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**, $1, 2, 3, \ldots$. 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, $-15 \in \mathbb{Z}$.

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 $A \subset A$.

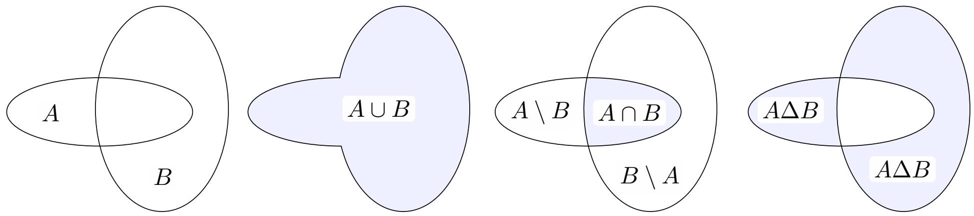

Figure 1 Venn diagrams of union, intersection, and differences

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."

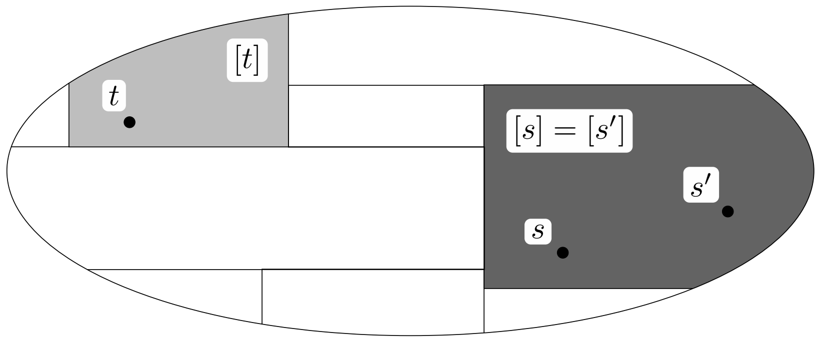

Figure 2 Equivalence classes and representatives

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

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

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:

$$a \Rightarrow b \quad \equiv \quad \neg a \Leftarrow \neg b \quad \equiv \quad \neg a \vee b. \tag{8}$$

$$\neg(a \Rightarrow b) \quad \equiv \quad a \ \& \ \neg b. \tag{9}$$

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.

<u>Case 1.</u> 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.

<u>Case 2.</u> 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.

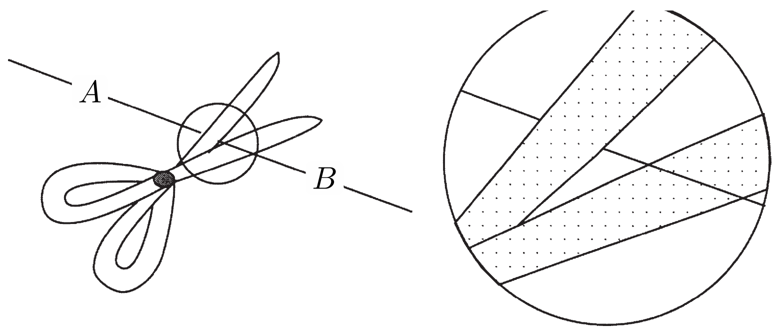

Figure 3 A Dedekind cut

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 $a \leq y \leq b$ for all $a \in \mathcal{A}$ and $b \in \mathcal{B}$. By trichotomy,

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

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 $\lim\limits_{n\to\infty} a_n = b$. 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 $b = \mathrm{l.u.b.}\, X$ 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. □

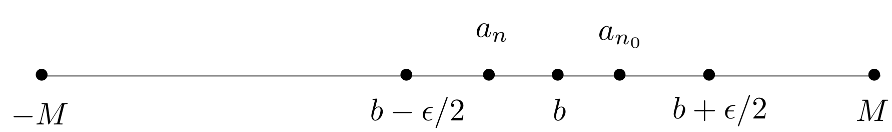

Figure 4 For all $n \geq N$ we have $|a_n - b| < \epsilon$.

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). $\square$

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.

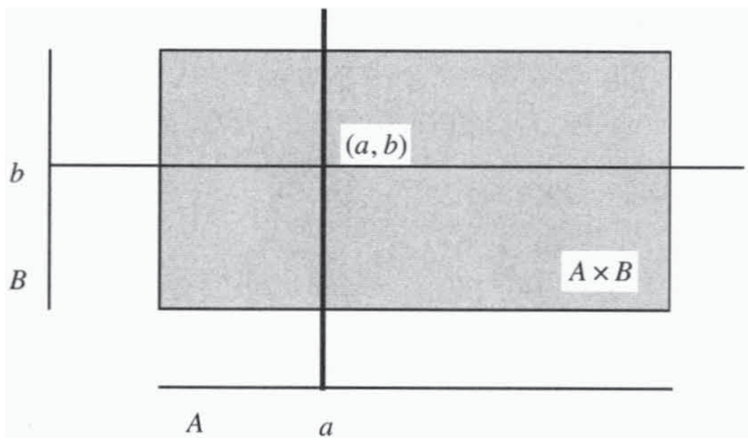

Figure 5 The Cartesian product $A \times B$

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.

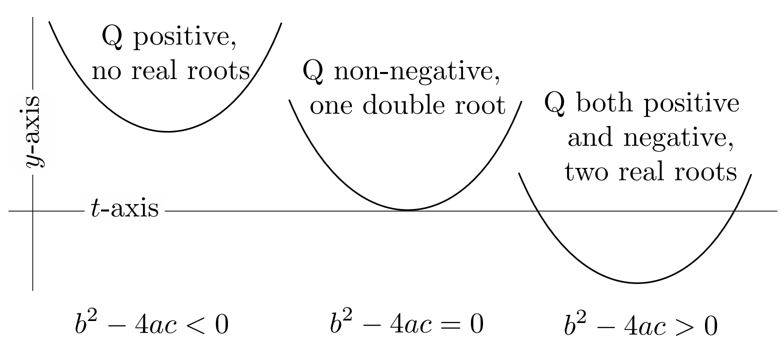

Figure 6 Quadratic graphs

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

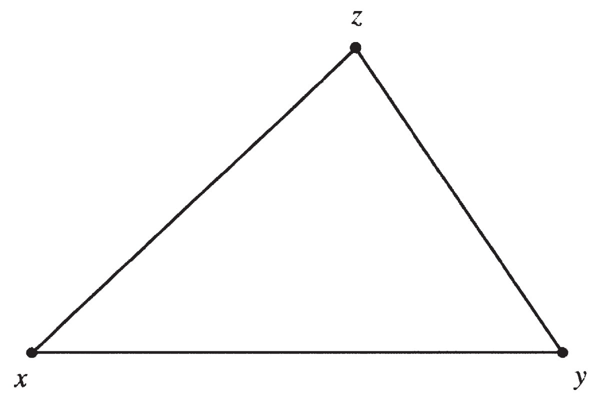

Figure 7 How the Triangle Inequality gets its name

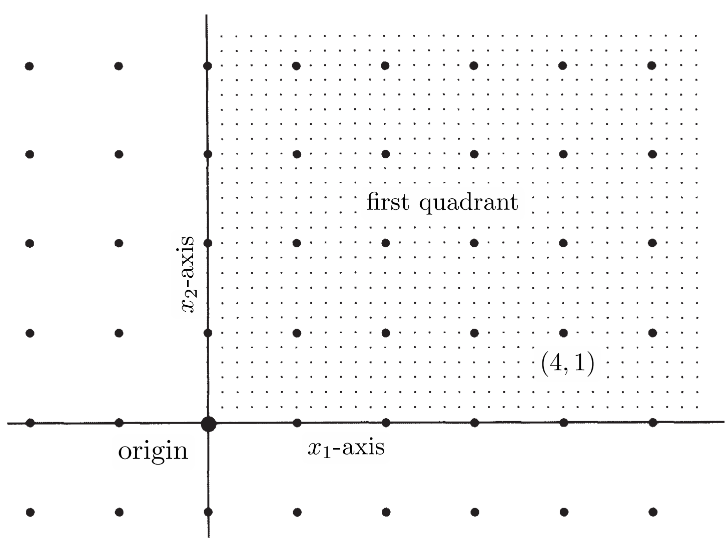

Figure 8 The integer lattice and first quadrant

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.

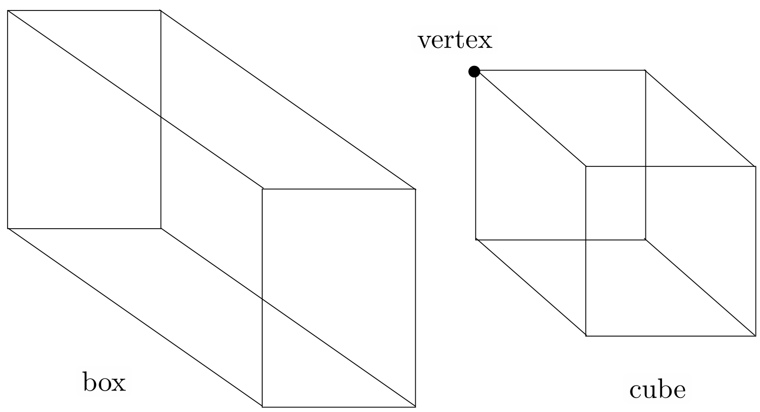

Figure 9 A box and a cube

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

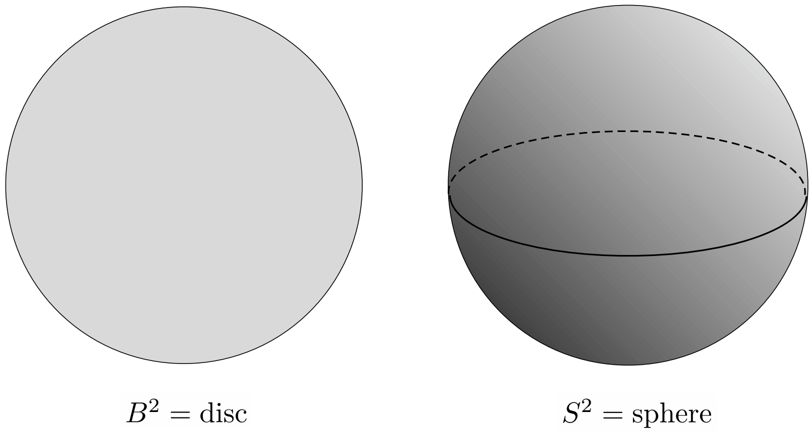

Figure 10 A 2-disc B^2 with its boundary circle, and a 2-sphere S^2 with its equator

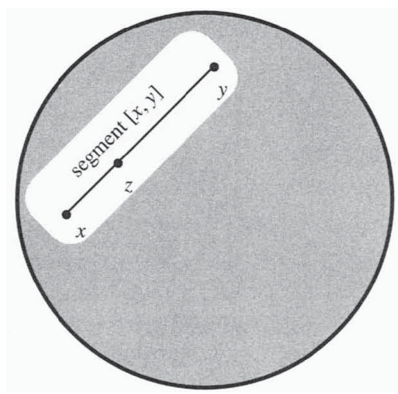

Figure 11 Convexity of the ball

$sx + ty$ with $s + t = 1$ and $0 \leq s, t \leq 1$ 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 $2st \geq 0$, we get

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

Taking the square root of both sides gives $|z| \leq 1$, 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**

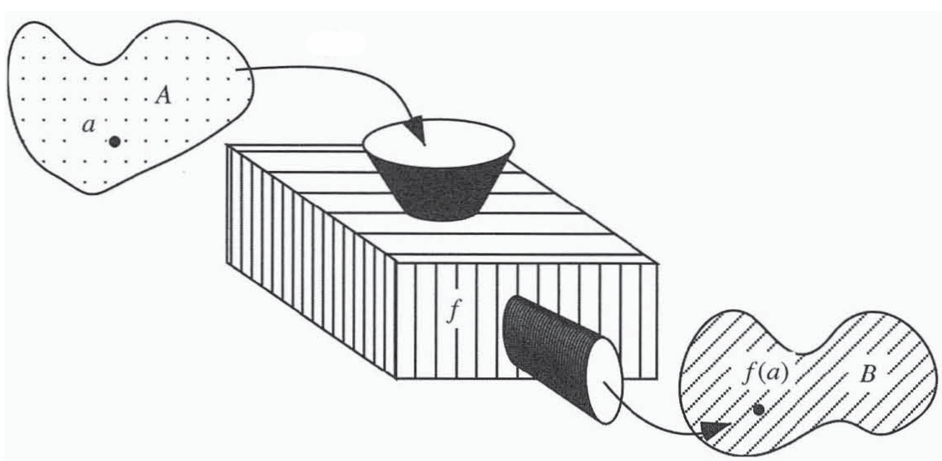

Figure 12 The function f as a machine

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.

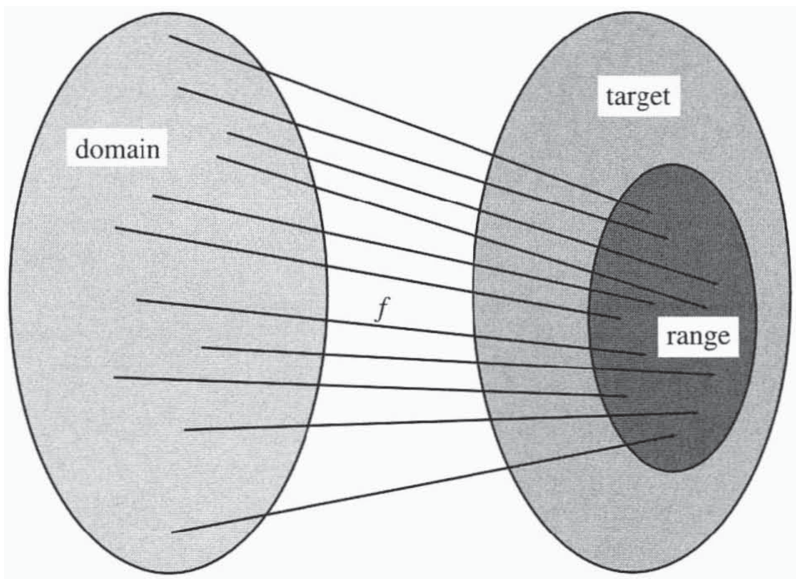

Figure 13 The domain, target, and range of a function

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 card $A = \text{card}\, B$ 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.

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

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.

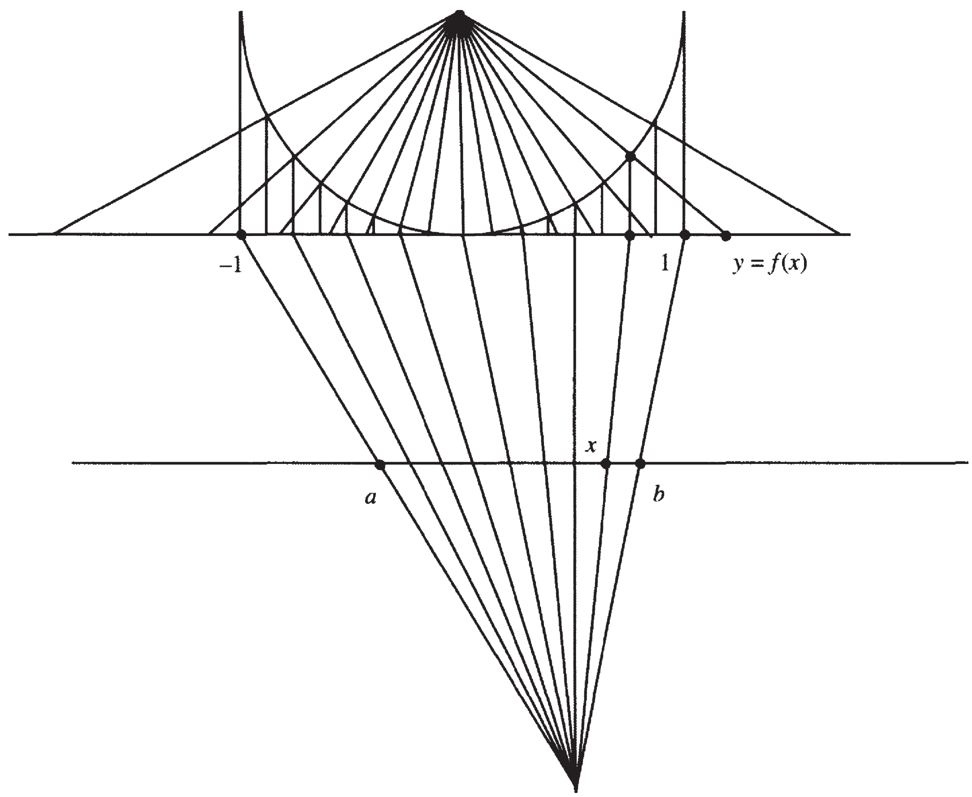

Figure 15 Equicardinality of (a, b), $(-1, 1)$, and $\mathbb{R}$

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. $\square$

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. $\square$

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

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

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. $\square$

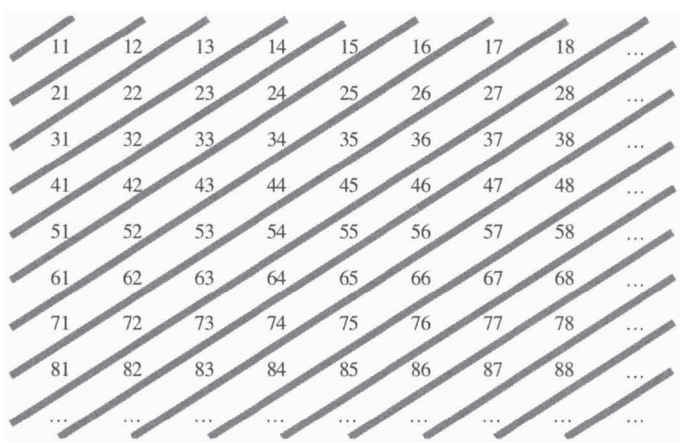

Figure 16 Counter-diagonals in an $\infty \times \infty$ matrix

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

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

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

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

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

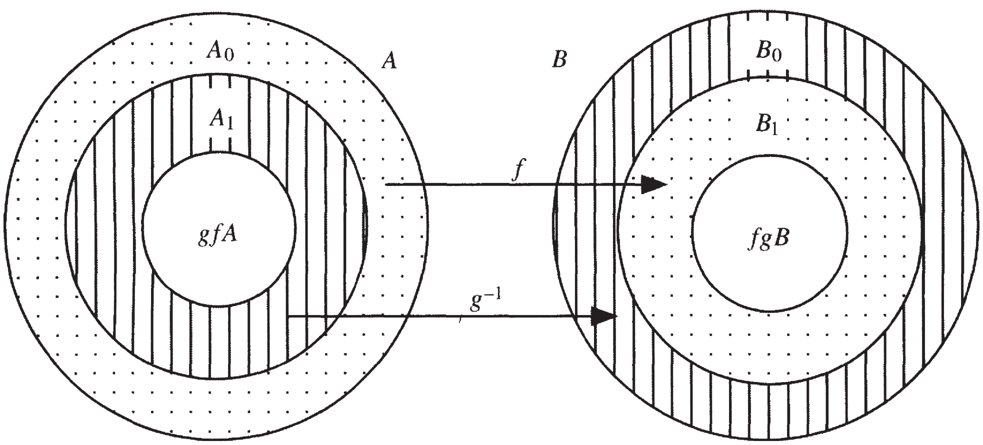

Figure 17 Pictorial proof of the Schroeder-Bernstein Theorem

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

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

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.

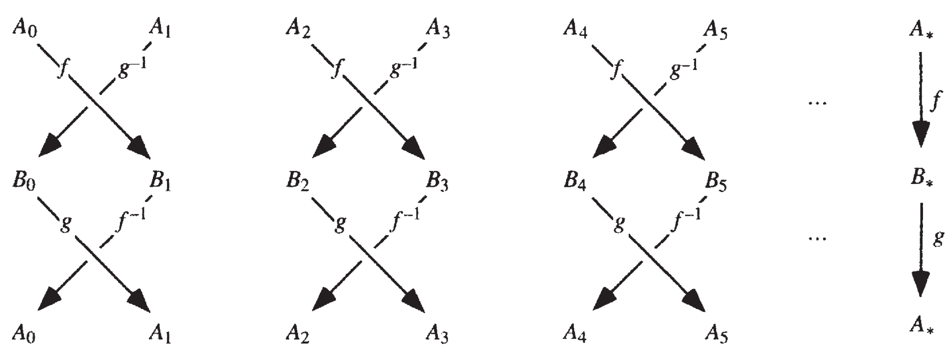

Figure 18 Diagramatic proof of the Schroeder-Bernstein Theorem

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.

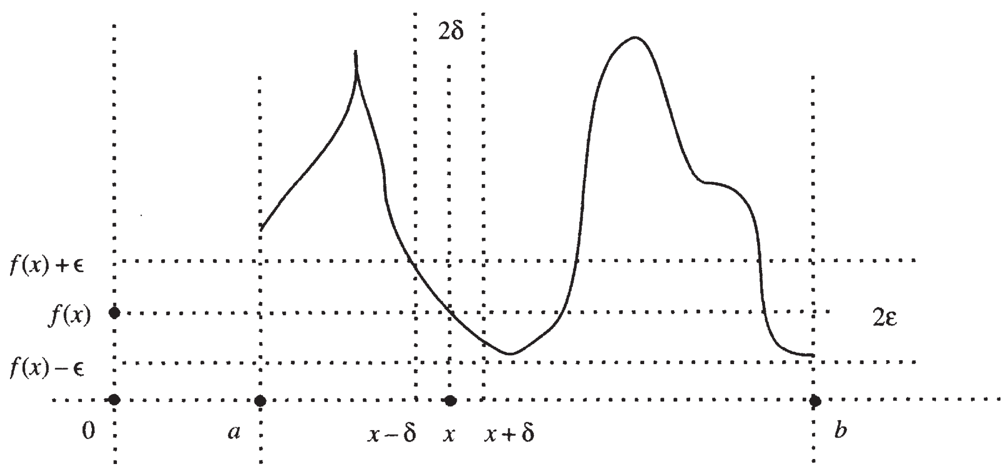

Figure 19 The graph of a continuous function of a real variable

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.

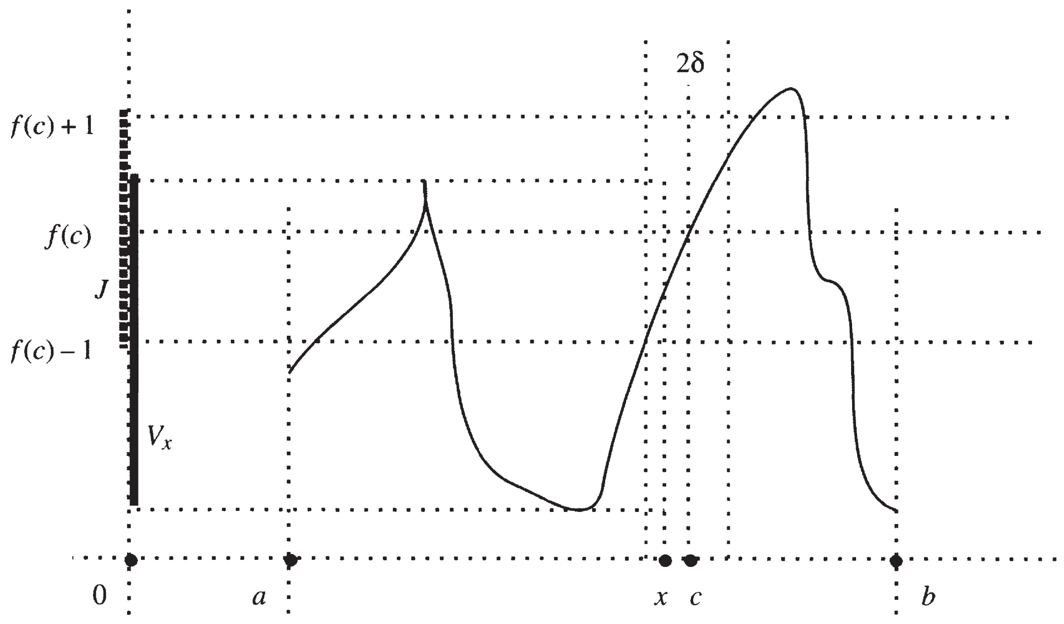

Figure 20 The value set V_x and the interval J

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

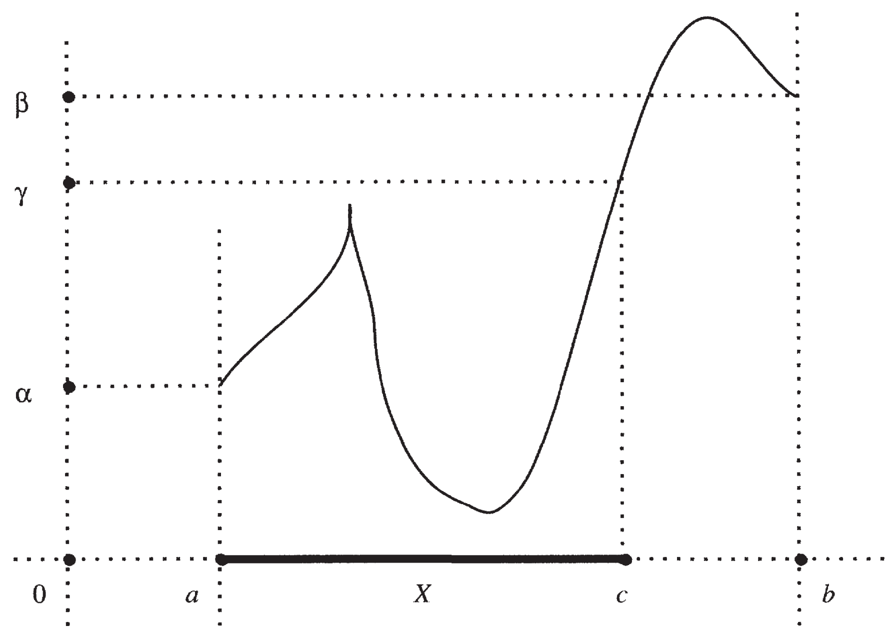

Figure 21 $x \in X$ implies that $f(x) \leq \gamma$.

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

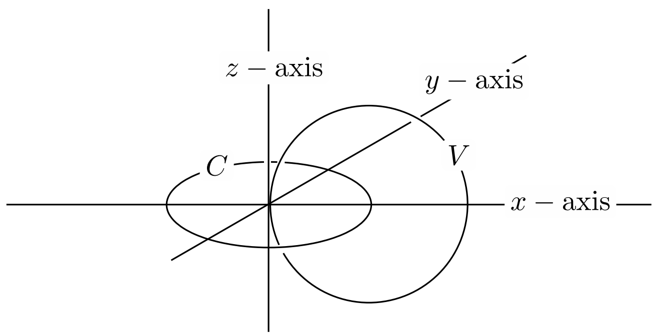

Figure 22 C and V are linked circles.

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 $0 \leq t \leq 1$. 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.

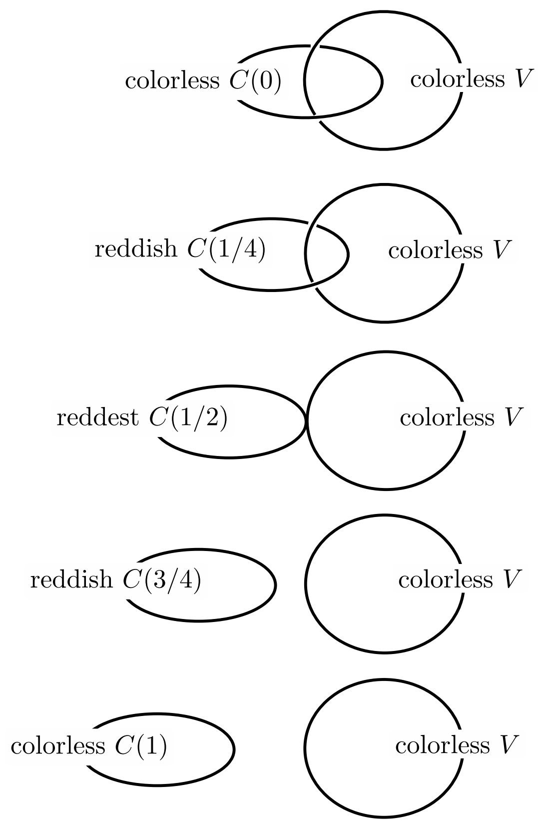

Figure 23 How to unlink linked circles using the fourth dimension

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

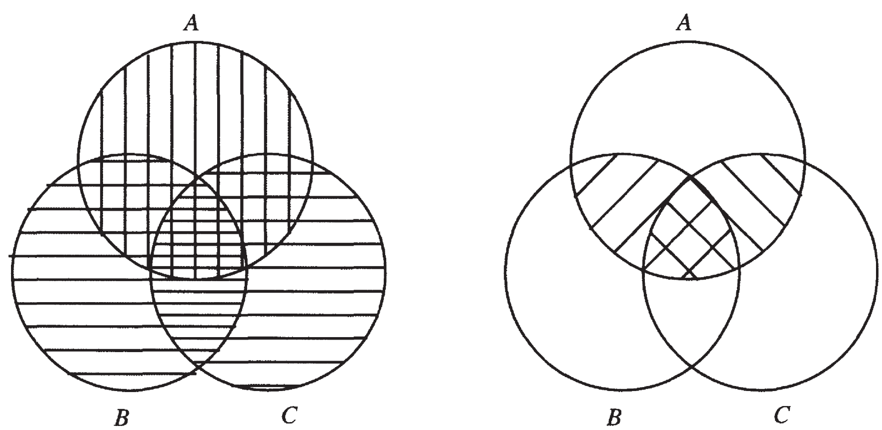

Figure 24 A is ruled vertically, B and C are ruled horizontally, $A \cap B$ is ruled diagonally, and $A \cap C$ is ruled counter-diagonally.

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
    $$b - \epsilon \ \leq \ s \ \leq \ b.$$
    (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 $A = \{r = \mathbb{Q} : r \leq 0 \text{ or } r^2 < 2\}$. [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
    $$y = \text{l. u. b.}\{s \in \mathbb{R} : s^n \leq x\}.$$
    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 $\leq x$, x_1 is the largest integer $\leq 10(x - N)$, x_2 is the largest integer $\leq 100(x - (N + x_1/10))$, 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 $\ell \geq k$ such that $x_\ell \neq 9$.
    (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 $0 \le s_1, \ldots, s_k \le 1$.

(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

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

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

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

(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 $\{(x, y) : f(x) \le y\}$.

*(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 $f''(x) \ge 0$ 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, $x_1 \le x_2$ implies $f(x_1) \le f(x_2)$.

(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

$$(a_1a_2a_3\ldots)\&(b_1b_2b_3\ldots) \to (a_1b_1a_2b_2a_3b_3\ldots).$$

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

$$\begin{aligned} A(\delta) &= \{u \in [a,b] : \text{if } x,t \in [a,u] \text{ and } |x-t| < \delta \\ &\qquad\qquad \text{then } |f(x) - f(t)| < \epsilon\} \\ A &= \bigcup_{\delta>0} A(\delta). \end{aligned}$$

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 $\limsup\limits_{n\to\infty} a_n$?
   (c) If $\lim\limits_{n\to\infty} a_n = -\infty$, how should we define $\limsup a_n$?

(d) When is it true that

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

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

(e) Define the **limit infimum**, or $\liminf$, 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

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

(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

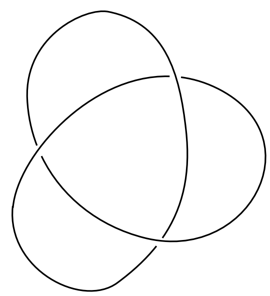

Figure 25 An overhand knot in 3-space

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.

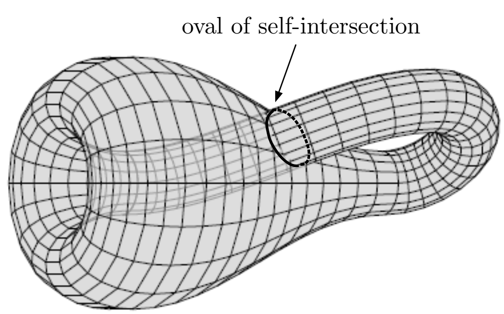

Figure 26 The Klein Bottle in 3-space has an oval of self-intersection.

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.]

2

A Taste of Topology

1 Metric Spaces

It may seem paradoxical at first, but a specific math problem can be harder to solve than some abstract generalization of it. For instance if you want to know how many roots the equation

$$t^5 - 4t^4 + t^3 - t + 1 = 0$$

can have then you could use calculus and figure it out. It would take a while. But thinking more abstractly, and with less work, you could show that every n^{th}-degree polynomial has at most n roots. In the same way many general results about functions of a real variable are more easily grasped at an abstract level – the level of metric spaces.

