Chapter 4

Connected Sets

We now know enough about sets to start a meaningful mathematical discussion about a piece of paper. We know that pieces of paper which are fresh from the pad have no rips or holes in them. Their surfaces are smooth and continuous. Thus, to describe this type of surface in mathematical terms, we need to develop a system of labelling that allows us to put a name to each spot on the page *without leaving any gaps.* If we cannot do this, the representation does not account for every smidge of paper and therefore is not true to what we know from experience. (Never mind that there are gaps between the atoms in the paper, you cannot see those anyway!)

This notion of "not leaving any gaps" in the paper is made more precise by the mathematical concepts of connectedness and continuity. Intuitively, a connected set is one that is "all in one piece". For example, in $\mathbb{R}$, $[0,1]$ is connected, while $[0,1]\cup[2,3]$ is not. In this chapter we will give a mathematically precise definition of "connected", and we will study some of the properties of connected sets. This study will serve as a prelude to our development of a naming system for the points on our piece of paper. We will undertake that task in Chapters 5 and 6.

The notion of connectedness is particularly useful for telling topological spaces apart. For example, if one topological space is connected and another is not, then there is an important sense in which the two spaces are not "alike". In this chapter, we will introduce a criterion for comparing topological spaces. This is important in discussing how our sheet can be crumpled up "without tearing it" (see Chapter **??**).

4.1 Separated Pairs

Let $(X, \mathbf{K})$ be a topological space and let $A, B \subset X$. We would like to define what it means for A and B to be "far apart", or "separated". Intuitively, $\mathbf{K}A$ is the set of all points close to A. If A and B are 'far apart, then no point in

B is close to A, so $\mathbf{K}A \cap B = \emptyset$. Likewise, if no point in A is close to B, then $A \cap \mathbf{K}B = \emptyset$. Thus we make the following definition.

Definition 4.1 *Let $(X, \mathbf{K})$ be a topological space and let $A, B \subset X$. We say that A and B are* **separated** *if*

$$\mathbf{K}A \cap B = A \cap \mathbf{K}B = \emptyset.$$

It goes without saying that we did not have to define "separated" in this way; however, our definition seems reasonable and satisfactory to the intuition. This said, let us see what this definition can tell us and then decide if we want to use it in the future.

The open intervals $A = (0, 1)$ and $B = (1, 2)$ are separated in $\mathbb{R}$, because

$$\mathbf{K}A \cap B = [0, 1] \cap (1, 2) = \emptyset,$$
$$A \cap \mathbf{K}B = (0, 1) \cap [1, 2] = \emptyset.$$

On the other hand, the intervals $A = [0, 1]$ and $B = (1, 2)$ are *not* separated in $\mathbb{R}$. Although

$$\mathbf{K}A \cap B = [0, 1] \cap (1, 2) = \emptyset,$$

the problem is that

$$A \cap \mathbf{K}B = [0, 1] \cap [1, 2] = \{1\} \neq \emptyset.$$

Exercises

1. Do you think it is possible to find two nonempty, separated sets A and B such that $A \cup B = \mathbb{R}$? (The answer will become clear later in the text.)

2. Prove that if A and B are closed, then $A - B$ and $B - A$ are separated.

4.2 Connected sets

Given the discussion of the previous section, the following definition should make intuitive sense.

Definition 4.2 *Let $(X, \mathbf{K})$ be a topological space, and let W be any subset of X. We say that W is* **disconnected** *if there are nonempty, separated sets A and B such that $W = A \cup B$. In this case, we call the pair (A, B) a* **separation** *of W. If no such separation exists, we say that W is* **connected**.

Note that when we say that a set W is connected, we must specify which topological space we are regarding W as a subset of. When this is not clear from context, we say something like "W is connected in $(X, \mathbf{K})$." This is necessary to avoid confusion since a set that is connected in one topological space may become separated if viewed in the context of another space.

The requirement in the above definition that A and B are nonempty is important. If we omitted it, then we would not have a very interesting definition: every set W would be disconnected, since $(W, \emptyset)$ would be a separation.

Let us look at some examples of this definition in action.

Example 4.3 The two-point set $\{0, 1\}$ is disconnected in $\mathbb{R}$.

Proof. I claim that $\{\{0\}, \{1\}\}$ is a separation of $\{0, 1\}$. Clearly $\{0\}$ and $\{1\}$ are nonempty, and their union is $\{\{0\}, \{1\}\}$. By Example 3.6, $\{0\}$ and $\{1\}$ are closed. Thus

$$\mathbf{K}\{0\} \cap \{1\} = \{0\} = \{1\} = \emptyset,$$

and likewise $\{0\} \cap \mathbf{K}\{1\} = \emptyset$.

□

Example 4.4 $\mathbb{Z}$ is disconnected in $\mathbb{R}$.

Proof. A separation is given by

$$(\{x \in \mathbb{Z} \mid x < 0\}\,,\ \{x \in \mathbb{Z} \mid x \geq 0\})\,.$$

(Indeed, each of these sets is closed in $\mathbb{R}$. We will omit the details.)

□

The integers $\mathbb{Z}$ provide an example of a **totally disconnected** set in the sense that it has no connected subset consisting of more than one point. Indeed, one can show that any set of integers is closed, and therefore that any two disjoint sets of integers are separated. (This should remind you of the discrete closure operator of Example 1.2.)

Our definition of connectedness talks about any subset W of X. In the special case $W = X$, the definition can be rephrased in a convenient way.

Proposition 4.5 *Let $(X, \mathbf{K})$ be a topological space. Then (A, B) is a separation of X if and only if $X = A \cup B$, $A \cap B = \emptyset$, A and B are nonempty, and A and B are closed.*

Proof. ($\Longrightarrow$) Let (A, B) be a separation of X. By definition, $X = A \cup B$, and A and B are nonempty. Also, by Axiom C1,

$$A \cap B \subset \mathbf{K}A \cap B = \emptyset.$$

Since $A \cap B = \emptyset$ and $X = A \cup B$, we have $A = X - B$. Since $\mathbf{K}A \cap B = \emptyset$,

$$\mathbf{K}A \subset X - B = A.$$

By Axiom C1, A is closed. Similarly, B is closed.

($\Longleftarrow$) Suppose $X = A \cup B$, $A \cap B = \emptyset$, A and B are nonempty, and A and B are closed. We just need to show that $\mathbf{K}A \cap B = A \cap \mathbf{K}B = \emptyset$. Since A is closed,

$$\mathbf{K}A \cap B = A \cap B = \emptyset.$$

Likewise $A \cap \mathbf{K}B = \emptyset$.

□

Corollary 4.6 *A topological space is connected if and only if it is not the union of two nonempty, disjoint closed sets.*

(Two sets A and B are **disjoint** if $A \cap B = \emptyset$.)

So far we have not yet exhibited any examples of connected sets. It turns out that closed intervals, open intervals, and $\mathbb{R}$ are all connected. The proofs of these facts require an understanding of some important subleties of the real numbers, so we will defer them to Chapter 6 where we will discuss the real numbers in more detail.

