Errett Bishop Douglas Bridges

Constructive Analysis

Springer-Verlag
Berlin Heidelberg NewYork Tokyo

Chapter 2. Calculus and the Real Numbers

Section 1 establishes some conventions about sets and functions. The next three sections are devoted to constructing the real numbers as certain Cauchy sequences of rational numbers, and investigating their order and arithmetic. The rest of the chapter deals with the basic ideas of the calculus of one variable. Topics covered include continuity, the convergence of sequences and series of continuous functions, differentiation, integration, Taylor's theorem, and the basic properties of the exponential and trigonometric functions and their inverses. Most of the material is a routine constructivization of the corresponding part of classical mathematics; for this reason it affords a good introduction to the constructive approach.

We assume that the reader is familiar with the order and arithmetic of the integers and the rational numbers. For us, a *rational number* will be an expression of the form p/q, where p and q are integers with $q \neq 0$. Two rational numbers p/q and p'/q' are *equal* if $pq' = p'q$. The integer n is identified with the rational number $n/1$.

There are geometric magnitudes which are not represented by rational numbers, and which can only be described by a sequence of rational approximations. Certain such approximating sequences are called real numbers. In this chapter we construct the real numbers and study their basic properties. Then we develop the fundamental ideas of the calculus.

1. Sets and Functions

Before constructing the real numbers, we introduce some notions which are basic to much of mathematics.

The totality of all mathematical objects constructed in accordance with certain requirements is called a *set*. The requirements of the construction, which vary with the set under consideration, determine the set. Thus the integers form a set, the rational numbers form a set, and (we anticipate here the formal definition of 'sequence') the collection of all sequences of integers is a set.

Each set will be endowed with a binary relation $=$ of *equality*. This relation is a matter of convention, except that it must be an *equivalence relation*; in other words, the following conditions must hold for all objects x, y, and z in the set:

(1.1) (i) $x=x$

(ii) If $x=y$, then $y=x$

(iii) If $x=y$ and $y=z$, then $x=z$.

The relation of equality given above for rational numbers is an equivalence relation. In this example there is a finite, mechanical procedure for deciding whether or not two given objects in the set are equal. Such a procedure will not exist in general: there are instances in which we are unable to decide whether or not two given elements of a set are equal; such an instance, in the theory of real numbers, will be given later.

We use the standard notation $a \in A$ to denote that a is an *element*, or *member*, of the set A, or that the construction defining a satisfies the requirements a construction must satisfy in order to define an object of A. We also use the notation $\{a_1, a_2, \ldots\}$ for a set whose elements can be written in a (possibly finite) list.

The dependence of one quantity on another is expressed by the basic notion of an operation. An *operation* from a set A into a set B is a finite routine f which assigns an element $f(a)$ of B to each given element a of A. This routine must afford an explicit, finite, mechanical reduction of the procedure for constructing $f(a)$ to the procedure for constructing a. If it is clear from the context what the sets A and B are, we sometimes denote f by $a \mapsto f(a)$, in order to bring out the form of $f(a)$ for a given element a of A. The set A is called the *domain* of the operation, and is denoted by $\mathrm{dmn} f$. In the most important case, we have $f(a)=f(a')$ whenever $a, a' \in A$ and $a=a'$; the operation f is then called a *function*, or a *mapping* of A into B, or a *map* of A into B. For two functions f, g from A into B, $f=g$ means that $f(a)=g(a)$ for each element a of A. Taken with this equality relation, the collection of all functions from A into B becomes a set.

The notation $f: A \to B$ indicates that f is a function from the set A to the set B.

A function x whose domain is the set $\mathbb{Z}^+$ of positive integers is called a *sequence*. The object $x_n \equiv x(n)$ is called the n^{th} *term* of the sequence. The finite routine x can be given explicitly, or it can be left to inference: for example, by writing the terms of the sequence in order

$$(x_1, x_2, \ldots)$$

until the rule of their formation becomes clear. Different notations for the sequence whose n^{th} term is x_n are: $n \mapsto x_n$, $(x_1, x_2, \ldots)$, $(x_n)_{n=1}^{\infty}$, and (x_n). Thus the sequence whose n^{th} term is n^2 can be written $n \mapsto n^2$, or $(1, 4, 9, \ldots)$, or $(n^2)_{n=1}^{\infty}$, or simply (n^2).

A *subsequence* of a sequence (x_n) consists of the sequence (x_n) and a sequence $(n_k)_{k=1}^{\infty}$ of positive integers such that $n_1 < n_2 < \ldots$. We identify such a subsequence with the sequence whose k^{th} term is x_{n_k}.

Sometimes we shall speak of sequences whose domain is some set of integers other than $\mathbb{Z}^+$. For example, we shall write $(x_n)_{n=0}^{\infty}$ to denote a mapping x from the set of nonnegative integers, where $x_n \equiv x(n)$ for each n.

Another example arises as follows. If n is a positive integer, then a *finite sequence* of *length* n is a function from the set $\{1, 2, \ldots, n\}$ into a set B.

The *cartesian product*, or simply the *product*, of sets $X_1, \ldots, X_n$ is defined to be the set

$$X \equiv X_1 \times X_2 \times \ldots \times X_n$$

of all ordered n-tuples $(x_1, \ldots, x_n)$ with $x_1 \in X_1$, $x_2 \in X_2, \ldots$, and $x_n \in X_n$. Elements $(x_1, \ldots, x_n)$ and $(x'_1, \ldots, x'_n)$ of the cartesian product are *equal* if the *coordinates* (or *components*) x_k and x'_k are equal elements of X_k for each k.

If x is a finite sequence of elements of a set B, then x can be identified with the element $(x(1), \ldots, x(n))$ of the cartesian product

$$B^n \equiv B \times B \times \ldots \times B,$$

where n is the length of x.

Returning to functions in general, we say that a function $f\colon A \to B$ maps A *onto* B if to each element b of B there corresponds an element a of A with $f(a) = b$. In other words, f maps A onto B if there is an operation g from B into A such that $f(g(b)) = b$ for each b in B. A set A is *countable* if there exists a mapping of $\mathbb{Z}^+$ onto A; intuitively, this means that the elements of A can be arranged in a sequence with possible duplications.

The elements of the cartesian product $\mathbb{Z} \times \mathbb{Z}$ of the set $\mathbb{Z}$ of integers with itself can be arranged in a sequence as follows. We order the elements (m, n) of $\mathbb{Z} \times \mathbb{Z}$, first according to the value of $|m| + |n|$, then according to the value of m, and finally according to the value of

n. This produces the sequence

(1.2) $((0,0),(-1,0),(0,-1),(0,1),(1,0),(-2,0),(-1,-1),\ldots)$,

in which each element of $\mathbb{Z}\times\mathbb{Z}$ occurs exactly once. In the sequence (1.2), omit every term (m,n) with $n=0$, and replace each term (m,n) with $n\neq 0$ by m/n; this produces the sequence

(1.3) $$(0/-1, 0/1, -1/-1, \ldots),$$

in which every expression p/q, with p and q integers and $q\neq 0$, occurs exactly once. Keeping only the term $0/1$ of (1.3), and those terms for which $q>0$, $p\neq 0$, and p is relatively prime to q, we obtain a sequence

(1.4) $$(0/1, -1/1, 1/1, \ldots)$$

which has the property that for any given rational number r there exists exactly one term equal to r.

For each positive integer n, let $\mathbb{Z}_n$ be the set $\{0, 1, \ldots, n-1\}$. If there is a mapping of $\mathbb{Z}_n$ onto the set A, then we say that A has *at most n elements*. A set with at most n elements for some n is said to be *subfinite*, or *finitely enumerable*. Note that every subfinite or countable set has at least one element.

Before we introduce stronger notions than countability and subfiniteness, we must discuss the composition of functions. The *composition* of two functions $f\colon A\to B$ and $g\colon B\to C$ is the function $g\circ f\colon A\to C$ defined by

$$(g\circ f)(a)\equiv g(f(a)) \qquad (a\in A).$$

Composition is associative:

$$h\circ(g\circ f)=(h\circ g)\circ f$$

whenever the compositions are defined.

If $f\colon A\to B$, $g\colon B\to A$, and $g(f(a))=a$ for all a in A, then the function g is called a *left inverse* of f, and the function f is called a *right inverse* of g. (Note that f has a left inverse if and only if it is *one-one*, in the sense that $a=a'$ for all elements a, a' of A with $f(a)=f(a')$.) When g is both a left and a right inverse of f, then it is simply called an *inverse* of f; f is then called a *one-one correspondence*, or a *bijection*, and the sets A and B are said to be *in one-one correspondence* with each other.

A set which is in one-one correspondence with the set $\mathbb{Z}^+$ of positive integers is said to be *countably infinite*. For example, let f be the sequence (1.4), and define a function g from the set $\mathbb{Q}$ of rational numbers to $\mathbb{Z}^+$ by writing $g(r)\equiv n$, where n is the unique positive integer for which $f(n)=r$. Then g is an inverse of f; so that the set $\mathbb{Q}$

is countably infinite. A similar proof using (1.2) shows that $\mathbb{Z} \times \mathbb{Z}$ is countably infinite.

A set which is in one-one correspondence with $\mathbb{Z}_n$ is said to *have n elements*, and to be *finite*. Every finite set is countable.

It is not true that every countable set is either countably infinite or subfinite. For example, let A consist of all positive integers n such that both n and $n+2$ are prime; then A is countable, but we do not know if it is either countably infinite or subfinite. This does not rule out the possibility that at some time in the future A will have become countably infinite or subfinite; it is possible that tomorrow someone will show that A is subfinite. This set A has the property that if it is subfinite, then it is finite. Not all sets have this property.

2. The Real Number System

The following definition is basic to everything that follows.

(2.1) **Definition.** A sequence (x_n) of rational numbers is *regular* if

(2.1.1) $$|x_m - x_n| \leqq m^{-1} + n^{-1} \quad (m, n \in \mathbb{Z}^+).$$

A *real number* is a regular sequence of rational numbers. Two real numbers $x \equiv (x_n)$ and $y \equiv (y_n)$ are *equal* if

(2.1.2) $$|x_n - y_n| \leqq 2n^{-1} \quad (n \in \mathbb{Z}^+).$$

The set of real numbers is denoted by $\mathbb{R}$.

(2.2) **Proposition.** *Equality of real numbers is an equivalence relation.*

Proof: Parts (i) and (ii) of (1.1) are obvious. Part (iii) is a consequence of the following lemma.

(2.3) **Lemma.** *The real numbers $x \equiv (x_n)$ and $y \equiv (y_n)$ are equal if and only if for each positive integer j there exists a positive integer N_j such that*

(2.3.1) $$|x_n - y_n| \leqq j^{-1} \quad (n \geqq N_j).$$

Proof: If $x = y$, then (2.3.1) holds with $N_j \equiv 2j$.

Assume conversely that for each j in $\mathbb{Z}^+$ there exists N_j satisfying (2.3.1). Consider a positive integer n. If m and j are any positive integers with $m \geqq \max\{j, N_j\}$, then

$$\begin{aligned}|x_n - y_n| &\leqq |x_n - x_m| + |x_m - y_m| + |y_m - y_n| \\ &\leqq (n^{-1} + m^{-1}) + j^{-1} + (n^{-1} + m^{-1}) < 2n^{-1} + 3j^{-1}.\end{aligned}$$

Since this holds for all j in $\mathbb{Z}^+$, (2.1.2) is valid. □

Notice that the proof of Lemma (2.3) singles out a specific N_j satisfying (2.3.1). This situation is typical: every proof of a theorem which asserts the existence of an object must embody, at least implicitly, a finite routine for the construction of the object.

The rational number x_n is called the n^{th} *rational approximation* to the real number $x \equiv (x_n)$. Note that the operation from $\mathbb{R}$ to $\mathbb{Q}$ which takes the real number x into its n^{th} rational approximation is not a function.

For later use we wish to associate with each real number $x \equiv (x_n)$ an integer K_x such that

$$|x_n| < K_x \quad (n \in \mathbb{Z}^+).$$

This is done by letting K_x be the least integer which is greater than $|x_1| + 2$. We call K_x the *canonical bound* for x.

The development of the arithmetic of the real numbers offers no surprises: we operate with real numbers by operating with their rational approximations.

(2.4) **Definition.** Let $x \equiv (x_n)$ and $y \equiv (y_n)$ be real numbers with respective canonical bounds K_x and K_y. Write

$$k \equiv \max\{K_x, K_y\}.$$

Let α be any rational number. We define

(a) $x + y \equiv (x_{2n} + y_{2n})_{n=1}^{\infty}$

(b) $xy \equiv (x_{2kn} y_{2kn})_{n=1}^{\infty}$

(c) $\max\{x, y\} \equiv (\max\{x_n, y_n\})_{n=1}^{\infty}$

(d) $-x \equiv (-x_n)_{n=1}^{\infty}$

(e) $\alpha^* \equiv (\alpha, \alpha, \alpha, \ldots)$.

(2.5) **Proposition.** *The sequences* $x+y$, xy, $\max\{x, y\}$, $-x$, *and* α^* *of Definition* (2.4) *are real numbers.*

Proof: (a) Write $z_n \equiv x_{2n} + y_{2n}$. Then $x + y \equiv (z_n)$. For all positive integers m and n,

$$\begin{aligned} |z_m - z_n| &\leqq |x_{2m} - x_{2n}| + |y_{2m} - y_{2n}| \\ &\leqq (2n)^{-1} + (2m)^{-1} + (2n)^{-1} + (2m)^{-1} = n^{-1} + m^{-1}. \end{aligned}$$

Thus $x + y$ is a real number.

(b) Write $z_n \equiv x_{2kn} y_{2kn}$. Then $xy \equiv (z_n)$. For all positive integers m and n,

$$\begin{aligned} |z_m - z_n| &= |x_{2km}(y_{2km} - y_{2kn}) + y_{2kn}(x_{2km} - x_{2kn})| \\ &\leqq k|y_{2km} - y_{2kn}| + k|x_{2km} - x_{2kn}| \\ &\leqq k((2km)^{-1} + (2kn)^{-1} + (2km)^{-1} + (2kn)^{-1}) = n^{-1} + m^{-1}. \end{aligned}$$

Thus xy is a real number.

(c) Write $z_n \equiv \max\{x_n, y_n\}$. Then $\max\{x, y\} \equiv (z_n)$. Consider positive integers m and n. For simplicity assume that

$$x_m = \max\{x_m, x_n, y_m, y_n\}.$$

Then

$$\begin{aligned} |z_m - z_n| &= |x_m - \max\{x_n, y_n\}| \\ &= x_m - \max\{x_n, y_n\} \leqq x_m - x_n \leqq n^{-1} + m^{-1}. \end{aligned}$$

Thus $\max\{x, y\}$ is a real number.

(d) For all positive integers m and n,

$$|-x_m - (-x_n)| = |x_m - x_n| \leqq m^{-1} + n^{-1}.$$

Thus $-x$ is a real number.

(e) This is obvious. □

There is no trouble in proving that $(x, y) \mapsto x + y$, $(x, y) \mapsto xy$, and $(x, y) \mapsto \max\{x, y\}$ are functions from $\mathbb{R} \times \mathbb{R}$ to $\mathbb{R}$; that $x \mapsto -x$ is a function from $\mathbb{R}$ to $\mathbb{R}$; and that $\alpha \mapsto \alpha^*$ is a function from $\mathbb{Q}$ to $\mathbb{R}$.

The operation

$$x \mapsto |x| \equiv \max\{x, -x\}$$

is therefore a function from $\mathbb{R}$ to $\mathbb{R}$, and the operation

$$(x, y) \mapsto \min\{x, y\} \equiv -\max\{-x, -y\}$$

is a function from $\mathbb{R} \times \mathbb{R}$ to $\mathbb{R}$.

The next proposition states that the real numbers obey the same rules of arithmetic as the rational numbers.

(2.6) **Proposition.** *For arbitrary real numbers x, y, and z and rational numbers α and β,*

(a) $x + y = y + x$, $xy = yx$

(b) $(x+y)+z=x+(y+z),\ x(yz)=(xy)z$

(c) $x(y+z)=xy+xz$

(d) $0^*+x=x,\ 1^*x=x$

(e) $x-x=0^*$

(f) $|xy|=|x||y|$

(g) $(\alpha+\beta)^*=\alpha^*+\beta^*$, $(\alpha\beta)^*=\alpha^*\beta^*$, *and* $(-\alpha)^*=-\alpha^*$.

We omit the simple proofs of these results.

We shall use standard notations, such as $x+y+z$ and $\max\{x,y,z\}$, without further comment.

