Chapter 1

Functions

1.1 Sets

If x is an **element** (or **member**) of a set X, we write $x \in X$. Set membership, by definition, does not involve multiplicity or order: thus the set $\{a, a\}$ is really $\{a\}$ and the sets $\{a, b\}$, $\{b, a\}$ are identical. The set with no elements is called the **empty set** and is denoted by $\varnothing$ or $\{\}$.

The elements of a set are determined by a common property such as that in the following problem.

Problem 1.1 (IMO 1971)**.** *Prove that for every natural number m, there exists a finite set S_m of points in a plane with the following property: For every point A in S_m, there are exactly m points in S_m which are at unit distance from A.*

Solution ([6], December 1971, p. 40). For $m = 1$ we take S_1 to be the end-points of a unit interval. For $m = 2$ we take S_2 to be the vertices of an equilateral triangle.

For $m > 2$ we construct S_m by induction. Given $S_k = \{\mathrm{A}_1, \mathrm{A}_2, \dots\}$, let S_{k+1} be the set that includes all points of S_k and Σ_k, where $\Sigma_k = \{\mathrm{A}'_1, \mathrm{A}'_2, \dots\}$ is the set of points obtained by translating the points of S_k by unit vector $\vec{u}$: $\mathrm{A}_i \overset{\vec{u}}{\mapsto} \mathrm{A}'_i$. That is, for each value of the label i, the points A_i, A'_i are at a unit distance. Each point in S_k and Σ_k has exactly k points at unit distance. By construction then, each point in S_{k+1} has at least $k+1$ points at unit distance. To ensure that there will be no more than $k + 1$ points at unit distance, the vector $\vec{u}$ must be selected appropriately to avoid any unwanted equilateral triangles. For each point A_i in S_k we draw a circle C_i of radius 1 and center A_i. Let P_i^1, P_i^2, ... be the intersection points of C_i with the other circles. The (finite) set of vectors $\overrightarrow{\mathrm{A}_i\mathrm{P}_i^\alpha}$ for all values of i and α are the vectors to be excluded as candidates for $\vec{u}$. $\square$

If every element of the set A also belongs to the set B, we say that A is a **subset** of B, and we write $A \subset B$. If two sets A, B satisfy $A \subset B$ and $B \subset A$, the two sets are **equal**; we write $A = B$.

The notation $A \subset B$ allows the possibility that $A = B$; in the rare event that we wish to exclude this possibility, we write $A \subsetneqq B$, and say that A is a **proper subset** of B. The notation $B \supset A$ is equivalent to $A \subset B$, and says that B is a **superset** of A.

Let $A, B \subset S$. The **union** of A and B, denoted by $A \cup B$, is defined as the set of $x \in S$ such that $x \in A$ or $x \in B$.

Let $A, B \subset S$. The **intersection** of A and B, denoted by $A \cap B$, is defined as the set of elements $x \in S$ such that $x \in A$ and $x \in B$. Two sets A, B are called **disjoint** if $A \cap B = \varnothing$.

Let $A, B \subset S$. The **difference** $B \setminus A$ is defined as the set of elements $x \in B$ that do not belong to A. In the particular case that $A \subset B$, we call $B \setminus A$ the **complement of** A **in** B, or simply the **complement of** A (if B can be inferred from the context). The complement of A is often denoted by $A^{\complement}$.

The **power set** of S, denoted by $\mathcal{P}(S)$, is the set of all subsets of S:

$$\mathcal{P}(S) := \{A : A \subset S\}.$$

The **Cartesian product** of an n-tuple of sets $(A_1, A_2, \ldots, A_n)$ is defined as the set whose elements consist of all n-tuples whose i-th entry is an element of A_i: in symbols,

$$A_1 \times A_2 \times \cdots \times A_n := \{\, (a_1, a_2, \ldots, a_n) : a_i \in A_i \text{ for } i = 1, 2, \ldots, n \,\}. \quad (1.1)$$

If $A_1 = A_2 = \cdots = A_n =: A$, we usually write

$$\underbrace{A \times A \times \cdots \times A}_{n \text{ times}} =: A^n.$$

Problem 1.2 (Putnam 1961). *Let Ω be a set of n points, where $n > 2$. Let Σ be a nonempty subset of $\mathcal{P}(\Omega)$ that is closed*[1] *with respect to unions, intersections, and complements. If k is the number of members of Σ, what are the possible values of k? Give a proof.*

Solution [21]. Since Σ is nonempty, there is a set A in Σ. Its complement $A^{\complement}$ is also in Σ and so is the intersection $A \cap A^{\complement} = \varnothing$ and the union $A \cup A^{\complement} = \Omega$.

Let $\mathcal{M}$ now be the following subset of Σ:

$$\mathcal{M} = \{M \in \Sigma : \text{there is no } A \in \Sigma,\ A \neq \varnothing, \text{ such that } A \subsetneqq M\}.$$

[1]That is, if A and B are members of Σ, then so are $A \cup B$, $A \cap B$, $A^{\complement} = \Omega \setminus A$, and $B^{\complement} = \Omega \setminus B$. This information was provided in the test.

In other words, $\mathcal{M}$ contains the 'minimal' sets of Σ. Obviously, $\mathcal{M}$ is a finite set.

Given two distinct minimal sets M_1 and M_2, they must be disjoint, since $M_1 \cap M_2 = B \in \Sigma$ and thus $B \subset M_1, M_2$, which, by definition, cannot be true for any nonempty set.

Now let M be the union of all minimal sets:

$$M = \bigcup_{M_i \in \mathcal{M}} M_i.$$

Its complement $M^{\complement}$ can be $\varnothing$ or some nonempty set. If it is nonempty, there is a minimal set M_0 such that $M_0 \subset M^{\complement}$; thus $M_0 \cap M^{\complement} \neq \varnothing$. However, $M_0 \cap M^{\complement} = M_0 \cap (\Omega \setminus M_0 \setminus \dots) = \varnothing$. A contradiction. Therefore $M^{\complement} = \varnothing$ or $M = \Omega$.

Finally, let C be any random set in Σ. Then

$$\begin{aligned} C = C \cap \Omega &= C \cap \Big(\bigcup_{M_i \in \mathcal{M}} M_i \Big) \\ &= \bigcup_{M_i \in \mathcal{M}} (C \cap M_i). \end{aligned}$$

Each of the intersections $C \cap M_i$ is either the empty set or M_i.

To summarize, we have proved that: Given a partition of Ω in p disjoint sets $M_1, M_2, \dots, M_p$, where $1 \leq p \leq n$, each member of Σ can be written as the union of a subcollection of the sets $M_1, M_2, \dots, M_p$. And since for each set M_i we have two choices (to include it or not to include it), we have 2^p possible subcollections. Therefore, $k = 2^p$, with all values $p = 1, 2, \dots, n$ allowed. □

Topological Concepts

Definition 1.3. Let $\mathcal{O}$ be a subset of the power set $\mathcal{P}(X)$ satisfying these properties:

1. $\varnothing \in \mathcal{O}, \ X \in \mathcal{O}$.
2. The intersection of a finite number of elements of $\mathcal{O}$ also belongs to $\mathcal{O}$.
3. The union of any number of elements of $\mathcal{O}$ also belongs to $\mathcal{O}$.

We say that $\mathcal{O}$ is a **topology** on X. The pair $(X, \mathcal{O})$ is called a **topological space**. We also say that X is a topological space, especially if the topology can be inferred from the context. It is common to refer to the elements of a topological space as **points**. The elements of the topology $\mathcal{O}$ are called the **open sets** of X (with respect to the topology). The complement $X \setminus O$ of an open set O is called a **closed** set.

Definition 1.4. Let $(X, \mathcal{O})$ be a topological space. A set $N \subset X$ is called a **neighborhood** of the point $p \in X$ if p lies in N and there exists an open set $O \in \mathcal{O}$ that contains p and is contained in N: in symbols, $p \in O \subset N$.

Nothing in this definition requires a neighborhood to be open. However, when we refer to a **neighborhood** of a point $a \in \mathbb{R}$, we shall have in mind—unless otherwise specified—an open interval of the form $(a-\delta, a+\delta)$, where $\delta > 0$ is called the **radius** of the neighborhood.

Given a topological space X and a subset S of X, the following definitions cover some of the most interesting points and sets related to S:

1. A point ℓ is called a **limit point** of the set S if every neighborhood of ℓ contains at least one point of S different from ℓ. The set S' of all limit points of S is called the **derived set** of S.
2. A point i is called an **interior point** if there exists a neighborhood of i that is contained in S. The set $\operatorname{Int} S$ of all interior points of S is called the **interior set** of S.
3. A point e is called an **exterior point** of S if there exists a neighborhood of e that is disjoint from S. The set $\operatorname{Ext} S$ of all exterior points of S is called the **exterior set** of S.
4. A point b is called a **boundary point** of S if every neighborhood of b intersects both S and $X \setminus S$. The set ∂S of all boundary points of S is called the **boundary** of S.
5. A point i_0 is called an **isolated point** of S if $i_0 \in S$ and i_0 lies in the exterior of $S \setminus \{i_0\}$. In other words, there is a neighborhood $N(i_0)$ of i_0 that does not contain any other point of S except i_0.
6. The smallest closed set $\bar{S}$ containing S is called the **closure** of S.
7. A subset S of X is called **dense** in X if $\bar{S} = X$.

There are many useful relations among the sets defined above. Let X be a topological space and let A, B, O, C be subsets of X. Then:

1. C is closed if and only if $\overline{C} = C$.
2. C is closed if and only if $C' \subset C$.
3. C is closed if and only if $\partial C \subset C$.
4. O is open if and only if $\operatorname{Int} O = O$.
5. O is open if and only if $\partial O \cap O = \varnothing$.
6. $\overline{A} = A \cup A'$.
7. $\overline{A \cup B} = \overline{A} \cup \overline{B}$.

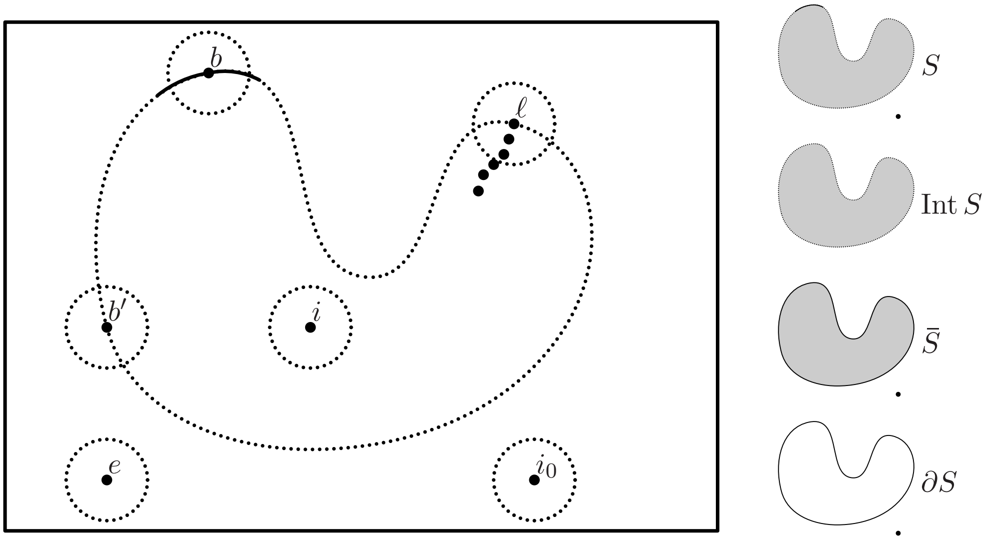


Figure 1.1: Terminology of points and sets in a topological space. The left side focuses on points: i is an interior point of the set S shown; e is exterior to S (and equivalently, interior to the complement of S); b and b' are boundary, the first being in S, the second in its complement. The set S was chosen to contain the isolated point i_0 in addition to the irregularly shaped region: note that $i_0 \in S \cap \overline{S \smallsetminus \{i_0\}}$. A limit point ℓ of S is the limit of a converging sequence of points in S distinct from ℓ.

8. $\overline{\overline{A}} = \overline{A}$.
9. $\overline{A \cap B} \subset \overline{A} \cap \overline{B}$.
10. $\overline{A \smallsetminus B} \subset \overline{A} \smallsetminus \overline{B}$.
11. $\overline{X \smallsetminus A} = \overline{X} - \operatorname{Int} A$.
12. $\overline{A} = X \smallsetminus \operatorname{Ext} A$.
13. $\partial A = X \smallsetminus (\operatorname{Int} A \cup \operatorname{Ext} A)$.
14. $\partial A = \overline{A} \cap \overline{(X \smallsetminus A)}$.
15. $\partial A = \overline{A} \smallsetminus \operatorname{Int} A$.

Example 1.5. An interesting set is the set of rational numbers

$$\mathbb{Q} = \{m/n : m \in \mathbb{Z},\, n \in \mathbb{Z}^*\}$$

as a subset of $\mathbb{R}$. Then $\overline{\mathbb{Q}} = \mathbb{R}$, $\partial\mathbb{Q} = \mathbb{R}$, $\operatorname{Int}\mathbb{Q} = \varnothing$. Since $\operatorname{Int}\mathbb{Q} \neq \mathbb{Q}$, $\mathbb{Q}$ is not open; since $\partial\mathbb{Q} \supset \mathbb{Q}$, $\mathbb{Q}$ is not closed either. □

Problem 1.6 (Putnam 1956)**.** *Suppose that each set X of points in the plane has an associated set $\overline{X}$ of points called its cover. Suppose further that*

$$\overline{X \cup Y} \supset \overline{\overline{X}} \cup \overline{Y} \cup Y, \tag{1.2}$$

where $\cup$ *designates point set sum (or union) and* $\supset$ *denotes set inclusion. Prove*

(a) $\overline{X} \supset X$,

(b) $\overline{\overline{X}} = \overline{X}$,

(c) $X \supset Y$ *implies* $\overline{X} \supset \overline{Y}$.

Prove conversely that (a), (b), and (c) imply (1.2).

Solution. In (1.2) we set $X = Y$ to find

$$\overline{X \cup X} \supset \overline{\overline{X}} \cup \overline{X} \cup X \Rightarrow \overline{X} \supset \overline{\overline{X}} \cup \overline{X} \cup X.$$

From the last relation we conclude that

$$\overline{X} \supset X$$

which is the relation (a) sought and

$$\overline{X} \supset \overline{\overline{X}}.$$

However, relation (a), after replacing X by $\overline{X}$, implies

$$\overline{\overline{X}} \supset \overline{X}.$$

The last two relations give the desired relation (b).

Finally, if $X \supset Y$ then $X \cup Y = X$. Therefore relation (1.2) is equivalent to

$$\overline{X} \supset \overline{\overline{X}} \cup \overline{Y} \cup Y,$$

from which we conclude $\overline{X} \supset \overline{Y}$.

