Chapter 3
Real Numbers

What are the real numbers? The usual answer is that they comprise the rational and irrational numbers. That is correct, but what are the irrational numbers? They are the numbers whose infinite decimal expansions are infinite and nonrepeating. But for this, we need to know precisely what an infinite decimal expansion is. We obtain an infinite decimal expansion by

1. executing a division algorithm indefinitely, e.g.,

$$\frac{1}{7} = 0{,}142857142857\ldots,$$

   or
2. locating a point on the number line, e.g.,

$$1 < \sqrt{2} < 2$$
$$1{,}4 < \sqrt{2} < 1{,}5$$
$$1{,}41 < \sqrt{2} < 1{,}42$$

   etc. Based on this, we can say the decimal expansion of $\sqrt{2}$ is $1.41\ldots$.

We note that the decimal expansion in 1. above also determines the location of the respective (rational) point on the number line. That is, a decimal expansion like $1.41421356\ldots$, always locates a point on the number line.

Now the question is the following: is the decimal expansion the *number itself*, or just a representation of it? The latter hypothesis is supported by the fact that we can write the number in different numeral systems to obtain different representations. For example, the number $1/2$ has the decimal form 0.5, while in binary, $1/2 = 0.1$; in ternary, $1/2 = 0.111\ldots$. But if decimal expansions are just a representation of the number, then what is the number itself? Perhaps it is the *point* that is represented by the decimal expansion?

M. Laczkovich, V.T. Sós, *Real Analysis*, Undergraduate Texts in Mathematics, DOI 10.1007/978-1-4939-2766-1_3

We can imagine the real numbers in various ways. But they will always be objects that we can use to measure distance, area, etc., and on which we can have operations (addition, multiplication, etc.). In the end, there are two ways in which we can clarify what real numbers are:

I. **Constructive approach.** We state that one of the above (or another) concepts defines the real numbers. For example, we can declare that the set of real numbers is the set of all infinite decimal expansions. The advantage of such a construction is that it answers the question as to what the real numbers are, albeit arbitrarily. But a disadvantage also arises: those who had a different image of real numbers will have a hard time following.
In a constructive approach, operations are usually not easy to define. (For example, it is not really clear what the product $2 \cdot 0.898899888999\ldots$ should be.) Properties of an operation, such as the distributive law $a(b+c) = ab+ac$, can also be inconvenient to check.

II. **Axiomatic approach.** In the axiomatic construction, we do not state *what* the real numbers are, but what properties they satisfy. Everyone can imagine what they want, but we fix some basic properties, and we refer only to these. Whoever accepts that the real numbers are *like* this also must accept any conclusions that we logically draw from the basic properties. Of course, these properties should be something that we should generally expect from the real numbers.
The axiomatic approach does not make direct constructions useless, since the question arises whether there exists such a construct satisfying the basic properties that we have stated. Several direct constructions are known. The construction by infinite decimal expansions can be seen in the book [1], while another construction will be touched upon in Remark 6.14. A third construction can be seen in Walter Rudin's textbook [7].
In the following sections, we follow an axiomatic approach to the real numbers. In particular, the notion of real numbers will be a fundamental concept. The basic properties that we accept without proof are called the *axioms* of the real numbers. We state the axioms in four groups. The first group consists of axioms regarding operations (addition and multiplication), while the second is made up of properties about ordering (less than, greater than). The third and fourth groups are just one axiom each. These will express that there are "arbitrarily large" natural numbers, as well as the fact that the set of real numbers is "complete".

I. Field Axioms

The first group of axioms requires some further fundamental concepts. We denote the set of real numbers by $\mathbb{R}$. We assume that there are two operations defined on the real numbers, called *addition* and *multiplication*. By this, we mean that for any two (not necessarily distinct) real numbers $a, b \in \mathbb{R}$, there correspond a number denoted by $a+b$ (the sum of a and b) as well as a number denoted by $a \cdot b$ (the product of

a and b). We also suppose that two distinct numbers 0 and 1 are specified. The first group of axioms deals with these concepts.

1. *Commutativity of addition:* $a+b=b+a$ *for each* $a,b \in \mathbb{R}$.
2. *Associativity of addition:* $(a+b)+c=a+(b+c)$ *for each* $a,b,c \in \mathbb{R}$.
3. $a+0=a$ *for each* $a \in \mathbb{R}$.
4. *For each* $a \in \mathbb{R}$*, there exists a* $b \in \mathbb{R}$ *such that* $a+b=0$.
5. *Commutativity of multiplication:* $a \cdot b = b \cdot a$ *for each* $a,b \in \mathbb{R}$.
6. *Associativity of multiplication:* $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ *for each* $a,b,c \in \mathbb{R}$.
7. $a \cdot 1 = a$ *for each* $a \in \mathbb{R}$.
8. *For each* $a \in \mathbb{R}$*,* $a \neq 0$*, there exists a* $b \in \mathbb{R}$ *such that* $a \cdot b = 1$.
9. *Distributivity:* $a \cdot (b+c) = a \cdot b + a \cdot c$ *for each* $a,b,c \in \mathbb{R}$.

If two operations that satisfy the nine axioms above are defined on a set, then we say that the set with the two operations is a **field**. (This includes the specification of the distinct elements 0 and 1.) The first group of axioms thus tells us that the real numbers form a field. This is why we call axioms 1–9 the **field axioms**.

It is easy to show that for each $a \in \mathbb{R}$, there is exactly one $b \in \mathbb{R}$ such that $a+b=0$ (see Theorem 3.28 in the first appendix). We denote this unique b by $-a$. Similarly, for each $a \neq 0$ there is exactly one $b \in \mathbb{R}$ such that $a \cdot b = 1$ (see Theorem 3.30 in the first appendix). This element b will be denoted by $\frac{1}{a}$ or $1/a$. If $c \neq 0$, then we denote $a \cdot (1/c)$ by $\frac{a}{c}$ or a/c.

We can show that all the usual properties of arithmetic we are used to and use without hesitation in solving algebraic expressions follow from the field axioms. We list these properties in the first appendix of the chapter, proving the most important ones.

II. Order axioms

The second group of axioms requires another fundamental concept. We suppose that there is a so-called **order relation**, denoted by $<$ (less than), defined on the real numbers. By this, we mean that for any two real numbers a and b, $a<b$ is either true or false. (We could also formulate this as follows: we have a map from the ordered pairs of real numbers to the logical statements "true" and "false." If (a,b) maps to "true," then we denote this by $a<b$.) The order axioms fix some properties of this order relation.

10. *Trichotomy: For any two real numbers* a *and* b*, exactly one of* $a<b$*,* $a=b$*,* $b<a$ *is true.*
11. *Transitivity: For each* $a,b,c \in \mathbb{R}$*, if* $a<b$ *and* $b<c$*, then* $a<c$.
12. *For each* $a,b,c \in \mathbb{R}$*, if* $a<b$*, then* $a+c<b+c$.
13. *For each* $a,b,c \in \mathbb{R}$*, if* $a<b$ *and* $0<c$*, then* $a \cdot c < b \cdot c$.

To talk about consequences of the order axioms, we need to introduce some notation. We say that $a \leq b$ (a is less than or equal to b) if $a < b$ or $a = b$. The notation $a > b$ is equivalent to $b < a$; $a \geq b$ is equivalent to $b \leq a$.

We call the number a **positive** if $a > 0$; **negative** if $a < 0$; **nonnegative** if $a \geq 0$; **nonpositive** if $a \leq 0$.

We call numbers that we get from adding 1 repeatedly **natural numbers**:

$$1,\ 1+1=2,\ 1+1+1=2+1=3,\ldots.$$

The set of natural numbers is denoted by $\mathbb{N}^+$.

The **integers** are the natural numbers, these times -1, and 0. The set of integers is denoted by $\mathbb{Z}$.

We denote the set of nonnegative integers by $\mathbb{N}$. That is, $\mathbb{N} = \mathbb{N}^+ \cup \{0\}$.

A real number is **rational** if we can write it as p/q, where p and q are integers and $q \neq 0$. We will denote the set of rational numbers by $\mathbb{Q}$.

A real number is called **irrational** if it is not rational.