There are three basic relations defined on the set of real numbers. The first of these, the equality relation, has already been defined. The remaining relations, which pertain to order, are best introduced in terms of certain subsets $\mathbb{R}^+$ and $\mathbb{R}^{0+}$ of $\mathbb{R}$.

(2.7) **Definition.** A real number $x\equiv(x_n)$ is *positive*, or $x\in\mathbb{R}^+$, if

(2.7.1) $$x_n>n^{-1}$$

for some n in $\mathbb{Z}^+$. A real number $x\equiv(x_n)$ is *nonnegative*, or $x\in\mathbb{R}^{0+}$, if

(2.7.2) $$x_n\geqq -n^{-1}\qquad (n\in\mathbb{Z}^+).$$

The following criteria are often useful.

(2.8) **Lemma.** *A real number $x\equiv(x_n)$ is positive if and only if there exists a positive integer N such that*

(2.8.1) $$x_m\geqq N^{-1}\qquad (m\geqq N).$$

A real number $x\equiv(x_n)$ is nonnegative if and only if for each n in $\mathbb{Z}^+$ there exists N_n in $\mathbb{Z}^+$ such that

(2.8.2) $$x_m\geqq -n^{-1}\qquad (m\geqq N_n).$$

Proof: Assume that $x\in\mathbb{R}^+$. Then $x_n>n^{-1}$ for some n in $\mathbb{Z}^+$. Choose N in $\mathbb{Z}^+$ with

$$2N^{-1}\leqq x_n-n^{-1}.$$

Then

$$\begin{aligned}x_m&\geqq x_n-|x_m-x_n|\geqq x_n-m^{-1}-n^{-1}\\&\geqq x_n-n^{-1}-N^{-1}>N^{-1}\end{aligned}$$

whenever $m\geqq N$. Therefore (2.8.1) is valid.

Conversely, if (2.8.1) is valid, then (2.7.1) holds with $n=N+1$. Therefore $x\in\mathbb{R}^+$.

Assume next that $x \in \mathbb{R}^{0+}$. Then for each positive integer n,

$$x_m \geqq -m^{-1} \geqq -n^{-1} \qquad (m \geqq n).$$

Therefore (2.8.2) is valid with $N_n \equiv n$.

Assume finally that (2.8.2) holds. Then if k, m, and n are positive integers with $m \geqq N_n$, we have

$$x_k \geqq x_m - |x_m - x_k| \geqq -n^{-1} - k^{-1} - m^{-1}.$$

Since m and n are arbitrary, this gives $x_k \geqq -k^{-1}$. Therefore $x \in \mathbb{R}^{0+}$. □

As a corollary of Lemma (2.8), we see that if x and y are equal real numbers, then x is positive if and only if y is positive, and x is nonnegative if and only if y is nonnegative.

It is not strictly correct to say that a real number (x_n) *is* an element of $\mathbb{R}^+$. An element of $\mathbb{R}^+$ consists of a real number (x_n) *and* a positive integer n such that $x_n > n^{-1}$, because an element of $\mathbb{R}^+$ is not presented until both (x_n) and n are given. One and the same real number (x_n) can be associated with two distinct (but equal) elements of $\mathbb{R}^+$. Nevertheless we shall continue to refer loosely to a positive real number (x_n). On those occasions when we need to refer to an n for which $x_n > n^{-1}$, we shall take the position that it was there implicitly all along.

The proof of the following proposition is now easy, and will be left to the reader. For convenience, $\mathbb{R}^*$ represents either $\mathbb{R}^+$ or $\mathbb{R}^{0+}$.

(2.9) **Proposition.** *Let x and y be real numbers. Then*

(a) $x+y \in \mathbb{R}^*$ *and* $xy \in \mathbb{R}^*$ *whenever* $x \in \mathbb{R}^*$ *and* $y \in \mathbb{R}^*$

(b) $x+y \in \mathbb{R}^+$ *whenever* $x \in \mathbb{R}^+$ *and* $y \in \mathbb{R}^{0+}$

(c) $|x| \in \mathbb{R}^{0+}$

(d) $\max\{x, y\} \in \mathbb{R}^*$ *whenever* $x \in \mathbb{R}^*$

(e) $\min\{x, y\} \in \mathbb{R}^*$ *whenever* $x \in \mathbb{R}^*$ *and* $y \in \mathbb{R}^*$.

We now define the order relations on $\mathbb{R}$.

(2.10) **Definition.** Let x and y be real numbers. We define

$$x > y \text{ (or } y < x) \quad \text{if } x - y \in \mathbb{R}^+$$

and

$$x \geqq y \text{ (or } y \leqq x) \quad \text{if } x - y \in \mathbb{R}^{0+}.$$

A real number x is *negative* if $x < 0^*$ – that is, if $-x$ is positive.

Consider real numbers x, x', y, and y' such that (i) $x=x'$, $y=y'$, and $x>y$. We have

$$x'-y'=x-y\in\mathbb{R}^+$$

and therefore (ii) $x'>y'$. We express the fact that (ii) holds whenever (i) is valid by saying that $>$ is a *relation* on $\mathbb{R}$. More formally, a *relation* on a set X is a subset S of $X\times X$ such that if x, x', y, y' are elements of X with $x=x'$, $y=y'$, and $(x,y)\in S$, then $(x',y')\in S$.

We express the fact that $x>y$ if and only if $y<x$ by saying that $>$ and $<$ are *transposed relations*. Similarly, $\geqq$ and $\leqq$ are transposed relations.

If $x<y$ or $x=y$, then $x\leqq y$. The converse is not valid: as we shall see later, it is possible that we have $x\leqq y$ without being able to prove that $x<y$ or $x=y$. For this reason it was necessary to define the relations $<$ and $\leqq$ independently of each other.

The following rules for manipulating inequalities are easily proved from Proposition (2.9). We omit the proofs.

(2.11) **Proposition.** *For all real numbers x, y, z, and t,*

(a) *$x<z$ whenever either $x<y$ and $y\leqq z$ or $x\leqq y$ and $y<z$*

(b) *$x\leqq z$ whenever $x\leqq y$ and $y\leqq z$*

(c) *$x+y\leqq z+t$ whenever $x\leqq z$ and $y\leqq t$*

(d) *$x+y<z+t$ whenever $x\leqq z$ and $y<t$*

(e) *$xy\leqq zy$ whenever $x\leqq z$ and $y\geqq 0^*$*

(f) *$xy<zy$ whenever $x<z$ and $y>0^*$*

(g) *if $x<y$, then $-x>-y$*

(h) *if $x\leqq y$, then $-x\geqq -y$*

(i) $\max\{x,y\}\geqq x$

(j) $\min\{x,y\}\leqq x$

(k) *if $x\leqq y$ and $y\leqq x$, then $x=y$*

(l) $|x|\geqq 0^*$

(m) $|x+y|\leqq|x|+|y|$.

An important property of the relation $<$, of which we shall make no use, is the *antisymmetry* property, which states that at most one of the relations $x<y$ and $y<x$ is valid for given real numbers x and y. This negative statement has no place in the affirmative mathematics we are trying to develop, except as motivation. Its place is taken by the affirmative statement (k) of Proposition (2.11). As a general principle, negative statements are only for counterexamples and motivation; they are not to be used in subsequent work.

(2.12) **Definition.** For real numbers x and y we write $x \neq y$ if and only if $x<y$ or $x>y$.

Inequality $\neq$ is a relation because both $<$ and $>$ are relations. As motivation we have the negative statement that at most one of the relations $x \neq y$, $x=y$ can hold for given real numbers x and y. In other words, at most one of the relations $x<y$, $x>y$, $x=y$ can hold. This is clear from the definitions.

The following proposition defines the *inverse* x^{-1} of a real number $x \neq 0^*$, and derives the basic properties of the operation $x \mapsto x^{-1}$.

(2.13) **Proposition.** *Let x be a nonzero real number (so that $|x| \in \mathbb{R}^+$). There exists a positive integer N with $|x_m| \geqq N^{-1}$ for $m \geqq N$. Define*

$$y_n \equiv (x_{N^3})^{-1} \qquad (n<N)$$

and

$$y_n \equiv (x_{nN^2})^{-1} \qquad (n \geqq N).$$

Then

$$x^{-1} \equiv (y_n)_{n=1}^{\infty}$$

is a real number which is positive if x is positive, and negative if x is negative; also $xx^{-1}=1^*$.

If t is any real number for which $xt=1^*$, then $t=x^{-1}$. The operation $x \mapsto x^{-1}$ is a function. If $x \neq 0$ and $y \neq 0$, then $(xy)^{-1}=x^{-1}y^{-1}$. If $\alpha \neq 0$ is rational, then $(\alpha^*)^{-1}=(\alpha^{-1})^*$. If $x \neq 0$, then $(x^{-1})^{-1}=x$.

Proof: Our definitions guarantee that $|y_n| \leqq N$ for all n.

Consider positive integers m and n. Write

$$j \equiv \max\{m, N\}, \qquad k \equiv \max\{n, N\}.$$

Then

$$\begin{aligned} |y_m - y_n| &= |y_m||y_n||x_{jN^2} - x_{kN^2}| \\ &\leqq N^2((jN^2)^{-1} + (kN^2)^{-1}) = j^{-1} + k^{-1} \leqq m^{-1} + n^{-1}. \end{aligned}$$

Therefore x^{-1} is a real number.

Assume now that $x>0^*$. Then by (2.8), $x_n>0$ for all sufficiently large n. Hence $y_n>K_x^{-1}$ (where K_x is the canonical bound for x) for all sufficiently large n. It follows from (2.8) that $x^{-1}>0^*$. A similar proof shows that $x^{-1}<0^*$ whenever $x<0^*$.

Let k be the maximum of the canonical bounds for x and x^{-1}. Write $xx^{-1} \equiv (z_n)$. Then

$$z_n \equiv x_{2nk} y_{2nk} \equiv x_{2nk}(x_{2nN^2k})^{-1} \qquad (n \geqq N).$$

Therefore

$$\begin{aligned} |z_n - 1^*| &= |x_{2nN^2k}|^{-1} |x_{2nk} - x_{2nN^2k}| \\ &\leqq |y_{2nk}|((2nk)^{-1} + (2nN^2k)^{-1}) \leqq n^{-1} \end{aligned}$$

for $n \geqq N$. It follows that $xx^{-1}=1^*$.

If t is any real number with $xt=1^*$, then

$$x^{-1}=x^{-1}(xt)=(x^{-1}x)t=(xx^{-1})t=t.$$

If $x=x'$, then

$$x'x^{-1}=xx^{-1}=1^*.$$

Therefore $x^{-1}=(x')^{-1}$. It follows that $x\mapsto x^{-1}$ is a function.

If $x\neq 0$ and $y\neq 0$, then

$$(xy)x^{-1}y^{-1}=xx^{-1}yy^{-1}=1^*.$$

Therefore $x^{-1}y^{-1}=(xy)^{-1}$.

If $\alpha\neq 0$ is rational, then $\alpha^*\equiv(\alpha,\alpha,\ldots)$. Therefore

$$(\alpha^*)^{-1}=(\alpha^{-1},\alpha^{-1},\ldots)=(\alpha^{-1})^*.$$

For each x in $\mathbb{R}^+$, x^{-1} is in $\mathbb{R}^+$, and thus $(x^{-1})^{-1}$ exists. Since $x^{-1}x=xx^{-1}=1^*$, it follows that $(x^{-1})^{-1}=x$. Similarly $(x^{-1})^{-1}=x$ if x is negative. Therefore $(x^{-1})^{-1}=x$ whenever $x\neq 0$. □

Of course, we often write x/y instead of xy^{-1} when x and y are real numbers with $y\neq 0$.

As the previous propositions show, $(\alpha\beta)^*=\alpha^*\beta^*$, $(\alpha+\beta)^*=\alpha^*+\beta^*$, $(-\alpha)^*=-\alpha^*$, $(|\alpha|)^*=|\alpha^*|$, and $(\alpha^{-1})^*=(\alpha^*)^{-1}$ for all rational numbers α and β. Also $\alpha\Delta\beta$ if and only if $\alpha^*\Delta\beta^*$, where Δ stands for any of the relations $=$, $<$, $>$, and $\neq$. This situation is expressed by saying that the map $\alpha\mapsto\alpha^*$ is an *order isomorphism* from $\mathbb{Q}$ into $\mathbb{R}$. This justifies identifying $\mathbb{Q}$ with a subset of $\mathbb{R}$, as we previously identified $\mathbb{Z}$ with a subset of $\mathbb{Q}$. Henceforth we make no distinction between a rational number α and the corresponding real number α^*.

The next lemma shows that the n^{th} rational approximation x_n to a real number $x\equiv(x_n)$ actually approximates x to within n^{-1}.

(2.14) **Lemma.** *For each real number $x\equiv(x_n)$, we have*

$$|x-x_n|\leqq n^{-1}\qquad(n\in\mathbb{Z}^+).$$

Proof: By (2.4) and the definition of $|\ |$, the m^{th} rational approximation to $n^{-1}-|x-x_n|$ is

$$n^{-1}-|x_{4m}-x_n|\geqq n^{-1}-((4m)^{-1}+n^{-1})=-(4m)^{-1}>-m^{-1}.$$

By (2.7), we have $n^{-1}-|x-x_n|\in\mathbb{R}^{0+}$. Therefore $|x-x_n|\leqq n^{-1}$. □

(2.15) **Lemma.** *If $x\equiv(x_n)$ and $y\equiv(y_n)$ are real numbers with $x<y$, then there exists a rational number α with $x<\alpha<y$.*

Proof: By (2.4), we have $y-x\equiv(y_{2n}-x_{2n})_{n=1}^{\infty}$. Since $y-x\in\mathbb{R}^+$, by (2.7) there exists n in $\mathbb{Z}^+$ with $y_{2n}-x_{2n}>n^{-1}$. Write

$$\alpha\equiv\tfrac{1}{2}(x_{2n}+y_{2n}).$$

Then

$$\alpha-x\geqq\alpha-x_{2n}-|x_{2n}-x|\geqq\tfrac{1}{2}(y_{2n}-x_{2n})-(2n)^{-1}>0.$$

Also,

$$y-\alpha\geqq y_{2n}-\alpha-|y_{2n}-y|\geqq\tfrac{1}{2}(y_{2n}-x_{2n})-(2n)^{-1}>0.$$

Therefore $x<\alpha<y$. □

As a corollary, for each x in $\mathbb{R}$ and r in $\mathbb{R}^+$ there exists α in $\mathbb{Q}$ with $|x-\alpha|<r$. Here is another corollary.

(2.16) **Proposition.** *If $x_1,\dots,x_n$ are real numbers with $x_1+\dots+x_n>0$, then $x_i>0$ for some i $(1\leqq i\leqq n)$.*

Proof: By (2.15), there exists a rational number α with $0<\alpha<x_1+\dots+x_n$. For $1\leqq i\leqq n$ let a_i be a rational number with

$$|x_i-a_i|<(2n)^{-1}\alpha.$$

Then

$$\sum_{i=1}^{n}a_i\geqq\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}|x_i-a_i|>\tfrac{1}{2}\alpha.$$

Therefore $a_i>(2n)^{-1}\alpha$ for some i. For this i it follows that

$$x_i\geqq a_i-|x_i-a_i|>0. \quad \square$$

(2.17) **Corollary.** *If x, y, and z are real numbers with $y<z$, then either $x<z$ or $x>y$.*

Proof: Since $z-x+x-y=z-y>0$, either $z-x>0$ or $x-y>0$, by (2.16). □

The next lemma gives an extremely useful method for proving inequalities of the form $x\leqq y$.

(2.18) **Lemma.** *Let x and y be real numbers such that the assumption $x>y$ implies that $0=1$. Then $x\leqq y$.*

Proof: Without loss of generality, we take $y=0$. For each n in $\mathbb{Z}^+$, either $x_n\leqq n^{-1}$ or $x_n>n^{-1}$. The case $x_n>n^{-1}$ is ruled out, since it implies that $x>0$. Therefore $-x_n\geqq-n^{-1}$ for all n, and so $-x\geqq 0$. Thus $x\leqq 0$. □

(2.19) **Theorem.** *Let (a_n) be a sequence of real numbers, and let x_0 and y_0 be real numbers with $x_0<y_0$. Then there exists a real number x such that $x_0 \leqq x \leqq y_0$ and $x \neq a_n$ for all n in $\mathbb{Z}^+$.*

Proof: We construct by induction sequences (x_n) and (y_n) of rational numbers such that

(i) $x_0 \leqq x_n \leqq x_m < y_m \leqq y_n \leqq y_0 \quad (m \geqq n \geqq 1)$

(ii) $x_n > a_n$ or $y_n < a_n \quad (n \geqq 1)$

(iii) $y_n - x_n < n^{-1} \quad (n \geqq 1)$.