To prove the converse, notice that for any two sets X and Y

$$X \cup Y \supset X \overset{(c)}{\Rightarrow} \overline{X \cup Y} \supset \overline{X} \overset{(b)}{\Rightarrow} \overline{X \cup Y} \supset \overline{\overline{X}}.$$

Similarly

$$X \cup Y \supset Y \overset{(c)}{\Rightarrow} \overline{X \cup Y} \supset \overline{Y}.$$

From the last two relations we see that

$$\overline{X \cup Y} \supset \overline{\overline{X}} \cup \overline{Y}.$$

And since $\overline{Y} \supset Y$, on the right-hand side we can write $\overline{Y} = \overline{Y} \cup Y$, thus obtaining (1.2). □

1.2 Relations

A **relation** on a set $A \neq \varnothing$ is a subset $R \subset A \times A$ of the Cartesian product $A \times A$. Usually, instead of writing $(x, y) \in R$, we write xRy; we say that x is **related** to y.

Definition 1.7. A relation $\sim$ on the set A is called an **equivalence relation** if it satisfies the following axioms for all $a, b, c \in A$:

1. *Reflexivity*: $a \sim a$.
2. *Symmetry*: $a \sim b$ implies $b \sim a$.
3. *Transitivity*: if $a \sim b$ and $b \sim c$ then $a \sim c$.

The expression $x \sim y$ is then read "x is equivalent to y".

Definition 1.8. Given an equivalence relation on A, the **equivalence class** of x is the set

$$[x] \equiv \{\, a \in A : a \sim x \,\}.$$

Exercise 1.9. *Let $[x]$ and $[y]$ be two equivalence classes. Show that either $[x] = [y]$ or $[x] \cap [y] = \varnothing$.*

A **partition** of a set A is a family of pairwise disjoint subsets $\{A_i\}$ of A such that $A = \bigcup_i A_i$. Using the last problem, we see that an equivalence relation in a set A defines a partition of A.

Example 1.10. In the set of integers $\mathbb{Z}$ we fix an integer k and we define an equivalence relation by

$$m \sim n \iff m - n \text{ is divisible by } k.$$

Since any integer n can be uniquely written in the form

$$n = n'\,k + r \quad \text{for } 0 \leq r < k \text{ and } n', k, r \in \mathbb{Z},$$

we conclude that two integers n, m belong to the same equivalence class if and only if they have the same residue r when divided by k. Therefore, this equivalence relation partitions the set of integers into k subsets:

$$A_r = \{k\,n' + r : n' \in \mathbb{Z}\} = [r] \quad \text{for } r = 0, 1, \ldots, k-1.$$

The set of equivalence classes is

$$\mathbb{Z}_k = \{[0], [1], \ldots, [k{-}1]\}.$$

Often we write $\mathbb{Z}_k = \mathbb{Z}/{\sim}$ to show that $\mathbb{Z}_k$ is the set $\mathbb{Z}$ with the relation $\sim$ "factored out". □

Definition 1.11. A relation $\preccurlyeq$ on the set A is called an **order relation** if it satisfies the following axioms for all $a, b, c \in A$:

1. Reflexivity: $a \preccurlyeq a$.
2. Symmetry: If $a \preccurlyeq b$ and $b \preccurlyeq a$, then $a = b$.
3. Transitivity: If $a \preccurlyeq b$ and $b \preccurlyeq c$, then $a \preccurlyeq c$.

The expression $x \preccurlyeq y$ is then read "x precedes y".

If every element of A is comparable to every other, that is, if for all $a, b \in A$ we have

4. Trichotomy: either $a \preccurlyeq b$ or $b \preccurlyeq a$,

then we say that $\preccurlyeq$ is a **total order**. If there are elements that are not comparable to one another, then we say the order is **partial**.

Example 1.12. Given a set A, let's consider on $\mathcal{P}(A)$ the relation of inclusion. That is, take $\preccurlyeq\ =\ \subset$, in the sense that for any $S, T \in \mathcal{P}(A)$, we have $S \preccurlyeq T$ if and only if $S \subset T$. This is a partial order on $\mathcal{P}(A)$ if A has more than one element, because two singlets $\{a\}$ and $\{b\}$ are not included in each other if $a \neq b$. □

Example 1.13. The set of reals $\mathbb{R}$ equipped with $\leq$ is totally ordered. □

1.3 Functions

The Notion of a Function

Suppose a subset F of $A \times B$ satisfies the following properties:

- $(x, y) \in F$ and $(x, y') \in F$ implies $y = y'$.
- For every $x \in A$, there is y such that $(x, y) \in F$.

We say that the triple $f = (F, A, B)$ is a **function** (or a **map**), and we write $f : A \to B$, pronounced "f maps A into B". We call A the **domain** of f, and B its **codomain**. Finally, the set of ordered pairs F is the graph of f; it is also denoted by Γ_f.

We think of the function f as a correspondence that assigns to an element x in the domain the element y of the codomain such that $(x, y) \in \Gamma_f$. We write also $f(x) = y$. Thus

$$\Gamma_f = \{\,(x, f(x)) : x \in A\,\} \subset A \times B.$$

It follows from the definition that the domain and codomain are integral parts of a function: for two functions to be equal, their domains, codomains, and graphs must be equal. That is, $f : A \to B$ and $g : C \to D$ are equal if and only if $A = C$, $B = D$, and $f(x) = g(x)$ for all $x \in A$.

For $x \in A$ and $y = f(x)$, we call y the **image** of x, and x the **preimage** of y. If $A' \subset A$, the subset of B given by $f(A') = \{f(a) : a \in A'\} \subset B$ is called the **image** of A'. The image $f(A)$ of the domain is also called the image of the function.[2]

Comment. Although the most common objects of our study will be real-valued functions ($B \subset \mathbb{R}$) of a real variable ($A \subset \mathbb{R}$), the definitions given here are more general. The elements of the domain may be vectors, matrices, or any other mathematical object, and likewise for the codomain. □

Example 1.14. The **identity function** $\mathrm{id}_A : A \to A$ maps each element of A to itself: $\mathrm{id}_A(x) = x$. □

Example 1.15. Suppose $c \in B$. The **constant function** with value c is the function $f : A \to B$ that maps the whole domain to c, that is, $f(x) = c$ for all $x \in A$. □

Suppose $f : A \to \mathbb{R}$ is a real-valued function defined on a subset $A \subset \mathbb{R}$. The graph $\Gamma_f = \{\, (x, f(x)) : x \in A \,\} \subset A \times B$ is often of help in visualizing the function's behavior (see Problem 1.17 on the next page). However, for many functions such a pictorial representation may not be possible and we must treat the graph as a set.

Example 1.16. The **Dirichlet function**, defined by

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q}, \\ 0 & \text{if } x \in \mathbb{R} \setminus \mathbb{Q}, \end{cases}$$

is one for which a sketch of the graph would not be much help: since both $\mathbb{Q}$ and its complement are dense in $\mathbb{R}$, the graph would look like two horizontal lines, no matter how much we magnify it. □

Problem 1.17. *Draw the graph of a function* $f : \mathbb{R} \to \mathbb{R}$ *that satisfies*

(a) $f(x + 1) = f(x) - 2$ *for all* $x \in \mathbb{R}$, *and*

(b) $f(x) = x^2$ *for* $x \in [0, 1)$.

[2] I avoid the term "range", which can mean either the codomain or the image of the domain.

Solution. Using induction it is easy to see that

$$f(x+n) = f(x) - 2n \quad \text{for all } n \in \mathbb{N}.$$

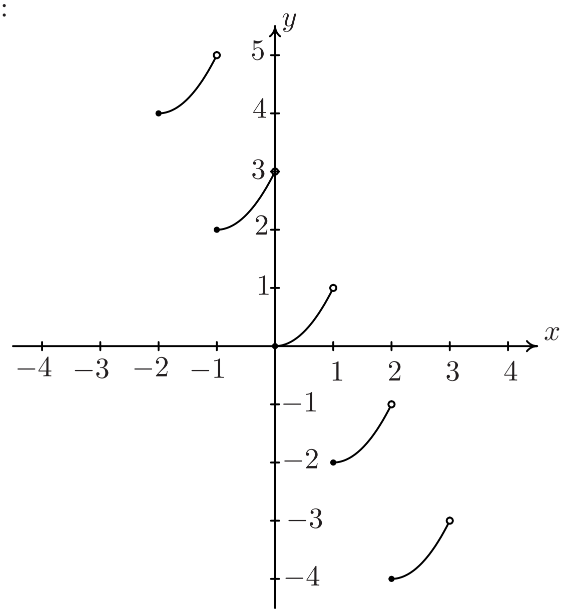
Let $y \in \mathbb{R}$ be an arbitrary real number. We can always write it as $y = x + n$, where n is the integer part of y, also denoted by $\lfloor y \rfloor$, and x is the fractional part of y, denoted by $\{y\}$ and belonging to the interval $[0, 1)$. Then, the previous equation gives

$$f(y) = f(\{y\}) - 2n \quad \text{for } n \leq y < n+1.$$

Since the fractional part satisfies the inequality $0 \leq \{y\} < 1$, condition (b) yields $f(\{y\}) = \{y\}^2$. However, $\{y\} = y - n$ and we finally have

$$f(y) = (y-n)^2 - 2n \quad \text{for } n \leq y < n+1.$$

Here is a sketch of the graph:



□

Properties of Functions

A function $f : A \to B$ is called **surjective** or **onto** if its image and codomain coincide: $f(A) = B$. Equivalently, every element in B has a preimage under f in A. Obviously, if a function $f : A \to B$ is not surjective, we can find a surjective function with the same "content" as f by just replacing the codomain B with $f(A)$.

Exercise 1.18. *Let $f : A \to B$ be a function and let A', B' be subsets of A. Prove:*

$$f(A' \cap B') \subset f(A') \cap f(B');$$
$$f(A' \cup B') = f(A') \cup f(B').$$

□

A function $f : A \to B$ is called **injective** or **one-to-one** if any two distinct elements in A are mapped to distinct elements in B:

$$x_1 \neq x_2 \;\Rightarrow\; f(x_1) \neq f(x_2) \quad \text{for all } x_1, x_2 \in A.$$

Equivalently, f is injective if

$$f(x_1) = f(x_2) \;\Rightarrow\; x_1 = x_2 \quad \text{for all } x_1, x_2 \in A.$$

A function f is called **bijective** if it is both injective and surjective.

The **restriction** of a function $f : A \to B$ to a set $A' \subset A$ is the function $g : A' \to B$ defined by $g(x) = f(x)$ for all $x \in A'$. It is often denoted by $f|_{A'}$. Equivalently, one can say that f is the **extension** of g to the set $A \supset A'$.

Let A, B be sets with total order, both orders being denoted by $\leq$. We say that $f : A \to B$ is **nondecreasing** if $f(x) \leq f(y)$ whenever $x, y \in A$ satisfy $x \leq y$. A nondecreasing function for which, in addition, $f(x) \neq f(y)$ whenever $x \neq y$ (equivalently, $f(x) < f(y)$ whenever $x < y$) is **strictly increasing**, or **order-preserving**. **Nonincreasing** functions and **strictly decreasing** (or **order-reversing**) functions are defined analogously. A (strictly) increasing or decreasing function is also called (strictly) **monotonic**.

A function $f : A \to B$ is said to be **upper bounded** if there exists $M \in B$ such that $f(x) \leq M$ for all $x \in A$. It is called **lower bounded** if there exists $m \in B$ such that $m \leq f(x)$ for all $x \in A$. The function is **bounded** if it is both lower and upper bounded. For a lower bounded function, we call the greatest of its lower bounds the **infimum**; for an upper bounded function, we call its least upper bound the **supremum**.

Operations on Functions

Given two real-valued functions $f : A \to \mathbb{R}$ and $g : A \to \mathbb{R}$ with the same domain, we define their sum $f + g$ (difference $f - g$, product $f \cdot g$, and so on) to be a function h defined on the same domain whose image $h(x)$ of $x \in A$ is the sum (difference, product, and so on) of the images $f(x)$, $g(x)$ of the original functions. (The quotient f/g is defined on the subset of A where g does not vanish.)

Let $f : A \to B$ and $g : B \to C$ be functions. We can define a new function by applying f and g sequentially; this new function $h : A \to C$, called the **composite** (or **composition**) of f and g, is given by $h(x) = g(f(x))$ for all $x \in A$. We also write $h = g \circ f$. Composition of functions satisfies associativity, that is, $(f \circ g) \circ h = f \circ (g \circ h)$, where the domains and codomains are assumed such that all operations are well defined.

For a function f whose domain contains the codomain, the notation $f^n(x)$ will indicate[3] the n-fold composition of f with itself: $f^2 = f \circ f$, and $f^n = f \circ f^{n-1}$ for $n > 2$. We call f^n the n-th **iterate** of f. For numerical functions, this is to be distinguished from the notation for powers of f; apart from the case of trigonometric functions (see footnote), we write $f(x)^2$ or $(f(x))^2$, and not $f^2(x)$, to indicate the square of $f(x)$.

Inverse Function

A bijective function $f : A \to B$ provides a unique association between the elements of the domain and codomain. This allows us to define a closely related function $g : B \to A$, by saying that $x = g(y)$ if and only if $y = f(x)$. This is called the **inverse function** of f and is typically denoted by f^{-1} instead of g. Obviously f^{-1} is also bijective.

Theorem 1.19. *A bijective function f has a unique inverse function f^{-1}.*

Comparing the graphs of f and f^{-1},

$$\begin{aligned} \Gamma_f &= \{(x,y) : y = f(x),\ x \in A\} \quad \text{and} \\ \Gamma_{f^{-1}} &= \{(y,x) : x = f^{-1}(y),\ y \in B\} \\ &= \{(y,x) : y = f(x),\ x \in A\}, \end{aligned}$$

we notice that $\Gamma_{f^{-1}}$ is obtained from Γ_f by reversing each ordered pair (x, y) in Γ_f, and conversely. Geometrically, the two graphs are the reflection of one another in the line $y = x$.

Exercise 1.20. *Let $f : A \to B$ and $g : B \to C$ be two bijective functions. Prove that* (a) $(f^{-1})^{-1} = f$; (b) $f^{-1} \circ f = \mathrm{id}_A$; (c) $f \circ f^{-1} = \mathrm{id}_B$; *and* (d) $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$. □

Problem 1.21 (IMO 1973)**.** *Let G be the set of functions $f : \mathbb{R} \to \mathbb{R}$ of the form $f(x) = a\,x + b$, where a and b are real numbers and $a \neq 0$. Suppose that G satisfies the following conditions:*

(a) *If $f, g \in G$ then $g \circ f \in G$.*

(b) *If $f \in G$ then $f^{-1} \in G$.*

(c) *For each $f \in G$ there exists a number $x_f \in \mathbb{R}$ such that $f(x_f) = x_f$.*

Prove that there exists a number $k \in \mathbb{R}$ such that $f(k) = k$ for all $f \in G$.

---

[3] I break this rule only for the trigonometric functions: $\sin^2 x$ stands for $(\sin x)^2$, not $\sin(\sin x)$. This tradition is so strong and widespread that I found it impossible to reverse it.