As with the case of field axioms, any properties and rules we have previously used to deal with inequalities follow from the order axioms above. We will discuss these further in the second appendix of this chapter. We recommend that the reader look over the Bernoulli inequality as well as the relationship between the arithmetic and geometric means, as well as the harmonic and arithmetic means (Theorems 2.5, 2.6, 2.7), and check to make sure that the properties used in the proofs all follow from the order axioms.

We highlight one consequence of the order axioms, namely that the natural numbers are positive and distinct; more precisely, they satisfy the inequalities $0 < 1 < 2 < \cdots$ (see Theorem 3.38 in the second appendix). Another important fact is that there are no neighboring real numbers. That is, *for arbitrary real numbers $a < b$, there exists a c such that $a < c < b$*, for example, $c = (a+b)/2$ (see Theorem 3.40).

Definition 3.1. The *absolute value* of a real number a, denoted by $|a|$, is defined as follows:

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

From the definition of absolute value and the previously mentioned consequences of the order axioms, it is easy to check the statements below for arbitrary real numbers a and b:

$|a| \geq 0$, and $|a| = 0$ if and only if $a = 0$;

$|a| = |-a|$;

$|a \cdot b| = |a| \cdot |b|$;

If $b \neq 0$, then $\left|\frac{1}{b}\right| = \frac{1}{|b|}$ and $\left|\frac{a}{b}\right| = \frac{|a|}{|b|}$;

Triangle Inequality: $|a+b| \leq |a| + |b|$; $\quad ||a| - |b|| \leq |a-b|$.

III. The Axiom of Archimedes

14. *For an arbitrary real number b, there exists a natural number n larger than b.*

If we replace b by b/a (where a and b are positive numbers) in the above axiom, we get the following consequence:

If a and b are arbitrary positive numbers, then there exists a natural number n such that $n \cdot a > b$.

And if instead of b we write $1/\varepsilon$, where $\varepsilon > 0$, then we get:

If ε *is an arbitrary positive number, then there exists a natural number n such that* $1/n < \varepsilon$.

An important consequence of the axiom of Archimedes is that the rational numbers are "everywhere dense" within the real numbers.

Theorem 3.2. *There exists a rational number between any two real numbers.*

Proof. Let $a < b$ be real numbers, and suppose first that $0 \le a < b$. By the axiom of Archimedes, there exists a positive integer n such that $1/n < b - a$. By the same axiom, there exists a positive integer m such that $a < m/n$. Let k be the smallest positive integer for which $a < k/n$. Then

$$\frac{k-1}{n} \le a < \frac{k}{n},$$

and thus

$$\frac{k}{n} - a \le \frac{k}{n} - \frac{k-1}{n} = \frac{1}{n} < b - a.$$

So what we have obtained is that $a < k/n < b$, so we have found a rational number between a and b.

Now suppose that $a < b \le 0$. Then $0 \le -b < -a$, so by the above argument, there is a rational number r such that $-b < r < -a$. Then $a < -r < b$, and the statement is again true.

Finally, if $a < 0 < b$, then there is nothing to prove, since 0 is a rational number. □

Remark 3.3. The axiom of Archimedes is indeed necessary, since it does not follow from the previous field and order axioms. We can show this by giving an example of an ordered field (a set with two operations and an order relation where the field and order axioms hold) where the axiom of Archimedes is not satisfied. We outline the construction.

We consider the set of polynomials in a single variable with integer coefficients, that is, expressions of the form $a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$, where $n \ge 0$, $a_0, \ldots, a_n$ are integers, and $a_n \ne 0$, and also the expression 0. We denote the set of polynomials with integer coefficients by $\mathbb{Z}[x]$. We say that the polynomial $a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ is nonzero if at least one of its coefficients a_i is nonzero. In this case, the leading coefficient of the polynomial is a_k, where k is the largest index with $a_k \neq 0$. Expressions of the form p/q, where $p,q \in \mathbb{Z}[x]$ and $q \neq 0$, are called rational expressions with integer coefficients (often simply rational expressions). We denote the set of rational expressions by $\mathbb{Z}(x)$. We consider the rational expressions p/q and r/s to be equal if $ps = qr$; that is, if ps and qr expand to the same polynomial.

We define addition and multiplication of rational expressions as

$$\frac{p}{q} + \frac{r}{s} = \frac{ps+qr}{qs} \qquad \text{and} \qquad \frac{p}{q} \cdot \frac{r}{s} = \frac{pr}{qs}.$$

One can show that with these operations, the set $\mathbb{Z}(x)$ becomes a field (where $0/1$ and $1/1$ play the role of the zero and the identity element respectively). We say that the rational expression p/q is positive if the signs of the leading coefficients of p and q agree. We say that the rational expression f is less than the rational expression g if $g - f$ is positive. We denote this by $f < g$. It is easy to check that $\mathbb{Z}(x)$ is an ordered field; that is, the order axioms are also satisfied. In this structure, the natural numbers will be the constant rational expressions $n/1$.

Now we show that in the above structure, the axiom of Archimedes is not met. We must show that there exists a rational expression that is greater than every natural number. We claim that the expression $x/1$ has this property. Indeed, $(n/1) < (x/1)$ for each n, since the difference $(x/1) - (n/1) = (x-n)/1$ is positive.

IV. Cantor's Axiom

The properties listed up until now (the 14 axioms and their consequences) still do not characterize the real numbers, since it is clear that the rational numbers satisfy those same properties. On the other hand, there are properties that we expect from the real numbers but are not true for rational numbers. For example, we expect a solution in the real numbers to the equation $x^2 = 2$, but we know that a rational solution does not exist (Theorem 2.1).

The last, so-called Cantor's axiom,[1] plays a central role in analysis. It expresses the fact that the set of real numbers is in some sense "complete."

To state the axiom, we need some definitions and notation. Let $a < b$. We call the set of real numbers x for which $a \leq x \leq b$ is true a **closed interval**, and we denote it by $[a,b]$. The set of x for which $a < x < b$ holds is denoted by (a,b), and we call it an **open interval**.

Let $I_n = [a_n, b_n]$ be a closed interval for every natural number n. We say that $I_1, I_2, \ldots$ form a **sequence of nested closed intervals** if $I_1 \supset I_2 \supset \ldots \supset I_n \supset \ldots$, that is, if

$$a_n \leq a_{n+1} < b_{n+1} \leq b_n$$

holds for each n. We can now formulate Cantor's axiom.

[1] Georg Cantor (1845–1918) German mathematician.

15. *Every sequence of nested closed intervals $I_1 \supset I_2 \supset \dots$ has a common point, that is, there exists a real number x such that $x \in I_n$ for each n.*

Remark 3.4. It is important that the intervals I_n be closed: a sequence of nested open intervals does not always contain a common element. If, for example, $J_n = (0, 1/n)$ ($n = 1, 2, \dots$), then $J_1 \supset J_2 \supset \dots$, but the intervals J_n do not contain a common element. For if $x \in J_1$, then $x > 0$. Then by the axiom of Archimedes, there exists an n for which $1/n < x$, and for this n, we see that $x \notin J_n$.

Let us see when the nested closed intervals $I_1, I_2, \dots$ have exactly one common point.

Theorem 3.5. *The sequence of nested closed intervals $I_1, I_2, \dots$ has exactly one common point if and only if there is no positive number that is smaller than every $b_n - a_n$, that is, if for every $\delta > 0$, there exists an n such that $b_n - a_n < \delta$.*

Proof. If x and y are both shared elements and $x < y$, then $a_n \le x < y \le b_n$, and so $y - x \le b_n - a_n$ for each n. In other words, if the closed intervals $I_1, I_2, \dots$ contain more than one common element, then there exists a positive number smaller than each $b_n - a_n$.

Conversely, suppose that $b_n - a_n \ge \delta > 0$ for all n. Let x be a common point of the sequence of intervals. If $b_n \ge x + (\delta/2)$ for every n, then $x + (\delta/2)$ is also a common point. Similarly, if $a_n \le x - (\delta/2)$ for all n, then $x - (\delta/2)$ is also a common point. One of these two cases must hold, since if $b_n < x + (\delta/2)$ for some n and $a_m > x - (\delta/2)$ for some m, then for $k \ge \max(n, m)$, we have that $x - (\delta/2) < a_k < b_k < x + (\delta/2)$ and $b_k - a_k < \delta$, which is impossible. □

Cantor's axiom concludes the axioms of the real numbers. With this axiomatic construction, by the real numbers we mean a structure that satisfies axioms 1–15. We can also express this by saying that

the real numbers form an Archimedean ordered field in which Cantor's axiom is satisfied.