Assume that $n \geqq 1$ and that $x_0, \ldots, x_{n-1}$, $y_0, \ldots, y_{n-1}$ have been constructed. Either $a_n > x_{n-1}$ or $a_n < y_{n-1}$. In case $a_n > x_{n-1}$, let x_n be any rational number with $x_{n-1} < x_n < \min\{a_n, y_{n-1}\}$, and let y_n be any rational number with $x_n < y_n < \min\{a_n, y_{n-1}, x_n + n^{-1}\}$. Then the relevant inequalities are satisfied. In case $a_n < y_{n-1}$, let y_n be any rational number with $\max\{a_n, x_{n-1}\} < y_n < y_{n-1}$, and x_n any rational number with $\max\{a_n, x_{n-1}, y_n - n^{-1}\} < x_n < y_n$. Again, the relevant inequalities are satisfied. This completes the induction.

From (i) and (iii) it follows that

$$|x_m - x_n| = x_m - x_n < y_n - x_n < n^{-1} \qquad (m \geqq n).$$

Similarly $|y_m - y_n| < n^{-1}$ for $m \geqq n$. Therefore $x \equiv (x_n)$ and $y \equiv (y_n)$ are real numbers. By (i) and (iii), they are equal. By (i), $x_n \leqq x$ and $y_n \geqq y$ for all n. If $a_n < x_n$, then $a_n < x$ and so $a_n \neq x$. If $a_n > y_n$, then $a_n > y = x$ and so $a_n \neq x$. Thus x has the required properties. $\square$

Theorem (2.19) is the famous theorem of Cantor, that the real numbers are uncountable. The proof is essentially Cantor's "diagonal" proof. Both Cantor's theorem and his method of proof are of great importance.

The time has come to consider some counterexamples. Let (n_k) be a sequence of integers, each of which is either 0 or 1, for which we are unable to prove either that $n_k = 1$ for some k or that $n_k = 0$ for all k. This corresponds to what Brouwer calls "a fugitive property of the natural numbers". For example, such a sequence can be defined as follows. Let n_k be 0 if $u^t + v^t \neq w^t$ for all integers u, v, w, t with $0 < u$, v, $w \leqq k$ and $3 \leqq t \leqq 2 + k$. Otherwise let n_k be 1. Then we are unable to prove $n_k = 1$ for some k, because this would disprove Fermat's last theorem. We are unable to prove $n_k = 0$ for all k, because this would prove Fermat's last theorem.

Now define $x_k \equiv 0$ if $n_j = 0$ for all $j \leqq k$, and $x_k \equiv 2^{-m}$ otherwise, where m is the least positive integer such that $n_m = 1$. Then $x \equiv (x_k)$ is a nonnegative real number, but we are unable to prove that $x>0$ or $x=0$. Since nothing is true unless and until it has been proved, it is untrue that $x>0$ or $x=0$.

Of course, if Fermat's last theorem is proved tomorrow, we shall probably still be able to define a fugitive sequence (n_k) of integers. Thus it is unlikely that there will ever exist a constructive proof that for every real number $x \geqq 0$ either $x>0$ or $x=0$. We express this fact by saying that there exists a real number $x \geqq 0$ such that it is *not* true that $x>0$ or $x=0$.

In much the same way we can construct a real number x such that it is *not* true that $x \geqq 0$ or $x \leqq 0$.

3. Sequences and Series of Real Numbers

We develop methods for defining a real number in terms of approximations by other real numbers.

(3.1) **Definition.** A sequence (x_n) of real numbers *converges* to a real number x_0 if for each k in $\mathbb{Z}^+$ there exists N_k in $\mathbb{Z}^+$ with

(3.1.1) $$|x_n - x_0| \leqq k^{-1} \quad (n \geqq N_k).$$

The real number x_0 is then called a *limit* of the sequence (x_n). To express the fact that (x_n) converges to x_0 we write

$$\lim_{n\to\infty} x_n = x_0$$

or

$$x_n \to x_0 \quad \text{as } n \to \infty$$

or simply $x_n \to x_0$.

A sequence (x_n) of real numbers is said to *converge*, or be *convergent*, if there exists a limit x_0 of (x_n).

It is easily seen that if (x_n) converges to both x_0 and x_0', then $x_0 = x_0'$.

A convergent sequence is *bounded:* there exists r in $\mathbb{R}^+$ such that $|x_n| \leqq r$ for all n.

A convergent sequence of real numbers is not determined until the limit x_0 and the sequence (N_k) are given, as well as the sequence (x_n) itself. Even when they are not mentioned explicitly, these quantities are implicitly present. Similar comments apply to many subsequent definitions, including the following.

(3.2) **Definition.** A sequence (x_n) of real numbers is a *Cauchy sequence* if for each k in $\mathbb{Z}^+$ there exists M_k in $\mathbb{Z}^+$ such that

(3.2.1) $$|x_m - x_n| \leqq k^{-1} \quad (m, n \geqq M_k).$$

(3.3) **Theorem.** *A sequence (x_n) of real numbers converges if and only if it is a Cauchy sequence.*

Proof: Assume that (x_n) converges to a real number x_0. Let the sequence (N_k) satisfy (3.1.1). Write $M_k \equiv N_{2k}$. Then

$$|x_m - x_n| \leqq |x_m - x_0| + |x_n - x_0| \leqq (2k)^{-1} + (2k)^{-1} = k^{-1}$$

for $m, n \geqq M_k$. Therefore (x_n) is a Cauchy sequence.

Assume conversely that (x_n) is a Cauchy sequence. Let the sequence (M_k) satisfy (3.2.1). Write $N_k \equiv \max\{3k, M_{2k}\}$. Then

$$|x_m - x_n| \leqq (2k)^{-1} \quad (m, n \geqq N_k).$$

Let y_k be the $(2k)^{\text{th}}$ rational approximation to x_{N_k}. For $m \geq n$,

$$\begin{aligned} |y_m - y_n| &\leq |y_m - x_{N_m}| + |x_{N_m} - x_{N_n}| + |x_{N_n} - y_n| \\ &\leq (2m)^{-1} + (2m)^{-1} + (2n)^{-1} + (2n)^{-1} = m^{-1} + n^{-1}. \end{aligned}$$

Therefore $y \equiv (y_n)$ is a real number. To see that (x_n) converges to y, we consider $n \geqq N_k$ and compute

$$\begin{aligned} |y - x_n| &\leqq |y - y_n| + |y_n - x_{N_n}| + |x_{N_n} - x_n| \\ &\leqq n^{-1} + (2n)^{-1} + (2k)^{-1} \leqq (3k)^{-1} + (6k)^{-1} + (2k)^{-1} = k^{-1}. \quad \square \end{aligned}$$

A subsequence of a convergent sequence converges to the same limit. If a sequence converges, then any sequence obtained from it by modifications (including, perhaps, insertions or deletions) which involve only a finite number of terms converges to the same limit.

If $x \equiv (x_n)$ is a regular sequence of rational numbers, then (x_n^*) converges to x, by (2.14).

A sequence (x_n) is *increasing* (respectively, *strictly increasing*) if $x_{n+1} \geqq x_n$ (respectively, $x_{n+1} > x_n$) for each n. *Decreasing* and *strictly decreasing* sequences are defined analogously, in the obvious way. A theorem of classical mathematics states that every bounded increasing sequence of real numbers converges. A counterexample to this statement is given by any increasing sequence (x_n) such that $x_n = 0$ or $x_n = 1$ for each n, but it is not known whether $x_n = 0$ for all n.

It is useful to supplement Definition (3.1) by writing

$$\lim_{n \to \infty} x_n = \infty$$

or

$$x_n \to \infty \quad \text{as} \quad n \to \infty$$

to express the fact that for each k in $\mathbb{Z}^+$ there exists N_k in $\mathbb{Z}^+$ with $x_n > k$ for all $n \geqq N_k$. We also define

$$\lim_{n\to\infty} x_n = -\infty$$

or

$$x_n \to -\infty \quad \text{as} \quad n\to\infty$$

to mean that $\lim\limits_{n\to\infty} -x_n = \infty$.

The next proposition shows that we may work with real numbers constructed as limits by working with their approximations.

(3.4) **Proposition.** *Assume that $x_n \to x_0$ as $n\to\infty$, and $y_n \to y_0$ as $n\to\infty$, where x_0 and y_0 are real numbers. Then*

(a) $x_n + y_n \to x_0 + y_0$ *as* $n\to\infty$

(b) $x_n y_n \to x_0 y_0$ *as* $n\to\infty$

(c) $\max\{x_n, y_n\} \to \max\{x_0, y_0\}$ *as* $n\to\infty$

(d) $x_0 = c$ *whenever* $x_n = c$ *for all* n

(e) *if* $x_0 \neq 0$ *and* $x_n \neq 0$ *for all* n, *then* $x_n^{-1} \to x_0^{-1}$ *as* $n\to\infty$

(f) *if* $x_n \leqq y_n$ *for all* n, *then* $x_0 \leqq y_0$.

Proof: (a) For each k in $\mathbb{Z}^+$ there exists N_k in $\mathbb{Z}^+$ such that

$$|x_n - x_0| \leq (2k)^{-1}, \quad |y_n - y_0| \leq (2k)^{-1} \qquad (n \geqq N_k).$$

Then

$$|x_n + y_n - (x_0 + y_0)| \leqq (2k)^{-1} + (2k)^{-1} = k^{-1} \qquad (n \geqq N_k).$$

Therefore $x_n + y_n \to x_0 + y_0$ as $n\to\infty$.

(b) Choose m in $\mathbb{Z}^+$ such that $|y_0| \leqq m$ and $|x_n| \leqq m$ for all n. For each k in $\mathbb{Z}^+$ choose N_k in $\mathbb{Z}^+$ with

$$|x_n - x_0| \leqq (2mk)^{-1}, \quad |y_n - y_0| \leqq (2mk)^{-1} \quad (n \geqq N_k).$$

Then for $n \geqq N_k$,

$$\begin{aligned} |x_n y_n - x_0 y_0| &\leqq |x_n(y_n - y_0)| + |y_0(x_n - x_0)| \\ &\leqq m(|y_n - y_0| + |x_n - x_0|) \leqq k^{-1}. \end{aligned}$$

Therefore $x_n y_n \to x_0 y_0$ as $n\to\infty$.

(c) Since

$$|\max\{x_n, y_n\} - \max\{x_0, y_0\}| \leqq \max\{|x_n - x_0|, |y_n - y_0|\},$$

it follows that

$$\max\{x_n, y_n\} \to \max\{x_0, y_0\} \quad \text{as } n\to\infty.$$

(d) If $x_n = c$ for all n, then (x_n) converges to c. Therefore $x_0 = c$.

(e) Since $|x_0| > 0$,

$$|x_n| \geqq |x_0| - |x_n - x_0| > \tfrac{1}{2}|x_0|$$

whenever n is large enough, say for $n \geqq n_0$. Let k and n be positive integers such that $n \geqq n_0$ and $|x_n - x_0| < (2k)^{-1}|x_0|^2$. Then

$$|x_n^{-1} - x_0^{-1}| = |x_n|^{-1}|x_0|^{-1}|x_n - x_0| \leqq 2|x_0|^{-2}(2k)^{-1}|x_0|^2 = k^{-1}.$$

Therefore $x_n^{-1} \to x_0^{-1}$ as $n \to \infty$.

(f) We compute

$$\begin{aligned} y_0 - x_0 &= \lim_{n\to\infty} y_n - \lim_{n\to\infty} x_n = \lim_{n\to\infty}(y_n - x_n) = \lim_{n\to\infty}|y_n - x_n| \\ &= \lim_{n\to\infty} \max\{y_n - x_n, x_n - y_n\} = \max\{y_0 - x_0, x_0 - y_0\} \geqq 0, \end{aligned}$$

by (a), (b), (c), and (d). □

For each sequence (x_n) of real numbers the number

$$s_n \equiv \sum_{k=1}^{n} x_k$$

is called the n^{th} *partial sum* of (x_n), and (s_n) is called the *sequence of partial sums* of the sequence (x_n). A *sum* s_0 of (x_n) is a limit of the sequence (s_n) of partial sums. We write

$$s_0 = \sum_{n=1}^{\infty} x_n$$

to indicate that s_0 is a sum of (x_n). A sequence which is meant to be summed is called a *series*. A series is said to *converge* to its sum. Thus the sequence $(2^{-n})_{n=1}^{\infty}$ converges to 0 as a sequence, but as a series it converges to $\sum_{n=1}^{\infty} 2^{-n} = 1$.

A convergent series remains convergent, but not necessarily to the same sum, after modification of finitely many of its terms.

The series (x_n) is often loosely referred to as *the series* $\sum_{n=1}^{\infty} x_n$.

If the series $\sum_{n=1}^{\infty} x_n$ converges, then $x_n \to 0$ as $n \to \infty$.

A series $\sum_{n=1}^{\infty} x_n$ is said to *converge absolutely* when the series $\sum_{n=1}^{\infty} |x_n|$ converges.

In classical analysis a series of nonnegative terms converges if the partial sums are bounded. This is not true in constructive analysis. However, we have the following result.

(3.5) **Proposition.** *If* $\sum_{n=1}^{\infty} y_n$ *is a convergent series of nonnegative terms, and if* $|x_n| \leqq y_n$ *for each* n, *then* $\sum_{n=1}^{\infty} x_n$ *converges.*

Proof: Since $\sum_{n=1}^{\infty} y_n$ is convergent, the sequence of partial sums is a Cauchy sequence. Therefore for each k in $\mathbb{Z}^+$ there exists an N_k in $\mathbb{Z}^+$ with

$$\sum_{j=n+1}^{m} y_j \leqq k^{-1} \qquad (m > n \geqq N_k).$$

Then

$$\left|\sum_{j=n+1}^{m} x_j\right| \leqq \sum_{j=n+1}^{m} y_j \leqq k^{-1} \qquad (m > n \geqq N_k).$$

Therefore the sequence of partial sums of the series $\sum_{n=1}^{\infty} x_n$ is a Cauchy sequence. By (3.3), the series converges. □

The criterion of Proposition (3.5) is known as the *comparison test.* It follows from the comparison test that every absolutely convergent series is convergent.

The terms of an absolutely convergent series $\sum_{n=1}^{\infty} x_n$ may be reordered without affecting the sum s_0 of the series. More precisely, if λ: $\mathbb{Z}^+ \to \mathbb{Z}^+$ is a bijection, then $\sum_{n=1}^{\infty} x_{\lambda(n)}$ exists and equals s_0. This may not be true if the series $\sum_{n=1}^{\infty} x_n$ is merely convergent.

A sequence (x_n) is said to *diverge* if there exists ε in $\mathbb{R}^+$ such that for each k in $\mathbb{Z}^+$ there exist m and n in $\mathbb{Z}^+$ with $m, n \geqq k$ and $|x_m - x_n| \geqq \varepsilon$. The motivation for this definition is, of course, that a sequence cannot be both convergent and divergent. A series is said to *diverge* if the sequence of its partial sums diverges.

The series $\sum_{n=1}^{\infty} n^{-1}$ diverges, because

$$\left|\sum_{n=1}^{2^{k+1}} n^{-1} - \sum_{n=1}^{2^k} n^{-1}\right| > \tfrac{1}{2} \qquad (k \in \mathbb{Z}^+).$$

The series $\sum_{n=1}^{\infty} x_n$ diverges whenever there exists r in $\mathbb{R}^+$ such that $|x_n| \geqq r$ for infinitely many values of n.

Let $\sum_{n=1}^{\infty} x_n$ and $\sum_{n=1}^{\infty} y_n$ be series of nonnegative terms. The *comparison test* for divergence is that $\sum_{n=1}^{\infty} x_n$ diverges whenever $\sum_{n=1}^{\infty} y_n$ diverges and there is a positive integer N with $x_n \geqq y_n$ for all $n \geqq N$.

The following very useful test for convergence and divergence is called the *ratio test.*

(3.6) **Proposition.** *Let* $\sum_{n=1}^{\infty} x_n$ *be a series,* c *a positive number, and* N *a positive integer. Then* $\sum_{n=1}^{\infty} x_n$ *converges if* $c<1$ *and*

(3.6.1) $$|x_{n+1}| \leqq c|x_n| \quad (n \geqq N),$$

and diverges if $c>1$ *and*

(3.6.2) $$|x_{n+1}| > c|x_n| \quad (n \geqq N).$$

Proof: Assume that $c<1$ and that (3.6.1) is valid. Then $|x_n| \leq c^{n-N}|x_N|$ for $n \geqq N$. By the comparison test, $\sum_{n=1}^{\infty} x_n$ converges.

