An Introduction to Real Analysis

John K. Hunter

Department of Mathematics, University of California at Davis

ABSTRACT. These are some notes on introductory real analysis. They cover the properties of the real numbers, sequences and series of real numbers, limits of functions, continuity, differentiability, sequences and series of functions, and Riemann integration. They don't include multi-variable calculus or contain any problem sets. Optional sections are starred.

Contents

Sets and Functions

> We understand a "set" to be any collection M of certain distinct objects of our thought or intuition (called the "elements" of M) into a whole. (Georg Cantor, 1895)

> In mathematics you don't understand things. You just get used to them. (Attributed to John von Neumann)

In this chapter, we define sets, functions, and relations and discuss some of their general properties. This material can be referred back to as needed in the subsequent chapters.

1.1. Sets

A set is a collection of objects, called the elements or members of the set. The objects could be anything (planets, squirrels, characters in Shakespeare's plays, or other sets) but for us they will be mathematical objects such as numbers, or sets of numbers. We write $x \in X$ if x is an element of the set X and $x \notin X$ if x is not an element of X.

If the definition of a "set" as a "collection" seems circular, that's because it is. Conceiving of many objects as a single whole is a basic intuition that cannot be analyzed further, and the the notions of "set" and "membership" are primitive ones. These notions can be made mathematically precise by introducing a system of axioms for sets and membership that agrees with our intuition and proving other set-theoretic properties from the axioms.

The most commonly used axioms for sets are the ZFC axioms, named somewhat inconsistently after two of their founders (Zermelo and Fraenkel) and one of their axioms (the Axiom of Choice). We won't state these axioms here; instead, we use "naive" set theory, based on the intuitive properties of sets. Nevertheless, all the set-theory arguments we use can be rigorously formalized within the ZFC system.

Sets are determined entirely by their elements. Thus, the sets X, Y are equal, written $X = Y$, if

$$x \in X \quad \text{if and only if} \quad x \in Y.$$

It is convenient to define the empty set, denoted by $\varnothing$, as the set with no elements. (Since sets are determined by their elements, there is only one set with no elements!) If $X \neq \varnothing$, meaning that X has at least one element, then we say that X is non-empty.

We can define a finite set by listing its elements (between curly brackets). For example,

$$X = \{2, 3, 5, 7, 11\}$$

is a set with five elements. The order in which the elements are listed or repetitions of the same element are irrelevant. Alternatively, we can define X as the set whose elements are the first five prime numbers. It doesn't matter how we specify the elements of X, only that they are the same.

Infinite sets can't be defined by explicitly listing all of their elements. Nevertheless, we will adopt a realist (or "platonist") approach towards arbitrary infinite sets and regard them as well-defined totalities. In constructive mathematics and computer science, one may be interested only in sets that can be defined by a rule or algorithm — for example, the set of all prime numbers — rather than by infinitely many arbitrary specifications, and there are some mathematicians who consider infinite sets to be meaningless without some way of constructing them. Similar issues arise with the notion of arbitrary subsets, functions, and relations.

1.1.1. Numbers. The infinite sets we use are derived from the natural and real numbers, about which we have a direct intuitive understanding.

Our understanding of the natural numbers $1, 2, 3, \ldots$ derives from counting. We denote the set of natural numbers by

$$\mathbb{N} = \{1, 2, 3, \ldots\}.$$

We define $\mathbb{N}$ so that it starts at 1. In set theory and logic, the natural numbers are defined to start at zero, but we denote this set by $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. Historically, the number 0 was later addition to the number system, primarily by Indian mathematicians in the 5th century AD. The ancient Greek mathematicians, such as Euclid, defined a number as a multiplicity and didn't consider 1 to be a number either.

Our understanding of the real numbers derives from durations of time and lengths in space. We think of the real line, or continuum, as being composed of an (uncountably) infinite number of points, each of which corresponds to a real number, and denote the set of real numbers by $\mathbb{R}$. There are philosophical questions, going back at least to Zeno's paradoxes, about whether the continuum can be represented as a set of points, and a number of mathematicians have disputed this assumption or introduced alternative models of the continuum. There are, however, no known inconsistencies in treating $\mathbb{R}$ as a set of points, and since Cantor's work it has been the dominant point of view in mathematics because of its precision, power, and simplicity.

We denote the set of (positive, negative and zero) integers by

$$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\},$$

and the set of rational numbers (ratios of integers) by

$$\mathbb{Q} = \{p/q : p, q \in \mathbb{Z} \text{ and } q \neq 0\}.$$

The letter "Z" comes from "zahl" (German for "number") and "Q" comes from "quotient." These number systems are discussed further in Chapter 2.

Although we will not develop any complex analysis here, we occasionally make use of complex numbers. We denote the set of complex numbers by

$$\mathbb{C} = \{x + iy : x, y \in \mathbb{R}\},$$

where we add and multiply complex numbers in the natural way, with the additional identity that $i^2 = -1$, meaning that i is a square root of -1. If $z = x + iy \in \mathbb{C}$, we call $x = \Re z$ the real part of z and $y = \Im z$ the imaginary part of z, and we call

$$|z| = \sqrt{x^2 + y^2}$$

the absolute value, or modulus, of z. Two complex numbers $z = x + iy$, $w = u + iv$ are equal if and only if $x = u$ and $y = v$.

