

A.4

Theory of the Real Numbers

A rigorous development of calculus is based on properties of the real numbers. Many results about functions, derivatives, and integrals would be false if stated for functions defined only on the rational numbers. In this appendix we briefly examine some basic concepts of the theory of the reals that hint at what might be learned in a deeper, more theoretical study of calculus.

Three types of properties make the real numbers what they are. These are the **algebraic**, **order**, and **completeness** properties. The algebraic properties involve addition and multiplication, subtraction and division. They apply to rational or complex numbers as well as to the reals.

The structure of numbers is built around a set with addition and multiplication operations. The following properties are required of addition and multiplication.

- A1** $a + (b + c) = (a + b) + c$ for all a, b, c .
- A2** $a + b = b + a$ for all a, b, c .
- A3** There is a number called “0” such that $a + 0 = a$ for all a .
- A4** For each number a , there is a b such that $a + b = 0$.
- M1** $a(bc) = (ab)c$ for all a, b, c .
- M2** $ab = ba$ for all a, b .
- M3** There is a number called “1” such that $a \cdot 1 = a$ for all a .
- M4** For each nonzero a , there is a b such that $ab = 1$.
- D** $a(b + c) = ab + bc$ for all a, b, c .

A1 and M1 are *associative laws*, A2 and M2 are *commutativity laws*, A3 and M3 are *identity laws*, and D is the *distributive law*. Sets that have these algebraic properties are examples of **fields**, and are studied in depth in the area of theoretical mathematics called abstract algebra.

The **order** properties allow us to compare the size of any two numbers. The order properties are

- O1** For any a and b , either $a \leq b$ or $b \leq a$ or both.
- 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 $ac \leq bc$.

O3 is the *transitivity law*, and O4 and O5 relate ordering to addition and multiplication.

We can order the reals, the integers, and the rational numbers, but we cannot order the complex numbers (see Appendix A.5). There is no reasonable way to decide whether a number like $i = \sqrt{-1}$ is bigger or smaller than zero. A field in which the size of any two elements can be compared as above is called an **ordered field**. Both the rational numbers and the real numbers are ordered fields, and there are many others.

We can think of real numbers geometrically, lining them up as points on a line. The **completeness property** says that the real numbers correspond to all points on the line, with no “holes” or “gaps.” The rationals, in contrast, omit points such as $\sqrt{2}$ and π , and the integers even leave out fractions like $1/2$. The reals, having the completeness property, omit no points.

What exactly do we mean by this vague idea of missing holes? To answer this we must give a more precise description of completeness. A number M is an **upper bound** for a set of numbers if all numbers in the set are smaller than or equal to M . M is a **least upper bound** if it is the smallest upper bound. For example, $M = 2$ is an upper bound for the negative numbers. So is $M = 1$, showing that 2 is not a least upper bound. The least upper bound for the set of negative numbers is $M = 0$. We define a **complete** ordered field to be one in which every nonempty set bounded above has a least upper bound.

If we work with just the rational numbers, the set of numbers less than $\sqrt{2}$ is bounded, but it does not have a rational least upper bound, since any rational upper bound M can be replaced by a slightly smaller rational number that is still larger than $\sqrt{2}$. So the rationals are not complete. In the real numbers, a set that is bounded above always has a least upper bound. The reals are a complete ordered field.

The completeness property is at the heart of many results in calculus. One example occurs when searching for a maximum value for a function on a closed interval $[a, b]$, as in Section 4.1. The function $y = x - x^3$ has a maximum value on $[0, 1]$ at the point x satisfying $1 - 3x^2 = 0$, or $x = \sqrt{1/3}$. If we limited our consideration to functions defined only on rational numbers, we would have to conclude that the function has no maximum, since $\sqrt{1/3}$ is irrational (Figure A.2). The Extreme Value Theorem (Section 4.1), which implies that continuous functions on closed intervals $[a, b]$ have a maximum value, is not true for functions defined only on the rationals.

The Intermediate Value Theorem implies that a continuous function f on an interval $[a, b]$ with $f(a) < 0$ and $f(b) > 0$ must be zero somewhere in $[a, b]$. The function values cannot jump from negative to positive without there being some point x in $[a, b]$ where $f(x) = 0$. The Intermediate Value Theorem also relies on the completeness of the real numbers and is false for continuous functions defined only on the rationals. The function $f(x) = 3x^2 - 1$ has $f(0) = -1$ and $f(1) = 2$, but if we consider f only on the rational numbers, it never equals zero. The only value of x for which $f(x) = 0$ is $x = \sqrt{1/3}$, an irrational number.