Next, assume that $c>1$ and that (3.6.2) holds. Then

$$|x_n| \geqq c^{n-N-1}|x_{N+1}| \geqq |x_{N+1}| \quad (n \geqq N+1)$$

and

$$|x_{N+1}| > c|x_N| \geqq 0.$$

Therefore $\sum_{n=1}^{\infty} x_n$ diverges. □

A corollary of the ratio test is that if the limit

$$L \equiv \lim_{n\to\infty} |x_{n+1} x_n^{-1}|$$

exists, then $\sum_{n=1}^{\infty} x_n$ converges whenever $L<1$ and diverges whenever $L>1$.

The ratio test says nothing in case $L=1$. To handle this case, we introduce stronger tests based on *Kummer's criterion.*

(3.7) **Lemma.** *Let* (a_n) *and* (x_n) *be sequences of positive numbers,* c *a positive number, and* N *a positive integer. Then* $\sum_{n=1}^{\infty} x_n$ *converges if* $a_n x_n \to 0$ *as* $n \to \infty$ *and*

(3.7.1) $$a_n x_n x_{n+1}^{-1} - a_{n+1} \geqq c \quad (n \geqq N),$$

while $\sum_{n=1}^{\infty} x_n$ *diverges if* $\sum_{n=1}^{\infty} a_n^{-1}$ *diverges and*

(3.7.2) $$a_n x_n x_{n+1}^{-1} - a_{n+1} \leqq 0 \quad (n \geqq N).$$

Proof: Assume that $a_n x_n \to 0$ and that (3.7.1) is valid. Let ε be an arbitrary positive number, and choose an integer $\nu \geqq N$ so that $a_k x_k - a_j x_j \leqq c\,\varepsilon$ whenever $j > k \geqq \nu$. For such j and k we have

$$\begin{aligned}\sum_{n=k+1}^{j} x_n &\leqq c^{-1} \sum_{n=k+1}^{j} x_n(a_{n-1} x_{n-1} x_n^{-1} - a_n)\\ &= c^{-1}(a_k x_k - a_j x_j) \leqq \varepsilon.\end{aligned}$$

Thus $\left(\sum_{n=1}^{j} x_n\right)_{j=1}^{\infty}$ is a Cauchy sequence, and so $\sum_{n=1}^{\infty} x_n$ converges.

Next assume that $\sum_{n=1}^{\infty} a_n^{-1}$ diverges and that (3.7.2) holds. Then for each $n \geqq N$, $x_n \geqq a_N x_N a_n^{-1}$. Thus $\sum_{n=1}^{\infty} x_n$ diverges, by comparison with $\sum_{n=1}^{\infty} a_n^{-1}$. □

(3.8) **Lemma.** *Let (y_n) be a sequence of positive numbers, c a positive number, and N a positive integer such that*

$$n(y_n y_{n+1}^{-1} - 1) \geqq c \qquad (n \geqq N).$$

Then $\lim_{n\to\infty} y_n = 0$.

Proof: For each $n > N$,

$$\begin{aligned}y_N y_n^{-1} &= (y_N y_{N+1}^{-1})(y_{N+1} y_{N+2}^{-1}) \ldots (y_{n-1} y_n^{-1})\\ &\geqq (1 + cN^{-1}) \ldots (1 + c(n-1)^{-1})\\ &\geqq 1 + c\sum_{k=N}^{n-1} k^{-1}.\end{aligned}$$

Given $\varepsilon > 0$, choose an integer $\nu > N$ so that $\sum_{k=N}^{n-1} k^{-1} > c^{-1}(\varepsilon^{-1} y_N - 1)$ for all $n \geqq \nu$. Then for such n we have $y_n < \varepsilon$. Hence $y_n \to 0$ as $n \to \infty$. □

The next convergence test is known as *Raabe's test.*

(3.9) **Proposition.** *Let $\sum_{n=1}^{\infty} x_n$ be a series of positive numbers such that $n(x_n x_{n+1}^{-1} - 1)$ converges to a limit L. Then $\sum_{n=1}^{\infty} x_n$ converges if $L > 1$, and diverges if $L < 1$.*

Proof: First note that

$$\begin{aligned}n(n x_n/(n+1) x_{n+1} - 1) &= n(n+1)^{-1}(n(x_n x_{n+1}^{-1} - 1) - 1)\\ &\to L - 1 \qquad \text{as } n \to \infty.\end{aligned}$$

If $L>1$, it follows from (3.8) that $nx_n \to 0$ as $n\to\infty$. We then obtain the convergence of $\sum_{n=1}^{\infty} x_n$ by taking $a_n \equiv n$ ($n\in\mathbb{Z}^+$) in Kummer's criterion. The same choice of a_n yields divergence of $\sum_{n=1}^{\infty} x_n$ in case $L<1$. □

Important real numbers represented by series are

$$e=1+\sum_{n=1}^{\infty}(n!)^{-1}$$

and

$$\pi=4\sum_{n=0}^{\infty}(-1)^n(2n+1)^{-1}.$$

The series for e converges by the ratio test. The convergence of the series for π is a consequence of the general result that a series $\sum_{n=1}^{\infty}(-1)^n x_n$ converges whenever (i) $x_n \geqq 0$ for all n and (ii) the sequence (x_n) is decreasing and converges to 0. To see this, consider positive integers m and n with $m\geqq n$. Then

$$\begin{aligned} 0 &\leqq (x_n - x_{n+1}) + (x_{n+2} - x_{n+3}) + \ldots + (-1)^{m+n} x_m \\ &= (-1)^n \sum_{k=n}^{m} (-1)^k x_k \\ &= x_n - (x_{n+1} - x_{n+2}) - \ldots + (-1)^{m+n} x_m \leqq x_n. \end{aligned}$$

It follows that the sequence of partial sums of the series is a Cauchy sequence. Therefore the series converges.

4. Continuous Functions

A property P which is applicable to the elements of a set S is defined by a statement of the requirements that an element of S must satisfy in order to have property P. To construct an element of S with property P we must construct an element of S, perform certain additional constructions which depend on the property P, and prove that the entities constructed satisfy certain requirements that are characteristic of the property P. Each property P applicable to elements of a set S determines a subset of S which is denoted by

$$\{x:\ x\in S,\ x \text{ has property } P\}$$

or

$$\{x\in S:\ x \text{ has property } P\}.$$

When the context makes it clear which set S is under discussion, we also write simply

$$\{x: x \text{ has property } P\}.$$

Properties applicable to elements of a set S, and subsets of S, are essentially the same things regarded from different points of view.

Among the most important subsets of $\mathbb{R}$ are the intervals.

(4.1) **Definition.** For all real numbers a and b we define

$$(a, b) \equiv \{x: x \in \mathbb{R},\ a < x < b\},$$
$$(a, b] \equiv \{x: x \in \mathbb{R},\ a < x \leqq b\},$$
$$[a, b) \equiv \{x: x \in \mathbb{R},\ a \leqq x < b\},$$
$$[a, b] \equiv \{x: x \in \mathbb{R},\ a \leqq x \leqq b\}.$$

Each of these sets in an *interval*, whose left and right *end points* are a and b, respectively. The interval (a, b) is *open*, $[a, b]$ is *closed*, $(a, b]$ is *half-open on the left*, and $[a, b)$ is *half-open* on the right. If $a < b$, then the intervals are said to be *proper*.

The above intervals are called *finite intervals*. We also introduce *infinite intervals*, by the following definitions:

$$(-\infty, a) \equiv \{x: x \in \mathbb{R},\ x < a\},$$
$$(-\infty, a] \equiv \{x: x \in \mathbb{R},\ x \leqq a\},$$
$$(a, \infty) \equiv \{x: x \in \mathbb{R},\ a < x\},$$
$$[a, \infty) \equiv \{x: x \in \mathbb{R},\ a \leqq x\}.$$

An interval I is *nonvoid* if we can construct a real number belonging to I. A nonvoid, closed, finite interval is called a *compact interval*. A nonvoid finite interval I with left and right end points a and b has *length* $|I| \equiv b - a$.

If an interval I is a subset of an interval J (that is, if every element of I also belongs to J), then we say that I is a *subinterval* of J.

The rules for manipulating intervals, which we use in the sequel without mention or proof, are implicit in Proposition (2.11).

(4.2) **Definition.** A nonvoid set A of real numbers is *bounded above* if there exists a real number b, called an *upper bound* of A, such that $x \leqq b$ for all x in A. A real number b is called a *supremum*, or *least upper bound*, of A if it is an upper bound of A, and if for each $\varepsilon > 0$ there exists x in A with $x > b - \varepsilon$.

We say that A is *bounded below* if there exists a real number b, called a *lower bound* of A, such that $b \leqq x$ for all x in A. A real number b is called an *infimum*, or *greatest lower bound*, of A if it is a lower bound of A, and if for each $\varepsilon>0$ there exists x in A with $x<b+\varepsilon$.

The supremum (respectively, infimum) of A is unique, if it exists, and is written $\sup A$ (respectively, $\inf A$).

A classical theorem asserts that every nonvoid set of real numbers that is bounded above has a supremum. A counterexample to this is provided by the set $\{x_n: n\in\mathbb{Z}^+\}$ where $x_n=0$ or $x_n=1$ for each n, but it is not known whether $x_n=0$ for all n.

We now prove the constructive *least-upper-bound principle*.

(4.3) **Proposition.** *Let A be a nonvoid set of real numbers that is bounded above. Then $\sup A$ exists if and only if for all x, y in $\mathbb{R}$ with $x<y$, either y is an upper bound of A or there exists a in A with $x<a$.*

Proof: If $\sup A$ exists and $x<y$, then either $\sup A<y$ or $x<\sup A$; in the latter case we can find a in A with

$$\sup A-(\sup A-x)<a$$

and hence $x<a$. Thus the stated condition is necessary.

Conversely, assume that the stated condition holds. Let a_1 be an element of A, and choose an upper bound b_1 of A with $b_1>a_1$. We construct recursively a sequence (a_n) in A and a sequence (b_n) of upper bounds of A such that for each n in $\mathbb{Z}^+$,

(i) $a_n\leqq a_{n+1}<b_{n+1}\leqq b_n$

and

(ii) $b_{n+1}-a_{n+1}\leqq\frac{3}{4}(b_n-a_n)$.

Having found $a_1,\ldots,a_n$ and $b_1,\ldots,b_n$, if $a_n+\frac{3}{4}(b_n-a_n)$ is an upper bound of A, we set $b_{n+1}\equiv a_n+\frac{3}{4}(b_n-a_n)$ and $a_{n+1}\equiv a_n$; while if there exists a in A with $a>a_n+\frac{1}{4}(b_n-a_n)$, we set $a_{n+1}\equiv a$ and $b_{n+1}\equiv b_n$. This completes the recursive construction.

By (i) and (ii), we have

$$0\leqq b_n-a_n\leqq(3/4)^{n-1}(b_1-a_1)\qquad(n\in\mathbb{Z}^+).$$

Hence the sequences (a_n) and (b_n) converge to a common limit ℓ with $a_n\leqq\ell\leqq b_n$ for each n in $\mathbb{Z}^+$. Since each b_n is an upper bound of A, so is ℓ. On the other hand, given $\varepsilon>0$, we can choose n so that $\ell\geqq a_n>\ell-\varepsilon$, where $a_n\in A$. Hence $\ell=\sup A$. □

In Proposition (4.3), if A is contained in some interval I, then in order to prove that $\sup A$ exists it is sufficient to consider arbitrary points x and y in I with $x<y$.

(4.4) **Corollary.** *Let the subset A of $\mathbb{R}$ have the property that for each $\varepsilon>0$ there exists a subfinite set $\{y_1, \ldots, y_n\}$ of points of A such that for each x in A at least one of the numbers $|x-y_1|, \ldots, |x-y_n|$ is less than ε. (Such a set A is called* totally bounded.*) Then* $\sup A$ *and* $\inf A$ *exist.*

Proof: It will suffice to prove that $\sup A$ exists. To this end, let x and y be real numbers with $x<y$, and set $\alpha\equiv\frac{1}{4}(y-x)$. Choose points $a_1, \ldots, a_N$ in A such that for each a in A at least one of the numbers $|a-a_1|, \ldots, |a-a_N|$ is less than α. For some n with $1\leq n\leq N$ we have

$$a_n>\max\{a_1, \ldots, a_N\}-\alpha.$$

Either $x<a_n$ or $a_n<x+2\alpha$. In the latter case, if $a\in A$ and we choose k with $|a-a_k|<\alpha$, we have

$$a\leqq a_k+|a-a_k|<a_n+\alpha+\alpha<x+4\alpha=y,$$

so that y is an upper bound of A. Thus $\sup A$ exists, by (4.3). $\square$

Often when one real number depends on another the dependence is smooth, or continuous. An exact description of what this means is given in the following definition.

(4.5) **Definition.** A real-valued function f defined on a compact interval I is *continuous* on I if for each $\varepsilon>0$ there exists $\omega(\varepsilon)>0$ such that $|f(x)-f(y)|\leqq\varepsilon$ whenever $x, y\in I$ and $|x-y|\leqq\omega(\varepsilon)$. The operation $\varepsilon \mapsto\omega(\varepsilon)$ is called a *modulus of continuity* for f.

A real-valued function f on an arbitrary interval J is *continuous* on J if it is continuous on every compact subinterval I of J.

For example, when a and b are real numbers with $a<b$, then f is continuous on (a, b) if and only if it is continuous on $[a+\delta, b-\delta]$ for each δ with $0<\delta<\frac{1}{2}(b-a)$.

A modulus of continuity ω is an indispensable part of the definition of a continuous function on a compact interval, although sometimes it is not mentioned explicitly. In the same way, moduli of continuity of the restrictions of f to each compact subinterval are indispensable parts of the definition of a continuous function f on a general interval.

Constant functions, and the identity function $x\mapsto x$, are continuous on every interval.

(4.6) **Proposition.** *If f: $[a, b]\to\mathbb{R}$ is a continuous function on a compact interval, then the quantities*

$$\sup f\equiv\sup\{f(x):\ x\in[a, b]\}$$

and

$$\inf f \equiv \inf\{f(x): \ x \in [a, b]\}$$

(called, respectively, the supremum *and the* infimum *of f on the interval* $[a, b]$*) exist.*

Proof: Consider any $\varepsilon > 0$. Choose real numbers $a = a_0 \leqq a_1 \leqq \ldots \leqq a_n = b$ such that $a_{i+1} - a_i \leqq \omega(\varepsilon)$ $(0 \leqq i \leqq n-1)$, where ω is a modulus of continuity for f. Then for each x in $[a, b]$ we have $|x - a_i| \leqq \omega(\varepsilon)$, and therefore $|f(x) - f(a_i)| \leqq \varepsilon$, for some i. Since ε is arbitrary, it follows that the set $\{f(x): x \in [a, b]\}$ is totally bounded. Therefore $\sup f$ and $\inf f$ exist, by (4.4). □

Let $f: A \to \mathbb{R}$ and $g: A \to \mathbb{R}$ be two functions defined on the same set A. We define the sum function $f + g: A \to \mathbb{R}$ by

$$(f+g)(x) \equiv f(x) + g(x) \qquad (x \in A).$$

Such functions as fg, $|f|$, and $\max\{f, g\}$ are defined similarly. If $g(x) \neq 0$ for each x in A, then $g^{-1}: A \to \mathbb{R}$ is defined by

$$g^{-1}(x) \equiv (g(x))^{-1} \qquad (x \in A).$$

We also write $1/g$ for g^{-1}, and f/g for the quotient function fg^{-1}.

The proof of the following proposition, which resembles the proof of Proposition (3.4), is left to the reader.

(4.7) **Proposition.** *Let f and g be continuous real-valued functions defined on an interval I. Then the functions $f+g$, fg, and* $\max\{f, g\}$ *are continuous on I. If f is bounded away from 0 on every compact subinterval J of I – that is, if $|f(x)| \geqq c$ for all x in J and some $c > 0$ (depending on J) – then f^{-1} is continuous on I.*

Proposition (4.7) implies that the quotient of continuous functions is continuous, provided that the denominator is bounded away from 0 on every compact subinterval. It also implies that a polynomial function

$$x \mapsto c_0 x^n + c_1 x^{n-1} + \ldots + c_n$$

is continuous on every interval, and that $|f|$ is continuous on each interval where f is continuous.

