CHAPTER 1 Introduction

The underlying space for all the analysis in this book is the set of real numbers. In this chapter we set down some basic properties of this set. These properties will serve as our axioms in the sense that it is possible to derive all the properties of the real numbers using only these axioms. However, we will avoid getting bogged down in this endeavor. Some readers may wish to refer to the appendix on set notation.

§1 The Set $\mathbb{N}$ of Natural Numbers

We denote the set $\{1, 2, 3, \ldots\}$ of all *positive integers* by $\mathbb{N}$. Each positive integer n has a successor, namely $n + 1$. Thus the successor of 2 is 3, and 37 is the successor of 36. You will probably agree that the following properties of $\mathbb{N}$ are obvious; at least the first four are.

N1. 1 belongs to $\mathbb{N}$.
N2. If n belongs to $\mathbb{N}$, then its successor $n + 1$ belongs to $\mathbb{N}$.
N3. 1 is not the successor of any element in $\mathbb{N}$.

K.A. Ross, *Elementary Analysis: The Theory of Calculus*,
Undergraduate Texts in Mathematics, DOI 10.1007/978-1-4614-6271-2_1,

N4. If n and m in $\mathbb{N}$ have the same successor, then $n = m$.
N5. A subset of $\mathbb{N}$ which contains 1, and which contains $n + 1$ whenever it contains n, must equal $\mathbb{N}$.

Properties N1 through N5 are known as the *Peano Axioms* or *Peano Postulates*. It turns out most familiar properties of $\mathbb{N}$ can be proved based on these five axioms; see [8] or [39].

Let's focus our attention on axiom N5, the one axiom that may not be obvious. Here is what the axiom is saying. Consider a subset S of $\mathbb{N}$ as described in N5. Then 1 belongs to S. Since S contains $n + 1$ whenever it contains n, it follows that S contains $2 = 1 + 1$. Again, since S contains $n+1$ whenever it contains n, it follows that S contains $3 = 2+1$. Once again, since S contains $n+1$ whenever it contains n, it follows that S contains $4 = 3+1$. We could continue this monotonous line of reasoning to conclude S contains any number in $\mathbb{N}$. Thus it seems reasonable to conclude $S = \mathbb{N}$. It is this reasonable conclusion that is asserted by axiom N5.

Here is another way to view axiom N5. Assume axiom N5 is false. Then $\mathbb{N}$ contains a set S such that

(i) $1 \in S$,
(ii) If $n \in S$, then $n + 1 \in S$,

and yet $S \neq \mathbb{N}$. Consider the smallest member of the set $\{n \in \mathbb{N} : n \notin S\}$, call it n_0. Since (i) holds, it is clear $n_0 \neq 1$. So n_0 is a successor to some number in $\mathbb{N}$, namely $n_0 - 1$. We have $n_0 - 1 \in S$ since n_0 is the smallest member of $\{n \in \mathbb{N} : n \notin S\}$. By (ii), the successor of $n_0 - 1$, namely n_0, is also in S, which is a contradiction. This discussion may be plausible, but we emphasize that we have not proved axiom N5 using the successor notion and axioms N1 through N4, because we implicitly used two unproven facts. We assumed every nonempty subset of $\mathbb{N}$ contains a least element and we assumed that if $n_0 \neq 1$ then n_0 is the successor to some number in $\mathbb{N}$.

Axiom N5 is the basis of mathematical induction. Let $P_1, P_2, P_3, \ldots$ be a list of statements or propositions that may or may not be true. The principle of mathematical induction asserts all the statements $P_1, P_2, P_3, \ldots$ are true provided

($\mathbf{I}_1$) P_1 is true,
($\mathbf{I}_2$) P_{n+1} is true whenever P_n is true.

We will refer to (I_1), i.e., the fact that P_1 is true, as the basis for induction and we will refer to (I_2) as the induction step. For a sound proof based on mathematical induction, properties (I_1) and (I_2) must both be verified. In practice, (I_1) will be easy to check.

Example 1
Prove $1 + 2 + \cdots + n = \frac{1}{2}n(n+1)$ for positive integers n. □

Solution
Our nth proposition is

$$P_n\text{: “}1 + 2 + \cdots + n = \frac{1}{2}n(n+1).\text{”}$$

Thus P_1 asserts $1 = \frac{1}{2} \cdot 1(1+1)$, P_2 asserts $1 + 2 = \frac{1}{2} \cdot 2(2+1)$, P_{37} asserts $1 + 2 + \cdots + 37 = \frac{1}{2} \cdot 37(37+1) = 703$, etc. In particular, P_1 is a true assertion which serves as our basis for induction.

For the induction step, suppose P_n is true. That is, we suppose

$$1 + 2 + \cdots + n = \tfrac{1}{2}n(n+1)$$

is true. Since we wish to prove P_{n+1} from this, we add $n+1$ to both sides to obtain

$$\begin{aligned} 1 + 2 + \cdots + n + (n+1) &= \tfrac{1}{2}n(n+1) + (n+1) \\ &= \tfrac{1}{2}[n(n+1) + 2(n+1)] = \tfrac{1}{2}(n+1)(n+2) \\ &= \tfrac{1}{2}(n+1)((n+1)+1). \end{aligned}$$

Thus P_{n+1} holds if P_n holds. By the principle of mathematical induction, we conclude P_n is true for all n. □

We emphasize that prior to the last sentence of our solution we did *not* prove “P_{n+1} is true.” We merely proved an implication: “if P_n is true, then P_{n+1} is true.” In a sense we proved an infinite number of assertions, namely: P_1 is true; if P_1 is true then P_2 is true; if P_2 is true then P_3 is true; if P_3 is true then P_4 is true; etc. Then we applied mathematical induction to conclude P_1 is true, P_2 is true, P_3 is true, P_4 is true, etc. We also confess that formulas like the one just proved are easier to prove than to discover. It can be a tricky matter to guess such a result. Sometimes results such as this are discovered by trial and error.

Example 2
All numbers of the form $5^n - 4n - 1$ are divisible by 16. □

Solution
More precisely, we show $5^n - 4n - 1$ is divisible by 16 for each n in $\mathbb{N}$. Our nth proposition is

$$P_n\text{: “}5^n - 4n - 1 \quad \text{is divisible by} \quad 16.\text{”}$$

The basis for induction P_1 is clearly true, since $5^1 - 4 \cdot 1 - 1 = 0$. Proposition P_2 is also true because $5^2 - 4 \cdot 2 - 1 = 16$, but note we didn't need to check this case before proceeding to the induction step. For the induction step, suppose P_n is true. To verify P_{n+1}, the trick is to write

$$5^{n+1} - 4(n+1) - 1 = 5(5^n - 4n - 1) + 16n.$$

Since $5^n - 4n - 1$ is a multiple of 16 by the induction hypothesis, it follows that $5^{n+1} - 4(n+1) - 1$ is also a multiple of 16. In fact, if $5^n - 4n - 1 = 16m$, then $5^{n+1} - 4(n+1) - 1 = 16 \cdot (5m + n)$. We have shown P_n implies P_{n+1}, so the induction step holds. An application of mathematical induction completes the proof. □

Example 3
Show $|\sin nx| \leq n|\sin x|$ for all positive integers n and all real numbers x. □

Solution
Our nth proposition is

$$P_n\text{: “}|\sin nx| \leq n|\sin x| \quad \text{for all real numbers} \quad x.\text{”}$$

The basis for induction is again clear. Suppose P_n is true. We apply the addition formula for sine to obtain

$$|\sin(n+1)x| = |\sin(nx + x)| = |\sin nx \cos x + \cos nx \sin x|.$$

Now we apply the Triangle Inequality and properties of the absolute value [see Theorems 3.7 and 3.5] to obtain

$$|\sin(n+1)x| \leq |\sin nx| \cdot |\cos x| + |\cos nx| \cdot |\sin x|.$$

Since $|\cos y| \leq 1$ for all y we see that

$$|\sin(n+1)x| \leq |\sin nx| + |\sin x|.$$

Now we apply the induction hypothesis P_n to obtain

$$|\sin(n+1)x| \leq n|\sin x| + |\sin x| = (n+1)|\sin x|.$$

Thus P_{n+1} holds. Finally, the result holds for all n by mathematical induction. $\square$

Exercises

1.1 Prove $1^2 + 2^2 + \cdots + n^2 = \frac{1}{6}n(n+1)(2n+1)$ for all positive integers n.

1.2 Prove $3 + 11 + \cdots + (8n-5) = 4n^2 - n$ for all positive integers n.

1.3 Prove $1^3 + 2^3 + \cdots + n^3 = (1+2+\cdots+n)^2$ for all positive integers n.

1.4 **(a)** Guess a formula for $1 + 3 + \cdots + (2n-1)$ by evaluating the sum for $n = 1$, 2, 3, and 4. [For $n = 1$, the sum is simply 1.]

(b) Prove your formula using mathematical induction.