Metric space theory can be seen as a special case of general topology, and many books present it that way, explaining compactness primarily in terms of open coverings. In my opinion, however, the sequence/subsequence approach provides the easiest and simplest route to mastering the subject. Accordingly it gets top billing throughout this chapter.

A **metric space** is a set M, the elements of which are referred to as points of M, together with a **metric** d having the three properties that distance has in Euclidean space. The metric $d = d(x, y)$ is a real number defined for all points $x, y \in M$ and $d(x, y)$ is called the **distance** from the point x to the point y. The three distance properties are as follows: For all $x, y, z \in M$ we have

(a) **positive definiteness**: $d(x, y) \geq 0$, and $d(x, y) = 0$ if and only if $x = y$.
(b) **symmetry**: $d(x, y) = d(y, x)$.
(c) **triangle inequality**: $d(x, z) \leq d(x, y) + d(y, z)$.

The function d is also called the **distance function**. Strictly speaking, it is the pair (M, d) which is a metric space, but we will follow the common practice of referring to "the metric space M," and leave to you the job of inferring the correct metric.

The main examples of metric spaces are $\mathbb{R}$, $\mathbb{R}^m$, and their subsets. The metric on $\mathbb{R}$ is $d(x, y) = |x - y|$ where $x, y \in \mathbb{R}$ and $|x - y|$ is the magnitude of $x - y$. The metric on $\mathbb{R}^m$ is the Euclidean length of $x - y$ where x, y are vectors in $\mathbb{R}^m$. Namely,

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

for $x = (x_1, \ldots, x_m)$ and $y = (y_1, \ldots, y_m)$.

Since Euclidean length satisfies the three distance properties, d is a bona fide metric and it makes $\mathbb{R}^m$ into a metric space. A subset $M \subset \mathbb{R}^m$ becomes a metric space when we declare the distance between points of M to be their Euclidean distance apart as points in $\mathbb{R}^m$. We say that M **inherits** its metric from $\mathbb{R}^m$ and is a **metric subspace** of $\mathbb{R}^m$. Figure 27 shows a few subsets of $\mathbb{R}^2$ to suggest some interesting metric spaces.

There is also one metric that is hard to picture but valuable as a source for counterexamples, the **discrete metric**. Given any set M, define the distance between distinct points of M to be 1 and the distance between every point and itself to be 0. This is a metric. See Exercise 4. If M consists of three points, say $M = \{a, b, c\}$, you can think of the vertices of the unit equilateral triangle as a model for M. See Figure 28. They have mutual distance 1 from each other. If M consists of one, two, or four points can you think of a model for the discrete metric on M? More challenging is to imagine the discrete metric on $\mathbb{R}$. All points, by definition of the discrete metric, lie at unit distance from each other.

Convergent Sequences and Subsequences

A sequence of points in a metric space M is a list $p_1, p_2, \ldots$ where the points p_n belong to M. Repetition is allowed, and not all the points of M need to appear in the list. Good notation for a sequence is (p_n), or $(p_n)_{n \in \mathbb{N}}$. The notation $\{p_n\}$ is also used but it is too easily confused with the set of points making up the sequence. The difference between $(p_n)_{n \in \mathbb{N}}$ and $\{p_n : n \in \mathbb{N}\}$ is that in the former case

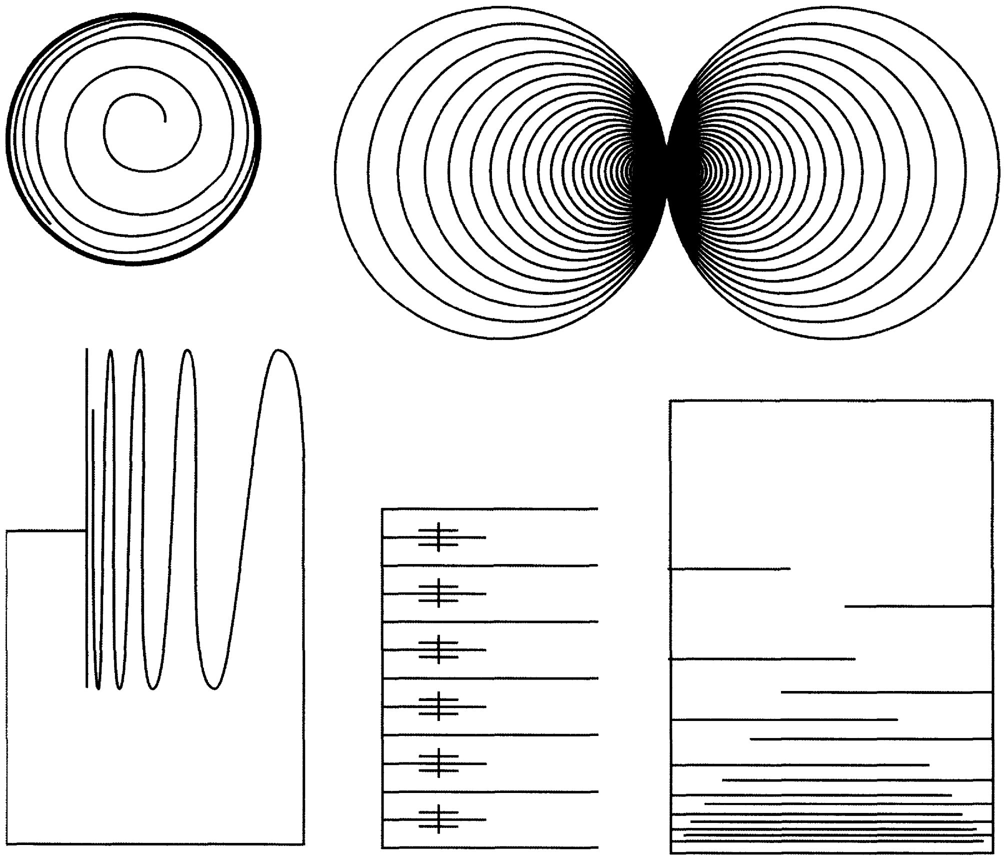

Figure 27 Five metric spaces – a closed outward spiral, a Hawaiian earring, a topologist's sine circle, an infinite television antenna, and Zeno's maze

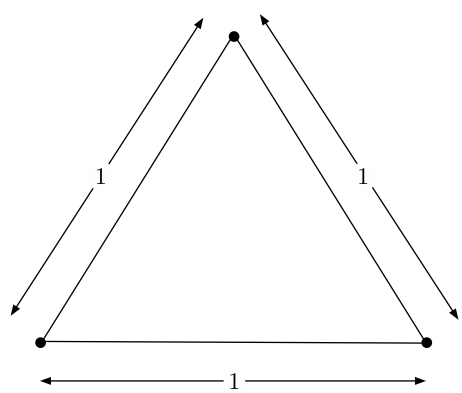

Figure 28 The vertices of the unit equilateral triangle form a discrete metric space.

the sequence prescribes an ordering of the points, while in the latter the points get jumbled together. For example, the sequences $1, 2, 3, \ldots$ and $1, 2, 1, 3, 2, 1, 4, 3, 2, 1, \ldots$ are different sequences but give the same set of points, namely $\mathbb{N}$.

Formally, a sequence in M is a function $f : \mathbb{N} \to M$. The n^{th} term in the sequence is $f(n) = p_n$. Clearly, every sequence defines a function $f : \mathbb{N} \to M$ and conversely, every function $f : \mathbb{N} \to M$ defines a sequence in M. The sequence (p_n) **converges to the limit** p in M if

$$\begin{aligned}&\forall \epsilon > 0 \ \exists N \in \mathbb{N} \text{ such that} \\ &n \in \mathbb{N} \text{ and } n \geq N \quad \Rightarrow \quad d(p_n, p) < \epsilon.\end{aligned}$$

Limits are unique in the sense that if (p_n) converges to p and if (p_n) also converges to p' then $p = p'$. The reason is this. Given any $\epsilon > 0$, there are integers N and N' such that if $n \geq N$ then $d(p_n, p) < \epsilon$, while if $n \geq N'$ then $d(p_n, p') < \epsilon$. Then for all $n \geq \max\{N, N'\}$ we have

$$d(p, p') \ \leq \ d(p, p_n) + d(p_n, p') \ < \ \epsilon + \epsilon \ = \ 2\epsilon.$$

But ϵ is arbitrary and so $d(p, p') = 0$ and $p = p'$. (This is the ϵ-principle again.)

We write $p_n \to p$, or $p_n \to p$ as $n \to \infty$, or

$$\lim_{n\to\infty} p_n \ = \ p$$

to indicate convergence. For example, in $\mathbb{R}$ the sequence $p_n = 1/n$ converges to 0 as $n \to \infty$. In $\mathbb{R}^2$ the sequence $(1/n, \ \sin n)$ does not converge as $n \to \infty$. In the metric space $\mathbb{Q}$ (with the metric it inherits from $\mathbb{R}$) the sequence 1, 1.4, 1.414, 1.4142, $\ldots$ does not converge.

Just as a set can have a subset, a sequence can have a subsequence. For example, the sequence $2, 4, 6, 8, \ldots$ is a subsequence of $1, 2, 3, 4, \ldots$. The sequence $3, 5, 7, 11, 13, 17, \ldots$ is a subsequence of $1, 3, 5, 7, 9, \ldots$, which in turn is a subsequence of $1, 2, 3, 4, \ldots$. In general, if $(p_n)_{n\in\mathbb{N}}$ and $(q_k)_{k\in\mathbb{N}}$ are sequences and if there is a sequence $n_1 < n_2 < n_3 < \ldots$ of positive integers such that for each $k \in \mathbb{N}$ we have $q_k = p_{n_k}$ then (q_k) is a **subsequence** of (p_n). Note that the terms in the subsequence occur in the same order as in the mother sequence.

1 Theorem *Every subsequence of a convergent sequence in M converges and it converges to the same limit as does the mother sequence.*

Proof Let (q_k) be a subsequence of (p_n), $q_k = p_{n_k}$, where $n_1 < n_2 < \ldots$. Assume that (p_n) converges to p in M. Given $\epsilon > 0$, there is an N such that for all $n \geq N$ we have $d(p_n, p) < \epsilon$. Since $n_1, n_2, \ldots$ are positive integers we have $k \leq n_k$ for all k. Thus, if $k \geq N$ then $n_k \geq N$ and $d(q_k, p) < \epsilon$. Hence (q_k) converges to p. □

A common way to state Theorem 1 is that limits are unaffected when we pass to a subsequence.

2 Continuity

In linear algebra the objects of interest are linear transformations. In real analysis the objects of interest are functions, especially continuous functions. A function f from the metric space M to the metric space N is just that; $f : M \to N$ and f sends points $p \in M$ to points $fp \in N$. The function maps M to N. The way you should think of functions – as devices, not formulas – is discussed in Section 4 of Chapter 1. The most common type of function maps M to $\mathbb{R}$. It is a real-valued function of the variable $p \in M$.

Definition A function $f : M \to N$ is **continuous** if it **preserves sequential convergence**: f sends convergent sequences in M to convergent sequences in N, limits being sent to limits. That is, for each sequence (p_n) in M which converges to a limit p in M, the image sequence $(f(p_n))$ converges to fp in N.

Here and in what follows, the notation fp is often used as convenient shorthand for $f(p)$. The metrics on M and N are d_M and d_N. We will often refer to either metric as just d.

2 Theorem *The composite of continuous functions is continuous.*

Proof Let $f : M \to N$ and $g : N \to P$ be continuous and assume that

$$\lim_{n\to\infty} p_n = p$$

in M. Since f is continuous, $\lim_{n\to\infty} f(p_n) = fp$. Since g is continuous, $\lim_{n\to\infty} g(f(p_n)) = g(fp)$ and therefore $g \circ f : M \to P$ is continuous. See Figure 29 on the next page. □

Moral The sequence condition is the easy way to tell at a glance whether a function is continuous.

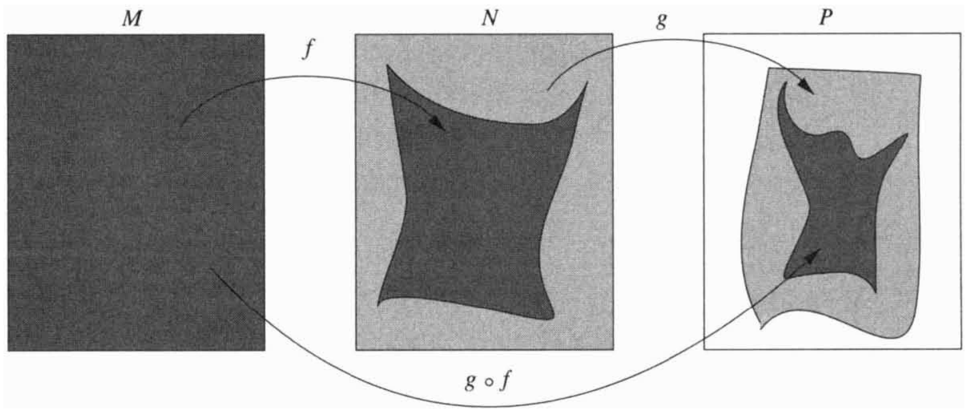

Figure 29 The composite function $g \circ f$

There are two "obviously" continuous functions.

3 Proposition *For every metric space M the identity map* $\mathrm{id} : M \to M$ *is continuous, and so is every constant function $f : M \to N$.*

Proof Let $p_n \to p$ in M. Then $\mathrm{id}(p_n) = p_n \to p = \mathrm{id}(p)$ as $n \to \infty$ which gives continuity of the identity map. Likewise, if $f(x) = q \in N$ for all $x \in M$ and if $p_n \to p$ in M then $fp = q$ and $f(p_n) = q$ for all n. Thus $f(p_n) \to fp$ as $n \to \infty$ which gives continuity of the constant function f. □

Homeomorphism

Vector spaces are isomorphic if there is a linear bijection from one to the other. When are metric spaces isomorphic? They should "look the same." The letters Y and T look the same; and they look different from the letter O. If $f : M \to N$ is a bijection and f is continuous and the inverse bijection $f^{-1} : N \to M$ is also continuous then f is a **homeomorphism**[†](or a "homeo" for short) and M, N are **homeomorphic**. We write $M \cong N$ to indicate that M and N are homeomorphic. $\cong$ is an equivalence relation: $M \cong M$ since the identity map is a homeomorphism $M \to M$; $M \cong N$ clearly implies that $N \cong M$; and the previous theorem shows that the composite of homeomorphisms is a homeomorphism.

Geometrically speaking, a homeomorphism is a bijection that can bend, twist, stretch, and wrinkle the space M to make it coincide with N, but it cannot rip,

[†]This is a rare case in mathematics in which spelling is important. Homeomorphism $\neq$ homomorphism.

puncture, shred, or pulverize M in the process. The basic questions to ask about metric spaces are:

(a) Given M and N, are they homeomorphic?
(b) What are the continuous functions from M to N?

A major goal of this chapter is to show you how to answer these questions in many cases. For example, is the circle homeomorphic to the interval? To the sphere? etc. Figure 30 indicates that the circle and the (perimeter of the) triangle are homeomorphic, while Figure 15 shows that (a, b), the semicircle, and $\mathbb{R}$ are homeomorphic.

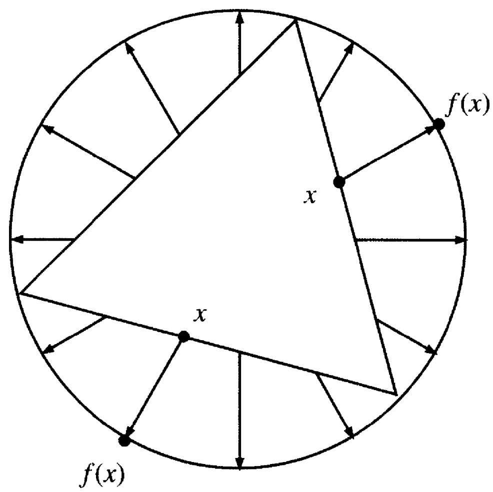

Figure 30 The circle and triangle are homeomorphic.

A natural question that should occur to you is whether continuity of f^{-1} is actually implied by continuity of a bijection f. It is not. Here is an instructive example.

Consider the interval $[0, 2\pi) = \{x \in \mathbb{R} : 0 \leq x < 2\pi\}$ and define $f : [0, 2\pi) \to S^1$ to be the mapping $f(x) = (\cos x, \sin x)$ where S^1 is the unit circle in the plane. The mapping f is a continuous bijection, but the inverse bijection is not continuous. For there is a sequence of points (z_n) on S^1 in the fourth quadrant that converges to $p = (1, 0)$ from below, and $f^{-1}(z_n)$ does not converge to $f^{-1}(p) = 0$. Rather it converges to 2π. Thus, f is a continuous bijection whose inverse bijection fails to be continuous. See Figure 31. In Exercises 49 and 50 you are asked to find worse examples of continuous bijections that are not homeomorphisms.

To build your intuition about continuous mappings and homeomorphisms, consider the following examples shown in Figure 32 – the unit circle in the plane, a trefoil knot in $\mathbb{R}^3$, the perimeter of a square, the surface of a donut (the 2-torus), the surface

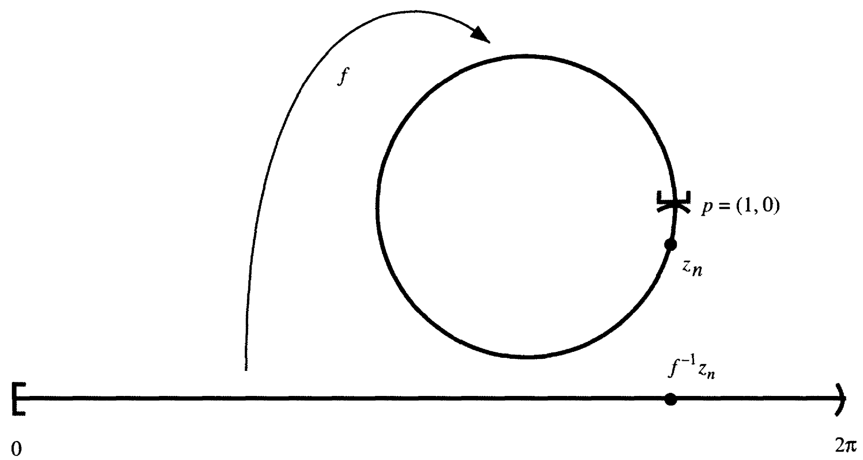

Figure 31 f wraps $[0, 2\pi)$ bijectively onto the circle.

of a ceramic coffee cup, the unit interval $[0, 1]$, the unit disc including its boundary. Equip all with the inherited metric. Which should be homeomorphic to which?

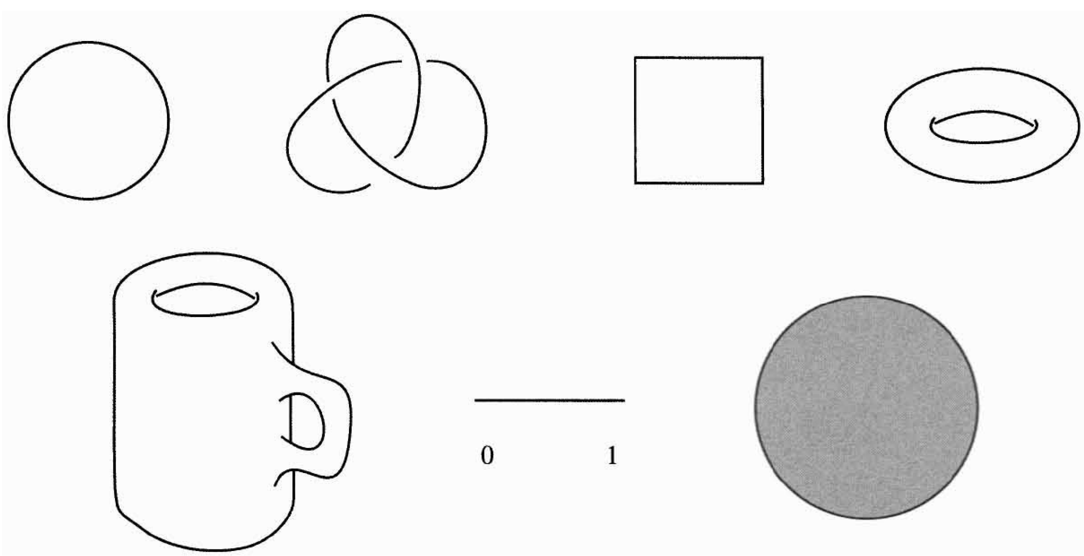

Figure 32 Seven metric spaces

The $(\boldsymbol{\epsilon}, \boldsymbol{\delta})$-Condition

The following theorem presents the more familiar (but equivalent!) definition of continuity using ϵ and δ. It corresponds to the definition given in Chapter 1 for real-valued functions of a real variable.

4 Theorem *$f : M \to N$ is continuous if and only if it satisfies the* $(\boldsymbol{\epsilon}, \boldsymbol{\delta})$***-condition****: For each $\epsilon > 0$ and each $p \in M$ there exists $\delta > 0$ such that if $x \in M$ and $d_M(x, p) < \delta$ then $d_N(fx, fp) < \epsilon$.*

Proof Suppose that f is continuous. It preserves sequential convergence. From the supposition that f fails to satisfy the (ϵ, δ)-condition at some $p \in M$ we will derive a contradiction. If the (ϵ, δ)-condition fails at p then there exists $\epsilon > 0$ such that for each $\delta > 0$ there is a point x with $d(x, p) < \delta$ and $d(fx, fp) \geq \epsilon$. Taking $\delta = 1/n$ we get a sequence (x_n) with $d(x_n, p) < 1/n$ and $d(f(x_n), fp) \geq \epsilon$, which contradicts preservation of sequential convergence. For $x_n \to p$ but $f(x_n)$ does not converge to fp, which contradicts the fact that f preserves sequential convergence. Since the supposition that f fails to satisfy the (ϵ, δ)-condition leads to a contradiction, f actually does satisfy the (ϵ, δ)-condition.

To check the converse, suppose that f satisfies the (ϵ, δ)-condition at p. For each sequence (x_n) in M that converges to p we must show $f(x_n) \to fp$ in N as $n \to \infty$. Let $\epsilon > 0$ be given. There is $\delta > 0$ such that $d_M(x, p) < \delta \Rightarrow d_N(fx, fp) < \epsilon$. Convergence of x_n to p implies there is an integer K such that for all $n \geq K$ we have $d_M(x_n, p) < \delta$, and therefore $d_N(f(x_n), fp) < \epsilon$. That is, $f(x_n) \to fp$ as $n \to \infty$. See also Exercise 13. $\square$

3 The Topology of a Metric Space

Now we come to two basic concepts in a metric space theory – closedness and openness. Let M be a metric space and let S be a subset of M. A point $p \in M$ is a **limit** of S if there exists a sequence (p_n) in S that converges to it.[†]

[†]A limit of S is also sometimes called a **limit point** of S. Take care though: Some mathematicians require that a limit point of S be the limit of a sequence of *distinct* points of S. They would say that a finite set has no limit points. We will *not* adopt their point of view. Another word used in this context, especially by the French, is "adherence." A point p **adheres** to the set S if and only if p is a limit of S. In more general circumstances, limits are defined using "nets" instead of sequences. They are like "uncountable sequences." You can read more about nets in graduate-level topology books such as *Topology* by James Munkres.

Definition S is a **closed set** if it contains all its limits.[†]

Definition S is an **open set** if for each $p \in S$ there exists an $r > 0$ such that

$$d(p,q) < r \quad \Rightarrow \quad q \in S.$$

5 Theorem *Openness is dual to closedness: The complement of an open set is a closed set and the complement of a closed set is an open set.*

Proof Suppose that $S \subset M$ is an open set. We claim that S^c is a closed set. If $p_n \to p$ and $p_n \in S^c$ we must show that $p \in S^c$. Well, if $p \notin S^c$ then $p \in S$ and, since S is open, there is an $r > 0$ such that

$$d(p,q) < r \quad \Rightarrow \quad q \in S.$$

Since $p_n \to p$, we have $d(p, p_n) < r$ for all large n, which implies that $p_n \in S$, contrary to the sequence being in S^c. Since the supposition that p lies in S leads to a contradiction, p actually does lie in S^c, proving that S^c is a closed set.

Suppose that S is a closed set. We claim that S^c is open. Take any $p \in S^c$. If there fails to exist an $r > 0$ such that

$$d(p,q) < r \quad \Rightarrow \quad q \in S^c$$

then for each $r = 1/n$ with $n = 1, 2, \ldots$ there exists a point $p_n \in S$ such that $d(p, p_n) < 1/n$. This sequence in S converges to $p \in S^c$, contrary to closedness of S. Therefore there actually does exist an $r > 0$ such that

$$d(p,q) < r \quad \Rightarrow \quad q \in S^c$$

which proves that S^c is an open set. $\square$

Most sets, like doors, are neither open nor closed, but ajar. Keep this in mind. For example neither $(a, b]$ nor its complement is closed in $\mathbb{R}$; $(a, b]$ is neither closed nor open. Unlike doors, however, sets can be both open and closed at the same time. For example, the empty set $\emptyset$ is a subset of every metric space and it is always closed. There are no sequences and no limits to even worry about. Similarly the full metric space M is a closed subset of itself: For it certainly contains the limit of

[†]Note how similarly algebraists use the word "closed." A field (or group or ring, etc.) is closed under its arithmetic operations: Sums, differences, products, and quotients of elements in the field still lie in the field. In our case it is limits. Limits of sequences in S must lie in S.

every sequence that converges in M. Thus, $\emptyset$ and M are closed subsets of M. Their complements, M and $\emptyset$, are therefore open: $\emptyset$ and M are both closed and open.

Subsets of M that are both closed and open are **clopen**. See also Exercise 125. It turns out that in $\mathbb{R}$ the only clopen sets are $\emptyset$ and $\mathbb{R}$. In $\mathbb{Q}$, however, things are quite different, sets such as $\{r \in \mathbb{Q} : -\sqrt{2} < r < \sqrt{2}\}$ being clopen in $\mathbb{Q}$. To summarize,

A subset of a metric space can be
closed, open, both, or neither.

You should expect the "typical" subset of a metric space to be neither closed nor open.

The **topology** of M is the collection $\mathcal{T}$ of all open subsets of M.

6 Theorem $\mathcal{T}$ *has three properties:*[†] *as a system it is closed under union, finite intersection, and it contains $\emptyset$, M. That is,*

(a) Every union of open sets is an open set.
(b) The intersection of finitely many open sets is an open set.
(c) $\emptyset$ and M are open sets.

Proof (a) If $\{U_\alpha\}$ is any collection[‡] of open subsets of M and $V = \bigcup U_\alpha$ then V is open. For if $p \in V$ then p belongs to at least one U_α and there is an $r > 0$ such that

$$d(p, q) < r \quad \Rightarrow \quad q \in U_\alpha.$$

Since $U_\alpha \subset V$, this implies that all such q lie in V, proving that V is open.

(b) If $U_1, \dots, U_n$ are open sets and $W = \bigcap U_k$ then W is open. For if $p \in W$ then for each k, $1 \le k \le n$, then there is an $r_k > 0$ such that

$$d(p, q) < r_k \quad \Rightarrow \quad q \in U_k.$$

Take $r = \min\{r_1, \dots, r_n\}$. Then $r > 0$ and

$$d(p, q) < r \quad \Rightarrow \quad q \in U_k,$$

[†] Any collection $\mathcal{T}$ of subsets of a set X that satisfies these three properties is called a topology on X, and X is called a **topological space**. Topological spaces are more general than metric spaces: There exist topologies that do not arise from a metric. Think of them as pathological. The question of which topologies can be generated by a metric and which cannot is discussed in *Topology* by Munkres. See also Exercise 30.

[‡] The α in the notation U_α "indexes" the sets. If $\alpha = 1, 2, \dots$ then the collection is countable, but we are just as happy to let α range through uncountable index sets.

for each k; i.e., $q \in W = \bigcap U_k$, proving that W is open.

(c) It is clear that $\emptyset$ and M are open sets. □

7 Corollary *The intersection of any number of closed sets is a closed set; the finite union of closed sets is a closed set; $\emptyset$ and M are closed sets.*

Proof Take complements and use De Morgan's laws. If $\{K_\alpha\}$ is a collection of closed sets then $U_\alpha = (K_\alpha)^c$ is open and

$$K = \bigcap K_\alpha = (\bigcup U_\alpha)^c .$$

Since $\bigcup U_\alpha$ is open, its complement K is closed. Similarly, a finite union of closed sets is the complement of the finite intersection of their complements, and is a closed set. □

What about an infinite union of closed sets? Generally, it is not closed. For example, the interval $[1/n, 1]$ is closed in $\mathbb{R}$, but the union of these intervals as n ranges over $\mathbb{N}$ is the interval $(0, 1]$ which is not closed in $\mathbb{R}$. Neither is the infinite intersection of open sets open in general.

Two sets whose closedness/openness properties are basic are:

$$\begin{aligned} \lim S &= \{p \in M : p \text{ is a limit of } S\} \\ M_r p &= \{q \in M : d(p,q) < r\}. \end{aligned}$$

The former is the **limit set** of S; the latter is the ***r*-neighborhood** of p.

8 Theorem $\lim S$ *is a closed set and* $M_r p$ *is an open set.*

Proof Simple but not immediate! See Figure 33.

Suppose that $p_n \to p$ and each p_n lies in $\lim S$. We claim that $p \in \lim S$. Since p_n is a limit of S there is a sequence $(p_{n,k})_{k\in\mathbb{N}}$ in S that converges to p_n as $k \to \infty$. Thus there exists $q_n = p_{n,k(n)} \in S$ such that

$$d(p_n, q_n) < \frac{1}{n}.$$

Then, as $n \to \infty$ we have

$$d(p, q_n) \le d(p, p_n) + d(p_n, q_n) \to 0$$

which implies that $q_n \to p$, so $p \in \lim S$, which completes the proof that $\lim S$ is a closed set.

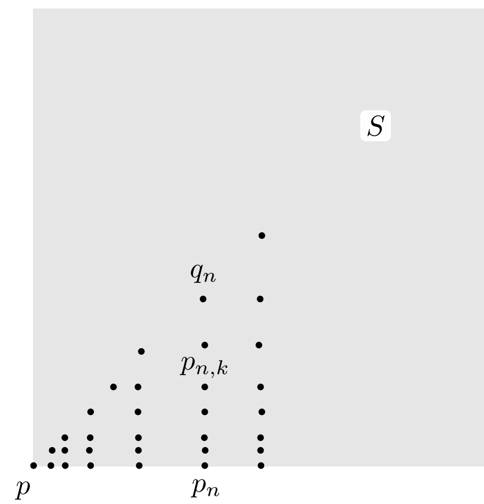

Figure 33 $S = (0,1) \times (0,1)$ and $p_n = (1/n, 0)$ converges to $p = (0,0)$ as $n \to \infty$. Each p_n is the limit of the sequence $p_{n,k} = (1/n, 1/k)$ as $k \to \infty$. The sequence $q_n = (1/n, 1/n)$ lies in S and converges to $(0,0)$. Hence: *The limits of limits are limits.*

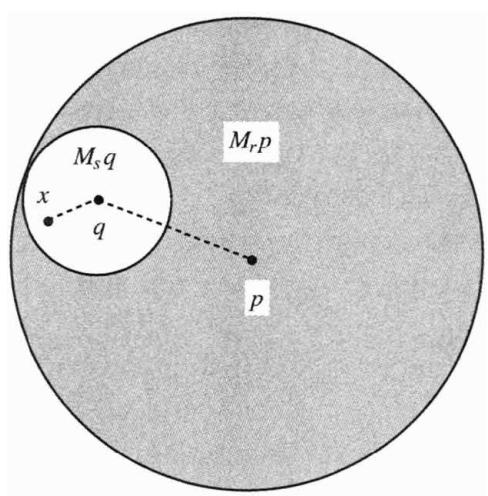

Figure 34 Why the r-neighborhood of p is an open set

To check that $M_r p$ is an open set, take any $q \in M_r p$ and observe that

$$s = r - d(p,q) > 0.$$

By the triangle inequality, if $d(q,x) < s$ then

$$d(p,x) \le d(p,q) + d(q,x) < r,$$

and hence $M_s q \subset M_r p$. See Figure 34. Since each $q \in M_r p$ has some $M_s q$ that is contained in $M_r p$, $M_r p$ is an open set. $\square$

9 Corollary *The interval (a,b) is open in $\mathbb{R}$ and so are $(-\infty,b)$, (a,∞), and $(-\infty,\infty)$. The interval $[a,b]$ is closed in $\mathbb{R}$.*

Proof (a,b) is the r-neighborhood of its midpoint $m = (a+b)/2$ where $r = (b-a)/2$. Therefore (a,b) is open in $\mathbb{R}$. Since the union of open sets is open we see that

$$\bigcup_{n\in\mathbb{N}} (b-n,\, b-1/n) = (-\infty,b)$$

is open. The same applies to (a,∞). The whole metric space $\mathbb{R} = (-\infty,\infty)$ is always open in itself.