Solution. If $a = 1$ for a function $f(x) = x + b$, then condition (c) requires that $b = 0$. In this case, all points k of $\mathbb{R}$ satisfy $f(k) = k$. Therefore we need to show the claim for $a \neq 1$.

Let $a, a' \neq 1$ and let

$$f(x) \;=\; a\,x + b, \quad g(x) \;=\; a'\,x + b',$$

be two functions in G. Then, condition (c) requires that there are two points x_f and x_g (not necessarily distinct) such that

$$f(x_f) \;=\; x_f \;\Rightarrow\; x_f \;=\; -\frac{b}{a-1},$$

and

$$g(x_g) \;=\; x_g \;\Rightarrow\; x_g \;=\; -\frac{b'}{a'-1}.$$

According to condition (b), both

$$f^{-1} \;=\; \frac{1}{a}\,x - \frac{b}{a} \quad \text{and} \quad g^{-1} \;=\; \frac{1}{a'}\,x - \frac{b'}{a'}$$

are in G. Then, according to condition (a),

$$f \circ g(x) \;=\; aa'\,x + ab' + b,$$

and

$$f^{-1} \circ g^{-1}(x) \;=\; \frac{1}{aa'}\,x - \frac{b' + ba'}{aa'},$$

and

$$f \circ g \circ f^{-1} \circ g^{-1} \;=\; x + (ab' + b) - (b' + ba')$$

are also elements of G. Since there is an x_0 for this latter function such that $f \circ g \circ f^{-1} \circ g^{-1}(x_0) = x_0$, we conclude that

$$(ab' + b) - (b' + ba') \;=\; 0 \;\Rightarrow\; \frac{b}{1-a} \;=\; \frac{b'}{1-a'} \;\Rightarrow\; x_f \;=\; x_g. \qquad \square$$

If $B' \subset B$, we denote by $f^{-1}(B')$ the set of preimages of elements of B':

$$f^{-1}(B') \;=\; \{x \in A : f(x) = y \text{ for some } y \in B'\}.$$

Exercise 1.22. *Let A' be a subset of A and B', B'' be two subsets of B. Prove:*

$$\begin{aligned} f^{-1}(f(A')) \;&\supset\; A'; \\ f(f^{-1}(B')) \;&\subset\; B'; \\ f^{-1}(B' \cap B'') \;&=\; f^{-1}(B') \cap f^{-1}(B''); \\ f^{-1}(B' \cup B'') \;&=\; f^{-1}(B') \cup f^{-1}(B''). \end{aligned} \qquad \square$$

1.4 Limits and Continuity

Limits

Consider a function $f : A \to \mathbb{R}$ defined on a subset A of $\mathbb{R}$, and a point ξ in the closure $\overline{A}$ of A. We say that f has a **limit** $\ell \in \mathbb{R}$ at ξ, and we write $\lim_{x\to\xi} f(x) = \ell$, if the following condition is satisfied: *For any $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$|f(x) - \ell| < \varepsilon \quad \textit{for all } x \in A \textit{ such that } 0 < |x - \xi| < \delta.$$

In other words, $f(x)$ gets arbitrarily close to ℓ as x approaches ξ. Note that ξ need not be in the domain of definition of f.

In the language of neighborhoods, f has a limit ℓ at ξ if and only if, given any neighborhood J of ℓ in $\mathbb{R}$, there exists a neighborhood I of ξ in $\mathbb{R}$ such that $f(A \cap I \setminus \{\xi\}) \subset J$.

This formulation immediately allows the notion of limit to be extended to functions from A into any topological space B, such as the Cartesian plane: $f : A \to B$ has a **limit** $\ell \in B$ at ξ if, given any neighborhood J of ℓ in B, there exists a neighborhood I of ξ in $\mathbb{R}$ such that $f(A \cap I \setminus \{\xi\}) \subset J$.

We say that f has a **right-hand limit** ℓ at ξ, written $\lim_{x\to\xi^+} f(x) = \ell$, if $f(x)$ gets arbitrarily close to ℓ as x approaches ξ *from the right*, regardless of what happens to the left of ξ. In terms of epsilons and deltas: *We write $\lim_{x\to\xi^+} f(x) = \ell$ if, for any $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$|f(x) - \ell| < \varepsilon \quad \textit{for all } x \in A \textit{ such that } x \in (\xi, \xi+\delta). \tag{1.5}$$

Note that $\lim_{x\to\xi^+} f(x) = \ell$ if and only if $\lim_{x\to\xi} \bar{f}(x) = \ell$, where $\bar{f}$ is the restriction of f to the subset $A \cap [\xi, +\infty)$.

The **left-hand limit**, denoted by $\lim_{x\to\xi^-} f(x)$, is defined similarly, using the condition $x \in (\xi-\delta, \xi)$ in (1.5), or the restriction of f to the subset $A \cap (-\infty, \xi]$.

Exercise 1.23. *Let the domains of f and g contain some neighborhood of $\xi \in \mathbb{R}$ (possibly with ξ itself removed). Suppose we have $\lim_{x\to\xi} f(x) = \ell_1$ and $\lim_{x\to\xi} g(x) = \ell_2$. Prove:*

(a) $\lim_{x\to\xi}(\lambda f(x)) = \lambda\,\ell_1$, *for any* $\lambda \in \mathbb{R}$.

(b) $\lim_{x\to\xi}(f(x) + g(x)) = \ell_1 + \ell_2$.

(c) $\lim_{x\to\xi} f(x)\,g(x) = \ell_1\ell_2$.

(d) $\lim_{x\to\xi} f(x)/g(x) = \ell_1/\ell_2$ *provided that* $\ell_2 \neq 0$ *and that, for some* $\delta > 0$, *we have* $g(x) \neq 0$ *whenever* $0 < |x - \xi| < \delta$. □

Theorem 1.24. *Let $f(x)$ and $g(x)$ exist and satisfy $f(x) \leq g(x)$ for all x in a neighborhood of ξ in $\mathbb{R}$ (possibly with ξ itself removed). If the limits of f, g as $x \to \xi$ exist, then*

$$\lim_{x\to\xi} f(x) \leq \lim_{x\to\xi} g(x).$$

Theorem 1.25 (Squeeze Theorem)**.** *Let $f(x)$, $g(x)$ and $h(x)$ exist and satisfy $f(x) \leq g(x) \leq h(x)$ for all x in a neighborhood of ξ in $\mathbb{R}$ (possibly with ξ itself removed). If $\lim_{x\to\xi} f(x) = \ell$ and $\lim_{x\to\xi} h(x) = \ell$, then*

$$\lim_{x\to\xi} g(x) = \ell.$$

Recall that a **sequence** is a function with domain $\mathbb{N}$, but we use notation such as x_n rather than $x(n)$ for the value of the function at n, and the sequence itself may be written $\{x_n\}_{n\in\mathbb{N}}$ or $\{x_n\}$. More generally, the domain of a sequence can be $\mathbb{Z} \cap [n_0, \infty)$, for some $n_0 \in \mathbb{Z}$.

A sequence $\{x_n\}_{n\in\mathbb{N}}$ of real numbers is said to **converge** to a limit $\ell \in \mathbb{R}$, and we write $\lim_{n\to\infty} x_n = \ell$ or $x_n \to \ell$, if for any $\varepsilon > 0$ there exists n_0 such that $|x_n - \ell| < \varepsilon$ for all $n \geq n_0$.

Results analogous to Theorems 1.24 and 1.25 also apply to sequences.

Continuity

Let A be a subset of $\mathbb{R}$. A function $f : A \to B$ is called **continuous** at $a \in A$ if $\lim_{x\to a} f(x) = f(a)$. It is called **right-continuous** at a if $\lim_{x\to a^+} f(x) = f(a)$, and **left-continuous** at a if $\lim_{x\to a^-} f(x) = f(a)$. We say that f is continuous (or right-continuous, or left-continuous) on an open interval $I \subset \mathbb{R}$ if f has the same property at every $x \in I$.

Using the notion of open sets, we can extend the concept of continuity to functions between two topological spaces. A function $f : X \to Y$ is said to be **continuous** at $a \in X$ if, given any open set J containing $f(a)$, there exists an open set I containing a and such that $f(I) \subset J$. The function f is continuous on X if it is continuous at every point $x \in X$.

Theorem 1.26. *Let B be any topological space and let $A \subset \mathbb{R}$ contain the interval $(\xi-\epsilon, \xi+\epsilon)$ for some $\xi \in \mathbb{R}$ and $\epsilon > 0$.*

(a) *A function $f : A \to B$ is continuous at ξ if and only if*

$$\lim_{n\to\infty} f(x_n) = f(\xi)$$

for every sequence $\{x_n\}$ converging to ξ.

(b) *A function $f : A \to B$ is right-continuous at ξ if and only if*

$$\lim_{n\to\infty} f(x_n) = f(\xi)$$

for every sequence $\{x_n\}$ converging to ξ from the right (this condition means that $\{x_n\}$ converges to ξ and $x_n \geq \xi$ for every n). A similar statement holds for left continuity.

A function that is not continuous at a point ξ of its domain is called **discontinuous** at ξ. A discontinuity can occur for one of the following reasons:

(i) The right-hand and left-hand limits are equal but different from the value of the function (a **removable discontinuity**).

(ii) The right-hand and left-hand limits exist but they have different values (a **jump discontinuity**).

(iii) The right-hand limit or the left-hand limit does not exist (an **essential discontinuity**).

Example 1.27. The function of Problem 1.17 has jump discontinuities at all integer points. □

We also apply the terminology above when ξ is in the closure of the domain, but not in the domain itself. By definition, f cannot be continuous at ξ in this case, but we talk of removable, jump, and essential discontinuities depending on whether the left-hand and right-hand limits exist and agree, or exist and disagree, or don't both exist.

Example 1.28. The function $f(x) = 1/x^2$, defined on $\mathbb{R}\setminus\{0\}$, has an essential discontinuity at $x = 0$: since $f(x)$ is unbounded in any interval $(0, \epsilon)$, there can be no limit of $f(x)$ at 0. □

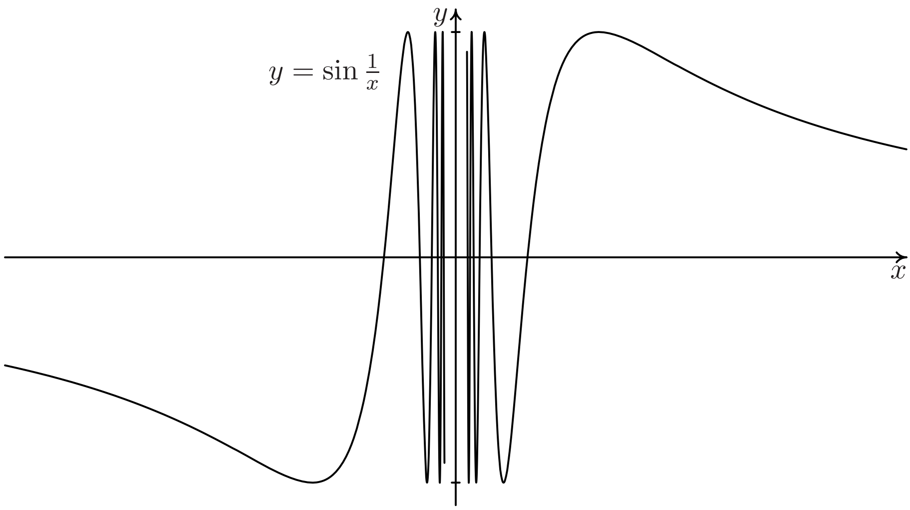
Example 1.29. The same is true of the function

$$f(x) = \sin\frac{1}{x},$$

again defined only for $x \neq 0$. We see from the graph above that $f(x)$ oscillates violently between -1 and 1 as x gets close to 0. To prove formally that f has no right-hand limit at 0, consider the sequences given by

$$a_n = \frac{2}{(4n+1)\pi} \quad \text{and} \quad b_n = \frac{2}{(4n-1)\pi}.$$

They both converge to 0 from the right, but $f(a_n) = \sin\bigl(2\pi n + \frac{\pi}{2}\bigr) = 1$ and $f(b_n) = \sin\bigl(2\pi n - \frac{\pi}{2}\bigr) = -1$, which would be in violation of Theorem 1.26(b) if there were a right-hand limit. □



Example 1.30. The function $f : \mathbb{R} \smallsetminus \{0\} \to \mathbb{R}$ given by

$$f(x) = x \sin \frac{1}{x}$$

has a discontinuity at $x = 0$, since this point does not belong to the domain. But the limit as $x \to 0$ exists and equals 0, as follows from Theorem 1.25, for instance. In other words, the right-hand and left-hand limits at 0 are both 0, and we are in case (i) of the previous page: a removable discontinuity.

The adjective "removable" is explained by the fact that, in this case, we can define (or redefine) the function at the point of discontinuity to obtain a continuous function. Indeed, let's extend f to a function $\bar{f}$ on all of $\mathbb{R}$, by setting $\bar{f}(x) = f(x)$ for $x \neq 0$, and $\bar{f}(0) = 0$. Then $\bar{f}$ is continuous at all points: we have removed the discontinuity. (In this situation we usually won't bother giving a different name to the extension: we will just say "set $f(0) = 0$", even though we're defining a new function distinct from f.) □

The previous discussion may have created the impression that functions are mostly continuous and they may have only a small set of points of discontinuity. However, this is not correct. The Dirichlet function of Example 1.16 is discontinuous *everywhere*. The function

$$f(x) = \begin{cases} x & \text{if } x \in \mathbb{Q}, \\ 0 & \text{if } x \in \mathbb{R} \smallsetminus \mathbb{Q}, \end{cases}$$

which is perhaps the simplest modification of the Dirichlet function, is continuous only at one point, $x = 0$.

Exercise 1.31. *Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be two continuous functions at $x = \xi$. Prove:*

(a) *$f + g$ is continuous at $x = \xi$.*

(b) *fg is continuous at $x = \xi$.*

(c) *f/g is continuous at $x = \xi$ provided that $g(\xi) \neq 0$.*

(d) *$f \circ g$ is continuous at $x = \xi$.* □

The next theorem, saying essentially that the graph of a continuous function cannot change signs unless it crosses the x-axis, finds many applications, as does its generalization, the Intermediate Value Theorem.

Theorem 1.32 (Bolzano)**.** *If a function f is continuous on the interval $[a, b]$ and $f(a)f(b) < 0$, then there exists a point $\xi \in (a, b)$ such that $f(\xi) = 0$.*

Theorem 1.33 (Intermediate Value Theorem)**.** *If a function $f(x)$ is continuous on the interval $[a, b]$, it takes on every value between $f(a)$ and $f(b)$.*

See Problem 1.17 for an example of how the conclusion of these theorems can fail to hold when the function is discontinuous.

1.5 Differentiation

First Derivative

A function $f : A \to \mathbb{R}$, where A contains ξ in its interior, is **differentiable** at ξ if the limit

$$\lim_{x \to \xi} \frac{f(x) - f(\xi)}{x - \xi} = \lim_{h \to 0} \frac{f(h + \xi) - f(\xi)}{h}$$

exists. This limit is then called the **derivative** of f at $x = \xi$ and is denoted by $f'(\xi)$ or $(df(x)/dx)|_\xi$.