1.5 Prove $1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^n} = 2 - \frac{1}{2^n}$ for all positive integers n.

1.6 Prove $(11)^n - 4^n$ is divisible by 7 when n is a positive integer.

1.7 Prove $7^n - 6n - 1$ is divisible by 36 for all positive integers n.

1.8 The principle of mathematical induction can be extended as follows. A list $P_m, P_{m+1}, \ldots$ of propositions is true provided (i) P_m is true, (ii) P_{n+1} is true whenever P_n is true and $n \geq m$.

(a) Prove $n^2 > n + 1$ for all integers $n \geq 2$.

(b) Prove $n! > n^2$ for all integers $n \geq 4$. [Recall $n! = n(n-1)\cdots 2 \cdot 1$; for example, $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$.]

1.9 **(a)** Decide for which integers the inequality $2^n > n^2$ is true.

(b) Prove your claim in (a) by mathematical induction.

1.10 Prove $(2n+1) + (2n+3) + (2n+5) + \cdots + (4n-1) = 3n^2$ for all positive integers n.

1.11 For each $n \in \mathbb{N}$, let P_n denote the assertion "$n^2 + 5n + 1$ is an even integer."

(a) Prove P_{n+1} is true whenever P_n is true.

(b) For which n is P_n actually true? What is the moral of this exercise?

1.12 For $n \in \mathbb{N}$, let $n!$ [read "n factorial"] denote the product $1 \cdot 2 \cdot 3 \cdots n$. Also let $0! = 1$ and define

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad \text{for} \quad k = 0, 1, \ldots, n. \tag{1.1}$$

The *binomial theorem* asserts that

$$(a+b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots + \binom{n}{n-1}ab^{n-1} + \binom{n}{n}b^n$$
$$= a^n + na^{n-1}b + \frac{1}{2}n(n-1)a^{n-2}b^2 + \cdots + nab^{n-1} + b^n.$$

(a) Verify the binomial theorem for $n = 1$, 2, and 3.

(b) Show $\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}$ for $k = 1, 2, \ldots, n$.

(c) Prove the binomial theorem using mathematical induction and part (b).

§2 The Set $\mathbb{Q}$ of Rational Numbers

Small children first learn to add and to multiply positive integers. After subtraction is introduced, the need to expand the number system to include 0 and negative integers becomes apparent. At this point the world of numbers is enlarged to include the set $\mathbb{Z}$ of all *integers*. Thus we have $\mathbb{Z} = \{0, 1, -1, 2, -2, \ldots\}$.

Soon the space $\mathbb{Z}$ also becomes inadequate when division is introduced. The solution is to enlarge the world of numbers to include all fractions. Accordingly, we study the space $\mathbb{Q}$ of all *rational numbers*, i.e., numbers of the form $\frac{m}{n}$ where $m, n \in \mathbb{Z}$ and $n \neq 0$. Note that $\mathbb{Q}$ contains all terminating decimals such as $1.492 = \frac{1{,}492}{1{,}000}$. The connection between decimals and real numbers is discussed in 10.3 on page 58 and in §16. The space $\mathbb{Q}$ is a highly satisfactory algebraic system in which the basic operations addition, multiplication, subtraction and division can be fully studied. No system is perfect, however, and $\mathbb{Q}$ is inadequate in some ways. In this section we will consider the defects of $\mathbb{Q}$. In the next section we will stress the good features of $\mathbb{Q}$ and then move on to the system of real numbers.

The set $\mathbb{Q}$ of rational numbers is a very nice algebraic system until one tries to solve equations like $x^2 = 2$. It turns out that no rational

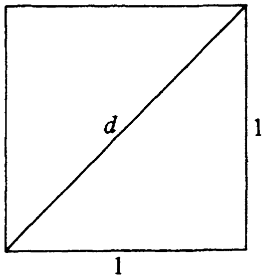

FIGURE 2.1

number satisfies this equation, and yet there are good reasons to believe some kind of number satisfies this equation. Consider, for example, a square with sides having length one; see Fig. 2.1. If d is the length of the diagonal, then from geometry we know $1^2+1^2 = d^2$, i.e., $d^2 = 2$. Apparently there is a positive length whose square is 2, which we write as $\sqrt{2}$. But $\sqrt{2}$ cannot be a rational number, as we will show in Example 2. Of course, $\sqrt{2}$ can be approximated by rational numbers. There are rational numbers whose squares are close to 2; for example, $(1.4142)^2 = 1.99996164$ and $(1.4143)^2 = 2.00024449$.

It is evident that there are lots of rational numbers and yet there are "gaps" in $\mathbb{Q}$. Here is another way to view this situation. Consider the graph of the polynomial x^2-2 in Fig. 2.2. Does the graph of x^2-2 cross the x-axis? We are inclined to say it does, because when we draw the x-axis we include "all" the points. We allow no "gaps." But notice that the graph of $x^2 - 2$ slips by all the rational numbers on the x-axis. The x-axis is our picture of the number line, and the set of rational numbers again appears to have significant "gaps."

There are even more exotic numbers such as π and e that are not rational numbers, but which come up naturally in mathematics. The number π is basic to the study of circles and spheres, and e arises in problems of exponential growth.

We return to $\sqrt{2}$. This is an example of what is called an algebraic number because it satisfies the equation $x^2 - 2 = 0$.

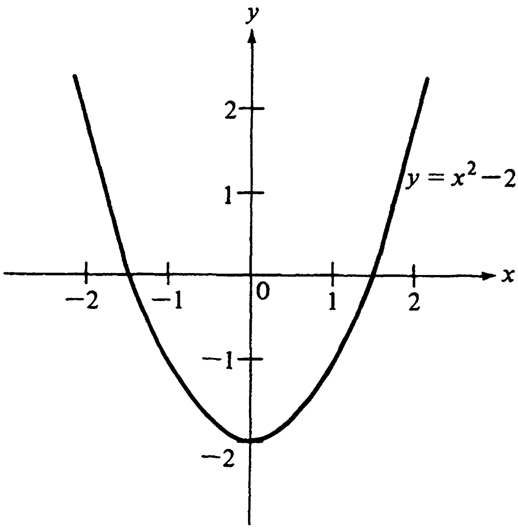

FIGURE 2.2

2.1 Definition.
A number is called an *algebraic number* if it satisfies a polynomial equation

$$c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0 = 0$$

where the coefficients $c_0, c_1, \ldots, c_n$ are integers, $c_n \neq 0$ and $n \geq 1$.

Rational numbers are always algebraic numbers. In fact, if $r = \frac{m}{n}$ is a rational number $[m, n \in \mathbb{Z}$ and $n \neq 0]$, then it satisfies the equation $nx - m = 0$. Numbers defined in terms of $\sqrt{\ }$, $\sqrt[3]{\ }$, etc. [or fractional exponents, if you prefer] and ordinary algebraic operations on the rational numbers are invariably algebraic numbers.

Example 1
$\frac{4}{17}$, $\sqrt{3}$, $\sqrt[3]{17}$, $\sqrt{2 + \sqrt[3]{5}}$ and $\sqrt{\frac{4-2\sqrt{3}}{7}}$ are algebraic numbers. In fact, $\frac{4}{17}$ is a solution of $17x - 4 = 0$, $\sqrt{3}$ is a solution of $x^2 - 3 = 0$, and $\sqrt[3]{17}$ is a solution of $x^3 - 17 = 0$. The expression $a = \sqrt{2 + \sqrt[3]{5}}$ means $a^2 = 2 + \sqrt[3]{5}$ or $a^2 - 2 = \sqrt[3]{5}$ so that $(a^2 - 2)^3 = 5$. Therefore we

have $a^6 - 6a^4 + 12a^2 - 13 = 0$, which shows $a = \sqrt{2 + \sqrt[3]{5}}$ satisfies the polynomial equation $x^6 - 6x^4 + 12x^2 - 13 = 0$.

Similarly, the expression $b = \sqrt{\frac{4-2\sqrt{3}}{7}}$ leads to $7b^2 = 4 - 2\sqrt{3}$, hence $2\sqrt{3} = 4 - 7b^2$, hence $12 = (4 - 7b^2)^2$, hence $49b^4 - 56b^2 + 4 = 0$. Thus b satisfies the polynomial equation $49x^4 - 56x^2 + 4 = 0$. □

The next theorem may be familiar from elementary algebra. It is the theorem that justifies the following remarks: the only possible rational solutions of $x^3 - 7x^2 + 2x - 12 = 0$ are $\pm1, \pm2, \pm3, \pm4, \pm6, \pm12$, so the only possible (rational) monomial factors of $x^3 - 7x^2 + 2x - 12$ are $x - 1$, $x + 1$, $x - 2$, $x + 2$, $x - 3$, $x + 3$, $x - 4$, $x + 4$, $x - 6$, $x + 6$, $x - 12$, $x + 12$. We won't pursue these algebraic problems; we merely make these observations in the hope they will be familiar.

