Chapter 5
Points, Lines, and the Structure of $\mathbb{R}$

Abstract In the previous chapter we saw how a large portion of mathematics can be formalized in first-order logic. The very fact that the construction of the classical number structures can be formalized this way makes first-order logic relevant, but is it necessary? For centuries mathematics has been developing successfully without much attention paid to formal rigor, and it is still practiced this way. When intuitions don't fail us, there is no need for excessive formalism, but what happens when they do? In modern mathematics intuition can be misleading, especially when actual infinity is involved. In this chapter, we will see how seemingly innocuous assumptions about actually infinite sets lead to consequences that are not easy to accept. Then, we will go back to our discussion of a formal approach that will help to make some sense out of it.

Keywords Square root of 2 · Irrational numbers · Real numbers · Dedekind cuts · Dedekind complete orderings · Banach-Tarski paradox · Infinite decimals

5.1 Density of Rational Numbers

The rational numbers are ordered densely. Between any two rational numbers, there is another one. To illustrate this graphically, we can start marking points on a straight line, starting with two, and continuing by marking new points in-between points already marked. Soon, no matter how fine our marks are, the line begins to look solid. If we imagine that all rational numbers have been marked, it seems that there should be no space left for any other marks. How reliable is that intuition? We should be careful. When we think of a point, we see it as a dot on a plane. We see lines with different widths, some can be thinner, some thicker. Such images are helpful visualizations of ideal geometric objects, but they are not the objects themselves. A geometric point cannot be seen. It has no length nor width. It is just an idea of a perfect, exact location. Similarly, lines have no width, they only have length. We will examine this more carefully in a moment, but now it is time for a short digression on sets that are actually infinite.

R. Kossak, *Mathematical Logic*, Springer Graduate Texts in Philosophy 3,
https://doi.org/10.1007/978-3-319-97298-5_5

The sets $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ share a common feature: they can be generated in an infinite step-by-step process. What is an infinite step-by-step process? We have already seen one. This is how the natural numbers are made: in the first step, we construct the number zero; in the second step the number one, and son on. In each next step we construct a number one larger than the previous one. This way, sooner or later, each natural number gets constructed. Now imagine that all those steps have been performed. This can be thought of as a magic limit step ω. In the first step, the first operation is performed, in the second—second, and so on. Then there is the step ω, marking the fact that all possible finite steps have been performed. It is the first infinite step. Here we speak about it metaphorically, later it will become a legitimate set-theoretic notion.

I have not explained what kind of constructions are allowed in step-by-step constructions, or even what it means to "construct a number," this will be made clearer in Chap. 6. Mathematical constructions are mental, although they are often modeled on actual operations, such as counting, performing a geometric construction, or calculating according to some formula. For now let us just think of any mental process in which steps are clearly understood, and that does not involve anything actually infinite.

Here is how we can construct $\mathbb{Z}$: in step one, we construct $\{0\}$ in step two $\{-1, 0, 1\}$, in step three, $\{-2, -1, 0, 1, 2\}$, ..., in step 100 we get the set $\{-99, -98, -97, \ldots, 97, 98, 99\}$, and so on. All these sets are finite approximations to our goal. We will have "constructed" all of $\mathbb{Z}$ in the limit step ω.

The process of generating the set of all fractions $\mathcal{F}$ step-by-step is a little more complicated. There are many ways to do it. Here is one.

- In step one, we make $\{0\}$.
- In step two, $\{-\frac{1}{1}, 0, \frac{1}{1}\}$.
- In step three, we make all fractions that can be written with numerators and denominators less than three: $\{-\frac{2}{2}, -\frac{1}{1}, -\frac{1}{2}, 0, \frac{1}{2}, \frac{1}{1}, \frac{2}{1}\}$.
- In step four we make the set of all fractions whose numerators and denominators are less than four.
- and so on...

In each step, we systematically add more and more fractions, with larger and larger numerators and denominators—finitely many at each step—making sure that every fraction will appear sooner or later. This is a well-defined process that will eventually generate all fractions, including all those different fractions that are equivalent to each other. In the limit step ω we will obtain all fractions. If we want $\mathbb{Q}$ as the final result, each step can be followed by erasing redundancies such as $\frac{5}{5}$ or $-\frac{2}{4}$.