As we mentioned before, such a field exists. A sketch of the construction of a field satisfying the conditions will be given in Remark 6.14.

Before beginning our detailed exposition of the theory of analysis, we give an important example of the application of Cantor's axiom. If $a \ge 0$ and k is a positive integer, then $\sqrt[k]{a}$ denotes the nonnegative number whose kth power is a. But it is not at all obvious that such a number exists. As we saw, the first 14 axioms do not even guarantee the existence of $\sqrt{2}$. We show that from the complete axiom system, the existence of $\sqrt[k]{a}$ follows.

Theorem 3.6. *If $a \ge 0$ and k is a positive integer, then there exists exactly one non-negative real number b for which $b^k = a$.*

Proof. We can suppose that $a > 0$. We give the proof only for the special case $k = 2$; the general case can be proved similarly. (Later, we give another proof of the theorem; see Corollary 10.59.)

We will find the b we want as the common point of a sequence of nested closed intervals. Let u_1 and v_1 be nonnegative numbers for which $u_1^2 \le a \le v_1^2$. (Such are, for example, $u_1 = 0$ and $v_1 = a+1$, since $(a+1)^2 > 2 \cdot a > a$.)

Suppose that $n \ge 1$ is an integer, and we have already defined the numbers u_n and v_n such that

$$u_n^2 \le a \le v_n^2 \tag{3.1}$$

holds. We distinguish two cases. If

$$\left(\frac{u_n + v_n}{2}\right)^2 < a,$$

then let

$$u_{n+1} = \frac{u_n + v_n}{2} \quad \text{and} \quad v_{n+1} = v_n.$$

But if

$$\left(\frac{u_n + v_n}{2}\right)^2 \ge a,$$

then let

$$u_{n+1} = u_n \quad \text{and} \quad v_{n+1} = \frac{u_n + v_n}{2}.$$

It is clear that in both cases, $[u_{n+1}, v_{n+1}] \subset [u_n, v_n]$ and

$$u_{n+1}^2 \le a \le v_{n+1}^2.$$

With this, we have defined u_n and v_n for each n. It follows from the definition that the intervals $[u_n, v_n]$ are nested closed intervals, so by Cantor's axiom, there exists a common point.

If b is a common point, then $u_n \le b \le v_n$, so

$$u_n^2 \le b^2 \le v_n^2 \tag{3.2}$$

holds for each n. Thus by (3.1), a and b^2 are both common points of the interval system $[u_n^2, v_n^2]$. We want to see that $b^2 = a$. By Theorem 3.5, it suffices to show that for every $\delta > 0$, there exists n such that $v_n^2 - u_n^2 < \delta$.

We obtained the $[u_{n+1}, v_{n+1}]$ interval by "halving" the interval $[u_n, v_n]$ and taking one of the halves. From this, we see clearly that $v_{n+1} - u_{n+1} = (v_n - u_n)/2$. Of course, we can also conclude this from the definition, since

$$v_n - \frac{u_n + v_n}{2} = \frac{u_n + v_n}{2} - u_n = \frac{v_n - u_n}{2}.$$

Then by induction, we see that $v_n - u_n = (v_1 - u_1)/2^{n-1}$ for each n. From this, we see that

$$v_n^2 - u_n^2 = (v_n - u_n) \cdot (v_n + u_n) \leq \frac{v_1 - u_1}{2^{n-1}} \cdot (v_1 + v_1) \leq \frac{2 \cdot v_1^2}{2^{n-1}} = \frac{4 \cdot v_1^2}{2^n} \leq \frac{4 \cdot v_1^2}{n} \tag{3.3}$$

for each n. Let δ be an arbitrary positive number. By the axiom of Archimedes, there exists n for which $4 \cdot v_1^2/n$ is smaller than δ. This shows that for suitable n, we have $v_n^2 - u_n^2 < \delta$, so by Theorem 3.5, $b^2 = a$.

The uniqueness of b is clear. For if $0 < b_1 < b_2$, then $b_1^2 < b_2^2$, so only one of b_1^2 and b_2^2 can be equal to a. □

Let us note that Theorem 2.1 became truly complete only now. In Chapter 1, when stating the theorem, we did not care whether the number denoted by $\sqrt{2}$ actually *exists*. In the proof of Theorem 2.1, we proved only that *if* $\sqrt{2}$ *exists*, then it cannot be rational. It is true, however, that in this proof we used only the field and order axioms (check).

Exercises

3.1. Consider the set $\{0, 1, \ldots, m-1\}$ with addition and multiplication modulo m. (By this, we mean that $i + j \equiv k$ if the remainders of $i + j$ and k on dividing by m are the same, and similarly $i \cdot j \equiv k$ if the remainders of $i \cdot j$ and k on dividing by m agree.) Show that this structure satisfies the field axioms if and only if m is prime.

3.2. Give an addition and a multiplication rule on the set $\{0, 1, a, b\}$ that satisfy the field axioms.

3.3. Let F be a subset of the real numbers such that $1 \in F$ and $F \neq \{0, 1\}$. Suppose that if $a, b \in F$ and $a \neq 0$, then $(1/a) - b \in F$. Prove that F is a field.

3.4. Prove that a finite field cannot be ordered in a way that it satisfies the order axioms. (H)

3.5. Using the field and order axioms, deduce the properties of the absolute value listed.

3.6. Check that the real numbers with the operation $(a, b) \mapsto a + b + 1$ satisfy the first four axioms. What is the zero element? Define a multiplication with which we get a field.

3.7. Check that the positive real numbers with the operation $(a, b) \mapsto a \cdot b$ satisfy the first four axioms. What is the zero element? Define a multiplication with which we get a field.

3.8. Check that the positive rational numbers with the operation $(a, b) \mapsto a \cdot b$ satisfy the first four axioms. What is the zero element? Prove that *there is no* multiplication here that makes it a field.

3.9. Check that the set of rational expressions with the operations and ordering given in Remark 3.3 satisfy the field and order axioms.

3.10. Is it true that the set of rational expressions with the given operations satisfies Cantor's axiom? (H)

3.11. In Cantor's axiom, we required the sequence of nested intervals to be made up of closed, bounded, and nonempty intervals. Check that the statement of Cantor's axiom is no longer true if any of these conditions are omitted.

3.1 Decimal Expansions: The Real Line

As we mentioned before, the decimal expansion of a real number "locates a point on the number line." We will expand on this concept now. First of all, we will give the exact definition of the decimal expansion of a real number. We will need the conventional notation for finite decimal expansions: if n is a nonnegative integer and each $a_1, \ldots a_k$ is one of $0, 1, \ldots, 9$, then $n.a_1 \ldots a_k$ denotes the sum

$$n + \frac{a_1}{10} + \frac{a_2}{10^2} + \cdots + \frac{a_k}{10^k}.$$

Let x be an arbitrary nonnegative real number. We say that the **decimal expansion** of x is $n.a_1 a_2 \ldots$ if

$$\begin{aligned} n &\le x \le n+1, \\ n.a_1 &\le x \le n.a_1 + \frac{1}{10}, \\ n.a_1 a_2 &\le x \le n.a_1 a_2 + \frac{1}{10^2}, \end{aligned} \tag{3.4}$$

and so on hold.

In other words, the decimal expansion of $x \ge 0$ is $n.a_1 a_2 \ldots$ if

$$n.a_1 \ldots a_k \le x \le n.a_1 \ldots a_k + \frac{1}{10^k} \tag{3.5}$$

holds for each positive integer k.

Several questions arise from the above definitions. Does every real number have a decimal expansion? Is the expansion unique? Is every decimal expansion the decimal expansion of a real number? (That is, for a given decimal expansion, does there exist a real number whose decimal expansion agrees with it?) The following theorems give answers to these questions.