The composition of continuous functions is continuous, in the sense that if $f: I \to J$ and $g: J \to \mathbb{R}$ are continuous, then $g \circ f$ is continuous, provided that f maps every compact subinterval of I into a compact subinterval of J. To prove this, it is sufficient to consider the case in which I and J are both compact. Let ω be a modulus of continuity for f, and σ a modulus of continuity for g. Then if $x, y \in I$, $\varepsilon > 0$, and $|x-y| \leqq \omega(\sigma(\varepsilon))$, we have $|f(x)-f(y)| \leqq \sigma(\varepsilon)$. Therefore

$$|g(f(x))-g(f(y))| \leqq \varepsilon.$$

It follows that $g \circ f$ is continuous, with modulus of continuity $\varepsilon \mapsto \omega(\sigma(\varepsilon))$.

Classically, a continuous function maps an interval onto an interval. We now prove a weak version of this result, known as the *intermediate value theorem.*

(4.8) **Theorem.** *Let f be a continuous map defined on an interval I, and let a, b be points of I with $f(a) < f(b)$. Then for each y in $[f(a), f(b)]$ and each $\varepsilon > 0$, there exists x in $[\min\{a, b\}, \max\{a, b\}]$ such that $|f(x)-y| < \varepsilon$.*

Proof: Since f is continuous, we must have $a \neq b$. We may assume that $a < b$. Consider y in $[f(a), f(b)]$ and $\varepsilon > 0$. Let

$$m \equiv \inf\{|f(x)-y| : a \leqq x \leqq b\},$$

which exists by (4.6). Suppose that $m > 0$. Then $f(a)-y \leqq -m$ and $f(b)-y \geqq m$. Let ω be a modulus of continuity for f on $[a, b]$, and choose points $a = x_0 \leqq x_1 \leqq \ldots \leqq x_n = b$ such that $x_{k+1} - x_k \leqq \omega(m)$ for $0 \leqq k \leqq n-1$. Then for such k we have

$$|f(x_{k+1})-y-(f(x_k)-y)| = |f(x_{k+1})-f(x_k)| \leqq m.$$

Since $|f(x)-y| \geqq m$ for all x in $[a, b]$, it follows that the quantities $f(x_k)-y$ and $f(x_{k+1})-y$ are either both positive or both negative. Therefore the quantities $f(x_i)-y$ $(0 \leqq i \leqq n)$ are either all positive or all negative. Hence $f(a)-y$ and $f(b)-y$ are either both positive or both negative. This contradiction ensures that the possibility $m > 0$ is ruled out; so that $m < \varepsilon$, and the desired conclusion follows. □

Under additional hypotheses satisfied by many of the common elementary functions of analysis, Theorem (4.8) can be strengthened to yield the conclusion that $f(x) = y$ for some x in $[\min\{a, b\}, \max\{a, b\}]$. For example, this strong conclusion obtains whenever f is *strictly increasing*, in the sense that $f(x) > f(x')$ for any two points x, x' of its domain with $x > x'$. For in that case, taking $a < b$ we can construct sequences (a_n), (b_n) in $[a, b]$ such that for each n in $\mathbb{Z}^+$,

(i) $a = a_1 \leqq a_2 \leqq \ldots \leqq a_n \leqq b_n \leqq \ldots \leqq b_2 \leqq b_1 = b$

(ii) $f(a_n) \leqq y \leqq f(b_n)$

(iii) $b_{n+1} - a_{n+1} \leqq (2/3)(b_n - a_n)$.

The sequences (a_n), (b_n) then converge to a common limit x in $[a,b]$ with $f(x)=y$.

Just as sequences of real numbers can converge to real numbers, sequences of continuous functions can converge to continuous functions. In fact, most of the important functions of analysis are defined as limits of sequences of continuous functions.

(4.9) **Definition.** A sequence (f_n) of continuous functions on a compact interval I *converges* on I to a continuous function f if for each $\varepsilon>0$ there exists N_ε in $\mathbb{Z}^+$ such that

(4.9.1) $$|f_n(x)-f(x)|\leqq\varepsilon \quad (x\in I,\ n\geqq N_\varepsilon).$$

A sequence (f_n) of continuous functions on an arbitrary interval J *converges* on J to a continuous function f if it converges to f on every compact subinterval I of J; in that case, f is called the *limit* of the sequence (f_n).

Definition (4.9) can be recast to bear a closer resemblance to Definition (3.1). To this end, we define the *norm* $\|f\|_I$ of a continuous function f on a compact interval I to be the supremum of $|f|$ on I. Then (f_n) converges to f on I if and only if for each k in $\mathbb{Z}^+$ there exists N_k in $\mathbb{Z}^+$ with

$$\|f_n-f\|_I\leqq k^{-1} \quad (n\geqq N_k).$$

(4.10) **Definition.** A sequence (f_n) of continuous functions on a compact interval I is a *Cauchy sequence* on I if for each $\varepsilon>0$ there exists M_ε in $\mathbb{Z}^+$ such that

(4.10.1) $$|f_m(x)-f_n(x)|\leqq\varepsilon \quad (x\in I;\ m,n\geqq M_\varepsilon).$$

A sequence of continuous functions on an arbitrary interval J is a *Cauchy sequence* on J if it is a Cauchy sequence on every compact subinterval of J.

The sequence (f_n) is a Cauchy sequence on the compact interval I if and only if for every k in $\mathbb{Z}^+$ there exists M_k in $\mathbb{Z}^+$ such that

$$\|f_m-f_n\|_I\leqq k^{-1} \quad (n,n\geqq M_k).$$

Notice that a sequence (c_n) of real numbers converges if and only if the corresponding sequence of constant functions, which we also denote by (c_n), converges on a given nonvoid interval I, and that a sequence of real numbers is a Cauchy sequence if and only if the corresponding sequence of constant functions is a Cauchy sequence on I. Because of these remarks, the following theorem is a generalization of Theorem (3.3).

(4.11) **Theorem.** *A sequence (f_n) of continuous functions on an interval J converges on J if and only if it is a Cauchy sequence on J.*

Proof: Assume that (f_n) converges to f on J. Let I be any compact subinterval of J. For each $\varepsilon>0$ choose N_ε in $\mathbb{Z}^+$ satisfying (4.9.1), and write $M_\varepsilon \equiv N_{\varepsilon/2}$. Then whenever $m, n \geqq M_\varepsilon$ and $x \in I$, we have

$$\begin{aligned}|f_m(x)-f_n(x)| &\leqq |f_m(x)-f(x)|+|f_n(x)-f(x)|\\ &\leqq \varepsilon/2+\varepsilon/2=\varepsilon.\end{aligned}$$

Therefore (f_n) is a Cauchy sequence on I. It follows that (f_n) is a Cauchy sequence on J.

Assume conversely that (f_n) is a Cauchy sequence on J. Then for each x in J, $(f_n(x))$ is a Cauchy sequence of real numbers, whose limit we denote by $f(x)$. We shall show that $f\colon J \to \mathbb{R}$ is a continuous function and that (f_n) converges to f on J. It is enough to show that f is continuous on each compact subinterval I of J, and that (f_n) converges to f on I. To this end, choose the positive integers M_ε such that (4.10.1) is valid, and for each n in $\mathbb{Z}^+$ let ω_n be a modulus of continuity for f_n on I. For each $\varepsilon>0$ write

$$\omega(\varepsilon) \equiv \omega_M(\varepsilon/3),$$

where $M \equiv M_{\varepsilon/3}$. Then whenever $x, y \in I$ and $|x-y| \leqq \omega(\varepsilon)$, we have

$$\begin{aligned}|f(x)-f(y)| &\leqq |f(x)-f_M(x)|+|f_M(x)-f_M(y)|+|f_M(y)-f(y)|\\ &= \lim_{n\to\infty}|f_n(x)-f_M(x)|+|f_M(x)-f_M(y)|+\lim_{n\to\infty}|f_M(y)-f_n(y)|\\ &\leqq \frac{\varepsilon}{3}+\frac{\varepsilon}{3}+\frac{\varepsilon}{3}=\varepsilon.\end{aligned}$$

Therefore f is continuous, with modulus of continuity ω. Finally, if $x \in I$, $\varepsilon>0$, and $n \geqq M_\varepsilon$, then

$$|f_n(x)-f(x)| = \lim_{m\to\infty}|f_n(x)-f_m(x)| \leqq \varepsilon.$$

Hence (f_n) converges to f on I. □

Notations to express the fact that (f_n) converges to f are

$$\lim_{n\to\infty} f_n = f$$

and

$$f_n \to f \quad \text{as } n \to \infty.$$

We also write simply $f_n \to f$.

To each sequence (f_n) of continuous functions on an interval I corresponds a sequence (g_n) of *partial sums*, defined by

$$g_n \equiv \sum_{k=1}^{n} f_k.$$

If (g_n) converges to a continuous function g on I, then g is the *sum* of the *series* $\sum_{n=1}^{\infty} f_n$,

$$g \equiv \sum_{n=1}^{\infty} f_n,$$

and the series is said to *converge* to g on I. If $\sum_{n=1}^{\infty} |f_n|$ converges on I, then $\sum_{n=1}^{\infty} f_n$ is said to *converge absolutely* on I. An absolutely convergent series of functions converges.

The comparison test and the ratio test for convergence carry over to series of functions. The *comparison test* states that if $\sum_{n=1}^{\infty} g_n$ is a convergent series of nonnegative continuous functions on an interval I, then the series $\sum_{n=1}^{\infty} f_n$ of continuous functions on I converges on I whenever $|f_n(x)| \leqq g_n(x)$ for all n in $\mathbb{Z}^+$ and all x in I.

The *ratio test* states that if $\sum_{n=1}^{\infty} f_n$ is a series of continuous functions on an interval J such that for each compact subinterval I of J there exist a constant c_I, $0 < c_I < 1$, and a positive integer N_I with

$$|f_{n+1}(x)| \leqq c_I |f_n(x)| \qquad (x \in I,\ n \geqq N_I),$$

then $\sum_{n=1}^{\infty} f_n$ converges absolutely on J.

A *power series* is a series of the form $\sum_{n=0}^{\infty} a_n(x-x_0)^n$, where $a_n(x-x_0)^n$ represents the function $x \mapsto a_n(x-x_0)^n$ and where $a_0(x-x_0)^0 \equiv a_0$ for all x. The ratio test has the following corollary.

(4.12) **Proposition.** *Let the power series* $\sum_{n=0}^{\infty} a_n(x-x_0)^n$ *have the property that there exist* $r>0$ *and* N *in* $\mathbb{Z}^+$ *such that* $|a_{n+1}| \leqq r^{-1}|a_n|$ *for all* $n \geqq N$. *Then the series converges absolutely on the interval* $(x_0 - r, x_0 + r)$.

Proof: If I is a compact subinterval of $(x_0 - r, x_0 + r)$, then there exists r_0 with $0 < r_0 < r$ such that $|x - x_0| \leqq r_0$ for all x in I. Then

$$|a_{n+1}(x-x_0)^{n+1}| \leqq r^{-1} r_0 |a_n(x-x_0)^n| \qquad (n \geqq N,\ x \in I).$$

By the ratio test, the series therefore converges absolutely on I. □

5. Differentiation

The rate at which a function is changing is a fundamental property of the function. Here is the precise definition of this concept.

(5.1) **Definition.** Let f and g be continuous functions on a proper compact interval I, and let δ be an operation from $\mathbb{R}^+$ to $\mathbb{R}^+$ such that

$$|f(y)-f(x)-g(x)(y-x)|\leqq\varepsilon|y-x|$$

whenever $\varepsilon>0$, $x, y\in I$, and $|y-x|\leq\delta(\varepsilon)$. Then f is said to be *differentiable* on I, g is called a *derivative* of f on I, and δ is called a *modulus of differentiability* for f on I.

If f and g are continuous functions on a proper interval J, then g is a *derivative* of f on J if it is a derivative of f on every proper compact subinterval of J; f is then said to be *differentiable* on J.

To express that g is a derivative of f we write

$$g=f,' \quad \text{or} \quad g=Df, \quad \text{or} \quad g(x)=\frac{df(x)}{dx}.$$

One way to interpret Definition (5.1) is that the *difference quotient*

$$(f(y)-f(x))(y-x)^{-1}$$

approaches $g(x)$ as y approaches x. In other words, g is the rate of change of f.

If f has two derivatives on I, then they are equal functions.

(5.2) **Theorem.** *Let f_1 and f_2 be differentiable functions on an interval I. Then f_1+f_2 and $f_1 f_2$ are differentiable on I. In case f_1 is bounded away from 0 on every compact subinterval of I, then f_1^{-1} is differentiable on I. The function $x\mapsto x$ is differentiable on $\mathbb{R}$. For each c in $\mathbb{R}$ the function $x\mapsto c$ is differentiable on $\mathbb{R}$. The derivatives in question are given by the following relations:*

(a) $D(f_1+f_2)=Df_1+Df_2$

(b) $D(f_1 f_2)=f_1 Df_2+f_2 Df_1$

(c) $Df_1^{-1}=-f_1^{-2}Df_1$

(d) $\dfrac{dx}{dx}=1$

(e) $\dfrac{dc}{dx}=0.$

Proof: It is enough to consider the case in which I is compact. Let δ_1 and δ_2 be moduli of differentiability for f_1 and f_2, respectively, on I, and ω_1 a modulus of continuity for f_1 on I.

(a) Whenever $x, y \in I$ and $|y-x| \leqq \delta(\varepsilon) \equiv \min\{\delta_1(\varepsilon/2), \delta_2(\varepsilon/2)\}$, we have

$$\begin{aligned}
&|f_1(y)+f_2(y)-(f_1(x)+f_2(x))-(f_1'(x)+f_2'(x))(y-x)| \\
&\qquad \leqq |f_1(y)-f_1(x)-f_1'(x)(y-x)|+|f_2(y)-f_2(x)-f_2'(x)(y-x)| \\
&\qquad \leqq \frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon.
\end{aligned}$$

Thus f_1+f_2 is differentable on I, with derivative $f_1'+f_2'$ and modulus of differentiability δ.

(b) Let M be a common bound for $|f_1|$, $|f_2|$, and $|f_2'|$ on the interval I. (For instance, define $M \equiv \max\{\|f_1\|_I, \|f_2\|_I, \|f_2'\|_I\}$.) Then whenever $x, y \in I$ and

$$|y-x| \leqq \delta(\varepsilon) \equiv \min\{\delta_1((3M)^{-1}\varepsilon), \delta_2((3M)^{-1}\varepsilon), \omega_1((3M)^{-1}\varepsilon)\},$$

we have

$$\begin{aligned}
&|f_1(y)\,f_2(y)-f_1(x)\,f_2(x)-(f_1(x)\,f_2'(x)+f_2(x)\,f_1'(x))(y-x)| \\
&\qquad \leqq |f_1(y)|\,|f_2(y)-f_2(x)-f_2'(x)(y-x)| \\
&\qquad\quad +|f_1(y)-f_1(x)|\,|f_2'(x)|\,|y-x| \\
&\qquad\quad +|f_2(x)|\,|f_1(y)-f_1(x)-f_1'(x)(y-x)| \\
&\qquad \leqq 3M(3M)^{-1}\varepsilon|y-x|=\varepsilon|y-x|.
\end{aligned}$$

Therefore $f_1 f_2$ is differentiable on I, with derivative $f_1 f_2'+f_2 f_1'$ and modulus of differentiability δ.

(c) For each $\varepsilon>0$ write

$$\delta(\varepsilon) \equiv \min\{\delta_1(\tfrac{1}{2}M^{-2}\varepsilon), \omega_1(\tfrac{1}{2}M^{-4}\varepsilon)\}$$

where $M \equiv \max\{\|f_1^{-1}\|_I, \|f_1'\|_I\}$. Then whenever $x, y \in I$ and $|y-x| \leqq \delta(\varepsilon)$, we have

$$\begin{aligned}
&|f_1^{-1}(y)-f_1^{-1}(x)+f_1^{-2}(x)\,f_1'(x)(y-x)| \\
&\quad =|f_1^{-1}(x)\,f_1^{-1}(y)|\,|f_1(y)-f_1(x)-f_1(y)\,f_1^{-1}(x)\,f_1'(x)(y-x)| \\
&\quad \leqq M^2|f_1(y)-f_1(x)-f_1'(x)(y-x)| \\
&\qquad +M^2|f_1'(x)\,f_1(x)^{-1}|\,|f_1(y)-f_1(x)|\,|y-x| \\
&\quad \leqq M^2(\tfrac{1}{2}M^{-2}\varepsilon)|y-x|+M^4(\tfrac{1}{2}M^{-4}\varepsilon)|y-x|=\varepsilon|y-x|.
\end{aligned}$$

Therefore f_1^{-1} is differentiable on I, with derivative $-f_1^{-2}f_1'$ and modulus of differentiability δ.

(d) This is obvious.