Exercises

1. Let $(X, \mathbf{K})$ be a topological space, and let W be a subset of X containing only one point. Prove that W is connected.

2. If $a < b < c$, show that the set $(a, b) \cup (b, c)$ is disconnected in $\mathbb{R}$.

3. Suppose (A, B) is a separation of Y, and suppose $W \subset Y$. Show that if $A \cap W$ and $B \cap W$ are nonempty, then $(A \cap W, B \cap W)$ is a separation of W.

4. Let $(X, \mathbf{K})$ be a topological space, and let $W \subset X$. Show that if W is connected, then $\mathbf{K}W$ is connected. *Hint:* Use proof by contradiction, and apply Exercise 3.

5. Show that if $(X, \mathbf{K})$ is a topological space, if W is a connected subset of X, and if $W \subset W' \subset \mathbf{K}W$, then W' is connected.

4.3 Unions of connected sets*

We will now prove that the union of a collection of connected sets, all of which intersect, is connected. This is true even if the collection of connected sets is infinite.

Figure 4.1: A and B are connected, but $A \cap B$ is disconnected.

Theorem 4.7 (Connected Union Theorem) *Let $(X, \mathbf{K})$ be a topological space, and suppose $\{Y_i \mid i \in I\}$ is a nonempty family of connected subsets of X such that $\bigcap_{i \in I} Y_i \neq \emptyset$. Then $\bigcup_{i \in I} Y_i$ is connected.*

Proof. We will use proof by contradiction. (See §5.1 for more about this style of proof.) Suppose (A, B) is a separation of $\bigcup_{i \in I} Y_i$. For each $i \in I$, either $A \cap Y_i$ or $B \cap Y_i$ is empty; otherwise $(A \cap Y_i, B \cap Y_i)$ would be a separation of Y_i, by Exercise 4.2.3. Choose a point $x \in \bigcap_{i \in I} Y_i$. Since $x \in \bigcup_{i \in I} Y_i = A \cup B$, without loss of generality $x \in A$. Then $A \cap Y_i \neq \emptyset$ for each $i \in I$, so $B \cap Y_i = \emptyset$ for each $i \in I$. This implies that $B \cap \bigcup_{i \in I} Y_i = \emptyset$, so $B = \emptyset$, which is a contradiction. □

One may ask other questions about combinining connected sets; for example, if A and B are connected, does it necessarily follow that $A \cap B$ connected? The answer is no. A counterexample in $\mathbb{R}^2$ is shown in Figure 4.1. (We will not have the tools to prove that A and B in the figure are connected until Chapter 6, but it seems plausible.)

Exercises

1. For each of the following, either give a proof or suggest a plausible counterexample.

   (a) If A is connnected and $B \subset A$, then B is connected.

   (b) If A is connected and B is connected, then $A - B$ is connected.

   (c) If A and B are disconnected then $A \cup B$ is disconnected.

2. Use induction (Appendix C) to show that if n is a positive integer and if $A_1, \dots, A_n$ are connected subsets of a topological space $(X, \mathbf{K})$ such that $A_i \cap A_{i+1} \neq \emptyset$ whenever $1 \leq i < n$, then $\bigcup_{i=1}^{n} A_n$ is connected.

4.4 Connected components

Intuitively, a set is disconnected when it has more than one "piece". In this section we will give a precise definition of "piece". We will do this using the concept of equivalence relation.

A **relation** $\sim$ on a set X is a statement which is either true or false for every ordered pair of elements of X. For example, '$<$' and '$>$' are relations on the set of real numbers, and $\iff$ is a relation that deals with statements. We could also say that '$\subset$' and '$=$' are relations between pairs of sets, although in modern set theory there is no set of all sets.

If we want to be extremely precise about what a relation is, we can use the notion of the **Cartesian product** of two sets. The Cartesian product of X and Y, denoted by $X \times Y$, is the set of all ordered pairs (x, y) where $x \in X$ and $y \in Y$. In short,

$$X \times Y \stackrel{\text{def}}{=} \{(x, y) \mid x \in X, y \in Y\}.$$

For example, $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. We can define a relation R on X to be a subset $R \subset X \times X$. An ordered pair of points in X is in this subset if and only if the relation holds between them. For example, the relation '$<$' on $\mathbb{R}$ corresponds to the subset $\{(x, y) \mid x < y\} \subset \mathbb{R} \times \mathbb{R}\}$.

For our purposes, the Cartesian product definition of a relation is overkill. However, this concept makes regular appeances in mathematical literature in the fields of set theory, analysis, and topology, so it is worth bearing in mind.

Definition 4.8 *A relation $\sim$ is an* **equivalence relation** *if it has the following three properties:*

Axiom E1 (reflexivity). For every x, $x \sim x$.

Axiom E2 (symmetry). For every x and y, $x \sim y \Longrightarrow y \sim x$.

Axiom E3 (transitivity). For every x, y, and z, if $x \sim y$ and $y \sim z$, then $x \sim z$.

Example 4.9 Equality (of elements of a given set) is an equivalence relation.

Example 4.10 Define a relation $\sim$ on $\mathbf{Z}$ as follows:

$$x \sim y \iff x - y \text{ is an integer multiple of } 3.$$

Then $\sim$ is an equivalence relation.

Proof. $\sim$ is reflexive because for any integer x, $x - x = 0$, and 0 is a multiple of 3, namely $3 \cdot 0$. To show that $\sim$ is symmetric, suppose $x \sim y$. This means that $x - y = 3k$ for some integer k, but then $y - x = 3(-k)$, so $y \sim x$. To show that $\sim$ is transitive, suppose $x \sim y$ and $y \sim z$. Then $x - y = 3k$ for some integer k, and $y - z = 3k'$ for some integer k', so

$$x - z = (x - y) + (y - z) = 3k + 3k' = 3(k + k').$$

Thus $x \sim z$.

□

An equivalence relation $\sim$ partitions a set X into **equivalence classes**, i.e., subsets which are equivalent to each other under the terms of the equivalence relation. Thus, for $x \in X$, define

$$C(x) = \{y \in X \mid x \sim y\}.$$

This is the equivalence class containing x. The reflexive property implies

$$x \in C(x).$$

The symmetric and transitive properties imply

$$x \sim y \Longrightarrow C(x) = C(y).$$

$$x \not\sim y \Longrightarrow C(x) \cap C(y) = \emptyset.$$

(The notation '$x \not\sim y$' means it is not true that $x \sim y$.)