A geometric line, or a line segment, is a single geometric object, but in modern mathematics we think of it as made of points, so it becomes an actual (not just potential) infinite set. How many points make a line? Certainly infinitely many, but we will try to be more precise. We can think of straight line as an ideal measuring tape. One point is designated to represent 0, then there are infinitely many points spread at equal distances in both directions away from 0. Those represent all integers; positive going in one direction, negative in the other. Usually we imagine this line positioned horizontally, with the positive numbers to the right, and negative to the left. We can mimic the step-by-step construction of the rational numbers and consecutively mark all rational numbers, with each number marked according to the magnitude it represents. For example $\frac{1}{2}$ is marked in the middle between 0 and 1, and the place for $\frac{4}{3}$ is found by dividing the segment between 1 and 2 into three equal pieces and marking $\frac{4}{3}$ at the end of the first piece to the right of 1. This is a laborious process, but it is easy to imagine how it all can be done by geometric means in a step-by-step fashion. It is interesting that the whole construction can be done just with a ruler and a compass (and infinite time on your hands). It is not immediately obvious how to divide a line segment into, for example, 13 equal pieces just with a ruler and a compass, but it can be neatly done. You can see it in animation at http://www.mathopenref.com/constdividesegment.html.

We will say that a point of the line is *rational* if it is one of the points marked in the process described above. The rational points are geometric representations of all rational numbers. The left-to-right ordering of the rational points agrees with the larger-smaller ordering of the rational numbers they represent. It is a common mathematical practice to identify the rational numbers with their geometric representations.

Is every point on the geometric line a rational point? This question would be hard to answer if we relied only on very basic geometric intuitions. We need a deeper insight. Consider a square with each side of length 1 positioned so that its lower left corner coincides with the point marked 0 on the number line (see Fig. 5.1). The circle whose center is at 0 and which passes through the upper right vertex of the square crosses the number line at a point. It was one of the great discoveries of Pythagorean mathematics (attributed to a member of the Pythagorean school named Hippasus) that point of intersection of the circle and the line is not rational. It follows that the rational numbers do not cover the whole line. The Pythagoreans did not think about it this way; what they proved was that the diagonal of a square is *incommensurable* with its side.[1] We will talk about incommensurability in the next section.

So now other questions suggest themselves. If there are points on the line that are not rational, how many such points are there? Moreover, since the line is not made of rational points, what else is it made of? If you think of the non-rational points as gaps in the line that is made only of the rational points, can those gaps be filled to complete the line? What does it mean to fill a gap? If a solid geometric line is a *continuum* without gaps, can it be made of points? If it is made of points, how is it made? If there are many points on the geometric line that do not correspond to rational numbers, do they correspond to some other numbers? What numbers? Where do those other numbers come from?

---

[1] This geometric fact has an interesting number theoretic variant: the sum of two equal square numbers is never a square number. $1^2 + 1^2 = 1 + 1 = 2$ (not a square), $2^2 + 2^2 = 4 + 4 = 8$ (not a square), $3^2 + 3^2 = 9 + 9 = 18$ (not a square), and so on. An elegant proof of this is in Appendix A.2.

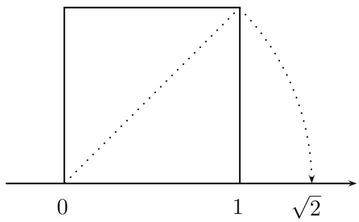

Fig. 5.1 $\sqrt{2}$ finds its place on the number line

5.2 What Are Real Numbers, Really?

Numbers which correspond to “gaps” in the line made of rational points are called *irrational*. All numbers, rational and irrational, that can be represented as distances between points on the real line, are called *real numbers*. The real line here is what in the previous section we called a geometric continuous line. This rather vague definition of the real numbers is usually given in entry level mathematics courses, and it suffices for many practical purposes. Actually, the intuition behind it suffices for much more, including some advanced mathematics. However, here we are concerned with a more careful approach. Notice that if we declare, as it is routinely done, that real numbers are distances from the origin (the point marking zero) to other points on the geometric line, then we are obliged to explain what we mean by a distance. Those distances are numbers, one usually hears, but this kind of answer will certainly not satisfy us.

The Pythagoreans discovered that the length of the diagonal of the unit square, is not commensurable with the base of the square. To explain what it means, let us consider the square in Fig. 5.1. What is the length of the diagonal of the square? By inspecting the picture, we can see that its length should be between 1 and 2, so its measure will not be a whole number. Will it be a fraction then? If the base of the square is divided into 10 equal intervals, one can check that 14, of such intervals, can be placed one by one on the diagonal, but the 15th sticks out. If the picture is large and precise enough, one can also see that the 14 intervals do not quite cover the diagonal. This means that the length of the diagonal is between 1.4 and 1.5. We could try a finer measure, say hundredths. When the base is divided into 100 intervals of equal length, one can verify that the diagonal can be almost covered with 141 such intervals, but the 142nd sticks out. This shows that the length of the diagonal is between 1.41 and 1.42. The meaning of incommensurability of the diagonal is that no mater what fraction of the base of the square we take as the unit measure, we can never find a number of intervals of this size that will exactly cover the diagonal. The last piece will either not reach the opposite corner, or will stick out. Surely though, everyone would agree that there should be a number that represents the length of the diagonal of the unit square. Let us call this number d.