The function f is called differentiable on an open interval $I \subset \mathbb{R}$ if f is differentiable at every $x \in I$. Geometrically, the derivative $f'(\xi)$ gives the slope of the tangent line at the point $(\xi, f(\xi))$ of Γ_f.

Using right-hand and left-hand limits, we can extend the concept of differentiation to *differentiation from the right* or *from the left.*

Theorem 1.34. *A function that is differentiable at a point ξ is continuous at ξ.*

Differentiability of the function $f(x)$ at x is equivalent to the continuity of the function

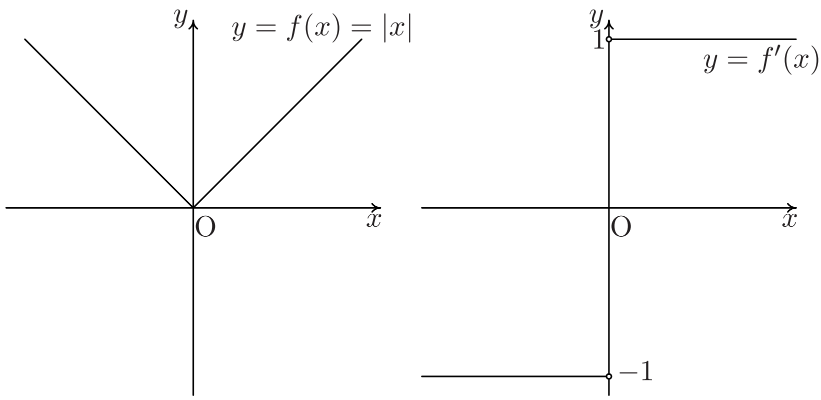
$$F(h) = \frac{f(h + x) - f(x)}{h}$$

at $h = 0$, with the variable x playing the role of a parameter specifying the point of interest. Therefore any discontinuity at $h = 0$, other than a removable one, signals problems.

Example 1.35. For the function $f(x) = |x|$, we have

$$F(h) = \frac{|x+h| - |x|}{h}.$$

For $x = 0$, this becomes $F(h) = |h|/h = \operatorname{sgn}(h)$, defined for $h \neq 0$. This is a discontinuous function with a jump discontinuity at $x = 0$. Therefore $f(x) = |x|$ does not have a derivative at $x = 0$. The discontinuity in $F(h)$ appears as a sharp bend in the graph of $f(x)$.



Note, however, that $|x|$ is differentiable from the right and from the left. □

Example 1.36. The function

$$f(x) = \begin{cases} x \sin \dfrac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$$

implies

$$F(h) = \sin \frac{1}{h}$$

when we set $x = 0$. This has an essential discontinuity at $h = 0$. Therefore $f(x)$ has no derivative at $x = 0$. In contrast, if we take

$$f(x) = \begin{cases} x^2 \sin \dfrac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$$

we obtain

$$F(h) = h \sin \frac{1}{h},$$

which is continuous at $h = 0$; therefore $f'(0)$ exists and equals 0. □

Continuous functions may appear to fail to be differentiable only at a small set of points. This view was held by mathematicians until Weierstrass[4] explicitly showed it to be incorrect. We present Weierstrass's example after some additional results on the derivative.

Exercise 1.37. *Let f and g be differentiable at x and let c be a constant. Prove:*

(a) $(f+g)'(x) = f'(x) + g'(x)$.

(b) $(cf)'(x) = cf'(x)$.

(c) $(fg)'(x) = f(x)g'(x) + f'(x)g(x)$.

(d) $(f/g)'(x) = \dfrac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$, *for* $g(x) \neq 0$.

(e) $(f(g(x))' = f'(g(x))\, g'(x)$. □

The following three theorems, each generalizing the previous one, are among the most useful results on differentiable functions.

Theorem 1.38 (Rolle)**.** *If f is continuous on $[a,b]$, is differentiable on (a,b) and satisfies $f(a) = f(b) = 0$, there is a $\xi \in (a,b)$ such that $f'(\xi) = 0$.*

That is, the graph of such an f, restricted to (a,b), has a horizontal tangent. One could try to use the zero derivative to show the existence of a local maximum or minimum for f between a and b (see page 28 for the connection). But in fact it's enough for f to be *continuous* to guarantee the existence of an *absolute* maximum or minimum in (a,b); see Theorem 1.47.

Theorem 1.39 (Lagrange's Mean Value Theorem)**.** *If f is continuous on $[a,b]$ and differentiable on (a,b), there is a $\xi \in (a,b)$ such that*

$$f'(\xi) = \frac{f(b) - f(a)}{b - a}.$$

That is, the graph of such an f has the property that every secant has at least one parallel tangent (the tangency occurring between the intersections of the secant).

Theorem 1.40 (Cauchy's Mean Value Theorem)**.** *If f, g are continuous on $[a,b]$, differentiable on (a,b), and $g(x) \neq 0$ for any $x \in (a,b)$, there is a $\xi \in (a,b)$ such that*

$$\frac{f'(\xi)}{g'(\xi)} = \frac{f(b) - f(a)}{g(b) - g(a)}.$$

[4] Bolzano seems to have done so before Weierstrass but his work was seemingly ignored.

Problem 1.41 (Putnam 1939). *Let $f(x)$ be defined for $x \in [a, b]$. Assuming appropriate properties of continuity and differentiability, prove for $x \in (a, b)$ that*

$$\frac{\frac{f(x)-f(a)}{x-a} - \frac{f(b)-f(a)}{b-a}}{x-b} = \frac{1}{2} f''(\xi),$$

where ξ is some number in (a, b).

Solution ([21]). Consider the function

$$F(y) = \begin{vmatrix} f(y) & y^2 & y & 1 \\ f(x) & x^2 & x & 1 \\ f(a) & a^2 & a & 1 \\ f(b) & b^2 & b & 1 \end{vmatrix},$$

where x is considered fixed and $y \in [a, b]$. Then $F(y)$ is continuous on $[a, b]$ and differentiable on (a, b). Moreover, $F(a) = F(x) = 0$ (since two rows of the determinant are equal to each other). Therefore the function $F(y)$ satisfies the conditions of Rolle's theorem on the interval $[a, x]$ and there exists a point $\alpha \in (a, x)$ such that $F'(\alpha) = 0$. Similarly, there exists a point $\beta \in (x, b)$ such that $F'(\beta) = 0$.

But, from the previous results, the function

$$F'(y) = \begin{vmatrix} f'(y) & 2y & 1 & 0 \\ f(x) & x^2 & x & 1 \\ f(a) & a^2 & a & 1 \\ f(b) & b^2 & b & 1 \end{vmatrix}$$

also satisfies the conditions of Rolle's theorem on the interval $[\alpha, \beta]$. Therefore, there exists $\xi \in (\alpha, \beta)$ such that $F''(\xi) = 0$, where

$$F''(y) = \begin{vmatrix} f''(y) & 2 & 0 & 0 \\ f(x) & x^2 & x & 1 \\ f(a) & a^2 & a & 1 \\ f(b) & b^2 & b & 1 \end{vmatrix}.$$

Upon expanding the determinant, we arrive at the result sought. □

Example 1.42 (Weierstrass's nowhere differentiable function)**.** Let

$$f(x) = \sum_{k=0}^{\infty} a^k \cos(b^k \pi x),$$

where $0 < a < 1$ and b is an odd positive integer. I leave the proof of continuity to the reader (who may simply accept it), and show only that f is nowhere differentiable, if ab is suitably large.

Let $F(h) = \big(f(x+h) - f(x)\big)/h$, as usual. Write $F = A_n + R_n$, where

$$A_n(h) = \sum_{k=0}^{n-1} a^k \,\frac{\cos[b^k\pi(x+h)] - \cos(b^k\pi x)}{h}$$

is the sum of the first n terms of the sum defining F, and

$$R_n = \sum_{k=n}^{\infty} a^k \,\frac{\cos[b^k\pi(x+h)] - \cos(b^k\pi x)}{h} \tag{1.6}$$

is the remainder. By the triangle inequality,

$$|A_n(h)| \le \sum_{k=0}^{n-1} a^k \left|\frac{\cos[b^k\pi(x+h)] - \cos(b^k\pi x)}{h}\right|.$$

For each k, we apply Lagrange's mean value theorem to the function $\cos(b^k\pi x)$ on $[x, x+h]$, to find $\xi \in (x, x+h)$ such that

$$\frac{\cos[b^k\pi(x+h)] - \cos(b^k\pi x)}{h} = b^k\pi\,\sin(b^k\pi\xi).$$

It follows that

$$\left|\frac{\cos[b^k\pi(x+h)] - \cos(b^k\pi x)}{h}\right| = b^k\pi\,|\sin(b^k\pi\xi)| \le b^k\pi;$$

thus

$$|A_n(h)| \le \pi \sum_{k=0}^{n-1} a^k b^k = \pi\,\frac{(ab)^n - 1}{ab - 1} < \pi\,\frac{(ab)^n}{ab-1},$$

if we take $ab > 1$.

Next we turn to the remainder term R_n. Set $b^n x = N_n + f_n$, where N_n is an integer part and f_n a fractional part, the latter defined here in the interval $[-1/2, 1/2]$. If $h = (1 - f_n)/b^n$, then $\frac{2}{3}b^n \le 1/h \le 2b^n$. Also, for any $k \ge n$,

$$\cos[b^k\pi(x+h)] = (-1)^{N_n+1} \quad\text{and}\quad \cos(b^k\pi x) = (-1)^{N_n+1}\,\cos(b^{k-n}\pi f_n).$$

Therefore, for such h, all summands in (1.6) are of the same sign, and we have

$$\begin{aligned} |R_n(h)| &= \left|\sum_{k=n}^{\infty} a^k \,\frac{1 - \cos(b^{k-n}\pi f_n)}{h}\right| \\ &> a^n\,\frac{1-\cos(\pi f_n)}{h} \ge \tfrac{2}{3}(ab)^n\,[1 - \cos(\pi f_n)] > \tfrac{2}{3}(ab)^n. \end{aligned}$$

Since $F = A_n + R_n$, we have, for some $h \leq \frac{3}{2}b^{-n}$,

$$|F(h)| \geq |R_n| - |A_n| > (ab)^n \frac{2ab - (2+3\pi)}{3(ab-1)}.$$

If $2ab > 2 + 3\pi$, the right-hand side grows without bound as $n \to \infty$, while h tends to 0. Therefore, F cannot be continuous, implying that f cannot be differentiable. Since x has been left unspecified, the result holds for any x in the domain of f. □

Higher Derivatives

The derivative f' of a function f may or may not be differentiable. If it is, we denote its derivative by f'', and so on. The n-th derivative of f is denoted by $f^{(n)}$, or by $d^n f/dx^n$.

Theorem 1.43 (Leibniz). *If f, g are n-times differentiable, then*

$$\frac{d^n}{dx^n}[f(x)g(x)] = \sum_{k=0}^{n} \binom{n}{k} f^{(k)}(x)\, g^{(n-k)}(x).$$

Theorem 1.44 (Taylor). *If f is $(n+1)$-times differentiable on $[a, x]$, with the first n derivatives continuous, then*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k + R_n,$$

where R_n, known as the ***remainder of order*** *n, is equal to*

$$R_n = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1} \quad \textit{for some } \xi \in (a, x).$$

If the derivative $f^{(n)}$ exists for any $n \geq 1$, the function is called **smooth**. For a smooth function, one can construct the **Taylor series**

$$\sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(a)\,(x-a)^n,$$

which does not necessarily converge. If it converges to $f(x)$ on some neighborhood I of a with radius r, we say that f is **analytic** at the point a. If $f(x)$ is analytic for all points of I, we say that f is analytic on I.

Problem 1.45 (Putnam 1998). *Let f be a real function on the real line with continuous third derivative. Prove that there exists a point a such that*

$$f(a) \cdot f'(a) \cdot f''(a) \cdot f'''(a) \geq 0.$$

Solution ([27]). If any of the $f(a), f'(a), f''(a), f'''(a)$ is zero, then the statement is true. We thus have to prove it in the case when $f(x)$,$f'(x)$,$f''(x)$,$f'''(x)$ are nonzero. In such a case, by continuity, all $f(x)$, $f'(x)$, $f''(x)$, $f'''(x)$ are strictly positive or strictly negative.

Without loss of generality, we assume that $f'''(x) > 0$ and $f''(x) > 0$—otherwise we make the changes of variable $x \mapsto -x$ and $f \mapsto -f$, respectively, which reverse the signs of these derivatives without changing the sign of the product $f(x)f'(x)f''(x)f'''(x)$.

Since $f'''(x) > 0$, the function f'' is strictly increasing. Then

$$f'(x) - f'(0) = \int_0^x f''(t)dt \geq \int_0^x f''(0)dt = xf''(0) \ ,$$

or $f'(x) \geq f'(0) + xf''(0)$. Similarly, since $f''(x) > 0$, the function f' is strictly increasing. Then

$$f(x) - f(0) = \int_0^x f'(t)dt \geq \int_0^x f'(0)dt = xf'(0) \ ,$$

or $f(x) \geq f(0) + xf'(0)$. We can now choose a large enough number $x = a$ such that $f'(a) > 0$ and $f(a) > 0$. □

Comment. As the authors of [27] observe, a function cannot be positive and concave everywhere nor negative and convex everywhere. Therefore, there is an interval for which

$$f(x)\,f''(x) > 0.$$

Assuming differentiability of f as many times as necessary and, by the above reasoning, substituting f successively by f', f'', …, $f^{(n)}$, we have

$$\begin{aligned} f'(x)\,f'''(x) &> 0,\\ f''(x)\,f^{(4)}(x) &> 0,\\ f'''(x)\,f^{(5)}(x) &> 0,\\ &\dots,\\ f^{(n)}(x)\,f^{(n+2)}(x) &> 0. \end{aligned}$$

Can you now say if there is a point a such that

$$f(a)f'(a)f^{(n)}(a)f^{(n+1)}(a) > 0, \quad \text{for } n \geq 3?$$

Applications of Derivatives

Definition 1.46. Let $f : D \to \mathbb{R}$ be a function, where D is a subset of $\mathbb{R}$, and let $x_0 \in D$. We say that f has

- an **absolute maximum** at x_0 if $f(x) \le f(x_0)$ for all $x \in D$;
- a **local maximum** at x_0 if $f(x) \le f(x_0)$ for all x in some neighborhood of x_0 in D;
- an **absolute minimum** at x_0 if $f(x) \ge f(x_0)$ for all $x \in D$;
- a **local minimum** at x_0 if $f(x) \ge f(x_0)$ for all x in some neighborhood of x_0 in D.

An **extremum** is either a minimum or a maximum (qualified as local or absolute as the case may be).