For example, under the equivalence relation $=$, we have $C(x) = \{x\}$. So each equivalence class contains exactly one element. In Example 4.10, there are three equivalence classes, and they are:

$$\{\ldots, -6, -3, 0, 3, 6, 9, \ldots\},$$

$$\{\ldots, -5, -2, 1, 4, 7, 10, \ldots\},$$

$$\{\ldots, -4, -1, 2, 5, 8, 11, \ldots\}.$$

Now let $(X, \mathbf{K})$ be a topological space. Define a relation $\sim$ on X as follows:

$$x \sim y \iff (\exists A \subset X)\ A \text{ is connected in } X \text{ and } x, y \in A.$$

Then $\sim$ is an equivalence relation (Exercise 1). The equivalence classes are called **connected components**. For example, as we will be able to prove later, the connected components of $(0, 2) - \{1\}$ in $\mathbb{R}$ are $(0, 1)$ and $(1, 2)$.

We leave it as an exercise to show that $C(x)$ is the largest connected set containing x. What this means is that $C(x)$ is a connected set containing x, but if D is connected and $x \in D$, then $D \subset C(x)$. Also, $C(x)$ is the only set with this property. (We should point out that largest sets with a given property do not always exist. For example, there is no largest finite set of integers. On the other hand, a largest set with a given property, if it exists, must be unique. Can you see why?)

Why have we bothered to abstract this seemingly simple notion of connected "pieces"? It is because it is not always easy to identify and enumerate all the connected components of a set. As an example, consider the Jordan Curve Theorem. This theorem states that if C is a simple closed curve in $\mathbb{R}^2$ (we will not give a precise definition of this now), then $\mathbb{R}^2 - C$ has exactly two connected components. Intuitively, these are the "inside" and "outside" of C. This statement is surprisingly hard to prove in general, although it is not so hard to prove for a particular curve C, such as a circle. You can learn about this in a more advanced topology course.

Exercises

1. Show that the relation $\sim$ defined above is an equivalence relation. *Hint:* to prove reflexivity, use Exercise 4.2.1. To prove transivity, use the Connected Union Theorem.

2. Show that $C(x)$ is the largest connected set containing x.

3. Take a guess as to what the connected components of $\mathbb{R} - \mathbb{Z}$ are.

4. Show that $C(x)$ is closed. *Hint:* Use Exercise 2 and an earlier exercise.

5. If $n \geq 0$, let S^n denote the **unit sphere**

$$S^n \stackrel{\text{def}}{=} \{x \in \mathbb{R}^{n+1} \mid d(x, 0) = 1\}.$$

   How many connected components do you think $\mathbb{R}^{n+1} - S^n$ has? (This depends on the value of n.)

Chapter 5

The Rationals are Not Connected

We mentioned in the previous chapter that the set of real numbers is connected, and that the proof of this is rather subtle. One indication of this subtlety is that the set of *rational* numbers is disconnected. Intuitively, this is because the rational numbers have "lots of holes" between them. We will elucidate this statement below. In the next chapter we will see how the real numbers "fill in the holes" in the set of rational numbers.

5.1 $\sqrt{2}$ is irrational

Recall that a **rational number** is a quotient a/b of integers a and b, where $b \neq 0$. An **irrational number** is a real number that is not rational. (As we mentioned in Chapter 3, we are taking the real numbers for granted for now, and we will discuss them more in the next chapter.)

Our first goal is to show that not all real numbers are rational. You already know this if you have done Exercise 2.4.**??**(b) and (d), but we will now give a specific example of an irrational number, namely $\sqrt{2}$.

To prove that $\sqrt{2}$ is irrational, we have to show that there do not exist integers a and b, with $b \neq 0$, such that $\sqrt{2} = a/b$. When we want to show that something does not exist, it is natural to try a **proof by contradiction**. We assume the exact opposite of what we are trying to prove, namely that such integers a and b do exist. We then try to deduce an absurd statement. If we succeed, then our assumption that a and b exist must be wrong, and we are done!

Theorem 5.1 $\sqrt{2}$ *is irrational.*

Proof. Suppose $\sqrt{2}$ is rational, i.e. $\sqrt{2} = a/b$ for some integers a and b. Then there exist positive integers a and b with this property, since if a and b are both negative, we can simply change their signs, and a and b cannot have opposite signs since $\sqrt{2}$ is positive. Let us choose positive integers a and b with $\sqrt{2} = a/b$, such that a is as small as possible. (We can do this by the Well-Ordering Principle, which says that every nonempty set of positive integers has a smallest element; see Appendix C.)

Squaring both sides of the equation $\sqrt{2} = a/b$ and multiplying both sides by b^2, we obtain $a^2 = 2b^2$. Since a^2 is even, it follows that a is even. Thus $a = 2k$ for some integer k, so $a^2 = 4k^2$, and hence $b^2 = 2k^2$. Since b^2 is even, it follows that b is even. Since a and b are both even, $a/2$ and $b/2$ are positive integers, and $\sqrt{2} = (a/2)/(b/2)$, because $(a/2)/(b/2) = a/b$. But we said before that a is as small as possible, so this is a contradiction. Therefore $\sqrt{2}$ cannot be rational. □

This particular type of proof by contradiction is known as **infinite descent**. If there exist positive integers a and b such that $a/b = \sqrt{2}$, then the above proof shows that we can find smaller positive integers a and b with the same property, and repeating this process, we will get an infinite descending sequence of positive integers, which is impossible.

Another way to produce irrational numbers is to use the theorem that a real number is rational if and only if its representation in decimal notation is eventually repeating (Exercise 4). An example of an eventually repeating decimal is

$$27.34\overline{581} = 27.34581581581581581581\cdots.$$

This is equal to $\frac{2734}{100} + \frac{581}{99900}$, which is clearly rational. The number

$$0.1010010001000010000010000001\cdots$$

is not eventually repeating (can you prove this?), and hence is not rational.

Exercises

1. (a) Prove that the sum of two rational numbers is rational.
   (b) Prove that the sum of a rational number and an irrational number is irrational. *Hint:* use proof by contradiction.
   (c) Is the sum of two irrational numbers necessarily irrational? Give a proof or a counterexample.
2. Prove that $\sqrt{6}$ is irrational.
3. Show that $0.1010010001000010000010000001\cdots$ is not eventually repeating. *Hint:* use proof by contradiction.
4. (may be tricky)

(a) Show that an eventually repeating decimal represents a rational number. *Hint:* How did we figure out that $27.34\overline{581} = \frac{2734}{100} + \frac{581}{99900}$? Generalize.

(b) Show that the decimal representation of a rational number is eventually repeating. *Hint:* Think carefully about how the long division process works. At each stage you take in certain data and then continue according to certain rules. Show that when you divide two particular integers, there are only finitely many possible values for this data. The process must eventually get stuck in a loop.