It follows from the Pythagorean theorem that $1^2 + 1^2 = d^2$. Hence $d^2 = 2$, and because of that we will say that $d = \sqrt{2}$, which means that d is that number whose square equals 2. So let us be clear. At his point we know that $\sqrt{2}$ represents the length of the diagonal. Calling it a *number* is a bit premature, since we do not know what kind of number it is, and since it is not rational, it is not clear at all what it means to square it. We will deal with this problem in a moment, but for now notice, that the task of assigning a number value to the length of the diagonal of the unit square forces us to extend the number system beyond the rational numbers. We need at least one new number d, but that forces us to add more numbers. How many? At least infinitely many. Here is why. Let d_n be the length of the diagonal of an n by n square, where n is a natural number. From the incommensurability of the diagonal of a square with its side, it follows that d_n is not a rational number. Geometry forces us to include all those new numbers in a larger number system, but what are those numbers and how can they be included?

Let us go back to the idea that all points on the continuous geometric line correspond to numbers. Georg Cantor—the inventor of set theory—proved in a paper published in 1874 that there are more points on the line than there are rational numbers. What this means can be explained as follows. As we have seen, the set of all rational numbers can be constructed in a step-by-step process. The particular choice of what we do in each step is not important; what matters is that each step involves only finitely many operations. Cantor proved that the real line cannot be built from points in a step-by-step construction.[2] There are far too many points on the line which do not correspond to rational numbers. Cantor's argument is presented in Appendix A.2.

5.3 Dedekind Cuts

If we cannot construct a geometric line step-by-step, then how can we do it? The line must be rather complex, but can it be somehow constructed or built up from simpler pieces? Can it be made of points?

In modern mathematics, there are essentially two standard constructions of the real line. One proceeds via Dedekind cuts, and the other via equivalence classes of Cauchy sequences. Cauchy's construction is useful in mathematical analysis but it is more technical, so we will only describe what Dedekind did. In preparation, recall that in the previous chapter we defined rational numbers to be pairs of integers. Objects of a new kind—the rational numbers—were built is a certain way, from previously introduced objects—the integers. So, in this sense, in the hierarchy of mathematical objects, the rational numbers are more complex than the integers. Something similar will happen now. We will build the real numbers from the rational ones. The construction will be much less direct.

[2] An interesting technical aspect is that we might as well allow a step-by-step construction in which each step is itself a step-by-step construction. Cantor's theorem applies in this case as well.

Look again at Fig. 5.1. The number we called the square root of 2 is not rational, but its place on the number line can be determined by its position with respect to all rational numbers around it. Some are below it, some above. This leads to an idea to *identify* $\sqrt{2}$ with its place in the ordered set or rational numbers, and, further, to *identify* this place with a pair of sets. The set of rational numbers to the left of that place, and the set rational numbers to the right. Let us call the set to the left the Dedekind cut of $\sqrt{2}$, and let us denote it by $D_{\sqrt{2}}$.

Guided by geometry, we were lead to consider lengths of linear segments that cannot be measured using rational numbers. It seems natural to identify such lengths with places on the number line made of rational numbers, especially if such places can be determined by geometric means. Notice that even though originally $\sqrt{2}$ can be thought of a gap, as an empty space in the line made of rational numbers, this place is uniquely determined by $D_{\sqrt{2}}$; hence it is not nothing. The cut $D_{\sqrt{2}}$ is a well-defined mathematical object, and it is this object that we can also think of as $\sqrt{2}$. We can think of the set $D_{\sqrt{2}}$ as a number and we will make it more precise next.

We defined $\sqrt{2}$ using the construction illustrated on Fig. 5.1. The next goal is not only to define more such numbers, but to create a whole number system. It seems that to do that one would be forced to consider all situations in geometry, and in other areas of mathematics, where irrational numbers can occur, and to extend the number system by adding new numbers by considering more Dedekind cuts, or perhaps by some other means as well. To foresee what other irrational numbers might be needed would require exact definitions of ways in which such new numbers could be introduced. That would be very hard to do, but, to a great relief, it turns out unnecessary. Following Richard Dedekind, we can do something else. We will extend the system of rational numbers, by adding the largest possible set of Dedekind cuts. Before we do it in full generality, let us take a closer look at $D_{\sqrt{2}}$.