Since the complement of $[a,b]$ is the open set $(-\infty,a) \cup (b,\infty)$, the interval $[a,b]$ is closed. $\square$

10 Corollary $\lim S$ *is the "smallest" closed set that contains S in the sense that if $K \supset S$ and K is closed then $K \supset \lim S$.*

Proof Obvious. K must contain the limit of each sequence in K that converges in M and therefore it must contain the limit of each sequence in $S \subset K$ that converges in M. These limits are exactly $\lim S$. $\square$

We refer to $\lim S$ as the **closure** of S and denote it also as $\overline{S}$. You start with S and make it closed by adding all its limits. You don't need to add any more points because limits of limits are limits. That is, $\lim(\lim S) = \lim S$. An operation like this is called **idempotent**. Doing the operation twice produces the same outcome as doing it once.

A **neighborhood** of a point p in M is any open set V that contains p. Theorem 8 implies that $V = M_r p$ is a neighborhood of p. Eventually, you will run across the phrase "closed neighborhood" of p, which refers to a closed set that contains an open set that contains p. However, until further notice all neighborhoods are open.

Usually, sets defined by strict inequalities are open while those defined by equalities or nonstrict inequalities are closed. Examples of closed sets in $\mathbb{R}$ are finite sets, $[a, b]$, $\mathbb{N}$, and the set $\{0\} \cup \{1/n : n \in N\}$. Each contains all its limits. Examples of open sets in $\mathbb{R}$ are open intervals, bounded or not.

Topological Description of Continuity

A property of a metric space or of a mapping between metric spaces that can be described solely in terms of open sets (or equivalently, in terms of closed sets) is called a **topological property**. The next result describes continuity topologically.

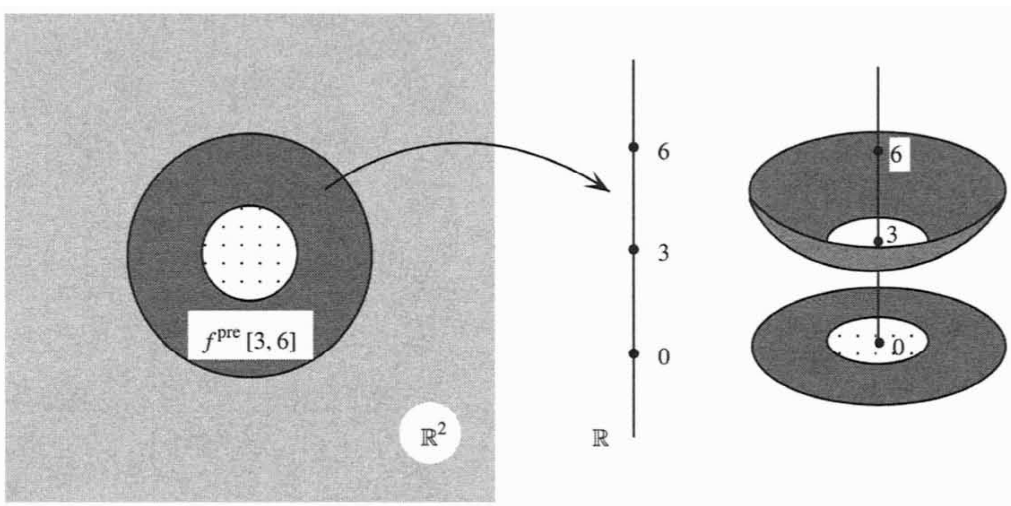

Figure 35 The function $f : (x, y) \mapsto x^2 + y^2 + 2$ and its graph over the preimage of $[3, 6]$

Let $f : M \to N$ be given. The **preimage**† of a set $V \subset N$ is

$$f^{\text{pre}}(V) = \{p \in M : f(p) \in V\}.$$

For example, if $f : \mathbb{R}^2 \to \mathbb{R}$ is the function defined by the formula

$$f(x, y) = x^2 + y^2 + 2$$

then the preimage of the interval $[3, 6]$ in $\mathbb{R}$ is the annulus in the plane with inner radius 1 and outer radius 2. Figure 35 shows the domain of f as $\mathbb{R}^2$ and the target

†The preimage of V is also called the **inverse image** of V and is denoted by $f^{-1}(V)$. Unless f is a bijection, this notation leads to confusion. There may be no map f^{-1} and yet expressions like $V \supset f(f^{-1}(V))$ are written that mix maps and nonmaps. By the way, if f sends no point of M into V then $f^{\text{pre}}(V)$ is the empty set.

as $\mathbb{R}$. The range is the set of real numbers ≥ 2. The graph of f is a paraboloid with lowest point $(0, 0, 2)$. The second part of the figure shows the portion of the graph lying above the annulus. You will find it useful to keep in mind the distinctions among the concepts: function, range, and graph.

11 Theorem *The following are equivalent for continuity of $f : M \to N$.*

(i) The (ϵ, δ)-condition.
(ii) The sequential convergence preservation condition.
*(iii) The **closed set condition**: The preimage of each closed set in N is closed in M.*
*(iv) The **open set condition**: The preimage of each open set in N is open in M.*

Proof Totally natural! By Theorem 4, (i) implies (ii).

(ii) implies (iii). If $K \subset N$ is closed in N and $p_n \in f^{\text{pre}}(K)$ converges to $p \in M$ then we claim that $p \in f^{\text{pre}}(K)$. By (ii), f preserves sequential convergence so

$$\lim_{n\to\infty} f(p_n) \;=\; fp.$$

Since K is closed in N, $fp \in K$, so $p \in f^{\text{pre}}(K)$, and we see that $f^{\text{pre}}(K)$ is closed in M. Thus (ii) implies (iii).

(iii) implies (iv). This follows by taking complements: $(f^{\text{pre}}(U))^c = f^{\text{pre}}(U^c)$.

(iv) implies (i). Let $\epsilon > 0$ and $p \in M$ be given. $N_\epsilon(fp)$ is open in N, so its preimage $U = f^{\text{pre}}(N_\epsilon(fp))$ is open in M. The point p belongs to the preimage so openness of U implies there is a $\delta > 0$ such that $M_\delta(p) \subset U$. Then

$$f(M_\delta(p)) \;\subset\; fU \;\subset\; N_\epsilon(fp)$$

gives the ϵ, δ condition, $d_M(p, x) < \delta \;\Rightarrow\; d_N(fp, fx) < \epsilon$. See Figure 36. $\square$

I hope you find the closed and open set characterizations of continuity elegant. Note that no explicit mention is made of the metric. The open set condition is purely topological. It would be perfectly valid to take as a *definition* of continuity that the preimage of each open set is open. In fact this is exactly what's done in general topology.

12 Corollary *A homeomorphism $f : M \to N$ bijects the collection of open sets in M to the collection of open sets in N. It bijects the topologies.*

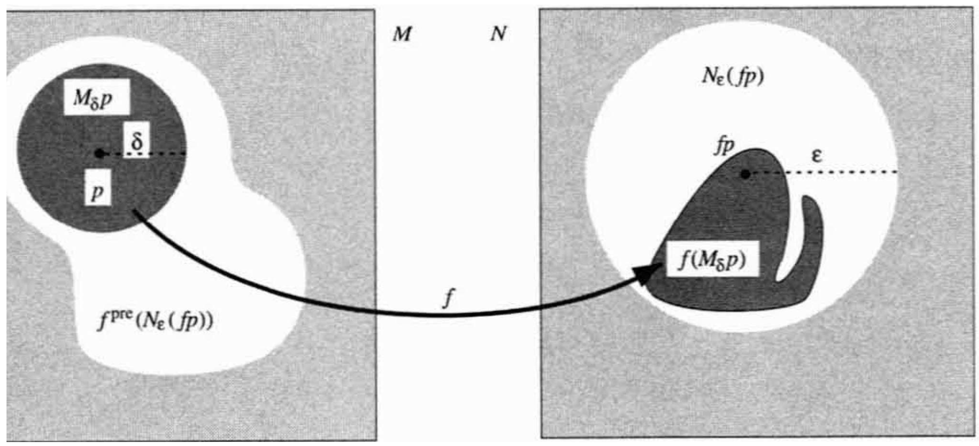

Figure 36 The ϵ, δ - condition for a continuous function $f : M \to N$

Proof Let V be an open set in N. By Theorem 11, since f is continuous, the preimage of V is open in M. Since f is a bijection, this preimage $U = \{p \in M : fp \in V\}$ is exactly the image of V by the inverse bijection, $U = f^{-1}(V)$. The same thing can be said about f^{-1} since f^{-1} is also a homeomorphism. That is, $V = fU$. Thus, sending U to fU bijects the topology of M to the topology of N. □

Because of this corollary, a homeomorphism is also called a **topological equivalence**.

In general, continuous maps do not need to send open sets to open sets. For example, the squaring map $x \mapsto x^2$ from $\mathbb{R}$ to $\mathbb{R}$ is continuous but it sends the open interval $(-1, 1)$ to the nonopen interval $[0, 1)$. See also Exercise 28.

Inheritance

If a set S is contained in a metric subspace $N \subset M$ you need to be careful when you say that S is open or closed. For example,

$$S = \{x \in \mathbb{Q} : -\pi < x < \pi\}$$

is a subset of the metric subspace $\mathbb{Q} \subset \mathbb{R}$. It is both open and closed with respect to $\mathbb{Q}$ but is neither open nor closed with respect to $\mathbb{R}$. To avoid this kind of ambiguity it is best to say that S is clopen "with respect to $\mathbb{Q}$ but not with respect to $\mathbb{R}$," or more briefly that S is clopen "in $\mathbb{Q}$ but not in $\mathbb{R}$." Nevertheless there is a simple relation between the topologies of M and N when N is a metric subspace of M.

13 Inheritance Principle *Every metric subspace N of M* ***inherits its topology*** *from M in the sense that each subset $V \subset N$ which is open in N is actually the intersection $V = N \cap U$ for some $U \subset M$ that is open in M, and vice versa.*

Proof It all boils down to the fact that for each $p \in N$ we have

$$N_r p = N \cap M_r p.$$

After all, $N_r p$ is the set of $x \in N$ such that $d_N(x, p) < r$ and this is exactly the same as the set of those $x \in M_r p$ that belong to N. Therefore N inherits its r-neighborhoods from M. Since its open sets are unions of its r-neighborhoods, N also inherits its open sets from M.

In more detail, if V is open in N then it is the union of those $N_r p$ with $N_r p \subset V$. Each such $N_r p$ is $N \cap M_r p$ and the union of these $M_r p$ is U, an open subset of M. The intersection $N \cap U$ equals V. Conversely, if U is any open subset of M and $p \in V = N \cap U$ then openness of U implies there is an $M_r p \subset U$. Thus $N_r p = N \cap M_r p \subset V$, which shows that V is open in N. □

14 Corollary *Every metric subspace of M inherits its closed sets from M.*

Proof By "inheriting its closed sets" we mean that each closed subset $L \subset N$ is the intersection $N \cap K$ for some closed subset $K \subset M$ and vice versa. The proof consists of two words: "Take complements." See also Exercise 34. □

Let's return to the example with $\mathbb{Q} \subset \mathbb{R}$ and $S = \{x \in \mathbb{Q} : -\pi < x < \pi\}$. The set S is clopen in $\mathbb{Q}$, so we should have $S = \mathbb{Q} \cap U$ for some open set $U \subset \mathbb{R}$ and $S = \mathbb{Q} \cap K$ for some closed set $K \subset \mathbb{R}$. In fact U and K are

$$U = (-\pi, \pi) \quad \text{and} \quad K = [-\pi, \pi].$$

15 Corollary *Assume that N is a metric subspace of M and also is a closed subset of M. A set $L \subset N$ is closed in N if and only if it is closed in M. Similarly, if N is a metric subspace of M and also is an open subset of M then $U \subset N$ is open in N if and only if it is open in M.*

Proof The proof is left to the reader as Exercise 34. □

Product Metrics

We next define a metric on the Cartesian product $M = X \times Y$ of two metric spaces. There are three natural ways to do so:

$$\begin{aligned} d_E(p,p') &= \sqrt{d_X(x,x')^2 + d_Y(y,y')^2} \\ d_{\max}(p,p') &= \max\{d_X(x,x'),\, d_Y(y,y')\} \\ d_{\text{sum}}(p,p') &= d_X(x,x') + d_Y(y,y') \end{aligned}$$

where $p = (x,y)$ and $p' = (x',y')$ belong to M. (d_E is the **Euclidean product metric**.) The proof that these expressions actually define metrics on M is left as Exercise 38.

16 Proposition $d_{\max} \le d_E \le d_{\text{sum}} \le 2d_{\max}$.

Proof Dropping the smaller term inside the square root shows that $d_{\max} \le d_E$; comparing the square of d_E and the square of d_{sum} shows that the latter has the terms of the former and the cross term besides, so $d_E \le d_{\text{sum}}$; and clearly d_{sum} is no larger than twice its greater term, so $d_{\text{sum}} \le 2d_{\max}$. □

17 Convergence in a Product Space *The following are equivalent for a sequence $p_n = (p_{1n}, p_{2n})$ in $M = M_1 \times M_2$:*

(a) (p_n) converges with respect to the metric d_{max}.
(b) (p_n) converges with respect to the metric d_E.
(c) (p_n) converges with respect to the metric d_{sum}.
(d) (p_{1n}) and (p_{2n}) converge in M_1 and M_2 respectively.

Proof This is immediate from Proposition 16. □

18 Corollary *If $f : M \to N$ and $g : X \to Y$ are continuous then so is their Cartesian product $f \times g$ which sends $(p,x) \in M \times X$ to $(fp, gx) \in N \times Y$.*

Proof If $(p_n, x_n) \to (p,x)$ in $M \times X$ then Theorem 17 implies $p_n \to p$ in M and $x_n \to x$ in X. By continuity, $f(p_n) \to fp$ and $g(x_n) \to gx$. By Theorem 17, $(f(p_n), g(x_n)) \to (fp, gx)$ which gives continuity of $f \times g$. □

The three metrics make sense in the obvious way for a Cartesian product of $m \ge 3$ metric spaces. The inequality

$$d_{\max} \le d_E \le d_{\text{sum}} \le m d_{\max}.$$

is proved in the same way, and we see that Theorem 17 holds also for the product of m metric spaces. This gives

19 Corollary (Convergence in $\mathbb{R}^m$) *A sequence of vectors (v_n) in $\mathbb{R}^m$ converges in $\mathbb{R}^m$ if and only if each of its component sequences (v_{in}) converges, $1 \le i \le m$. The limit of the vector sequence is the vector*

$$v = \lim_{n\to\infty} v_n = \left(\lim_{n\to\infty} v_{1n}, \ \lim_{n\to\infty} v_{2n}, \ldots, \lim_{n\to\infty} v_{mn}\right).$$

The distance function $d : M \times M \to \mathbb{R}$ is a function from the metric space $M \times M$ to the metric space $\mathbb{R}$, so the following assertion makes sense.

20 Theorem *d is continuous.*

Proof We check (ϵ, δ)-continuity with respect to the metric d_{sum}. Given $\epsilon > 0$ we take $\delta = \epsilon$. If $d_{\text{sum}}((p,q),(p',q')) < \delta$ then the triangle inequality gives

$$\begin{array}{rcccl} d(p,q) & \le & d(p,p') + d(p',q') + d(q',q) & < & d(p',q') + \epsilon \\ d(p',q') & \le & d(p',p) + d(p,q) + d(q,q') & < & d(p,q) + \epsilon \end{array}$$

which gives

$$d(p,q) - \epsilon \ < \ d(p',q') \ < \ d(p,q) + \epsilon.$$

Thus $|d(p',q') - d(p,q)| < \epsilon$ and we get continuity with respect to the metric d_{sum}. By Theorem 17 it does not matter which metric we use on $\mathbb{R} \times \mathbb{R}$. $\square$

As you can see, the use of d_{sum} simplifies the proof by avoiding square root manipulations. The sum metric is also called the **Manhattan metric** or the **taxicab metric**. Figure 37 shows the "unit discs" with respect to these metrics in $\mathbb{R}^2$. See also Exercise 2.

21 Corollary *The metrics $d_{\max}$, d_E, and d_{sum} are continuous.*

Proof Theorem 20 asserts that all metrics are continuous. $\square$

22 Corollary *The absolute value is a continuous mapping $\mathbb{R} \to \mathbb{R}$. In fact the norm is a continuous mapping from any normed space to $\mathbb{R}$.*

Proof $\|v\| = d(v, 0)$. $\square$

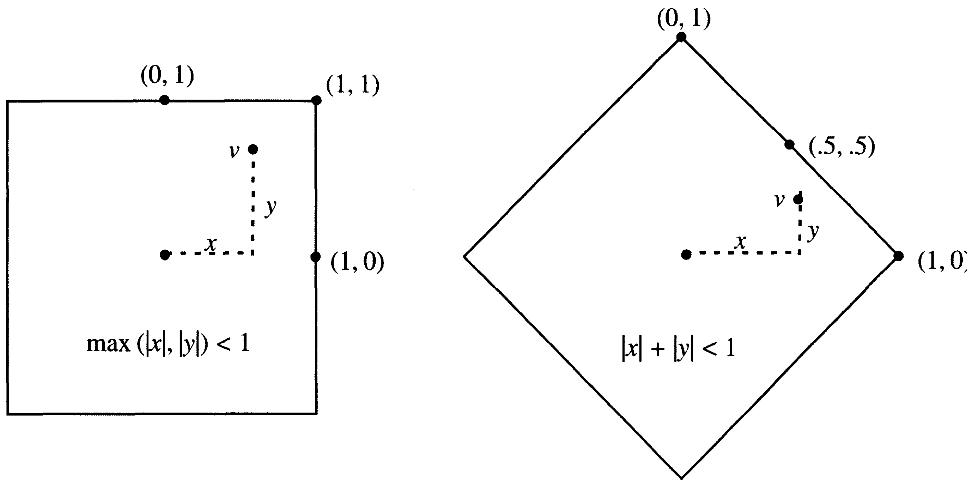

Figure 37 The unit disc in the max metric is a square, and in the sum metric it is a rhombus.

Completeness

In Chapter 1 we discussed the Cauchy criterion for convergence of a sequence of real numbers. There is a natural way to carry these ideas over to a metric space M. The sequence (p_n) in M satisfies a **Cauchy condition** provided that for each $\epsilon > 0$ there is an integer N such that for all $k, n \geq N$ we have $d(p_k, p_n) < \epsilon$, and (p_n) is said to be a **Cauchy sequence**. In symbols,

$$\forall \epsilon > 0 \ \exists N \text{ such that } k, n \geq N \ \Rightarrow \ d(p_k, p_n) < \epsilon.$$

The terms of a Cauchy sequence "bunch together" as $n \to \infty$. Each convergent sequence (p_n) is Cauchy. For if (p_n) converges to p as $n \to \infty$ then, given $\epsilon > 0$, there is an N such that for all $n \geq N$ we have

$$d(p_n, p) < \frac{\epsilon}{2}.$$

By the triangle inequality, if $k, n \geq N$ then

$$d(p_k, p_n) \ \leq \ d(p_k, p) + d(p, p_n) < \epsilon,$$

so convergence $\Rightarrow$ Cauchy.

Theorem 1.5 states that the converse is true in the metric space $\mathbb{R}$. Every Cauchy sequence in $\mathbb{R}$ converges to a limit in $\mathbb{R}$. In the general metric space, however, this need not be true. For example, consider the metric space $\mathbb{Q}$ of rational numbers, equipped with the inherited metric $d(x, y) = |x - y|$, and consider the sequence

$$(r_n) = (1.4,\ 1.41,\ 1.414,\ 1.4142,\ \ldots).$$

It is Cauchy. Given $\epsilon > 0$, choose $N > -\log_{10} \epsilon$. If $k, n \geq N$ then $|r_k - r_n| \leq 10^{-N} < \epsilon$. Nevertheless, (r_n) refuses to converge in $\mathbb{Q}$. After all, as a sequence in $\mathbb{R}$ it converges to $\sqrt{2}$, and if it also converges to some $r \in \mathbb{Q}$, then by uniqueness of limits in $\mathbb{R}$ we have $r = \sqrt{2}$, something we know is false. In brief, convergence $\Rightarrow$ Cauchy but not conversely.

A metric space M is **complete** if each Cauchy sequence in M converges to a limit in M. Theorem 1.5 states that $\mathbb{R}$ is complete.

23 Theorem $\mathbb{R}^m$ *is complete.*

Proof Let (p_n) be a Cauchy sequence in $\mathbb{R}^m$. Express p_n in components as

$$p_n = (p_{1n}, \ldots,\ p_{mn}).$$

Because (p_n) is Cauchy, each component sequence $(p_{in})_{n \in \mathbb{N}}$ is Cauchy. Completeness of $\mathbb{R}$ implies that the component sequences converge, and therefore the vector sequence converges. $\square$

24 Theorem *Every closed subset of a complete metric space is a complete metric subspace.*

Proof Let A be a closed subset of the complete metric space M and let (p_n) be a Cauchy sequence in A with respect to the inherited metric. It is of course also a Cauchy sequence in M and therefore it converges to a limit p in M. Since A is closed we have $p \in A$. $\square$

25 Corollary *Every closed subset of Euclidean space is a complete metric space.*

Proof Obvious from the previous theorem and completeness of $\mathbb{R}^m$. $\square$

Remark Completeness is *not* a topological property. For example, consider $\mathbb{R}$ with its usual metric and $(-1, 1)$ with the metric it inherits from $\mathbb{R}$. Although they are homeomorphic metric spaces, $\mathbb{R}$ is complete but $(-1, 1)$ is not.

In Section 10 we explain how every metric space can be completed.

4 Compactness

Compactness is the single most important concept in real analysis. It is what reduces the infinite to the finite.

Definition A subset A of a metric space M is (sequentially) **compact** if every sequence (a_n) in A has a subsequence (a_{n_k}) that converges to a limit in A.

The empty set and finite sets are trivial examples of compact sets. For a sequence (a_n) contained in a finite set repeats a term infinitely often, and the corresponding constant subsequence converges.

Compactness is a *good* feature of a set. We will develop criteria to decide whether a set is compact. The first is the most often used, but beware! − its converse is generally false.

26 Theorem *Every compact set is closed and bounded.*

Proof Suppose that A is a compact subset of the metric space M and that p is a limit of A. Does p belong to A? There is a sequence (a_n) in A converging to p. By compactness, some subsequence (a_{n_k}) converges to some $q \in A$, but every subsequence of a convergent sequence converges to the same limit as does the mother sequence, so $q = p$ and $p \in A$. Thus A is closed.

To see that A is bounded, choose and fix any point $p \in M$. Either A is bounded or else for each $n \in \mathbb{N}$ there is a point $a_n \in A$ such that $d(p, a_n) \geq n$. Compactness implies that some subsequence (a_{n_k}) converges. Convergent sequences are bounded, which contradicts the fact that $d(p, a_{n_k}) \to \infty$ as $k \to \infty$. Therefore (a_n) cannot exist and for some large r we have $A \subset M_r p$, which is what it means that A is bounded.□

27 Theorem *The closed interval* $[a, b] \subset \mathbb{R}$ *is compact.*

Proof Let (x_n) be a sequence in $[a, b]$ and set

$$C = \{x \in [a, b] : x_n < x \text{ only finitely often}\}.$$

Equivalently, for all but finitely many n, $x_n \geq x$. Since $a \in C$ we know that $C \neq \emptyset$. Clearly b is an upper bound for C. By the least upper bound property of $\mathbb{R}$ there exists $c = \text{l.u.b.}\, C$ with $c \in [a, b]$. We claim that a subsequence of (x_n) converges to c. Suppose not, i.e., no subsequence of (x_n) converges to c. Then for some $r > 0$, x_n lies in $(c - r,\ c + r)$ only finitely often, which implies that $c + r \in C$, contrary to c being an upper bound for C. Hence some subsequence of (x_n) does converge to c and $[a, b]$ is compact. □

To pass from $\mathbb{R}$ to $\mathbb{R}^m$ we think about compactness for Cartesian products.

28 Theorem *The Cartesian product of two compact sets is compact.*

Proof Let $(a_n, b_n) \in A \times B$ be given where $A \subset M$ and $B \subset N$ are compact. There exists a subsequence (a_{n_k}) that converges to some point $a \in A$ as $k \to \infty$. The subsequence (b_{n_k}) has a sub-subsequence $(b_{n_{k(\ell)}})$ that converges to some $b \in B$ as $\ell \to \infty$. The sub-subsequence $(a_{n_{k(\ell)}})$ continues to converge to the point a. Thus

$$(a_{n_{k(\ell)}}, b_{n_{k(\ell)}}) \to (a, b)$$

as $\ell \to \infty$. This implies that $A \times B$ is compact. □

29 Corollary *The Cartesian product of m compact sets is compact.*

Proof Write $A_1 \times A_2 \times \cdots \times A_m = A_1 \times (A_2 \times \cdots \times A_m)$ and perform induction on m. (Theorem 28 handles the bottom case $m = 2$.) □

30 Corollary *Every box $[a_1, b_1] \times \cdots \times [a_m, b_m]$ in $\mathbb{R}^m$ is compact.*

Proof Obvious from Theorem 27 and the previous corollary. □

An equivalent formulation of these results is the

31 Bolzano-Weierstrass Theorem *Every bounded sequence in $\mathbb{R}^m$ has a convergent subsequence.*

Proof A bounded sequence is contained in a box, which is compact, and therefore the sequence has a subsequence that converges to a limit in the box. See also Exercise 11. □

Here is a simple fact about compacts.

32 Theorem *Every closed subset of a compact is compact.*

Proof If A is a closed subset of the compact set K and if (a_n) is a sequence of points in A then clearly (a_n) is also a sequence of points in K, so by compactness of K there is a subsequence (a_{n_k}) converging to a limit $p \in K$. Since A is closed, p lies in A which proves that A is compact. □

Now we come to the first partial converse to Theorem 26.

33 Heine-Borel Theorem *Every closed and bounded subset of* $\mathbb{R}^m$ *is compact.*

Proof Let $A \subset \mathbb{R}^m$ be closed and bounded. Boundedness implies that A is contained in some box, which is compact. Since A is closed, Theorem 32 implies that A is compact. See also Exercise 11. □

The Heine-Borel Theorem states that closed and bounded subsets of Euclidean space are compact, but it is *vital*† to remember that a closed and bounded subset of a general metric space may fail to be compact. For example, the set $\mathbb{N}$ of natural numbers equipped with the discrete metric is a complete metric space, it is closed in itself (as is every metric space), and it is bounded. But it is not compact. After all, what subsequence of $1, 2, 3, \ldots$ converges?

A more striking example occurs in the metric space $C([0,1], \mathbb{R})$ whose metric is $d(f, g) = \max\{|f(x) - g(x)|\}$. The space is complete but its closed unit ball is not compact. For example, the sequence of functions $f_n = x^n$ has no subsequence that converges with respect to the metric d. In fact every closed ball is noncompact.

Ten Examples of Compact Sets

1. Any finite subset of a metric space, for instance the empty set.
2. Any closed subset of a compact set.
3. The union of finitely many compact sets.
4. The Cartesian product of finitely many compact sets.
5. The intersection of arbitrarily many compact sets.
6. The closed unit ball in $\mathbb{R}^3$.
7. The boundary of a compact set, for instance the unit 2-sphere in $\mathbb{R}^3$.
8. The set $\{x \in \mathbb{R} : \exists\, n \in \mathbb{N} \text{ and } x = 1/n\} \cup \{0\}$.
9. The Hawaiian earring. See page 58.
10. The Cantor set. See Section 8.

Nests of Compacts

If $A_1 \supset A_2 \supset \cdots \supset A_n \supset A_{n+1} \supset \ldots$ then (A_n) is a **nested sequence** of sets. Its intersection is

$$\bigcap_{n=1}^{\infty} A_n = \{p : \text{for each } n \text{ we have } p \in A_n\}.$$

†I have asked variants of the following True or False question on every analysis exam I've given: "Every closed and bounded subset of a complete metric space is compact." You would be surprised at how many students answer "True."

See Figure 38.

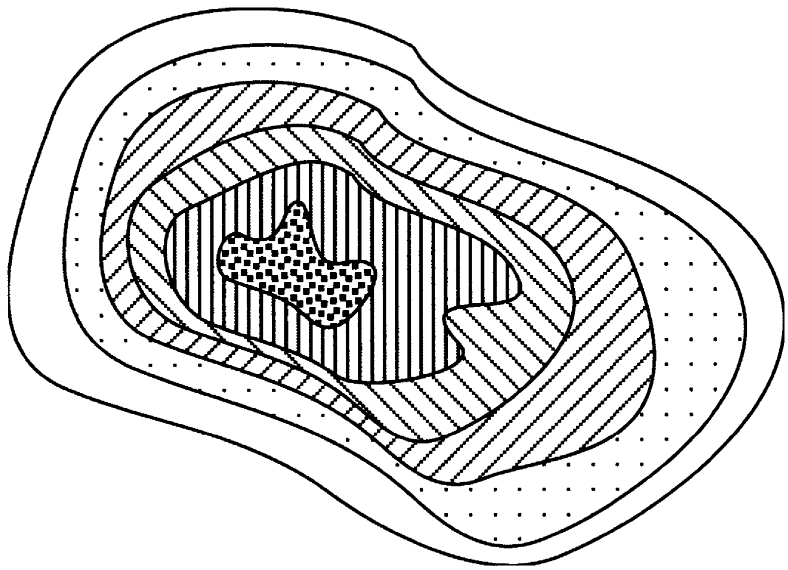

Figure 38 A nested sequence of sets

For example, we could take A_n to be the disc $\{z \in \mathbb{R}^2 : |z| \leq 1/n\}$. The intersection of all the sets A_n is then the singleton $\{0\}$. On the other hand, if A_n is the ball $\{z \in \mathbb{R}^3 : |z| \leq 1 + 1/n\}$ then $\bigcap A_n$ is the closed unit ball B^3.

34 Theorem *The intersection of a nested sequence of compact nonempty sets is compact and nonempty.*

Proof Let (A_n) be such a sequence. By Theorem 26, A_n is closed. The intersection of closed sets is always closed. Thus, $\bigcap A_n$ is a closed subset of the compact set A_1 and is therefore compact. It remains to show that the intersection is nonempty.

A_n is nonempty, so for each $n \in \mathbb{N}$ we can choose $a_n \in A_n$. The sequence (a_n) lies in A_1 since the sets are nested. Compactness of A_1 implies that (a_n) has a subsequence (a_{n_k}) converging to some point $p \in A_1$. The limit p also lies in the set A_2 since except possibly for the first term, the subsequence (a_{n_k}) lies in A_2 and A_2 is a closed set. The same is true for A_3 and for all the sets in the nested sequence. Thus, $p \in \bigcap A_n$ and $\bigcap A_n$ is shown to be nonempty. $\square$

The **diameter** of a nonempty set $S \subset M$ is the supremum of the distances $d(x, y)$ between points of S.

35 Corollary *If in addition to being nested, nonempty, and compact, the sets A_n have diameter that tends to* 0 *as $n \to \infty$ then $A = \bigcap A_n$ is a single point.*

Proof For each $n \in \mathbb{N}$, A is a subset of A_n, which implies that A has diameter zero. Since distinct points lie at positive distance from each other, A consists of at most one point, while by Theorem 34 it consists of at least one point. See also Exercise 52. $\square$

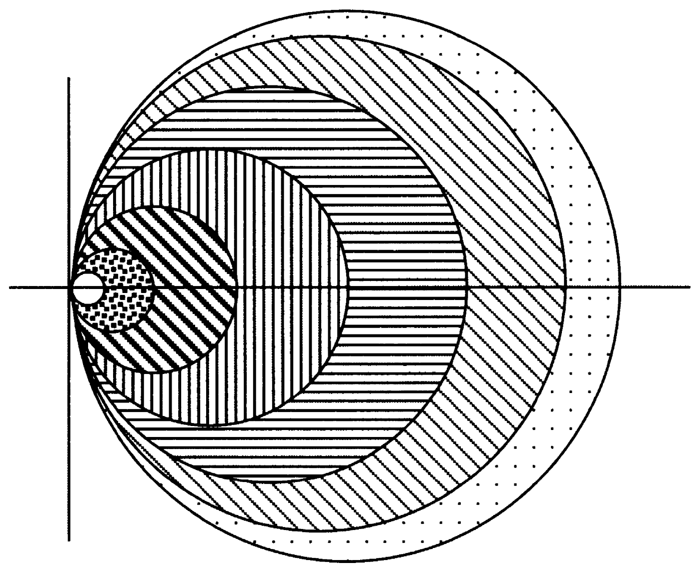

Figure 39 This nested sequence has empty intersection.