5.2 A separation of the rationals

Theorem 5.2 *The set $\mathbb{Q}$ of rational numbers is disconnected in $\mathbb{R}$.*

Proof. Let

$$\begin{aligned} A &= \{x \in \mathbb{Q} \mid x < \sqrt{2}\}, \\ B &= \{x \in \mathbb{Q} \mid x > \sqrt{2}\}. \end{aligned}$$

I claim that (A, B) is a separation of $\mathbb{Q}$. Since $\sqrt{2}$ is irrational, $\mathbb{Q} = A \cup B$. Also, it is clear that A and B are nonempty. We need to show that

$$\mathbf{K}A \cap B = A \cap \mathbf{K}B = \emptyset.$$

We introduce the following notation:

$$\begin{aligned} (-\infty, \sqrt{2}) &\stackrel{\text{def}}{=} \{x \in \mathbb{R} \mid x < \sqrt{2}\}, \\ (-\infty, \sqrt{2}] &\stackrel{\text{def}}{=} \{x \in \mathbb{R} \mid x \leq \sqrt{2}\}. \end{aligned}$$

Clearly $A \subset (-\infty, \sqrt{2})$. Furthermore, by an argument very similar to the proof of Example 3.7,

$$\mathbf{K}(-\infty, \sqrt{2}) = (-\infty, \sqrt{2}].$$

(Make sure you understand this proof.) So by Theorem 1.9,

$$\mathbf{K}A \subset (-\infty, \sqrt{2}].$$

By definition of B,

$$(-\infty, \sqrt{2}] \cap B = \emptyset.$$

It follows from the above two lines that

$$\mathbf{K}A \cap B = \emptyset.$$

The proof that $A \cap \mathbf{K}B = \emptyset$ is analogous. □

In fact, the rational numbers are totally disconnected. One can show that between any two rational numbers, there is an irrational number.

5.3 Approximating $\sqrt{2}$ with rational numbers*

Although $\sqrt{2}$ is irrational, it is possible to find rational numbers that are very close to $\sqrt{2}$. We would now like to mention several ways of doing this. This is just for "fun", although Newton's method will reappear later in the course.

Continued fractions. Write

$$\sqrt{2} = 1 + (\sqrt{2} - 1) = 1 + \frac{1}{\frac{1}{\sqrt{2}-1}} = 1 + \frac{1}{\frac{(\sqrt{2}+1)}{(\sqrt{2}+1)(\sqrt{2}-1)}} = 1 + \frac{1}{1+\sqrt{2}}.$$

We can substitute this expression for $\sqrt{2}$ into itself to get

$$\sqrt{2} = 1 + \frac{1}{1+\sqrt{2}} = 1 + \frac{1}{1 + \left(1 + \frac{1}{1+\sqrt{2}}\right)} = 1 + \frac{1}{2 + \frac{1}{1+\left(1+\frac{1}{1+\sqrt{2}}\right)}}.$$

Continuing this process, we get

$$\sqrt{2} = 1 + \frac{1}{2 + \frac{1}{2+\frac{1}{2+\frac{1}{2+\dots}}}}$$

This funny expression is called a **continued fraction**.

It is not at all clear what an infinite expression like this means. But if we are lucky, if we evaluate a finite piece of it we should get a number close to $\sqrt{2}$. Let us try:

$$\sqrt{2} \approx 1 + \frac{1}{2 + \frac{1}{2+\frac{1}{2+\frac{1}{2}}}} = 1 + \frac{1}{2 + \frac{1}{2+\frac{2}{5}}} = 1 + \frac{1}{2 + \frac{5}{12}} = 1 + \frac{12}{29} = \frac{41}{29}.$$

Squaring $41/29$, we get $1681/841$, which is quite close to 2 since 2 times 841 is 1682. If we want to improve our level of accuracy, we write

$$\sqrt{2} = \frac{1}{1+\sqrt{2}} \approx \frac{1}{1+\frac{41}{29}} = 1 + \frac{29}{70} = \frac{99}{70}.$$

Squaring this, we get $9801/4900$, which is even closer to 2. In general, if x is an approximation to $\sqrt{2}$, we can improve it by replacing x with $1 + \frac{1}{1+x}$. Each such replacement is called an **iteration**. Iterating some more, we get $239/169$, $577/408$, $1393/985$, and $3363/2378$. The last of these numbers is within one fifteen millionth of $\sqrt{2}$.

Newton's method. Suppose x is close to $\sqrt{2}$. Let $\varepsilon = \sqrt{2} - x$. We then have

$$x + \varepsilon = \sqrt{2}.$$

Squaring both sides, we get

$$x^2 + 2x\varepsilon + \varepsilon^2 = 2.$$

Since ε is small, ε^2 is very small, so we can discard it and write

$$x^2 + 2x\varepsilon \approx 2,$$

whence

$$\varepsilon \approx \frac{2 - x^2}{2x}.$$

Thus

$$\sqrt{2} = x + \varepsilon \approx x + \frac{2 - x^2}{2x} = \frac{2 + x^2}{2x}.$$

So if x is close to $\sqrt{2}$, we can replace it with $\frac{2+x^2}{2x}$, which is hopefully closer.

Let us try this method and see how well it works. A reasonable starting value is $x = 1$. We then replace x with

$$\frac{2 + x^2}{2x} = \frac{2 + (1)^2}{2(1)} = \frac{3}{2}.$$

We then replace this with

$$\frac{2 + (3/2)^2}{2(3/2)} = \frac{17}{12}.$$

Iterating again, we get $577/408$. Notice that this is one of the numbers we got from the continued fraction, but this time we obtained it in fewer steps. The next iteration gives $\sqrt{2} \approx 665857/470832$, which is accurate to about twelve decimal digits. In fact, every iteration of this procedure roughly doubles the number of digits of accuracy.

Divide and average. Notice that if $x = \sqrt{2}$, then $2/x = x$. If x is close to $\sqrt{2}$ but a little smaller, than $2/x$ is a little larger. Likewise if x is a little bigger than $\sqrt{2}$, then $2/x$ is a little smaller than $\sqrt{2}$. This suggests that if x is close to $\sqrt{2}$, we might try replacing x with the average of x and $2/x$, i.e.

$$x \longmapsto \frac{x + \frac{2}{x}}{2} = \frac{x^2 + 2}{2x}.$$

We see that this turns out to be exactly the same as Newton's method.

Arithmetic, geometric, and harmonic means. Observe that $\sqrt{2}$ is the geometric mean of 1 and 2. (See Exercise 3.5.1.) We know that the geometric mean of 1 and 2 lies somewhere between the harmonic and arithmetic means, which are 4/3 and 3/2. So to get an approximation to $\sqrt{2}$, we might try the mean of 4/3 and 3/2. But which mean — arithmetic, geometric, or harmonic? It turns out that $\sqrt{2}$ is exactly equal to the geometric mean of 4/3 and 3/2. More generally, the geometric mean of two numbers is equal to the geometric mean of the harmonic and arithmetic means of the two numbers. In other words,

$$\sqrt{ab} = \sqrt{\frac{2}{\frac{1}{a}+\frac{1}{b}}\frac{a+b}{2}};$$

and you can easily verify this.

So to approximate $\sqrt{2}$, we let $a_0 = 1$ and $b_0 = 2$. If a_n and b_n have been found, we let a_{n+1} be the harmonic mean of a_n and b_n, and we let b_{n+1} be the arithmetic mean of a_n and b_n. The closed intervals $[a_0, b_0], [a_1, b_1], [a_2, b_2], \ldots$ then "close in" on $\sqrt{2}$. In fact, one can show that the numbers $b_1, b_2, \ldots$ are the same numbers we obtained before by Newton's method.