To define $D_{\sqrt{2}}$, one does not have to appeal to Fig. 5.1. Since $\sqrt{2}$ is not rational, for every rational number p either $p^2 < 2$ or $p^2 > 2$. We can define the Dedekind cut $D_{\sqrt{2}}$ as the set of rational number p such that $p^2 < 2$. The cut $D_{\sqrt{2}}$ defined this way does not have a largest element, and there is no smallest element above it. To see that this is the case, one must prove that for any rational p such that $p^2 < 2$, there is a rational q such that $p < q$ and $q^2 < 2$; and that any rational p such that $p^2 > 2$, there is a rational q such that $q < p$ and $2 < q^2$. Using elementary algebra, one can check that for $q = \dfrac{2p+2}{p+2}$, if $p^2 < 2$, then $p < q$, and $q^2 < 2$; and if $2 < p^2$, then $q < p$, and $2 < q^2$.

For a rational number p, let the Dedekind cut of p, denoted by D_p, be the set of all rational number that are less that p. For example, D_0 is the set of all rational negative numbers. Now we will consider a new structure whose elements are all sets D_p, and whose relations are defined in accordance with the ordering, addition, and multiplication in $\mathbb{Q}$. For example, for rational p, q, and r, we define that $D_p + D_q = D_r$ if and only if $p + q = r$. This way we obtain a copy of the $(\mathbb{Q}, \mathsf{Add}_{\mathbb{Q}}, \mathsf{Mult}_{\mathbb{Q}})$

in which each number p is replaced with the set of numbers D_p. The elements have changed, but they stand in a one-to-one correspondence with the old elements, and the correspondence $p \leftrightarrow D_p$ preserves the arithmetic relations. The new structure is an *isomorphic copy* of the old one. In this sense it is the same old $\mathbb{Q}$, with its addition and multiplication, and we will treat is as such. The point of this maneuver is to allow a smoother transition to the extension of $\mathbb{Q}$ to the larger set of real numbers.

For each rational p, there is a smallest rational number that is not in D_p. That number is p itself. The cut $D_{\sqrt{2}}$ is different. As we saw above, there is no smallest rational number p such that $2 < p^2$.

In general, a subset D of $\mathbb{Q}$ *Dedekind cut* if D is nonempty set, D is not the whole $\mathbb{Q}$, for every p, if p is in D then so are all rational numbers that are less than p, and D has no largest number.

Now we can say precisely what the real numbers are. They are exactly the Dedekind cuts. The real numbers are special sets of rational numbers. Some of them, those of the form D_p, for rational p, represent rational numbers, and all other, such that there is no smallest rational number above them, are *irrational*. We will use $\mathbb{R}$ as the name for the set of all Dedekind cuts. To make the set $\mathbb{R}$ a number structure, we need to say how the Dedekind cuts are added and multiplied. We will only define addition. This definition of multiplication is similar, but a bit more complicated. For details see [32].

If D and E are Dedekind cuts, then their sum $D+E$ is defined to be the Dedekind cut consisting of all sums $p+q$, where p is in D and q is in E. One has to check that the addition thus defined has all the required properties, such as $D + E = E + D$. It all works out, and as a result we have extended our number system to include all real numbers, and it is a huge accomplishment.

5.3.1 Dedekind Complete Orderings

If D and E are Dedekind cuts, then D is less that E if there is a p in E such that $q < p$, for all q in D. This gives us a new linearly ordered set $(\mathbb{R}, \mathsf{Less}_{\mathbb{R}})$. The ordering relation $\mathsf{Less}_{\mathbb{R}}$ is both dense and complete. Density of has already been defined. The notion of completeness can be formalized in different ways. We will present one.

In an attempt to keep the separation of syntax and semantics clearly visible, when comparing numbers, instead of the usual $x < y$, have been using $(x, y) \in \mathsf{Less}_{\mathcal{S}}$, where $\mathcal{S}$ is $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, or $\mathbb{R}$. This is done to make clear that, an order relation is a set of ordered pairs of elements of the given structure. It is important to keep this in mind, but with more structures to consider and a more detailed discussion to follow, keeping the formal notation absolutely correct all the time becomes a hindrance to reading (and writing). From now on, in accord with the usual mathematical practice, we will be taking liberties with notation. In particular, we will go back to the usual symbol $<$ for ordering in structures. When I say that $(A, <)$ and $(B, <)$ are ordered sets, I mean that in the first structure $<$ is a set of ordered pairs of elements of the set A, and the second, $<$ is the set of ordered pairs of elements of B. We will also use $x \leq y$ to abbreviate $x < y \vee x = y$. Moreover, in first-order statements we will no longer use atomic formulas $L(x, y)$, and just write $x < y$ instead. Sometimes it can lead to confusion, but it is a necessary compromise.