The next theorem also allows one to prove algebraic numbers that do not look like rational numbers are usually not rational numbers. Thus $\sqrt{4}$ is obviously a rational number, while $\sqrt{2}$, $\sqrt{3}$, $\sqrt{5}$, etc. turn out to be nonrational. See the examples following the theorem. Also, compare Exercise 2.7. Recall that an integer k is a *factor* of an integer m or *divides* m if $\frac{m}{k}$ is also an integer.

If the next theorem seems complicated, first read the special case in Corollary 2.3 and Examples 2–5.

2.2 Rational Zeros Theorem.
Suppose $c_0, c_1, \ldots, c_n$ are integers and r is a rational number satisfying the polynomial equation

$$c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0 = 0 \tag{1}$$

where $n \geq 1$, $c_n \neq 0$ and $c_0 \neq 0$. Let $r = \frac{c}{d}$ where c, d are integers having no common factors and $d \neq 0$. Then c divides c_0 and d divides c_n.

In other words, the only rational *candidates* for solutions of (1) have the form $\frac{c}{d}$ where c divides c_0 and d divides c_n.

Proof
We are given

$$c_n \left(\frac{c}{d}\right)^n + c_{n-1} \left(\frac{c}{d}\right)^{n-1} + \cdots + c_1 \left(\frac{c}{d}\right) + c_0 = 0.$$

We multiply through by d^n and obtain

$$c_n c^n + c_{n-1}c^{n-1}d + c_{n-2}c^{n-2}d^2 + \cdots + c_2c^2d^{n-2} + c_1cd^{n-1} + c_0d^n = 0. \tag{2}$$

If we solve for c_0d^n, we obtain

$$c_0d^n = -c[c_nc^{n-1} + c_{n-1}c^{n-2}d + c_{n-2}c^{n-3}d^2 + \cdots + c_2cd^{n-2} + c_1d^{n-1}].$$

It follows that c divides c_0d^n. But c and d^n have no common factors, so c divides c_0. This follows from the basic fact that if an integer c divides a product ab of integers, and if c and b have no common factors, then c divides a. See, for example, Theorem 1.10 in [50].

Now we solve (2) for c_nc^n and obtain

$$c_nc^n = -d[c_{n-1}c^{n-1} + c_{n-2}c^{n-2}d + \cdots + c_2c^2d^{n-3} + c_1cd^{n-2} + c_0d^{n-1}].$$

Thus d divides c_nc^n. Since c^n and d have no common factors, d divides c_n. ■

2.3 Corollary.
Consider the polynomial equation

$$x^n + c_{n-1}x^{n-1} + \cdots + c_1x + c_0 = 0,$$

where the coefficients $c_0, c_1, \ldots, c_{n-1}$ are integers and $c_0 \neq 0$.[1] *Any rational solution of this equation must be an integer that divides c_0.*

Proof
In the Rational Zeros Theorem 2.2, the denominator of r must divide the coefficient of x^n, which is 1 in this case. Thus r is an integer and it divides c_0. ■

Example 2
$\sqrt{2}$ is not a rational number. □

Proof
By Corollary 2.3, the only rational numbers that could possibly be solutions of $x^2 - 2 = 0$ are $\pm 1, \pm 2$. [Here $n = 2$, $c_2 = 1$, $c_1 = 0$, $c_0 = -2$. So the rational solutions have the form $\frac{c}{d}$ where c divides

[1] Polynomials like this, where the highest power has coefficient 1, are called *monic* polynomials.

$c_0 = -2$ and d divides $c_2 = 1$.] One can substitute each of the four numbers $\pm 1, \pm 2$ into the equation $x^2 - 2 = 0$ to quickly eliminate them as possible solutions of the equation. Since $\sqrt{2}$ is a solution of $x^2 - 2 = 0$, it cannot be a rational number. ■

Example 3
$\sqrt{17}$ is not a rational number. □

Proof
The only possible rational solutions of $x^2 - 17 = 0$ are $\pm 1, \pm 17$, and none of these numbers are solutions. ■

Example 4
$\sqrt[3]{6}$ is not a rational number. □

Proof
The only possible rational solutions of $x^3 - 6 = 0$ are $\pm 1, \pm 2, \pm 3, \pm 6$. It is easy to verify that none of these eight numbers satisfies the equation $x^3 - 6 = 0$. ■

Example 5
$a = \sqrt{2 + \sqrt[3]{5}}$ is not a rational number. □

Proof
In Example 1 we showed a is a solution of $x^6 - 6x^4 + 12x^2 - 13 = 0$. By Corollary 2.3, the only possible rational solutions are $\pm 1, \pm 13$. When $x = 1$ or -1, the left hand side of the equation is -6 and when $x = 13$ or -13, the left hand side of the equation turns out to equal 4,657,458. This last computation could be avoided by using a little common sense. Either observe a is "obviously" bigger than 1 and less than 13, or observe

$$13^6 - 6 \cdot 13^4 + 12 \cdot 13^2 - 13 = 13(13^5 - 6 \cdot 13^3 + 12 \cdot 13 - 1) \neq 0$$

since the term in parentheses cannot be zero: it is one less than some multiple of 13. ■

Example 6
$b = \sqrt{\frac{4 - 2\sqrt{3}}{7}}$ is not a rational number. □

Proof

In Example 1 we showed b is a solution of $49x^4 - 56x^2 + 4 = 0$. By Theorem 2.2, the only possible rational solutions are

$$\pm 1, \pm 1/7, \pm 1/49, \pm 2, \pm 2/7, \pm 2/49, \pm 4, \pm 4/7, \pm 4/49.$$

To complete our proof, all we need to do is substitute these 18 candidates into the equation $49x^4 - 56x^2 + 4 = 0$. This prospect is so discouraging, however, that we choose to find a more clever approach. In Example 1, we also showed $12 = (4 - 7b^2)^2$. Now if b were rational, then $4 - 7b^2$ would also be rational [Exercise 2.6], so the equation $12 = x^2$ would have a rational solution. But the only possible rational solutions to $x^2 - 12 = 0$ are $\pm 1, \pm 2, \pm 3, \pm 4, \pm 6, \pm 12$, and these all can be eliminated by mentally substituting them into the equation. We conclude $4 - 7b^2$ cannot be rational, so b cannot be rational. ■

As a practical matter, many or all of the rational candidates given by the Rational Zeros Theorem can be eliminated by approximating the quantity in question. It is nearly obvious that the values in Examples 2 through 5 are not integers, while all the rational candidates are. The number b in Example 6 is approximately 0.2767; the nearest rational candidate is $+2/7$ which is approximately 0.2857.

It should be noted that not all irrational-looking expressions are actually irrational. See Exercise 2.7.

2.4 Remark.

While admiring the efficient Rational Zeros Theorem for finding rational zeros of polynomials with integer coefficients, you might wonder how one would find other zeros of these polynomials, or zeros of other functions. In §31, we will discuss the most well-known method, called Newton's method, and its cousin, the secant method. That discussion can be read now; only the proof of the theorem uses material from §31.

Exercises

2.1 Show $\sqrt{3}$, $\sqrt{5}$, $\sqrt{7}$, $\sqrt{24}$, and $\sqrt{31}$ are not rational numbers.

2.2 Show $\sqrt[3]{2}$, $\sqrt[7]{5}$ and $\sqrt[4]{13}$ are not rational numbers.

2.3 Show $\sqrt{2+\sqrt{2}}$ is not a rational number.

2.4 Show $\sqrt[3]{5-\sqrt{3}}$ is not a rational number.

2.5 Show $[3+\sqrt{2}]^{2/3}$ is not a rational number.

2.6 In connection with Example 6, discuss why $4-7b^2$ is rational if b is rational.

2.7 Show the following irrational-looking expressions are actually rational numbers: (a) $\sqrt{4+2\sqrt{3}}-\sqrt{3}$, and (b) $\sqrt{6+4\sqrt{2}}-\sqrt{2}$.

2.8 Find all rational solutions of the equation $x^8-4x^5+13x^3-7x+1=0$.

§3 The Set $\mathbb{R}$ of Real Numbers

The set $\mathbb{Q}$ is probably the largest system of numbers with which you really feel comfortable. There are some subtleties but you have learned to cope with them. For example, $\mathbb{Q}$ is not simply the set of symbols m/n, where $m, n \in \mathbb{Z}, n \neq 0$, since we regard some pairs of different looking fractions as equal. For example, $\frac{2}{4}$ and $\frac{3}{6}$ represent the same element of $\mathbb{Q}$. A rigorous development of $\mathbb{Q}$ based on $\mathbb{Z}$, which in turn is based on $\mathbb{N}$, would require us to introduce the notion of equivalence classes. In this book we assume a familiarity with and understanding of $\mathbb{Q}$ as an algebraic system. However, in order to clarify exactly what we need to know about $\mathbb{Q}$, we set down some of its basic axioms and properties.

The basic algebraic operations in $\mathbb{Q}$ are addition and multiplication. Given a pair a, b of rational numbers, the sum $a+b$ and the product ab also represent rational numbers. Moreover, the following properties hold.