Theorem 1.47 (Extreme Value Theorem)**.** *If f is continuous on $[a, b]$, then f attains an absolute maximum and an absolute minimum somewhere in $[a, b]$. If, in addition, $f(a) = f(b)$, then f attains an absolute maximum or an absolute minimum somewhere in (a, b).*

The second statement follows easily from the first: either the maximum and the minimum in $[a, b]$ are equal — and then the function is constant and any point in (a, b) attains the maximum — or the maximum (or the minimum) is distinct from $f(a) = f(b)$, and so must be achieved in (a, b).

Problem 1.48 (Bulgaria 2000)**.** *Let*

$$f(x) = \frac{x^2 + 4x + 3}{x^2 + 7x + 14}, \quad g(x) = \frac{x^2 - 5x + 10}{x^2 + 5x + 20}.$$

(a) *Find the maximum value of $f(x)$.*

(b) *Find the maximum value of $g(x)$.*

(c) *Find the maximum value of $g(x)^{f(x)}$.*

Solution. The quadratic polynomials $x^2 + 7x + 14$, $x^2 - 5x + 10$, $x^2 + 5x + 20$ have negative discriminants and therefore they are positive for all values of x. Thus $g(x) > 0$ and the function $g(x)^{f(x)}$ is well defined.

The quadratic polynomial $x^2 + 4x + 3$ has roots -1 and -3. Therefore $f(x)$ is positive for $x \in (-\infty, -3) \cup (-1, +\infty)$ and negative for $x \in (-3, -1)$.

Finally,

$$g(x) = \frac{x^2 - 5x + 10}{x^2 + 5x + 20} = \frac{(x^2 + 5x + 20) - 10(x + 1)}{x^2 + 5x + 20} = 1 - 10\,\frac{x + 1}{x^2 + 5x + 20}.$$

That is, $g(x) \le 1$ for $x \ge -1$ and $g(x) \ge 1$ for $x \le -1$.

(a) The maximum value of $f(x)$ is 2. Indeed,

$$\begin{aligned} f(x) \le 2 \quad &\Leftrightarrow \quad \frac{x^2 + 4x + 3}{x^2 + 7x + 14} \le 2 \\ &\Leftrightarrow \quad x^2 + 4x + 3 \le 2x^2 + 14x + 28 \\ &\Leftrightarrow \quad 0 \le x^2 + 10x + 25 = (x + 5)^2. \end{aligned}$$

The maximum value is attained for $x = -5$.

(b) As in part (a),

$$g(x) \leq 3 \;\Leftrightarrow\; 0 \leq (x+5)^2.$$

The function $g(x)$ attains a maximum value of 3 for $x = -5$.

(c) For $x \in (-\infty, -3]$, since $g(x) > 0$, it follows that $\ln g(x) \leq \ln 3$. Also $f(x) \leq 2$. Then

$$f(x)\,\ln g(x) \;\leq\; 2\ln 3, \quad \text{so} \quad g(x)^{f(x)} \leq 9.$$

For $x \in [-3, -1]$, we have $g(x) \geq 1$ and $f(x) \leq 0$. Then

$$g(x)^{|f(x)|} \geq 1, \quad \text{so} \quad g(x)^{f(x)} \leq 1.$$

For $x \in [-1, +\infty]$, we have $0 < g(x) \leq 1$ and $f(x) \geq 0$. Then

$$g(x)^{f(x)} \leq 1.$$

So, the maximum value of $g(x)^{f(x)}$ is 9, attained at $x = -5$. The reader should notice that we could have simply said that, since the maxima of f and g both occur at $x = -5$, then the maximum of g^f is equal to $(\max g)^{(\max f)} = 3^2 = 9$. □

This problem demonstrates that quite some work is necessary to find extrema of a function by algebraic methods. In addition, having numbers that always conspire to simplify the calculations is impossible. The study of extrema is systematized by the derivatives of a function. We shall call **critical points** of a function f those points c at which $f'(c) = 0$ or f' does not exist.

Theorem 1.49. *Extreme values of a function can only occur at critical points and endpoints.*

However, a critical point may not necessarily be a point of extreme value. The next two theorems provide sufficient criteria for a critical point to lead to an extremum.

Theorem 1.50 (First Derivative Test)**.** *Let f be differentiable in a neighborhood of a critical point c in the interior of the domain.*

If the derivative f' changes sign as it crosses c, then f has a local extremum at c: a local maximum if $f' > 0$ for $x < c$ and $f' < 0$ for $x > c$, and a local minimum in the opposite case.

Corollary. *Let $f : A \to \mathbb{R}$ be defined on an* ***interval*** *$A \subset \mathbb{R}$ and differentiable in the interior of A. Then:*

(a) *f is strictly increasing if and only if $f' > 0$.*

(b) *f is nondecreasing if and only if $f' \geq 0$.*

(c) *f is nonincreasing if and only if $f' \leq 0$.*

(d) *f is strictly decreasing if and only if $f' < 0$.*

In each case, the inequality is supposed to hold at all points in the domain.

Theorem 1.51 (Second Derivative Test). *Let f be defined and differentiable in a neighborhood of a critical point c. Then f has a local minimum at c if $f''(c) > 0$. It has a local maximum if $f''(c) < 0$.*

Exercise 1.52. *If $f''(x_0) = f'''(x_0) = \cdots = f^{(2n)}(x_0) = 0$, but $f^{(2n+1)}(x_0) \neq 0$, discuss the behavior of f in the neighborhood of x_0. The point x_0 is called a* ***point of inflection****.* □

1.6 Solved Problems

We now present some solved problems on functions, selected for their relevance to later chapters.

Problem 1.53 (Singapore 2002). *Let $f(x)$ be a function which satisfies*

$$f(29 + x) = f(29 - x) \quad \text{for all } x \in \mathbb{R}.$$

If $f(x)$ has exactly three real roots α, β, γ, determine the value of $\alpha + \beta + \gamma$.

Solution. Consider the function F given by $F(x) = f(29+x)$. (If we translate the graph of f to the left by 29 units, we obtain the graph of F; alternatively, we can think of translating the coordinate axes to the right).

Clearly, the functional equation satisfied by the function becomes

$$F(x) = F(-x);$$

that is, F is an even function. Therefore its roots must be located symmetrically with respect to the origin. Since there are three roots, one must be 0, and the others must be x_0 and $-x_0$ for some x_0. Returning to the original function f, its roots are obtained from those of F by adding 29, since $0 = F(x)$ means $0 = f(29 + x)$. Thus the roots of f are 29, $x_0 + 29$ and $-x_0 + 29$, and their sum is 87. □

Problem 1.54 ([14], Problem 7)**.** *Let* $f_0(x) = \dfrac{1}{1-x}$*, and* $f_n(x) = f_0(f_{n-1}(x))$*,* $n = 1, 2, 3, \ldots$*. Evaluate* $f_{1976}(1976)$*.*

Solution. We notice that

$$\begin{aligned} f_0(x) &= \frac{1}{1-x}, \\ f_1(x) &= f_0(f_0(x)) = \frac{1-x}{-x}, \\ f_2(x) &= f_0(f_1(x)) = x, \\ f_3(x) &= f_0(f_2(x)) = \frac{1}{1-x} = f_0(x). \end{aligned}$$

From these results we conclude inductively that

$$f_{3k+r}(x) = f_r(x) \quad \text{for } k = 0, 1, 2, \ldots \text{ and } r = 0, 1, 2.$$

Therefore

$$f_{1976}(x) = f_2(x).$$

In particular, $f_{1976}(1976) = f_2(1976) = 1976$. □

Problem 1.55. *Let* $f(x) = x^2 - 2$ *with* $x \in [-2, 2]$*. Show that the equation*

$$f^n(x) = x$$

has 2^n *real roots.*

Solution. Since $x \in [-2, 2]$ we set $x = 2\cos\theta$, with $0 \leq \theta \leq 2\pi$. Then

$$\begin{aligned} f(\cos\theta) &= 2[2\cos^2\theta - 1] = 2\cos(2\theta), \\ f(f(\cos\theta)) &= [2\cos(2\theta)]^2 - 2 = 2[2\cos^2(4\theta) - 1] = 2\cos(4\theta), \end{aligned}$$

and so on. By induction, we easily verify that

$$f^n(\cos\theta) = 2\cos(2^n\theta).$$

The given equation, in the new notation, becomes

$$2\cos(2^n\theta) = 2\cos\theta,$$

with solutions $2^n\theta = 2k\pi \pm \theta$, $k \in \mathbb{Z}$ or

$$\theta_k^- = k\,\frac{2\pi}{2^n - 1} \quad \text{and} \quad \theta_k^+ = k\,\frac{2\pi}{2^n + 1} \quad \text{for } k \in \mathbb{Z}.$$

The distinct solutions are those for which $0 \le \theta < 2\pi$. Therefore,

$$\theta_k^- = k\,\frac{2\pi}{2^n - 1} \quad \text{for } k = 0, 1, \ldots, 2^{n-1} - 1,$$
$$\theta_k^+ = k\,\frac{2\pi}{2^n + 1} \quad \text{for } k = 1, \ldots, 2^{n-1}.$$

Counting, these are exactly 2^n in number. □

This problem actually appeared in an International Mathematical Olympiad:

Problem 1.56 (IMO 1976). *Let $P_1(x) = x^2 - 2$ and $P_j(x) = P_1(P_{j-1}(x))$ for $j = 2, 3, \ldots$. Show that for any positive integer n, the roots of the equation $P_n(x) = x$ are real and distinct.*

In this statement the domain is not specified to be $[-2, 2]$. The problem is actually taken from the theory of *orthogonal polynomials*. The functions $T_N(\cos\theta) = \cos(N\theta)$ are known as **Chebyshev polynomials** (of the first kind). When written in terms of the variable $x = \cos\theta$, they are indeed polynomials. The function $f^n(x)$ of the problem is just $2\,T_{2^n}(x)$.

Here is a related problem that was proposed for the IMO but did not make the cut:

Problem 1.57 (IMO 1978 Longlist). *Given the expression*

$$P_n(x) = \frac{1}{2^n}\left[\left(x + \sqrt{x^2 - 1}\right)^n + \left(x - \sqrt{x^2 - 1}\right)^n\right],$$

prove that $P_n(x)$ satisfies the identity

$$P_n(x) - x\,P_{n-1}(x) + \tfrac{1}{4}\,P_{n-2}(x) = 0,$$

and that $P_n(x)$ is a polynomial in x of degree n.

The Chebyshev polynomials $T_n(x)$ satisfy the recursion relation

$$T_n(x) - 2xT_{n-1}(x) + T_{n-2}(x) = 0,$$

and are given by the expression

$$T_n(x) = \tfrac{1}{2}\big[\big(x + \sqrt{x^2 - 1}\big)^n + \big(x - \sqrt{x^2 - 1}\big)^n\big].$$

Thus the problem just involves a multiplicative rewriting of the Chebyshev polynomials:

$$P_n(x) = 2^{n-1}\,T_n(x).$$

You can try to solve it anyway without this information.

Problem 1.56, in the way stated, is not immediately related to the Chebyshev polynomials, and it takes some time — even for a more experienced person — to make the connection.[5] However, Problem 1.57 is a standard exercise from college textbooks. Here is another variant:

Problem 1.58 (Sweden 1996). *For all integers $n \geq 1$ the functions p_n are defined for $x \geq 1$ by*

$$p_n(x) = \tfrac{1}{2}\left[\left(x + \sqrt{x^2 - 1}\right)^n + \left(x - \sqrt{x^2 - 1}\right)^n\right].$$

Show that $p_n(x) \geq 1$ and that $p_{mn}(x) = p_m(p_n(x))$.

Again, you can try to solve it independently of what is presented here.

The next problem and its solution share several common ideas with the solution of the previous problem.

Problem 1.59 (Turkey 1998). *Let $\{a_n\}$ be the sequence of real numbers defined by $a_1 = t$ and*

$$a_{n+1} = 4a_n(1 - a_n) \quad \textit{for } n \geq 1.$$

For how many distinct values of t do we have $a_{1998} = 0$?

Solution. We define the function $f(x) = 4x(1 - x)$. Then the given sequence becomes[6] $a_1 = t$, $a_2 = f(t)$, $a_3 = f^2(t)$, and so on. Therefore, the problem asks us to find the distinct roots of the equation $f^{1997}(t) = 0$.

The image $f(x)$ of x will be in $[0, 1]$ if $x \in [0, 1]$. To have any roots, t must be in $[0, 1]$. In such a case we can set $t = \sin^2\theta$, with $\theta \in [0, \pi/2]$. Then

$$f(t) = f(\sin^2\theta) = 4\sin^2\theta(1 - \sin^2\theta) = (2\sin\theta\,\cos\theta)^2 = \sin^2(2\theta).$$

Inductively,

$$\begin{aligned} f^2(t) &= f(f(t)) = f(\sin^2(2\theta)) = \sin^2(4\theta), \\ \cdots &\quad \cdots \quad \cdots \\ f^n(t) &= \sin^2(2^n\theta). \end{aligned}$$

The roots of $f^n(t) = 0$ are, then, those satisfying $2^n\theta = k\pi$, $k \in \mathbb{Z}$ or, more precisely,

$$\theta = \frac{k\pi}{2^n} \quad \text{for } k = 0, 1, 2, \ldots, 2^{n-1}.$$

For $n = 1997$, this gives $2^{1996} + 1$ distinct values of t. □

---

[5] At least that was the situation in 1976. After this, the theory of iterations became increasingly known and popular and the IMO 1976 problem also became another textbook problem. We discuss the topic of iterations in Chapter 16.

[6] For more details on this idea, see Chapter 16.

The function $f(x) = \lambda x(1-x)$ is known as the **logistic function** and plays an important role in the topic of *chaos*. (See Section 16.7.) Sometimes it is also called the **population growth model**. This name is motivated by the following interpretation. Let $n = 1, 2, \ldots$ count the generations of a species and let p_n be the population of the species at the n-th generation. If the population enjoys an unlimited food supply and habitat, it will grow according to the law $p_{n+1} = A\,p_n$ (a geometric progression). However, as the population grows, stress develops if resources are limited. As a result, a fraction of the population dies. This "removed" population is given by $-B\,p_n^2$. Therefore, the population of each generation follows the law $p_{n+1} = Ap_n - Bp_n^2$. If we define $a_n = \lambda p_n$ and $\lambda = A/B$, then the last equation is written equivalently as $a_{n+1} = \lambda a_n(1-a_n)$.

Here is an extension of Bolzano's theorem to some discontinuous functions:

Problem 1.60 ([1], Problem E1336, November 1958). *For the function $f : [0,1] \to \mathbb{R}$, $f(0) > 0$, $f(1) < 0$ and there exists a continuous function g such that $h = f + g$ is increasing. Prove that there exists a $\xi \in (0,1)$ such that $f(\xi) = 0$.*

Solution ([1], May 1959). Set $E = \{x : f(x) \geq 0\}$. This set is nonempty since $0 \in E$. It is also bounded since at most it can be $[0,1)$. Let $s \leq 1$ be its supremum. Since h is increasing, we have, for any $x \in E$ and $s \geq x$,

$$h(s) \geq h(x) = f(x) + g(x) \geq g(x).$$

By taking the limit $x \to s$ of this inequality, we get $h(s) \geq g(s)$, since g is continuous; this, in turn, implies $f(s) \geq 0$.