Figure 39 shows a nested sequence of nonempty *noncompact* sets with empty intersection. They are the open discs with center $(1/n, 0)$ on the x-axis and radius $1/n$. They contain no common point. (Their closures do intersect at a common point, the origin.)

Continuity and Compactness

Next we discuss how compact sets behave under continuous transformations.

36 Theorem *If $f : M \to N$ is continuous and A is a compact subset of M then fA is a compact subset of N. That is, the continuous image of a compact is compact.*

Proof Suppose that (b_n) is a sequence in fA. For each $n \in \mathbb{N}$ choose a point $a_n \in A$ such that $f(a_n) = b_n$. By compactness of A there exists a subsequence (a_{n_k}) that converges to some point $p \in A$. By continuity of f it follows that

$$b_{n_k} = f(a_{n_k}) \to fp \in fA$$

as $k \to \infty$. Thus, every sequence (b_n) in fA has a subsequence converging to a limit in fA, which shows that fA is compact. $\square$

From Theorem 36 follows the natural generalization of the min/max theorem in Chapter 1 which concerns continuous real-valued functions defined on an interval $[a, b]$. See Theorem 1.23.

37 Corollary *A continuous real-valued function defined on a compact set is bounded; it assumes maximum and minimum values.*

Proof Let $f : M \to \mathbb{R}$ be continuous and let A be a compact subset of M. Theorem 36 implies that fA is a compact subset of $\mathbb{R}$, so by Theorem 26 it is closed and bounded. Thus, the greatest lower bound, v, and least upper bound, V, of fA exist and belong to fA; i.e., there exist points $p, P \in A$ such that for all $a \in A$ we have $v = fp \leq fa \leq fP = V$. $\square$

Homeomorphisms and Compactness

A homeomorphism is a bicontinuous bijection. Originally, compactness was called bicompactness. This is reflected in the next theorem.

38 Theorem *If M is compact and M is homeomorphic to N then N is compact. Compactness is a topological property.*

Proof N is the continuous image of M, so by Theorem 36 it is compact. $\square$

39 Corollary $[0, 1]$ *and* $\mathbb{R}$ *are not homeomorphic.*

Proof One is compact and the other isn't. $\square$

40 Theorem *If M is compact then a continuous bijection $f : M \to N$ is a homeomorphism – its inverse bijection $f^{-1} : N \to M$ is automatically continuous.*

Proof Suppose that $q_n \to q$ in N. Since f is a bijection, $p_n = f^{-1}(q_n)$ and $p = f^{-1}(q)$ are well defined points in M. To check continuity of f^{-1} we must show that $p_n \to p$.

If (p_n) refuses to converge to p then there is a subsequence (p_{n_k}) and a $\delta > 0$ such that for all k we have $d(p_{n_k}, p) \geq \delta$. Compactness of M gives a sub-subsequence $(p_{n_{k(\ell)}})$ that converges to a point $p^* \in M$ as $\ell \to \infty$.

Necessarily, $d(p, p^*) \geq \delta$, which implies that $p \neq p^*$. Since f is continuous we have

$$f(p_{n_{k(\ell)}}) \to f(p^*)$$

as $\ell \to \infty$. The limit of a convergent sequence is unchanged by passing to a subsequence, and so $f(p_{n_{k(\ell)}}) = q_{n_{k(\ell)}} \to q$ as $\ell \to \infty$. Thus, $f(p^*) = q = f(p)$, contrary to f being a bijection. It follows that $p_n \to p$ and therefore that f^{-1} is continuous. $\square$

If M is not compact then Theorem 40 becomes false. For example, the function $f : [0, 2\pi) \to \mathbb{R}^2$ defined by $f(x) = (\cos x, \sin x)$ is a continuous bijection onto the unit circle in the plane, but it is not a homeomorphism. This useful example was discussed on page 65. Not only does this f fail to be a homeomorphism, but there is no homeomorphism at all from $[0, 2\pi)$ to S^1. The circle is compact while $[0, 2\pi)$ is not. Therefore they are not homeomorphic. See also Exercises 49 and 50.

Embedding a Compact

Not only is a compact space M closed in itself, as is every metric space, but it is also a closed subset of each metric space in which it is embedded. More precisely we say that $h : M \to N$ **embeds** M into N if h is a homeomorphism from M onto hM. (The metric on hM is the one it inherits from N.) Topologically M and hM are equivalent. A property of M that holds also for every embedded copy of M is an **absolute** or **intrinsic** property of M.

41 Theorem *A compact is absolutely closed and absolutely bounded.*

Proof Obvious from Theorems 26 and 36. $\square$

For example, no matter how the circle is embedded in $\mathbb{R}^3$, its image is always closed and bounded. See also Exercises 48 and 120.

Uniform Continuity and Compactness

In Chapter 1 we defined the concept of uniform continuity for real-valued functions of a real variable. The definition in metric spaces is analogous. A function $f : M \to N$ is **uniformly continuous** if for each $\epsilon > 0$ there exists a $\delta > 0$ such that

$$p, q \in M \text{ and } d_M(p, q) < \delta \quad \Rightarrow \quad d_N(fp, fq) < \epsilon.$$

42 Theorem *Every continuous function defined on a compact is uniformly continuous.*

Proof Suppose not, and $f : M \to N$ is continuous, M is compact, but f fails to be uniformly continuous. Then there is some $\epsilon > 0$ such that no matter how small

δ is, there exist points $p, q \in M$ with $d(p,q) < \delta$ but $d(fp, fq) \geq \epsilon$. Take $\delta = 1/n$ and let (p_n) and (q_n) be sequences of points in M such that $d(p_n, q_n) < 1/n$ while $d(f(p_n), f(q_n)) \geq \epsilon$. Compactness of M implies that there is a subsequence p_{n_k} which converges to some $p \in M$ as $k \to \infty$. Since $d(p_n, q_n) < 1/n \to 0$ as $n \to \infty$, (q_{n_k}) converges to the same limit as does (p_{n_k}) as $k \to \infty$; namely $q_{n_k} \to p$. Continuity at p implies that $f(p_{n_k}) \to fp$ and $f(q_{n_k}) \to fp$. If k is large then

$$d(f(p_{n_k}), f(q_{n_k})) \ \leq \ d(f(p_{n_k}), fp) \ + \ d(fp, f(q_{n_k})) \ < \ \epsilon,$$

contrary to the supposition that $d(f(p_n), f(q_n)) \geq \epsilon$ for all n. $\square$

Theorem 42 gives a second proof that continuity implies uniform continuity on an interval $[a, b]$. For $[a, b]$ is compact.

5 Connectedness

As another application of these ideas, we consider the general notion of connectedness. Let A be a subset of a metric space M. If A is neither the empty set nor M then A is a **proper** subset of M. Recall that if A is both closed and open in M it is said to be clopen. The complement of a clopen set is clopen. The complement of a proper subset is proper.

If M has a proper clopen subset A then M is **disconnected**. For there is a **separation** of M into proper, disjoint clopen subsets,

$$M \ = \ A \sqcup A^c.$$

(The notation $\sqcup$ indicates disjoint union.) M is **connected** if it is not disconnected, i.e., it contains no proper clopen subset. Connectedness of M does not mean that M is connected *to* something, but rather that M is one unbroken thing. See Figure 40.

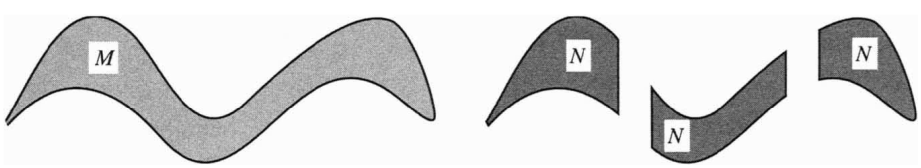

Figure 40 M and N illustrate the difference between being connected and being disconnected.

43 Theorem *If M is connected, $f : M \to N$ is continuous, and f is onto then N is connected. The continuous image of a connected is connected.*

Proof Simple! If A is a clopen proper subset of N then, according to the open and closed set conditions for continuity, $f^{\text{pre}}(A)$ is a clopen subset of M. Since f is onto and $A \neq \emptyset$, we have $f^{\text{pre}}(A) \neq \emptyset$. Similarly, $f^{\text{pre}}(A^c) \neq \emptyset$. Therefore $f^{\text{pre}}(A)$ is a proper clopen subset of M, contrary to M being connected. It follows that A cannot exist and that N is connected. $\square$

44 Corollary *If M is connected and M is homeomorphic to N then N is connected. Connectedness is a topological property.*

Proof N is the continuous image of M, so Theorem 43 implies it is connected. $\square$

45 Corollary (Generalized Intermediate Value Theorem) *Every continuous real-valued function defined on a connected domain has the intermediate value property.*

Proof Assume that $f : M \to \mathbb{R}$ is continuous and M is connected. If f assumes values $\alpha < \beta$ in $\mathbb{R}$ and if it fails to assume some value γ with $\alpha < \gamma < \beta$, then

$$M = \{x \in M : f(x) < \gamma\} \sqcup \{x \in M : f(x) > \gamma\}$$

is a separation of M, contrary to connectedness. $\square$

46 Theorem $\mathbb{R}$ *is connected.*

Proof If $U \subset \mathbb{R}$ is nonempty and clopen we claim that $U = \mathbb{R}$. Choose some $p \in U$ and consider the set

$$X = \{x \in U : \text{the open interval } (p, x) \text{ is contained in } U\}.$$

X is nonempty since U is open. Let s be the supremum of X. If s is finite (i.e., X is bounded above) then $s = \text{l. u. b.}\, X$ and s is a limit of X. Since $X \subset U$ and U is closed we have $s \in U$. Since U is open there is an interval $(s - r, s + r) \subset U$, which gives an interval $(p, s + r) \subset U$, contrary to s being an upper bound for X. Hence $s = \infty$ and $U \supset (p, \infty)$. The same reasoning gives $U \supset (-\infty, p)$, so $U = \mathbb{R}$ as claimed. Thus there are no proper clopen subsets of $\mathbb{R}$ and $\mathbb{R}$ is connected. $\square$

47 Corollary (Intermediate Value Theorem for $\mathbb{R}$) *Every continuous function $f : \mathbb{R} \to \mathbb{R}$ has the intermediate value property.*

Proof Immediate from the Generalized Intermediate Value Theorem and connectedness of $\mathbb{R}$. $\square$

48 Corollary *The following metric spaces are connected: The intervals (a, b), $[a, b]$, the circle, and all capital letters of the Roman alphabet.*

Proof The interval (a, b) is homeomorphic to $\mathbb{R}$, while $[a, b]$ is the continuous image of $\mathbb{R}$ under the map whose graph is shown in Figure 41. The circle is the continuous image of $\mathbb{R}$ under the map $t \mapsto (\cos t, \sin t)$. Connectedness of the letters A, ..., Z is equally clear. $\square$

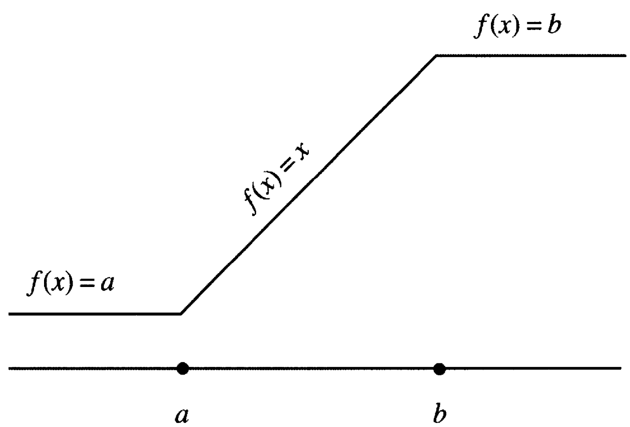

Figure 41 The function f surjects $\mathbb{R}$ continuously to $[a, b]$.

Connectedness properties give a good way to distinguish nonhomeomorphic sets.

Example The union of two disjoint closed intervals is not homeomorphic to a single interval. One set is disconnected and the other is connected.

Example The closed interval $[a, b]$ is not homeomorphic to the circle S^1. For removal of a point $x \in (a, b)$ disconnects $[a, b]$ while the circle remains connected upon removal of any point. More precisely, suppose that $h : [a, b] \to S^1$ is a homeomorphism. Choose a point $x \in (a, b)$, and consider $X = [a, b] \setminus \{x\}$. The restriction of h to X is a homeomorphism from X onto Y, where Y is the circle with the point hx removed. But X is disconnected while Y is connected. Hence h cannot exist and the segment is not homeomorphic to the circle.

Example The circle is not homeomorphic to the figure eight. Removing any two points of the circle disconnects it, but this is not true of the figure eight. Or, removing the crossing point disconnects the figure eight but removing any point of the circle leaves it connected.

Example The circle is not homeomorphic to the disc. For removing two points disconnects the circle but does not disconnect the disc.

As you can see, it is useful to be able to recognize disconnected subsets S of a metric space M. By definition, S is a disconnected subset of M if it is disconnected when considered in its own right as a metric space with the metric it inherits from M; i.e., it has a separation $S = A \sqcup B$ such that A and B are proper clopen subsets of S. The sets A, B are separated in S but they need not be separated in M. Their closures in M may intersect.

Example The punctured interval $X = [a, b] \setminus \{c\}$ is disconnected if $a < c < b$. For $X = [a, c) \sqcup (c, b]$ is a separation of X. The closures of the two sets with respect to the metric space X do not intersect, even though their closures with respect to $\mathbb{R}$ do intersect. Pay attention to this phenomenon which is related to the Inheritance Principle.

Example Any subset Y of the punctured interval is disconnected if it meets both $[a, c)$ and $(c, b]$. For $Y = ([a, c) \cap Y) \sqcup ((c, b] \cap Y)$ is a separation of Y.

49 Theorem *The closure of a connected set is connected. More generally, if $S \subset M$ is connected and $S \subset T \subset \overline{S}$ then T is connected.*

Proof It is equivalent to show that if T is disconnected then S is disconnected. Disconnectedness of T implies that

$$T = A \sqcup B$$

where A, B are clopen and proper in T. It is natural to expect that

$$S = K \sqcup L$$

is a separation of S where $K = A \cap S$ and $L = B \cap S$. The sets K and L are disjoint, their union is S, and by the Inheritance Principle they are clopen. But are they proper?

If $K = \emptyset$ then $A \subset S^c$. Since A is proper there exists $p \in A$. Since A is open in T, there exists a neighborhood $M_r p$ such that

$$T \cap M_r p \ \subset \ A \ \subset \ S^c.$$

The neighborhood $M_r p$ contains no points of S, which is contrary to p belonging to $\overline{S}$. Thus, $K \neq \emptyset$. Similarly, $L = B \cap S \neq \emptyset$, so $S = K \sqcup L$ is a separation of S, proving that S is disconnected. $\square$

Example The outward spiral expressed in polar coordinates as

$$S = \{(r, \theta) : (1 - r)\theta = 1 \text{ and } \theta \geq \pi/2\}$$

has $\overline{S} = S \cup S^1$, where S^1 is the unit circle. Since S is connected, so is $\overline{S}$. (Recall that $\overline{S}$ is the closure of S.) See Figure 27.

50 Theorem *The union of connected sets sharing a common point p is connected.*

Proof Let $S = \bigcup S_\alpha$, where each S_α is connected and $p \in \bigcap S_\alpha$. If S is disconnected then it has a separation $S = A \sqcup A^c$ where A, A^c are proper and clopen. One of them contains p; say it is A. Then $A \cap S_\alpha$ is a nonempty clopen subset of S_α. Since S_α is connected, $A \cap S_\alpha = S_\alpha$ for each α, and $A = S$. This implies that $A^c = \emptyset$, a contradiction. Therefore S is connected. $\square$

Example The 2-sphere S^2 is connected. For S^2 is the union of great circles, each passing through the poles.

Example Every convex set C in $\mathbb{R}^m$ (or in any metric space with a compatible linear structure) is connected. If we choose a point $p \in C$ then each $q \in C$ lies on a line segment $[p, q] \subset C$. Thus, C is the union of connected sets sharing the common point p. It is connected.

Definition A **path** joining p to q in a metric space M is a continuous function $f : [a, b] \to M$ such that $fa = p$ and $fb = q$. If each pair of points in M can be joined by a path in M then M is **path-connected**. See Figure 42.

51 Theorem *Path-connected implies connected.*

Proof Assume that M is path-connected but not connected. Then $M = A \sqcup A^c$ for some proper clopen $A \subset M$. Choose $p \in A$ and $q \in A^c$. There is a path $f : [a, b] \to M$ from p to q. The separation $f^{\text{pre}}(A) \sqcup f^{\text{pre}}(A^c)$ contradicts connectedness of $[a, b]$. Therefore M is connected. $\square$

Example All connected subsets of $\mathbb{R}$ are path-connected. See Exercise 67.

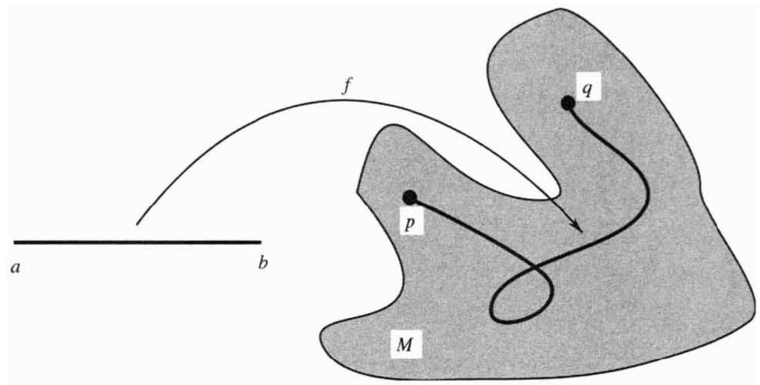

Figure 42 A path f in M that joins p to q

Example Every open connected subset of $\mathbb{R}^m$ is path-connected. See Exercises 61 and 66.

Example The **topologist's sine curve** is a compact connected set that is not path-connected. It is $M = G \cup Y$ where

$$\begin{aligned} G &= \{(x,y) \in \mathbb{R}^2 : y = \sin 1/x \text{ and } 0 < x \leq 1/\pi\} \\ Y &= \{(0,y) \in \mathbb{R}^2 : -1 \leq y \leq 1\}. \end{aligned}$$

See Figure 43. The metric on M is just Euclidean distance. Is M connected? Yes!

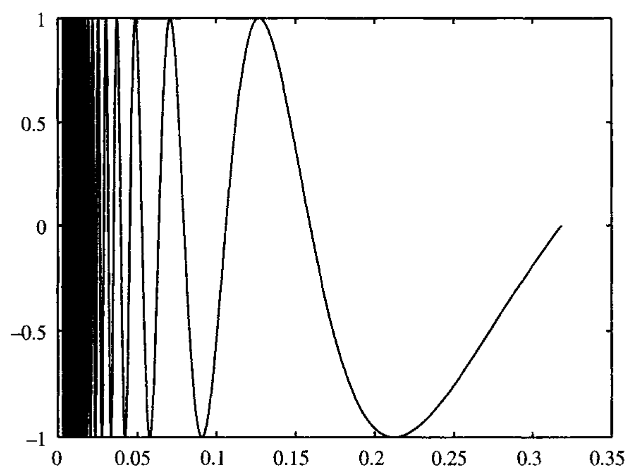

Figure 43 The topologist's sine curve M is a connected set. It includes the vertical segment Y at $x = 0$.

The graph G is connected and $M = \overline{G}$. By Theorem 49 M is connected.

6 Other Metric Space Concepts

Here are a few standard metric space topics related to what appears above. If $S \subset M$ then its **closure** is the smallest closed subset of M that contains S, its **interior** is the largest open subset of M contained in S, and its **boundary** is the difference between its closure and its interior. Their notations are

$$\overline{S} = \operatorname{cl} S = \text{ closure of } S \quad \operatorname{int} S = \text{ interior of } S \quad \partial S = \text{ boundary of } S.$$

To avoid inheritance ambiguity it would be better (but too cumbersome) to write $\operatorname{cl}_M S$, $\operatorname{int}_M S$, and $\partial_M S$ to indicate the ambient space M. In Exercise 95 you are asked to check various simple facts about them, such as $\overline{S} = \lim S =$ the intersection of all closed sets that contain S.

Clustering and Condensing

Two concepts similar to limits are clustering and condensing. The set S "clusters" at p (and p is a **cluster point**† of S) if each $M_r p$ contains infinitely many points of S. The set S **condenses** at p (and p is a **condensation point** of S) if each $M_r p$ contains uncountably many points of S. Thus, S limits at p, clusters at p, or condenses at p according to whether each $M_r p$ contains some, infinitely many, or uncountably many points of S. See Figure 44.

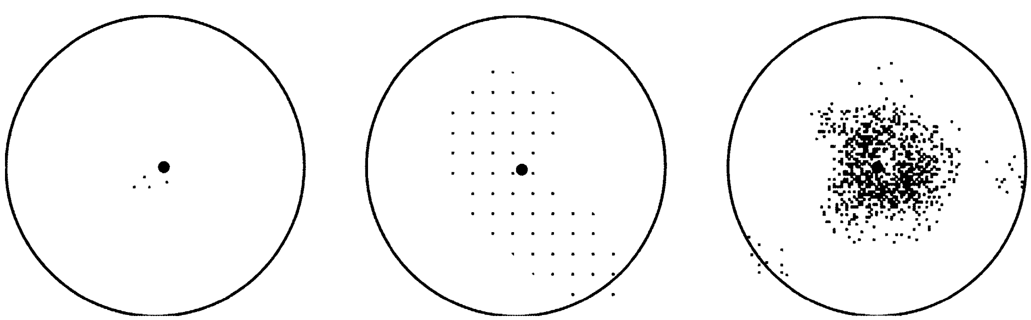

Figure 44 Limiting, clustering, and condensing behavior

†Cluster points are also called **accumulation points**. As mentioned above, they are also sometimes called limit points, a usage that conflicts with the limit idea. A finite set S has no cluster points, but of course, each of its points p is a limit of S since the constant sequence $(p, p, p, \ldots)$ converges to p.

52 Theorem *The following are equivalent conditions to S clustering at p.*

(i) There is a sequence of distinct points in S that converges to p.

(ii) Each neighborhood of p contains infinitely many points of S.

(iii) Each neighborhood of p contains at least two points of S.

(iv) Each neighborhood of p contains at least one point of S other than p.

Proof Clearly (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv), and (ii) is the definition of clustering. It remains to check (iv) $\Rightarrow$ (i).

Assume (iv) is true: Each neighborhood of p contains a point of S other than p. In $M_1 p$ choose a point $p_1 \in (S \setminus \{p\})$. Set $r_2 = \min(1/2, d(p_1, p))$, and in the smaller neighborhood $M_{r_2} p$, choose $p_2 \in (S \setminus \{p\})$. Proceed inductively: Set $r_n = \min(1/n, d(p_{n-1}, p))$ and in $M_{r_n} p$, choose $p_n \in (S \setminus \{p\})$. Since $r_n \to 0$ the sequence (p_n) converges to p. The points p_n are distinct since they have different distances to p,

$$d(p_1, p) \geq r_2 > d(p_2, p) \geq r_3 > d(p_3, p) \geq \ldots .$$

Thus (iv) $\Rightarrow$ (i) and the four conditions are equivalent. $\square$

Condition (iv) is the form of the definition of clustering most frequently used, although it is the hardest to grasp. It is customary to denote by S' the set of cluster points of S.

53 Proposition $S \cup S' = \overline{S}$.

Proof A cluster point is a type of limit of S, so $S' \subset \lim S = \overline{S}$ and

$$S \cup S' \subset \overline{S}$$

On the other hand, if $p \in \overline{S}$ then either $p \in S$ or else $p \notin S$ and each neighborhood of p contains points of S other than p. This implies that $p \in S \cup S'$, so $\overline{S} \subset S \cup S'$, and the two sets are equal. $\square$

54 Corollary *S is closed if and only if $S' \subset S$.*

Proof S is closed if and only if $S = \overline{S}$. Since $\overline{S} = S \cup S'$, equivalent to $S' \subset S$ is $\overline{S} = S$. $\square$

55 Corollary *The least upper bound and greatest lower bound of a nonempty bounded set $S \subset \mathbb{R}$ belong to the closure of S. Thus, if S is closed then they belong to S.*

Proof If $b = \text{l.u.b.}\, S$ then each interval $(b - r, b]$ contains points of S. The same is true for intervals $[a, a + r)$ where $a = \text{g.l.b.}\, S$ $\square$

Perfect Metric Spaces

A metric space M is **perfect** if $M' = M$, i.e., each $p \in M$ is a cluster point of M. Recall that M clusters at p if each $M_r p$ is an infinite set. For example $[a, b]$ is perfect and $\mathbb{Q}$ is perfect. $\mathbb{N}$ is not perfect since none of its points are cluster points.

56 Theorem *Every nonempty, perfect, complete metric space is uncountable.*

Proof Suppose not: Assume M is nonempty, perfect, complete, and countable. Since M consists of cluster points it must be denumerable and not finite. Say

$$M = \{x_1, x_2, \ldots\}$$

is a list of all the elements of M. We will derive a contradiction by finding a point of M not in the list. Define

$$\widehat{M_r p} = \{q \in M : d(p, q) \leq r\}.$$

It is the **closed neighborhood** of radius r at p. Choose any $y_1 \in M$ with $y_1 \neq x_1$ and choose $r_1 > 0$ so that $Y_1 = \widehat{M_{r_1}}(y_1)$ "excludes" x_1 in the sense that $x_1 \notin Y_1$. We can take r_1 as small as we want, say $r_1 < 1$.

Since M clusters at y_1 we can choose $y_2 \in M_{r_1}(y_1)$ with $y_2 \neq x_2$ and choose $r_2 > 0$ so that $Y_2 = \widehat{M_{r_2}}(y_2)$ excludes x_2. Taking r_2 small ensures $Y_2 \subset Y_1$. (Here we are using openness of $M_{r_1}(y_1)$.) Also we take $r_2 < 1/2$. Since $Y_2 \subset Y_1$, it excludes x_1 as well as x_2. See Figure 45.

Nothing stops us from continuing inductively, and we get a nested sequence of closed neighborhoods $Y_1 \supset Y_2 \supset Y_3 \ldots$ such that Y_n excludes $x_1, \ldots, x_n$, and has radius $r_n \leq 1/n$. Thus the center points y_n form a Cauchy sequence. Completeness of M implies that

$$\lim_{n \to \infty} y_n = y \in M$$

exists. Since the sets Y_n are closed and nested, $y \in Y_n$ for each n. Does y equal x_1? No, for Y_1 excludes x_1. Does it equal x_2? No, for Y_2 excludes x_2. In fact, for each n we have $y \neq x_n$. The point y is nowhere in the supposedly complete list of elements of M, a contradiction. Hence M is uncountable. $\square$

57 Corollary $\mathbb{R}$ *and* $[a, b]$ *are uncountable.*

Proof $\mathbb{R}$ is complete and perfect, while $[a, b]$ is compact, therefore complete, and perfect. Neither is empty. $\square$

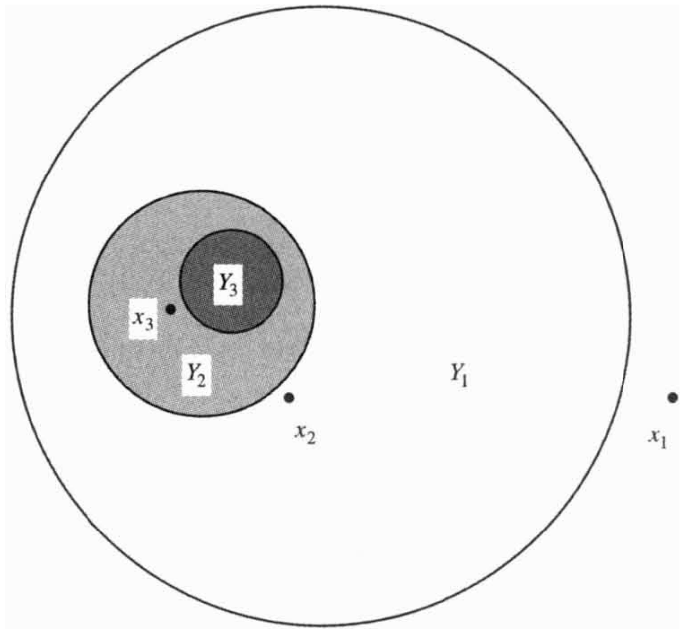

Figure 45 The exclusion of successively more points of the sequence (x_n) that supposedly lists all the elements of M

58 Corollary *Every nonempty perfect complete metric space is everywhere uncountable in the sense that each r-neighborhood is uncountable.*

Proof The $r/2$-neighborhood $M_{r/2}(p)$ is perfect: It clusters at each of its points. The closure of a perfect set is perfect. Thus, $\overline{M_{r/2}(p)}$ is perfect. Being a closed subset of a complete metric space, it is complete. According to Theorem 56, $\overline{M_{r/2}(p)}$ is uncountable. Since $\overline{M_{r/2}(p)} \subset M_r p$, $M_r p$ is uncountable. $\square$

Continuity of Arithmetic in $\mathbb{R}$

Addition is a mapping $\text{Sum} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ that assigns to (x, y) the real number $x + y$. Subtraction and multiplication are also such mappings. Division is a mapping $\mathbb{R} \times (\mathbb{R} \setminus \{0\}) \to \mathbb{R}$ that assigns to (x, y) the number x/y.

59 Theorem *The arithmetic operations of $\mathbb{R}$ are continuous.*

60 Lemma *For each real number c the function* $\text{Mult}_c : \mathbb{R} \to \mathbb{R}$ *that sends x to cx is continuous.*

Proof If $c = 0$ the function is constantly equal to 0 and is therefore continuous. If $c \neq 0$ and $\epsilon > 0$ is given, choose $\delta = \epsilon / |c|$. If $|x - y| < \delta$ then

$$|\mathrm{Mult}_c(x) - \mathrm{Mult}_c(y)| = |c|\,|x - y| < |c|\,\delta = \epsilon$$

which shows that Mult_c is continuous. $\square$

Proof of Theorem 59 We use the preservation of sequential convergence criterion for continuity. It's simplest. Let $(x_n, y_n) \to (x, y)$ as $n \to \infty$.

By the triangle inequality we have

$$|\mathrm{Sum}(x_n, y_n) - \mathrm{Sum}(x, y)| \leq |x_n - x| + |y_n - y| = d_{\mathrm{sum}}((x_n, y_n), (x, y)).$$

By Corollary 21 d_{sum} is continuous, so $d_{\mathrm{sum}}((x_n, y_n), (x, y)) \to 0$ as $n \to \infty$, which completes the proof that Sum is continuous. (By Theorem 17 it does not matter which metric we use on $\mathbb{R} \times \mathbb{R}$.)

Subtraction is the composition of continuous functions

$$\mathrm{Sub}(x, y) = \mathrm{Sum} \circ (\mathrm{id} \times \mathrm{Mult}_{-1})(x, y)$$

and is therefore continuous. (Proposition 3 implies id is continuous, Lemma 60 implies Mult_{-1} is continuous, and Corollary 18 implies $\mathrm{id} \times \mathrm{Mult}_{-1}$ is continuous.)

Multiplication is continuous since

$$\begin{aligned} |\mathrm{Mult}(x_n, y_n) - \mathrm{Mult}(x, y)| &= |x_n y_n - xy| \\ &\leq |x_n - x|\,|y_n| + |x|\,|y_n - y| \\ &\leq B(|x - x_n| + |y - y_n|) \\ &= \mathrm{Mult}_B(d_{\mathrm{sum}}((x_n, y_n), (x, y))) \to 0 \end{aligned}$$

as $n \to \infty$, where we use the fact that convergent sequences are bounded to write $|y_n| + |x| \leq B$ for all n.

Reciprocation is the function $\mathrm{Rec} : \mathbb{R} \setminus \{0\} \to \mathbb{R} \setminus \{0\}$ that sends x to $1/x$. If $x_n \to x \neq 0$ then there is a constant $b > 0$ such that for all large n we have $|1/x_n| \leq b$ and $|1/x| \leq b$. Since

$$|\mathrm{Rec}(x_n) - \mathrm{Rec}(x)| = \left| \frac{1}{x_n} - \frac{1}{x} \right| = \frac{|x_n - x|}{|x x_n|} \leq \mathrm{Mult}_{b^2}(|x_n - x|) \to 0$$

as $n \to \infty$ we see that Rec is continuous.