Dedekind cuts in an arbitrary linearly ordered set $(A, <)$ can be defined the same way as we did it for $(\mathbb{Q}, <)$, i.e. D is a Dedekind cut in $(A, <)$ if D is nonempty subsets of A, such that $D \neq A$, for all a in D, if $b < a$, then b is in D, and D has no largest element. A Dedekind cut D of $(A, <)$ is *rational* if there is an a in A, such that $D = \{x : x < a\}$. Otherwise D is *irrational.*

It is easy to see that $(\mathbb{N}, <)$ and $(\mathbb{Z}, <)$ have no Dedekind cuts.

Now comes a crucial definition.

Definition 5.1 A linearly ordered set $(A, <)$ is *Dedekind complete* if all Dedekind cuts of $(A, <)$ are rational.

As we have seen, $(\mathbb{Q}, <)$ is not Dedekind complete. It turns out that $(\mathbb{R}, <)$ is, and this has important consequences. One can think of the irrational Dedekind cuts as objects filling in all irrational gaps in linearly ordered sets—they *complete* the ordering of $\mathbb{Q}$.

From the Dedekind incomplete $(\mathbb{Q}, <)$ we constructed a larger ordered set $(\mathbb{R}, <)$. This set is larger, so potentially it could have new irrational Dedekind cuts. Then, one could move to a larger structure, in which these cuts become rational, but this new structure could also have new irrational cuts that would need to be filled in a larger structure, and so on. Such a process could have no end, but in the case of $(\mathbb{R}, <)$ it does end abruptly after just one step. It turns out that all Dedekind cuts in $(\mathbb{R}, <)$ are already rational (in $(\mathbb{R}, <)$!). This fact is not difficult to prove, but a proof requires more technicalities, and I do not want to make the text any heavier with notation than it is already. The more mathematically inclined reader may want to try to think how to prove it. The key is that rational numbers are dense in the set of real numbers, i.e. between any two real numbers, rational or not, there is always a rational number.

The ordering of $\mathbb{Q}$ is dense and has no largest and no smallest element. Moreover, $\mathbb{Q}$ can be constructed step-by-step. Cantor proved that any ordered set $(D, <)$ that is dense, has no largest and no smallest element, and can be constructed step-by-step, is an isomorphic copy of $(\mathbb{Q}, <)$. This means that there is a function f that to every rational number p assigns exactly one element $f(p)$ of D; every element of D is assigned to some p in $\mathbb{Q}$, and f preserves the order, i.e. for all p and q in $\mathbb{Q}$, $p < q$ if and only if $f(p) < f(q)$.

5.3.2 Summary

Let us summarize. We have just seen a construction of a new structure

$$(\mathbb{R}, \mathsf{Less}_{\mathbb{R}}, \mathsf{Add}_{\mathbb{R}}, \mathsf{Mult}_{\mathbb{R}})$$

with the domain $\mathbb{R}$ and the set of three relations on it. After identifying the rational numbers with the rational Dedekind cuts (in $\mathbb{Q}$), we can see that the new structure extends $(\mathbb{Q}, \mathsf{Less}_{\mathbb{Q}}, \mathsf{Add}_{\mathbb{Q}}, \mathsf{Mult}_{\mathbb{Q}})$, in the sense that the ordering, addition, and multiplication on $\mathbb{Q}$ do not change when we pass to $\mathbb{R}$. Moreover, the rational numbers $\mathbb{Q}$ form a dense subset of $\mathbb{R}$, and the ordered structure $(\mathbb{R}, \mathsf{Less}_{\mathbb{R}})$ is Dedekind complete. The last two properties are crucial.

Let us go back to geometry for a while. In geometry, basic objects, such as points and lines, are often treated axiomatically. The axioms of geometry do not explain what these objects are; they just tell us how they interact. For example, an axiom can say "any two distinct points lie on exactly one line."

Think of a geometric line as an infinite measuring tape. It has a point marked by 0, and points marking the integers at equal distances between the consecutive ones. It is an infinitely precise tape, meaning that it has points marking all rational numbers. Between any two such points, there is another point. So there is a dense set of points marking all rational numbers, and there still is room for points representing irrational numbers such as $\sqrt{2}$ and many other. This is all fine, but if we want to understand how the geometric line is actually *made* of points, simple geometric intuition does not suffice.

What we have done so far is this. Starting with the idea of counting we constructed the number systems of natural, integer and rational numbers. Each new structure, except the real numbers, was formally defined in terms of first-order logic in the previously constructed one. Moreover, we saw that each of the sets $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ was built in step-by-step process. Even if one is not convinced that infinite processes can actually be completed via the magic limit step ω, these constructions can be carried out forever, producing larger and larger fragments of the intended structures. In this sense, we can say that we understand how those structures are built. We can almost see them. The situation changes dramatically when turn to $(\mathbb{R}, \mathsf{Less}_{\mathbb{R}}, \mathsf{Add}_{\mathbb{R}}, \mathsf{Mult}_{\mathbb{R}})$. This structure is also defined formally, but the definition is of a different character, and there is no step-by-step procedure in sight. There will be more about it in the next section, but let us concentrate on the positive aspects of the construction first.