Exercises

1. Use a continued fraction to approximate $\sqrt{3}$.

2. Let $a_0 = b_0 = 1$, and let
$$\frac{a_{n+1}}{b_{n+1}} = 1 + \frac{1}{1+\frac{a_n}{b_n}},$$
so that a_n/b_n is the n^{th} approximation to $\sqrt{2}$ obtained by the method of continued fractions. Compute the first few values of a_n and b_n, and compare a_n^2 with b_n^2. Do you notice a pattern? Can you make a conjecture? Can you prove it? (You might want to find formulas for a_{n+1} and b_{n+1} in terms of a_n and b_n and then try to prove your conjecture by induction.)

Chapter 6

Completeness of the Real Numbers

As we mentioned when we first introduced them, the real numbers form a "complete, archimedean, ordered field." What has this meant to us so far? Up to this point, we have looked at the real numbers from a geometric perspective, as points on a line that are given to us with the positions zero and one marked off. Using a straightedge and a compass, we can add, subtract, multiply, and divide any two points (as long as we do not divide by zero, of course). In other words, we are using the operations which are permitted by the field axioms. Also, given any two numbers a and b, we have supposed that we can always tell when $a > b$, i.e., that we can distinguish the ordering of the set.

However, we have not actually used all the properties of the real numbers: our operations so far would make as much sense if we simply used the rationals. However, we showed in the last chapter that the rationals will not suffice to describe our sheet of paper, since numbering a line with them would leave holes, places like $\sqrt{2}$ that cannot be named by rational numbers and that leave the rationals disconnected. The further assumption we need about the line that distinguishes the real numbers from the rationals will be called "completeness." In this chapter we will develop several equivalent ways of saying what it means for an ordered field to be complete. We will then present a way of defining the real numbers so that the completeness condition is satisfied. Finally, we will show that the real line understood in this way really is connected. It will turn out that completeness as we define it here for ordered fields is a special case of a definition that classifies any metric space as either compete or incomplete.

6.1 Completeness

In the last chapter, we discovered the existence of "holes" in the system of rational numbers, and we also developed ways of finding rational numbers that come closer and closer to filling in these "holes." However, if the number we seek is irrational, these approximation procedures would continue indefinitely without ever specifying the desired number precisely.

More realistically, suppose we stop after a certain decimal place is reached. What does this truncated decimal tell us? For example, if we stop approximating $\sqrt{2}$ at the tenths place, we are left with the number 1.4. This means that *somewhere* in the closed interval $[1.4, 1.5]$ lies the number $\sqrt{2}$. If we go to the next decimal place, we discover that $\sqrt{2}$ now lies somewhere in the (smaller) interval $[1.41, 1.42]$, as you can easily verify by squaring these numbers. In general, we can keep making our intervals smaller and smaller thus:

$$[1,2]\,,\ [1.4,1.5]\,,\ [1.41,1.42]\,,\ [1.414,1.415],\ \ldots\,.$$

In order to be sure numbers like $\sqrt{2}$ exist, we must assume that there will be *something* inside all these intervals, for otherwise, our decimal expansion would lead nowhere. This motivates the following definition.

Definition 6.1 *Let $\mathbb{F}$ be an ordered Archimdean field containing $\mathbb{Z}^+$. We say that $\mathbb{F}$ satisfies* **the Principle of Nested Closed Intervals** *if, whenever $a_1, a_2, \ldots$ and $b_1, b_2, \ldots$ belong to $\mathbb{F}$ with*

$$a_1 \le a_2 \le a_3 \le \cdots \le b_3 \le b_2 \le b_1,$$

then $\bigcap_{n=1}^{\infty}[a_n, b_n]$ is a nonembpty subset of $\mathbb{F}$ where

$$[a_n, b_n] \stackrel{\text{def}}{=} \{x \in \mathbb{F} \mid a_n \le x \le b_n\}.$$

The rationals $\mathbb{Q}$ are an ordered Archimdean field containing $\mathbb{Z}^+$ which does not satisfy the Principle of Nested Closed Intervals. This follows from our construction of nested closed intervals bracketing $\sqrt{2}$ whose intersection contains no member of the rationals. We would like to define the real numbers as the smallest field containing $\mathbb{Q}$ that satisfies the Principle of Nested Closed Intervals. That is what the reals will turn out to be, but this approach makes it hard to see that such a set exists. We could, of course, try to form such a set by declaring that it consists of all the intersections of sequences of nested closed intervals in $\mathbb{Q}$. In other words, just throw in the missing things we need by considering the procedure for constructing one as the object itself, a common strategy in mathematics. Among the complications in this instance would be, first, that the intersection of a sequence of nested closed intervals may contain many different real numbers and, second, that many different sequences of nested closed intervals could represent the same real number.

The first complication, that $\bigcap_{n=1}^{\infty}[a_n, b_n]$ could contain more than one number, can be avoided if we restrict our attention to cases where the intervals

$[a_n, b_n]$ get small fast enough. This happens, for example, when each closed interval is half as large as the one preceding it — that is, when

$$b_{n+1} - a_{n+1} = \frac{1}{2}(b_n - a_n)$$

for each n (see Exercise 1).

More generally, if there is a fixed integer $q \geq 2$ such that

$$b_{n+1} - a_{n+1} = \frac{1}{q}(b_n - a_n)$$

for all n, then $\cap_{n=1}^{\infty}[a_n, b_n]$ consists of at most one point. A popluar choice for q is ten, as in the following defintion.

Definition 6.2 *A sequence of nested closed intervals $I_0 \supset I_1 \supset I_2 \supset \dots$ is called* **decimating** *if $I_0 = [d_0, d_0 + 1]$ for some $d_0 \in \mathbb{Z}$, and, for each $n \in \mathbb{Z}^+$, if we divide I_{n-1} into ten subintervals of equal length, then I_n is one of these 10. In other words, for all $k \in \mathbb{Z}^+$, there is an integer d_k with $0 \leq d_k \leq 9$ such that*

$$I_n = \left[\ \sum_{k=0}^{n} \frac{d_k}{10^k},\ \sum_{k=0}^{n} \frac{d_k}{10^k} + \frac{1}{10^n}\ \right]$$

holds for every nonnegative integer n.

The process of approximating a point using a decimating sequence of nested closed intervals is exactly what happens when you use a metric ruler. Starting with a fundamental length of one meter, you can specify a point to the nearest decimeter, centimeter, millimeter, ..., Angstroms, Femtometers, and so on. Of course, the integers d_k described above line up to form the decimal expansion of the number we seek. All this suggests that, perhaps we do not have to worry about the intersections of every sequence of nested closed intervals, but only the intersections of decimating sequences.