Division is continuous on $\mathbb{R} \times (\mathbb{R} \setminus \{0\})$ since it is the composite of continuous mappings $\mathrm{Mult} \circ (\mathrm{id} \times \mathrm{Rec}) : (x, y) \mapsto (x, 1/y) \mapsto x \cdot 1/y$. $\square$

The absolute value is a mapping $\mathrm{Abs} : \mathbb{R} \to \mathbb{R}$ that sends x to $|x|$. It is continuous since it is $d(x, 0)$ and the distance function is continuous. The maximum and minimum are functions $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$ given by the formulas

$$\max(x, y) = \frac{x + y}{2} + \frac{|x - y|}{2} \qquad \min(x, y) = \frac{x + y}{2} - \frac{|x - y|}{2},$$

so they are also continuous.

61 Corollary *The sums, differences, products, and quotients, absolute values, maxima, and minima of real-valued continuous functions are continuous. (The denominator functions should not equal zero.)*

Proof Take, for example, the sum $f + g$ where $f, g : M \to \mathbb{R}$ are continuous. It is the composite of continuous functions

$$\begin{array}{ccccc} M & \xrightarrow{f \times g} & \mathbb{R} \times \mathbb{R} & \xrightarrow{\mathrm{Sum}} & \mathbb{R} \\ x & \mapsto & (fx, gx) & \mapsto & \mathrm{Sum}(fx, gx), \end{array}$$

and is therefore continuous. The same applies to the other operations. □

62 Corollary *Polynomials are continuous functions.*

Proof Proposition 3 states that constant functions and the identity function are continuous. Thus Corollary 61 and induction imply that the polynomial $a_0 + a_1 x + \cdots + a_n x^n$ is continuous. □

The same reasoning shows that polynomials of m variables are continuous functions $\mathbb{R}^m \to \mathbb{R}$.

Boundedness

A subset S of a metric space M is **bounded** if for some $p \in M$ and some $r > 0$,

$$S \subset M_r p.$$

A set which is not bounded is **unbounded**. For example, the elliptical region $4x^2 + y^2 < 4$ is a bounded subset of $\mathbb{R}^2$, while the hyperbola $xy = 1$ is unbounded. It is easy to see that if S is bounded then for each $q \in M$ there is an s such that $M_s q$ contains S.

Distinguish the word "bounded" from the word "finite." The first refers to physical size, the second to the number of elements. The concepts are totally different.

Also, boundedness has little connection to the existence of a boundary – a clopen subset of a metric space has empty boundary, but some clopen sets are bounded, others not.

Exercise 39 asks you to show that every convergent sequence is bounded, and to decide whether it is also true that every Cauchy sequence is bounded, even when the metric space is not complete.

Boundedness is not a topological property. For example, $(-1, 1)$ and $\mathbb{R}$ are homeomorphic although $(-1, 1)$ is bounded and $\mathbb{R}$ is unbounded. The same example shows that completeness is not a topological property.

A function from M to another metric space N is a **bounded function** if its range is a bounded subset of N. That is, there exist $q \in N$ and $r > 0$ such that

$$fM \subset N_r q.$$

Note that a function can be bounded even though its graph is not. For example, $x \mapsto \sin x$ is a bounded function $\mathbb{R} \to \mathbb{R}$ although its graph, $\{(x, y) \in \mathbb{R}^2 : y = \sin x\}$, is an unbounded subset of $\mathbb{R}^2$.

7 Coverings

For the sake of simplicity we have postponed discussing compactness in terms of open coverings until this point. Typically, students find coverings a challenging concept. It is central, however, to much of analysis – for example, measure theory.

Definition A collection $\mathcal{U}$ of subsets of M **covers** $A \subset M$ if A is contained in the union of the sets belonging to $\mathcal{U}$. The collection $\mathcal{U}$ is a **covering** of A. If $\mathcal{U}$ and $\mathcal{V}$ both cover A and if $\mathcal{V} \subset \mathcal{U}$ in the sense that each set $V \in \mathcal{V}$ belongs also to $\mathcal{U}$ then we say that $\mathcal{U}$ **reduces to** $\mathcal{V}$, and that $\mathcal{V}$ is a **subcovering** of A.

Definition If all the sets in a covering $\mathcal{U}$ of A are open then $\mathcal{U}$ is an **open covering** of A. If every open covering of A reduces to a finite subcovering of A then we say that A is **covering compact**†.

The idea is that if A is covering compact and $\mathcal{U}$ is an open covering of A then just a finite number of the open sets are actually doing the work of covering A. The rest are redundant.

†You will frequently find it said that an open covering of A *has* a finite subcovering. “Has” means “reduces to.”

A covering $\mathcal{U}$ of A is also called a **cover** of A. The members of $\mathcal{U}$ are *not* called covers. Instead, you could call them **scraps** or **patches**. Imagine the covering as a patchwork quilt that covers a bed, the quilt being sewn together from overlapping scraps of cloth. See Figure 46.

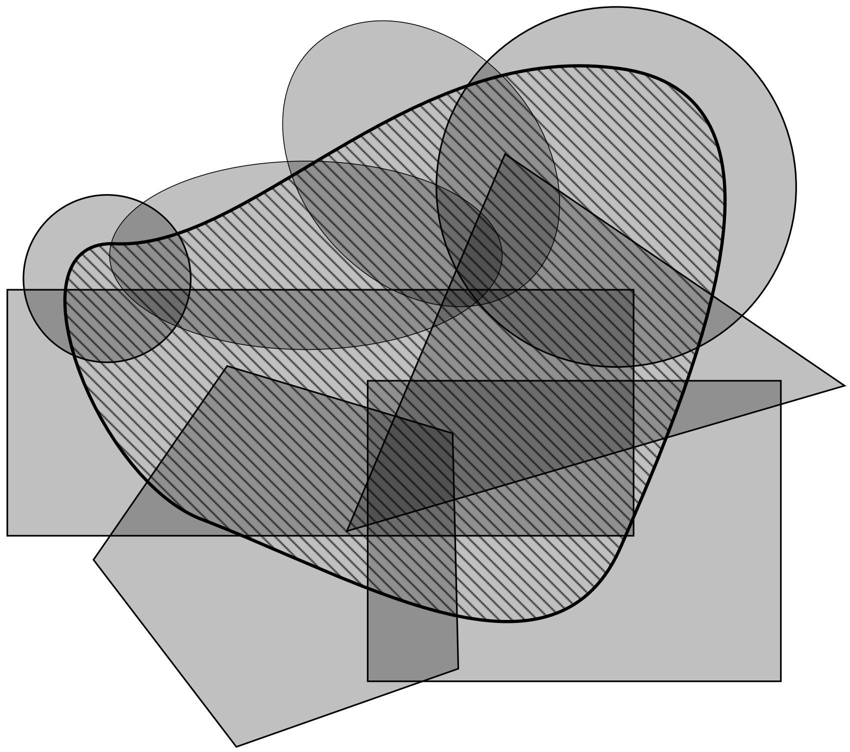

Figure 46 A covering of A by eight scraps. The set A is cross-hatched. The scraps are two discs, two rectangles, two ellipses, a pentagon, and a triangle. Each point of A belongs to at least one scrap.

The mere existence of a finite open covering of A is trivial and utterly worthless. *Every* set A has such a covering, namely the single open set M. Rather, for A to be covering compact, each and every open covering of A must reduce to a finite subcovering of A. Deciding directly whether this is so is daunting. How could you hope to verify the finite reducibility of all open coverings of A? There are so many of them. For this reason we concentrated on sequential compactness; it is relatively easy to check by inspection whether every sequence in a set has a convergent subsequence.

To check that a set is not covering compact it suffices to find an open covering which fails to reduce to a finite subcovering. Occasionally this is simple. For example, the set $(0, 1]$ is not covering compact in $\mathbb{R}$ because its covering

$$\mathcal{U} = \{(1/n, 2) : n \in \mathbb{N}\}$$

fails to reduce to a finite subcovering.

63 Theorem *For a subset A of a metric space M the following are equivalent:*

(a) A is covering compact.
(b) A is sequentially compact.

Proof that (a) implies (b) We assume A is covering compact and prove it is sequentially compact. Suppose not. Then there is a sequence (p_n) in A, no subsequence of which converges in A. Each point $a \in A$ therefore has some neighborhood $M_r a$ such that $p_n \in M_r a$ only finitely often. (The radius r may depend on the point a.) The collection $\{M_r a : a \in A\}$ is an open covering of A and by covering compactness it reduces to a finite subcovering

$$\{M_{r_1}(a_1), M_{r_2}(a_2), \ldots, M_{r_k}(a_k)\}$$

of A. Since p_n appears in each of these finitely many neighborhoods $M_{r_i}(a_i)$ only finitely often, it follows from the pigeonhole principle that (p_n) has only finitely many terms, a contradiction. Thus (p_n) cannot exist, and A is sequentially compact. $\square$

The following presentation of the proof that (b) implies (a) appears in Royden's book, *Real Analysis*. A **Lebesgue number** for a covering $\mathcal{U}$ of A is a positive real number λ such that for each $a \in A$ there is some $U \in \mathcal{U}$ with $M_\lambda a \subset U$. Of course, the choice of this U depends on a. It is crucial, however, that the Lebesgue number λ is independent of $a \in A$.

The idea of a Lebesgue number is that we know each point $a \in A$ is contained in some $U \in \mathcal{U}$, and if λ is extremely small then $M_\lambda a$ is just a slightly swollen point – so the same should be true for it too. No matter where in A the neighborhood $M_\lambda a$ is placed, it should lie wholly in some member of the covering. See Figure 47.

If A is noncompact then it may have open coverings with no positive Lebesgue number. For example, let $A = (0, 1) \subset \mathbb{R} = M$. The singleton collection $\{A\}$ is an open covering of A, but there is no $\lambda > 0$ such that for every $a \in A$ we have $(a - \lambda, a + \lambda) \subset A$. See Exercise 86.

64 Lebesgue Number Lemma *Every open covering of a sequentially compact set has a Lebesgue number $\lambda > 0$.*

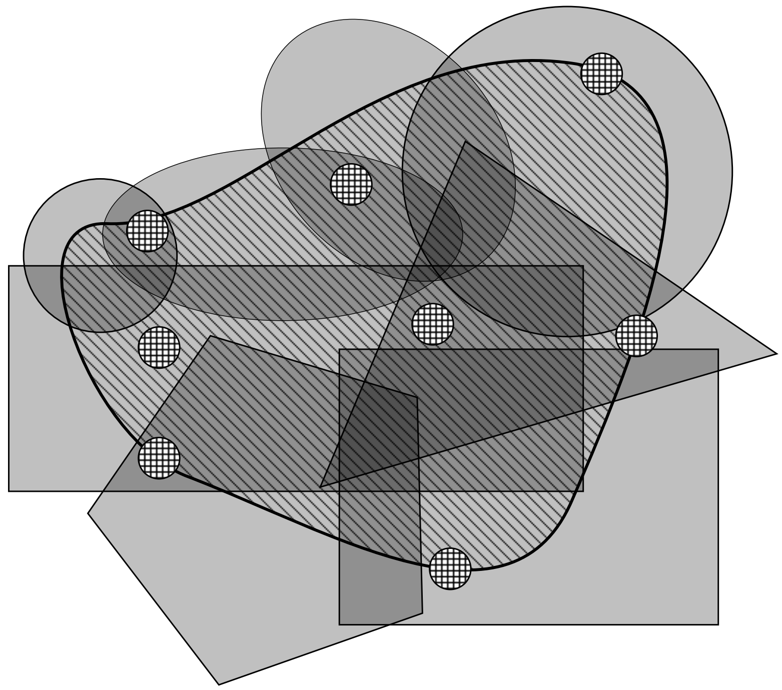

Figure 47 Small neighborhoods are like swollen points. $\mathcal{U}$ has a positive Lebesgue number λ. The λ-neighborhood of each point in the cross-hatched set A is wholly contained in at least one member of the covering.

Proof Suppose not: $\mathcal{U}$ is an open covering of a sequentially compact set A, and yet for each $\lambda > 0$ there exists an $a \in A$ such that no $U \in \mathcal{U}$ contains $M_\lambda a$. Take $\lambda = 1/n$ and let $a_n \in A$ be a point such that no $U \in \mathcal{U}$ contains $M_{1/n}(a_n)$. By sequential compactness, there is a subsequence (a_{n_k}) converging to some point $p \in A$. Since $\mathcal{U}$ is an open covering of A, there exist $r > 0$ and $U \in \mathcal{U}$ with $M_r p \subset U$. If k is large then $d(a_{n_k}, p) < r/2$ and $1/n_k < r/2$, which implies by the triangle inequality that

$$M_{1/n_k}(a_{n_k}) \subset M_r p \subset U,$$

contrary to the supposition that no $U \in \mathcal{U}$ contains $M_{1/n}(a_n)$. We conclude that, after all, $\mathcal{U}$ does have a Lebesgue number $\lambda > 0$. See Figure 48. $\square$

Proof that (b) implies (a) in Theorem 63 Let $\mathcal{U}$ be an open covering of the sequentially compact set A. We want to reduce $\mathcal{U}$ to a finite subcovering. By the Lebesgue Number Lemma, $\mathcal{U}$ has a Lebesgue number $\lambda > 0$. Choose any $a_1 \in A$ and some $U_1 \in \mathcal{U}$ such that

$$M_\lambda(a_1) \subset U_1.$$

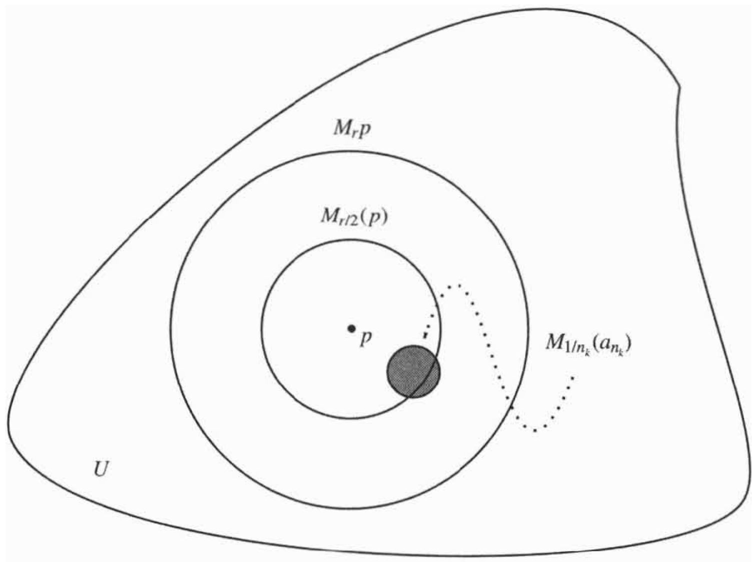

Figure 48 The neighborhood $M_r p$ engulfs the smaller neighborhood $M_{1/n_k}(a_{n_k})$.

If $U_1 \supset A$ then $\mathcal{U}$ reduces to the finite subcovering $\{U_1\}$ consisting of a single set, and the implication (b) $\Rightarrow$ (a) is proved. On the other hand, as is more likely, if U_1 does not contain A then we choose a point $a_2 \in A \setminus U_1$ and $U_2 \in \mathcal{U}$ such that

$$M_\lambda(a_2) \subset U_2.$$

Either $\mathcal{U}$ reduces to the finite subcovering $\{U_1, U_2\}$ (and the proof is finished) or else we can continue, eventually producing a sequence (a_n) in A and a sequence (U_n) in $\mathcal{U}$ such that

$$M_\lambda(a_n) \subset U_n \text{ and } a_{n+1} \in (A \setminus (U_1 \cup \cdots \cup U_n)).$$

We will show that such sequences (a_n), (U_n) lead to a contradiction. By sequential compactness, there is a subsequence (a_{n_k}) that converges to some $p \in A$. For a large k we have $d(a_{n_k}, p) < \lambda$ and

$$p \in M_\lambda(a_{n_k}) \subset U_{n_k}.$$

See Figure 49.

All a_{n_ℓ} with $\ell > k$ lie outside U_{n_k}, which contradicts their convergence to p. Thus, at some finite stage the process of choosing points a_n and sets U_n terminates, and $\mathcal{U}$

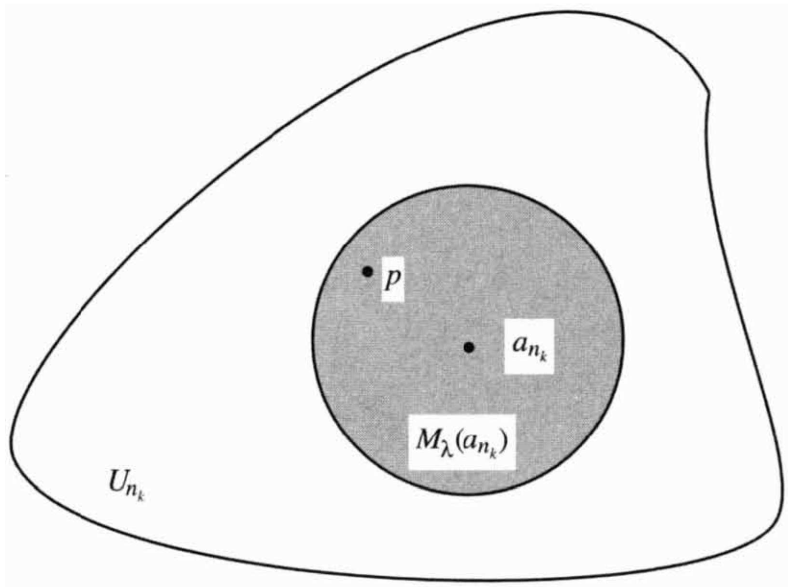

Figure 49 The point a_{n_k} is so near p that the neighborhood $M_\lambda(a_{n_k})$ engulfs p.

reduces to a finite subcovering $\{U_1, \ldots, U_n\}$ of A, which implies that A is covering compact. See also the remark on page 421. □

Upshot In light of Theorem 63, the term "compact" may now be applied equally to any set obeying (a) or (b).

Total Boundedness

The Heine-Borel Theorem states that a subset of $\mathbb{R}^m$ is compact if and only if it is closed and bounded. In more general metric spaces, such as $\mathbb{Q}$, the assertion is false. But what if the metric space is complete? As remarked on page 81 it is still false.

But mathematicians do not quit easily. The Heine-Borel Theorem ought to generalize beyond $\mathbb{R}^m$ somehow. Here is the concept we need: A set $A \subset M$ is **totally bounded** if for each $\epsilon > 0$ there exists a finite covering of A by ϵ-neighborhoods. No mention is made of a covering reducing to a subcovering. How close total boundedness is to the worthless fact that every metric space has a finite open covering!

65 Generalized Heine-Borel Theorem *A subset of a complete metric space is compact if and only if it is closed and totally bounded.*

Proof Let A be a compact subset of M. Therefore it is closed. To see that it is totally bounded, let $\epsilon > 0$ be given and consider the covering of A by ϵ-neighborhoods,

$$\{M_\epsilon x \,:\, x \in A\}.$$

Compactness of A implies that this covering reduces to a finite subcovering and therefore A is totally bounded.

Conversely, assume that A is a closed and totally bounded subset of the complete metric space M. We claim that A is sequentially compact. That is, every sequence (a_n) in A has a subsequence that converges in A. Set $\epsilon_k = 1/k,\ k = 1, 2, \ldots$. Since A is totally bounded we can cover it by finitely many ϵ_1-neighborhoods

$$M_{\epsilon_1}(q_1),\ \ldots,\ M_{\epsilon_1}(q_m).$$

By the pigeonhole principle, terms of the sequence a_n lie in at least one of these neighborhoods infinitely often, say it is $M_{\epsilon_1}(p_1)$. Choose

$$a_{n_1} \in A_1 \,=\, A \cap M_{\epsilon_1}(p_1).$$

Every subset of a totally bounded set is totally bounded, so we can cover A_1 by finitely many ϵ_2-neighborhoods. For one of them, say $M_{\epsilon_2}(p_2)$, a_n lies in $A_2 = A_1 \cap M_{\epsilon_2}(p_2)$ infinitely often. Choose $a_{n_2} \in A_2$ with $n_2 > n_1$.

Proceeding inductively, cover A_{k-1} by finitely many ϵ_k-neighborhoods, one of which, say $M_{\epsilon_k}(p_k)$, contains terms of the sequence (a_n) infinitely often. Then choose $a_{n_k} \in A_k = A_{k-1} \cap M_{\epsilon_k}(p_k)$ with $n_k > n_{k-1}$. Then (a_{n_k}) is a subsequence of (a_n). It is Cauchy. For if $\epsilon > 0$ is given we choose N such that $2/N < \epsilon$. If $k, \ell \geq N$ then

$$a_{n_k}, a_{n_\ell} \in A_N \quad \text{and} \quad \operatorname{diam} A_N \leq 2\epsilon_N \,=\, \frac{2}{N} < \epsilon,$$

which shows that (a_{n_k}) is Cauchy. Completeness of M implies that (a_{n_k}) converges to some $p \in M$ and since A is closed we have $p \in A$. Hence A is compact. $\square$

66 Corollary *A metric space is compact if and only if it is complete and totally bounded.*

Proof Every compact metric space M is complete. This is because, given a Cauchy sequence (p_n) in M, compactness implies that some subsequence converges in M, and if a Cauchy sequence has a convergent subsequence then the mother sequence converges too. As observed above, compactness immediately gives total boundedness.

Conversely, assume that M is complete and totally bounded. Every metric space is closed in itself. By Theorem 65, M is compact. $\square$

8 Cantor Sets

Cantor sets are fascinating examples of compact sets that are maximally disconnected. (To emphasize the disconnectedness, one sometimes refers to a Cantor set as "Cantor dust.") Here is how to construct the standard **Cantor set**. Start with the unit interval $[0,1]$ and remove its open middle third, $(1/3, 2/3)$. Then remove the open middle third from the remaining two intervals, and so on. This gives a nested sequence $C^0 \supset C^1 \supset C^2 \supset \ldots$ where $C^0 = [0,1]$, C^1 is the union of the two intervals $[0,1/3]$ and $[2/3,1]$, C^2 is the union of four intervals $[0,1/9]$, $[2/9,1/3]$, $[2/3,7/9]$, and $[8/9,1]$, C^3 is the union of eight intervals, and so on. See Figure 50.

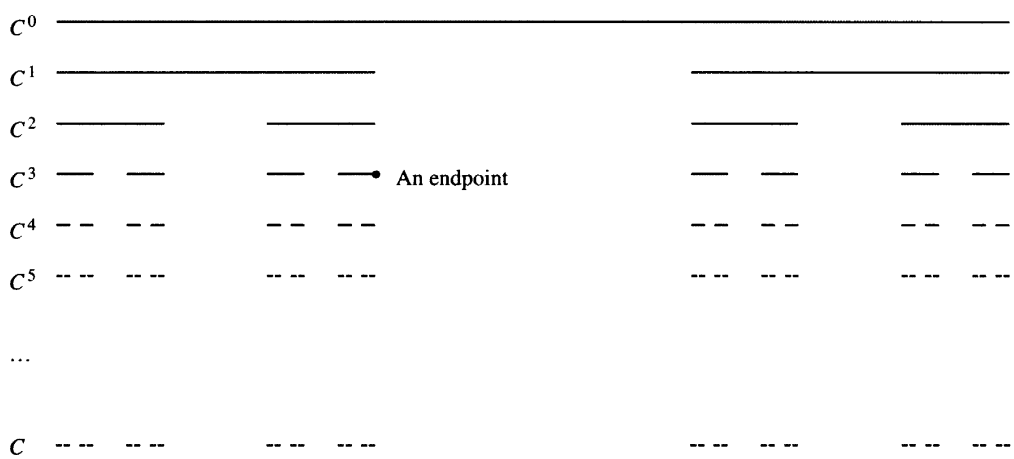

Figure 50 The construction of the standard middle-thirds Cantor set C

In general C^n is the union of 2^n closed intervals, each of length $1/3^n$. Each C^n is compact. The **standard middle thirds Cantor set** is the nested intersection

$$C = \bigcap_{n=0}^{\infty} C^n.$$

We refer to C as "the" Cantor set. Clearly it contains the endpoints of each of the intervals comprising C^n. Actually, it contains uncountably many more points than these endpoints! There are other Cantor sets defined by removing, say, middle fourths, pairs of middle tenths, etc. All Cantor sets turn out to be homeomorphic to the standard Cantor set. See Section 9.

A metric space M is **totally disconnected** if each point $p \in M$ has arbitrarily small clopen neighborhoods. That is, given $\epsilon > 0$ and $p \in M$, there exists a clopen set U such that

$$p \in U \subset M_\epsilon p.$$

For example, every discrete space is totally disconnected. So is $\mathbb{Q}$.

67 Theorem *The Cantor set is a compact, nonempty, perfect, and totally disconnected metric space.*

Proof The metric on C is the one it inherits from $\mathbb{R}$, the usual distance $|x-y|$. Let E be the set of endpoints of all the C^n-intervals,

$$E = \{0,\ 1,\ 1/3,\ 2/3,\ 1/9,\ 2/9,\ 7/9,\ 8/9,\ \ldots\}.$$

Clearly E is denumerable and contained in C, so C is nonempty and infinite. It is compact because it is the intersection of compacts.

To show C is perfect and totally disconnected, take any $x \in C$ and any $\epsilon > 0$. Fix n so large that $1/3^n < \epsilon$. The point x lies in one of the 2^n intervals I of length $1/3^n$ that comprise C^n. Fix this I. The set $E \cap I$ is infinite and contained in the interval $(x-\epsilon,\ x\ +\epsilon)$. Thus C clusters at x and C is perfect. See Figure 51.

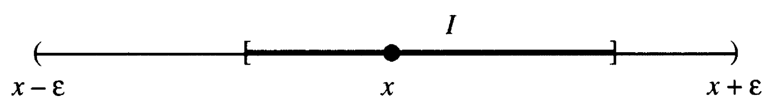

Figure 51 The endpoints of C cluster at x.

The interval I is closed in $\mathbb{R}$ and therefore in C^n. The complement $J = C^n \setminus I$ consists of finitely many closed intervals and is therefore closed too. Thus, I and J are clopen in C^n. By the Inheritance Principle their intersections with C are clopen in C, so $C \cap I$ is a clopen neighborhood of x in C which is contained in the ϵ-neighborhood of x, completing the proof that C is totally disconnected. $\square$

68 Corollary *The Cantor set is uncountable.*

Proof Being compact, C is complete, and by Theorem 56, every complete, perfect, nonempty metric space is uncountable. $\square$

A more direct way to see that the Cantor set is uncountable involves a geometric coding scheme. Take the code $0 =$ left and $2 =$ right. Then

$$C_0 = \text{ left interval } = [0, 1/3] \qquad C_2 = \text{ right interval } = [2/3, 1],$$

and $C^1 = C_0 \cup C_2$. Similarly, the left and right subintervals of C_0 are coded C_{00} and C_{02}, while the left and right subintervals of C_2 are C_{20} and C_{22}. This gives

$$C^2 = C_{00} \sqcup C_{02} \quad \sqcup \quad C_{20} \sqcup C_{22}.$$

The intervals that comprise C^3 are specified by strings of length 3. For instance, C_{220} is the left subinterval of C_{22}. In general an interval of C^n is coded by an **address string** of n symbols, each a 0 or a 2. Read it like a zip code. The first symbol gives the interval's gross location (left or right), the second symbol refines the location, the third refines it more, and so on.

Imagine now an **infinite address string** $\omega = \omega_1\omega_2\omega_3\ldots$ of zeros and twos. Corresponding to ω, we form a nested sequence of intervals

$$C_{\omega_1} \supset C_{\omega_1\omega_2} \supset C_{\omega_1\omega_2\omega_3} \supset \cdots \supset C_{\omega_1\ldots\omega_n} \supset \ldots,$$

the intersection of which is a point $p = p(\omega) \in C$. Specifically,

$$p(\omega) = \bigcap_{n \in \mathbb{N}} C_{\omega|n}$$

where $\omega|n = \omega_1 \ldots \omega_n$ **truncates** ω to an address of length n. See Theorem 34.

As we have observed, each infinite address string defines a point in the Cantor set. Conversely, each point $p \in C$ has an address $\omega = \omega(p)$: its first n symbols $\alpha = \omega|n$ are specified by the interval C_α of C^n in which p lies. A second point q has a different address, since there is some n for which p and q lie in distinct intervals C_α and C_β of C^n.

In sum, the Cantor set is in one-to-one correspondence with the set Ω of addresses. Each address $\omega \in \Omega$ defines a point $p(\omega) \in C$ and each point $p \in C$ has a unique address $\omega(p)$. The set Ω is uncountable. In fact it corresponds bijectively to $\mathbb{R}$. See Exercise 112.

If $S \subset M$ and $\overline{S} = M$ then S is **dense** in M. For example, $\mathbb{Q}$ is dense in $\mathbb{R}$. The set S is **somewhere dense** if there exists an open nonempty set $U \subset M$ such that $\overline{S \cap U} \supset U$. If S is not somewhere dense then it is **nowhere dense**.

69 Theorem *The Cantor set contains no interval and is nowhere dense in* $\mathbb{R}$.

Proof Suppose not and C contains (a, b). Then $(a, b) \subset C^n$ for all $n \in \mathbb{N}$. Take n with $1/3^n < b - a$. Since (a, b) is connected it lies wholly in a single C^n-interval, say I. But I has smaller length than (a, b), which is absurd, so C contains no interval.

Next, suppose C is dense in some nonempty open set $U \subset \mathbb{R}$, i.e., the closure of $C \cap U$ contains U. Thus

$$C = \overline{C} \supset \overline{C \cap U} \supset U \supset (a, b),$$

contrary to the fact that C contains no interval. $\square$

The existence of an uncountable nowhere dense set is astonishing. Even more is true: The Cantor set is a **zero set** – it has "outer measure zero." By this we mean that, given any $\epsilon > 0$, there is a countable covering of C by open intervals (a_k, b_k), and the **total length** of the covering is

$$\sum_{k=1}^{\infty} b_k - a_k < \epsilon.$$

(Outer measure is one of the central concepts of Lebesgue Theory. See Chapter 6.) After all, C is a subset of C^n, which consists of 2^n closed intervals, each of length $1/3^n$. If n is large enough then $2^n/3^n < \epsilon$. Enlarging each of these closed intervals to an open interval keeps the sum of the lengths $< \epsilon$, and it follows that C is a zero set.

If we discard subintervals of $[0,1]$ in a different way, we can make a **fat Cantor set** – one that has positive outer measure. Instead of discarding the middle-thirds of intervals at the n^{th} stage in the construction, we discard only the middle $1/n!$ portion. The discards are grossly smaller than the remaining intervals. See Figure 52. The total amount discarded from $[0,1]$ is < 1, and the total amount remaining, the outer measure of the fat Cantor set, is positive. See Exercise 3.31.

Figure 52 In forming a fat Cantor set, the gap intervals occupy a progressively smaller proportion of the Cantor set intervals.

9* Cantor Set Lore

In this section, we explore some arcane features of Cantor sets.

Although the continuous image of a connected set is connected, the continuous image of a disconnected set may well be connected. Just crush the disconnected set to a single point. Nevertheless, I hope you find the following result striking, for it means that the Cantor set C is the **universal compact metric space**, of which all others are merely shadows.

70 Cantor Surjection Theorem *Given a compact nonempty metric space M, there is a continuous surjection of C onto M.*

See Figure 53. Exercise 114 suggests a direct construction of a continuous surjection $C \to [0,1]$, which is already an interesting fact. The proof of Theorem 70

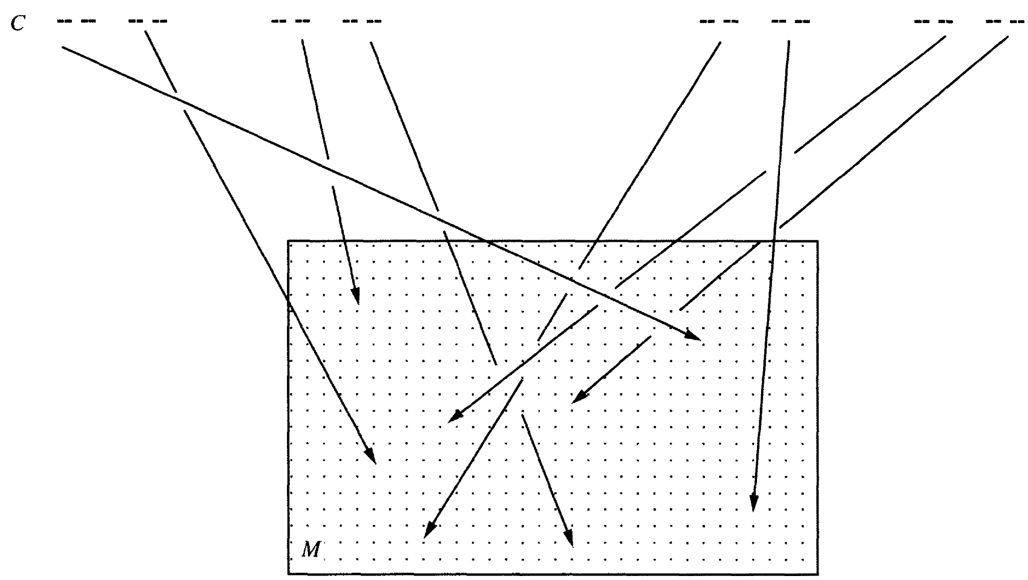

Figure 53 σ surjects C onto M.

involves a careful use of the address notation from Section 8 and the following simple lemma about dividing a compact metric space M into small pieces. A **piece** of M is any compact nonempty subset of M.