A1. $a+(b+c)=(a+b)+c$ for all a, b, c.
A2. $a+b=b+a$ for all a, b.
A3. $a+0=a$ for all a.
A4. For each a, there is an element $-a$ such that $a+(-a)=0$.
M1. $a(bc)=(ab)c$ for all a, b, c.
M2. $ab=ba$ for all a, b.

M3. $a \cdot 1 = a$ for all a.
M4. For each $a \neq 0$, there is an element a^{-1} such that $aa^{-1} = 1$.
DL $a(b + c) = ab + ac$ for all a, b, c.

Properties A1 and M1 are called the *associative laws*, and properties A2 and M2 are the *commutative laws*. Property DL is the *distributive law*; this is the least obvious law and is the one that justifies "factorization" and "multiplying out" in algebra. A system that has more than one element and satisfies these nine properties is called a *field*. The basic algebraic properties of $\mathbb{Q}$ can proved solely on the basis of these field properties. We will not pursue this topic in any depth, but we illustrate our claim by proving some familiar properties in Theorem 3.1 below.

The set $\mathbb{Q}$ also has an order structure $\leq$ satisfying

O1. Given a and b, either $a \leq b$ or $b \leq a$.
O2. If $a \leq b$ and $b \leq a$, then $a = b$.
O3. If $a \leq b$ and $b \leq c$, then $a \leq c$.
O4. If $a \leq b$, then $a + c \leq b + c$.
O5. If $a \leq b$ and $0 \leq c$, then $ac \leq bc$.

Property O3 is called the *transitive law*. This is the characteristic property of an ordering. A field with an ordering satisfying properties O1 through O5 is called an *ordered field*. Most of the algebraic and order properties of $\mathbb{Q}$ can be established for an ordered field. We will prove a few of them in Theorem 3.2 below.

The mathematical system on which we will do our analysis will be the set $\mathbb{R}$ of all *real numbers*. The set $\mathbb{R}$ will include all rational numbers, all algebraic numbers, π, e, and more. It will be a set that can be drawn as the real number line; see Fig. 3.1. That is, every real number will correspond to a point on the number line, and every point on the number line will correspond to a real number. In particular, unlike $\mathbb{Q}$, $\mathbb{R}$ will not have any "gaps." We will also see that real numbers have decimal expansions; see 10.3 on page 58 and §16. These remarks help describe $\mathbb{R}$, but we certainly have not defined $\mathbb{R}$ as a precise mathematical object. It turns out that $\mathbb{R}$ can be defined entirely in terms of the set $\mathbb{Q}$ of rational numbers; we indicate in the enrichment §6 one way this can be done. But then it is a long and tedious task to show how to add and multiply the

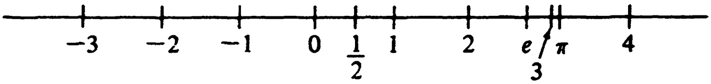

FIGURE 3.1

objects defined in this way and to show that the set $\mathbb{R}$, with these operations, satisfies all the familiar algebraic and order properties we expect to hold for $\mathbb{R}$. To develop $\mathbb{R}$ properly from $\mathbb{Q}$ in this way and to develop $\mathbb{Q}$ properly from $\mathbb{N}$ would take us several chapters. This would defeat the purpose of this book, which is to accept $\mathbb{R}$ as a mathematical system and to study some important properties of $\mathbb{R}$ and functions on $\mathbb{R}$. Nevertheless, it is desirable to specify exactly what properties of $\mathbb{R}$ we are assuming.

Real numbers, i.e., elements of $\mathbb{R}$, can be added together and multiplied together. That is, given real numbers a and b, the sum $a+b$ and the product ab also represent real numbers. Moreover, these operations satisfy the field properties A1 through A4, M1 through M4, and DL. The set $\mathbb{R}$ also has an order structure $\leq$ that satisfies properties O1 through O5. Thus, like $\mathbb{Q}$, $\mathbb{R}$ is an ordered field.

In the remainder of this section, we will obtain some results for $\mathbb{R}$ that are valid in any ordered field. In particular, these results would be equally valid if we restricted our attention to $\mathbb{Q}$. These remarks emphasize the similarities between $\mathbb{R}$ and $\mathbb{Q}$. We have not yet indicated how $\mathbb{R}$ can be distinguished from $\mathbb{Q}$ as a mathematical object, although we have asserted that $\mathbb{R}$ has no "gaps." We will make this observation much more precise in the next section, and then we will give a "gap filling" axiom that finally will distinguish $\mathbb{R}$ from $\mathbb{Q}$.

3.1 Theorem.
The following are consequences of the field properties:

(i) $a+c=b+c$ *implies* $a=b$;
(ii) $a\cdot 0=0$ *for all* a;
(iii) $(-a)b=-ab$ *for all* a,b;
(iv) $(-a)(-b)=ab$ *for all* a,b;
(v) $ac=bc$ *and* $c\neq 0$ *imply* $a=b$;
(vi) $ab=0$ *implies either* $a=0$ *or* $b=0$;

for $a,b,c\in\mathbb{R}$.

Proof

(i) $a+c=b+c$ implies $(a+c)+(-c)=(b+c)+(-c)$, so by A1, we have $a+[c+(-c)]=b+[c+(-c)]$. By A4, this reduces to $a+0=b+0$, so $a=b$ by A3.

(ii) We use A3 and DL to obtain $a\cdot 0=a\cdot(0+0)=a\cdot 0+a\cdot 0$, so $0+a\cdot 0=a\cdot 0+a\cdot 0$. By (i) we conclude $0=a\cdot 0$.

(iii) Since $a+(-a)=0$, we have $ab+(-a)b=[a+(-a)]\cdot b=0\cdot b=0=ab+(-(ab))$. From (i) we obtain $(-a)b=-(ab)$.

(iv) and (v) are left to Exercise 3.3.

(vi) If $ab=0$ and $b\neq 0$, then $0=b^{-1}\cdot 0=0\cdot b^{-1}=(ab)\cdot b^{-1}=a(bb^{-1})=a\cdot 1=a$. ∎

3.2 Theorem.
The following are consequences of the properties of an ordered field:

(i) *If* $a\leq b$, *then* $-b\leq -a$;

(ii) *If* $a\leq b$ *and* $c\leq 0$, *then* $bc\leq ac$;

(iii) *If* $0\leq a$ *and* $0\leq b$, *then* $0\leq ab$;

(iv) $0\leq a^2$ *for all* a;

(v) $0<1$;

(vi) *If* $0<a$, *then* $0<a^{-1}$;

(vii) *If* $0<a<b$, *then* $0<b^{-1}<a^{-1}$;

for $a,b,c\in\mathbb{R}$.

Note $a<b$ means $a\leq b$ and $a\neq b$.

Proof

(i) Suppose $a\leq b$. By O4 applied to $c=(-a)+(-b)$, we have $a+[(-a)+(-b)]\leq b+[(-a)+(-b)]$. It follows that $-b\leq -a$.

(ii) If $a\leq b$ and $c\leq 0$, then $0\leq -c$ by (i). Now by O5 we have $a(-c)\leq b(-c)$, i.e., $-ac\leq -bc$. From (i) again, we see $bc\leq ac$.

(iii) If we put $a=0$ in property O5, we obtain: $0\leq b$ and $0\leq c$ imply $0\leq bc$. Except for notation, this is exactly assertion (iii).

(iv) For any a, either $a\geq 0$ or $a\leq 0$ by O1. If $a\geq 0$, then $a^2\geq 0$ by (iii). If $a\leq 0$, then we have $-a\geq 0$ by (i), so $(-a)^2\geq 0$, i.e., $a^2\geq 0$.

(v) Is left to Exercise 3.4.

(vi) Suppose $0 < a$ but $0 < a^{-1}$ fails. Then we must have $a^{-1} \leq 0$ and $0 \leq -a^{-1}$. Now by (iii) $0 \leq a(-a^{-1}) = -1$, so that $1 \leq 0$, contrary to (v).

(vii) Is left to Exercise 3.4. ■

Another important notion that should be familiar is that of absolute value.

3.3 Definition.
We define

$$|a| = a \quad \text{if} \quad a \geq 0 \qquad \text{and} \qquad |a| = -a \quad \text{if} \quad a \leq 0.$$

$|a|$ is called the *absolute value of* a.

Intuitively, the absolute value of a represents the distance between 0 and a, but in fact we will *define* the idea of "distance" in terms of the "absolute value," which in turn was defined in terms of the ordering.

3.4 Definition.
For numbers a and b we define $\mathrm{dist}(a, b) = |a-b|$; $\mathrm{dist}(a, b)$ represents the *distance between a and b.*

The basic properties of the absolute value are given in the next theorem.