(e) This is obvious too. □

(5.3) **Corollary.** *For all positive integers n,*

(5.3.1)
$$\frac{dx^n}{dx} = n\,x^{n-1}.$$

Proof: The proof is by induction on n. When $n=1$, (5.3.1) is just (d) of Theorem (5.2). If (5.3.1) is true for a given value of n, then

$$\frac{dx^{n+1}}{dx} = \frac{d(x \cdot x^n)}{dx} = x^n + x(n\,x^{n-1}) = (n+1)\,x^n,$$

by (b) of Theorem (5.2). Therefore (5.3.1) is true for all n. □

Theorem (5.2) and its corollary imply the formula

$$D(f_1 f_2^{-1}) = f_2^{-2}(f_2 D f_1 - f_1 D f_2)$$

for the derivative of a quotient, and the formula

$$D\left(\sum_{k=0}^{n} a_{n-k} x^k\right) = \sum_{k=1}^{n} k\, a_{n-k}\, x^{k-1}$$

for the derivative of a polynomial.

The next theorem is the so-called *chain rule* for the derivative of a composite function. Its intuitive meaning is that the rate of change of quantity C with respect to quantity A is the product of the rate of change of C with respect to some third quantity B by the rate of change of B with respect to A.

(5.4) **Theorem.** *Let $f\colon I \to \mathbb{R}$ and $g\colon J \to \mathbb{R}$ be differentiable functions such that f maps each compact subinterval of I into a compact subinterval of J. Then $g \circ f$ is differentiable, and*

(5.4.1)
$$(g \circ f)' = (g' \circ f)\, f'.$$

Proof: It is no loss of generality to assume that I and J are compact. Let δ_f be a modulus of differentiability and ω_f a modulus of continuity for f on I. Let δ_g be a modulus of differentiability for g on J. For each $\varepsilon > 0$ write

$$\delta(\varepsilon) \equiv \min\{\omega_f(\delta_g(\alpha)),\ \delta_f(\beta)\},$$

where

$$\alpha \equiv (1 + \|f'\|_I)^{-1} \frac{\varepsilon}{2} \quad \text{and} \quad \beta \equiv (\alpha + \|g'\|_J)^{-1} \frac{\varepsilon}{2}.$$

Then for x, y in I and $|y-x| \leqq \delta(\varepsilon)$ we have $|f(y)-f(x)| \leqq \delta_g(\alpha)$, so that

$$|g(f(y))-g(f(x))-g'(f(x))(f(y)-f(x))| \leqq \alpha|f(y)-f(x)|.$$

Also

$$|f(y)-f(x)| \leqq \|f'\|_I|y-x|+|f(y)-f(x)-f'(x)(y-x)|$$

and

$$|f(y)-f(x)-f'(x)(y-x)| \leqq \beta|y-x|.$$

Using these inequalities and noting that $\alpha\|f'\|_I<\varepsilon/2$, we compute

$$\begin{aligned}
&|g(f(y))-g(f(x))-g'(f(x))\,f'(x)(y-x)| \\
&\quad \leqq |g(f(y))-g(f(x))-g'(f(x))(f(y)-f(x))| \\
&\qquad +|g'(f(x))|\,|f(y)-f(x)-f'(x)(y-x)| \\
&\quad \leqq \alpha|f(y)-f(x)|+\|g'\|_J|f(y)-f(x)-f'(x)(y-x)| \\
&\quad \leqq \alpha\|f'\|_I|y-x|+(\alpha+\|g'\|_J)|f(y)-f(x)-f'(x)(y-x)| \\
&\quad <\frac{\varepsilon}{2}|y-x|+\frac{\varepsilon}{2}|y-x|=\varepsilon|y-x|.
\end{aligned}$$

It follows that $g\circ f$ is differentiable on I, with derivative $(g'\circ f)f'$ and modulus of differentiability δ. □

The next lemma is known as *Rolle's theorem.*

(5.5) **Lemma.** *Let f be differentiable on the interval $[a,b]$, and let $f(a)=f(b)$. Then for each $\varepsilon>0$ there exists x in $[a,b]$ with $|f'(x)|\leqq\varepsilon$.*

Proof: Let δ be a modulus of differentiability for f on $[a,b]$. Let

$$m\equiv\inf\{|f'(x)|: x\in[a,b]\},$$

which exists, by (4.6). Suppose that $m>0$. We may assume that $f'(a)\geqq m$. For each x in $[a,b]$ we have $f'(x)\geqq m$. For if $f'(x)<m$, then $f'(x)\leqq -m$, so that, by the intermediate value theorem (4.8), there exists ξ in $[a,b]$ with $|f'(\xi)|<m$; this contradicts the definition of m. Now choose points $a=x_0\leqq x_1\leqq\ldots\leqq x_n=b$ so that $x_{k+1}-x_k\leqq\delta(\frac{1}{2}m)$ $(0\leqq k\leqq n-1)$. Then

$$\begin{aligned}
0&=f(b)-f(a) \\
&=\sum_{k=0}^{n-1}(f(x_{k+1})-f(x_k)) \\
&=\sum_{k=0}^{n-1}f'(x_k)(x_{k+1}-x_k)+\sum_{k=0}^{n-1}(f(x_{k+1})-f(x_k)-f'(x_k)(x_{k+1}-x_k)) \\
&\geqq\sum_{k=0}^{n-1}m(x_{k+1}-x_k)-\sum_{k=0}^{n-1}\tfrac{1}{2}m(x_{k+1}-x_k) \\
&=\tfrac{1}{2}m(b-a)>0.
\end{aligned}$$

This contradiction ensures that $m=0$. The desired conclusion follows immediately. $\square$

Rolle's theorem implies the *mean value theorem*, which gives a basic estimate for the difference of two values of a differentiable function.

(5.6) **Theorem.** *Let f be differentiable on the interval $[a,b]$. Then for each $\varepsilon>0$ there exists x in $[a,b]$ with*

$$|f(b)-f(a)-f'(x)(b-a)|\leqq\varepsilon.$$

Proof: Define the function h on $[a,b]$ by

$$h(x)\equiv(x-a)(f(b)-f(a))-f(x)(b-a)\qquad(x\in[a,b]).$$

Then $h(b)=h(a)=-f(a)(b-a)$. By (5.5), there exists x in $[a,b]$ with

$$\varepsilon\geqq|h'(x)|=|f(b)-f(a)-f'(x)(b-a)|.\quad\square$$

A function f on a proper interval I is *increasing* (respectively, *strictly increasing*) if $f(x)\geqq f(y)$ (respectively, $f(x)>f(y)$) whenever $x,y\in I$ and $x>y$. We say that f is *decreasing* (respectively, *strictly decreasing*) if $-f$ is increasing (respectively, strictly increasing). It follows from Theorem (5.6) that if $f:I\to\mathbb{R}$ is differentiable on I and $f'(x)\geqq 0$ (respectively, $f'(x)\leqq 0$) for all x in I, then f is increasing (respectively, decreasing) on I.

(5.7) **Definition.** Let $f,f^{(1)},f^{(2)},\dots,f^{(n-1)}$ be differentiable functions on a proper interval I such that

$$Df=f^{(1)},\ Df^{(1)}=f^{(2)},\dots,Df^{(n-2)}=f^{(n-1)},$$

and set $f^{(n)}\equiv Df^{(n-1)}$. Then $f^{(n)}$ is called the n^{th} *derivative* of f on I, and is also written D^nf; f is then said to be *n times differentiable* on I. The function f itself may be written $f^{(0)}$ or D^0f.

A natural way to simplify a continuous function and set it up for computation is to replace it by a polynomial approximation. The basic result on polynomial approximation of differentiable functions is *Taylor's theorem* ((5.10) below). To see the motivation for Taylor's theorem, consider an n times differentiable function f on an interval I, and a point a in I. It is natural to approximate f by a polynomial of degree n whose derivatives of orders $0,1,\dots,n$ at a have the same values as the corresponding derivatives of f at a. The unique such polynomial is obviously

(5.8)
$$\sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k.$$

This approximation is useful for a given value b of x only when there exists a good estimate for the remainder

(5.9)
$$R \equiv f(b) - \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(b-a)^k.$$

The purpose of Taylor's theorem is to obtain such an estimate.

(5.10) **Theorem.** *Let f be an $(n+1)$ times differentiable function on an interval I, let ε be a positive number, and let a and b be points of I. Then there exists c with* $\min\{a,b\} \leqq c \leqq \max\{a,b\}$, *such that*

$$\left| R - \frac{f^{(n+1)}(c)}{n!}(b-c)^n(b-a) \right| \leqq \varepsilon,$$

where R is given by (5.9).

Proof: Let

$$M \equiv 1 + \max\{f^{(1)}(a)/1!, f^{(2)}(a)/2!, \ldots, f^{(n)}(a)/n!\}$$

and let ω be a modulus of continuity for f on $[\min\{a,b\}, \max\{a,b\}]$. Let

$$0 < \delta < \min\{1, \varepsilon/2nM, \omega(\tfrac{1}{2}\varepsilon)\}.$$

Either $0 < |a-b|$ or $|a-b| < \delta$. In the latter case,

$$\begin{aligned} |R| &\leqq |f(b)-f(a)| + \sum_{k=1}^{n} |f^{(k)}(a)/k!|\,|b-a|^k \\ &\leqq \tfrac{1}{2}\varepsilon + M\sum_{k=1}^{n} \delta^k < \tfrac{1}{2}\varepsilon + M\sum_{k=1}^{n} (\varepsilon/2nM) = \varepsilon, \end{aligned}$$

and so we can take $c \equiv b$. In case $0 < |a-b|$, define the differentiable function g on I by

$$\begin{aligned} g(x) \equiv f(b) - f(x) - \frac{f'(x)}{1!}(b-x) - \frac{f''(x)}{2!}(b-x)^2 - \\ \cdots - \frac{f^{(n)}(x)}{n!}(b-x)^n - R(b-x)(b-a)^{-1}. \end{aligned}$$

Then $g(a) = g(b) = 0$ and

$$\begin{aligned} g'(x) &= -f'(x) + f'(x) - f''(x)(b-x) + f''(x)(b-x) - \\ &\quad \cdots + \frac{f^{(n)}(x)}{(n-1)!}(b-x)^{n-1} - \frac{f^{(n+1)}(x)}{n!}(b-x)^n + R(b-a)^{-1} \\ &= -\frac{f^{(n+1)}(x)}{n!}(b-x)^n + R(b-a)^{-1}, \end{aligned}$$

since the terms cancel in pairs except for the last two. By Rolle's theorem, there exists c with $\min\{a,b\} \leqq c \leqq \max\{a,b\}$ and $|g'(c)| \leqq \varepsilon(b-a)^{-1}$. The required estimate for R now follows. □

An important special case occurs when f is *infinitely differentiable* – that is, $f^{(n)}$ exists for all positive integers n – on an open interval $I \equiv (a-t, a+t)$. In that case, if

$$\frac{r^n f^{(n+1)}}{n!} \to 0 \quad \text{as } n \to \infty \ (0<r<t), \tag{5.11}$$

then the *Taylor series* for f about a,

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n, \tag{5.12}$$

converges to f on I. Note that although condition (5.11) enables us to prove the convergence of (5.12) by the comparison test, Theorem (5.10) is needed to show that the sum of the Taylor series is f.

In fact, the series (5.12) converges to f under the weaker condition

$$\frac{r^n f^{(n)}}{n!} \to 0 \quad \text{as } n \to \infty \ (0<r<t).$$

To prove this, we need a different estimate of the remainder term, which is derived using the theory of integration from Section 6 below. (See Problem 23.)

Examples of (5.12) will occur later, when some of the common functions of analysis are expanded in Taylor series.

6. Integration

Differentiation deals with the instantaneous behavior of a function. We now look at a process, called *integration*, which deals with the behavior of a function on the average; in fact, it is essentially a rigorous formulation of the averaging process. Differentiation and integration will turn out to be inverse processes in a certain precise sense.

If $a = a_0 \leqq a_1 \leqq \ldots \leqq a_n = b$ are real numbers, the finite sequence $P \equiv (a_0, \ldots, a_n)$ is called a *partition* of the interval $[a,b]$, n is called the *length* of P, and

$$\operatorname{mesh} P \equiv \max\{a_{i+1} - a_i : 0 \leqq i \leqq n-1\}$$

is called the *mesh* of P. A partition $Q \equiv (a'_0, \ldots, a'_m)$ of $[a,b]$ is a *refinement* of P if for each i $(0 \leqq i \leqq n)$ there exists j with $a'_j = a_i$.

For each partition $P \equiv (a_0, a_1, \ldots, a_n)$ of $[a,b]$ and each continuous function f on $[a,b]$, $S(f,P)$ will denote an arbitrary sum of the type

$$\sum_{i=0}^{n-1} f(x_i)(a_{i+1}-a_i) \tag{6.1}$$

where $x_i \in [a_i, a_{i+1}]$ $(0 \leqq i \leqq n-1)$. In particular, if $a_i = a + i(b-a)/n$ $(0 \leqq i \leqq n)$, the quantity

$$S(f,a,b,n) \equiv \frac{b-a}{n} \sum_{i=0}^{n-1} f\left(a + i\frac{b-a}{n}\right) \tag{6.2}$$

is one of the numbers $S(f,P)$.

Two partitions $P \equiv (a_0, \ldots, a_n)$ and $Q \equiv (b_0, \ldots, b_m)$ of $[a,b]$ are *separated* if $a_0 < a_1 < \ldots < a_n$, $b_0 < b_1 < \ldots < b_m$, and $a_i \neq b_j$ for all i and j with $1 \leqq i \leqq n-1$ and $1 \leqq j \leqq m-1$.

(6.3) **Theorem.** *If f is a continuous function on a compact interval $[a,b]$, then the sequence*

$$(S(f,a,b,n))_{n=1}^{\infty} \tag{6.3.1}$$

converges to a limit, which is called the integral of f from a to b *and is written*

$$\int_a^b f(x)\,dx. \tag{6.3.2}$$

Moreover, if ω is a modulus of continuity for f, then for any $\varepsilon > 0$ and any partition P of $[a,b]$ with mesh $P \leqq \omega(\varepsilon)$, *we have*

$$\left| S(f,P) - \int_a^b f(x)\,dx \right| \leqq \varepsilon(b-a). \tag{6.3.3}$$

Proof: Let ε and δ be arbitrary positive numbers. Consider partitions $P \equiv (a_0, \ldots, a_n)$ and $Q \equiv (b_0, \ldots, b_m)$ of $[a,b]$ with mesh $P \leqq \omega(\varepsilon)$ and mesh $Q \leqq \omega(\delta)$. To begin with, assume that $b-a > 0$ and that P and Q are separated. Let $R \equiv (c_0, \ldots, c_r)$ be the common refinement of P and Q obtained by ordering the numbers $a_0, \ldots, a_n, b_1, \ldots, b_{m-1}$ in a strictly increasing sequence. Let $z_k \in [c_k, c_{k+1}]$ $(0 \leqq k \leqq r-1)$. For each i in $\{0, \ldots, n-1\}$, let $\sum_i$ denote summation over those indices k for which $a_i \leqq c_k < a_{i+1}$. Then

$$\begin{aligned}
|S(f,P) - S(f,R)| &= \left| \sum_{i=0}^{n-1} f(x_i)(a_{i+1}-a_i) - \sum_{k=0}^{r-1} f(z_k)(c_{k+1}-c_k) \right| \\
&= \left| \sum_{i=0}^{n-1} f(x_i) \sum_i (c_{k+1}-c_k) - \sum_{i=0}^{n-1} \sum_i f(z_k)(c_{k+1}-c_k) \right| \\
&\leqq \sum_{i=0}^{n-1} \sum_i |f(x_i) - f(z_k)|\,(c_{k+1}-c_k) \leqq \sum_{i=0}^{n-1} \sum_i \varepsilon(c_{k+1}-c_k) = \varepsilon(b-a).
\end{aligned}$$

Similarly,

$$|S(f,Q)-S(f,R)|\leqq\delta(b-a).$$

It now follows that

$$|S(f,P)-S(f,Q)|\leqq(\varepsilon+\delta)(b-a). \tag{6.3.4}$$

Now consider the general case, and let α be an arbitrary positive number. If $b-a$ is small enough, then

$$|S(f,P)-S(f,Q)|\leqq(\varepsilon+\delta)(b-a)+\alpha. \tag{6.3.5}$$

If $b-a>0$, then applying the first part of the proof to separated partitions which are sufficiently close approximations to P and Q, we again obtain (6.3.5). As $\alpha>0$ is arbitrary, we now see that (6.3.4) holds in the general case also. In particular,

$$|S(f,a,b,j)-S(f,a,b,k)|\leqq 2\varepsilon(b-a)$$

for all positive integers $j,k\geqq\omega(\varepsilon)^{-1}(b-a)$. Therefore (6.3.1) is a Cauchy sequence, and hence converges. For each integer $k\geqq\omega(\delta)^{-1}(b-a)$, inequality (6.3.4) entails

$$|S(f,P)-S(f,a,b,k)|\leqq(\varepsilon+\delta)(b-a).$$

Passing to the limit as $k\to\infty$ and $\delta\to 0$ gives (6.3.3). □

If f and g are real-valued functions on a set X such that $f(x)\leqq g(x)$ for all x in X, then we write $f\leqq g$.

