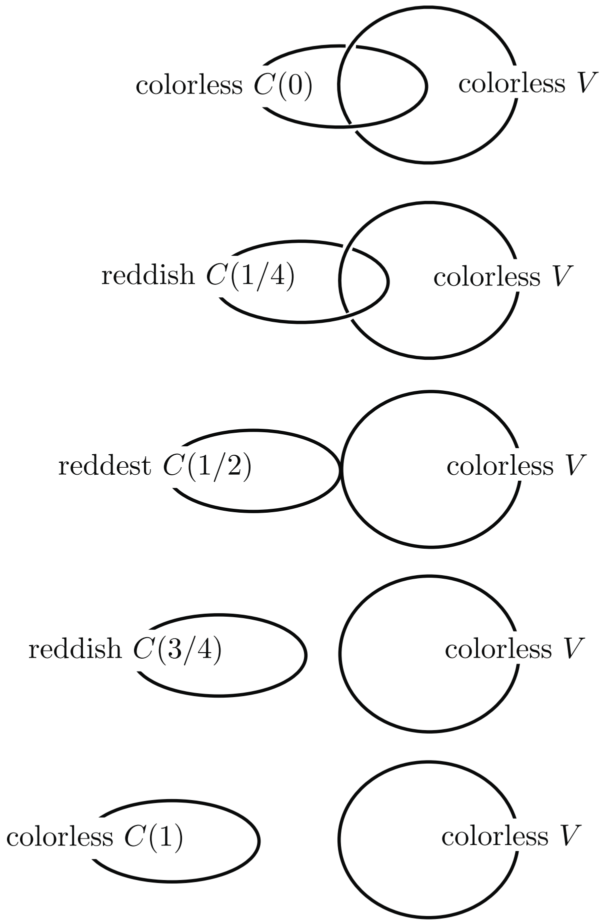

Figure 23 How to unlink linked circles using the fourth dimension

Exercises

0. Prove that for all sets A, B, C the formula

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

is true. Here is the solution written out in gory detail. *Imitate this style in writing out proofs in this course.* See also the guidelines for writing a rigorous proof on page 5. Follow them!
Hypothesis. A, B, C are sets.
Conclusion. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
Proof. To prove two sets are equal we must show that every element of the first set is an element of the second set and vice versa. Referring to Figure 24, let x denote an element of the set $A \cap (B \cup C)$. It belongs to A and it belongs to B or to C. Therefore x belongs to $A \cap B$ or it belongs to $A \cap C$. Thus x belongs to the set $(A \cap B) \cup (A \cap C)$ and we have shown that every element of the first set $A \cap (B \cup C)$ belongs to the second set $(A \cap B) \cup (A \cap C)$.

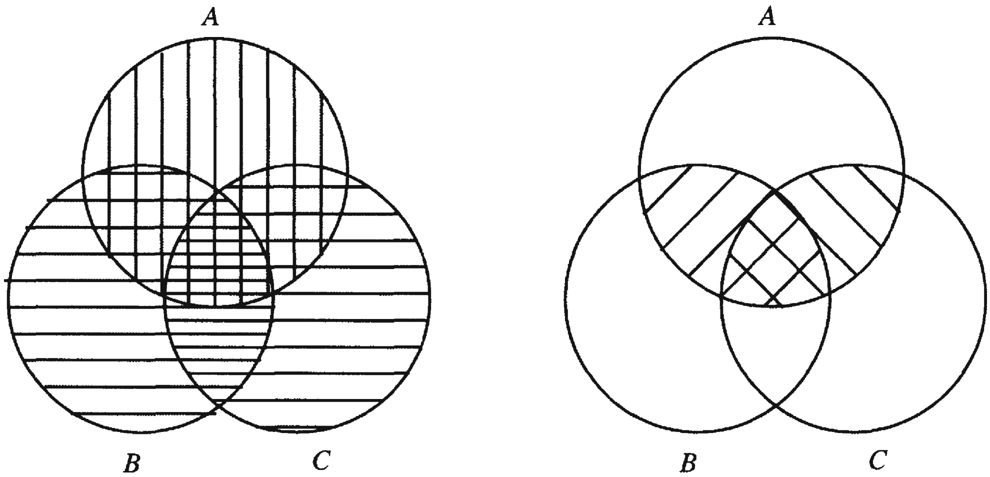

Figure 24 A is ruled vertically, B and C are ruled horizontally, $A \cap B$ is ruled diagonally, and $A \cap C$ is ruled counter-diagonally.