Theorem 3.7. *Every nonnegative real number has a decimal expansion.*

Proof. Let $x \ge 0$ be a given real number. By the axiom of Archimedes, there exists a positive integer larger than x. If k is the smallest positive integer larger than x, and

$n = k - 1$, then $n \le x < n + 1$. Since $n + (a/10)$ $(a = 0, 1, \ldots, 10)$ is not larger than x for $a = 0$ but larger for $a = 10$, there is an $a_1 \in \{0, 1, \ldots, 9\}$ for which $n.a_1 \le x < (n.a_1) + 1/10$. Since $(n.a_1) + (a/10^2)$ $(a = 0, 1, \ldots, 10)$ is not larger than x for $a = 0$ but larger for $a = 10$, there is an $a_2 \in \{0, 1, \ldots, 9\}$ for which $n.a_1a_2 \le x < (n.a_1a_2) + 1/10^2$. Repeating this process, we get the digits $a_1, a_2, \ldots$, which satisfy (3.5) for each k. □

Let us note that in the above theorem, we saw that for each $x \ge 0$, there is a decimal expansion $n.a_1a_2\ldots$ for which the stronger inequality

$$n.a_1 \ldots a_k \le x < n,a_1 \ldots a_k + \frac{1}{10^k} \tag{3.6}$$

holds for each positive integer k. Decimal expansions with this property are unique, since there is only one nonnegative integer for which $n \le x < n+1$, only one digit a_1 for which $n.a_1 \le x < (n.a_1) + \frac{1}{10}$, and so on.

Now if x has another decimal expansion $m.b_1b_2\ldots$, then this cannot satisfy (3.6), so either $x = m + 1$ or $x = m.b_1 \ldots b_k + (1/10^k)$ for some k. It is easy to check that in the case of $x = m + 1$, we have $n = m + 1$, $a_i = 0$ and $b_i = 9$ for each i; in the case of $x = m.b_1 \ldots b_k + (1/10^k)$ we get $a_i = 0$ and $b_i = 9$ for each $i > k$. We have then the following theorem.

Theorem 3.8. *Positive numbers with a finite decimal expansions have two infinite decimal expansions: one has all* 0 *digits from a certain point on, while the other repeats the digit* 9 *after some point. Every other nonnegative real number has a unique decimal expansion.*

The next theorem expresses the fact that the decimal expansions of real numbers contain all formal decimal expansions.

Theorem 3.9. *For arbitrary* $n \in \mathbb{N}$ *and* $(a_1, a_2, \ldots)$ *consisting of digits* $\{0, 1, \ldots, 9\}$, *there exists exactly one nonnegative real number whose decimal expansion is* $n.a_1a_2\ldots$.

Proof. The condition (3.5) expresses the fact that x is an element of

$$I_k = \left[n,a_1 \ldots a_k,\ n,a_1 \ldots a_k + \frac{1}{10^k}\right]$$

for each k. Since these are nested closed intervals, Cantor's axiom implies that there exists a real number x such that $x \in I_k$ for each k. The fact that this x is unique follows from Theorem 3.5, since by the axiom of Archimedes, for each $\delta > 0$, there exists a k such that $1/k < \delta$, and then $1/10^k < 1/k < \delta$. □

Remarks 3.10. **1.** Being able to write real numbers in decimal form has several interesting consequences. The most important corollary concerns the question of how accurately the axioms of the real numbers describe the real numbers. The question is whether we "forgot" something from the axioms; is it not possible that further properties are needed? To understand the answer, let us recollect the idea behind the axiomatic approach. Recall that we pay no attention to what the real numbers actually are, just the properties they satisfy. Instead of the set $\mathbb{R}$ we are used to, we could take another set $\mathbb{R}'$, assuming that there are two operations and a relation defined on it that satisfy the 15 axioms.

The fact that we were able to deduce our results about decimal expansions using only the 15 axioms means that these are true in both $\mathbb{R}$ and $\mathbb{R}'$. So we can pair each nonnegative element of x with the $x' \in \mathbb{R}'$ that has the same decimal expansion. We extend this association to $\mathbb{R}$ by setting $(-x)' = -x'$. By the above theorems regarding decimal expansions, we have a one-to-one correspondence[2] between $\mathbb{R}$ and $\mathbb{R}'$.

It can also be shown (although we do not go into detail here) that this correspondence commutes with the operations, that is, $(x+y)' = x' + y'$ and $(x \cdot y)' = x' \cdot y'$ for each $x, y \in \mathbb{R}$, and moreover, $x < y$ holds if and only if $x' < y'$. The existence of such a correspondence (or isomorphism, for short) shows us that $\mathbb{R}$ and $\mathbb{R}'$ are "indistinguishable": if a statement holds in one, then it holds in the other as well. This fact is expressed in mathematical logic by saying that *any two models of the axioms of real numbers are isomorphic.* So if our goal is to describe the properties of the real numbers as precisely as possible, then we have reached this goal; including other axioms cannot restrict the class of models satisfying the axioms of real numbers any further.

2. Another important consequence of being able to express real numbers as infinite decimals is that we get *a one-to-one correspondence between the set of real numbers and points on a line.*

Let e be a line, and let us pick two different points P and Q that lie on e. Call the direction $\overrightarrow{PQ}$ positive, and the opposite direction negative. In our correspondence we assign the number 0 to P, and 1 to Q. From the point P, we can then measure multiples of the distance $\overline{PQ}$ in both positive and negative directions, and assign the integers to the corresponding points. If we subdivide each segment determined by these integer points into k smaller equal parts (for each k), we get the points that we can map to rational numbers. Let x be a nonnegative real number, and let its decimal expansion be $n.a_1a_2\ldots$. Let A_k and B_k be the points that we mapped to $n.a_1\ldots a_k$, and $(n.a_1\ldots a_k)+(1/10^k)$ respectively. It follows from the properties of the line that there is exactly one common point of the segments A_kB_k. This will be the point A that we map to x. Finally, measuring the segment PA in the negative direction from P yields us the point that will be mapped to $-x$.

It can be shown that this has given us a one-to-one correspondence between the points of e and the real numbers. The real number corresponding to a certain point on the line is called its **coordinate**, and the line itself is called a **number line**, or the **real line**. In the future, when we talk about a number x, we will sometimes refer to it as the point with the coordinate x, or simply the point x.

---

[2] We say that a function f is a **one-to-one correspondence** (or a **bijective map**, or a **bijection**) between sets A and B if different points of A get mapped to different points of B (that is, if $a_1 \neq a_2$ then $f(a_1) \neq f(a_2)$), and for every $b \in B$ there exists an $a \in A$ such that $f(a) = b$.

The benefit of this correspondence is that we can better see or understand certain statements and properties of the real numbers when they are viewed as the number line. Many times, we get ideas for proofs by looking at the real line. It is, however, important to note that properties that we can "see" on the number line cannot be taken for granted, or as proved; in fact, statements that seem to be true on the real line might turn out to be false. In our proofs, we can refer only to the fundamental principles of real numbers (that is the axioms) and the theorems already established.

Looking at the real numbers as the real line suggests many concepts that prove to be important regardless of our visual interpretation. Such is, for example, the notion of an everywhere dense set.

Definition 3.11. We say that a set of real numbers H is *everywhere dense* in $\mathbb{R}$ if every open interval contains elements of H; that is, if for every $a < b$, there exists an $x \in H$ for which $a < x < b$.

So for example, by Theorem 3.2, the set of rational numbers is everywhere dense in $\mathbb{R}$. We now show that the same is true for the set of irrational numbers.

Theorem 3.12. *The set of irrational numbers is everywhere dense in* $\mathbb{R}$.

Proof. Let $a < b$ be arbitrary. Since the set of rational numbers is everywhere dense and $a - \sqrt{2} < b - \sqrt{2}$, there is a rational number r such that $a - \sqrt{2} < r < b - \sqrt{2}$. Then $a < r + \sqrt{2} < b$, and the open interval (a,b) contains the irrational number $r + \sqrt{2}$. The irrationality of $r + \sqrt{2}$ follows from the fact that if it were rational, then $\sqrt{2} = (r + \sqrt{2}) - r$ would also be rational, whereas it is not. □

Motivated by the visual representation of the real line, we shall sometimes use the word *segment* instead of *interval*. Now we will expand the concept of intervals (or segments).