3.5 Theorem.
(i) $|a| \geq 0$ *for all* $a \in \mathbb{R}$.
(ii) $|ab| = |a| \cdot |b|$ *for all* $a, b \in \mathbb{R}$.
(iii) $|a + b| \leq |a| + |b|$ *for all* $a, b \in \mathbb{R}$.

Proof

(i) is obvious from the definition. [The word "obvious" as used here signifies the reader should be able to quickly see why the result is true. Certainly if $a \geq 0$, then $|a| = a \geq 0$, while $a < 0$ implies $|a| = -a > 0$. We will use expressions like "obviously" and "clearly" in place of very simple arguments, but we will not use these terms to obscure subtle points.]

(ii) There are four easy cases here. If $a \geq 0$ and $b \geq 0$, then $ab \geq 0$, so $|a| \cdot |b| = ab = |ab|$. If $a \leq 0$ and $b \leq 0$, then $-a \geq 0$, $-b \geq 0$ and $(-a)(-b) \geq 0$ so that $|a| \cdot |b| = (-a)(-b) = ab = |ab|$. If $a \geq 0$ and $b \leq 0$, then $-b \geq 0$ and $a(-b) \geq 0$ so that $|a| \cdot |b| = a(-b) = -(ab) = |ab|$. If $a \leq 0$ and $b \geq 0$, then $-a \geq 0$ and $(-a)b \geq 0$ so that $|a| \cdot |b| = (-a)b = -ab = |ab|$.

(iii) The inequalities $-|a| \leq a \leq |a|$ are obvious, since either $a = |a|$ or else $a = -|a|$. Similarly $-|b| \leq b \leq |b|$. Now four applications of O4 yield

$$-|a| + (-|b|) \leq a + b \leq |a| + b \leq |a| + |b|$$

so that

$$-(|a| + |b|) \leq a + b \leq |a| + |b|.$$

This tells us $a + b \leq |a| + |b|$ and also $-(a + b) \leq |a| + |b|$. Since $|a + b|$ is equal to either $a + b$ or $-(a + b)$, we conclude $|a + b| \leq |a| + |b|$. ■

3.6 Corollary.
$\mathrm{dist}(a, c) \leq \mathrm{dist}(a, b) + \mathrm{dist}(b, c)$ for all $a, b, c \in \mathbb{R}$.

Proof
We can apply inequality (iii) of Theorem 3.5 to $a - b$ and $b - c$ to obtain $|(a - b) + (b - c)| \leq |a - b| + |b - c|$ or $\mathrm{dist}(a, c) = |a - c| \leq |a - b| + |b - c| \leq \mathrm{dist}(a, b) + \mathrm{dist}(b, c)$. ■

The inequality in Corollary 3.6 is very closely related to an inequality concerning points **a**, **b**, **c** in the plane, and the latter inequality can be interpreted as a statement about triangles: the length of a side of a triangle is less than or equal to the sum of the lengths of the other two sides. See Fig. 3.2. For this reason, the inequality in Corollary 3.6 and its close relative (iii) in Theorem 3.5 are often called the *Triangle Inequality*.

3.7 Triangle Inequality.
$|a + b| \leq |a| + |b|$ for all a, b.

A useful variant of the triangle inequality is given in Exercise 3.5(b).

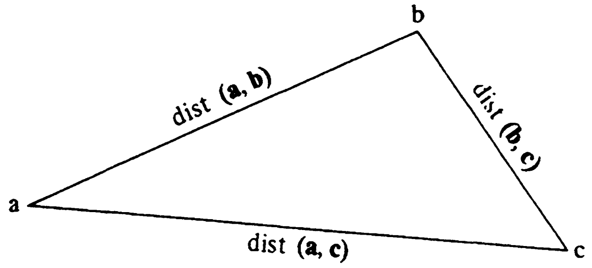

FIGURE 3.2

Exercises

3.1 **(a)** Which of the properties A1–A4, M1–M4, DL, O1–O5 fail for $\mathbb{N}$?

(b) Which of these properties fail for $\mathbb{Z}$?

3.2 **(a)** The commutative law A2 was used in the proof of (ii) in Theorem 3.1. Where?

(b) The commutative law A2 was also used in the proof of (iii) in Theorem 3.1. Where?

3.3 Prove (iv) and (v) of Theorem 3.1.

3.4 Prove (v) and (vii) of Theorem 3.2.

3.5 **(a)** Show $|b| \leq a$ if and only if $-a \leq b \leq a$.

(b) Prove $||a| - |b|| \leq |a - b|$ for all $a, b \in \mathbb{R}$.

3.6 **(a)** Prove $|a+b+c| \leq |a|+|b|+|c|$ for all $a, b, c \in \mathbb{R}$. *Hint*: Apply the triangle inequality twice. Do *not* consider eight cases.

(b) Use induction to prove

$$|a_1 + a_2 + \cdots + a_n| \leq |a_1| + |a_2| + \cdots + |a_n|$$

for n numbers $a_1, a_2, \ldots, a_n$.

3.7 **(a)** Show $|b| < a$ if and only if $-a < b < a$.

(b) Show $|a - b| < c$ if and only if $b - c < a < b + c$.

(c) Show $|a - b| \leq c$ if and only if $b - c \leq a \leq b + c$.

3.8 Let $a, b \in \mathbb{R}$. Show if $a \leq b_1$ for every $b_1 > b$, then $a \leq b$.

§4 The Completeness Axiom

In this section we give the completeness axiom for $\mathbb{R}$. This is the axiom that will assure us $\mathbb{R}$ has no "gaps." It has far-reaching consequences and almost every significant result in this book relies on it. Most theorems in this book would be false if we restricted our world of numbers to the set $\mathbb{Q}$ of rational numbers.

4.1 Definition.
Let S be a nonempty subset of $\mathbb{R}$.

(a) If S contains a largest element s_0 [that is, s_0 belongs to S and $s \leq s_0$ for all $s \in S$], then we call s_0 the *maximum of* S and write $s_0 = \max S$.

(b) If S contains a smallest element, then we call the smallest element the *minimum of* S and write it as $\min S$.

Example 1

(a) Every finite nonempty subset of $\mathbb{R}$ has a maximum and a minimum. Thus

$$\max\{1,2,3,4,5\} = 5 \quad \text{and} \quad \min\{1,2,3,4,5\} = 1,$$
$$\max\{0,\pi,-7,e,3,4/3\} = \pi \quad \text{and} \quad \min\{0,\pi,-7,e,3,4/3\} = -7,$$
$$\max\{n \in \mathbb{Z} : -4 < n \leq 100\} = 100 \quad \text{and}$$
$$\min\{n \in \mathbb{Z} : -4 < n \leq 100\} = -3.$$

(b) Consider real numbers a and b where $a < b$. The following notation will be used throughout

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

$[a,b]$ is called a *closed interval*, (a,b) is called an *open interval*, while $[a,b)$ and $(a,b]$ are called *half-open* or *semi-open intervals*. Observe $\max[a,b] = b$ and $\min[a,b] = a$. The set (a,b) has no maximum and no minimum, since the endpoints a and b do not belong to the set. The set $[a,b)$ has no maximum, but a is its minimum.

(c) The sets $\mathbb{Z}$ and $\mathbb{Q}$ have no maximum or minimum. The set $\mathbb{N}$ has no maximum, but $\min \mathbb{N} = 1$.

(d) The set $\{r \in \mathbb{Q} : 0 \le r \le \sqrt{2}\}$ has a minimum, namely 0, but no maximum. This is because $\sqrt{2}$ does not belong to the set, but there are rationals in the set arbitrarily close to $\sqrt{2}$.

(e) Consider the set $\{n^{(-1)^n} : n \in \mathbb{N}\}$. This is shorthand for the set

$$\{1^{-1}, 2, 3^{-1}, 4, 5^{-1}, 6, 7^{-1}, \dots\} = \{1, 2, \tfrac{1}{3}, 4, \tfrac{1}{5}, 6, \tfrac{1}{7}, \dots\}.$$

The set has no maximum and no minimum. □

4.2 Definition.

Let S be a nonempty subset of $\mathbb{R}$.

(a) If a real number M satisfies $s \le M$ for all $s \in S$, then M is called an *upper bound of* S and the set S is said to be *bounded above.*

(b) If a real number m satisfies $m \le s$ for all $s \in S$, then m is called a *lower bound of* S and the set S is said to be *bounded below.*

(c) The set S is said to be *bounded* if it is bounded above and bounded below. Thus S is bounded if there exist real numbers m and M such that $S \subseteq [m, M]$.

Example 2

(a) The maximum of a set is always an upper bound for the set. Likewise, the minimum of a set is always a lower bound for the set.

(b) Consider a, b in $\mathbb{R}$, $a < b$. The number b is an upper bound for each of the sets $[a, b]$, (a, b), $[a, b)$, $(a, b]$. Every number larger than b is also an upper bound for each of these sets, but b is the smallest or least upper bound.

(c) None of the sets $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{N}$ is bounded above. The set $\mathbb{N}$ is bounded below; 1 is a lower bound for $\mathbb{N}$ and so is any number less than 1. In fact, 1 is the largest or greatest lower bound.