On the other hand let y denote an element of the set $(A \cap B) \cup (A \cap C)$. It belongs to $A \cap B$ or it belongs to $A \cap C$. Therefore it belongs to A and it belongs to B or to C. Thus y belongs to $A \cap (B \cup C)$ and we have shown that every element of the second set $(A \cap B) \cup (A \cap C)$ belongs to the first set $A \cap (B \cup C)$. Since each element of the first set belongs to the second set and each element of the second belongs to the first, the two sets are equal, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$. QED

1. Prove that for all sets A, B, C the formula

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

is true.

2. If several sets $A, B, C, \ldots$ all are subsets of the same set X then the differences $X \setminus A,\ X \setminus B,\ X \setminus C, \ldots$ are the **complements** of $A, B, C, \ldots$ in X and are denoted $A^c, B^c, C^c, \ldots$. The symbol A^c is read "A complement."
   (a) Prove that $(A^c)^c = A$.
   (b) Prove **De Morgan's Law**: $(A \cap B)^c = A^c \cup B^c$ and derive from it the law $(A \cup B)^c = A^c \cap B^c$.
   (c) Draw Venn diagrams to illustrate the two laws.
   (d) Generalize these laws to more than two sets.
3. Recast the following English sentences in mathematics, using correct mathematical grammar. Preserve their meaning.
   (a) 2 is the smallest prime number.
   (b) The area of any bounded plane region is bisected by some line parallel to the x-axis.
   *(c) "All that glitters is not gold."

*4. What makes the following sentence ambiguous? "A death row prisoner can't have too much hope."

5. Negate the following sentences in English using correct mathematical grammar.
   (a) If roses are red, violets are blue.
   *(b) He will sink unless he swims.
6. Why is the square of an odd integer odd and the square of an even integer even? What is the situation for higher powers? [Hint: Prime factorization.]
7. (a) Why does 4 divide every even integer square?
   (b) Why does 8 divide every even integer cube?
   (c) Why can 8 never divide twice an odd cube?
   (d) Prove that the cube root of 2 is irrational.
8. Suppose that the natural number k is not a perfect n^{th} power.
   a Prove that its n^{th} root is irrational.
   b Infer that the n^{th} root of a natural number is either a natural number or it is irrational. It is never a fraction.
9. Let $x = A|B$, $x' = A'|B'$ be cuts in $\mathbb{Q}$. We defined

$$x + x' = (A + A') \mid \text{ rest of } \mathbb{Q}.$$

   (a) Show that although $B + B'$ is disjoint from $A + A'$, it may happen in degenerate cases that $\mathbb{Q}$ is not the union of $A + A'$ and $B + B'$.
   (b) Infer that the definition of $x + x'$ as $(A + A') \mid (B + B')$ would be incorrect.
   (c) Why did we not define $x \cdot x' = (A \cdot A') \mid$ rest of $\mathbb{Q}$?
10. Prove that for each cut x we have $x + (-x) = 0^*$. [This is not entirely trivial.]
11. A multiplicative inverse of a nonzero cut $x = A|B$ is a cut $y = C|D$ such that $x \cdot y = 1^*$.

(a) If $x > 0^*$, what are C and D?
(b) If $x < 0^*$, what are they?
(c) Prove that x uniquely determines y.