Definition 6.3 *Let $\mathbb{F}$ be an ordered Archimdean field containing $\mathbb{Z}^+$. To say that $\mathbb{F}$ satisfies the* **Decimal Expansion Principle** *means that x is a nonnegative member of $\mathbb{F}$ if and only if there exists a sequence of nonnegative integers $d_0, d_1, d_2, \dots$ with $0 \leq d_k \leq 9$ for k in $\mathbb{Z}^+$ such that*

$$\{x\} = \bigcap_{n=0}^{\infty}\ \left[\ \sum_{k=0}^{n} \frac{d_k}{10^k},\ \sum_{k=0}^{n} \frac{d_k}{10^k} + \frac{1}{10^n}\ \right]$$

.

In a field with the Decimal Expansion Property, it is tempting to write that a nonegative x equals the infinite series $\sum_{k=0}^{\infty} \frac{d_k}{10^k}$ where the d_k are determined as above. Strictly speaking, we only know how to add up a finite number of summands. Suppose, more generally, we try to assign meaning to a series $\sum_{k=0}^{\infty} a_k$ with each $a_k \geq 0$. Notice that, if we truncate infinite series after n terms, the finite sums we obtain become bigger and bigger with n. Form the set

$$A = \Big\{ \sum_{k=0}^{n} a_k \mid n \in \mathbb{Z}^+ \Big\}.$$

Then number we are after should be at least as big as every element of A, but no bigger than it has to be in order to be at least as big as every element of A. If such a number exists, we call it the least upper bound of A as in the following definition.

Definition 6.4 *Let $\mathbb{F}$ be an ordered field. We say that $b \in \mathbb{F}$ is an upper bound for a set $A \subset \mathbb{F}$ if $a \leq b$ for all $a \in A$. When such a b exists, we say A is bounded above.*

We call $c \in \mathbb{F}$ a **least upper bound** *of A and write $c = \sup(A)$ to stand for "the supremum of A" if both:*

(i) c is an upper bound for A; and

(ii) if d is also an upper bound for A, then $c \leq d$.

To say that $\mathbb{F}$ satisfies the **Least Upper Bound Principle** *means that whenever $A \subset \mathbb{F}$ is nonempty and bounded above, then there exists a $c \in \mathbb{F}$ that is a least upper bound for A.*

Again, the rationals $\mathbb{Q}$ do not satisfy this principle either as you can see by considering the set $A = \{x \in \mathbb{Q} \mid x^2 < 2\}$. In fact, the next theorem asserts that a given $\mathbb{F}$ satisfies either all or none of the three principles defined in this section.

Theorem 6.5 *Let $\mathbb{F}$ be an ordered Archimdean field containing $\mathbb{Z}^+$. Then the following are equivalent:*

1. *The Least Upper Bound Principle (LUBP) holds for $\mathbb{F}$.*

2. *The Principle of Nested Closed Closed Intervals (PNCI) holds for $\mathbb{F}$.*

3. *The Decimal Expansion Principle (DEP) holds for $\mathbb{F}$.*

If any (and hence all) of these principles are satisfied by $\mathbb{F}$, we say that $\mathbb{F}$ is **complete** *and write*

$$x = \sum_{k=0}^{\infty} \frac{d_k}{10^k} = d_0.d_1d_2d_3\ldots$$

for nonnegative $x \in \mathbb{F}$ to mean that

$$x \; = \; \sup \quad \left\{ \sum_{k=0}^{n} \frac{d_k}{10^k} \mid n \in \mathbb{Z}^+ \right\}$$

where $d_0, d_1, d_2, \ldots$ *is a sequence of nonnegative integers with* $0 \le d_k \le 9$ *for* k *in* $\mathbb{Z}^+$ *such that*

$$\{x\} = \bigcap_{n=0}^{\infty} \left[\sum_{k=0}^{n} \frac{d_k}{10^k}, \sum_{k=0}^{n} \frac{d_k}{10^k} + \frac{1}{10^n} \right].$$

Proof. Notice that it is not necesary to provide all the arguments for establishing directly that every pair of assertions is equivalent. Rather, it is enough to show that the first statement implies the second, the second implies the third, and the third implies the first.

Step 1: LUBP $\Longrightarrow$ PCNI

Let $I_1, I_2, I_3, \ldots$ denote a sequence of nested closed intervals with $I_n = [a_n, b_n]$. The set $\{a_n | n \in \mathbb{Z}^+\}$ is therefore nonempty and bounded above by b_1, so it has a least upper bound $a \in \mathbb{F}$. Similarly, the set $\{-b_n | n \in \mathbb{Z}^+\}$ is nonempty and bounded above by $-a_1$, so it has a least upper bound which we write as $-b$ for some $b \in \mathbb{F}$. (In other words, b is the greatest lower bound for $\{b_n | n \in \mathbb{Z}\}$). Then $a \le b$, so $I = [a, b]$ is nonempty with $I \subset \bigcap_{n=1}^{\infty} I_n$ as desired.

Step 2: PCNI $\Longrightarrow$ DEP

Assuming PCNI holds, we must show x is a nonnegative element of $\mathbb{F}$ if and if there exists a sequence of nonnegative integers $d_0, d_1, d_2, \ldots$ with $0 \le d_k \le 9$ for k in $\mathbb{Z}^+$ such that

$$\{x\} = \bigcap_{n=0}^{\infty} \left[\sum_{k=0}^{n} \frac{d_k}{10^k}, \sum_{k=0}^{n} \frac{d_k}{10^k} + \frac{1}{10^n} \right]$$

. If x is the only element of such an intersection, then PCNI asserts that $x \in \mathbb{F}$ since each of the endpoints of the intervals being intersected must belong to any field containing $\mathbb{Z}^+$. Conversely, suppose that $x \in \mathbb{F}$. It follows from the the Archimedean Property of $\mathbb{F}$ that there exists a nonnegative integer d_0 such that x belongs to $I_0 = [d_0, d_0 + 1]$. We can then define inductively a decimating sequence $I_0 \subset I_1 \subset I_2 \subset \ldots$ of the desired form by picking d_n given d_{n-1} from among the integers between zero and nine so that

$$x \in I_n = \left[\sum_{k=0}^{n} \frac{d_k}{10^k}, \sum_{k=0}^{n} \frac{d_k}{10^k} + \frac{1}{10^n} \right]$$

. By construction, we have $x \in \cap_{n=0}^{\infty} I_n$. Suppose that $y \in \mathbb{F}$ also belongs to $\cap_{n=0}^{\infty} I_n$. Then $x = y + \varepsilon$ for some nonzero ε is impossible because, for n chosen large enough so that $\frac{1}{10^n} < |\varepsilon|$, we could not have both x and y in I_n. Hence $x = y$ and so we can write $x = \cap_{n=0}^{\infty} I_n$ as desired.