71 Lemma *If M is a nonempty compact metric space and $\epsilon > 0$ is given then M can be expressed as the finite union of pieces, each of diameter $\leq \epsilon$.*

Proof Reduce the covering $\{M_{\epsilon/2}(x) : x \in M\}$ of M to a finite subcovering and take the closure of each member of the subcovering. □

We say that M **divides into** these small pieces. The metaphor is imperfect because the pieces may overlap. The strategy of the proof of Theorem 70 is to divide M into large pieces, divide the large pieces into small pieces, divide the small pieces into smaller pieces and continue indefinitely. Labeling the pieces coherently with words in two letters leads to the Cantor surjection.

Let $W(n)$ be the set of words in two letters, say a and b, having length n. Then $\#W(n) = 2^n$. For example $W(2)$ consists of the four words aa, bb, ab, and ba.

Using Lemma 71 we divide M into a finite number of pieces of diameter ≤ 1 and we denote by $\mathcal{M}_1$ the collection of these pieces. We choose n_1 with $2^{n_1} \geq \#\mathcal{M}_1$ and choose any surjection $w_1 : W(n_1) \to \mathcal{M}_1$. Since there are enough words in $W(n_1)$, w_1 exists. We say w_1 **labels** $\mathcal{M}_1$ and if $w_1(\alpha) = L$ then α is a **label** of L.

Then we divide each $L \in \mathcal{M}_1$ into finitely many smaller pieces. Let $\mathcal{M}_2(L)$ be the collection of these smaller pieces and let

$$\mathcal{M}_2 = \bigcup_{L \in \mathcal{M}_1} \mathcal{M}_2(L).$$

Choose n_2 such that $2^{n_2} \geq \max\{\#\mathcal{M}_2(L) : L \in \mathcal{M}_1\}$ and label $\mathcal{M}_2$ with words $\alpha\beta \in W(n_1 + n_2)$ such that

$$\begin{aligned} \text{If } L = \; & w_1(\alpha) \text{ then } \alpha\beta \text{ labels the pieces } S \in \mathcal{M}_2(L) \\ & \text{as } \beta \text{ varies in } W(n_2). \end{aligned}$$

This labeling amounts to a surjection $w_2 : W(n_1 + n_2) \to \mathcal{M}_2$ that is **coherent** with w_1 in the sense that $\beta \mapsto w_2(\alpha\beta)$ labels the pieces $S \in w_1(\alpha)$. Since there are enough words in $W(n_2)$, w_2 exists. If there are other labels α' of $L \in \mathcal{M}_1$ then we get other labels $\alpha'\beta'$ for the pieces $S \in \mathcal{M}_2(L)$. We make no effort to correlate them.

Proceeding by induction we get finer and finer divisions of M coherently labeled with longer and longer words. More precisely there is a sequence of divisions $(\mathcal{M}_k)$ and surjections $w_k : W_k = W(n_1 + \cdots + n_k) \to \mathcal{M}_k$ such that

(a) The maximum diameter of the pieces $L \in \mathcal{M}_k$ tends to zero as $k \to \infty$.

(b) $\mathcal{M}_{k+1}$ **refines** $\mathcal{M}_k$ in the sense that each $S \in \mathcal{M}_{k+1}$ is contained in some $L \in \mathcal{M}_k$. ("The small pieces S are contained in the large pieces L.")

(c) If $L \in \mathcal{M}_k$ and $\mathcal{M}_{k+1}(L)$ denotes $\{S \in \mathcal{M}_{k+1} : S \subset L\}$ then

$$L = \bigcup_{S \in \mathcal{M}_{k+1}(L)} S.$$

(d) The labelings w_k are **coherent** in the sense that if $w_k(\alpha) = L \in \mathcal{M}_k$ then $\beta \mapsto w_{k+1}(\alpha\beta)$ labels $\mathcal{M}_{k+1}(L)$ as β varies in $W(n_{k+1})$.

See Figure 54.

Proof of the Cantor Surjection Theorem We are given a nonempty compact metric space M and we seek a continuous surjection $\sigma : C \to M$ where C is the standard Cantor set.

$C = \bigcap C^n$ where C^n is the disjoint union of 2^n closed intervals of length $1/3^n$. In Section 8 we labeled these C^n-intervals with words in the letters 0 and 2 having length n. (For instance C_{220} is the left C^3-interval of $C_{22} = [8/9, 1]$, namely $C_{220} = [8/9, 25/27]$.) We showed there is a natural bijection between C and the set of all infinite words in the letters 0 and 2 defined by

$$p = \bigcap_{n \in \mathbb{N}} C_{\omega|n}.$$

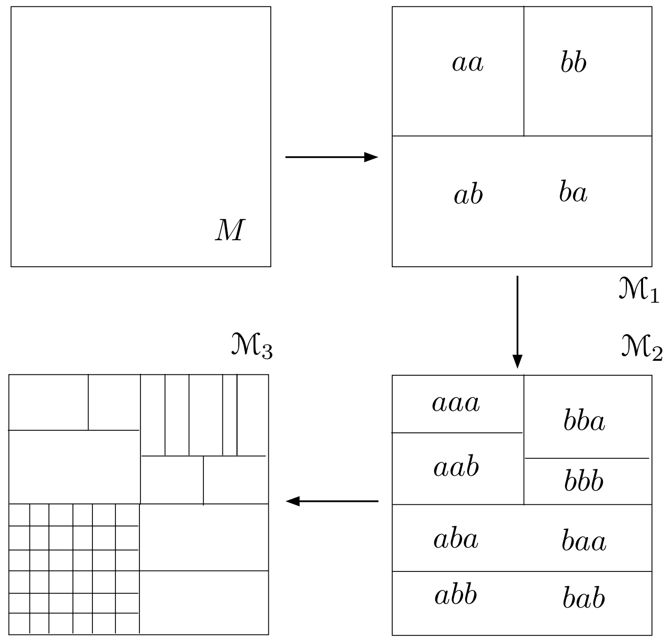

Figure 54 Coherently labeled successive divisions of M. They have $n_1 = 2$, $n_2 = 1$, and $n_3 = 6$. Note that overlabeling is necessary.

We referred to $\omega = \omega(p)$ as the address of p. ($\omega|n$ is the truncation of ω to its first n letters.) See page 107.

For $k = 1, 2, \ldots$ let $\mathcal{M}_k$ be the fine divisions of M constructed above, coherently labeled by w_k. They obey (a)-(d). Given $p \in C$ we look at the nested sequence of pieces $L_k(p) \in \mathcal{M}_k$ such that $L_k(p) = w_k(\omega|(n_1 + \cdots + n_k))$ where $\omega = \omega(p)$. That is, we truncate $\omega(p)$ to its first $n_1 + \cdots + n_k$ letters and look at the piece in $\mathcal{M}_k$ with this label. (We replace the letters 0 and 2 with a and b.) Then $(L_k(p))$ is a nested decreasing sequence of nonempty compact sets whose diameters tend to 0 as $k \to \infty$. Thus $\bigcap L_k(p)$ is a well defined point in M and we set

$$\sigma(p) \ = \ \bigcap_{k \in \mathbb{N}} L_k(p).$$

We must show that σ is a continuous surjection $C \to M$. Continuity is simple. If $p, p' \in C$ are close together then for large n the first n entries of their addresses are equal. This implies that $\sigma(p)$ and $\sigma(p')$ belong to a common L_k and k is large. Since the diameter of L_k tends to 0 as $k \to \infty$ we get continuity.

Surjectivity is also simple. Each $q \in M$ is the intersection of at least one nested sequence of pieces $L_k \in \mathcal{M}_k$. For q belongs to some piece $L_1 \in \mathcal{M}_1$, and it also belongs to some subpiece $L_2 \in \mathcal{M}_2(L_1)$, etc. Coherence of the labeling of the $\mathcal{M}_k$ implies that for each nested sequence (L_k) there is an infinite word $\alpha = \alpha_1\alpha_2\alpha_3\ldots$ such that $\alpha_i \in W(n_i)$ and $L_k = w_k(\alpha_1 \ldots \alpha_m)$ with $m = n_1 + \cdots + n_k$. The point $p \in C$ with address α is sent by σ to q. $\square$

Peano Curves

72 Theorem *There exists a* ***Peano curve****, a continuous path in the plane which is* ***space-filling*** *in the sense that its image has nonempty interior. In fact there is a Peano curve whose image is the closed unit disc B^2.*

Proof Let $\sigma : C \to B^2$ be a continuous surjection supplied by Theorem 70. Extend σ to a map $\tau : [0, 1] \to B^2$ by setting

$$\tau(x) = \begin{cases} \sigma(x) & \text{if } x \in C \\ (1-t)\sigma(a) + t\sigma(b) & \text{if } x = (1-t)a + tb \in (a,b) \\ & \text{and } (a,b) \text{ is a gap interval.} \end{cases}$$

A **gap interval** is an interval $(a, b) \subset C^c$ such that $a, b \in C$. Because σ is continuous, $|\sigma(a) - \sigma(b)| \to 0$ as $|a - b| \to 0$. Hence τ is continuous. Its image includes the disc B^2 and thus has nonempty interior. In fact the image of τ is exactly B^2, since the disc is convex and τ just extends σ via linear interpolation. See Figure 55. $\square$

This Peano curve cannot be one-to-one since C is not homeomorphic to B^2. (C is disconnected while B^2 is connected.) In fact no Peano curve τ can be one-to-one. See Exercise 102.

Cantor Spaces

We say that M is a **Cantor space** if, like the standard Cantor set C, it is compact, nonempty, perfect, and totally disconnected.

73 Moore-Kline Theorem *Every Cantor space M is homeomorphic to the standard middle-thirds Cantor set C.*

A **Cantor piece** is a nonempty clopen subset S of a Cantor space M. It is easy to see that S is also a Cantor space. See Exercise 100. Since a Cantor space is totally disconnected, each point has a small clopen neighborhood N. Thus, a Cantor space can always be divided into two disjoint Cantor pieces, $M = U \sqcup U^c$.

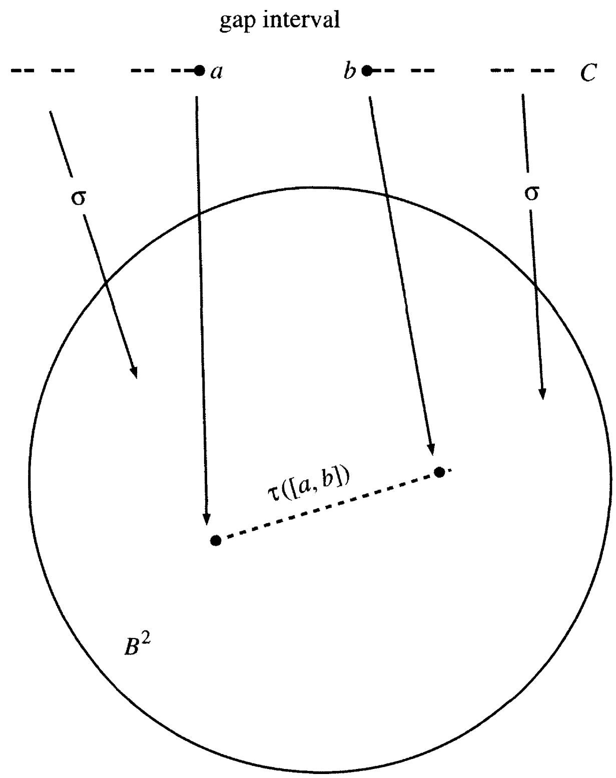

Figure 55 Filling in the Cantor surjection σ to make a Peano space-filling curve τ

74 Cantor Partition Lemma *Given a Cantor space M and $\epsilon > 0$, there is a number N such that for each $d \geq N$ there is a partition of M into d Cantor pieces of diameter $\leq \epsilon$. (We care most about dyadic d.)*

Proof A **partition** of a set is a division of it into disjoint subsets. In this case the small Cantor pieces form a partition of the Cantor space M. Since M is totally disconnected and compact, we can cover it with finitely many clopen neighborhoods $U_1, \ldots, U_m$ having diameter $\leq \epsilon$. To make the sets U_i disjoint, define

$$\begin{aligned} V_1 &= U_1 \\ V_2 &= U_1 \setminus U_2 \\ &\cdots \\ V_m &= U_m \setminus (U_1 \cup \cdots \cup U_{m-1}). \end{aligned}$$

If any V_i is empty, discard it. This gives a partition $M = X_1 \sqcup \cdots \sqcup X_N$ into $N \leq m$ Cantor pieces of diameter $\leq \epsilon$.

If $d = N$ this finishes the proof. If $d > N$ then we inductively divide X_N into two, and then three, and eventually $d - N + 1$ disjoint Cantor pieces; say

$$X_N = Y_1 \sqcup \cdots \sqcup Y_{d-N+1}.$$

The partition $M = X_1 \sqcup \cdots \sqcup X_{N-1} \sqcup Y_1 \sqcup \cdots \sqcup Y_{d-N+1}$ finishes the proof. □

Proof of the Moore-Kline Theorem We are given a Cantor space M and we seek a homeomorphism from the standard Cantor set C onto M.

By Lemma 74 there is a partition $\mathcal{M}_1$ of M into d_1 nonempty Cantor pieces where $d_1 = 2^{n_1}$ is dyadic and the pieces have diameter ≤ 1. Thus there is a bijection $w_1 : W_1 \to \mathcal{M}_1$ where $W_1 = W(n_1)$.

According to the same lemma, each $L \in \mathcal{M}_1$ can be partitioned into $N(L)$ Cantor pieces of diameter $\leq 1/2$. Choose a dyadic number

$$d_2 = 2^{n_2} \geq \max\{N(L) : L \in \mathcal{M}_1\}$$

and use the lemma again to partition each L into d_2 smaller Cantor pieces. These pieces constitute $\mathcal{M}_2(L)$, and we set $\mathcal{M}_2 = \bigcup_L \mathcal{M}_2(L)$. It is a partition of M having cardinality $d_1 d_2$ and in the natural way described in the proof of Theorem 70 it is coherently labeled by $W_2 = W(n_1 + n_2)$. Specifically, for each $L \in \mathcal{M}_1$ there is a bijection $w_L : W(n_2) \to \mathcal{M}_2(L)$ and we define $w_2 : W_2 \to \mathcal{M}_2$ by $w_2(\alpha\beta) = S \in \mathcal{M}_2$ if and only if $w_1(\alpha) = L$ and $w_L(\beta) = S$. This w_2 is a bijection.

Proceeding in exactly the same way, we pass from 2 to 3, from 3 to 4, and eventually from k to $k+1$, successively refining the partitions and extending the bijective labelings.

The Cantor surjection constructed in the proof of Theorem 70 is

$$\sigma(p) = \bigcap_k L_k(p)$$

where $L_k(p) \in \mathcal{M}_k$ has label $\omega(p)|m$ with $m = n_1 + \cdots + n_k$. Distinct points $p, p' \in C$ have distinct addresses ω, ω'. Because the labelings w_k are bijections and the divisions $\mathcal{M}_k$ are partitions, $\omega \neq \omega'$ implies that for some k, $L_k(p) \neq L_k(p')$, and thus $\sigma(p) \neq \sigma(p')$. That is, σ is a continuous bijection $C \to M$. A continuous bijection from one compact to another is a homeomorphism. □

75 Corollary *Every two Cantor spaces are homeomorphic.*

Proof Immediate from the Moore-Kline Theorem: Each is homeomorphic to C. □

76 Corollary *The fat Cantor set is homeomorphic to the standard Cantor set.*

Proof Immediate from the Moore-Kline Theorem. □

77 Corollary *A Cantor set is homeomorphic to its own Cartesian square; that is,* $C \cong C \times C$*.*

Proof It is enough to check that $C \times C$ is a Cantor space. It is. See Exercise 99. □

The fact that a nontrivial space is homeomorphic to its own Cartesian square is disturbing, is it not?

Ambient Topological Equivalence

Although all Cantor spaces are homeomorphic to each other when considered as abstract metric spaces, they can present themselves in very different ways as subsets of Euclidean space. Two sets A, B in $\mathbb{R}^m$ are **ambiently homeomorphic** if there is a homeomorphism of $\mathbb{R}^m$ to itself that sends A onto B. For example, the sets

$$A = \{0\} \cup [1,2] \cup \{3\} \quad \text{and} \quad B = \{0\} \cup \{1\} \cup [2,3]$$

are homeomorphic when considered as metric spaces, but there is no ambient homeomorphism of $\mathbb{R}$ that carries A to B. Similarly, the trefoil knot in $\mathbb{R}^3$ is homeomorphic but not ambiently homeomorphic in $\mathbb{R}^3$ to a planar circle. See also Exercise 105.

78 Theorem *Every two Cantor spaces in* $\mathbb{R}$ *are ambiently homeomorphic.*

Let M be a Cantor space contained in $\mathbb{R}$. According to Theorem 73, M is homeomorphic to the standard Cantor set C. We want to find a homeomorphism of $\mathbb{R}$ to itself that carries C to M.

The **convex hull** of $S \subset \mathbb{R}^m$ is the smallest convex set H that contains S. When $m = 1$, H is the smallest interval that contains S.

79 Lemma *A Cantor space* $M \subset \mathbb{R}$ *can be divided into two Cantor pieces whose convex hulls are disjoint.*

Proof Obvious from one-dimensionality of $\mathbb{R}$: Choose a point $x \in \mathbb{R} \setminus M$ such that some points of M lie to the left of x and others lie to its right. Then

$$M \;=\; M \cap (-\infty, x) \;\sqcup\; (x, \infty) \cap M$$

divides M into disjoint Cantor pieces whose convex hulls are disjoint closed intervals. □

Proof of Theorem 78 Let $M \subset \mathbb{R}$ be a Cantor space. We will find a homeomorphism $\tau : \mathbb{R} \to \mathbb{R}$ sending C to M. Lemma 79 leads to Cantor divisions $\mathcal{M}_k$ such that the convex hulls of the pieces in each $\mathcal{M}_k$ are disjoint. With respect to the left/right order of $\mathbb{R}$, label these pieces in the same way that the Cantor middle third intervals are labeled: L_0 and L_2 in $\mathcal{M}_1$ are the left and right pieces of M, L_{00} and L_{02} are the left and right pieces of L_0, and so on. Then the homeomorphism $\sigma : C \to M$ constructed in Theorems 70 and 73 is automatically monotone increasing. Extend σ across the gap intervals affinely as was done in the proof of Theorem 72, and extend it to $\mathbb{R} \setminus [0, 1]$ in any affine increasing fashion such that $\tau(0) = \sigma(0)$ and $\tau(1) = \sigma(1)$. Then $\tau : \mathbb{R} \to \mathbb{R}$ extends σ to $\mathbb{R}$. The monotonicity of σ implies that τ is one-to-one, while the continuity of σ implies that τ is continuous. $\tau : \mathbb{R} \to \mathbb{R}$ is a homeomorphism that carries C onto M.

If $M' \subset \mathbb{R}$ is a second Cantor space and $\tau' : \mathbb{R} \to \mathbb{R}$ is a homeomorphism that sends C onto M' then $\tau' \circ \tau^{-1}$ is a homeomorphism of $\mathbb{R}$ that sends M onto M'. $\square$

As an example, one may construct a Cantor set in $\mathbb{R}$ by removing from $[0, 1]$ its middle third, then removing from each of the remaining intervals nine symmetrically placed subintervals; then removing from each of the remaining twenty intervals, four asymmetrically placed subintervals; and so forth. In the limit (if the lengths of the remaining intervals tend to zero) we get a nonstandard Cantor set M. According to Theorem 78, there is a homeomorphism of $\mathbb{R}$ to itself sending the standard Cantor set C onto M.

Another example is the fat Cantor set mentioned on page 108. It too is ambiently homeomorphic to C.

Theorem *Every two Cantor spaces in* $\mathbb{R}^2$ *are ambiently homeomorphic.*

We do not prove this theorem here. The key step is to show M has a dyadic disc partition. That is, M can be divided into a dyadic number of Cantor pieces, each piece contained in the interior of a small topological disc D_i, the D_i being mutually disjoint. (A topological disc is any homeomorph of the closed unit disc B^2. Smallness refers to diam D_i.) The proofs I know of the existence of such dyadic partitions are tricky cut-and-paste arguments and are beyond the scope of this book. See Moise's book, *Geometric Topology in Dimensions 2 and 3* and also Exercise 138.

Antoine's Necklace

A Cantor space $M \subset \mathbb{R}^m$ is **tame** if there is an ambient homeomorphism $h : \mathbb{R}^m \to \mathbb{R}^m$ that carries the standard Cantor set C (imagined to lie on the x_1-axis in $\mathbb{R}^m$) onto M. If M is not tame it is **wild**. Cantor spaces contained in the line or plane are tame. In 3-space, however, there are wild ones, Cantor sets A so badly embedded in $\mathbb{R}^3$ that they act like curves. It is the lack of a "ball dyadic partition lemma" that causes the problem.

The first wild Cantor set was discovered by Louis Antoine, and is known as **Antoine's Necklace**. The construction involves the solid torus or anchor ring, which is homeomorphic to the Cartesian product $B^2 \times S^1$. It is easy to imagine a necklace of solid tori: Take an ordinary steel chain and modify it so its first and last links are also linked. See Figure 56.

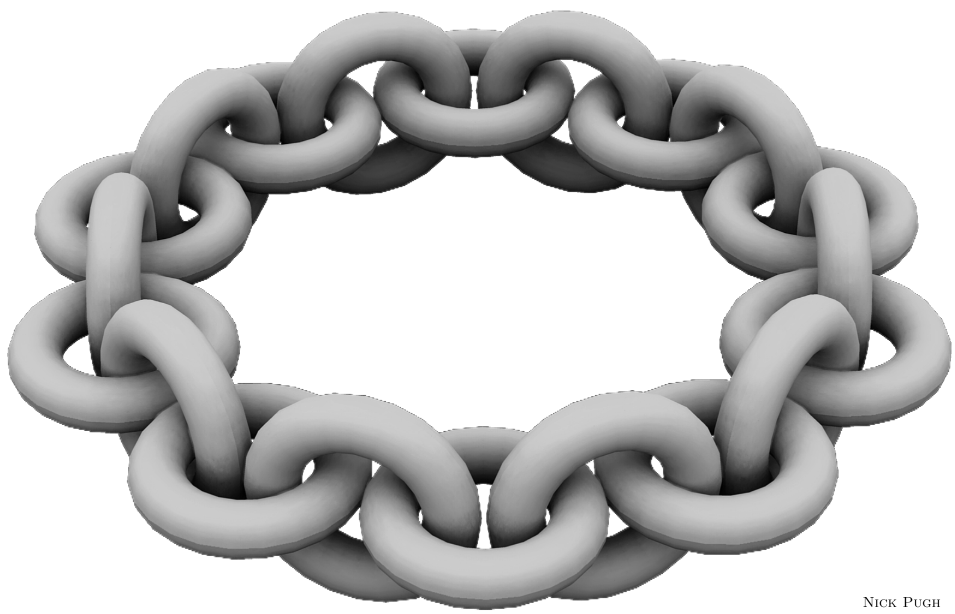

Figure 56 A necklace of twenty solid tori

Antoine's construction then goes like this. Draw a solid torus A^0. Interior to A^0, draw a necklace A^1 of several small solid tori, and make the necklace encircle the hole of A^0. Repeat the construction on each solid torus T comprising A^1. That is, interior to each T, draw a necklace of very small solid tori so that it encircles the hole of T. The result is a set $A^2 \subset A^1$ which is a necklace of necklaces. In Figure 56, A^2 would consist of 400 solid tori. Continue indefinitely, producing a nested decreasing sequence $A^0 \supset A^1 \supset A^2 \supset \dots$. The set A^n is compact and consists of a large number (20^n) of extremely small solid tori arranged in a hierarchy of necklaces. It is an n^{th} order necklace. The intersection $A = \bigcap A^n$ is a Cantor space, since it is compact, perfect, nonempty, and totally disconnected. It is homeomorphic to C. See Exercise 139.

Certainly A is bizarre, but is it wild? Is there no ambient homeomorphism h of $\mathbb{R}^3$ that sends the standard Cantor set C onto A? The reason that h cannot exist is explained next.

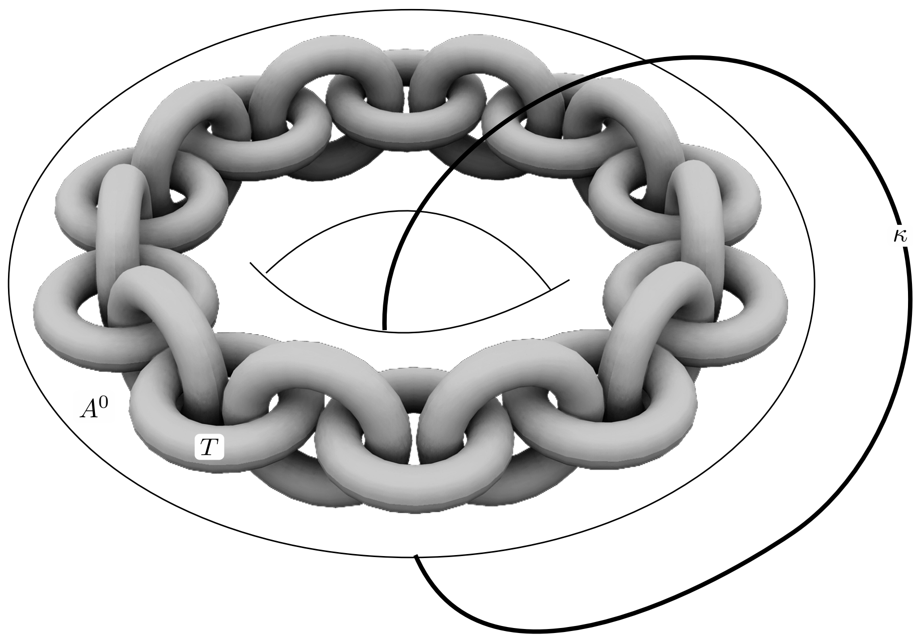

Figure 57 κ loops through A^0, which contains the necklace of solid tori.

Referring to Figure 57, the loop κ passing through the hole of A^0 cannot be continuously shrunk to a point in $\mathbb{R}^3$ without hitting A. For if such a motion of κ avoids A then, by compactness, it also avoids one of the high-order necklaces A^n. In $\mathbb{R}^3$ it is impossible to continuously de-link two linked loops, and it is also impossible to continuously de-link a loop from a necklace of loops. (These facts are intuitively believable but hard to prove. See Dale Rolfsen's book, *Knots and Links*.)

On the other hand, each loop λ in $\mathbb{R}^3 \setminus C$ can be continuously shrunk to a point without hitting C. For there is no obstruction to pushing λ through the gap intervals of C.

Now suppose that there is an ambient homeomorphism h of $\mathbb{R}^3$ that sends C to A. Then $\lambda = h^{-1}(\kappa)$ is a loop in $\mathbb{R}^3 \setminus C$, and it can be shrunk to a point in $\mathbb{R}^3 \setminus C$, avoiding C. Applying h to this motion of λ continuously shrinks κ to a point, avoiding A, which we have indicated is impossible. Hence h cannot exist, and A is wild.

10* Completion

Many metric spaces are complete (for example, every closed subset of Euclidean space is complete), and completeness is a reasonable property to require of a metric space, especially in light of the following theorem.

80 Completion Theorem *Every metric space can be completed.*

This means that just as $\mathbb{R}$ completes $\mathbb{Q}$, we can take any metric space M and find a complete metric space $\widehat{M}$ containing M whose metric extends the metric of M. To put it another way, M is always a metric subspace of a complete metric space. In a natural sense the completion is uniquely determined by M.

81 Lemma *Given four points $p, q, x, y \in M$, we have*

$$|d(p,q) - d(x,y)| \;\leq\; d(p,x) + d(q,y).$$

Proof The triangle inequality implies that

$$\begin{aligned} d(x,y) &\leq d(x,p) + d(p,q) + d(q,y) \\ d(p,q) &\leq d(p,x) + d(x,y) + d(y,q), \end{aligned}$$

and hence

$$-(d(p,x) + d(q,y)) \;\leq\; d(p,q) - d(x,y) \;\leq\; (d(p,x) + d(q,y)).$$

A number sandwiched between $-k$ and k has magnitude $\leq k$, which completes the proof. $\square$

Proof of the Completion Theorem 80 We consider the collection $\mathcal{C}$ of all Cauchy sequences in M, convergent or not, and convert *it* into the completion of M. (This is a bold idea, is it not?) Cauchy sequences (p_n) and (q_n), are **co-Cauchy** if $d(p_n, q_n) \to 0$ as $n \to \infty$. Co-Cauchyness is an equivalence relation on $\mathcal{C}$. (This is easy to check.)

Define $\widehat{M}$ to be $\mathcal{C}$ modulo the equivalence relation of being co-Cauchy. Points of $\widehat{M}$ are equivalence classes $P = [(p_n)]$ such that (p_n) is a Cauchy sequence in M. The metric on $\widehat{M}$ is

$$D(P,Q) \;=\; \lim_{n\to\infty} d(p_n, q_n),$$

where $P = [(p_n)]$ and $Q = [(q_n)]$. It only remains to verify three things:

(a) D is a well defined metric on $\widehat{M}$.
(b) $M \subset \widehat{M}$.
(c) $\widehat{M}$ is complete.

None of these assertions is really hard to prove, although the details are somewhat messy because of possible equivalence class/representative ambiguity.

(a) By Lemma 81

$$|d(p_m, q_m) - d(p_n, q_n)| \ \le \ d(p_m, p_n) + d(q_m, q_n).$$

Thus $(d(p_n, q_n))$ is a Cauchy sequence in $\mathbb{R}$, and because $\mathbb{R}$ is complete,

$$L \ = \ \lim_{n\to\infty} d(p_n, q_n)$$

exists. Let (p'_n) and (q'_n) be sequences that are co-Cauchy with (p_n) and (q_n), and let

$$L' \ = \ \lim_{n\to\infty} d(p'_n, q'_n).$$

Then

$$|L - L'| \ \le \ |L - d(p_n, q_n)| + |d(p_n, q_n) - d(p'_n, q'_n)| + |d(p'_n, q'_n) - L'|.$$

As $n \to \infty$, the first and third terms tend to 0. By Lemma 81, the middle term is

$$|d(p_n, q_n) - d(p'_n, q'_n)| \ \le \ d(p_n, p'_n) + d(q_n, q'_n),$$

which also tends to 0 as $n \to \infty$. Hence $L = L'$ and D is well defined on $\widehat{M}$. The d-distance on M is symmetric and satisfies the triangle inequality. Taking limits, these properties carry over to D on $\widehat{M}$, while positive definiteness follows directly from the co-Cauchy definition.

(b) Think of each $p \in M$ as a constant sequence, $\overline{p} = (p, p, p, p, \ldots)$. Clearly it is Cauchy and clearly the D-distance between two constant sequences $\overline{p}$ and $\overline{q}$ is the same as the d-distance between the points p and q. In this way M is naturally a metric subspace of $\widehat{M}$.

(c) Let $(P_k)_{k\in\mathbb{N}}$ be a Cauchy sequence in $\widehat{M}$. We must find $Q \in \widehat{M}$ to which P_k converges as $k \to \infty$. (Note that (P_k) is a sequence of equivalence classes, not a sequence of points in M, and convergence refers to D not d.) Because D is well defined we can use a trick to shorten the proof. Observe that every subsequence of a Cauchy sequence is Cauchy, and it and the mother sequence are co-Cauchy. For all the terms far along in the subsequence are also far along in the mother sequence. This lets us take a representative of P_k *all* of whose terms are at distance $< 1/k$ from each other. Call this sequence $(p_{k,n})_{n\in\mathbb{N}}$. We have $[(p_{k,n})] = P_k$.

Set $q_n = p_{n,n}$. We claim that (q_n) is Cauchy and $D(P_k, Q) \to 0$ as $k \to \infty$, where $Q = [(q_n)]$. That is, $\widehat{M}$ is complete.