(d) Any nonpositive real number is a lower bound for $\{r \in \mathbb{Q} : 0 \le r \le \sqrt{2}\}$ and 0 is the set's greatest lower bound. The least upper bound is $\sqrt{2}$.

(e) The set $\{n^{(-1)^n} : n \in \mathbb{N}\}$ is not bounded above. Among its many lower bounds, 0 is the greatest lower bound. □

We now formalize two notions that have already appeared in Example 2.

4.3 Definition.
Let S be a nonempty subset of $\mathbb{R}$.

(a) If S is bounded above and S has a least upper bound, then we will call it the *supremum of* S and denote it by $\sup S$.
(b) If S is bounded below and S has a greatest lower bound, then we will call it the *infimum of* S and denote it by $\inf S$.

Note that, unlike $\max S$ and $\min S$, $\sup S$ and $\inf S$ need not belong to S. Note also that a set can have at most one maximum, minimum, supremum and infimum. Sometimes the expressions "least upper bound" and "greatest lower bound" are used instead of the Latin "supremum" and "infimum" and sometimes $\sup S$ is written lub S and $\inf S$ is written glb S. We have chosen the Latin terminology for a good reason: We will be studying the notions "lim sup" and "lim inf" and this notation is completely standard; no one writes "lim lub" for instance.

Observe that if S is bounded above, then $M = \sup S$ if and only if (i) $s \le M$ for all $s \in S$, and (ii) whenever $M_1 < M$, there exists $s_1 \in S$ such that $s_1 > M_1$.

Example 3

(a) If a set S has a maximum, then $\max S = \sup S$. A similar remark applies to sets that have infimums.
(b) If $a, b \in \mathbb{R}$ and $a < b$, then

$$\sup[a,b] = \sup(a,b) = \sup[a,b) = \sup(a,b] = b.$$

(c) As noted in Example 2, we have $\inf \mathbb{N} = 1$.
(d) If $A = \{r \in \mathbb{Q} : 0 \le r \le \sqrt{2}\}$, then $\sup A = \sqrt{2}$ and $\inf A = 0$.
(e) We have $\inf\{n^{(-1)^n} : n \in \mathbb{N}\} = 0$. □

Example 4

(a) The set $A = \{\frac{1}{n^2} : n \in \mathbb{N} \text{ and } n \ge 3\}$ is bounded above and bounded below. It has a maximum, namely $\frac{1}{9}$, but it has no minimum. In fact, $\sup(A) = \frac{1}{9}$ and $\inf(A) = 0$.

(b) The set $B = \{r \in \mathbb{Q} : r^3 \leq 7\}$ is bounded above, by 2 for example. It does not have a maximum, because $r^3 \neq 7$ for all $r \in \mathbb{Q}$, by the Rational Zeros Theorem 2.2. However, $\sup(B) = \sqrt[3]{7}$. The set B is not bounded below; if this isn't obvious, think about the graph of $y = x^3$. Clearly B has no minimum. Starting with the next section, we would write $\inf(B) = -\infty$.

(c) The set $C = \{m + n\sqrt{2} : m, n \in \mathbb{Z}\}$ isn't bounded above or below, so it has no maximum or minimum. We could write $\sup(C) = +\infty$ and $\inf(C) = -\infty$.

(d) The set $D = \{x \in \mathbb{R} : x^2 < 10\}$ is the open interval $(-\sqrt{10}, \sqrt{10})$. Thus it is bounded above and below, but it has no maximum or minimum. However, $\inf(D) = -\sqrt{10}$ and $\sup(D) = \sqrt{10}$. □

Note that, in Examples 2–4, every set S that is bounded above possesses a least upper bound, i.e., $\sup S$ exists. This is not an accident. Otherwise there would be a "gap" between the set S and the set of its upper bounds.

4.4 Completeness Axiom.
Every nonempty subset S of $\mathbb{R}$ that is bounded above has a least upper bound. In other words, $\sup S$ *exists and is a real number.*

The completeness axiom for $\mathbb{Q}$ would assert that every nonempty subset of $\mathbb{Q}$, that is bounded above by some rational number, has a least upper bound that is a rational number. The set $A = \{r \in \mathbb{Q} : 0 \leq r \leq \sqrt{2}\}$ is a set of rational numbers and it is bounded above by some rational numbers [3/2 for example], but A has no least upper bound that is a rational number. Thus the completeness axiom does not hold for $\mathbb{Q}$! Incidentally, the set A can be described entirely in terms of rationals: $A = \{r \in \mathbb{Q} : 0 \leq r \text{ and } r^2 \leq 2\}$.

The completeness axiom for sets bounded below comes free.

4.5 Corollary.
Every nonempty subset S of $\mathbb{R}$ that is bounded below has a greatest lower bound $\inf S$.

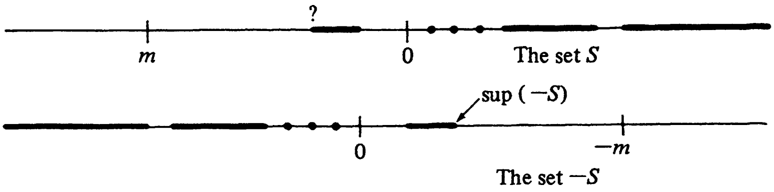

FIGURE 4.1

Proof

Let $-S$ be the set $\{-s : s \in S\}$; $-S$ consists of the negatives of the numbers in S. Since S is bounded below there is an m in $\mathbb{R}$ such that $m \le s$ for all $s \in S$. This implies $-m \ge -s$ for all $s \in S$, so $-m \ge u$ for all u in the set $-S$. Thus $-S$ is bounded above by $-m$. The Completeness Axiom 4.4 applies to $-S$, so $\sup(-S)$ exists. Figure 4.1 suggests we prove $\inf S = -\sup(-S)$.

Let $s_0 = \sup(-S)$; we need to prove

$$-s_0 \le s \quad \text{for all} \quad s \in S, \tag{1}$$

and

$$\text{if } t \le s \quad \text{for all} \quad s \in S, \quad \text{then} \quad t \le -s_0. \tag{2}$$

The inequality (1) will show $-s_0$ is a lower bound for S, while (2) will show $-s_0$ is the *greatest* lower bound, that is, $-s_0 = \inf S$. We leave the proofs of (1) and (2) to Exercise 4.9. ■

It is useful to know:

$$\text{if } a > 0, \quad \text{then} \quad \frac{1}{n} < a \quad \text{for some positive integer} \quad n, \tag{*}$$

and

$$\text{if } b > 0, \quad \text{then} \quad b < n \quad \text{for some positive integer} \quad n. \tag{**}$$

These assertions are not as obvious as they may appear. If fact, there exist ordered fields that do not have these properties. In other words, there exists a mathematical system satisfying all the properties A1–A4, M1–M4, DL and O1–O5 in §3 and yet possessing elements $a > 0$ and $b > 0$ such that $a < 1/n$ and $n < b$ *for all* n. On the other hand, such strange elements cannot exist in $\mathbb{R}$ or $\mathbb{Q}$. We next prove this; in view of the previous remarks we *must* expect to use the Completeness Axiom.

4.6 Archimedean Property.
If $a > 0$ and $b > 0$, then for some positive integer n, we have $na > b$.

This tells us that, even if a is quite small and b is quite large, some integer multiple of a will exceed b. Or, to quote [4], given enough time, one can empty a large bathtub with a small spoon. [Note that if we set $b = 1$, we obtain assertion (*), and if we set $a = 1$, we obtain assertion (**).]

Proof
Assume the Archimedean property fails. Then there exist $a > 0$ and $b > 0$ such that $na \leq b$ for all $n \in \mathbb{N}$. In particular, b is an upper bound for the set $S = \{na : n \in \mathbb{N}\}$. Let $s_0 = \sup S$; this is where we are using the completeness axiom. Since $a > 0$, we have $s_0 < s_0 + a$, so $s_0 - a < s_0$. [To be precise, we obtain $s_0 \leq s_0 + a$ and $s_0 - a \leq s_0$ by property O4 and the fact that $a + (-a) = 0$. Then we conclude $s_0 - a < s_0$ since $s_0 - a = s_0$ implies $a = 0$ by Theorem 3.1(i).] Since s_0 is the least upper bound for S, $s_0 - a$ cannot be an upper bound for S. It follows that $s_0 - a < n_0 a$ for some $n_0 \in \mathbb{N}$. This implies $s_0 < (n_0 + 1)a$. Since $(n_0 + 1)a$ is in S, s_0 is not an upper bound for S and we have reached a contradiction. Our assumption that the Archimedean property fails was wrong. ■

We give one more result that seems obvious from our experience with the real number line, but which cannot be proved for an arbitrary ordered field.

4.7 Denseness of $\mathbb{Q}$.
If $a, b \in \mathbb{R}$ and $a < b$, then there is a rational $r \in \mathbb{Q}$ such that $a < r < b$.