Step 3: DEP $\Longrightarrow$ LUBP

Let $A \subset \mathbb{F}$ be nonempty and bounded above. Without loss of generality, suppose there is an positive element $x \in A$ since otherwise we could just translate A to make this so. By the Archimedean Property, there is an integer in $\mathbb{Z}^+$ that bounds A from above. By the well-ordering principle, there is a least such integer d_0 which therefore satisfies

$$A \cap \ [d_0 - 1, d_0] \neq \emptyset \ \text{ but } \ \mathrm{A} \ \cap \ [\mathrm{d}_0, \infty) = \emptyset.$$

Continuing inductively, choose d_n given d_{n-1} from among the integers between zero and nine so that

$$A \cap \left[\sum_{k=0}^{n} \frac{d_k}{10^k} - \frac{1}{10^n}, \sum_{k=0}^{n} \frac{d_k}{10^k}\right] \neq \emptyset \ \text{ but } \ \mathrm{A} \cap \left[\sum_{\mathrm{k}=0}^{\mathrm{n}} \frac{\mathrm{d}_\mathrm{k}}{10^\mathrm{k}}, \infty\right) = \emptyset.$$

By DEP, there is an $x \in \mathbb{F}$ such that

$$x = \bigcap_{n=0}^{\infty} \ \left[\sum_{k=0}^{n} \frac{d_k}{10^k}, \sum_{k=0}^{n} \frac{d_k}{10^k} + \frac{1}{10^n}\right].$$

Now this x is certainly an upper bound for A since, by construction,

$$a \leq \sum_{k=0}^{n} \frac{d_k}{10^k}$$

for all $a \in A$ and every nonnegative integer n. Moreover, we claim that x is the least such upper bound. For suppose that y is another with $x - y = \varepsilon > 0$. By the Archimedean property, we can find an $n \in \mathbb{Z}^+$ such that $\frac{1}{10^n} < \frac{\varepsilon}{2}$. But then there must exist an a in A such that

$$y = x - \varepsilon < x - \frac{2}{10^n} < \sum_{k=0}^{n} \frac{d_k}{10^k} - \frac{1}{10^n} < a$$

which contradicts the assertion that y is an upper bound for A. We conclude that $x = \sup(A)$ as claimed.

□

Exercises

1. Prove that if $a_1, a_2, \ldots$ and $b_1, b_2, \ldots$ are members of an ordered Archimedean field containing $\mathbb{Z}^+$ such that $a_1 \leq a_2 \leq a_3 \leq \cdots \leq b_3 \leq b_2 \leq b_1$, and if $b_{n+1} - a_{n+1} = \frac{1}{2}(b_n - a_n)$ for each n, then there exists only one point in $\cap_{n=1}^{\infty}[a_n, b_n]$. *Hint:* use proof by contradiction.

2. Is there a "principle of nested open intervals"? Prove or give a counterexample to the following statement. If $a_1, a_2, \ldots$ and $b_1, b_2, \ldots$ are real numbers such that
$$a_1 \le a_2 \le \cdots \le b_3 \le b_2 \le b_1,$$
and if $a_n < b_n$ for each n, then
$$\cap_{n=1}^{\infty}(a_n, b_n) \neq \emptyset.$$

3. Can two different elements of an ordered field $\mathbb{F}$ both be least upper bounds of the same set $A \subset \mathbb{F}$?

4. How would you define the greatest upper bound d of a subset A of an ordered field $\mathbb{F}$? The usual notation is $d = inf(A)$, which stands for the "infimum of A." Can you reformulate the competeness of $\mathbb{F}$ in terms of the existence of greatest lower bounds?

5. Justify the assertion in the theorem that
$$x = \sup \left\{ \sum_{k=0}^{n} \frac{d_k}{10^k} \mid n \in \mathbb{Z}^+ \right\}.$$
In particular, how do you know this least upper bound exists?

6. How could you define the sum of a series $\sum_{k=0}^{\infty} a_k$ with each $a_k \geq 0$? Must this sum exist? Illustrate how the story becomes more complicated if we drop the requirement that each terms must be nonnegative.

7. Fix an integer $q \geq 2$. Show that the Decimal Expansion Principle is equivalent to a statement about q-ary expansions of the form $\sum_{k=0}^{\infty} \frac{e_k}{q^k}$ where q plays the role of base instead of 10.

6.2 The Reals

We would like to define the reals as the smallest complete field containing the rationals. Does such a set exist? To see that this is a serious question, notice that the notion of a least upper bound has an impredicative definition since $\sup(A)$ is determined as a member of a set to which it belongs, namely the set all upper bounds of A.

One strategy for constructing the reals is to think of adding on to the rationals all the least upper bounds that are missing. You know you need to fill in a point whenever there is a set $A \subset \mathbb{Q}$ that is bounded above but has no least upper bound in $\mathbb{Q}$. So why not let such a set A actually be the missing point? In this way, a point in $\mathbb{R}$ will turn out to be a subset of $\mathbb{Q}$ in need of a least upper bound. We will work with a certain family of subsets of $\mathbb{Q}$ called Dedekind cuts chosen so that each cut determines one and only one point in $\mathbb{R}$.

Definition 6.6 *A subset A of the rational numbers $\mathbb{Q}$ is called a* **Dedekind cut** *if the following conditions hold:*

1. *$A \neq \emptyset$ and $A \neq \mathbb{Q}$.*

2. *If $r \in A$ and $s \in \mathbb{Q}$ satisfies $s < r$, then $s \in A$.*

3. *A does not contain a largest element.*

You can picture a Dedekind cut as consisting of all the rationals to the left of some geometric point on the line. Thus $A = \{x \in \mathbb{Q} | x < 2\}$ and $B = \{x \in \mathbb{Q} \mid x^2 < 2 or x < 0\}$ are examples of Dedekind cuts. Given $r \in \mathbb{Q}$, we let $r^* = \{s \in \mathbb{Q} \mid s < r\}$ denote the Dedekind cut corresponding to r. In this way, the set of all Dedekind cuts contains a copy of the rationals. We claim that it is possible, in a way compatible with the operations on these rationals, to define how to take an ordered pair of cuts and determine their sum, difference, product, and quotient as long as the second is nonzero, as well as their ordering. For example, if A and B are Dedekind cuts, set

$$A + B = \{a + b \mid a \in A \text{ and } b \in B\}$$

and check that $(r+s)^* = r^* + s^*$ for $r, s \in \mathbb{Q}$. Multiplication is defined similarly, but you must take care to distinguish different cases to keep the signs straight. As for the ordering, we simply declare that $A \leq B$ means that $A \subset B$ so that, for $s, r \in \mathbb{Q}$, we have $s^* \leq r^*$ if and only if $s \leq r$.

Thus, we use what we know about the rationals to make the set of all Dedekind cuts into an ordered field containing a copy of $\mathbb{Z}^+$. It is easy to see that this field is Archimedean. Moreover, it satisfies the Least Upper Bound Principle because if

$$\{A_i \mid i \in I\}$$

is a nonempty set of Dedekind cuts bounded above by a cut B, then the Dedekind cut $\bigcup_{i \in I} A_i$ is its least upper bound. The set of all Dedekind cuts is therefore complete. In fact, it must be the smallest complete field containing a copy of $\mathbb{Q}$ because, if we were to remove even a single cut A, that $A \subset \mathbb{Q}$ would be a bounded above but have no least upper bound. We summarize the arguments sketched here in the following result.