12. Prove that there exists no smallest positive real number. Does there exist a smallest positive rational number? Given a real number x, does there exist a smallest real number $y > x$?
13. Let $b = \text{l.u.b.}\, S$, where S is a bounded nonempty subset of $\mathbb{R}$.
(a) Given $\epsilon > 0$ show that there exists an $s \in S$ with
$$b - \epsilon \ \le\ s\ \le\ b.$$
(b) Can $s \in S$ always be found so that $b - \epsilon < s < b$?
(c) If $x = A|B$ is a cut in $\mathbb{Q}$, show that $x = \text{l.u.b.}A$.
14. Prove that $\sqrt{2} \in \mathbb{R}$ by showing that $x \cdot x = 2$ where $x = A|B$ is the cut in $\mathbb{Q}$ with $A = \{r = \mathbb{Q} : r \le 0 \text{ or } r^2 < 2\}$. [Hint: Use Exercise 13. See also Exercise 16, below.]
15. Given $y \in \mathbb{R}$, $n \in \mathbb{N}$, and $\epsilon > 0$, show that for some $\delta > 0$, if $u \in \mathbb{R}$ and $|u - y| < \delta$ then $|u^n - y^n| < \epsilon$. [Hint: Prove the inequality when $n = 1$, $n = 2$, and then do induction on n using the identity
$$u^n - y^n = (u - y)(u^{n-1} + u^{n-2}y + \ldots + y^{n-1}).]$$
16. Given $x > 0$ and $n \in \mathbb{N}$, prove that there is a unique $y > 0$ such that $y^n = x$. That is, the n^{th} root of x exists and is unique. [Hint: Consider
$$y = \text{l. u. b.}\{s \in \mathbb{R} : s^n \le x\}.$$
Then use Exercise 15 to show that y^n can be neither $< x$ nor $> x$.]
17. Let $x, y \in \mathbb{R}$ and $n \in \mathbb{N}$ be given.
(a) Prove that $x < y$ if and only if $x^n < y^n$.
(b) Infer from Exercise 16 that $x < y$ if and only if the n^{th} root of x is less than the n^{th} root of y.
18. Prove that real numbers correspond bijectively to decimal expansions not terminating in an infinite strings of nines, as follows. The decimal expansion of $x \in \mathbb{R}$ is $N.x_1x_2\ldots$, where N is the largest integer $\le x$, x_1 is the largest integer $\le 10(x - N)$, x_2 is the largest integer $\le 100(x - (N + x_1/10))$, and so on.
(a) Show that each x_k is a digit between 0 and 9.
(b) Show that for each k there is an $\ell \ge k$ such that $x_\ell \ne 9$.
(c) Conversely, show that for each such expansion $N.x_1x_2\ldots$ not terminating in an infinite string of nines, the set
$$\{N,\ N + \frac{x_1}{10},\ N + \frac{x_1}{10} + \frac{x_2}{100},\ \ldots\}$$

is bounded and its least upper bound is a real number x with decimal expansion $N.x_1x_2\ldots$.

(d) Repeat the exercise with a general base in place of 10.

19. Formulate the definition of the **greatest lower bound** (g.l.b.) of a set of real numbers. State a g.l.b. property of $\mathbb{R}$ and show it is equivalent to the l.u.b. property of $\mathbb{R}$.
20. Prove that limits are unique, i.e., if (a_n) is a sequence of real numbers that converges to a real number b and also converges to a real number b', then $b = b'$.
21. Let $f : A \to B$ be a function. That is, f is some rule or device which, when presented with any element $a \in A$, produces an element $b = f(a)$ of B. The **graph** of f is the set S of all pairs $(a, b) \in A \times B$ such that $b = f(a)$.
    (a) If you are given a subset $S \subset A \times B$, how can you tell if it is the graph of some function? (That is, what are the set theoretic properties of a graph?)
    (b) Let $g : B \to C$ be a second function and consider the composed function $g \circ f : A \to C$. Assume that $A = B = C = [0, 1]$, draw $A \times B \times C$ as the unit cube in 3-space, and try to relate the graphs of f, g, and $g \circ f$ in the cube.
22. A **fixed-point** of a function $f : A \to A$ is a point $a \in A$ such that $f(a) = a$. The **diagonal** of $A \times A$ is the set of all pairs (a, a) in $A \times A$.
    (a) Show that $f : A \to A$ has a fixed-point if and only if the graph of f intersects the diagonal.
    (b) Prove that every continuous function $f : [0, 1] \to [0, 1]$ has at least one fixed-point.
    (c) Is the same true for continuous functions $f : (0, 1) \to (0, 1)$?[†]
    (d) Is the same true for discontinuous functions?
23. A rational number p/q is **dyadic** if q is a power of 2, $q = 2^k$ for some nonnegative integer k. For example, $0, 3/8, 3/1, -3/256$, are dyadic rationals, but $1/3, 5/12$ are not. A dyadic interval is $[a, b]$ where $a = p/2^k$ and $b = (p + 1)/2^k$. For example, [.75, 1] is a dyadic interval but $[1, \pi]$, $[0, 2]$, and $[.25, .75]$ are not. A dyadic cube is the product of dyadic intervals having equal length. The set of dyadic rationals may be denoted as $\mathbb{Q}_2$ and the dyadic lattice as $\mathbb{Q}_2^m$.
    (a) Prove that any two dyadic squares (i.e., planar dyadic cubes) of the same size are either identical, intersect along a common edge, intersect at a common vertex, or do not intersect at all.
    (b) Show that the corresponding intersection property is true for dyadic cubes in $\mathbb{R}^m$.