Let H be a closed or open interval. Then clearly, H contains every segment whose endpoints are contained in H (this property is called convexity). It is reasonable to call every set that satisfies this property an interval. Other than closed and open intervals, the following are intervals as well.

Let $a < b$. The set of all x for which $a \le x < b$ is denoted by $[a,b)$ and is called a **left-closed, right-open interval**. The set of all x that satisfy $a < x \le b$ is denoted by $(a,b]$ and is called a **left-open, right-closed interval**. That is,

$$[a,b) = \{x : a \le x < b\} \text{ and } (a,b] = \{x : a < x \le b\}.$$

Intervals of the form $[a,b]$, (a,b), $[a,b)$, and $(a,b]$ are called **bounded** (or finite) intervals. We also introduce the notation

$$\begin{aligned} (-\infty,a] &= \{x : x \le a\}, & [a,\infty) &= \{x : x \ge a\}, \\ (-\infty,a) &= \{x : x < a\}, & (a,\infty) &= \{x : x > a\}, \end{aligned} \tag{3.7}$$

as well as $(-\infty,\infty) = \mathbb{R}$.

Intervals of the form $(-\infty, a]$, $(-\infty, a)$, $[a, \infty)$, (a, ∞) as well as $(-\infty, \infty) = \mathbb{R}$ itself are called **unbounded** (or infinite) intervals. Out of these, $(-\infty, a]$ and $[a, \infty)$ are **closed half-lines** (or closed rays), while $(-\infty, a)$ and (a, ∞) are **open half-lines** (or open rays).

We consider the empty set and a set consisting of a single point (a singleton) as intervals too; these are the **degenerate** intervals. Singletons are considered to be closed intervals, which is expressed by the notation $[a, a] = \{a\}$.

Remark 3.13. The symbol ∞ appearing in the unbounded intervals does not represent a specific object. The notation should be thought of as an abbreviation. For example, $[a, \infty)$ is merely a shorter (and more expressive) way of writing the set $\{x\colon x \geq a\}$. The symbol ∞ will pop up many more times. In every case, the same is true, so we give meaning (in each case clearly defined) only to the whole expression.

Exercises

3.12. Prove that

(a) If x and y are rational, then $x + y$ is rational.
(b) If x is rational and y is irrational, then $x + y$ is irrational.

Is it true that if x and y are irrational then $x + y$ is irrational?

3.13. Prove that the decimal expansion of a positive real number is periodic if and only if the number is rational.

3.14. Prove that the set of numbers having finite decimal expansions and their negatives is everywhere dense.

3.15. Partition the number line into the union of infinitely many pairwise disjoint everywhere dense sets.

3.16. Let $H \subset \mathbb{R}$ be a nonempty set that contains the difference of every pair of its (not necessarily distinct) elements. Prove that either there is a real number a such that $H = \{n \cdot a : n \in \mathbb{Z}\}$, or H is everywhere dense. ($*$H)

3.17. Prove that if α is irrational, then the set $\{n \cdot \alpha + k : n, k \in \mathbb{Z}\}$ is everywhere dense.

3.2 Bounded Sets

If a set of real numbers A is finite and nonempty, then there has to be a largest element among them (we can easily check this by induction on the number of elements in A). If, however, a set has infinitely many elements, then there is not necessarily a largest element. Clearly, none of the sets $\mathbb{R}$, $[a,\infty)$, and $\mathbb{N}$ have a largest element. These sets all share the property that they have elements larger than every given real number (in the case of $\mathbb{N}$, this is the axiom of Archimedes). We see that such a set can never have a largest element.

Sets without the above property are more interesting, that is, sets whose elements are all less than some real number. Such is, for example, the open interval (a,b), whose elements are all smaller than b. However the set (a,b) does not have a largest element: if $x \in (a,b)$, then $x < b$, so by Theorem 3.40, there exists a real number y such that $x < y < b$, so automatically $y \in (a,b)$. Another example is the set

$$B = \left\{ \frac{1}{2}, \frac{2}{3}, \ldots, \frac{n-1}{n}, \ldots \right\}. \tag{3.8}$$

Every element of this is less than 1, but there is no largest element, since for every element $(n-1)/n$, the next one, $n/(n+1)$ is larger.

To understand these phenomena better, we introduce some definitions and notation. If a set A has a largest (or in other words, maximal) element, we denote it by $\max A$. If a set A has a smallest (or in other words, minimal) element, we denote it by $\min A$. If the set A is finite, $A = \{a_1, \ldots a_n\}$, then instead of $\max A$, we can write $\max_{1 \le i \le n} a_i$, and similarly, we can write $\min_{1 \le i \le n} a_i$ instead of $\min A$.

Definition 3.14. We say that a set of real numbers A is *bounded from above* if there exists a number b such that for every element x of A, $x \le b$ holds (that is no element of A is larger than b). Every number b with this property is called an *upper bound* of A. So the set A is bounded from above exactly when it has an upper bound.

The set A is *bounded from below* if there exists a number c such that for every element x of A, the inequality $x \ge c$ holds (that is, no element of A is smaller than c). Every number c with such a property is called a *lower bound* of A. The set A is bounded from below if and only if it has a lower bound.

We say A is *bounded* when it is bounded both from above and from below.

Let us look at the previous notions using these new definitions. If $\max A$ exists, then it is clearly also an upper bound of A, so A is bounded from above. The following implications hold:

$$(A \text{ is finite and nonempty}) \Rightarrow (\max A \text{ exists}) \Rightarrow (A \text{ is bounded from above}).$$

The reverse implications are not usually true, since, for example, the closed interval $[a,b]$ has a largest element but is not finite, while (a,b) is bounded from above but does not have a largest element.

Further examples: The set $\mathbb{N}^+$ is bounded from below (in fact, it has a smallest element) but is not bounded from above.

$\mathbb{Z}$ is not bounded from above or from below.

Every finite interval is a bounded set.

Half-lines are not bounded.

The set $C = \{1, \frac{1}{2}, \dots, \frac{1}{n} \dots\}$ is bounded, since the greatest element is 1, so it is also an upper bound. On the other hand, every element of C is nonnegative, so 0 is a lower bound. The set does not have a smallest element.

Let us note that 0 is the *greatest* lower bound of C. Clearly, if $\varepsilon > 0$, then by the axiom of Archimedes, there is an n such that $1/n < \varepsilon$. Since $1/n \in C$, this means that ε is not a lower bound.

We can see similar behavior in the case of the open interval (a,b). As we saw, the set does not have a greatest element. But we can again find a *least* upper bound: it is easy to see that the number b is the least upper bound of (a,b). Or take the set B defined in (3.8). This does not have a largest element, but it is easy to see that among the upper bounds, there is a smallest one, namely 1. The following important theorem expresses the fact that this is true for every (nonempty) set bounded from above. Before stating the theorem, let us note that if b is an upper bound of a set H, then every number greater than b is also an upper bound. Thus a set bounded from above always has infinitely many upper bounds.

Theorem 3.15. *Every nonempty set that is bounded from above has a least upper bound.*

Proof. Let A be a nonempty set that is bounded from above. We will obtain the least upper bound of A as a common point of a sequence of nested closed intervals (similarly to the proof of Theorem 3.6).

Let v_1 be an upper bound of A. Let, moreover, a_0 be an arbitrary element of A (this exists, for we assumed $A \neq \emptyset$), and pick an arbitrary number $u_1 < a_0$. Then u_1 is not an upper bound of A, and $u_1 < v_1$ (since $u_1 < a_0 \le v_1$).

Let us suppose that $n \ge 1$ is an integer and we have defined the numbers $u_n < v_n$ such that u_n is not an upper bound, while v_n is an upper bound of A. We distinguish two cases. If $(u_n + v_n)/2$ is not an upper bound of A, then let

$$u_{n+1} = \frac{u_n + v_n}{2} \quad \text{and} \quad v_{n+1} = v_n.$$

However, if $(u_n + v_n)/2$ is an upper bound of A, then let

$$u_{n+1} = u_n \quad \text{and} \quad v_{n+1} = \frac{u_n + v_n}{2}.$$

It is clear that in both cases, $[u_{n+1}, v_{n+1}] \subset [u_n, v_n]$, and u_{n+1} is not an upper bound, while v_{n+1} is an upper bound of A.