We have captured the desired properties of the reals by saying that the real numbers are a complete ordered field. But we’re not quite finished. Greek mathematicians in the school of Pythagoras tried to impose another property on the numbers of the real line, the condition that all numbers are ratios of integers. They learned that their effort was doomed when they discovered irrational numbers such as $\sqrt{2}$. How do we know that our efforts to specify the real numbers are not also flawed, for some unseen reason? The artist Escher drew optical illusions of spiral staircases that went up and up until they rejoined themselves at the bottom. An engineer trying to build such a staircase would find that no structure realized the plans the architect had drawn. Could it be that our design for the reals contains some subtle contradiction, and that no construction of such a number system can be made?

We resolve this issue by giving a specific description of the real numbers and verifying that the algebraic, order, and completeness properties are satisfied in this model. This

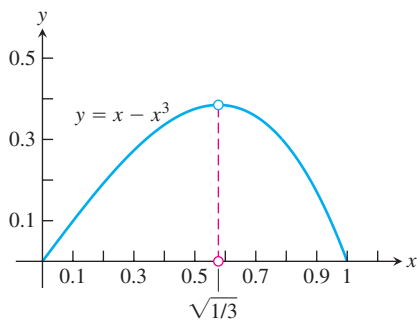


FIGURE A.2 The maximum value of $y = x - x^3$ on $[0, 1]$ occurs at the irrational number $x = \sqrt{1/3}$.

is called a **construction** of the reals, and just as stairs can be built with wood, stone, or steel, there are several approaches to constructing the reals. One construction treats the reals as all the infinite decimals,

$$a.d_1d_2d_3d_4\dots$$

In this approach a real number is an integer a followed by a sequence of decimal digits d_1, d_2, d_3, \dots , each between 0 and 9. This sequence may stop, or repeat in a periodic pattern, or keep going forever with no pattern. In this form, 2.00, 0.3333333... and 3.1415926535898... represent three familiar real numbers. The real meaning of the dots “...” following these digits requires development of the theory of sequences and series, as in Chapter 11. Each real number is constructed as the limit of a sequence of rational numbers given by its finite decimal approximations. An infinite decimal is then the same as a series

$$a + \frac{d_1}{10} + \frac{d_2}{100} + \dots$$

This decimal construction of the real numbers is not entirely straightforward. It's easy enough to check that it gives numbers that satisfy the completeness and order properties, but verifying the algebraic properties is rather involved. Even adding or multiplying two numbers requires an infinite number of operations. Making sense of division requires a careful argument involving limits of rational approximations to infinite decimals.

A different approach was taken by Richard Dedekind (1831–1916), a German mathematician, who gave the first rigorous construction of the real numbers in 1872. Given any real number x , we can divide the rational numbers into two sets: those less than or equal to x and those greater. Dedekind cleverly reversed this reasoning and defined a real number to be a division of the rational numbers into two such sets. This seems like a strange approach, but such indirect methods of constructing new structures from old are common in theoretical mathematics.

These and other approaches (see Appendix A.5) can be used to construct a system of numbers having the desired algebraic, order, and completeness properties. A final issue that arises is whether all the constructions give the same thing. Is it possible that different constructions result in different number systems satisfying all the required properties? If yes, which of these is the real numbers? Fortunately, the answer turns out to be no. The reals are the only number system satisfying the algebraic, order, and completeness properties.

Confusion about the nature of real numbers and about limits caused considerable controversy in the early development of calculus. Calculus pioneers such as Newton, Leibniz, and their successors, when looking at what happens to the difference quotient

$$\frac{\Delta y}{\Delta x} = \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

as each of Δy and Δx approach zero, talked about the resulting derivative being a quotient of two infinitely small quantities. These “infinitesimals,” written dx and dy , were thought to be some new kind of number, smaller than any fixed number but not zero. Similarly, a definite integral was thought of as a sum of an infinite number of infinitesimals

$$f(x) \cdot dx$$

as x varied over a closed interval. While the approximating difference quotients $\Delta y/\Delta x$ were understood much as today, it was the quotient of infinitesimal quantities, rather than

a limit, that was thought to encapsulate the meaning of the derivative. This way of thinking led to logical difficulties, as attempted definitions and manipulations of infinitesimals ran into contradictions and inconsistencies. The more concrete and computable difference quotients did not cause such trouble, but they were thought of merely as useful calculation tools. Difference quotients were used to work out the numerical value of the derivative and to derive general formulas for calculation, but were not considered to be at the heart of the question of what the derivative actually was. Today we realize that the logical problems associated with infinitesimals can be avoided by *defining* the derivative to be the limit of its approximating difference quotients. The ambiguities of the old approach are no longer present, and in the standard theory of calculus, infinitesimals are neither needed nor used.