(6.4) **Proposition.** *Let f and g be continuous functions on a compact interval $[a,b]$, and let α,β be real numbers. Then*

$$\int_a^b(\alpha f(x)+\beta g(x))\,dx=\alpha\int_a^b f(x)\,dx+\beta\int_a^b g(x)\,dx. \tag{6.4.1}$$

Moreover, if $f\leqq g$, then

$$\int_a^b f(x)\,dx\leqq\int_a^b g(x)\,dx. \tag{6.4.2}$$

Proof: Equation (6.4.1) follows from the fact that

$$S(\alpha f+\beta g,a,b,n)=\alpha S(f,a,b,n)+\beta S(g,a,b,n)$$

for all n in $\mathbb{Z}^+$, and (6.4.2) follows from the inequalities

$$S(f,a,b,n)\leqq S(g,a,b,n)\qquad(n\in\mathbb{Z}^+).\quad\square$$

As a corollary we have

$$c_1(b-a)\leqq\int_a^b f(x)\,dx\leqq c_2(b-a)$$

whenever $c_1 \leqq f(x) \leqq c_2$ for all x in $[a,b]$. By setting $-c_1 = c_2 = \|f\|_{[a,b]}$, we obtain

(6.5)
$$\left|\int_a^b f(x)\,dx\right| \leqq \|f\|_{[a,b]}(b-a).$$

(6.6) **Proposition.** *If f is a continuous function on a compact interval $[a,b]$, then*

(6.6.1)
$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx \qquad (a \leqq c \leqq b).$$

Proof. Let $a \leqq c \leqq b$. For arbitrary partitions $P \equiv (a, \ldots, c)$ of $[a,c]$ and $Q \equiv (c, \ldots, b)$ of $[c,b]$ we have

$$S(f,P) + S(f,Q) = S(f,R)$$

with appropriate choices of the approximating sums, where R is the partition $(a, \ldots, c, \ldots, b)$ of $[a,b]$ formed by combining the partitions P and Q. Since mesh $R = \max\{\text{mesh}\, P,\ \text{mesh}\, Q\}$, equality (6.6.1) follows by passage to the limit as mesh $P \to 0$ and mesh $Q \to 0$. □

For arbitrary real numbers a and b and each continuous function f on $[\min\{a,b\}, \max\{a,b\}]$ we define

(6.7)
$$\int_a^b f(x)\,dx \equiv \int_\lambda^b f(x)\,dx - \int_\lambda^a f(x)\,dx,$$

where $\lambda \equiv \min\{a,b\}$. In case $a \leqq b$ this agrees with the definition already given.

Equality (6.4.1) is valid for arbitrary real numbers a and b, provided that the integrals are defined (which means that f and g are continuous on the interval $[\min\{a,b\}, \max\{a,b\}]$). Equality (6.6.1) is valid whenever f is continuous on the interval $[\min\{a,b,c\}, \max\{a,b,c\}]$. In particular, setting $b = a$ in (6.6.1) gives

$$\int_a^c f(x)\,dx = -\int_c^a f(x)\,dx$$

whenever the integrals are defined. For arbitrary real numbers a and b equality (6.5) holds with $\|f\|_{[a,b]}$ replaced by the norm of f on $[\min\{a,b\}, \max\{a,b\}]$.

We now prove that differentiation and integration are inverse processes, a result whose imposing title is *the fundamental theorem of calculus.*

(6.8) **Theorem.** *Let f be a continuous function on a proper interval I, let a be a point of I, and write*

$$g(x) \equiv \int_a^x f(t)\,dt \qquad (x \in I).$$

Then $g' = f$. Also, if g_0 is any differentiable function on I with $g_0' = f$, then the difference $g - g_0$ is a constant function.

Proof: There is no loss of generality in taking I to be compact. Let ω be a modulus of continuity for f on I. For each $\varepsilon > 0$ and each pair of points x, y in I with $|x-y| \leqq \omega(\varepsilon)$, we have

$$\begin{aligned} |g(y) - g(x) - f(x)(y-x)| &= \left| \int_x^y f(t)\,dt - \int_x^y f(x)\,dt \right| \\ &= \left| \int_x^y (f(t) - f(x))\,dt \right| \leqq \varepsilon |y-x|. \end{aligned}$$

Therefore g is differentiable on I, with derivative f and modulus of differentiability ω. If also $g_0' = f$ on I, then $(g-g_0)' = 0$ on I. By the mean-value theorem, $(g-g_0)(a) = (g-g_0)(b)$ whenever a, $b \in I$ and $a < b$. Therefore $g - g_0$ is a constant function. □

Our next lemma has an interest of its own.

(6.9) **Lemma.** *Let (f_n) be a sequence of continuous functions converging to a function f on a nonvoid interval I. Let a be any point of I, and write*

$$g(x) \equiv \int_a^x f(t)\,dt \qquad (x \in I)$$

and

$$g_n(x) \equiv \int_a^x f_n(t)\,dt \qquad (x \in I)$$

for each positive integer n. Then $g_n \to g$ on I as $n \to \infty$.

Proof: It is no loss of generality to assume that I is compact. Let r be the length of I. Then for each x in I we have

$$\begin{aligned} |g_n(x) - g(x)| &= \left| \int_a^x (f_n(t) - f(t))\,dt \right| \\ &\leqq r \, \| f_n - f \|_I \to 0 \qquad \text{as } n \to \infty. \end{aligned}$$

Therefore (g_n) converges to g on I. □

(6.10) **Theorem.** *Let (f_n) be a sequence of differentiable functions which converges to a continuous f on a proper interval I. Assume that the sequence (f_n') of derivatives converges to a continuous function g on I. Then $f' = g$ on I.*

Proof: It is no loss of generality to assume that I is compact. Let a be any point of I. For each x in I write

$$h(x) \equiv \int_a^x g(t)\,dt$$

and

$$h_n(x) \equiv \int_a^x f_n'(t)\,dt.$$

By (6.8), we have $h' = g$ and $h_n' = f_n'$ $(n \in \mathbb{Z}^+)$. Hence $f_n - h_n$ is constant on I. By (6.9), we have $h_n \to h$ on I as $n \to \infty$. Therefore

$$f - h = \lim_{n \to \infty} (f_n - h_n)$$

is constant on I. It follows that $f' = h' = g$ on I, as was to be proved. □

7. Certain Important Functions

A good point of departure for the study of the exponential function is obtained from an analysis of its rate of change. Intuitively the exponential function is that differentiable function on $\mathbb{R}$, normalized to have the value 1 at 0, which increases at a rate equal to its size. In other words, it is the solution of the differential equation $Df = f$ normalized by the condition $f(0) = 1$. Clearly, such a function f, it it exists, is infinitely differentiable, and $f^{(n)}(0) = 1$ for all nonnegative integers n. Thus the Taylor series for f about $x = 0$ is

$$\sum_{n=0}^{\infty} \frac{x^n}{n!}. \tag{7.1}$$

For each interval $[-c, c]$ with $c > 0$, $c^n \|f\|_{[-c,c]}/n! \to 0$ as $n \to \infty$, and hence the series converges to f on $[-c, c]$. Therefore if f exists, it is unique, and is given by (7.1). On the other hand, (7.1) converges on $\mathbb{R}$, by the ratio test. We therefore define the *exponential function* exp by means of the series (7.1):

$$\exp(x) \equiv \sum_{n=0}^{\infty} \frac{x^n}{n!} \qquad (x \in \mathbb{R}). \tag{7.2}$$

Clearly $\exp(0)=1$. Also,

(7.3) $$\frac{d\exp(x)}{dx}=\sum_{n=0}^{\infty}\frac{nx^{n-1}}{n!}=\sum_{n=1}^{\infty}\frac{x^{n-1}}{(n-1)!}=\exp(x),$$

by (6.10). The function exp has the desired properties, and is unique.

(7.4) **Proposition.** *The function* exp *satisfies the* functional equation

(7.4.1) $$\exp(x+y)=\exp(x)\exp(y) \qquad (x, y\in\mathbb{R}).$$

Proof: Let z be any positive real number. Then $\exp(z)\neq 0$, by (7.2). Define the differentiable function $f\colon \mathbb{R}\to\mathbb{R}$ by

$$f(x)\equiv(\exp(z))^{-1}\exp(x+z).$$

The derivative of f is

$$\frac{df(x)}{dx}=(\exp(z))^{-1}\exp(x+z)\frac{d(x+z)}{dx}=f(x),$$

by (5.4). Also, $f(0)=1$. By the uniqueness of the exponential function, $f(x)=\exp(x)$ – that is,

(7.4.2) $$\exp(x+z)=\exp(x)\exp(z).$$

Consider now arbitrary real numbers x and y. Choose z with $z>0$ and $z+y>0$. Then, by (7.4.2),

$$\begin{aligned}\exp(x+y)\exp(z)&=\exp(x+y+z)\\&=\exp(x)\exp(y+z)=\exp(x)\exp(y)\exp(z),\end{aligned}$$

which is equivalent to (7.4.1). □

It can be shown that exp is the unique continuous function from $\mathbb{R}$ to $\mathbb{R}$ that has the value 1 at 0 and satisfies (7.4.1).

Equality (7.4.1) implies that $\exp(-x)=(\exp(x))^{-1}$. Also,

(7.5) $$\exp(x)>0 \qquad (x\in\mathbb{R}).$$

For, given x in $\mathbb{R}$ and choosing $z>0$ so that $x+z>0$, we have $\exp(z)\geqq 1$ and $\exp(x+z)\geqq 1$; whence

$$\exp(x)=\exp(x+z)(\exp(z))^{-1}>0.$$

The theory of the exponential function can also be approached as follows. To construct a function f with $f(0)=1$ and $f'=f$, notice that $f'=f$ implies, by the mean-value theorem, that $f(x+\delta)$ is approximately equal to $f(x)+\delta f'(x)=(1+\delta)f(x)$ when δ is small. Therefore, if t is any real number and n is a large positive integer, $f(t/n)$

is approximately $1+(t/n)$, $f(2t/n)=f(t/n+t/n)$ is approximately $(1+(t/n))^2$, and so forth, and thus $f(t)=f(n(t/n))$ is approximately $(1+(t/n))^n$. This heuristic argument suggests that we define

$$\exp(t)=\lim_{n\to\infty}\left(1+\frac{t}{n}\right)^n.$$

It is not hard to see that

$$\left(1+\frac{t}{n}\right)^n=1+n\frac{t}{n}+\frac{n(n-1)}{2!}\left(\frac{t}{n}\right)^2+\ldots$$

converges on $\mathbb{R}$ to $\sum_{n=0}^{\infty} t^n/n!$ as $n\to\infty$. This approach to the theory of the exponential function would therefore lead to the same series representation.

We want the logarithmic function to be the inverse function to the exponential function, but we shall not define it in that way. Arguing heuristically on the basis of the chain rule (5.4), for each $x>0$ we want

$$1=\frac{dx}{dx}=\frac{d}{dx}(\exp(\ln(x))=\exp(\ln(x))\frac{d}{dx}\ln(x)=x\frac{d}{dx}\ln(x).$$

This gives

(7.6) $$\frac{d}{dx}\ln(x)=x^{-1} \qquad (x>0).$$

We therefore define $\ln(x)$ to be the integral of x^{-1}; specifically

(7.7) $$\ln(x)\equiv\int_1^x t^{-1}\,dt \qquad (x>0).$$

By (6.8), ln is differentiable on $(0,\infty)$ and (7.6) is valid.

Let y be any positive real number. By (5.4.1), the derivative of the function

(7.8) $$x\mapsto\ln(xy)-\ln(y)$$

is x^{-1}. Also, (7.8) vanishes at 1. By the last statement of (6.8), $\ln(x)=\ln(xy)-\ln(y)$; in other words,

(7.9) $$\ln(xy)=\ln(x)+\ln(y) \qquad (x,y>0).$$

This is the functional equation for the logarithmic function, corresponding to the functional equation (7.4.1) for the exponential function.

By (7.5) and (5.4), the composite function $\ln\circ\exp$ exists and is differentiable everywhere on $\mathbb{R}$, and

$$\frac{d}{dx}(\ln(\exp(x))=1 \qquad (x\in\mathbb{R}).$$

Since also $\ln(\exp(0))=\ln(1)=0$, we have

$$\text{(7.10)} \qquad \ln(\exp(x))=x \qquad (x\in\mathbb{R}),$$

by the last part of (6.8). Consider $x>0$ and write $y\equiv\exp(\ln(x))$. Then

$$\ln(y)=\ln(\exp(\ln(x)))=\ln(x),$$

by (7.10). Thus

$$0=\ln(y)-\ln(x)=\int_x^y t^{-1}\,dt.$$

If $y>x$, this gives $0\geqq y^{-1}(y-x)$, a contradiction. By (2.18), we therefore have $y\leqq x$. Similarly, $x\leqq y$; whence $x=y$. Thus $\exp(\ln(x))=x$ for all $x>0$. It follows that the functions $\exp\colon\mathbb{R}\to\mathbb{R}^+$ and $\ln\colon\mathbb{R}^+\to\mathbb{R}$ are inverse to each other.

For any $a>0$ we now define

$$a^x\equiv\exp(x\ln(a)) \qquad (x\in\mathbb{R}).$$

We leave the reader to confirm that the map $x\mapsto a^x$ has the familiar properties.

Note that for a fixed $x>0$ the map $t\mapsto t^x$ of $\mathbb{R}^+$ into $\mathbb{R}^+$ extends to a continuous map of $\mathbb{R}^{0+}$ into $\mathbb{R}^{0+}$ with $0^x=0$; this follows from the fact that t^x is arbitrarily small for all positive numbers t sufficiently close to 0.

The trigonometric functions sin and cos also can be approached via an intuitive analysis of their rates of change. From this analysis we are led to believe that

$$\text{(7.11)} \qquad \frac{d}{dx}\cos x=-\sin x, \qquad \frac{d}{dx}\sin x=\cos x \qquad (x\in\mathbb{R}).$$

We therefore define these functions by power series constructed in such a way that (7.11) will hold. Remembering that $\cos(0)=1$ and $\sin(0)=0$, we are forced to define

$$\text{(7.12)} \qquad \left.\begin{aligned}\cos x&\equiv\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n}}{(2n)!},\\ \sin x&\equiv\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n+1}}{(2n+1)!}\end{aligned}\right\} \quad (x\in\mathbb{R}).$$

Theorem (6.10) implies that (7.11) is valid.

(7.13) **Proposition.** *The functions* sin *and* cos *satisfy the functional equations*

(7.13.1) $$\sin(x+y)=\sin x\cos y+\cos x\sin y$$

(7.13.2) $$\cos(x+y)=\cos x\cos y-\sin x\sin y$$

for all x, y *in* $\mathbb{R}$.

Proof: Consider both sides of (7.13.1) as functions of x, say $f(x)$ and $g(x)$. The functions f and g have Taylor series about $x=0$ which converge on $\mathbb{R}$. The coefficients of these series are expressed in terms of the various derivatives of f and g at $x=0$. Therefore to show that $f=g$ on $\mathbb{R}$ it is enough to show that $f^{(n)}(0)=g^{(n)}(0)$ for all nonnegative integers n. Since $f^{(2)}=-f$ and $g^{(2)}=-g$, it is enough to notice that

$$f(0)=\sin y=\sin 0\cos y+\cos 0\sin y=g(0)$$

and

$$f'(0)=\cos y=\cos 0\cos y+(-\sin 0)\sin y=g'(0).$$

Therefore (7.13.1) is valid. Equation (7.13.2) follows from (7.13.1) by differentiation with respect to x. □

Putting $y=-x$ in (7.13.2) gives the useful identity

(7.14) $$\cos^2 x+\sin^2 x=1.$$

Further study of the trigonometric functions depends on properties of the number π, which we construct as twice the first positive zero of the cosine function.

(7.15) **Theorem.** *The sequence* (x_n) *of real numbers defined inductively by the equations*

(i) $$x_1\equiv 1$$

(ii) $$x_{n+1}\equiv x_n+\cos x_n \qquad (n\geqq 1)$$

is increasing, and converges to a limit $\pi/2$ *such that* $\cos(\pi/2)=0$ *and* $\cos x>0$ *for all* x *with* $0\leqq x<\pi/2$.