[†] A question posed in this manner means that, as well as answering the question with a "yes" or a "no," you should give a proof if your answer is "yes" or a specific counterexample if your answer is "no." Also, to do this exercise you should read Theorems 22, 23, 24.

24. Given a cube in $\mathbb{R}^m$, what is the largest ball it contains? Given a ball in $\mathbb{R}^m$, what is the largest cube it contains? What are the largest ball and cube contained in a given box in $\mathbb{R}^m$?

25. (a) Given $\epsilon > 0$, show that the unit disc contains finitely many dyadic squares whose total area exceeds $\pi - \epsilon$, and which intersect each other only along their boundaries.

    **(b) Show that the assertion remains true if we demand that the dyadic squares are disjoint.

    (c) Formulate (a) in dimension $m = 3$ and $m \geq 4$.

    **(d) Do the analysis with squares and discs interchanged. That is, given $\epsilon > 0$ prove that finitely many disjoint closed discs can be drawn inside the unit square so that the total area of the discs exceeds $1 - \epsilon$. [Hint: The Pile of Sand Principle. On the first day of work, take away 1/16 of a pile of sand. On the second day take away 1/16 of the remaining pile of sand. Continue. What happens to the pile of sand after n days when $n \to \infty$? Instead of sand, think of your obligation to place finitely many disjoint dyadic squares (or discs) that occupy at least 1/16 of the area of the unit disc (or unit square).]

*26. Let $b(R)$ and $s(R)$ be the number of integer unit cubes in $\mathbb{R}^m$ that intersect the ball and sphere of radius R, centered at the origin.

    (a) Let $m = 2$ and calculate the limits

    $$\lim_{R\to\infty} \frac{s(R)}{b(R)} \quad \text{and} \quad \lim_{R\to\infty} \frac{s(R)^2}{b(R)}.$$

    (b) Take $m \geq 3$. What exponent k makes the limit

    $$\lim_{R\to\infty} \frac{s(R)^k}{b(R)}$$

    interesting?

    (c) Let $c(R)$ be the number of integer unit cubes that are contained in the ball of radius R, centered at the origin. Calculate

    $$\lim_{R\to\infty} \frac{c(R)}{b(R)}$$

    (d) Shift the ball to a new, arbitrary center (not on the integer lattice) and re-calculate the limits.

27. Prove that the interval $[a, b]$ in $\mathbb{R}$ is the same as the segment $[a, b]$ in $\mathbb{R}^1$. That is,

$$\begin{aligned} & \{x \in \mathbb{R} : a \leq x \leq b\} \\ = \ & \{y \in \mathbb{R} : \exists\, s, t \in [0, 1] \text{ with } s + t = 1 \text{ and } y = sa + tb\}. \end{aligned}$$

[Hint: How do you prove that two sets are equal?]