Proof
We need to show $a < \frac{m}{n} < b$ for some integers m and n where $n > 0$, and thus we need

$$an < m < bn. \tag{1}$$

Since $b - a > 0$, the Archimedean property shows there exists an $n \in \mathbb{N}$ such that

$$n(b-a) > 1, \quad \text{and hence} \quad bn - an > 1. \tag{2}$$

From this, it is fairly evident that there is an integer m between an and bn, so that (1) holds. However, the proof that such an m exists is a little delicate. We argue as follows. By the Archimedean property again, there exists an integer $k > \max\{|an|, |bn|\}$, so that

$$-k < an < bn < k.$$

Then the sets $K = \{j \in \mathbb{Z} : -k \le j \le k\}$ and $\{j \in K : an < j\}$ are finite, and they are nonempty, since they both contain k. Let $m = \min\{j \in K : an < j\}$. Then $-k < an < m$. Since $m > -k$, we have $m - 1$ in K, so the inequality $an < m - 1$ is false by our choice of m. Thus $m - 1 \le an$ and, using (2), we have $m \le an + 1 < bn$. Since $an < m < bn$, (1) holds. ■

Exercises

4.1 For each set below that is bounded above, list three upper bounds for the set.[2] Otherwise write "NOT BOUNDED ABOVE" or "NBA."

(a) $[0,1]$
(b) $(0,1)$
(c) $\{2,7\}$
(d) $\{\pi, e\}$
(e) $\{\frac{1}{n} : n \in \mathbb{N}\}$
(f) $\{0\}$
(g) $[0,1] \cup [2,3]$
(h) $\cup_{n=1}^{\infty}[2n, 2n+1]$
(i) $\cap_{n=1}^{\infty}[-\frac{1}{n}, 1 + \frac{1}{n}]$
(j) $\{1 - \frac{1}{3^n} : n \in \mathbb{N}\}$
(k) $\{n + \frac{(-1)^n}{n} : n \in \mathbb{N}\}$
(l) $\{r \in \mathbb{Q} : r < 2\}$
(m) $\{r \in \mathbb{Q} : r^2 < 4\}$
(n) $\{r \in \mathbb{Q} : r^2 < 2\}$
(o) $\{x \in \mathbb{R} : x < 0\}$
(p) $\{1, \frac{\pi}{3}, \pi^2, 10\}$
(q) $\{0, 1, 2, 4, 8, 16\}$
(r) $\cap_{n=1}^{\infty}(1 - \frac{1}{n}, 1 + \frac{1}{n})$
(s) $\{\frac{1}{n} : n \in \mathbb{N} \text{ and } n \text{ is prime}\}$
(t) $\{x \in \mathbb{R} : x^3 < 8\}$
(u) $\{x^2 : x \in \mathbb{R}\}$
(v) $\{\cos(\frac{n\pi}{3}) : n \in \mathbb{N}\}$
(w) $\{\sin(\frac{n\pi}{3}) : n \in \mathbb{N}\}$

4.2 Repeat Exercise 4.1 for lower bounds.

4.3 For each set in Exercise 4.1, give its supremum if it has one. Otherwise write "NO sup."

[2] An integer $p \ge 2$ is a *prime* provided the only positive factors of p are 1 and p.

4.4 Repeat Exercise 4.3 for infima [plural of infimum].

4.5 Let S be a nonempty subset of $\mathbb{R}$ that is bounded above. Prove if $\sup S$ belongs to S, then $\sup S = \max S$. *Hint*: Your proof should be very short.

4.6 Let S be a nonempty bounded subset of $\mathbb{R}$.

(a) Prove $\inf S \leq \sup S$. *Hint*: This is almost obvious; your proof should be short.

(b) What can you say about S if $\inf S = \sup S$?

4.7 Let S and T be nonempty bounded subsets of $\mathbb{R}$.

(a) Prove if $S \subseteq T$, then $\inf T \leq \inf S \leq \sup S \leq \sup T$.

(b) Prove $\sup(S \cup T) = \max\{\sup S, \sup T\}$. *Note*: In part (b), do *not* assume $S \subseteq T$.

4.8 Let S and T be nonempty subsets of $\mathbb{R}$ with the following property: $s \leq t$ for all $s \in S$ and $t \in T$.

(a) Observe S is bounded above and T is bounded below.

(b) Prove $\sup S \leq \inf T$.

(c) Give an example of such sets S and T where $S \cap T$ is nonempty.

(d) Give an example of sets S and T where $\sup S = \inf T$ and $S \cap T$ is the empty set.

4.9 Complete the proof that $\inf S = -\sup(-S)$ in Corollary 4.5 by proving (1) and (2).

4.10 Prove that if $a > 0$, then there exists $n \in \mathbb{N}$ such that $\frac{1}{n} < a < n$.

4.11 Consider $a, b \in \mathbb{R}$ where $a < b$. Use Denseness of $\mathbb{Q}$ 4.7 to show there are infinitely many rationals between a and b.

4.12 Let $\mathbb{I}$ be the set of real numbers that are not rational; elements of $\mathbb{I}$ are called *irrational numbers*. Prove if $a < b$, then there exists $x \in \mathbb{I}$ such that $a < x < b$. *Hint*: First show $\{r + \sqrt{2} : r \in \mathbb{Q}\} \subseteq \mathbb{I}$.

4.13 Prove the following are equivalent for real numbers a, b, c. [*Equivalent* means that either all the properties hold or none of the properties hold.]

(i) $|a - b| < c$,

(ii) $b - c < a < b + c$,

(iii) $a \in (b - c, b + c)$.

Hint: Use Exercise 3.7(b).

4.14 Let A and B be nonempty bounded subsets of $\mathbb{R}$, and let $A + B$ be the set of all sums $a + b$ where $a \in A$ and $b \in B$.

(a) Prove $\sup(A+B) = \sup A + \sup B$. *Hint*: To show $\sup A + \sup B \leq \sup(A + B)$, show that for each $b \in B$, $\sup(A + B) - b$ is an upper bound for A, hence $\sup A \leq \sup(A + B) - b$. Then show $\sup(A + B) - \sup A$ is an upper bound for B.

(b) Prove $\inf(A + B) = \inf A + \inf B$.

4.15 Let $a, b \in \mathbb{R}$. Show if $a \leq b + \frac{1}{n}$ for all $n \in \mathbb{N}$, then $a \leq b$. Compare Exercise 3.8.

4.16 Show $\sup\{r \in \mathbb{Q} : r < a\} = a$ for each $a \in \mathbb{R}$.

§5 The Symbols $+\infty$ and $-\infty$

The symbols $+\infty$ and $-\infty$ are extremely useful even though they are *not* real numbers. We will often write $+\infty$ as simply ∞. We will adjoin $+\infty$ and $-\infty$ to the set $\mathbb{R}$ and extend our ordering to the set $\mathbb{R} \cup \{-\infty, +\infty\}$. Explicitly, we will agree that $-\infty \leq a \leq +\infty$ for all a in $\mathbb{R} \cup \{-\infty, \infty\}$. This provides the set $\mathbb{R} \cup \{-\infty, +\infty\}$ with an ordering that satisfies properties O1, O2 and O3 of §3. We emphasize we will *not* provide the set $\mathbb{R} \cup \{-\infty, +\infty\}$ with any algebraic structure. We may use the symbols $+\infty$ and $-\infty$, but we must continue to remember they do not represent real numbers. Do *not* apply a theorem or exercise that is stated for real numbers to the symbols $+\infty$ or $-\infty$.

It is convenient to use the symbols $+\infty$ and $-\infty$ to extend the notation established in Example 1(b) of §4 to unbounded intervals. For real numbers $a, b \in \mathbb{R}$, we adopt the following

$$[a, \infty) = \{x \in \mathbb{R} : a \leq x\}, \qquad (a, \infty) = \{x \in \mathbb{R} : a < x\},$$
$$(-\infty, b] = \{x \in \mathbb{R} : x \leq b\}, \qquad (-\infty, b) = \{x \in \mathbb{R} : x < b\}.$$

We occasionally also write $(-\infty, \infty)$ for $\mathbb{R}$. $[a, \infty)$ and $(-\infty, b]$ are called *closed intervals* or *unbounded closed intervals*, while (a, ∞) and

$(-\infty, b)$ are called *open intervals* or *unbounded open intervals*. Consider a nonempty subset S of $\mathbb{R}$. Recall that if S is bounded above, then $\sup S$ exists and represents a real number by the completeness axiom 4.4. We define

$$\sup S = +\infty \quad \text{if } S \text{ is not bounded above.}$$

Likewise, if S is bounded below, then $\inf S$ exists and represents a real number [Corollary 4.5]. And we define

$$\inf S = -\infty \quad \text{if } S \text{ is not bounded below.}$$

For emphasis, we recapitulate:

Let S be any nonempty subset of $\mathbb{R}$. The symbols $\sup S$ *and* $\inf S$ *always make sense. If S is bounded above, then* $\sup S$ *is a real number; otherwise* $\sup S = +\infty$. *If S is bounded below, then* $\inf S$ *is a real number; otherwise* $\inf S = -\infty$. *Moreover, we have* $\inf S \leq \sup S$.