Proof: To begin with, we prove that

(7.15.1) $$\cos t>0 \qquad (0\leqq t\leqq x_n,\ n\in\mathbb{Z}^+).$$

Consider first the case $n=1$, when we have $0\leqq t\leqq x_1=1$. Then

$$\cos t=\left(1-\frac{t^2}{2!}\right)+\left(\frac{t^4}{4!}-\frac{t^6}{6!}\right)+\ldots\geqq 1-\tfrac{1}{2}>0.$$

Next assume that (7.15.1) is true for a particular value of n; then $x_{n+1} > x_n$. Consider a value of t with $x_n < t \leqq x_{n+1}$. Now $|\sin x| \leqq 1$ for all x in $\mathbb{R}$, by (7.14); and $|\sin x_n| < 1$, because $\cos x_n > 0$. Hence

$$\cos t = \cos x_n - \int_{x_n}^{t} \sin x\, dx > \cos x_n - (x_{n+1} - x_n) = 0.$$

Since cos is a continuous function and $\cos t > 0$ whenever $0 \leqq t \leqq x_n$ or $x_n < t \leqq x_{n+1}$, $\cos t > 0$ whenever $0 \leqq t \leqq x_{n+1}$. By induction, (7.15.1) holds for all n. Since $\cos x_n > 0$ for all n, the sequence (x_n) is increasing.

From the mean-value theorem and (7.15.1) we obtain

$$\sin x_n \geqq \sin 0 = 0 \tag{7.15.2}$$

for all n, and $\sin t \geqq \sin 1$ whenever $1 \leqq t \leqq x_n$ for some n. Also by the mean-value theorem, for each $n \geqq 2$ there exists t in $[x_{n-1}, x_n]$ such that

$$\begin{aligned} x_{n+1} - x_n &= x_n + \cos x_n - (x_{n-1} + \cos x_{n-1}) \\ &= x_n - x_{n-1} + \cos x_n - \cos x_{n-1} \end{aligned}$$

is arbitrarily near to

$$x_n - x_{n-1} - (\sin t)(x_n - x_{n-1}) \leqq (x_n - x_{n-1})(1 - \sin 1).$$

Therefore

$$x_{n+1} - x_n \leqq (1 - \sin 1)(x_n - x_{n-1}).$$

This gives

$$x_{n+1} - x_n \leqq (1 - \sin 1)^{n-1}(x_2 - x_1)$$

for all $n \geqq 1$. Thus (x_n) is a Cauchy sequence, whose limit we denote by $\pi/2$.

If $0 \leqq x < \pi/2$, then $x < x_n$ for some n, and thus $\cos x > 0$, by (7.15.1). Taking limits on both sides of (ii) as $n \to \infty$ we see that $\cos(\pi/2) = 0$. □

By (7.15.2), $\sin(\pi/2) \geqq 0$. Since $\cos(\pi/2) = 0$, it follows from (7.14) that $\sin(\pi/2) = 1$, and hence from (7.13.1) that $\sin(x + \pi/2) = \cos x$ for each x in $\mathbb{R}$. Similarly, $\sin(\pi/2 - x) = \cos x$ and $\cos(x + \pi/2) = -\sin x$. Hence $\sin(x + \pi) = \cos(x + \pi/2) = -\sin x$, and thus $\sin(x + 2\pi) = \sin x$. Similarly, $\cos(x + \pi) = -\cos x$ and $\cos(x + 2\pi) = \cos x$. On the other hand, if $\alpha > 0$ and $\sin(x + \alpha) = \sin x$ for all x, then $\cos \alpha = \sin(\pi/2 + \alpha) = \sin(\pi/2) = 1$. Since (as the reader may prove) cos is a strictly decreasing function on $[0, \pi]$ and a strictly increasing function on $[\pi, 2\pi]$, and since $\cos 0 = \cos 2\pi = 1$, it follows that $\alpha \geqq 2\pi$. Thus the function sin, and similarly the function cos, is periodic, with period 2π.

Next we study the function arc sin, which is inverse to the function sin. For motivation we take the derivative of the equation

$$\sin(\arcsin x) = x$$

and get

$$\cos(\arcsin x)\frac{d}{dx}(\arcsin x) = 1.$$

Since

$$1 = \cos^2(\arcsin x) + \sin^2(\arcsin x) = \cos^2(\arcsin x) + x^2,$$

it follows that

$$\frac{d}{dx}(\arcsin x) = \pm(1-x^2)^{-\frac{1}{2}}.$$

With this motivation, we define

$$\arcsin x \equiv \int_0^x (1-t^2)^{-\frac{1}{2}}\,dt \qquad (-1<x<1).$$

Then arc sin is differentiable on $(-1, 1)$, and

$$\frac{d}{dx}(\arcsin x) = (1-x^2)^{-\frac{1}{2}}.$$

To discuss the range of arc sin, we first observe that the function sin is strictly increasing on $[-\pi/2, \pi/2]$. Since $\sin(-\pi/2) = -1$ and $\sin(\pi/2) = 1$, it follows from the remarks after (4.8), and the continuity of sin, that for each x in $(-1, 1)$ there exists y in $(-\pi/2, \pi/2)$ with $x = \sin y$. For any y in $(-\pi/2, \pi/2)$ we have

$$\frac{d}{dy}\arcsin(\sin y) = (1-\sin^2 y)^{-\frac{1}{2}}\cos y = 1.$$

Hence, as $\arcsin(\sin 0) = 0$,

(7.16) $$\arcsin(\sin y) = y \qquad (-\pi/2 < y < \pi/2).$$

It follows that the function arc sin maps the interval $(-1, 1)$ into the interval $(-\pi/2, \pi/2)$. Moreover, if $-1<x<1$ and $z = \sin(\arcsin x)$, then $\arcsin z = \arcsin x$, by (7.16); so that

$$\int_x^z (1-t^2)^{-\frac{1}{2}}\,dt = 0.$$

Therefore $x = z = \sin(\arcsin x)$. It follows that the functions

$$\sin : (-\pi/2, \pi/2) \to (-1, 1)$$

and

$$\arcsin : (-1, 1) \to (-\pi/2, \pi/2)$$

are inverses.

Problems

The word "*not*" will be used throughout in the sense of the discussion at the end of Section 2.

1. Construct a set A such that $x=y$ for all elements x and y of A, but A is *not* subfinite or void.

2. Construct a mapping f from a set A to a set B such that f is onto B, but there does *not* exist a mapping $g: B \to A$ with $f(g(b))=b$ for all b in B.

3. Prove that $(x, y) \mapsto xy$ is a function from $\mathbb{R} \times \mathbb{R}$ into $\mathbb{R}$.

4. Let x and y be real numbers such that $x \neq y$ entails $0=1$. Show that $x=y$.

5. Show that if $x_1, \ldots, x_n$ are real numbers such that $x_1 \ldots x_n < 0$, then $x_i < 0$ for some i.

6. Call a pair (S, T) of nonvoid subsets of the set Q of rational numbers a *Dedekind cut* if $s<t$ for all s in S and t in T, and if for arbitrary rational numbers x, y with $x<y$ either $x \in S$ or $y \in T$. Show that for each Dedekind cut (S, T) there exists a unique real number $x(S, T)$ such that $s \leq x(S, T)$ for all s in S, and $t \geq x(S, T)$ for all t in T. Show that for each real number x there exists a Dedekind cut (S, T) with $x(S, T)=x$.

7. Show that the real number system can be constructed from the Dedekind cuts. In other words, define directly equality, order, addition, and multiplication of Dedekind cuts in such a way that the map $(S, T) \mapsto x(S, T)$ from the set of all Dedekind cuts into $\mathbb{R}$ is a bijection which preserves order and the operations of addition and multiplication.

8. Construct a real number that is *not* positive or negative or equal to 0.

9. Construct a real number that does *not* have a decimal expansion.

10. Construct a real number that is *not* rational and not irrational. (A real number x is *irrational* if $x \neq r$ for each rational number r.)

11. Construct a continuous function $f: [0,1] \to \mathbb{R}$ such that there does *not* exist a point x in $[0,1]$ with $f(x)$ equal to the supremum of f.

12. Show that a continuous function on a compact interval admits a modulus of continuity which is a continuous function.

13. A mapping f of an interval I into $\mathbb{R}$ is *weakly injective* if $x=y$ whenever $f(x)=f(y)$, and is *injective* if $f(x) \neq f(y)$ whenever $x \neq y$. Prove that the following statements are equivalent.

(i) Every weakly injective map $f\colon [0,1] \to \mathbb{R}$ is injective.
(ii) If x is a real number such that the equality $x=0$ is impossible, then $x \neq 0$.
(iii) For each sequence (x_n) in $\{0,1\}$, if it is impossible that $x_n=0$ for all n, then there exists n with $x_n=1$ (*Markov's principle*).

(There is no known proof or counterexample for Markov's principle.)

14. Let $f\colon [0,1] \to \mathbb{R}$ be a continuous function with $f(0)<f(1)$. Show that there is a sequence (y_n) of elements of $[f(0), f(1)]$ such that if $f(0) \leqq y \leqq f(1)$ and if $y \neq y_n$ for each n, then there exists x in $[0,1]$ with $f(x)=y$.

15. Let $f\colon [0,1] \to \mathbb{R}$ be a continuous function with $f(0)<0$ and $f(1)>0$, such that for arbitrary real numbers a and b with $0 \leqq a < b \leqq 1$ there exists x in $[a,b]$ with $f(x) \neq 0$. Show that there exists x in $[0,1]$ with $f(x)=0$.

16. Let $f\colon [0,1] \to \mathbb{R}$ be a continuous function with $f(0)<0$ and $f(1)>0$, such that there exists a positive integer n with $|f(x)|+|f'(x)| + \ldots + |f^{(n)}(x)| > 0$ for all x in $[0,1]$. Show that there exists x in $[0,1]$ with $f(x)=0$.

17. Let $f\colon [0,1] \to \mathbb{R}$ be a polynomial function with $f(0)<0$ and $f(1)>0$. Show that there exists x in $[0,1]$ with $f(x)=0$.

18. Show that if $f\colon [0,1] \to [0,1]$ is a continuous function, then for each $\varepsilon>0$ there exists x in $[0,1]$ with $|f(x)-x|<\varepsilon$.

19. Let $f\colon [0,1] \to \mathbb{R}$ be differentiable and *locally nonconstant*, in the sense that if $x, y \in [0,1]$ and $x<y$, then there exists z in (x,y) with $f(z) \neq f(x)$. Suppose also that $f(0)=f(1)$. Prove that there exists x in $[0,1]$ with $f'(x)=0$.

20. Construct a continuous function $f\colon [0,1] \to \mathbb{R}$ such that f' does not exist on any proper compact subinterval of $[0,1]$. (In other words,

the assumption that f' exists on a proper compact subinterval leads to a contradiction.)

21. Let f and g be continuous functions on the compact interval $[a,b]$, such that $g(x)\geqq 0$ for all x in $[a,b]$. Prove that for each $\varepsilon>0$ there exists c in $[a,b]$ such that

$$\left|\int_a^b f(x)\,g(x)\,dx-f(c)\int_a^b g(x)\,dx\right|<\varepsilon.$$

22. Prove that under the hypotheses of Taylor's theorem,

$$R=\int_a^b \frac{f^{(n+1)}(t)}{n!}(b-t)^n\,dt.$$

Hence prove that for each $\varepsilon>0$ there exists c with $\min\{a,b\}\leqq c \leqq \max\{a,b\}$, such that

$$\left|R-\frac{f^{(n+1)}(c)}{(n+1)!}(b-a)^{n+1}\right|<\varepsilon.$$

23. Prove that if f is infinitely differentiable on the open interval $I\equiv(a-t,a+t)$, and if

$$\frac{r^n f^{(n)}}{n!}\to 0 \quad \text{as } n\to\infty \ (0<r<t),$$

then the Taylor series for f about a converges to f on I.

Notes

As the definitions of Section 1 indicate, a single concept of classical mathematics may split into two or more distinct concepts when looked at constructively. For example, the distinction between finite and subfinite sets does not exist classically; nor does the distinction between weakly injective and injective mappings (Problem 13).

In [7] what we have called a map of A *onto* B is called a *surjection;* a function $f: A\to B$ is there said to map A onto B only if it has a right inverse. Although there is a meaningful distinction involved here (see Problem 2), we prefer to use the simple expression 'onto' for the more general situation. When a right inverse map exists we can say so explicitly if we need to.

Problem 2 indicates that the axiom of choice is *not* true in constructive mathematics. In certain formal systems attempting to describe constructive mathematics it can be shown that the axiom of choice entails the principle of omniscience [42].

An alternative to Definition (2.1) would be to define real numbers by the method of Dedekind cuts (see Problem 6). Another possibility would be to define a real number as a sequence (x_n) of rational numbers *and* a sequence (N_k) of integers, such that $|x_m - x_n| \leqq k^{-1}$ whenever $m, n \geqq N_k$.

Notice that a real number *is* a regular sequence of rational numbers, not an equivalence class of regular sequences of rational numbers. To define a real number to be such an equivalence class would be either pointless or incorrect: pointless (but correct) if the equivalence class is required to be specified by giving some particular regular sequence of rational numbers that belongs to it, and incorrect otherwise.

Precisely what is meant by a subset (in particular, what it means for $\mathbb{R}^+$ and $\mathbb{R}^{0+}$ to be subsets of $\mathbb{R}$) will be discussed in Chapter 3.

In classical mathematics, a proof like that of Lemma (2.18), in which we assume a certain hypothesis, derive a contradiction, and therefore conclude that the negation of our hypothesis is true, is called an indirect proof. It is necessary to exercise caution in thinking of Lemma (2.18) in this way. Although $x \leqq y$ is equivalent to the negation of $x > y$, it is *not* true that $x > y$ is equivalent to the negation of $x \leqq y$.

There is a paradox growing out of Theorem (2.19) which the reader should resolve. Since every regular sequence of rational numbers can presumably be described by a phrase in the English language, and since the phrases in the English language can be sequentially ordered, the regular sequences of rational numbers can be sequentially ordered, in contradiction to Theorem (2.19).

The counterexamples following Theorem (2.19) are deceitful. The reader is asked not to form the impression that the purpose of constructive mathematics is to consider pathological numbers like the ones discussed here. The only reason for discussing such numbers is to show that certain statements (in this case, the statement that for each x in $\mathbb{R}^{0+}$ either $x > 0$ or $x = 0$) are not constructively valid. To appreciate this point better, the reader should examine the counterexamples given in Chapter 1, where the treatment is more detailed.

Sometimes a mathematician, when confronted with a counterexample of this sort, will say he is not interested in real numbers that have been artificially twisted out of shape. One reply to this is to point out that the set of real numbers x such that either $x > 0$ or $x \leqq 0$ is *not* complete (nor is it closed under addition).

The 'choice' involved in the proof of (b) of Proposition (3.4) is performed according to a definite rule.

Strictly speaking, a continuous function on a compact interval I is a pair (f, ω) consisting of a function $f: I \to \mathbb{R}$ and a modulus of continuity ω for f on I; two continuous functions (f_1, ω_1) and (f_2, ω_2) are *equal* if f_1 and f_2 are equal functions. In practice, as in Definition (4.5), we call f itself a continuous function, in the spirit of the remarks preceding Proposition (2.9).

Notice that divergence is a positive notion.

As we remarked in Chapter 1, a nonvoid bounded set of real numbers need *not* have a least upper bound. This, of course, makes some parts of constructive mathematics harder than the corresponding classical material, but it also makes some parts more interesting.

The concept introduced in Definition (4.5) is classically called uniform continuity. We abbreviate this to continuity, since the classical version of continuity (that is, pointwise continuity) is not used anywhere in the book. Classically, one proves that on a compact interval, pointwise continuity implies uniform continuity. Constructively, there is no known proof of this result, and so we assume uniform continuity from the start. Nothing essential is lost by this assumption, and a certain simplicity is gained. Similar remarks apply to Definitions (4.9) and (5.1).

It would be more correct to call Lemma (5.5) a constructive substitute for Rolle's theorem: the classical result states that under the hypotheses of Lemma (5.5), $f'(x)=0$ for some x in (a, b). It follows from Problem 19 that as long as we are concerned with the familiar elementary functions of real analysis, the sharp classical form of Rolle's theorem is constructively valid.

The proof of Theorem (5.10) illustrates the minor inconveniences caused by not being able to compare arbitrary real numbers.

For Riemann integration of possibly discontinuous functions see Problem 10 of Chapter 6.