With the above, we have defined u_n and v_n for every n. It follows from the definition that the sequence of intervals $[u_n, v_n]$ forms a sequence of nested closed intervals, so by Cantor's axiom, they have a common point. We can also see that there is only one common point. By Theorem 3.5, it suffices to see that for every number $\delta > 0$, there exists an n for which $v_n - u_n < \delta$. But it is clear that (just as in the proof of Theorem 3.6) $v_n - u_n = (v_1 - u_1)/2^{n-1}$ for each n. So if n is large enough that $(v_1 - u_1)/2^{n-1} < \delta$ (which will happen for some n by the axiom of Archimedes), then $v_n - u_n < \delta$ will also hold.

So we see that the sequence of intervals $[u_n, v_n]$ has one common point. Let this be denoted by b. We want to show that b is an upper bound of A. Let $a \in A$ be arbitrary, and suppose that $b < a$. Then $u_n \le b < a \le v_n$ for each n. (Here the third inequality follows from the fact that v_n is an upper bound.) This means that a is a common point of the sequence of intervals $[u_n, v_n]$, which is impossible. This shows that $a \le b$ for each $a \in A$, making b an upper bound.

Finally, we show that b is the least upper bound. Let c be another upper bound, and suppose that $c < b$. Then $u_n < c < b \le v_n$ for each n. (Here the first inequality follows from the fact that $u_n \ge c$ would mean that u_n is an upper bound, but that cannot be.) This means that c is a common point of the interval sequence $[u_n, v_n]$, which is impossible. Therefore, we have $b \le c$ for each upper bound c, making b the least upper bound. □

A straightforward modification of the above proof yields Theorem 3.15 for lower bounds.

Theorem 3.16. *Every nonempty set that is bounded from below has a greatest lower bound.*

We can reduce Theorem 3.16 to Theorem 3.15. For let A be a nonempty set bounded from below. It is easy to check that the set $B = \{-x : x \in A\}$ is nonempty and bounded from above. By Theorem 3.15, B has a least upper bound. If b is the least upper bound of B, then it is easy to see that $-b$ will be the greatest lower bound of A.

The above argument uses the field axioms (since even the existence of the numbers $-x$ requires the first four axioms). It is worth noting that Theorem 3.16 has a proof that uses only the order axioms in addition to Theorem 3.15 (see Exercise 3.25).

Definition 3.17. We call the least upper bound of a nonempty set A that is bounded from above the *supremum* of A, and denote it by $\sup A$. The greatest lower bound of a nonempty set A that is bounded from below is called the *infimum* of A, and is denoted by $\inf A$.

For completeness, we extend the notion of infimum and supremum to sets that are not bounded.

Definition 3.18. If the set A is not bounded from above, then we say that the least upper bound or supremum of A is infinity, and we denote it by $\sup A = \infty$. If the set A is not bounded from below, then we say that the greatest lower bound or infimum of A is negative infinity, and we denote it by $\inf A = -\infty$.

Remark 3.19. Clearly, $\sup A \in A$ (respectively $\inf A \in A$) holds if and only if A has a largest (respectively smallest) element, and then $\max A = \sup A$ (respectively $\min A = \inf A$). The infimum and supremum of a nonempty set A agree if and only if A is a singleton.

For any two sets $A, B \subset \mathbb{R}$, the **sumset**, denoted by $A+B$, is the set $\{a+b : a \in A,\ b \in B\}$. The following relationships hold between the supremum and infimum of sets and their sumsets.

Theorem 3.20. *If A, B are nonempty sets, then* $\sup(A+B) = \sup A + \sup B$ *and* $\inf(A+B) = \inf A + \inf B$.

If either one of $\sup A$ and $\sup B$ is infinity, then what we mean by the statement is that $\sup(A+B)$ is infinity. Similarly, if either one of $\inf A$ and $\inf B$ is negative infinity, then the statement is to be understood to mean that $\inf(A+B)$ is negative infinity.

Proof. We prove only the statement regarding the supremum. If $\sup A = \infty$, then A is not bounded from above. It is clear that $A+B$ is then also not bounded from above, so $\sup(A+B) = \infty$.

Now suppose that both $\sup A$ and $\sup B$ are finite. If $a \in A$ and $b \in B$, then we know that $a+b \le \sup A + \sup B$, so $\sup A + \sup B$ is an upper bound of $A+B$.

On the other hand, if c is an upper bound of $A+B$, then for arbitrary $a \in A$ and $b \in B$, we have $a+b \le c$, that is, $a \le c-b$, so $c-b$ is an upper bound for A. Then $\sup A \le c-b$, that is, $b \le c - \sup A$ for each $b \in B$, meaning that $c - \sup A$ is an upper bound of B. From this, we get that $\sup B \le c - \sup A$, that is, $\sup A + \sup B \le c$, showing that $\sup A + \sup B$ is the least upper bound of the set $A+B$. □

Exercises

3.18. Let H be a set of real numbers. Which properties of H do the following statements express?

(a) $(\forall x \in \mathbb{R})(\exists y \in H)(x < y)$;
(b) $(\forall x \in H)(\exists y \in \mathbb{R})(x < y)$;
(c) $(\forall x \in H)(\exists y \in H)(x < y)$.

3.19. Prove that

$$\max(a,b) = \frac{|a-b|+a+b}{2} \quad \text{and} \quad \min(a,b) = \frac{-|a-b|+a+b}{2}.$$

3.20. Let $A \cap B \neq \emptyset$. What can we say about the relationships between $\sup A$, $\sup B$; $\sup(A \cup B)$, $\sup(A \cap B)$; and $\sup(A \setminus B)$?

3.21. Let $A = (0,1)$, $B = [-\sqrt{2}, \sqrt{2}]$ and

$$C = \left\{ \frac{1}{2^n} + \frac{1}{2^m} : n \in \mathbb{N}^+,\ m \in \mathbb{N}^+ \right\}.$$

Find—if they exist—the supremum, infimum, maximum, and minimum of the above sets.

3.22. Let $A \cdot B = \{a \cdot b : a \in A,\ b \in B\}$ for arbitrary sets $A, B \subset \mathbb{R}$. What kind of relationships can we find between $\sup A$, $\sup B$, $\inf A$, $\inf B$ and $\inf(A \cdot B)$, $\sup(A \cdot B)$? What if we suppose $A, B \subset (0, \infty)$?

3.23. Let A be an arbitrary set of numbers, and let

$$B = \{-b : b \in A\}, \quad C = \{1/c : c \in A,\ c \neq 0\}.$$

What kind of relationships can we find between $\sup A$, $\inf A$, $\sup B$, $\inf B$, $\sup C$, $\inf C$?

3.24. Prove that if $a > 0$, $k \in \mathbb{N}^+$, $H_- = \{x > 0 \colon x^k < a\}$, and $H_+ = \{x > 0 : x^k > a\}$, then $\sup H_- = \inf H_+ = \sqrt[k]{a}$, that is, $(\sup H_-)^k = (\inf H_+)^k = a$. (This can also show us the existence of $\sqrt[k]{a}$ for $a > 0$.)

3.25. Let X be an ordered set. (This means that we have a relation $<$ given on X that satisfies the first two of the order axioms, trichotomy and transitivity.) Suppose that whenever a nonempty subset of X has an upper bound, it has a least upper bound. Show that if a nonempty subset of X has a lower bound, then it has a greatest lower bound. (H)

3.26. We say a set $H \subset \mathbb{R}$ is convex if $[x, y] \subset H$ whenever $x, y \in H$, $x < y$. Prove that a set is convex if and only if it is an interval. (Use the theorem about the existence of a least upper bound and greatest lower bound.) Show that without assuming Cantor's axiom, this would not be true.