28. A **convex combination** of $w_1, \ldots, w_k \in \mathbb{R}^m$ is a vector sum

$$w = s_1 w_1 + \cdots + s_k w_k$$

such that $s_1 + \cdots + s_k = 1$ and $0 \leq s_1, \ldots, s_k \leq 1$.
   (a) Prove that if a set E is convex then E contains the convex combination of any finite number of points in E.
   (b) Why is the converse obvious?
29. (a) Prove that the ellipsoid

$$E = \{(x, y, z) \in \mathbb{R}^3 : \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \leq 1\}$$

   is convex. [Hint: E is the unit ball for a different dot product. What is it? Does the Cauchy-Schwarz inequality not apply to all dot products?]
   (b) Prove that all boxes in $\mathbb{R}^m$ are convex.
30. A function $f : (a, b) \to \mathbb{R}$ is a **convex function** if for all $x, y \in (a, b)$ and all $s, t \in [0, 1]$ with $s + t = 1$ we have

$$f(sx + ty) \ \leq \ sf(x) + tf(y).$$

   (a) Prove that f is convex if and only if the set S of points above its graph is convex in $\mathbb{R}^2$. The set S is $\{(x, y) : f(x) \leq y\}$.
   *(b) Prove that every convex function is continuous.
   (c) Suppose that f is convex and $a < x < u < b$. The slope σ of the line through $(x, f(x))$ and $(u, f(u))$ depends on x and u, say $\sigma = \sigma(x, u)$. Prove that σ increases when x increases, and σ increases when u increases.
   (d) Suppose that f is second-order differentiable. That is, f is differentiable and its derivative f' is also differentiable. As is standard, we write $(f')' = f''$. Prove that f is convex if and only if $f''(x) \geq 0$ for all $x \in (a, b)$.
   (e) Formulate a definition of convexity for a function $f : M \to \mathbb{R}$ where $M \subset \mathbb{R}^m$ is a convex set. [Hint: Start with $m = 2$.]

*31. Suppose that a function $f : [a, b] \to \mathbb{R}$ is monotone nondecreasing. That is, $x_1 \leq x_2$ implies $f(x_1) \leq f(x_2)$.
   (a) Prove that f is continuous except at a countable set of points. [Hint: Show that at each $x \in (a, b)$, f has **right limit** $f(x+)$ and a **left limit** $f(x-)$, which are limits of $f(x + h)$ as h tends to 0 through positive and negative values respectively. The **jump** of f at x is $f(x+) - f(x-)$. Show that f is continuous at x if and only if it has zero jump at x. At how many points can f have jump ≥ 1? At how many points can the jump be between 1/2 and 1? Between 1/3 and 1/2?]
   (b) Is the same assertion true for a monotone function defined on all of $\mathbb{R}$?

*32. Suppose that E is a convex region in the plane bounded by a curve C.

(a) Show that C has a tangent line except at a countable number of points. [For example, the circle has a tangent line at all its points. The triangle has a tangent line except at three points, and so on.]

(b) Similarly, show that a convex function has a derivative except at a countable set of points.

*33. Let $f(k, m)$ be the number of k-dimensional faces of the m-cube. See Table 1.

	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$\cdots$	m	$m+1$
$k=0$	2	4	8	$f(0,4)$	$f(0,5)$	$\cdots$	$f(0,m)$	$f(0,m+1)$
$k=1$	1	4	12	$f(1,4)$	$f(1,5)$	$\cdots$	$f(1,m)$	$f(1,m+1)$
$k=2$	0	1	6	$f(2,4)$	$f(2,5)$	$\cdots$	$f(2,m)$	$f(2,m+1)$
$k=3$	0	0	1	$f(3,4)$	$f(3,5)$	$\cdots$	$f(3,m)$	$f(3,m+1)$
$k=4$	0	0	0	$f(4,4)$	$f(4,5)$	$\cdots$	$f(4,m)$	$f(4,m+1)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

Table 1 $f(k, m)$ is the number of k-dimensional faces of the m-cube.

(a) Verify the numbers in the first three columns.

(b) Calculate the columns $m = 4$, $m = 5$, and give the formula for passing from the m^{th} column to the $(m + 1)^{\text{st}}$.

(c) What would an $m = 0$ column mean?