Example 1

For nonempty bounded subsets A and B of $\mathbb{R}$, Exercise 4.14 asserts

$$\sup(A+B) = \sup A + \sup B \quad \text{and} \quad \inf(A+B) = \inf A + \inf B. \quad (1)$$

We verify the first equality is true even if A or B is unbounded, and Exercise 5.7 asks you to do the same for the second equality.

Consider $x \in A+B$, so that $x = a+b$ for some $a \in A$ and $b \in B$. Then $x = a + b \leq \sup A + \sup B$. Since x is any element in $A+B$, $\sup A + \sup B$ is an upper bound for $A+B$; hence $\sup(A+B) \leq \sup A + \sup B$. It remains to show $\sup(A+B) \geq \sup A + \sup B$.

Since the sets here are nonempty, the suprema here are not equal to $-\infty$, so we're not in danger of encountering the undefined sum $-\infty + \infty$. If $\sup A + \sup B = +\infty$, then at least one of the suprema, say $\sup B$, equals $+\infty$. Select some a_0 in A. Then $\sup(A+B) \geq \sup(a_0 + B) = a_0 + \sup B = +\infty$, so (1) holds in this case. Otherwise $\sup A + \sup B$ is finite. Consider $\epsilon > 0$. Then there exists $a \in A$ and $b \in B$, so that $a > \sup A - \frac{\epsilon}{2}$ and $b > \sup B - \frac{\epsilon}{2}$. Then $a + b \in A + B$ and so $\sup(A+B) \geq a + b > \sup A + \sup B - \epsilon$. Since $\epsilon > 0$ is arbitrary, we conclude $\sup(A+B) \geq \sup A + \sup B$. □

The exercises for this section clear up some loose ends. Most of them extend results in §4 to sets that are not necessarily bounded.

Exercises

5.1 Write the following sets in interval notation:
(a) $\{x \in \mathbb{R} : x < 0\}$ **(b)** $\{x \in \mathbb{R} : x^3 \leq 8\}$
(c) $\{x^2 : x \in \mathbb{R}\}$ **(d)** $\{x \in \mathbb{R} : x^2 < 8\}$

5.2 Give the infimum and supremum of each set listed in Exercise 5.1.

5.3 Give the infimum and supremum of each unbounded set listed in Exercise 4.1.

5.4 Let S be a nonempty subset of $\mathbb{R}$, and let $-S = \{-s : s \in S\}$. Prove $\inf S = -\sup(-S)$. *Hint*: For the case $-\infty < \inf S$, simply state that this was proved in Exercise 4.9.

5.5 Prove $\inf S \leq \sup S$ for every nonempty subset of $\mathbb{R}$. Compare Exercise 4.6(a).

5.6 Let S and T be nonempty subsets of $\mathbb{R}$ such that $S \subseteq T$. Prove $\inf T \leq \inf S \leq \sup S \leq \sup T$. Compare Exercise 4.7(a).

5.7 Finish Example 1 by verifying the equality involving infimums.

§6 * A Development of $\mathbb{R}$

There are several ways to give a careful development of $\mathbb{R}$ based on $\mathbb{Q}$. We will briefly discuss one of them and give suggestions for further reading on this topic. [See the remarks about enrichment sections in the preface.]

To motivate our development we begin by observing

$$a = \sup\{r \in \mathbb{Q} : r < a\} \quad \text{for each} \quad a \in \mathbb{R};$$

see Exercise 4.16. Note the intimate relationship: $a \leq b$ if and only if $\{r \in \mathbb{Q} : r < a\} \subseteq \{r \in \mathbb{Q} : r < b\}$ and, moreover, $a = b$ if and only if $\{r \in \mathbb{Q} : r < a\} = \{r \in \mathbb{Q} : r < b\}$. Subsets α of $\mathbb{Q}$ having the form $\{r \in \mathbb{Q} : r < a\}$ satisfy these properties:

(i) $\alpha \neq \mathbb{Q}$ and α is not empty,
(ii) If $r \in \alpha$, $s \in \mathbb{Q}$ and $s < r$, then $s \in \alpha$,
(iii) α contains no largest rational.

Moreover, every subset α of $\mathbb{Q}$ that satisfies (i)–(iii) has the form $\{r \in \mathbb{Q} : r < a\}$ for some $a \in \mathbb{R}$; in fact, $a = \sup \alpha$. Subsets α of $\mathbb{Q}$ satisfying (i)–(iii) are called *Dedekind cuts*.

The remarks in the last paragraph relating real numbers and Dedekind cuts are based on our knowledge of $\mathbb{R}$, including the completeness axiom. But they can also motivate a development of $\mathbb{R}$ based solely on $\mathbb{Q}$. In such a development we make no a priori assumptions about $\mathbb{R}$. We assume only that we have the ordered field $\mathbb{Q}$ and that $\mathbb{Q}$ satisfies the Archimedean property 4.6. A Dedekind cut is a subset α of $\mathbb{Q}$ satisfying (i)–(iii). The set $\mathbb{R}$ of real numbers is *defined* as the space of all Dedekind cuts. Thus elements of $\mathbb{R}$ are *defined* as certain subsets of $\mathbb{Q}$. The rational numbers are identified with certain Dedekind cuts in the natural way: each rational s corresponds to the Dedekind cut $s^* = \{r \in \mathbb{Q} : r < s\}$. In this way $\mathbb{Q}$ is regarded as a subset of $\mathbb{R}$, that is, $\mathbb{Q}$ is identified with the set $\mathbb{Q}^* = \{s^* : s \in \mathbb{Q}\}$.

The set $\mathbb{R}$ defined in the last paragraph is given an order structure as follows: if α and β are Dedekind cuts, then we define $\alpha \leq \beta$ to signify $\alpha \subseteq \beta$. Properties O1, O2 and O3 in §3 hold for this ordering. Addition is defined in $\mathbb{R}$ as follows: if α and β are Dedekind cuts, then

$$\alpha + \beta = \{r_1 + r_2 : r_1 \in \alpha \text{ and } r_2 \in \beta\}.$$

It turns out that $\alpha + \beta$ is a Dedekind cut [hence in $\mathbb{R}$] and this definition of addition satisfies properties A1–A4 in §3. Multiplication of Dedekind cuts is a tedious business and has to be defined first for Dedekind cuts $\geq 0^*$. For a naive attempt, see Exercise 6.4. After the product of Dedekind cuts has been defined, the remaining properties of an ordered field can be verified for $\mathbb{R}$. The ordered field $\mathbb{R}$ constructed in this manner from $\mathbb{Q}$ is complete: the completeness property in 4.4 can be *proved* rather than taken as an axiom.

The real numbers are developed from Cauchy sequences in $\mathbb{Q}$ in [31, §5]. A thorough development of $\mathbb{R}$ based on Peano's axioms is given in [39].

Exercises

6.1 Consider $s, t \in \mathbb{Q}$. Show

(a) $s \leq t$ if and only if $s^* \subseteq t^*$;

(b) $s = t$ if and only if $s^* = t^*$;

(c) $(s + t)^* = s^* + t^*$. Note that $s^* + t^*$ is a sum of Dedekind cuts.

6.2 Show that if α and β are Dedekind cuts, then so is $\alpha + \beta = \{r_1 + r_2 : r_1 \in \alpha \text{ and } r_2 \in \beta\}$.

6.3 **(a)** Show $\alpha + 0^* = \alpha$ for all Dedekind cuts α.

(b) We claimed, without proof, that addition of Dedekind cuts satisfies property A4. Thus if α is a Dedekind cut, there is a Dedekind cut $-\alpha$ such that $\alpha + (-\alpha) = 0^*$. How would you define $-\alpha$?

6.4 Let α and β be Dedekind cuts and define the "product": $\alpha \cdot \beta = \{r_1 r_2 : r_1 \in \alpha \text{ and } r_2 \in \beta\}$.

(a) Calculate some "products" of Dedekind cuts using the Dedekind cuts 0^*, 1^* and $(-1)^*$.

(b) Discuss why this definition of "product" is totally unsatisfactory for defining multiplication in $\mathbb{R}$.

6.5 **(a)** Show $\{r \in \mathbb{Q} : r^3 < 2\}$ is a Dedekind cut, but $\{r \in \mathbb{Q} : r^2 < 2\}$ is not a Dedekind cut.

(b) Does the Dedekind cut $\{r \in \mathbb{Q} : r^3 < 2\}$ correspond to a rational number in $\mathbb{R}$?

(c) Show $0^* \cup \{r \in \mathbb{Q} : r \geq 0 \text{ and } r^2 < 2\}$ is a Dedekind cut. Does it correspond to a rational number in $\mathbb{R}$?