3.27. Assume the field axioms, the order axioms, and the statement that if a nonempty set has an upper bound, then it has a least upper bound. Deduce from this the axiom of Archimedes and Cantor's axiom. (∗ S)

3.3 Exponentiation

We denote the n-fold product $a \cdot \ldots \cdot a$ by a^n, and call it the nth **power** of a (we call n the **exponent** and a the **base**). It is clear that for all real numbers a, b and positive integers x, y, the equations

$$(ab)^x = a^x \cdot b^x, \qquad a^{x+y} = a^x \cdot a^y, \qquad (a^x)^y = a^{xy} \tag{3.9}$$

hold. Our goal is to extend the notion of exponentiation in a way that satisfies the above identities. If $a \neq 0$, the equality $a^{x+y} = a^x \cdot a^y$ can hold if and only if we declare a^0 to be 1, and a^{-n} to be $1/a^n$ for every positive integer n. Accepting this definition, we see that the three identities of (3.9) hold for each $a, b \neq 0$ and $x, y \in \mathbb{Z}$. In the following, we concern ourselves only with powers of numbers different from zero; *we define only the positive integer powers of zero (for now).*

To extend exponentiation to rational exponents, we use Theorem 3.6, guaranteeing the existence of roots. If $a < 0$ and k is odd, then we denote the number $-\sqrt[k]{|a|}$ by $\sqrt[k]{a}$. Then $(\sqrt[k]{a})^k = a$ clearly holds.

Let r be rational, and suppose that $r = p/q$, where p, q are relatively prime integers, and $q > 0$. If $a \neq 0$, then by $(a^x)^y = a^{xy}$, if $a^r = b$, then $b^q = a^p$ holds. If q is odd, then this uniquely determines b: the only possible value is $b = \sqrt[q]{a^p}$. If q is even, then p is odd, and $a^p = b^q$ must be positive. This is possible only if $a > 0$, and then $b = \pm\sqrt[q]{a^p}$. Since it is natural that every power of a positive number should be positive, the logical conclusion is that $b = \sqrt[q]{a^p}$. We accept the following definition.

Definition 3.21. Let p and q be relatively prime integers, and suppose $q > 0$. If $a > 0$, then the value of $a^{p/q}$ is defined to be $\sqrt[q]{a^p}$. We define $a^{p/q}$ similarly if $a < 0$ and q is odd.

In the following, we deal only with powers of positive numbers.

Theorem 3.22. *If $a > 0$, n, m are integers, and $m > 0$, then $a^{n/m} = \sqrt[m]{a^n}$. (Note that we did not require n and m to be relatively prime.)*

Proof. Let $n/m = p/q$, where p, q are relatively prime integers and $q > 0$. We have to show that $\sqrt[m]{a^n} = \sqrt[q]{a^p}$. Since both sides are positive, it suffices to show that $\left(\sqrt[m]{a^n}\right)^{mq} = \left(\sqrt[q]{a^p}\right)^{mq}$, that is, $a^{nq} = a^{mp}$. However, this is clear, since $n/m = p/q$, so $nq = mp$. □

Theorem 3.23. *The identities in (3.9) hold for each positive a, b and rational x, y.*

Proof. We prove only the first identity, since the rest can be proved similarly. Let $x = p/q$, where p, q are relatively prime integers and $q > 0$. Then $(ab)^x = \sqrt[q]{(ab)^p}$ and $a^x \cdot b^x = \sqrt[q]{a^p} \cdot \sqrt[q]{b^p}$. Since these are all positive numbers, it suffices to show that

$$\left(\sqrt[q]{(ab)^p}\right)^q = \left(\sqrt[q]{a^p} \cdot \sqrt[q]{b^p}\right)^q.$$

The left side equates to $(ab)^p$. To compute the right side, we apply the first identity in (3.9) to the integer powers q and then p.

We get that

$$\left(\sqrt[q]{a^p} \cdot \sqrt[q]{b^p}\right)^q = \left(\sqrt[q]{a^p}\right)^q \cdot \left(\sqrt[q]{b^p}\right)^q = a^p \cdot b^p = (ab)^p.$$

□

Theorem 3.24. *$a^r > 0$ for each positive a and rational r. If $r_1 < r_2$, then $a > 1$ implies $a^{r_1} < a^{r_2}$, while $0 < a < 1$ leads to $a^{r_1} > a^{r_2}$.*

Proof. The inequality $a^r > 0$ is clear from the definition. If $a > 1$ and p, q are positive integers, then $a^{p/q} = \sqrt[q]{a^p} > 1$, and so $a^r > 1$ for each positive rational r. So if $r_1 < r_2$, then $a^{r_2} = a^{r_1} a^{r_2 - r_1} > a^{r_1}$. The statement for $0 < a < 1$ can be proved the same way. □

To extend exponentiation to irrational powers, we will keep in mind the monotone property shown in the previous theorem. Let $a > 1$. If we require that whenever $x \le y$, we also have $a^x \le a^y$, then a^x needs to satisfy $a^r \le a^x \le a^s$ whenever s and r are rational numbers such that $r \le x \le s$. We show that this restriction uniquely defines a^x.

Theorem 3.25. *If $a > 1$, then for an arbitrary real number x, we have*

$$\sup\{a^r : r \in \mathbb{Q},\ r < x\} = \inf\{a^s : s \in \mathbb{Q},\ s > x\}. \tag{3.10}$$

If $0 < a < 1$, then for an arbitrary real number x, we have

$$\inf\{a^r : r \in \mathbb{Q},\ r < x\} = \sup\{a^s : s \in \mathbb{Q},\ s > x\}. \tag{3.11}$$

Proof. Let $a > 1$. The set $A = \{a^r : r \in \mathbb{Q},\ r < x\}$ is nonempty and bounded from above, since a^s is an upper bound for each rational s greater than x. Thus $\alpha = \sup A$ is finite, and $\alpha \le a^s$ whenever $s > x$ for rational s. Then α is a lower bound of the set $B = \{a^s : s \in \mathbb{Q},\ s > x\}$, so $\beta = \inf B$ is finite and $\alpha \le \beta$. We show that $\alpha = \beta$.

Suppose that $\beta > \alpha$, and let $(\beta/\alpha) = 1 + h$, where $h > 0$. For each positive integer n, there exists an integer k for which $k/n \le x < (k+1)/n$. Then we know that $(k-1)/n < x < (k+1)/n$, and so $a^{(k-1)/n} \le \alpha < \beta \le a^{(k+1)/n}$. We get

$$\frac{\beta}{\alpha} \le \frac{a^{(k+1)/n}}{a^{(k-1)/n}} = a^{2/n},$$

and by applying Bernoulli's inequality,

$$a^2 \ge \left(\frac{\beta}{\alpha}\right)^n = (1+h)^n \ge 1 + nh.$$

However, this is impossible if $n > a^2/h$. Thus $\alpha = \beta$, which proves (3.10). The second statement can be proved similarly. □

Definition 3.26. Let $a > 1$. For an arbitrary real number x, the number a^x denotes $\sup\{a^r : r \in \mathbb{Q},\ r < x\} = \inf\{a^s : s \in \mathbb{Q},\ s > x\}$. If $0 < a < 1$, then the value of a^x is $\inf\{a^r : r \in \mathbb{Q},\ r < x\} = \sup\{a^s : s \in \mathbb{Q},\ s > x\}$. We define the exponent 1^x to be 1 for each x.

Let us note that if x is rational, then the above definition agrees with the previously defined value by Theorem 3.25.

Theorem 3.27. *$a^x > 0$ for each positive a and real number x. If $x_1 < x_2$, then $a > 1$ implies $a^{x_1} < a^{x_2}$, while $0 < a < 1$ leads to $a^{x_1} > a^{x_2}$.*

Proof. Let $a > 1$. For an arbitrary x, we can pick a rational $r < x$ such that $a^x \ge a^r > 0$. If $x_1 < x_2$, then let r_1 and r_2 be rational numbers for which $x_1 < r_1 < r_2 < x_2$. Then $a^{x_1} \le a^{r_1} < a^{r_2} \le a^{x_2}$. We can argue similarly in the case $0 < a < 1$. □

Later, we will see (refer to Theorem 11.4) that all three identities in (3.9) hold for every $a, b > 0$ and $x, y \in \mathbb{R}$.