We defined the real numbers as Dedekind cuts. These cuts—rather complex objects in themselves—are the elements of the new structure. A structure is a set with a set of relations on it. The set is $\mathbb{R}$. Each element of $\mathbb{R}$, as a Dedekind cut, and as such it has its own structure, but in the arithmetic structure of the real numbers it only matters how the elements of the domain are related to one another by the ordering, addition, and multiplication. Notice however, that we did use the structure of individual Dedekind cuts to define those relations.

The ordering of the reals is Dedekind complete; there are no gaps in it. One can say that $(\mathbb{R}, <)$ fulfills all expectations one has of the continuous geometric line. We defined a linearly ordered set that has the properties that we think a geometric line should have. We have built a model, an explicitly defined representation. Moreover, not only $(\mathbb{R}, <)$ provides a model of a geometric line, $\mathbb{R}$ is also equipped with the arithmetic structure given by $\mathsf{Add}_{\mathbb{R}}$, and $\mathsf{Mult}_{\mathbb{R}}$, so this justifies referring to the points on that line as numbers.

5.4 Dangerous Consequences

The construction of the real line $\mathbb{R}$ involves *all* Dedekind cuts, and because of that it is not really a construction in any practical sense. We can not perform infinitely many operations, but at least, if a the structure is build step-by-step in a well-defined manner, we can imagine the process continuing forever. No such conceptual comfort is available to describe filling in the gaps in the line made of rational points. It is a mathematical construction of a set-theoretic nature. For now, let us just accept that the real number line has been constructed, all gaps filled, and it can serve as an ideal measuring tape from which both the planar and the three dimensional coordinate systems can be constructed. We still do not quite understand how a line is made of points, but we at least have a chance to ground elusive geometric considerations in a seemingly more solid domain of sets. Unfortunately, it is not that simple. Some new and rather unexpected problems arise and they need to be analyzed carefully.

Consider two circles centered at the same point, one with radius 1 and the other with radius 2. The circumference of the first one is 2π, and the circumference of the second is 4π, so the second circle is twice as large as the first. For each point on the smaller circle, draw a line from the center passing through that point (see Fig. 5.2). The line will touch the larger circle at one point. You can also draw a line starting from a point on a larger circle and the line through this point and the center will touch the smaller circle at one point. We are using the fact that both circles are continuous curves hence there are no gaps in them. If we choose distinct points on the smaller circle, the corresponding points on the larger circle will also be distinct, and the same happens if we choose distinct points on the larger circle, the corresponding point on the smaller one are distinct. This reveals an interesting fact: there is a one-to-one correspondence between the set of points of the smaller circle and the set of points of the larger one. In other words, even though their sizes are different, both circles have the same *number* of points. We have not yet said what could be the "number" of points of an infinite set, but we have used the following plausible principle: if there is a one-to-one correspondence between elements of sets, then the number of elements in those sets should be the same. How is it that sets of the same size (the number of elements), can have different measures? We will not discuss it in detail, but let me just say that the main reason is that while the number of the elements of an infinite set, as defined by Cantor, is infinite, the measure of a set is a (finite) real number.[3]

In 1924, two Polish mathematicians, Stefan Banach and Alfred Tarski, published the following theorem. A solid ball in 3-dimensional space can be decomposed into five pieces. Each of the pieces can be rotated and shifted without changing distances between its points, and by such moves the pieces can be reassembled to form two

[3]Real numbers understood as Dedekind cuts, or, in the usual representation, as sequences of digits, are infinite objects. Here we mean that real numbers are finite in the sense that they measure finite quantities. For example, the area of a circle with radius 1 is π. It is definitely finite (less than 4), but π is a real number with an infinite decimal representation.

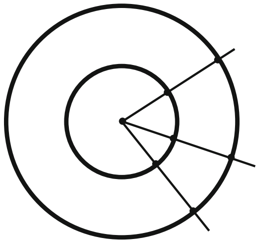

Fig. 5.2 One-to-one correspondence between points of a smaller and a larger circles