Let $\epsilon > 0$ be given. There exists $N \geq 3/\epsilon$ such that if $k, \ell \geq N$ then

$$D(P_k, P_\ell) \leq \frac{\epsilon}{3}$$

and

$$\begin{aligned} d(q_k, q_\ell) &= d(p_{k,k}, p_{\ell,\ell}) \\ &\leq d(p_{k,k}, p_{k,n}) + d(p_{k,n}, p_{\ell,n}) + d(p_{\ell,n}, p_{\ell,\ell}) \\ &\leq \frac{1}{k} + d(p_{k,n}, p_{\ell,n}) + \frac{1}{\ell} \\ &\leq \frac{2\epsilon}{3} + d(p_{k,n}, p_{\ell,n}). \end{aligned}$$

The inequality is valid for all n and the left-hand side, $d(q_k, q_\ell)$, does not depend on n. The limit of $d(p_{k,n}, p_{\ell,n})$ as $n \to \infty$ is $D(P_k, P_\ell)$, which we know to be $< \epsilon/3$. Thus, if $k, \ell \geq N$ then $d(q_k, q_\ell) < \epsilon$ and (q_n) is Cauchy. Similarly we see that $P_k \to Q$ as $k \to \infty$. For, given $\epsilon > 0$, we choose $N \geq 2/\epsilon$ such that if $k, n \geq N$ then $d(q_k, q_n) < \epsilon/2$, from which it follows that

$$\begin{aligned} d(p_{k,n}, q_n) &\leq d(p_{k,n}, p_{k,k}) + d(p_{k,k}, q_n) \\ &= d(p_{k,n}, p_{k,k}) + d(q_k, q_n) \\ &\leq \frac{1}{k} + \frac{\epsilon}{2} < \epsilon. \end{aligned}$$

The limit of the left-hand side of this inequality, as $n \to \infty$, is $D(P_k, Q)$. Thus

$$\lim_{k \to \infty} P_k = Q$$

and $\widehat{M}$ is complete. □

Uniqueness of the completion is not surprising, and is left as Exercise 106. A different proof of the Completion Theorem is sketched in Exercise 4.39.

A Second Construction of $\mathbb{R}$ from $\mathbb{Q}$

In the particular case that the metric space M is $\mathbb{Q}$, the Completion Theorem leads to a construction of $\mathbb{R}$ from $\mathbb{Q}$ via Cauchy sequences. Note, however, that applying the theorem as it stands involves circular reasoning, for its proof uses completeness of $\mathbb{R}$ to define the metric D. Instead, we use only the Cauchy sequence *strategy.*

Convergence and Cauchyness for sequences of rational numbers are concepts that make perfect sense without a priori knowledge of $\mathbb{R}$. Just take all epsilons and deltas in the definitions to be rational. The **Cauchy completion** $\widehat{\mathbb{Q}}$ of $\mathbb{Q}$ is the collection $\mathcal{C}$ of Cauchy sequences in $\mathbb{Q}$ modulo the equivalence relation of being co-Cauchy.

We claim that $\widehat{\mathbb{Q}}$ is a complete ordered field. That is, $\widehat{\mathbb{Q}}$ is just another version of $\mathbb{R}$. The arithmetic on $\widehat{\mathbb{Q}}$ is defined by

$$\begin{aligned} P+Q &= [(p_n+q_n)] & \qquad P-Q=[(p_n-q_n)] \\ PQ &= [(p_nq_n)] & \qquad P/Q=[(p_n/q_n)] \end{aligned}$$

where $P=[(p_n)]$ and $Q=[(q_n)]$. Of course $Q \neq [(0,0,\ldots)]$ in the fraction P/Q. Exercise 134 asks you to check that these natural definitions make $\widehat{\mathbb{Q}}$ a field. Although there are many things to check – well definedness, commutativity, and so forth – all are effortless. There are no sixteen case proofs as with cuts. Also, just as with metric spaces, $\mathbb{Q}$ is naturally a subfield of $\widehat{\mathbb{Q}}$ when we think of $r \in \mathbb{Q}$ as the constant sequence $\overline{r} = [(r,r,\ldots)]$.

That's the easy part – now the rest.

To define the order relation on $\widehat{\mathbb{Q}}$ we rework some of the cut ideas. If $P \in \widehat{\mathbb{Q}}$ has a representative $[(p_n)]$, such that for some $\epsilon > 0$, we have $p_n \geq \epsilon$ for all n then P is positive. If $-P$ is positive then P is negative.

Then we define $P \prec Q$ if $Q-P$ is positive. Exercise 135 asks you to check that this defines an order on $\widehat{\mathbb{Q}}$, consistent with the standard order $<$ on $\mathbb{Q}$ in the sense that for all $p,q \in \mathbb{Q}$ we have $p < q \iff \overline{p} \prec \overline{q}$. In particular, you are asked to prove the trichotomy property: Each $P \in \widehat{\mathbb{Q}}$ is either positive, negative, or zero, and these possibilities are mutually exclusive.

Combining Cauchyness with the definition of $\prec$ gives

$$(1) \qquad P=[(p_n)] \prec Q=[(q_n)] \iff \begin{array}{l} \text{there exist } \epsilon > 0 \text{ and } N \in \mathbb{N} \\ \text{such that for all } m,n \geq N, \\ \text{we have } p_m+\epsilon < q_n. \end{array}$$

It remains to check the least upper bound property. Let $\mathcal{P}$ be a nonempty subset of $\widehat{\mathbb{Q}}$ that is bounded above. We must find a least upper bound for $\mathcal{P}$.

First of all, since $\mathcal{P}$ is bounded there is a $B=(b_n) \in \widehat{\mathbb{Q}}$ such that $P \prec B$ for all $P \in \mathcal{P}$. We can choose B so its terms lie at distance ≤ 1 from each other. Set $b=b_1+1$. Then $\overline{b}$ is an upper bound for $\mathcal{P}$. Since $\mathbb{Q}$ is Archimedean there is an integer $m \geq b$, and $\overline{m}$ is also an upper bound for $\mathcal{P}$. By the same reasoning $\mathcal{P}$ has upper bounds $\overline{r}$ such that r is a dyadic fraction with arbitrarily large denominator 2^n.

Since $\mathcal{P}$ is nonempty, the same reasoning shows that there are dyadic fractions s with large denominators such that $\overline{s}$ is not an upper bound for $\mathcal{P}$.

We assert that the least upper bound for $\mathcal{P}$ is the equivalence class Q of the following Cauchy sequence $(q_0, q_1, q_2, \ldots)$.

(a) q_0 is the smallest integer such that $\overline{q_0}$ is an upper bound for $\mathcal{P}$.
(b) q_1 is the smallest fraction with denominator 2 such that $\overline{q_1}$ is an upper bound for $\mathcal{P}$.
(c) q_2 is the smallest fraction with denominator 4 such that $\overline{q_2}$ is an upper bound for $\mathcal{P}$.
(d) $\ldots$
(e) q_n is the smallest fraction with denominator 2^n such that $\overline{q_n}$ is an upper bound for $\mathcal{P}$.

The sequence (q_n) is well defined because some but not all dyadic fractions with denominator 2^n are upper bounds for $\mathcal{P}$. By construction (q_n) is monotone decreasing and $q_{n-1} - q_n \le 1/2^n$. Thus, if $m \le n$ then

$$\begin{aligned} 0 \le q_m - q_n &= q_m - q_{m+1} + q_{m+1} - q_{m+2} + \cdots + q_{n-1} - q_n \\ &\le \frac{1}{2^{m+1}} + \cdots + \frac{1}{2^n} < \frac{1}{2^m}. \end{aligned}$$

It follows that (q_n) is Cauchy and $Q = [(q_n)] \in \widehat{\mathbb{Q}}$.

Suppose that Q is *not* an upper bound for $\mathcal{P}$. Then there is some $P = [(p_n)] \in \mathcal{P}$ with $Q \prec P$. By (1), there is an $\epsilon > 0$ and an N such that for all $n \ge N$,

$$q_N + \epsilon < p_n.$$

It follows that $\overline{q_N} \prec P$, a contradiction to $\overline{q_N}$ being an upper bound for $\mathcal{P}$.

On the other hand suppose there is a smaller upper bound for $\mathcal{P}$, say $R = (r_n) \prec Q$. By (1) there are $\epsilon > 0$ and N such that for all $m, n \ge N$,

$$r_m + \epsilon < q_n.$$

Fix a $k \ge N$ with $1/2^k < \epsilon$. Then for all $m \ge N$,

$$r_m < q_k - \epsilon < q_k - \frac{1}{2^k}.$$

By (1), $R \prec \overline{q_k - 1/2^k}$. Since R is an upper bound for $\mathcal{P}$, so is $\overline{q_k - 1/2^k}$, a contradiction to q_k being the *smallest* fraction with denominator 2^k such that $\overline{q_k}$ is an upper bound for $\mathcal{P}$. Therefore, Q is indeed a least upper bound for $\mathcal{P}$.

This completes the verification that the Cauchy completion of $\mathbb{Q}$ is a complete ordered field. Uniqueness implies that it is isomorphic to the complete ordered field $\mathbb{R}$ constructed by means of Dedekind cuts in Section 2 of Chapter 1. Decide for yourself which of the two constructions of the real number system you like better – cuts or Cauchy sequences. Cuts make least upper bounds straightforward and algebra awkward, while with Cauchy sequences it is the reverse.

Exercises

1. An ant walks on the floor, ceiling, and walls of a cubical room. What metric is natural for the ant's view of its world? What metric would a spider consider natural? If the ant wants to walk from a point p to a point q, how could it determine the shortest path?
2. Why is the sum metric on $\mathbb{R}^2$ called the Manhattan metric and the taxicab metric?
3. What is the set of points in $\mathbb{R}^3$ at distance exactly 1/2 from the unit circle S^1 in the plane,

$$\begin{aligned} T \quad = \quad \{p \in \mathbb{R}^3 : \exists\, q \in S^1 &\text{ and } d(p,q) = 1/2 \\ &\text{and for all } q' \in S^1 \text{ we have } d(p,q) \leq d(p,q')\}? \end{aligned}$$

4. Write out a proof that the discrete metric on a set M is actually a metric.
5. For $p, q \in S^1$, the unit circle in the plane, let

$$d_a(p,q) = \min\{|\measuredangle(p) - \measuredangle(q)|\,,\ 2\pi - |\measuredangle(p) - \measuredangle(q)|\}$$

where $\measuredangle(z) \in [0, 2\pi)$ refers to the angle that z makes with the positive x-axis. Use your geometric talent to prove that d_a is a metric on S^1.
6. For $p, q \in [0, \pi/2)$ let

$$d_s(p,q) = \sin|p - q|\,.$$

Use your calculus talent to decide whether d_s is a metric.
7. Prove that every convergent sequence (p_n) in a metric space M is bounded, i.e., that for some $r > 0$, some $q \in M$, and all $n \in \mathbb{N}$, we have $p_n \in M_r q$.
8. Consider a sequence (x_n) in the metric space $\mathbb{R}$.
   (a) If (x_n) converges in $\mathbb{R}$ prove that the sequence of absolute values $(|x_n|)$ converges in $\mathbb{R}$.
   (b) State the converse.
   (c) Prove or disprove it.
9. A sequence (x_n) in $\mathbb{R}$ **increases** if $n < m$ implies $x_n \leq x_m$. It **strictly increases** if $n < m$ implies $x_n < x_m$. It **decreases** or **strictly decreases** if $n < m$ always implies $x_n \geq x_m$ or always implies $x_n > x_m$. A sequence is **monotone** if it increases or it decreases. Prove that every sequence in $\mathbb{R}$ which is monotone and bounded converges in $\mathbb{R}$.[†]
10. Prove that the least upper bound property is equivalent to the "monotone sequence property" that every bounded monotone sequence converges.

[†]This is nicely is expressed by Pierre Teilhard de Chardin, *"Tout ce qui monte converge,"* in a different context.

11. Let (x_n) be a sequence in $\mathbb{R}$.
    *(a) Prove that (x_n) has a monotone subsequence.
    (b) How can you deduce that every bounded sequence in $\mathbb{R}$ has a convergent subsequence?
    (c) Infer that you have a second proof of the Bolzano-Weierstrass Theorem in $\mathbb{R}$.
    (d) What about the Heine-Borel Theorem?
12. Let (p_n) be a sequence and $f : \mathbb{N} \to \mathbb{N}$ be a bijection. The sequence $(q_k)_{k\in\mathbb{N}}$ with $q_k = p_{f(k)}$ is a **rearrangement** of (p_n).
    (a) Are limits of a sequence unaffected by rearrangement?
    (b) What if f is an injection?
    (c) A surjection?
13. Assume that $f : M \to N$ is a function from one metric space to another which satisfies the following condition: If a sequence (p_n) in M converges then the sequence $(f(p_n))$ in N converges. Prove that f is continuous. [This result improves Theorem 4.]
14. The simplest type of mapping from one metric space to another is an **isometry**. It is a bijection $f : M \to N$ that preserves distance in the sense that for all $p, q \in M$ we have
    $$d_N(fp, fq) = d_M(p, q).$$
    If there exists an isometry from M to N then M and N are said to be **isometric**, $M \equiv N$. You might have two copies of a unit equilateral triangle, one centered at the origin and one centered elsewhere. They are isometric. Isometric metric spaces are indistinguishable as metric spaces.
    (a) Prove that every isometry is continuous.
    (b) Prove that every isometry is a homeomorphism.
    (c) Prove that $[0, 1]$ is not isometric to $[0, 2]$.
15. Prove that isometry is an equivalence relation: If M is isometric to N, show that N is isometric to M; show that each M is isometric to itself (what mapping of M to M is an isometry?); if M is isometric to N and N is isometric to P, show that M is isometric to P.
16. Is the perimeter of a square isometric to the circle? Homeomorphic? Explain.
17. Which capital letters of the Roman alphabet are homeomorphic? Are any isometric? Explain.
18. Is $\mathbb{R}$ homeomorphic to $\mathbb{Q}$? Explain.
19. Is $\mathbb{Q}$ homeomorphic to $\mathbb{N}$? Explain.
20. What function (given by a formula) is a homeomorphism from $(-1, 1)$ to $\mathbb{R}$? Is every open interval homeomorphic to $(0, 1)$? Why or why not?
21. Is the plane minus four points on the x-axis homeomorphic to the plane minus four points in an arbitrary configuration?

22. If every closed and bounded subset of a metric space M is compact, does it follow that M is complete? (Proof or counterexample.)
23. $(0,1)$ is an open subset of $\mathbb{R}$ but not of $\mathbb{R}^2$, when we think of $\mathbb{R}$ as the x-axis in $\mathbb{R}^2$. Prove this.
24. For which intervals $[a,b]$ in $\mathbb{R}$ is the intersection $[a,b] \cap \mathbb{Q}$ a clopen subset of the metric space $\mathbb{Q}$?
25. Prove directly from the definition of closed set that every singleton subset of a metric space M is a closed subset of M. Why does this imply that every finite set of points is also a closed set?
26. Prove that a set $U \subset M$ is open if and only if none of its points are limits of its complement.
27. If $S, T \subset M$, a metric space, and $S \subset T$, prove that
    (a) $\overline{S} \subset \overline{T}$.
    (b) $\operatorname{int}(S) \subset \operatorname{int}(T)$.
28. A map $f : M \to N$ is **open** if for each open set $U \subset M$, the image set $f(U)$ is open in N.
    (a) If f is open, is it continuous?
    (b) If f is a homeomorphism, is it open?
    (c) If f is an open, continuous bijection, is it a homeomorphism?
    (d) If $f : \mathbb{R} \to \mathbb{R}$ is a continuous surjection, must it be open?
    (e) If $f : \mathbb{R} \to \mathbb{R}$ is a continuous, open surjection, must it be a homeomorphism?
    (f) What happens in (e) if $\mathbb{R}$ is replaced by the unit circle S^1?
29. Let $\mathcal{T}$ be the collection of open subsets of a metric space M, and $\mathcal{K}$ the collection of closed subsets. Show that there is a bijection from $\mathcal{T}$ onto $\mathcal{K}$.
30. Consider a two-point set $M = \{a, b\}$ whose topology consists of the two sets, M and the empty set. Why does this topology not arise from a metric on M?
31. Prove the following.
    (a) If U is an open subset of $\mathbb{R}$ then it consists of countably many disjoint intervals $U = \bigsqcup U_i$. (Unbounded intervals $(-\infty, b)$, (a, ∞), and $(-\infty, \infty)$ are permitted.)
    (b) Prove that these intervals U_i are uniquely determined by U. In other words, there is only one way to express U as a disjoint union of open intervals.
    (c) If $U, V \subset \mathbb{R}$ are both open, so $U = \bigsqcup U_i$ and $V = \bigsqcup V_j$ where U_i and V_j are open intervals, show that U and V are homeomorphic if and only if there are equally many U_i and V_j.
32. Show that every subset of $\mathbb{N}$ is clopen. What does this tell you about every function $f : \mathbb{N} \to M$, where M is a metric space?

33. (a) Find a metric space in which the boundary of $M_r p$ is not equal to the sphere of radius r at p, $\partial(M_r p) \neq \{x \in M : d(x,p) = r\}$.
    (b) Need the boundary be contained in the sphere?
34. Use the Inheritance Principle to prove Corollary 15.
35. Prove that S clusters at p if and only if for each $r > 0$ there is a point $q \in M_r p \cap S$, such that $q \neq p$.
36. Construct a set with exactly three cluster points.
37. Construct a function $f : \mathbb{R} \to \mathbb{R}$ that is continuous only at points of $\mathbb{Z}$.
38. Let X, Y be metric spaces with metrics d_X, d_Y, and let $M = X \times Y$ be their Cartesian product. Prove that the three natural metrics d_E, $d_{\max}$, and d_{sum} on M are actually metrics. [Hint: Cauchy-Schwarz.]
39. (a) Prove that every convergent sequence is bounded. That is, if (p_n) converges in the metric space M, prove that there is some neighborhood $M_r q$ containing the set $\{p_n : n \in \mathbb{N}\}$.
    (b) Is the same true for a Cauchy sequence in an incomplete metric space?
40. Let M be a metric space with metric d. Prove that the following are equivalent.
    (a) M is homeomorphic to M equipped with the discrete metric.
    (b) Every function $f : M \to M$ is continuous.
    (c) Every bijection $g : M \to M$ is a homeomorphism.
    (d) M has no cluster points.
    (e) Every subset of M is clopen.
    (f) Every compact subset of M is finite.
41. Let $\| \quad \|$ be any norm on $\mathbb{R}^m$ and let $B = \{x \in \mathbb{R}^m : \|x\| \leq 1\}$. Prove that B is compact. [Hint: It suffices to show that B is closed and bounded with respect to the Euclidean metric.]
42. What is wrong with the following "proof" of Theorem 28? "Let $((a_n, b_n))$ be any sequence in $A \times B$ where A and B are compact. Compactness implies the existence of subsequences (a_{n_k}) and (b_{n_k}) converging to $a \in A$ and $b \in B$ as $k \to \infty$. Therefore $((a_{n_k}, b_{n_k}))$ is a subsequence of $((a_n, b_n))$ that converges to a limit in $A \times B$, proving that $A \times B$ is compact."
43. Assume that the Cartesian product of two nonempty sets $A \subset M$ and $B \subset N$ is compact in $M \times N$. Prove that A and B are compact.
44. Consider a function $f : M \to \mathbb{R}$. Its graph is the set
$$\{(p, y) \in M \times \mathbb{R} : y = fp\}.$$
    (a) Prove that if f is continuous then its graph is closed (as a subset of $M \times \mathbb{R}$).
    (b) Prove that if f is continuous and M is compact then its graph is compact.
    (c) Prove that if the graph of f is compact then f is continuous.
    (d) What if the graph is merely closed? Give an example of a discontinuous function $f : \mathbb{R} \to \mathbb{R}$ whose graph is closed.

45. Draw a Cantor set C on the circle and consider the set A of all chords between points of C.
    (a) Prove that A is compact.
    *(b) Is A convex?
46. Assume that A, B are compact, disjoint, nonempty subsets of M. Prove that there are $a_0 \in A$ and $b_0 \in B$ such that for all $a \in A$ and $b \in B$ we have
$$d(a_0, b_0) \leq d(a, b).$$
[The points a_0, b_0 are closest together.]
47. Suppose that $A, B \subset \mathbb{R}^2$.
    (a) If A and B are homeomorphic, are their complements homeomorphic?
    *(b) What if A and B are compact?
    ***(c) What if A and B are compact and connected?
48. Prove that there is an embedding of the line as a closed subset of the plane, and there is an embedding of the line as a bounded subset of the plane, but there is no embedding of the line as a closed and bounded subset of the plane.
*49. Construct a subset $A \subset \mathbb{R}$ and a continuous bijection $f : A \to A$ that is not a homeomorphism. [Hint: By Theorem 36 A must be noncompact.]
**50. Construct nonhomeomorphic connected, closed subsets $A, B \subset \mathbb{R}^2$ for which there exist continuous bijections $f : A \to B$ and $g : B \to A$. [Hint: By Theorem 36 A and B must be noncompact.]
***51. Do there exist nonhomeomorphic closed sets $A, B \subset \mathbb{R}$ for which there exist continuous bijections $f : A \to B$ and $g : B \to A$?
52. Let (A_n) be a nested decreasing sequence of nonempty closed sets in the metric space M.
    (a) If M is complete and $\operatorname{diam} A_n \to 0$ as $n \to \infty$, show that $\bigcap A_n$ is exactly one point.
    (b) To what assertions do the sets $[n, \infty)$ provide counterexamples?
53. Suppose that (K_n) is a nested sequence of compact nonempty sets, $K_1 \supset K_2 \supset \ldots$, and $K = \bigcap K_n$. If for some $\mu > 0$, $\operatorname{diam} K_n \geq \mu$ for all n, is it true that $\operatorname{diam} K \geq \mu$?
54. If $f : A \to B$ and $g : C \to B$ such that $A \subset C$ and for each $a \in A$ we have $f(a) = g(a)$ then g **extends** f. We also say that f **extends to** g. Assume that $f : S \to \mathbb{R}$ is a uniformly continuous function defined on a subset S of a metric space M.
    (a) Prove that f extends to a uniformly continuous function $\overline{f} : \overline{S} \to \mathbb{R}$.
    (b) Prove that $\overline{f}$ is the unique continuous extension of f to a function defined on $\overline{S}$.
    (c) Prove the same things when $\mathbb{R}$ is replaced with a complete metric space N.

55. The **distance** from a point p in a metric space M to a nonempty subset $S \subset M$ is defined to be $\operatorname{dist}(p, S) = \inf\{d(p,s) : s \in S\}$.
    (a) Show that p is a limit of S if and only if $\operatorname{dist}(p, S) = 0$.
    (b) Show that $p \mapsto \operatorname{dist}(p, S)$ is a uniformly continuous function of $p \in M$.
56. Prove that the 2-sphere is not homeomorphic to the plane.
57. If S is connected, is the interior of S connected? Prove this or give a counterexample.
58. Theorem 49 states that the closure of a connected set is connected.
    (a) Is the closure of a disconnected set disconnected?
    (b) What about the interior of a disconnected set?

*59. Prove that every countable metric space (not empty and not a singleton) is disconnected. [Astonishingly, there exists a countable topological space which is connected. Its topology does not arise from a metric.]

60. (a) Prove that a continuous function $f : M \to \mathbb{R}$, all of whose values are integers, is constant provided that M is connected.
    (b) What if all the values are irrational?
61. Prove that the (double) cone $\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = z^2\}$ is path-connected.
62. Prove that the annulus $A = \{z \in \mathbb{R}^2 : r \leq |z| \leq R\}$ is connected.
63. A subset E of $\mathbb{R}^m$ is **starlike** if it contains a point p_0 (called a **center** for E) such that for each $q \in E$, the segment between p_0 and q lies in E.
    (a) If E is convex and nonempty prove that it is starlike.
    (b) Why is the converse false?
    (c) Is every starlike set connected?
    (d) Is every connected set starlike? Why or why not?

*64. Suppose that $E \subset \mathbb{R}^m$ is open, bounded, and starlike, and p_0 is a center for E.
    (a) Is it true or false that all points p_1 in a small enough neighborhood of p_0 are also centers for E?
    (b) Is the set of centers convex?
    (c) Is it closed as a subset of E?
    (d) Can it consist of a single point?

65. Suppose that $A, B \subset \mathbb{R}^2$ are convex, closed, and have nonempty interiors.
    (a) Prove that A, B are the closure of their interiors.
    (b) If A, B are compact, prove that they are homeomorphic.
    [Hint: Draw a picture.]
66. (a) Prove that every connected open subset of $\mathbb{R}^m$ is path-connected.
    (b) Is the same true for open connected subsets of the circle?
    (c) What about connected nonopen subsets of the circle?
67. List the convex subsets of $\mathbb{R}$ up to homeomorphism. How many are there and how many are compact?

68. List the closed convex sets in $\mathbb{R}^2$ up to homeomorphism. There are nine. How many are compact?

*69. Generalize Exercises 65 and 68 to $\mathbb{R}^3$; to $\mathbb{R}^m$.

70. Prove that (a, b) and $[a, b)$ are not homeomorphic metric spaces.

71. Let M and N be nonempty metric spaces.
    (a) If M and N are connected prove that $M \times N$ is connected.
    (b) What about the converse?
    (c) Answer the questions again for path-connectedness.

72. Let H be the hyperbola $\{(x, y) \in \mathbb{R}^2 : xy = 1 \text{ and } x, y > 0\}$ and let X be the x-axis.
    (a) Is the set $S = X \cup H$ connected?
    (b) What if we replace H with the graph G of any continuous positive function $f : \mathbb{R} \to (0, \infty)$; is $X \cup G$ connected?
    (c) What if f is everywhere positive but discontinuous at just one point.

73. Is the disc minus a countable set of points connected? Path-connected? What about the sphere or the torus instead of the disc?

74. Let $S = \mathbb{R}^2 \setminus \mathbb{Q}^2$. (Points $(x, y) \in S$ have at least one irrational coordinate.) Is S connected? Path-connected? Prove or disprove.

*75. An **arc** is a path with no self-intersection. Define the concept of arc-connectedness and prove that a metric space is path-connected if and only if it is arc-connected.

76. (a) The intersection of connected sets need not be connected. Give an example.
    (b) Suppose that $S_1, S_2, S_3, \ldots$ is a sequence of connected, closed subsets of the plane and $S_1 \supset S_2 \supset \ldots$. Is $S = \bigcap S_n$ connected? Give a proof or counterexample.
    *(c) Does the answer change if the sets are compact?
    (d) What is the situation for a nested decreasing sequence of compact path-connected sets?

77. If a metric space M is the union of path-connected sets S_α, all of which have the nonempty path-connected set K in common, is M path-connected?

78. $(p_1, \ldots, p_n)$ is an **ϵ-chain** in a metric space M if for each i we have $p_i \in M$ and $d(p_i, p_{i+1}) < \epsilon$. The metric space is **chain-connected** if for each $\epsilon > 0$ and each pair of points $p, q \in M$ there is an ϵ-chain from p to q.
    (a) Show that every connected metric space is chain-connected.
    (b) Show that if M is compact and chain-connected then it is connected.
    (c) Is $\mathbb{R} \setminus \mathbb{Z}$ chain-connected?
    (d) If M is complete and chain-connected, is it connected?

79. Prove that if M is nonempty, compact, locally path-connected, and connected then it is path-connected. (See Exercise 143, below.)

80. The **Hawaiian earring** is the union of circles of radius $1/n$ and center $x = \pm 1/n$ on the x-axis, for $n \in \mathbb{N}$. See Figure 27 on page 58.
   (a) Is it connected?
   (b) Path-connected?
   (c) Is it homeomorphic to the one-sided Hawaiian earring?

*81. The topologist's sine curve is the set

$$\{(x, y) : x = 0 \text{ and } |y| \leq 1 \text{ or } 0 < x \leq 1 \text{ and } y = \sin 1/x\}.$$

See Figure 43. The **topologist's sine circle** is shown in Figure 58. (It is the union of a circular arc and the topologist's sine curve.) Prove that it is path-connected but not locally path-connected. (M is **locally path-connected** if for each $p \in M$ and each neighborhood U of p there is a path-connected subneighborhood V of p.)

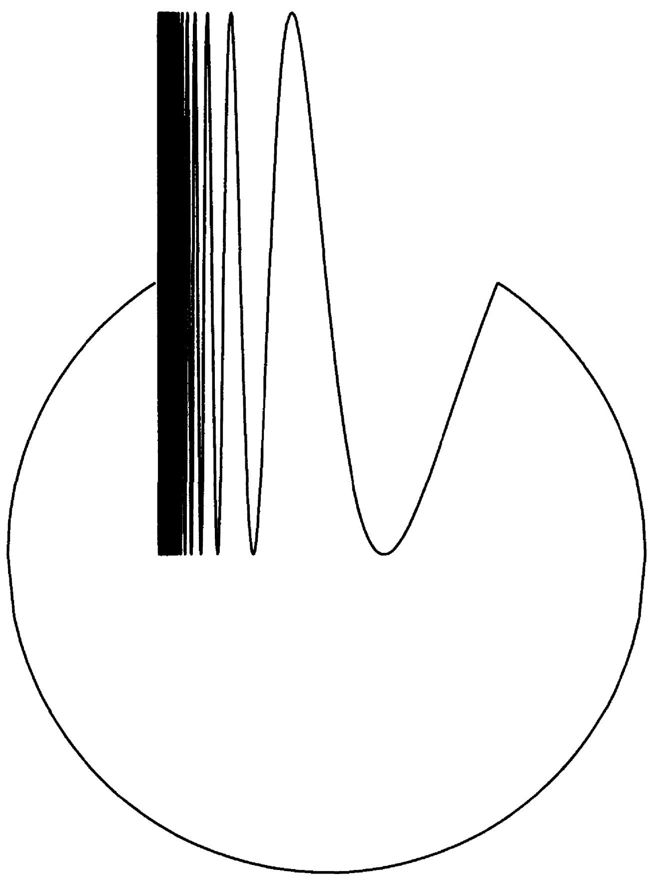

Figure 58 The topologist's sine circle

82. The graph of $f : M \to \mathbb{R}$ is the set $\{(x, y) \in M \times \mathbb{R} : y = fx\}$.
    (a) If M is connected and f is continuous, prove that the graph of f is connected.
    (b) Give an example to show that the converse is false.
    (c) If M is path-connected and f is continuous, show that the graph is path-connected.
    (d) What about the converse?
83. The open cylinder is $(0, 1) \times S^1$. The punctured plane is $\mathbb{R}^2 \setminus \{0\}$.
    (a) Prove that the open cylinder is homeomorphic to the punctured plane.
    (b) Prove that the open cylinder, the double cone, and the plane are not homeomorphic.
84. Is the closed strip $\{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 1\}$ homeomorphic to the closed half-plane $\{(x, y) \in \mathbb{R}^2 : x \geq 0\}$? Prove or disprove.
85. Suppose that M is compact and that $\mathcal{U}$ is an open covering of M which is "redundant" in the sense that each $p \in M$ is contained in at least two members of $\mathcal{U}$. Show that $\mathcal{U}$ reduces to a finite subcovering with the same property.
86. Suppose that every open covering of M has a positive Lebesgue number. Give an example of such an M that is not compact.

Exercises 87–94 treat the basic theorems in the chapter, avoiding the use of sequences. The proofs will remain valid in general topological spaces.

87. Give a direct proof that $[a, b]$ is covering compact. [Hint: Let $\mathcal{U}$ be an open covering of $[a, b]$ and consider the set

$$C = \{x \in [a, b] : \text{ finitely many members of } \mathcal{U} \text{ cover } [a, x]\}.$$

    Use the least upper bound principle to show that $b \in C$.]
88. Give a direct proof that a closed subset A of a covering compact set K is covering compact. [Hint: If $\mathcal{U}$ is an open covering of A, adjoin the set $W = M \setminus A$ to $\mathcal{U}$. Is $\mathcal{W} = \mathcal{U} \cup \{W\}$ an open covering of K? If so, so what?]
89. Give a proof of Theorem 36 using open coverings. That is, assume A is a covering compact subset of M and $f : M \to N$ is continuous. Prove directly that fA is covering compact. [Hint: What is the criterion for continuity in terms of preimages?]
90. Suppose that $f : M \to N$ is a continuous bijection and M is covering compact. Prove directly that f is a homeomorphism.
91. Suppose that M is covering compact and that $f : M \to N$ is continuous. Use the Lebesgue number lemma to prove that f is uniformly continuous. [Hint: Consider the covering of N by $\epsilon/2$-neighborhoods $\{N_{\epsilon/2}(q) : q \in N\}$ and its preimage in M, $\{f^{\mathrm{pre}}(N_{\epsilon/2}(q)) : q \in N\}$.]