Again from the increasing property of h, we have $h(s) \leq h(1)$ with $h(1) = g(1) + f(1) < g(1)$ and $h(s) = g(s) + f(s) \geq g(s)$. Therefore $g(1) > h(1) \geq h(s) \geq g(s)$. Since g is continuous, it takes all values between $g(1)$ and $g(s)$. In particular it takes the value $h(s)$. That is, there exists a $\xi \geq s$ such that $g(\xi) = h(s)$.

Now we observe that

$$h(\xi) \geq h(s) \;\Rightarrow\; h(\xi) \geq g(\xi) \;\Rightarrow\; f(\xi) \geq 0.$$

In other words, $\xi \in E$. However, by the definition of s, t cannot be greater than s; thus $\xi = s$. Consequently, the equation $g(\xi) = h(s)$ gives $f(\xi) = 0$. □

Problem 1.61 (IMO 1968). *Let f be a real-valued function defined for all real numbers x such that, for some positive a, the equation*

$$f(x+a) = \tfrac{1}{2} + \sqrt{f(x) - f(x)^2} \tag{1.7}$$

holds for all x.

(a) *Prove that the function $f(x)$ is periodic.*

(b) *For $a = 1$ give an example of a non-constant function with the required properties.*

Solution. (a) Setting $x = y + a$ in equation (1.7), we have

$$\begin{aligned}
f(y+2a) &= \tfrac{1}{2} + \sqrt{f(y+a) - f(y+a)^2} \\
&= \tfrac{1}{2} + \sqrt{\tfrac{1}{2} + \sqrt{f(y) - f(y)^2} - \left(\tfrac{1}{2} + \sqrt{f(y) - f(y)^2}\right)^2} \\
&= \tfrac{1}{2} + \sqrt{\tfrac{1}{2} + \sqrt{f(y) - f(y)^2} - \tfrac{1}{4} - f(y) + f(y)^2 - \sqrt{f(y) - f(y)^2}} \\
&= \tfrac{1}{2} + \sqrt{\tfrac{1}{4} - f(y) + f(y)^2} \\
&= \tfrac{1}{2} + \sqrt{\left(\tfrac{1}{2} - f(y)\right)^2} \\
&= \tfrac{1}{2} + \left|\tfrac{1}{2} - f(y)\right|.
\end{aligned}$$

From this equation (or the given one) we see that $f(x) \geq \frac{1}{2}$. Therefore, the absolute value on the last line is equal to $f(y) - \frac{1}{2}$ and thus

$$f(y + 2a) = f(y).$$

The function $f(x)$ is periodic with period $2a$.

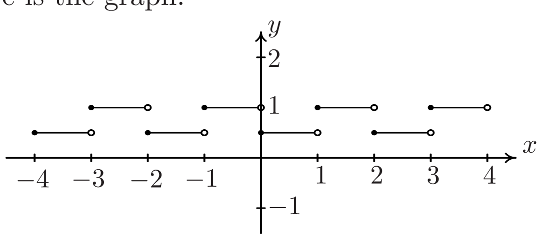
(b) The simplest nonconstant function is one taking two values. So, let

$$f(x) = \begin{cases} \frac{1}{2} & \text{if } x \in [0,1), \\ 1 & \text{if } x \in [1,2), \end{cases}$$

with the rest of the values determined by periodicity:

$$f(x) = \begin{cases} \frac{1}{2} & \text{if } x \in [2n,\, 2n+1), \\ 1 & \text{if } x \in [2n+1,\, 2n+2), \end{cases}$$

for all $n \in \mathbb{Z}$. Here is the graph:

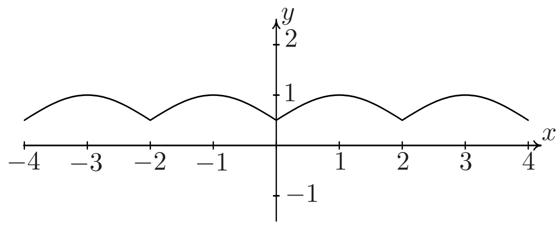


One might dislike this function being discontinuous. If a continuous function is sought, one can use

$$f(x) = \frac{1}{2} + \frac{1}{2}\sin\left(\frac{\pi x}{2}\right) \quad \text{for } x \in [0,2),$$

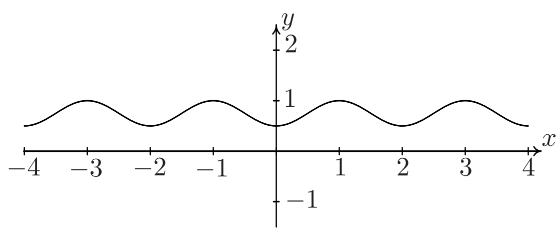
with the rest of the values determined by periodicity. The result can be written nicely in the form

$$f(x) = \frac{1}{2} + \frac{1}{2}\left|\sin\left(\frac{\pi x}{2}\right)\right| \quad \text{for } x \in \mathbb{R}.$$



Still, the function is not differentiable at the points $x = 2n$, $n \in \mathbb{Z}$. We can give another example that is differentiable at these points too:

$$f(x) = \frac{1}{2} + \frac{1}{2}\sin^2\left(\frac{\pi x}{2}\right) \quad \text{for } x \in \mathbb{R}.$$



□

Problem 1.62 ([1], Problem 11233, June/July 2006)**.** *Show that for positive integer n and for $x \neq 0$,*

$$\frac{d^n}{dx^n}\left(x^{n-1}\sin\frac{1}{x}\right) = \frac{(-1)^n}{x^{n+1}}\sin\left(\frac{1}{x} + \frac{n\pi}{2}\right).$$

Solution. We can prove this easily by induction. The identity is true for $n = 1$:

$$\begin{aligned}\frac{d}{dx}\sin\frac{1}{x} &= -\frac{1}{x^2}\cos\frac{1}{x}\\ &= -\frac{1}{x^2}\sin\left(\frac{1}{x} + \frac{\pi}{2}\right).\end{aligned}$$

Let it be true for some $n = k$:

$$\frac{d^k}{dx^k}\left(x^{k-1}\sin\frac{1}{x}\right) = \frac{(-1)^k}{x^{k+1}}\sin\left(\frac{1}{x} + \frac{k\pi}{2}\right).$$

Using this and Theorem 1.43, we find

$$\begin{aligned}\frac{d^{k+1}}{dx^{k+1}}\left(x^k \sin\frac{1}{x}\right) &= \frac{d^{k+1}}{dx^{k+1}}\left(x\,x^{k-1}\sin\frac{1}{x}\right)\\ &= (k{+}1)\,\frac{d^k}{dx^k}\left(x^{k-1}\sin\frac{1}{x}\right) + x\,\frac{d^{k+1}}{dx^{k+1}}\left(x^{k-1}\sin\frac{1}{x}\right)\\ &= (k{+}1)\,\frac{(-1)^k}{x^{k+1}}\sin\left(\frac{1}{x}+\frac{k\pi}{2}\right) + x\,\frac{d}{dx}\,\frac{(-1)^k}{x^{k+1}}\sin\left(\frac{1}{x}+\frac{k\pi}{2}\right).\end{aligned}$$

The derivative of the product (second summand on the right-hand side) gives two terms, one of which exactly cancels the remaining term. So, finally

$$\begin{aligned}\frac{d^{k+1}}{dx^{k+1}}\left(x^k \sin\frac{1}{x}\right) &= \frac{(-1)^{k+1}}{x^{k+2}}\cos\left(\frac{1}{x}+\frac{k\pi}{2}\right)\\ &= \frac{(-1)^{k+1}}{x^{k+2}}\sin\left(\frac{1}{x}+\frac{(k+1)\pi}{2}\right).\end{aligned}$$

Therefore the identity is true for $n = k{+}1$ and consequently for all $n \in \mathbb{N}^*$. □

By the same technique, you may generalize the previous result to the following.

Problem 1.63. *If f is an n-times differentiable function, then*

$$\frac{d^n}{dx^n}\left[x^{n-1} f\left(\frac{1}{x}\right)\right] = \frac{(-1)^n}{x^{n+1}}\, f^{(n)}\left(\frac{1}{x}\right).$$

If this is too straightforward, a more challenging problem would be:

Problem 1.64. *If f is an n-times differentiable function, then find*

$$\frac{d^n}{dx^n}\left[x^m f\left(\frac{1}{x}\right)\right].$$

If you fail, the answer (but not the proof) is found in the January 2008 issue of [1].

Problem 1.65 ([1], Problem E3214, June/July 1987)**.** *Let f be a real function with $n+1$ derivatives on $[a, b]$. Suppose $f^{(k)}(a) = f^{(k)}(b) = 0$ for $k = 0, 1, \ldots, n$. Prove that there is a number $\xi \in (a, b)$ such that $f^{(n+1)}(\xi) = f(\xi)$.*

Following the solution given by R. Brooks ([1], vol. **96**, p. 740), we split the solution into two parts: a simple lemma (the case of $n = 0$) and the proof for a general n.

Lemma 1.66. *Let f be a differentiable function on $[a, b]$. Suppose $f(a) = f(b) = 0$. Prove that there is a number $\xi \in (a, b)$ such that $f'(\xi) = f(\xi)$.*

Proof. Consider the function

$$g(x) = e^{-x} f(x),$$

which satisfies the conditions of Rolle's theorem on $[a, b]$. Then there exists a $\xi \in (a, b)$ such that $g'(\xi) = 0$ or

$$e^{-\xi} \left(f(\xi) - f'(\xi) \right) = 0.$$

From this, it is obvious that $f'(\xi) = f(\xi)$. □

Using the lemma above we can now prove the statement in the general case.

Solution. Consider the function

$$g(x) = \sum_{k=0}^{n} f^{(k)}(x),$$

which satisfies the conditions of the previous lemma. Then there is a $\xi \in (a, b)$ such that $g'(\xi) = g(\xi)$, which is easily seen to give

$$f^{(n+1)}(\xi) = f(\xi).$$ □

L. M. Levine has shown that the result holds true without requiring the vanishing of the derivatives at $x = b$. (See [1], vol. **96**, p. 740.)

Chapter 5

Cauchy's Equations

In this chapter we study four functional equations that are solved by the linear, power, exponential, and logarithmic functions. These equations were studied by Augustin Louis Cauchy, and since then they have formed the cornerstone of the theory. The derivation of the solutions of these functional equations serves as a vehicle to introduce the reader to the central ideas in functional equations. It is imperative that these simple results and the methodology behind them should be memorized (this can be done with minimal effort), as they appear often in functional problems.

5.1 The First Cauchy Equation

Problem 5.1. *Find all continuous functions $f : \mathbb{R} \to \mathbb{R}$ that satisfy the functional relation*

$$f(x+y) = f(x) + f(y) \quad \textit{for all } x, y \in \mathbb{R}. \tag{5.1}$$

Comment. Equation (5.1) is known as the **first** (or **linear**) **Cauchy functional equation**.

Solution. Setting $x = y = 0$ in the defining equation, we see that $f(0) = 2f(0)$; hence $f(0) = 0$. Then, setting $x = -y$, we obtain $f(0) = f(x) + f(-x)$, which implies

$$f(-x) = -f(x). \tag{5.2}$$

Using induction, we now check that for any natural number $n > 0$,

$$f(x_1 + x_2 + \cdots + x_n) = f(x_1) + f(x_2) + \cdots + f(x_n). \tag{5.3}$$

For $n = 2$, this is true by definition. For the induction step, assume that (5.3) is true for $n = n_0$:

$$f(x_1 + x_2 + \cdots + x_{n_0}) = f(x_1) + f(x_2) + \cdots + f(x_{n_0}).$$

To complete the induction, it suffices to check that (5.3) is true for $n = n_0 + 1$:

$$\begin{aligned} f(x_1 + x_2 + \cdots + x_{n_0} + x_{n_0+1}) &= f((x_1 + x_2 + \cdots + x_{n_0}) + x_{n_0+1}) \\ &= f(x_1 + x_2 + \cdots + x_{n_0}) + f(x_{n_0+1}) \\ &= f(x_1) + f(x_2) + \cdots + f(x_{n_0}) + f(x_{n_0+1}). \end{aligned}$$

Having proved (5.3), we now use it with $x_i = x$ for all i, obtaining

$$f(nx) = n\, f(x). \tag{5.4}$$

Next, if $x = \dfrac{m}{n}\, z$, with $m \in \mathbb{N}$ and $n \in \mathbb{N}^*$, we get

$$f(mz) = n\, f\left(\frac{m}{n} z\right),$$

which, by (5.4), yields

$$\frac{m}{n}\, f(z) = f\left(\frac{m}{n} z\right). \tag{5.5}$$

From this and (5.2), we conclude that

$$f\left(-\frac{m}{n} z\right) = -\frac{m}{n}\, f(z). \tag{5.6}$$

So far we have proved that

$$f(qz) = q\, f(z) \quad \text{for all } z \in \mathbb{R} \text{ and } q \in \mathbb{Q}. \tag{5.7}$$

In particular, if we set $c = f(1)$, we find that

$$f(q) = c\, q \quad \text{for all } q \in \mathbb{Q}.$$

Now take $r \in \mathbb{R}$. There is a sequence of rational numbers (q_n) such that

$$\lim_{n \to +\infty} q_n = r.$$

For the terms of this sequence we have $f(q_n) = c\, q_n$; since f is continuous,

$$f(r) = f\big(\lim_{n \to +\infty} q_n\big) = \lim_{n \to +\infty} f(q_n) = \lim_{n \to +\infty} c\, q_n = c\, r.$$

Thus the solutions to the problem must be of the form $f(x) = c\, x$, for fixed $c \in \mathbb{R}$. It is trivial to see that any $c \in \mathbb{R}$ works. □

Exercise 5.2. *In the first Cauchy equation, restrict the domain to $\mathbb{R}_+$; that is, consider continuous functions $f : \mathbb{R}_+ \to \mathbb{R}$ that satisfy the functional relation*

$$f(x + y) = f(x) + f(y) \quad \textit{for all } x, y \in \mathbb{R}_+.$$

Prove that any such function is still of the form $f(x) = c\, x$; that is, no new solutions are gained by restricting the domain.

5.2 The Second Cauchy Equation

Problem 5.3. *Find all continuous functions $f : \mathbb{R} \to \mathbb{R}$ that satisfy the functional relation*

$$f(x+y) = f(x)\,f(y) \quad \textit{for all } x, y \in \mathbb{R}. \tag{5.8}$$

Comment. Equation (5.8) is known as the **second** (or **exponential**) **Cauchy functional equation**.

Clearly, $f(x) \equiv 0$ gives a solution, but it's an uninteresting one. So from now on we assume f is not identically zero. We then show that f is *nowhere* zero. Indeed, if $f(x_0) = 0$ for some x_0, then

$$f(x) = f(x - x_0 + x_0) = f(x - x_0)\,f(x_0) = 0 \quad \text{for all } x \in \mathbb{R},$$

and f vanishes identically.