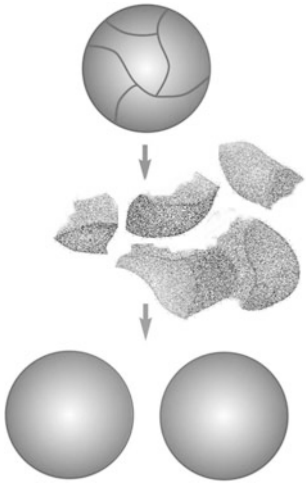

Fig. 5.3 Banach-Tarski paradox

solid balls, each of the same radius as the initial ball (see Fig. 5.3). In other words, one can cut a billiard ball into five pieces, and glue those pieces together to get two balls of the same size as the initial one. This result is known as the Banach-Tarski paradox. What is going on here is well-understood, and the result does not contradict anything we know about sets and geometry.[4] The explanation rests on a deeper understanding of properties of sets of (large) infinite size. If the pieces used in the Banach-Tarski decomposition are subjected to rigid motions in space, then their measures should not change. How can it be then, that the five pieces put together in one way make one ball, and in another two balls of the same size? The conclusion is that those pieces are not measurable. To be measurable is a technical term of measure theory. If arbitrary sets of points are admitted as objects, then, under some assumptions, there always will be objects that cannot be assigned a measure. Such are the pieces in the Banach-Tarski decomposition, but simpler examples of non-measurable sets had already been discovered earlier. One example, the Vitali set, is in the exercises to this chapter.

By the time Banach and Tarski proved their theorem, many other seemingly paradoxical phenomena had been discovered and studied. There is a bounded curve

[4]For full details see [37].

of infinite length. There is a continuous line completely filling a square. About an early result showing that the number of points in a unit circle is the same as the number of points in the whole 3-dimensional space, Cantor wrote to Dedekind: "I see it, but I don't believe it!" All those developments strongly indicated that there was a genuine need to base mathematics on solid axiomatic foundations. Since rigorous proofs can be given of results strikingly contradicting geometric intuitions, one would like to base mathematics on self-evident, undeniable truths, and then try reconstruct it rigorously just from those basic truths (the axioms). Not everyone agrees, but most mathematicians believe that task has been accomplished, and we will see how in the next chapter.

5.5 Infinite Decimals

This is a section for those who may wonder how decimal numbers are related to the subject of this chapter. In a college textbook we read: "The real numbers are the numbers that can be written in decimal notation, including those that require an infinite decimal expansion." What does that mean? What is an *infinite decimal expansion*? The talk here is about something actually infinite, so we prick up our ears. Let us start with some examples. The long division algorithm can be applied to find the decimal representation of any rational number. By dividing 8 into 1, we get 0.125. This means that $\frac{1}{8} = \frac{1}{10} + \frac{2}{100} + \frac{5}{1000}$. If we try to divide 3 into 1, the answer is less satisfying. Long division never ends, and we declare that the decimal representation of $\frac{1}{3}$ is the *infinite* decimal 0.3333333..., it is 0 followed by an infinite string 3's. But what could it possibly mean? The simple statement $\frac{1}{3} = 0.3333333\ldots$ stands for something rather advanced. It expresses that the infinite sequence of decimal fractions $\frac{3}{10}, \frac{33}{100}, \frac{333}{1000} \ldots$ *converges* to the number $\frac{1}{3}$. Each of these numbers is smaller than $\frac{1}{3}$, but for each number r less than $\frac{1}{3}$, no matter how close to $\frac{1}{3}$, there is a number in that sequence that is larger than r.

The infinite sequence 0.3333333.... is an example of a *repeating* decimal. It is often denoted by 0.(3), or $0.\bar{3}$, indicating that 3 repeats endlessly. If you randomly pick a fraction, there is a good chance that it is a repeating decimal. For example, $\frac{1}{7} = 0.142857142857\cdots = 0.(142857)$. Even more interesting is $\frac{1}{17}$. Its decimal representation is 0.(0588235294117647). The number 0.1234343434... =0.12(34) is also an example of a repeating infinite decimal. It is eventually repeating.

Here is an interesting fact: every repeating infinite decimal converges—in the sense described above—to a rational number. For example, let us see what 0.(123) converges to. Let x be the real number to which the decimal 0.(123) converges.

$$x = 0.123123\ldots$$

Then, it follows that $1000x$ converges to 123.123123.... Since 123.123123... = 123+0.123123..., we get:

$$1000x = 123 + x$$

This equation can be easily solved, giving us the answer $x = \frac{123}{999}$. This fraction can be reduced. It is equal (equivalent) to $\frac{41}{333}$.

If you are familiar with the kind of algebra that we used above, a moment's reflection will convince you that the procedure just described works for any eventually repeating decimal. This is interesting in itself, but there is also a consequence that is actually quite fascinating. Since $\sqrt{2}$ is irrational, its decimal representation is not repeating. The infinite decimal representation of $\sqrt{2}$ always changes its pattern. It never runs into a loop of repeating digits. The fascinating part of this conclusion is that, even though we can only survey finite parts of the infinite sequence of digits in the representation of $\sqrt{2}$, we do know something about the whole complete infinite sequence. We know for sure something that we could never verify by direct checking.