Exercises

3.28. Prove that if $n \in \mathbb{N}$, then $\sqrt{n}$ is either an integer or irrational.

3.29. Prove that if $n, k \in \mathbb{N}^+$, then $\sqrt[k]{n}$ is either an integer or irrational.

3.30. Let $a > 0$ be rational. Prove that if a^a is rational, then a is an integer.

3.31. Let a and b be rational numbers, $0 < a < b$. Prove that $a^b = b^a$ holds if and only if there exists an $n \in \mathbb{N}^+$ such that

$$a = \left(1 + \frac{1}{n}\right)^n \qquad \text{and} \qquad b = \left(1 + \frac{1}{n}\right)^{n+1}. \tag{H}$$

3.32. Prove that if $0 < a \le b$ are real numbers and $r > 0$ is rational, then $a^r \le b^r$. (S)

3.33. Prove that if $x > -1$ and $b \ge 1$, then $(1+x)^b \ge 1 + bx$. But if $x > -1$ and $0 \le b \le 1$, then $(1+x)^b \le 1 + bx$. (H S)

3.4 First Appendix: Consequences of the Field Axioms

Theorem 3.28. *If $a + b = 0$ and $a + c = 0$, then $b = c$.*

Proof. Using the first three axioms, we get that

$$c = c + 0 = c + (a + b) = (c + a) + b = (a + c) + b = 0 + b = b + 0 = b.$$

□

If we compare the result of the previous theorem with axiom 4, then we get that for every $a \in \mathbb{R}$, there is exactly one b such that $a + b = 0$. We denote this unique b by $-a$.

Theorem 3.29. *For every a and b, there is exactly one x such that $a = b + x$.*

Proof. If $x = (-b) + a$, then

$$b + x = b + ((-b) + a) = (b + (-b)) + a = 0 + a = a + 0 = a.$$

On the other hand, if $a = b + x$, then

$$x = x + 0 = 0 + x = ((-b) + b) + x = (-b) + (b + x) = (-b) + a.$$

□

From now on, we will denote the element $(-b) + a = a + (-b)$ by $a - b$.

Theorem 3.30. *If $a \cdot b = 1$ and $a \cdot c = 1$, then $b = c$.*

This can be proven in the same way as Theorem 3.28, using axioms 5–8 here. Comparing Theorem 3.30 with axiom 8, we obtain that for every $a \neq 0$, there exists exactly one b such that $a \cdot b = 1$. We denote this unique element b by $\frac{1}{a}$ or $1/a$.

Theorem 3.31. *For any a and $b \neq 0$, there is exactly one x such that $a = b \cdot x$.*

Proof. Mimicking the proof of Theorem 3.29 will show us that $x = a \cdot (1/b)$ is the unique real number that satisfies the condition of the theorem. □

If $a, b \in \mathbb{R}$ and $b \neq 0$, then we denote the number $a \cdot (1/b)$ by $\frac{a}{b}$ or a/b.

Theorem 3.32. *Every real number a satisfies $a \cdot 0 = 0$.*

Proof. Let $a \cdot 0 = b$. By axioms 3 and 9,

$$b = a \cdot 0 = a \cdot (0 + 0) = (a \cdot 0) + (a \cdot 0) = b + b.$$

Since $b + 0 = b$ also holds, Theorem 3.29 implies $b = 0$. □

It is easy to check that each of the following identities follows from the field axioms:

$$-a = (-1) \cdot a, \qquad (a - b) - c = a - (b + c),$$

$$(-a) \cdot b = -(a \cdot b), \qquad \frac{1}{a/b} = \frac{b}{a} \;\; (a, b \neq 0), \qquad \frac{a}{b} \cdot \frac{c}{d} = \frac{a \cdot c}{b \cdot d} \;\; (b, d \neq 0).$$

With the help of induction, it is easy to justify that putting parentheses anywhere in a sum or product of several terms does not change the value of the sum or product. For example, $(a + b) + (c + d) = (a + (b + c)) + d$ and $(a \cdot b) \cdot (c \cdot d) = a \cdot ((b \cdot c) \cdot d)$. Therefore, we can omit parentheses in sums or products; by the sum $a_1 + \cdots + a_n$ and the product $a_1 \cdot \ldots \cdot a_n$, we mean the number that we would get by putting parentheses at arbitrary places in the sum (or product), thereby adding or multiplying two numbers at a time.

3.5 Second Appendix: Consequences of the Order Axioms

Theorem 3.33. *If $a < b$ and $c < d$, then $a + c < b + d$. If $a \leq b$ and $c \leq d$, then $a + c \leq b + d$.*

Proof. Let $a < b$ and $c < d$. Applying axiom 12 and commutativity twice, we get that $a + c < b + c = c + b < d + b = b + d$. The second statement is a clear consequence of this. □

We can similarly show that if $0 < a < b$ and $0 < c < d$, then $a \cdot c < b \cdot d$; moreover, if $0 \leq a \leq b$ and $0 \leq c \leq d$, then $a \cdot c \leq b \cdot d$. (We need to use Theorem 3.32 for the proof of the second statement.)

Theorem 3.34. *If $a < b$ then $-a > -b$.*

Proof. By axiom 12, we have $-b = a + (-a - b) < b + (-a - b) = -a$. □

Theorem 3.35. *If $a < b$ and $c < 0$, then $a \cdot c > b \cdot c$. If $a \leq b$ and $c \leq 0$, then $a \cdot c \geq b \cdot c$.*

Proof. Let $a < b$ and $c < 0$. By the previous theorem, $-c > 0$, so by axiom 13, we have $-a \cdot c = a \cdot (-c) < b \cdot (-c) = -b \cdot c$. Thus using the previous theorem again, we obtain $a \cdot c > b \cdot c$. The second statement is a simple consequence of this one, using Theorem 3.32. □

Theorem 3.36. $1 > 0$.

Proof. By axiom 10, it suffices to show that neither the statement $1 = 0$ nor $1 < 0$ holds. We initially assumed the numbers 0 and 1 to be distinct, so all we need to exclude is the case $1 < 0$. Suppose that $1 < 0$. Then by Theorem 3.35, we have $1 \cdot 1 > 0 \cdot 1$, that is, $1 > 0$. However, this contradicts our assumption. Thus we conclude that our assumption was false, and so $1 > 0$. □

Theorem 3.37. *If $a > 0$, then $1/a > 0$. If $0 < a < b$, then $1/a > 1/b$. If $a \neq 0$, then $a^2 > 0$.*

Proof. Let $a > 0$. If $1/a \leq 0$, then by Theorems 3.35 and 3.32, we must have $1 = (1/a) \cdot a \leq 0 \cdot a = 0$, which is impossible. Thus by axiom 10, only $1/a > 0$ is possible.

Now suppose that $0 < a < b$. If $1/a \leq 1/b$, then by axiom 13, $b = (1/a) \cdot a \cdot b \leq (1/b) \cdot a \cdot b = a$, which is impossible.

The third statement is clear by axiom 13 and Theorem 3.35. □

Theorem 3.38. *The natural numbers are positive and distinct.*

Proof. By Theorem 3.36, we have that $0 < 1$. Knowing this, axiom 12 then implies $1 = 0 + 1 < 1 + 1 = 2$. Applying axiom 12 again yields $2 = 1 + 1 < 2 + 1 = 3$, and so on. Finally, we get that $0 < 1 < 2 < \cdots$, and using transitivity, both statements of the theorem are clear. □

Theorem 3.39. *If n is a natural number, then for every real number a, we have*

$$\underbrace{a + \cdots + a}_{n \text{ terms}} = n \cdot a.$$

Proof. $\underbrace{a + \cdots + a}_{n \text{ terms}} = \underbrace{1 \cdot a + \cdots + 1 \cdot a}_{n \text{ terms}} = (\underbrace{1 + \cdots + 1}_{n \text{ terms}}) \cdot a = n \cdot a.$ □

Theorem 3.40. *There are no neighboring real numbers. That is, for arbitrary real numbers $a < b$, there exists a real number c such that $a < c < b$. One such number, for example, is $c = (a + b)/2$.*

Proof. By Theorem 3.39 and axiom 12, $2 \cdot a = a + a < a + b$. If we multiply this inequality by $1/2$ (which is positive by Theorems 3.37 and 3.38), then we get that $a < c$. A similar argument shows that $c < b$. □