Solution 1. We next show that f is everywhere positive. For any $x \in \mathbb{R}$, the defining relation gives $f(x) = f(x/2)^2$, and so $f(x) \geq 0$; but then $f(x) > 0$.

Now, we already know *one* function that satisfies (5.8): the exponential function $x \mapsto e^x$. This suggests that things may simplify if we express f as an exponential; that is, we should try introducing $g(x) = \ln f(x)$, which is equivalent to $f(x) = e^{g(x)}$. Note that g is well defined since $f(x) > 0$ for all x.

The defining equation (5.8) translates in terms of g as follows:

$$g(x+y) = \ln f(x+y) = \ln(f(x)f(y)) = \ln f(x) + \ln f(y) = g(x) + g(y);$$

that is,

$$g(x+y) = g(x) + g(y) \quad \text{for all } x, y \in \mathbb{R}.$$

We already know that the continuous solutions of this equation are of the form $g(x) = c\,x$, with $c = g(1)$ any real number. From this, the function f is found to be

$$f(x) = a^x,$$

where we have set $a = e^c = f(1)$; here a can be any positive number. □

In this solution we used as a stepping stone our results for the linear Cauchy functional equation (5.1). However, one might like a proof that is independent of the results of the previous section; we do this next.

Solution 2. Setting $x = y = 0$ in the defining equation, we get $f(0) = f(0)^2$, whose solutions are $f(0) = 0$ and $f(0) = 1$. Under our assumption that f does not vanish identically, therefore, we have $f(0) = 1$.

Then, setting $x = -y$ in the defining equation, we get $f(0) = f(x)\,f(-x)$, which implies

$$f(-x) = f(x)^{-1}. \tag{5.9}$$

Also, it follows from (5.8) by induction that, for any natural number $n > 0$,

$$f(x_1 + x_2 + \cdots + x_n) = f(x_1)\,f(x_2)\,\ldots\,f(x_n). \tag{5.10}$$

This is checked in the same way as (5.3).

Setting $x_i = x$ for all i in (5.10), we obtain

$$f(nx) = f(x)^n. \tag{5.11}$$

Next, if $x = \dfrac{m}{n}\,z$, with $m \in \mathbb{N}$ and $n \in \mathbb{N}^*$, we get

$$f(my) = f\left(\frac{my}{n}\right)^n,$$

which, by (5.11), yields $f(y)^m = f\left(\dfrac{m}{n}y\right)^n$; that is to say

$$f\left(\frac{m}{n}y\right) = f(y)^{m/n}. \tag{5.12}$$

From this and (5.9), we conclude that

$$f\left(-\frac{m}{n}y\right) = f(y)^{-m/n}. \tag{5.13}$$

So far we have proved that

$$f(qz) = f(z)^q \quad \text{for all } z \in \mathbb{R} \text{ and } q \in \mathbb{Q}.$$

In particular, if we set $a = f(1)$, we find that

$$f(q) = a^q \quad \text{for all } q \in \mathbb{Q}.$$

Exactly as in the case of the first Cauchy equation, the continuity of f and the continuity of the function $q \mapsto a^q$ imply that $f(r) = a^r$ for all $r \in \mathbb{R}$. □

5.3 The Third Cauchy Equation

Problem 5.4. *Find all continuous functions $f : \mathbb{R}_+^* \to \mathbb{R}$ that satisfy the functional relation*

$$f(xy) = f(x) + f(y) \quad \textit{for all } x, y \in \mathbb{R}_+^*. \tag{5.14}$$

Comment. Equation (5.14) is known as the **third** (or **logarithmic**) **Cauchy functional equation**.

Note that here the domain is restricted to the positive reals. If we replace $\mathbb{R}_+^*$ by $\mathbb{R}$ in the problem, the only solution is $f(x) \equiv 0$. That solution is of course valid in the restricted domain as well, but again we declare it uninteresting. (See Example 2.3.)

Solution 1. Since the independent variable x takes values in $\mathbb{R}_+^*$, the change of variable $w = \ln x$ leads to a variable taking values in $\mathbb{R}$. Then the function $g : \mathbb{R} \to \mathbb{R}$ given by $g(w) = f(e^w)$, or equivalently $f(x) = g(\ln x)$, satisfies the relation

$$g(\ln x) + g(\ln y) = g\left(\ln(xy)\right) \;=\; g(\ln x + \ln y) \quad \text{for all } x, y \in \mathbb{R}_+^*.$$

Since the logarithmic function is surjective onto $\mathbb{R}$, we can write the preceding condition as

$$g(w_1) + g(w_2) = g(w_1 + w_2) \quad \text{for all } w_1, w_2 \in \mathbb{R}.$$

We know from Section 5.1 that the solutions to this functional relation are of the form $g(w) = c\,w$, with $c = g(0) = f(1)$. From this we can work back to the function f:

$$f(x) = f(e^w) = g(w) = c\,w = c \ln x = \frac{\ln x}{1/c} = \frac{\ln x}{\ln e^{1/c}},$$

where, in the last two equalities, we have excluded the case $c = 0$ (corresponding to f being identically zero). The reason for these last manipulations becomes apparent when we introduce $\gamma = e^{1/c} = e^{1/f(1)}$, which leads to

$$f(x) = \frac{\ln x}{\ln \gamma} = \log_\gamma x,$$

the general form of the solution. You should check that γ can be any positive number apart from 1. □

This solution depends on the results of Section 5.1 for the linear Cauchy functional equation (5.1). At the danger of being repetitive, we also write down a solution that does not use those results directly (along the lines of Solution 2 in Section 5.2).

Solution 2. Setting $x = y = 1$ in the defining equation (5.14), we see that

$$f(1) = 2f(1) \;\Rightarrow\; f(1) = 0\ .$$

Also, if $y = 1/x$,

$$f(1) = f(x) + f(x^{-1}) \;\Rightarrow\; f(x^{-1}) = -f(x) \; . \tag{5.15}$$

We now notice that, for any natural number $n > 0$,

$$f(x_1 x_2 \cdots x_n) = f(x_1) + f(x_2) + \cdots + f(x_n) \; . \tag{5.16}$$

This is verified using the method of induction. For $n = 2$, equation (5.16) is true by definition. Assuming that it is true for some n_0,

$$f(x_1 x_2 \cdots x_{n_0}) = f(x_1) + f(x_2) + \cdots + f(x_{n_0}),$$

we see that it is true for $n_0 + 1$:

$$\begin{aligned} f(x_1 x_2 \cdots x_{n_0} x_{n_0+1}) &= f\left((x_1 x_2 \cdots x_{n_0}) x_{n_0+1}\right) \\ &= f(x_1 x_2 \cdots x_{n_0})\, f(x_{n_0+1}) \\ &= f(x_1)\, f(x_2) \,\cdots\, f(x_{n_0})\, f(x_{n_0+1}) \; . \end{aligned}$$

Setting $x_i = x$, for all i in (5.16),

$$f(x^n) = n\, f(x) \; . \tag{5.17}$$

If, moreover, $x = y^{m/n}$, $m \in \mathbb{N}$, $n \in \mathbb{N}^*$,

$$f(y^m) = n\, f\left(y^{\frac{m}{n}}\right) \overset{(5.17)}{\Rightarrow} m\, f(y) = n\, f\left(y^{\frac{m}{n}}\right) \;\Rightarrow\; f\left(y^{\frac{m}{n}}\right) = \frac{m}{n}\, f(y) \; . \tag{5.18}$$

From this equation and (5.15), we conclude that

$$f\left(y^{-\frac{m}{n}}\right) = -\frac{m}{n}\, f(y) \; . \tag{5.19}$$

In other words, so far we have proved that

$$f(y^q) = q\, f(y) \; , \quad \forall y \in \mathbb{R} \; , \quad q \in \mathbb{Q} \; .$$

Now, let $r \in \mathbb{R}$. There is a sequence of rationals $\{q_n\}$ such that

$$\lim_{n \to \infty} q_n = r \; .$$

For the terms of the sequence $\{q_n\}$, the function f gives

$$f(y^{q_n}) = q_n\, f(y) \; ,$$

and since it is continuous,

$$f(y^r) = \lim_{n \to \infty} f(y^{q_n}) = \lim_{n \to \infty} q_n\, f(y) = r\, f(y) \; .$$

Finally, let's set $y = a = \text{const}$ and $x = a^r$. Then we find the function f in the final form:

$$f(x) = f(a)\, \log_a x = \log_b x \; ,$$

where we defined a constant b such that $f(a) = 1/\log_a b$. □

Exercise 5.5. *Solve the third Cauchy functional equation with domain* $\mathbb{R}^*$*; that is, find all continuous functions* $f : \mathbb{R}^* \to \mathbb{R}$ *that satisfy the functional relation*

$$f(xy) = f(x) + f(y) \quad \textit{for all } x, y \in \mathbb{R}^*.$$

(Recall that, by definition, the continuity requirement in this case boils down to continuity in each of the intervals $(-\infty, 0)$ and $(0, \infty)$.)

5.4 The Fourth Cauchy Equation

Problem 5.6. *Find all continuous functions* $f : \mathbb{R}_+ \to \mathbb{R}$ *that satisfy the functional relation*

$$f(xy) = f(x)\, f(y) \quad \textit{for all } x, y \in \mathbb{R}_+. \tag{5.20}$$

Comment. Equation (5.20) is known as the **fourth** (or **power**) **Cauchy functional equation**.

Solution 1. Again, $f(x) \equiv 0$ gives an uninteresting solution, so suppose f is not identically zero. We first show that f is nonzero except perhaps at the origin. Indeed, suppose $f(x_0) = 0$ for some $x_0 > 0$. Then

$$f(x) = f\Big(\frac{x}{x_0}\, x_0\Big) = f\Big(\frac{x}{x_0}\Big) f(x_0) = 0 \quad \text{for all } x \in \mathbb{R}_+,$$

and f vanishes identically.

Further, f must be nonnegative everywhere and positive away from 0. Indeed, for any $x \in \mathbb{R}_+$, we can write $f(x) = f(\sqrt{x})^2$.

Once more we introduce $g(x) = \ln f(x)$, but this time only for $x > 0$, since $f(0)$ may vanish. The defining relation (5.20) can be rewritten in terms of g by taking the logarithm of both sides, yielding

$$g(x\,y) = g(x) + g(y) \quad \text{for all } x, y \in \mathbb{R}_+^*.$$

We know from the previous section that the solutions of this equation are $g(x) \equiv 0$ and $g(x) = \log_\gamma x = \ln x / \ln \gamma$, for $\gamma \neq 1$. In the first case we have $f(x) \equiv 1$; in the second we have

$$f(x) = e^{g(x)} = x^c, \tag{5.21}$$

where $c = 1/\ln \gamma \neq 0$. In fact, the function $f(x) \equiv 1$ on $\mathbb{R}_+^*$ is also of the form $f(x) = x^c$, with $c = 0$, so we settle on (5.21) as the general form of our solution on $\mathbb{R}_+^*$.

There remains to extend f to all of $\mathbb{R}_+$. By continuity, we should pick $f(0) = 0$ if $c > 0$, and $f(0) = 1$ if $c = 0$. Clearly these extensions satisfy

(5.20). If $c < 0$, the function $f(x) = x^c$ cannot be extended continuously to the origin, so such values of c do not yield solutions of the fourth Cauchy problem as stated (they do on the domain $\mathbb{R}_+^*$). □

Once more we write down a careful alternative solution independent of the results in Section 5.1, along the lines of Solution 2 in Section 5.2.

Solution 2. One first proves that $f(x) > 0$ (as above). Setting $x = y = 1$ in the defining equation, we see that

$$f(1) = f(1)^2 \Rightarrow f(1)\,(f(1) - 1) = 0 \Rightarrow f(1) = 0,\ 1\ ,$$

and therefore we must have $f(1) = 1$.

Also, if $y = 1/x$ for $x \neq 0$,

$$f(1) = f(x)\,f(x^{-1}) \Rightarrow f(x^{-1}) = f(x)^{-1}\ . \tag{5.22}$$

We now notice that for any natural number $n > 0$

$$f(x_1 x_2 \cdots x_n) = f(x_1)\,f(x_2)\,\cdots\,f(x_n)\ . \tag{5.23}$$

This is verified using the method of induction. For $n = 2$, equation (5.23) is true by definition. Assuming, that it is true for some n_0,

$$f(x_1 x_2 \cdots x_{n_0}) = f(x_1)\,f(x_2)\,\cdots\,f(x_{n_0})\ ,$$

we see that it is true for $n_0 + 1$:

$$\begin{aligned} f(x_1 x_2 \cdots x_{n_0} x_{n_0+1}) &= f\left((x_1 x_2 \cdots x_{n_0}) x_{n_0+1}\right) \\ &= f(x_1 x_2 \cdots x_{n_0})\,f(x_{n_0+1}) \\ &= f(x_1)\,f(x_2)\,\cdots\,f(x_{n_0})\,f(x_{n_0+1})\ . \end{aligned}$$

Setting $x_i = x,\ \forall i$ in (5.23),

$$f(x^n) = f(x)^n\ . \tag{5.24}$$

If, moreover, $x = y^{m/n}$, $m \in \mathbb{N}$, $n \in \mathbb{N}^*$,

$$f(y^m) = f\left(y^{\frac{m}{n}}\right)^n \overset{(5.24)}{\Rightarrow} f(y)^m = f\left(y^{\frac{m}{n}}\right)^n \Rightarrow f\left(y^{\frac{m}{n}}\right) = f(y)^{m/n}\ . \tag{5.25}$$

From this equation and (5.22), we conclude that

$$f\left(y^{-\frac{m}{n}}\right) = f(y)^{-m/n}\ . \tag{5.26}$$

In other words, so far we have proved that

$$f(y^q) = f(y)^q \ , \quad \forall y \in \mathbb{R}_+^* \ , \quad q \in \mathbb{Q} \ .$$

Now, let $r \in \mathbb{R}$. There is a sequence of rationals $\{q_n\}$ such that

$$\lim_{n\to\infty} q_n = r \ .$$

For the terms of the sequence $\{q_n\}$, the function f gives

$$f(y^{q_n}) = f(y)^{q_n} \ ,$$

and since it is continuous,

$$f(y^r) = \lim_{n\to\infty} f(y^{q_n}) = \lim_{n\to\infty} f(y)^{q_n} = f(y)^r \ .$$

Finally, let's set $y = a = \text{const}$ and $x = a^r$. Then we find the function f in the form:

$$f(x) = f(a)^{\log_a x} \ .$$

If $f(a) = 1$, then $f(x) = 1$. If $f(a) \neq 1$, we define a constant c such that $f(a) = a^c$. Then

$$f(x) = x^c \ , \quad c \neq 0 \ .$$

□

Question. *Define the sign function* $\text{sgn}(x)$ *by*

$$\text{sgn}(x) = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ +1 & \text{if } x > 0. \end{cases}$$

Show that it satisfies $\text{sgn}(x\,y) = \text{sgn}(x)\,\text{sgn}(y)$*; in other words, it is a solution of the functional equation* (5.20)*. However, it is not a power function. Explain.*

Answer. The sign function is not continuous: it has a jump discontinuity at $x = 0$. When we presented the definition of functional equations in Section 2.1, we mentioned that the set of solutions strongly depends on the conditions imposed on the functions. When continuity is assumed, the Cauchy functional equations have a restricted set of solutions, exactly as we have seen above. If this condition is relaxed, then the set of solutions greatly enlarges. Discontinuous solutions of the Cauchy equations will be studied in Chapter 11.