92. Give a direct proof that the nested decreasing intersection of nonempty covering compact sets is nonempty. [Hint: If $A_1 \supset A_2 \supset \ldots$ are covering compact, consider the open sets $U_n = A_n^c$. If $\bigcap A_n = \emptyset$, what does $\{U_n\}$ cover?]
93. Generalize Exercise 92 as follows. Suppose that M is covering compact and $\mathcal{C}$ is a collection of closed subsets of M such that every intersection of finitely many members of $\mathcal{C}$ is nonempty. (Such a collection $\mathcal{C}$ is said to have the **finite intersection property**.) Prove that the **grand intersection** $\bigcap_{C \in \mathcal{C}} C$ is nonempty. [Hint: Consider the collection of open sets $\mathcal{U} = \{C^c : C \in \mathcal{C}$.]
94. If every collection of closed subsets of M which has the finite intersection property also has a nonempty grand intersection, prove that M is covering compact. [Hint: Given an open covering $\mathcal{U} = \{U_\alpha\}$, consider the collection of closed sets $\mathcal{C} = \{U_\alpha^c\}$.]
95. Let S be a subset of a metric space M. With respect to the definitions on page 92 prove the following.
    (a) The closure of S is the intersection of all closed subsets of M that contain S.
    (b) The interior of S is the union of all open subsets of M that are contained in S.
    (c) The boundary of S is a closed set.
    (d) Why does (a) imply the closure of S equals $\lim S$?
    (e) If S is clopen, what is ∂S?
    (f) Give an example of $S \subset \mathbb{R}$ such that $\partial(\partial S) \neq \emptyset$, and infer that "the boundary of the boundary $\partial \circ \partial$ is not always zero."
96. If $A \subset B \subset C$, A is dense in B, and B is dense in C prove that A is dense in C.
97. Is the set of dyadic rationals (the denominators are powers of 2) dense in $\mathbb{Q}$? In $\mathbb{R}$? Does one answer imply the other? (Recall that A is dense in B if $A \subset B$ and $\overline{A} \supset B$.)
98. Show that $S \subset M$ is somewhere dense in M if and only if $\operatorname{int}(\overline{S}) \neq \emptyset$. Equivalently, S is nowhere dense in M if and only if its closure has empty interior.
99. Let M, N be nonempty metric spaces and $P = M \times N$.
    (a) If M, N are perfect prove that P is perfect.
    (b) If M, N are totally disconnected prove that P is totally disconnected.
    (c) What about the converses?
    (d) Infer that the Cartesian product of Cantor spaces is a Cantor space. (We already know that the Cartesian product of compacts is compact.)
    (e) Why does this imply that $C \times C = \{(x, y) \in \mathbb{R}^2 : x \in C \text{ and } y \in C\}$ is homeomorphic to C, C being the standard Cantor set?
100. Prove that every Cantor piece is a Cantor space. (Recall that M is a Cantor space if it is compact, nonempty, totally disconnected and perfect, and that $A \subset M$ is a Cantor piece if it is nonempty and clopen.)

*101. Let Σ be the set of all infinite sequences of zeroes and ones. For example, $(100111000011111\ldots) \in \Sigma$. Define the metric

$$d(a,b) = \sum \frac{|a_n - b_n|}{2^n}$$

where $a = (a_n)$ and $b = (b_n)$ are points in Σ.

(a) Prove that Σ is compact.

(b) Prove that Σ is homeomorphic to the Cantor set.

102. Prove that no Peano curve is one-to-one. (Recall that a Peano curve is a continuous map $f : [0,1] \to \mathbb{R}^2$ whose image has a nonempty interior.)

103. Prove that there is a continuous surjection $\mathbb{R} \to \mathbb{R}^2$. What about $\mathbb{R}^m$?

104. Find two nonhomeomorphic compact subsets of $\mathbb{R}$ whose complements are homeomorphic.

105. As on page 115, consider the subsets of $\mathbb{R}$,

$$A = \{0\} \cup [1, 2] \cup \{3\} \quad \text{and} \quad B = \{0\} \cup \{1\} \cup [2, 3].$$

(a) Why is there no ambient homeomorphism of $\mathbb{R}$ to itself that carries A onto B?

(b) Thinking of $\mathbb{R}$ as the x-axis, is there an ambient homeomorphism of $\mathbb{R}^2$ to itself that carries A onto B?

106. Prove that the completion of a metric space is unique in the following natural sense: A completion of a metric space M is a complete metric X space containing M as a metric subspace such that M is dense in X. That is, every point of X is a limit of M.

(a) Prove that M is dense in the completion $\widehat{M}$ constructed in the proof of Theorem 80.

(b) If X and X' are two completions of M prove that there is an isometry $i : X \to X'$ such that $i(p) = p$ for all $p \in M$.

(c) Prove that i is the unique such isometry.

(d) Infer that $\widehat{M}$ is unique.

107. If M is a metric subspace of a complete metric space S prove that $\overline{M}$ is a completion of M.

*108. Consider the identity map id : $C_{\max} \to C_{\text{int}}$ where $C_{\max}$ is the metric space $C([0,1],\mathbb{R})$ of continuous real-valued functions defined on $[0,1]$, equipped with the max-metric $d_{\max}(f,g) = \max|f(x) - g(x)|$, and C_{int} is $C([0,1],\mathbb{R})$ equipped with the integral metric,

$$d_{\text{int}}(f,g) = \int_0^1 |f(x) - g(x)|\, dx.$$

Show that id is a continuous linear bijection (an isomorphism) but its inverse is not continuous.

*109. A metric on M is an **ultrametric** if for all $x, y, z \in M$ we have

$$d(x,z) \ \leq \max\{d(x,y), d(y,z)\}.$$

(Intuitively this means that the trip from x to z cannot be broken into shorter legs by making a stopover at some y.)

(a) Show that the ultrametric property implies the triangle inequality.

(b) In an ultrametric space show that "all triangles are isosceles."

(c) Show that a metric space with an ultrametric is totally disconnected.

(d) Define a metric on the set Σ of strings of zeroes and ones in Exercise 101 as

$$d_*(a,b) = \begin{cases} \dfrac{1}{2^n} & \text{if } n \text{ is the smallest index for which } a_n \neq b_n \\ \\ 0 & \text{if } a = b. \end{cases}$$

Show that d_* is an ultrametric and prove that the identity map is a homeomorphism $(\Sigma, d) \to (\Sigma, d_*)$.

*110. $\mathbb{Q}$ inherits the Euclidean metric from $\mathbb{R}$ but it also carries a very different metric, the ***p*-adic** metric. Given a prime number p and an integer n, the p-adic norm of n is

$$|n|_p = \frac{1}{p^k}$$

where p^k is the largest power of p that divides n. (The norm of 0 is by definition 0.) The more factors of p, the smaller the p-norm. Similarly, if $x = a/b$ is a fraction, we factor x as

$$x = p^k \cdot \frac{r}{s}$$

where p divides neither r nor s, and we set

$$|x|_p = \frac{1}{p^k}.$$

The p-adic metric on $\mathbb{Q}$ is

$$d_p(x,y) = |x - y|_p.$$

(a) Prove that d_p is a metric with respect to which $\mathbb{Q}$ is perfect – every point is a cluster point.

(b) Prove that d_p is an ultrametric.

(c) Let $\mathbb{Q}_p$ be the metric space completion of $\mathbb{Q}$ with respect to the metric d_p, and observe that the extension of d_p to $\mathbb{Q}_p$ remains an ultrametric. Infer from Exercise 109 that $\mathbb{Q}_p$ is totally disconnected.

(d) Prove that $\mathbb{Q}_p$ is locally compact, in the sense that every point has small compact neighborhoods.

(e) Infer that $\mathbb{Q}_p$ is covered by neighborhoods homeomorphic to the Cantor set. See Gouvêa's book, *p-adic Numbers*.

111. Let $M = [0,1]$ and let $\mathcal{M}_1$ be its division into two intervals [0, 1/2] and [1/2, 1]. Let $\mathcal{M}_2$ be its division into four intervals [0, 1/4], [1/4, 1/2], [1/2, 3/4], and [3/4, 1]. Continuing these bisections generates natural divisions of $[0,1]$. The pieces are intervals. We label them with words using the letters 0 and 1 as follows: 0 means "left" and 1 means "right," so the four intervals in $\mathcal{M}_2$ are labeled as 00, 01, 10, and 11 respectively.

(a) Verify that all endpoints of the intervals (except 0 and 1) have two addresses. For instance,

$$\bigcap_k \left[\frac{2^{k-1}-1}{2^k},\ \frac{1}{2}\right] = \left\{\frac{1}{2}\right\} = \bigcap_k \left[\frac{1}{2},\ \frac{2^{k-1}+1}{2^k}\right].$$

(b) Verify that the points $0, 1$, and all nonendpoints have unique addresses.

*112. Prove that $\#C = \#\mathbb{R}$. [Hint: According to the Schroeder-Bernstein Theorem from Chapter 1 it suffices to find injections $C \to \mathbb{R}$ and $\mathbb{R} \to C$. The inclusion $C \subset \mathbb{R}$ is an injection $C \to \mathbb{R}$. Each $t \in [0,1)$ has a unique base-2 expansion $\tau(t)$ that does not terminate in an infinite string of ones. Replacing each 1 by 2 converts $\tau(t)$ to $\omega(t)$, an infinite address in the symbols 0 and 2. It does not terminate in an infinite string of twos. Set $h(t) = \sum_{i=1}^{\infty} \omega_i/3^i$ and verify that $h : [0,1) \to C$ is an injection. Since there is an injection $\mathbb{R} \to [0,1)$, conclude that there is an injection $\mathbb{R} \to C$, and hence that $\#C = \#\mathbb{R}$.]

Remark The Continuum Hypothesis states that if S is any uncountable subset of $\mathbb{R}$ then S and $\mathbb{R}$ have equal cardinality. The preceding coding shows that C is not only uncountable (as is implied by Theorem 56) but actually has the same cardinality as $\mathbb{R}$. That is, C is not a counterexample to the Continuum Hypothesis. The same is true of all uncountable closed subsets of $\mathbb{R}$. See Exercise 151.

113. Let M be the standard Cantor set C. In the notation of Section 8, C^n is the collection of 2^n Cantor intervals of length $1/3^n$ that nest down to C as $n \to \infty$. Verify that setting $\mathcal{C}_k = C \cap C^k$ gives divisions of C into disjoint clopen pieces.

*114. (a) Prove directly that there is a continuous surjection of the middle-thirds Cantor set C onto the closed interval $[0,1]$. [Hint: Each $x \in C$ has a base 3 expansion (x_n), all of whose entries are zeroes and twos. (For example, $2/3 = (2\overline{0})_{\text{base } 3}$ and $1/3 = (0\overline{2})_{\text{base } 3}$. Write $y = (y_n)$ by replacing the twos in (x_n) by ones and interpreting the answer base 2. Show that the map $x \mapsto y$ works.]

(b) Compare this surjection to the one constructed from the bisection divisions in Exercise 113.

115. Rotate the unit circle S^1 by a fixed angle α, say $R : S^1 \to S^1$. (In polar coordinates, the transformation R sends $(1, \theta)$ to $(1, \theta + \alpha)$.)
  (a) If α/π is rational, show that each orbit of R is a finite set.
  *(b) If α/π is irrational, show that each orbit is infinite and has closure equal to S^1.

116. A metric space M with metric d can always be remetrized so the metric becomes bounded. Simply define the **bounded metric**

$$\rho(p, q) = \frac{d(p, q)}{1 + d(p, q)}.$$

  (a) Prove that ρ is a metric. Why is it obviously bounded?
  (b) Prove that the identity map $M \to M$ is a homeomorphism from M with the d-metric to M with the ρ-metric.
  (c) Infer that boundedness of M is not a topological property.
  (d) Find homeomorphic metric spaces, one bounded and the other not.

117. Fold a piece of paper in half.
  (a) Is this a continuous transformation of one rectangle into another?
  (b) Is it injective?
  (c) Draw an open set in the target rectangle, and find its preimage in the original rectangle. Is it open?
  (d) What if the open set meets the crease?

The **baker's transformation** is a similar mapping. A rectangle of dough is stretched to twice its length and then folded back on itself. Is the transformation continuous? A formula for the baker's transformation in one variable is $f(x) = 1 - |1 - 2x|$. The $\boldsymbol{n^{\text{th}}}$ **iterate** of f is $f^n = f \circ f \circ \ldots \circ f$, n times. The **orbit** of a point x is

$$\{x,\ f(x),\ f^2(x),\ \ldots,\ f^n(x),\ \ldots\}.$$

[For clearer but more awkward notation one can write $f^{\circ n}$ instead of f^n. This distinguishes composition $f \circ f$ from multiplication $f \cdot f$.]
  (e) If x is rational prove that the orbit of x is a finite set.
  (f) If x is irrational what is the orbit?

*118. The implications of compactness are frequently equivalent to it. Prove
  (a) If every continuous function $f : M \to \mathbb{R}$ is bounded then M is compact.
  (b) If every continuous bounded function $f : M \to \mathbb{R}$ achieves a maximum or minimum then M is compact.
  (c) If every continuous function $f : M \to \mathbb{R}$ has compact range fM then M is compact.

(d) If every nested decreasing sequence of nonempty closed subsets of M has nonempty intersection then M is compact.

Together with Theorems 63 and 65, (a)–(d) give seven equivalent definitions of compactness. [Hint: Reason contrapositively. If M is not compact then it contains a sequence (p_n) that has no convergent subsequence. It is fair to assume that the points p_n are distinct. Find radii $r_n > 0$ such that the neighborhoods $M_{r_n}(p_n)$ are disjoint and no sequence $q_n \in M_{r_n}(p_n)$ has a convergent subsequence. Using the metric define a function $f_n : M_{r_n}(p_n) \to \mathbb{R}$ with a spike at p_n, such as

$$f_n(x) = \frac{r_n - d(x, p_n)}{a_n + d(x, p_n)}$$

where $a_n > 0$. Set $f(x) = f_n(x)$ if $x \in M_{r_n}(p_n)$, and $f(x) = 0$ if x belongs to no $M_{r_n}(p_n)$. Show that f is continuous. With the right choice of a_n show that f is unbounded. With a different choice of a_n, it is bounded but achieves no maximum, and so on.]

119. Let M be a metric space of diameter ≤ 2. The **cone** for M is the set

$$C = C(M) = \{p_0\} \cup M \times (0, 1]$$

with the **cone metric**

$$\begin{aligned} \rho((p,s),\,(q,t)) &= |s-t| + \min\{s,t\}d(p,q) \\ \rho((p,s),\,p_0) &= s \\ \rho(p_0,p_0) &= 0. \end{aligned}$$

The point p_0 is the vertex of the cone. Prove that ρ is a metric on C. [If M is the unit circle, think of it in the plane $z = 1$ in $\mathbb{R}^3$ centered at the point $(0, 0, 1)$. Its cone is the 45-degree cone with vertex the origin.]

120. Recall that if for each embedding of M, $h : M \to N$, hM is closed in N then M is said to be absolutely closed. If each hM is bounded then M is absolutely bounded. Theorem 41 implies that compact sets are absolutely closed and absolutely bounded. Prove:

(a) If M is absolutely bounded then M is compact.

*(b) If M is absolutely closed then M is compact.

Thus these are two more conditions equivalent to compactness. [Hint: From Exercise 118(a), if M is noncompact there is a continuous function $f : M \to \mathbb{R}$ that is unbounded. For Exercise 120(a), show that $F(x) = (x, f(x))$ embeds M onto a nonbounded subset of $M \times \mathbb{R}$. For 120(b), justify the additional assumption that the metric on M is bounded by 2. Then use Exercise 118(b) to show that if M is noncompact then there is a continuous function $g : M \to (0, 1]$ such that for some nonclustering sequence (p_n), we have $g(p_n) \to 0$ as $n \to \infty$. Finally, show that $G(x) = (x, gx)$ embeds M onto a nonclosed subset S of the cone $C(M)$ discussed in Exercise 119. S will be nonclosed because it limits at p_0 but does not contain it.]

121. (a) Prove that every function defined on a discrete metric space is uniformly continuous.
   (b) Infer that it is false to assert that if every continuous function $f : M \to \mathbb{R}$ is uniformly continuous then M is compact.
   (c) Prove, however, that if M is a metric subspace of a compact metric space K and every continuous function $f : M \to \mathbb{R}$ is uniformly continuous then M is compact.

122. Recall that p is a cluster point of S if each $M_r p$ contains infinitely many points of S. The set of cluster points of S is denoted as S'. Prove:
   (a) If $S \subset T$ then $S' \subset T'$.
   (b) $(S \cup T)' = S' \cup T'$.
   (c) $S' = (\overline{S})'$.
   (d) S' is closed in M; that is, $S'' \subset S'$ where $S'' = (S')'$.
   (e) Calculate $\mathbb{N}'$, $\mathbb{Q}'$, $\mathbb{R}'$, $(\mathbb{R} \setminus \mathbb{Q})'$, and $\mathbb{Q}''$.
   (f) Let T be the set of points $\{1/n : n \in \mathbb{N}\}$. Calculate T' and T''.
   (g) Give an example showing that S'' can be a proper subset of S'.

123. Recall that p is a condensation point of S if each $M_r p$ contains uncountably many points of S. The set of condensation points of S is denoted as S^*. Prove:
   (a) If $S \subset T$ then $S^* \subset T^*$.
   (b) $(S \cup T)^* = S^* \cup T^*$.
   (c) $S^* \subset \overline{S}^*$ where $\overline{S}^* = (\overline{S})^*$
   (d) S^* is closed in M; that is, $S^{*\prime} \subset S^*$ where $S^{*\prime} = (S^*)'$.
   (e) $S^{**} \subset S^*$ where $S^{**} = (S^*)^*$.
   (f) Calculate $\mathbb{N}^*$, $\mathbb{Q}^*$, $\mathbb{R}^*$, and $(\mathbb{R} \setminus \mathbb{Q})^*$.
   (g) Give an example showing that S^* can be a proper subset of $(\overline{S})^*$. Thus, (c) is not in general an equality.
   **(h) Give an example that S^{**} can be a proper subset of S^*. Thus, (e) is not in general an equality. [Hint: Consider the set M of all functions $f : [a, b] \to [0, 1]$, continuous or not, and let the metric on M be the sup metric, $d(f, g) = \sup\{|f(x) - g(x)| : x \in [a, b]\}$. Consider the set S of all "δ-functions with rational values."]
   **(i) Give examples that show in general that S^* neither contains nor is contained in S'^* where $S'^* = (S')^*$. [Hint: δ-functions with values $1/n$, $n \in \mathbb{N}$.]

124. Recall that p is an interior point of $S \subset M$ if some $M_r p$ is contained in S. The set of interior points of S is the interior of S and is denoted $\operatorname{int} S$. For all subsets S, T of the metric space M prove:
   (a) $\operatorname{int} S = S \setminus \partial S$.
   (b) $\operatorname{int} S = (\overline{S^c})^c$.

(c) $\text{int}(\text{int}\, S) = \text{int}\, S$.
(d) $\text{int}(S \cap T) = \text{int}(S \cap \text{int}\, T$.
(e) What are the dual equations for the closure?
(f) Prove that $\text{int}(S \cup T) \supset \text{int}\, S \cup \text{int}\, T$. Show by example that the inclusion can be strict, i.e., not an equality.

125. A point p is a boundary point of a set $S \subset M$ if every neighborhood $M_r p$ contains points of both S and S^c. The boundary of S is denoted ∂S. For all subsets S, T of a metric space M prove:
(a) S is clopen if and only if $\partial S = \emptyset$.
(b) $\partial S = \partial S^c$.
(c) $\partial\partial S \subset \partial S$.
(d) $\partial\partial\partial S = \partial\partial S$.
(e) $\partial(S \cup T) \quad \subset \quad \partial S \cup \partial T$.
(f) Give an example in which (c) is a strict inclusion, $\partial\partial S \neq \partial S$.
(g) What about (e)?

*126. Suppose that E is an uncountable subset of $\mathbb{R}$. Prove that there exists a point $p \in \mathbb{R}$ at which E condenses. [Hint: Use decimal expansions. Why must there be an interval $[n, n+1)$ containing uncountably many points of E? Why must it contain a decimal subinterval with the same property? (A decimal subinterval $[a, b)$ has endpoints $a = n+k/10$, $b = n+(k+1)/10$ for some digit k, $0 \leq k \leq 9$.) Do you see lurking the decimal expansion of a condensation point?] Generalize to $\mathbb{R}^2$ and to $\mathbb{R}^m$.

127. The metric space M is **separable** if it contains a countable dense subset. [Note the confusion of language: "Separable" has nothing to do with "separation."]
(a) Prove that $\mathbb{R}^m$ is separable.
(b) Prove that every compact metric space is separable.

128. *(a) Prove that every metric subspace of a separable metric space is separable, and deduce that every metric subspace of $\mathbb{R}^m$ or of a compact metric space is separable.
(b) Is the property of being separable topological?
(c) Is the continuous image of a separable metric space separable?

129. Think up a nonseparable metric space.

130. Let $\mathcal{B}$ denote the collection of all ϵ-neighborhoods in $\mathbb{R}^m$ whose radius ϵ is rational and whose center has all coordinates rational.
(a) Prove that $\mathcal{B}$ is countable.
(b) Prove that every open subset of $\mathbb{R}^m$ can be expressed as the countable union of members of $\mathcal{B}$.

(The union need not be disjoint, but it is at most a countable union because there are only countably many members of $\mathcal{B}$. A collection such as $\mathcal{B}$ is called a **countable base** for the topology of $\mathbb{R}^m$.)

131. (a) Prove that every separable metric space has a countable base for its topology, and conversely that every metric space with a countable base for its topology is separable.
(b) Infer that every compact metric space has a countable base for its topology.

*132. Referring to Exercise 123, assume now that M is separable, $S \subset M$, and, as before S' is the set of cluster points of S while S^* is the set of condensation points of S. Prove:
(a) $S^* \subset (S')^* = (\overline{S})^*$.
(b) $S^{**} = S^{*\prime} = S^*$.
(c) Why is (a) not in general an equality?

[Hints: For (a) write $S \subset (S \setminus S') \cup S'$ and $\overline{S} = (S \setminus S') \cup S'$, show that $(S \setminus S')^* = \emptyset$, and use Exercise 123(a). For (b), Exercise 123(d) implies that $S^{**} \subset S^{*\prime} \subset S^*$. To prove that $S^* \subset S^{**}$, write $S \subset (S \setminus S^*) \cup S^*$ and show that $(S \setminus S^*)^* = \emptyset$.]

*133. Prove that
(a) An uncountable subset of $\mathbb{R}$ clusters at some point of $\mathbb{R}$.
(b) An uncountable subset of $\mathbb{R}$ clusters at some point of itself.
(c) An uncountable subset of $\mathbb{R}$ condenses at uncountably many points of itself.
(d) What about $\mathbb{R}^m$ instead of $\mathbb{R}$?
(e) What about any compact metric space?
(f) What about any separable metric space?

[Hint: Review Exercise 126.]

*134. Prove that $\widehat{\mathbb{Q}}$, the Cauchy sequences in $\mathbb{Q}$ modulo the equivalence relation of being co-Cauchy, is a field with respect to the natural arithmetic operations defined on page 122, and that $\mathbb{Q}$ is naturally a subfield of $\widehat{\mathbb{Q}}$.

135. Prove that the order on $\widehat{\mathbb{Q}}$ defined on page 122 is a bona fide order which agrees with the standard order on $\mathbb{Q}$.

*136. Let M be the square $[0,1]^2$, and let aa, ba, bb, ab label its four quadrants – upper right, upper left, lower left, and lower right.
(a) Define nested bisections of the square using this pattern repeatedly, and let τ_k be a curve composed of line segments that visit the k^{th}-order quadrants systematically. Let $\tau = \lim_k \tau_k$ be the resulting Peano curve à la the Cantor Surjection Theorem.
(b) Compare τ to the Peano curve $f : I \to I^2$ directly constructed on pages 271- 274 of the second edition of Munkres' book *Topology*.

*137. Let P be a closed perfect subset of a separable complete metric space M. Prove that each point of P is a condensation point of P. In symbols, $P = P' \Rightarrow P = P^*$.

**138. Given a Cantor space $M \subset \mathbb{R}^2$, given a line segment $[p,q] \subset \mathbb{R}^2$ with $p, q \notin M$, and given an $\epsilon > 0$, prove that there exists a path A in the ϵ-neighborhood of $[p, q]$ that joins p to q and is disjoint from M. [Hint: Think of A as a bisector of M. From this bisection fact a dyadic disc partition of M can be constructed, which leads to the proof that M is tame.]

139. To prove that Antoine's Necklace A is a Cantor set, you need to show that A is compact, perfect, nonempty, and totally disconnected.
   (a) Do so. [Hint: What is the diameter of any connected component of A^n, and what does that imply about A?]
   **(b) If, in the Antoine construction two linked solid tori are placed *very cleverly* inside each larger solid torus, show that the intersection $A = \bigcap A^n$ is a Cantor set.

*140. Consider the **Hilbert cube**

$$H = \{(x_1, x_2, \ldots) \in [0,1]^\infty : \text{ for each } n \in \mathbb{N} \text{ we have } |x_n| \leq 1/2^n\}.$$

Prove that H is compact with respect to the metric

$$d(x, y)) = \sup_n |x_n - y_n|$$

where $x = (x_n),\ y = (y_n)$. [Hint: Sequences of sequences.]

Remark Although compact, H is infinite-dimensional and is homeomorphic to no subset of $\mathbb{R}^m$.

141. Prove that the Hilbert cube is perfect and homeomorphic to its Cartesian square, $H \cong H \times H$.

***142. Assume that M is compact, nonempty, perfect, and homeomorphic to its Cartesian square, $M \cong M \times M$. Must M be homeomorphic to the Cantor set, the Hilbert cube, or some combination of them?

143. A **Peano space** is a metric space M that is the continuous image of the unit interval: There is a continuous surjection $\tau : [0, 1] \to M$. Theorem 72 states the amazing fact that the 2-disc is a Peano space. Prove that every Peano space is
   (a) compact,
   (b) nonempty,
   (c) path-connected,
   *(d) and **locally path-connected**, in the sense that for each $p \in M$ and each neighborhood U of p there is a smaller neighborhood V of p such that any two points of V can be joined by a path in U.

*144. The converse to Exercise 143 is the **Hahn-Mazurkiewicz Theorem**. Assume that a metric space M is a compact, nonempty, path-connected, and locally path-connected. Use the Cantor Surjection Theorem 70 to show that M is a Peano space. [The key is to make uniformly short paths to fill in the gaps of $[0, 1] \setminus C$.]

145. One of the famous theorems in plane topology is the **Jordan Curve Theorem**. A **Jordan curve** J is a homeomorph of the unit circle in the plane. (Equivalently it is $f([a,b])$ where $f : [a,b] \to \mathbb{R}^2$ is continuous, $f(a) = f(b)$, and for no other pair of distinct $s, t \in [a,b]$ does $f(s)$ equal $f(t)$. It is also called a **simple closed curve**.) The Jordan Curve Theorem asserts that $\mathbb{R}^2 \setminus J$ consists of two disjoint, connected open sets, its inside and its outside, and every path between them must meet J. Prove the Jordan Curve Theorem for the circle, the square, the triangle, and – if you have courage – every simple closed polygon.
146. The **utility problem** gives three houses 1, 2, 3 in the plane and the three utilities, Gas, Water, and Electricity. You are supposed to connect each house to the three utilities without crossing utility lines. (The houses and utilities are disjoint.)
    - (a) Use the Jordan curve theorem to show that there is no solution to the utility problem in the plane.
    - *(b) Show also that the utility problem cannot be solved on the 2-sphere S^2.
    - *(c) Show that the utility problem can be solved on the surface of the torus.
    - *(d) What about the surface of the Klein bottle?
    - ***(e) Given utilities $U_1, \ldots, U_m$ and houses $H_1, \ldots, H_n$ located on a surface with g handles, find necessary and sufficient conditions on m, n, g so that the utility problem can be solved.
147. Let M be a metric space and let $\mathcal{K}$ denote the class of nonempty compact subsets of M. The r-neighborhood of $A \in \mathcal{K}$ is

$$M_r A = \{x \in M : \exists a \in A \text{ and } d(x,a) < r\} = \bigcup_{a \in A} M_r a.$$

For $A, B \in \mathcal{K}$ define

$$D(A,B) = \inf\{r > 0 : A \subset M_r B \text{ and } B \subset M_r A\}.$$

- (a) Show that D is a metric on $\mathcal{K}$. (It is called the **Hausdorff metric** and $\mathcal{K}$ is called the **hyperspace** of M.)
- (b) Denote by $\mathcal{F}$ the collection of finite nonempty subsets of M and prove that $\mathcal{F}$ is dense in $\mathcal{K}$. That is, given $A \in \mathcal{K}$ and given $\epsilon > 0$ show there exists $F \in \mathcal{F}$ such that $D(A,F) < \epsilon$.
- *(c) If M is compact prove that $\mathcal{K}$ is compact.
- (d) If M is connected prove that $\mathcal{K}$ is connected.
- **(e) If M is path-connected is $\mathcal{K}$ path-connected?
- (f) Do homeomorphic metric spaces have homeomorphic hyperspaces?

Remark The converse to (f), $\mathcal{K}(M) \cong \mathcal{K}(N) \Rightarrow M \cong N$ is false. The hyperspace of every Peano space is the Hilbert cube. This is a difficult result but a good place to begin reading about hyperspaces is Sam Nadler's book *Continuum Theory*.

**148. Start with a set $S \subset \mathbb{R}$ and successively take its closure, the complement of its closure, the closure of that, and so on: S, $\mathrm{cl}(S)$, $(\mathrm{cl}(S))^c, \ldots$. Do the same to S^c. In total, how many distinct subsets of $\mathbb{R}$ can be produced this way? In particular decide whether each chain S, $\mathrm{cl}(S)$, $\ldots$ consists of only finitely many sets. For example, if $S = \mathbb{Q}$ then we get $\mathbb{Q}$, $\mathbb{R}$, $\emptyset$, $\emptyset$, $\mathbb{R}$, $\mathbb{R}, \ldots$ and $\mathbb{Q}^c$, $\mathbb{R}$, $\emptyset$, $\emptyset$, $\mathbb{R}$, $\mathbb{R}, \ldots$ for a total of four sets.

**149. Consider the letter T.

(a) Prove that there is no way to place uncountably many copies of the letter T disjointly in the plane. [Hint: First prove this when the unit square replaces the plane.]

(b) Prove that there is no way to place uncountably many homeomorphic copies of the letter T disjointly in the plane.

(c) For which other letters of the alphabet is this true?

(d) Let U be a set in $\mathbb{R}^3$ formed like an umbrella: It is a disc with a perpendicular segment attached to its center. Prove that uncountably many copies of U cannot be placed disjointly in $\mathbb{R}^3$.

(e) What if the perpendicular segment is attached to the boundary of the disc?

150. Let M be a complete, separable metric space such as $\mathbb{R}^m$. Prove the **Cupcake Theorem: Each closed set $K \subset M$ can be expressed uniquely as the disjoint union of a countable set and a perfect closed set,

$$C \sqcup P = K.$$

**151. Let M be an uncountable compact metric space.

(a) Prove that M contains a homeomorphic copy of the Cantor set. [Hint: Imitate the construction of the standard Cantor set C.]

(b) Infer that Cantor sets are ubiquitous. There is a continuous surjection $\sigma : C \to M$ and there is a continuous injection $i : C \to M$.

(c) Infer that every uncountable closed set $S \subset \mathbb{R}$ has $\#S = \#\mathbb{R}$, and hence that the Continuum Hypothesis is valid for closed sets in $\mathbb{R}$. [Hint: Cupcake and Exercise 112.]

(d) Is the same true if M is separable, uncountable, and complete?

**152. Write jingles at least as good as the following. Pay attention to the meter as well as the rhyme.

When a set in the plane
is closed and bounded,
you can always draw
a curve around it.

Peter Přibik

If a clopen set can be detected,
Your metric space is disconnected.

David Owens

A coffee cup feeling quite dazed,
said to a donut, amazed,
an open surjective continuous injection,
You'd be plastic and I'd be glazed.

Norah Esty

'Tis a most indisputable fact
If you want to make something compact
Make it bounded and closed
For you're totally hosed
If either condition you lack.
Lest the reader infer an untruth
(Which I think would be highly uncouth)
I must hasten to add
There are sets to be had
Where the converse is false, fo'sooth.

Karla Westfahl

For ev'ry a and b in S
if there exists a path that's straight
from a to b and it's inside
then "S must be convex," we state.

Alex Wang