(d) Prove that the alternating sum of the entries in any column is 1. That is, $2-1 = 1$, $4-4+1 = 1$, $8-12+6-1 = 1$, and in general $\sum(-1)^k f(k, m) = 1$. This alternating sum is called the **Euler characteristic**.

34. Find an exact formula for a bijection $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$. Is one

$$f(i, j) = j + (1 + 2 + \cdots + (i + j - 2)) = \frac{i^2 + j^2 + i(2j - 3) - j + 2}{2}?$$

35. Prove that the union of denumerably many sets B_k, each of which is countable, is countable. How could it happen that the union is finite?

*36. Without using the Schroeder-Bernstein Theorem,

(a) Prove that $[a, b] \sim (a, b] \sim (a, b)$.

(b) More generally, prove that if C is countable then

$$\mathbb{R} \setminus C \sim \mathbb{R} \sim \mathbb{R} \cup C.$$

(c) Infer that the set of irrational numbers has the same cardinality as $\mathbb{R}$, i.e., $\mathbb{R} \setminus \mathbb{Q} \sim \mathbb{R}$. [Hint: Imagine that you are the owner of denumerably many hotels, $H_1, H_2, \ldots$, all fully occupied, and that a traveler arrives and asks you for accommodation. How could you re-arrange your current guests to make room for the traveler?]

*37. Prove that $\mathbb{R}^2 \sim \mathbb{R}$. [Hint: Think of shuffling two digit strings

$$(a_1 a_2 a_3 \ldots) \& (b_1 b_2 b_3 \ldots) \to (a_1 b_1 a_2 b_2 a_3 b_3 \ldots).$$

In this way you could transform a pair of reals into a single real. Be sure to face the nines-termination issue.]

38. Let S be a set and let $\mathcal{P} = \mathcal{P}(S)$ be the collection of all subsets of S. [$\mathcal{P}(S)$ is called the **power set** of S.] Let $\mathcal{F}$ be the set of functions $f : S \to \{0, 1\}$.
    (a) Prove that there is a natural bijection from $\mathcal{F}$ onto $\mathcal{P}$ defined by

$$f \mapsto \{s \in S : f(s) = 1\}.$$

    *(b) Prove that the cardinality of $\mathcal{P}$ is greater than the cardinality of S, even when S is empty or finite.

[Hints: The notation Y^X is sometimes used for the set of all functions $X \to Y$. In this notation $\mathcal{F} = \{0, 1\}^S$ and assertion (b) becomes $\#(S) < \#(\{0, 1\}^S)$. The empty set has one subset, itself, whereas it has no elements, so $\#(\emptyset) = 0$, while $\#(\{0, 1\}^\emptyset) = 1$, which proves (b) for the empty set. Assume there is a bijection from S onto $\mathcal{P}$. Then there is a bijection $\beta : S \to \mathcal{F}$, and for each $s \in S$, $\beta(s)$ is a function, say $f_s : S \to \{0, 1\}$. Think like Cantor and try to find a function which corresponds to no s. Infer that β could not have been onto.]

39. A real number is **algebraic** if it is a root of a nonconstant polynomial with integer coefficients.
    (a) Prove that the set A of algebraic numbers is denumerable. [Hint: Each polynomial has how many roots? How many linear polynomials are there? How many quadratics? ...]
    (b) Repeat the exercise for roots of polynomials whose coefficients belong to some fixed, arbitrary denumerable set $S \subset \mathbb{R}$.
    *(c) Repeat the exercise for roots of trigonometric polynomials with integer coefficients.
    (d) Real numbers that are not algebraic are said to be **transcendental**. Trying to find transcendental numbers is said to be like looking for hay in a haystack. Why?

40. A **finite word** is a finite string of letters, say from the roman alphabet.
    (a) What is the cardinality of the set of all finite words, and thus of the set of all possible poems and mathematical proofs?
    (b) What if the alphabet had only two letters?
    (c) What if it had countably many letters?
    (d) Prove that the cardinality of the set Σ_n of all infinite words formed using a finite alphabet of n letters, $n \geq 2$, is equal to the cardinality of $\mathbb{R}$.