Exercise 5.7. *Solve the fourth Cauchy functional equation with domain* $\mathbb{R}$*; that is, find all continuous functions* $f : \mathbb{R} \to \mathbb{R}$ *that satisfy the functional relation*

$$f(xy) = f(x)f(y) \quad \text{for all } x, y \in \mathbb{R}.$$

5.5 Solved Problems

Often, functional equations can be solved by a reduction to one of Cauchy's functional equations. In this section, we present several examples.

Problem 5.8 ([46]). *Find the continuous solutions $f : \mathbb{R} \to \mathbb{R}$ of the functional relation*

$$f(x+y) = A^y f(x) + A^x f(y) \quad \text{for all } x, y \in \mathbb{R},$$

where A is a positive constant.

Solution. Assume f is a continuous solution, and define a continuous function g by $g(x) = A^{-x} f(x)$, or equivalently $f(x) = A^x g(x)$. To reexpress the defining functional equation in terms of g, we observe that

$$g(x+y) = A^{-(x+y)} f(x+y) = A^{-x} f(x) + A^{-y} f(y) = g(x) + g(y).$$

We have thus reduced the problem to the first Cauchy functional equation, whose solutions are of the form $g(x) = c\,x$, where c is any constant. Therefore $f(x) = c\,x\,A^x$. □

Problem 5.9. *Find the continuous solutions $f : \mathbb{R} \to \mathbb{R}$ of the functional equation*

$$f\left(\frac{x+y}{2}\right) = \frac{f(x) + f(y)}{2} \quad \text{for all } x, y \in \mathbb{R}. \tag{5.27}$$

Comment. This is known as the **Jensen functional equation**. It can also be written in the equivalent form

$$f(x+y) + f(x-y) = 2\,f(x) \quad \text{for all } x, y \in \mathbb{R}.$$ □

Solution. For $y = 0$, equation (5.27) gives

$$f\left(\frac{x}{2}\right) = \frac{f(x) + f(0)}{2} = \frac{f(x) + b}{2},$$

where $b = f(0)$. In this relation we substitute $y + z$ for x to obtain

$$f\left(\frac{y+z}{2}\right) = \frac{f(y+z) + b}{2}.$$

The left-hand side can be rewritten from the defining relation (5.27) and thus

$$\frac{f(y) + f(z)}{2} = \frac{f(y+z) + b}{2},$$

or

$$f(y+z) = f(y) + f(z) - b.$$

Now consider the continuous function g defined by $g(x) = f(x) - b$. In terms of it, the functional equation becomes

$$g(x+y) = g(x) + g(y),$$

whose solutions are of the form $g(x) = c\,x$, where c is any constant. Therefore $f(x) = c\,x + b$, where c and b are arbitrary. □

Problem 5.10. *Find the continuous solutions $f : \mathbb{R} \to \mathbb{R}$ of the functional equation*

$$f\left(\sqrt{\frac{x^2+y^2}{2}}\right) = \sqrt{\frac{f(x)^2+f(y)^2}{2}} \quad \textit{for all } x, y \in \mathbb{R}.$$

Solution. From the defining equation we see that $f(x) \geq 0$ for all $x \in \mathbb{R}_+$, since the square root of a number is a nonnegative number. Define $F(x) = f(\sqrt{x})^2$ for $x \geq 0$. Then the defining equation becomes

$$F\left(\frac{u+v}{2}\right) = \frac{F(u)+F(v)}{2},$$

where $u = x^2$, $v = y^2$. Therefore,

$$F(u) = c\,u + b.$$

The function $F(u)$ will be positive for all values of its argument u if both b and c are nonnegative constants. Therefore,

$$f(x) = \sqrt{c\,x^2 + b}, \quad b, c \geq 0,$$

for $x \geq 0$.

To find the values of f at negative values, we observe that the defining equation is unchanged under the substitution $x \mapsto -x$. This leads to $f(x)^2 = f(-x)^2$; hence $f(-x) = \pm f(x)$. Since f is continuous, its graph for negative values consists of a series of odd and even parts (relative to the branch of positive x) with transitions happening at the roots of f. And since for each nonpositive root we must have a nonnegative root, we conclude that the only possible root is $x = 0$ or $f(x) \equiv 0$. From this discussion we arrive at the following solutions of the given functional equation:

$$\begin{aligned} f(x) &= 0, \\ f(x) &= \sqrt{c\,x^2 + b}, \quad c \geq 0,\ b > 0, \\ f(x) &= a\,|x|, \quad a > 0, \\ f(x) &= a\,x, \qquad a > 0. \end{aligned}$$

□

Problem 5.11. *Find the continuous solutions $f : \mathbb{R} \to \mathbb{R}$ of the functional equation*

$$f(x+y) = \alpha^{xy} f(x) f(y) \quad \text{for all } x, y \in \mathbb{R},$$

where α is a positive constant.

Solution. Define the continuous function $g(x) = \alpha^{-x^2/2} f(x)$. Then the functional equation takes the form

$$g(x+y) = g(x)\, g(y),$$

which has the solutions $g(x) = 0$ or $g(x) = b^x$, where b is a positive constant. Therefore $f(x) = 0$ or $f(x) = b^x \alpha^{x^2/2}$. □

Problem 5.12. *Find the continuous solutions $f : \mathbb{R} \to \mathbb{R}$ of the functional equation*

$$f(x+y) = f(x) + f(y) + f(x) f(y) \quad \text{for all } x, y \in \mathbb{R}.$$

Solution. Define the continuous function $g(x) = f(x)+1$. Then the functional equation takes the form

$$g(x+y) = g(x)\, g(y),$$

which has the solutions $g(x) = 0$ or $g(x) = b^x$, where b a positive constant. Therefore $f(x) = -1$ or $f(x) = b^x - 1$. □

Problem 5.13 (Putnam 1947)**.** *Find the continuous solutions $f : \mathbb{R} \to \mathbb{R}$ of the functional equation*

$$f(\sqrt{x^2+y^2}) = f(x) f(y) \quad \text{for all } x, y \in \mathbb{R}.$$

Solution. If f vanishes identically, it satisfies the equation. In the opposite case, there exists $x_0 \in \mathbb{R}$ such that $f(x_0) \neq 0$; then

$$f(\sqrt{x_0^2+y^2}) = f(x_0)\, f(y) = f(x_0)\, f(-y),$$

or $f(y) = f(-y)$, for all $y \in \mathbb{R}$. This shows that f is an even function. It is enough to search for the form of $f(x)$ when $x \geq 0$.

Define the continuous function $g(x) = f(\sqrt{x})$, $x \geq 0$. Then the functional equation takes the Cauchy form

$$g(u+v) = g(u)\, g(v) \quad \text{for all } u, v \in \mathbb{R}_+$$

(via the substitutions $u = x^2$ and $v = y^2$). This has the nonvanishing solutions $g(u) = a^u$, where a is any positive constant. Therefore $f(x) = a^{x^2}$. □

Problem 5.14. *Find the continuous solutions $f : \mathbb{R} \to \mathbb{R}$ of the functional equation*

$$f(\sqrt[n]{x^n + y^n}) = f(x) + f(y) \quad \textit{for all } x, y \in \mathbb{R}, \tag{5.28}$$

where n is a given positive natural number.

Solution. Here, the reduction to a Cauchy equation is achieved by "factoring" the function f as a composition of functions, one of which is $x \mapsto x^n$. Let $g : \mathbb{R} \to \mathbb{R}$ be defined by $g(x^n) = f(x)$. Then (5.28) is equivalent to

$$g(x^n + y^n) = g(x^n) + g(y^n) \quad \text{for all } x, y \in \mathbb{R}. \tag{5.29}$$

Take first the case of n odd, where $x \mapsto x^n$ is one-to-one and onto. Then, since x^n and y^n can take any real value, 5.29 is equivalent to

$$g(x + y) = g(x) + g(y) \quad \text{for all } x, y \in \mathbb{R},$$

whose solutions, as we know, are of the form $g(x) = c\,x$ for fixed $c \in \mathbb{R}$. Translating back to f, we get $f(x) = g(x^n) = c\,x^n$.

If n is even we use a similar reasoning, but we must consider that the equation $g(x^n) = f(x)$ only defines g for nonnegative values of the argument. Because x^n and y^n range over $[0, +\infty)$ only, (5.29) is equivalent to

$$g(x + y) = g(x) + g(y) \quad \text{for all } x, y \in \mathbb{R}_+.$$

Fortunately this is still enough to guarantee that $g(x) = c\,x$ for fixed $c \in \mathbb{R}$ (see Exercise 5.2). Therefore the solutions in the case of n even are again the functions f of the form $f(x) = g(x^n) = c\,x^n$. ☐

Problem 5.15 (Romania 1997)**.** *Find all continuous solutions $f : \mathbb{R} \to [0, +\infty)$ such that*

$$f(x^2 + y^2) = f(x^2 - y^2) + f(2xy).$$

Solution. The key here is to observe that the three arguments are related by the equation

$$(x^2 + y^2)^2 = (x^2 - y^2)^2 + (2xy)^2.$$

Thus, if we set $\alpha = x^2 - y^2$ and $\beta = 2xy$, the defining equation is equivalent to

$$f(\sqrt{\alpha^2 + \beta^2}) = f(\alpha) + f(\beta) \quad \text{for } \alpha, \beta \in \mathbb{R}.$$

(Why are all values of α and β achieved?) From the previous problem we know that the general solution is $f(x) = a\,x^2$, where a is a real constant. ☐

Problem 5.16. *Find the continuous solutions $f : \mathbb{R} \to \mathbb{R}$ of the functional equation*

$$f(x+y) = x^2y + xy^2 - 2xy + f(x) + f(y) \quad \textit{for all } x, y \in \mathbb{R}.$$

Solution. Define the continuous function $g(x) = f(x) - x^3/3 + x^2$. Then the functional equation takes the Cauchy form

$$g(x+y) = g(x) + g(y).$$

This equation has the general solution $g(x) = c\,x$, where a is any real constant. Therefore $f(x) = c\,x + x^3/3 - x^2$. □

Problem 5.17. *Show that the functional equation*

$$f\left(\frac{x+y}{2}\right)^2 = f(x)\,f(y) \tag{5.30}$$

is equivalent to the functional equation

$$f(x)^2 = f(x+y)\,f(x-y). \tag{5.31}$$

Then find the continuous solutions $f : \mathbb{R} \to \mathbb{R}$ of (5.30).

Comment. Equation (5.31) is known as the **Lobachevsky functional equation**.

Solution. Setting $x = a + b$ and $y = a - b$ in (5.30), we find (5.31).

Setting $y = 0$ in (5.30), we find

$$f\left(\frac{x}{2}\right)^2 = f(x)\,f(0).$$

If $f(0) = 0$, then $f(x/2)^2 = 0$, for all $x \in \mathbb{R}$. This implies that $f(x) = 0$ for all $x \in \mathbb{R}$. If $f(0) \neq 0$, we replace x with $x + y$; hence

$$f\left(\frac{x+y}{2}\right)^2 = f(x+y)\,f(0),$$

or, after using (5.30),

$$f(x)\,f(y) = f(x+y)\,f(0).$$

Set $b = f(0)$ and define the continuous function $g(x) = f(x)/b$. Then the functional equation takes the Cauchy form

$$g(x+y) = g(x)\,g(y).$$

This equation has the solutions $g(x) = 0$ or $g(x) = a^x$, for $a \in \mathbb{R}_+^*$. Working our way back to f we conclude that the general form of the solution is $f(x) = b\,a^x$, with $a \in \mathbb{R}_+^*$ and $b \in \mathbb{R}$. □

Here is an alternative formulation of the previous problem:

Exercise 5.18. *Find all continuous functions that map three successive terms of any arithmetic progression to three successive terms of a geometric progression.*

Can you see the relation?

Problem 5.19 ([46]). *Find all continuous solutions $f : \mathbb{R} \to \mathbb{R}$ of the functional equation*

$$f(x+y) = f(x) + f(y) + a\,(1-b^x)\,(1-b^y), \tag{5.32}$$

where a, b are real constants and $b > 0$.

Solution. Set

$$g(x) = f(x) - a\,(b^x - 1). \tag{5.33}$$

Then the function $g(x)$ satisfies

$$g(x+y) = g(x) + g(y),$$

with solution $g(x) = c\,x$ with c a real constant. Therefore

$$f(x) = a\,(b^x - 1) + c\,x. \qquad \square$$

Problem 5.20. *Let $a, b \in \mathbb{R}$. Find the continuous solutions $f : \mathbb{R} \to \mathbb{R}$ of the functional equation*

$$f(x+y+a) + b = f(x) + f(y) \quad \textit{for all } x, y \in \mathbb{R}. \tag{5.34}$$

Solution. Define the continuous function $g(x) = f(x-a) - b$. Then (5.34) takes the form

$$g(x+y+2a) = g(x+a) + g(y+a),$$

or

$$g(u+v) = g(u) + g(v),$$

where $u = x + a$ and $v = y + a$. By (5.1), $g(u) = c\,u$, where c is a constant. Therefore $f(x) = c(x+a) + b$. $\square$

Problem 5.21 (Romania TST 2006). *Let $r, s \in \mathbb{Q}$. Find all functions $f : \mathbb{Q} \to \mathbb{Q}$ such that for all $x, y \in \mathbb{Q}$ we have*

$$f\,(x + f(y)) = f(x+r) + y + s. \tag{5.35}$$

Solution. It is easy to verify that $f(x) = \pm(x+s)+r$ are solutions. We show that they are the only solutions.

Adding $z-r$ to both sides of (5.35) and finding the value of f at the points thus obtained, we get

$$f\big(z-r+f\left(x+f(y)\right)\big) = f\big(z+y+s-r+f(x+r)\big). \tag{5.36}$$

Applying (5.35) to each side of (5.36), we obtain

$$f(z)+x+f(y)+s = f(z+y+s)+x+r+s,$$

or

$$f(y)+f(z) = f(y+z+s)+r,$$

which is (5.34). Therefore $f(x) = c(x+s)+r$, where c is a constant. Substituting $f(x) = c(x+s)+r$ into (5.35), we find

$$c\big(x+c(y+s)+r+s\big)+r = c(x+r+s)+r+y+s,$$

or

$$(c^2-1)(y+s) = 0.$$

Hence $c = \pm 1$ and $f(x) = \pm(x+s)+r$. □

As an easy extension of the previous problem, the reader should try:

Exercise 5.22. *Let $a, b \in \mathbb{R}$. Show that $f(x) = \pm(x+a)+b$ are the only continuous solutions to the functional equation*

$$f\left(x+f(y)\right) = f(x+a)+y+b \quad \textit{for all } x, y \in \mathbb{R}.$$