How does the decimal expansion of $\sqrt{2}$ relate to the Dedekind cut representing this number? Recall that the cut representing $\sqrt{2}$ is $D_{\sqrt{2}} = \{p : p \in \mathbb{Q} \wedge p^2 < 2\}$. This representation gives us an effective procedure for finding the decimal expansion of $\sqrt{2}$. By direct calculations, one can check that

- 1 is the largest integer whose square is less than 2;
- 1.4 is the largest two digit decimal whose square is less than 2;
- 1.41 is the largest three digit decimal whose square is less than 2;
- 1.414 is the largest four digit decimal whose square is less than 2;
- 1.4 142 is the largest five digit decimal whose square is less than 2;
- and so on…

Proceeding this way, one can accurately compute arbitrarily long decimal approximations to $\sqrt{2}$. Of course, to say that one can compute, is a stretch. After a few steps, such calculations become too tedious to perform. The algorithm described above is not efficient, there are much better algorithms.

Exercises

Exercise 5.1 * *Prove that the ordered set* $(\mathbb{R}, \mathsf{Less}_{\mathbb{R}})$ *is Dedekind complete.*

Exercise 5.2 * *Write a formal definition of multiplication of real numbers represented as Dedekind cuts.*

Exercise 5.3 *We know that* $\sqrt{2}$ *is irrational. Prove that* $\sqrt{2}+1$ *and* $2\sqrt{2}$ *are also irrational. Hint: Assume that* $\sqrt{2}+1 = \frac{p}{q}$*, where* p *and* q *are integers, and use simple algebra to derive a contradiction; similarly for* $2\sqrt{2}$.

Exercise 5.4 *Prove that if* a *and* b *are rational numbers and* $a \neq 0$*, then* $a\sqrt{2}+b$ *is irrational.*

Exercise 5.5 *Give an example of two irrational numbers* a *and* b *such that* $a+b$ *is rational. Hint: For all* a*,* $a + (-a) = 0$.

Exercise 5.6 *Give an example of two irrational numbers a and b such that* $a \cdot b$ *is rational. Hint: For all* $a \geq 0$, $\sqrt{a} \cdot \sqrt{a} = a$.

Exercise 5.7 *Use long division to find the decimal representations of* $\frac{1}{7}$ *and* $\frac{1}{17}$.

Exercise 5.8 ** **The Vitali set**. In this exercise you are asked to fill in details of the construction of a set of real numbers that is not measurable.*

- *We define a relation* E *on the set of real numbers in the interval* $[0, 1] = \{x : x \in \mathbb{R} \wedge 0 \leq x \leq 1\}$ *by saying that the numbers a and b are related, if the distance between them is a rational number. Show that* E *is an equivalence relation.*
- *For each equivalence class* $[a]_E = \{b : aEb\}$, *we select one element* c_a *in it. The Vitali set* V *is the set* $\{c_a : a \in [0, 1]\}$.
- *For each rational number* $p \in [0, 1]$ *we define the set* V_p *to be the union of two sets* $\{x + p : x \in V \wedge x + p \leq 1\}$ *and* $\{x + p - 1 : x \in V \wedge x + p > 1\}$. *In other words, for each* p, *the set* V_p *is obtained by shifting the set* V *to the right by* p, *and cutting and pasting the part of it that sticks out beyond* 1 *at the left end of the interval* $[0, 1]$.
- *Here are two important properties of the sets* V_p:

  (1) *If* $p \neq q$, *then* V_p *and* V_q *are disjoint.*
  (2) *For each* $x \in [0, 1]$ *there is a rational number* p *such that either* $x + p$ *or* $x - p$ *is in* V.

- *Suppose the set* V *can be assigned some measure* m. *There are two possibilities: either* $m = 0$ *(yes, there can be nonempty sets with measure is* 0, *for example a single point, or a finite collection of points is like that) or* $m > 0$. *If* $m = 0$, *then, since* V_p *is obtained from* V *by rigid motions, the measure of* V_p *is also* 0. *It follows that the whole interval* $[0, 1]$ *is covered by the sets* V_p, *which are all disjoint, and whose measure is* 0. *The problem now is that the set of rational numbers can be constructed in a step-by-step process, and it follows that the whole interval* $[0, 1]$ *would be covered by disjoint sets of measure* 0 *in a step-by-step process. Measure theory does not allow this to happen, because then the measure of the whole interval* $[0, 1]$ *would be* 0, *and that is a contradiction. If* $m > 0$, *then by a similar argument, it follows that the interval* $[0, 1]$ *would be covered by infinitely many disjoint sets, each of measure* m; *hence the measure of* $[0, 1]$ *would have to be infinite, and that is a contradiction as well.*