(e) Give a solution to Exercise 37 by justifying the equivalence chain

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} \quad \sim \quad \Sigma_2 \times \Sigma_2 \quad \sim \quad \Sigma_4 \times \Sigma_4 \quad \sim \quad \mathbb{R}.$$

(f) How many decimal expansions terminate in an infinite string of 9's? How many don't?

41. If v is a value of a continuous function $f : [a, b] \to \mathbb{R}$ use the Least Upper Bound Property to prove that there are smallest and largest $x \in [a, b]$ such that $f(x) = v$.

42. A function defined on an interval $[a, b]$ or (a, b) is **uniformly continuous** if for each $\epsilon > 0$ there exists a $\delta > 0$ such that $|x - t| < \delta$ implies that $|f(x) - f(t)| < \epsilon$. (Note that this δ cannot depend on x, it can only depend on ϵ. With ordinary continuity, the δ can depend on both x and ϵ.)
   (a) Show that a uniformly continuous function is continuous but continuity does not imply uniform continuity. (For example, prove that $\sin(1/x)$ is continuous on the interval $(0, 1)$ but is not uniformly continuous there. Graph it.)
   (b) Is the function $2x$ uniformly continuous on the unbounded interval $(-\infty, \infty)$?
   (c) What about x^2?

*43. Prove that a continuous function defined on an interval $[a, b]$ is uniformly continuous. [Hint: Let $\epsilon > 0$ be given. Think of ϵ as fixed and consider the sets

$$A(\delta) = \{u \in [a, b] : \text{if } x, t \in [a, u] \text{ and } |x - t| < \delta \text{ then } |f(x) - f(t)| < \epsilon\}$$

$$A = \bigcup_{\delta > 0} A(\delta).$$

Using the Least Upper Bound Property, prove that $b \in A$. Infer that f is uniformly continuous. The fact that continuity on $[a, b]$ implies uniform continuity is one of the important, fundamental principles of continuous functions.]

*44. Define injections $f : \mathbb{N} \to \mathbb{N}$ and $g : \mathbb{N} \to \mathbb{N}$ by $f(n) = 2n$ and $g(n) = 2n$. From f and g, the Schroeder-Bernstein Theorem produces a bijection $\mathbb{N} \to \mathbb{N}$. What is it?

*45. Let (a_n) be a sequence of real numbers. It is **bounded** if the set $A = \{a_1, a_2, \ldots\}$ is bounded. The **limit supremum**, or lim sup, of a bounded sequence (a_n) as $n \to \infty$ is

$$\limsup_{n \to \infty} a_n = \lim_{n \to \infty} \left(\sup_{k \geq n} a_k \right)$$

   (a) Why does the lim sup exist?
   (b) If $\sup\{a_n\} = \infty$, how should we define $\limsup_{n \to \infty} a_n$?
   (c) If $\lim_{n \to \infty} a_n = -\infty$, how should we define $\limsup a_n$?

(d) When is it true that

$$\begin{aligned} \limsup_{n\to\infty}(a_n + b_n) &\le \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n \\ \limsup_{n\to\infty} ca_n &= c\limsup_{n\to\infty} a_n? \end{aligned}$$

When is it true they are unequal? Draw pictures that illustrate these relations.

(e) Define the **limit infimum**, or lim inf, of a sequence of real numbers, and find a formula relating it to the limit supremum.

(f) Prove that $\lim_{n\to\infty} a_n$ exists if and only if the sequence (a_n) is bounded and $\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n$.

**46. The unit ball with respect to a norm $\| \;\|$ on $\mathbb{R}^2$ is

$$\{v \in \mathbb{R}^2 : \|v\| \le 1\}.$$

(a) Find necessary and sufficient geometric conditions on a subset of $\mathbb{R}^2$ that it is the unit ball for some norm.

(b) Give necessary and sufficient geometric conditions that a subset be the unit ball for a norm arising from an inner product.

(c) Generalize to $\mathbb{R}^m$. [You may find it useful to read about closed sets in the next chapter, and to consult Exercise 41 there.]