Theorem 6.7 *The set of all Dedekind cuts is the smallest field containing a copy of the rationals that is ordered, Archimedean, and complete. We therefore denote this set $\mathbb{R}$ and consider each Dedekind cut to be a real number.*

Here again, the objects we wish to study turn out to be defined as sets of other objects. Once we know that something called the reals exists with all the properties we expect, we can ignore in practice that each real number is itself a set of rationals, and go back to thinking of them as points on the line, decimals, or anything else that works for you. For example, another way to construct the reals that generalizes to nonordered spaces is to think of each real as an equivalence class of certain sequences of rationals called Cauchy sequences. For now, the completeness of the reals can enter proofs by citing whichever of the three equivalent principles introduced above seems most convenient.

Exercises

1. Explain why the repeating decimal $0.999\bar{9}\ldots$ must equal the number one. What does this tell you about what you would have to do in order to define the reals as decimals?

2. Fill in the details concerning addition, subtraction, multiplication, and nonzero division on the set of all Dedekind cuts.

3. State and prove a theorem that describes when the union of Dedekind cuts is a Dedekind cut.

4. Given $A, B \subset \mathbb{R}$ with $x = \sup(A)$ and $y = \sup(B)$, does it follow that $x \pm y = \sup(A \pm B)$?

5. Let $A \subset \mathbb{R}$ be a nonempty finite set. State and prove a theorem characterizing the least upper bound of A.

6. Show that if $A \subset \mathbb{R}$ has a least upper bound $x = \sup(A)$, then x belongs to $\mathbf{K}A$.

7. Show that there are lots more real numbers than rationals by proving that the interval (0, 1) in the reals is not countable even though the set of rationals between zero and one is countable. Cantor's argument begins by supposing on the countrary that you can make a list of all the decimal expansions of the reals between zero and one. Let $.d_{i,1}d_{i,2}d_{i,3}\ldots$ denote the ith row of such a list. Then produce a number not on the list by going down the diagonal and making the ith entry of the new number's decimal expansion different from $d_{i,i}$.

8. A subset A of a topological space $(X, \mathbf{K})$ is said to be **dense** in a subset B of the same space if $\mathbf{K}A = B$. Show that the rationals are dense in the reals and that the irrationals are dense in the reals. Why can't we just define $\mathbb{R}$ as the closure of $\mathbb{Q}$? A topological space containing a countable dense set is said to be **separable**. Find examples of spaces that are and are not separable.

6.3 Intervals are connected

Now that we know that the reals exist and are complete, we should be able to prove that they are connected. Notice that the metric on the reals that induces the Euclidean closure operator on $\mathbb{R}$ is still $d(x, y) = |x - y|$, which yields a nonnegative Dedekind cut equalling either plus or minus the cut obtained by taking the difference of x and y.

Definition 6.8 *An* **interval** *is a set $S \subset \mathbb{R}$ such that if $a < b < c$ and $a, c \in S$, then $b \in S$.*

For example, $\mathbb{R}$, closed intervals, and open intervals are intervals according to this definition. We will now prove that every interval is connected. The proof technique is to use the principle of nested closed intervals to "zoom in" on a point with a special property.

Theorem 6.9 *If S is an interval, then S is connected.*

Proof. Suppose S is an interval and suppose (A, B) is a separation of S. We will show that there must be a contradiction.

Since A and B are nonempty, we can choose $a_1 \in A$ and $b_1 \in B$. Without loss of generality, $a_1 < b_1$. Since S is an interval, and since a_1 and b_1 are in S, the average $(a_1 + b_1)/2$ is also in S. Since $S = A \cup B$, the point $(a_1 + b_1)/2$ is either in A or in B. If $(a_1 + b_1)/2 \in A$, let $a_2 = (a_1 + b_1)/2$ and let $b_2 = b_1$. If $(a_1 + b_1)/2 \in B$, let $a_2 = a_1$ and let $b_2 = (a_1 + b_1)/2$. Notice that $a_2 \in A$ and $b_2 \in B$, but

$$b_2 - a_2 = \frac{1}{2}(b_1 - a_1).$$

Since S is an interval and $a_2, b_2 \in S$, it follows that $(a_2 + b_2)/2) \in S$. If $(a_2 + b_2)/2 \in A$, let $a_3 = (a_2 + b_2)/2$ and let $b_3 = b_2$; otherwise let $a_3 = a_2$ and let $b_3 = (a_2 + b_2)/2$. Now $a_3 \in A$ and $b_3 \in B$, but $b_3 - a_3 = (b_2 - a_2)/2$. Continuing this process, we can find numbers $a_1 \leq a_2 \leq \cdots \leq b_2 \leq b_1$ such that $a_n \in A$, $b_n \in B$, and

$$b_{n+1} - a_{n+1} = \frac{1}{2}(b_n - a_n)$$

for each n.

By the Principle of Nested Closed Intervals, there exists $x \in \bigcap_{n=1}^{\infty}[a_n, b_n]$. We claim that $x \in \mathbf{K}A$ and $x \in \mathbf{K}B$. To show that $x \in \mathbf{K}A$, let $\varepsilon > 0$ be given. As in Exercise 1, we can choose n such that $b_n - a_n < \varepsilon$. Since $x \in [a_n, b_n]$, it follows easily that $d(x, a_n) < \varepsilon$. So a_n is a point in A within distance ε of x. Since we can do this for any ε, it follows that $x \in \mathbf{K}A$. Similarly $x \in \mathbf{K}B$.

Now $x \in S$, since $a_1, b_1 \in S$, $x \in [a_1, b_1]$, and S is an interval. So $x \in A$ or $x \in B$. But if $x \in A$ then $A \cap \mathbf{K}B \neq \emptyset$, which is a contradiction. On the other hand, if $x \in B$ then $\mathbf{K}A \cap B \neq \emptyset$, which is also a contradiction. □

Exercises

1. Prove that if S is a connected subset of $\mathbb{R}$, then S is an interval.

2. Prove that intervals are connected in $\mathbb{R}$ directly from either one of the other two characterizations of completeness.

3. By bisecting instead of decimating, describe how to construct the binary expansion of a number x in $[0, 1]$. Give examples. What real number does $0.111\bar{1}\ldots$ represent, and why?

4. A subset A of a topological space is called **totally disconnected** if its only connected components are singletons. Show that a discrete topological space is totally disconnected. Show that the Cantor Middle Thirds set is totally disconnected. Hint: What can you say about the base-three expansion of points in the Cantor set?

5. A subset A of a topological space $(X, \mathbf{K})$ is called **perfect** if, for each $x \in A$, we have $x \in \mathbf{K}(X - A)$. Give some examples of sets that are perfect and sets that are not. Prove that the Cantor Middle Thirds set is perfect. In general, we call A a Cantor set if it is closed, perfect, and totally disconnected. Can you construct others besides his Middle Thirds set?