47. Assume that V is an inner product space whose inner product induces a norm as $|x| = \sqrt{\langle x, x\rangle}$.

(a) Show that $| \;|$ obeys the **parallelogram law**

$$|x + y|^2 + |x - y|^2 = 2|x|^2 + 2|y|^2$$

for all $x, y \in V$.

*(b) Show that any norm obeying the parallelogram law arises from a unique inner product. [Hints: Define the prospective inner product as

$$\langle x, y\rangle = \left|\frac{x + y}{2}\right|^2 - \left|\frac{x - y}{2}\right|^2$$

Checking that $\langle\;,\;\rangle$ satisfies the inner product properties of symmetry and positive definiteness is easy. Also, it is immediate that $|x|^2 = \langle x, x\rangle$, so $\langle\;,\;\rangle$ induces the given norm. Checking bilinearity is another story.

(i) Let $x, y, z \in V$ be arbitrary. Show that the parallelogram law implies

$$\langle x + y,\, z\rangle + \langle x - y,\, z\rangle = 2\langle x, y\rangle,$$

and infer that $\langle 2x, z\rangle = 2\langle x, z\rangle$. For arbitrary $u, v \in V$ set $x = \frac{1}{2}(u+v)$ and $y = \frac{1}{2}(u-v)$, plug in to the previous equation, and deduce

$$\langle u, z\rangle + \langle v, z\rangle = \langle u+v, z\rangle,$$

which is additive bilinearity in the first variable. Why does it now follow at once that $\langle\ ,\ \rangle$ is also additively bilinear in the second variable?

(ii) To check multiplicative bilinearity, prove by induction that if $m \in \mathbb{Z}$ then $m\langle x, y\rangle = \langle mx, y\rangle$, and if $n \in \mathbb{N}$ then $\frac{1}{n}\langle x, y\rangle = \langle \frac{1}{n}x, y\rangle$. Infer that $r\langle x, y\rangle = \langle rx, y\rangle$ when r is rational. Is $\lambda \mapsto \langle \lambda x, y\rangle - \lambda\langle x, y\rangle$ a continuous function of $\lambda \in \mathbb{R}$, and does this give multiplicative bilinearity?]

48. Consider a knot in 3-space as shown in Figure 25. In 3-space it cannot be

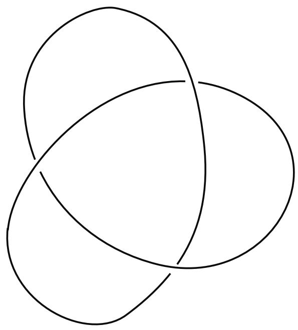

Figure 25 An overhand knot in 3-space

unknotted. How can you unknot it in 4-space?

*49. Prove that there exists no continuous three dimensional motion de-linking the two circles shown in Figure 22 which keeps both circles flat at all times.

50. The Klein bottle is a surface that has an oval of self intersection when it is shown in 3-space. See Figure 26. It can live in 4-space with no self-intersection.

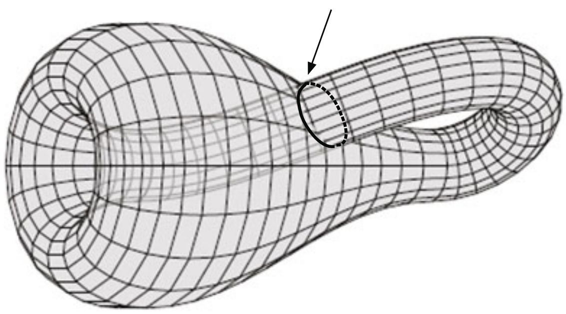

Figure 26 The Klein Bottle in 3-space has an oval of self-intersection.

How?

51. Read *Flatland* by Edwin Abbott. Try to imagine a Flatlander using color to visualize 3-space.
52. Can you visualize a 4-dimensional cube – its vertices, edges, and faces? [Hint: It may be easier (and equivalent) to picture a 4-dimensional parallelepiped whose eight red vertices have xyz-coordinates that differ from the xyz-coordinates of its eight colorless vertices. It is a 4-dimensional version of a rectangle or parallelogram whose edges are not parallel to the coordinate axes.]