1.1.2. Subsets. A set A is a subset of a set X, written $A \subset X$ or $X \supset A$, if every element of A belongs to X; that is, if

$$x \in A \text{ implies that } x \in X.$$

We also say that A is included in X.[1] For example, if P is the set of prime numbers, then $P \subset \mathbb{N}$, and $\mathbb{N} \subset \mathbb{R}$. The empty set $\varnothing$ and the whole set X are subsets of any set X. Note that $X = Y$ if and only if $X \subset Y$ and $Y \subset X$; we often prove the equality of two sets by showing that each one includes the other.

In our notation, $A \subset X$ does not imply that A is a proper subset of X (that is, a subset of X not equal to X itself), and we may have $A = X$. This notation for non-strict inclusion is not universal; some authors use $A \subset X$ to denote strict inclusion, in which $A \neq X$, and $A \subseteq X$ to denote non-strict inclusion, in which $A = X$ is allowed.

Definition 1.1. The power set $\mathcal{P}(X)$ of a set X is the set of all subsets of X.

Example 1.2. If $X = \{1, 2, 3\}$, then

$$\mathcal{P}(X) = \{\varnothing, \{1\}, \{2\}, \{3\}, \{2, 3\}, \{1, 3\}, \{1, 2\}, \{1, 2, 3\}\}.$$

The power set of a finite set with n elements has 2^n elements because, in defining a subset, we have two independent choices for each element (does it belong to the subset or not?). In Example 1.2, X has 3 elements and $\mathcal{P}(X)$ has $2^3 = 8$ elements.

The power set of an infinite set, such as $\mathbb{N}$, consists of all finite and infinite subsets and is infinite. We can define finite subsets of $\mathbb{N}$, or subsets with finite

---

[1]By contrast, we say that an element $x \in X$ is *contained* in X, in which cases the singleton set $\{x\}$ is *included* in X. This terminological distinction is not universal, but it is almost always clear from the context whether one is referring to an element of a set or a subset of a set. In fact, before the development of the contemporary notation for set theory, Dedekind [**3**] used the same symbol ($\subseteq$) to denote both membership of elements and inclusion of subsets.

complements, by listing finitely many elements. Some infinite subsets, such as the set of primes or the set of squares, can be defined by giving a definite rule for membership. We imagine that a general subset $A \subset \mathbb{N}$ is "defined" by going through the elements of $\mathbb{N}$ one by one and deciding for each $n \in \mathbb{N}$ whether $n \in A$ or $n \notin A$.

If X is a set and P is a property of elements of X, we denote the subset of X consisting of elements with the property P by $\{x \in X : P(x)\}$.

Example 1.3. The set

$$\left\{ n \in \mathbb{N} : n = k^2 \text{ for some } k \in \mathbb{N} \right\}$$

is the set of perfect squares $\{1, 4, 9, 16, 25, \dots\}$. The set

$$\{ x \in \mathbb{R} : 0 < x < 1 \}$$

is the open interval $(0, 1)$.

1.1.3. Set operations. The intersection $A \cap B$ of two sets A, B is the set of all elements that belong to both A and B; that is

$$x \in A \cap B \text{ if and only if } x \in A \text{ and } x \in B.$$

Two sets A, B are said to be disjoint if $A \cap B = \varnothing$; that is, if A and B have no elements in common.

The union $A \cup B$ is the set of all elements that belong to A or B; that is

$$x \in A \cup B \text{ if and only if } x \in A \text{ or } x \in B.$$

Note that we always use 'or' in an inclusive sense, so that $x \in A \cup B$ if x is an element of A or B, or both A and B. (Thus, $A \cap B \subset A \cup B$.)

The set-difference of two sets B and A is the set of elements of B that do not belong to A,

$$B \setminus A = \{x \in B : x \notin A\} .$$

If we consider sets that are subsets of a fixed set X that is understood from the context, then we write $A^c = X \setminus A$ to denote the complement of $A \subset X$ in X. Note that $(A^c)^c = A$.

Example 1.4. If

$$A = \{2, 3, 5, 7, 11\} , \qquad B = \{1, 3, 5, 7, 9, 11\}$$

then

$$A \cap B = \{3, 5, 7, 11\} , \qquad A \cup B = \{1, 2, 3, 5, 7, 9, 11\} .$$

Thus, $A \cap B$ consists of the natural numbers between 1 and 11 that are both prime and odd, while $A \cup B$ consists of the numbers that are either prime or odd (or both). The set differences of these sets are

$$B \setminus A = \{1, 9\} , \qquad A \setminus B = \{2\} .$$

Thus, $B \setminus A$ is the set of odd numbers between 1 and 11 that are not prime, and $A \setminus B$ is the set of prime numbers that are not odd.

These set operations may be represented by Venn diagrams, which can be used to visualize their properties. In particular, if $A, B \subset X$, we have De Morgan's laws:

$$(A \cup B)^c = A^c \cap B^c, \qquad (A \cap B)^c = A^c \cup B^c.$$

The definitions of union and intersection extend to larger collections of sets in a natural way.

Definition 1.5. Let $\mathcal{C}$ be a collection of sets. Then the union of $\mathcal{C}$ is

$$\bigcup \mathcal{C} = \{x : x \in X \text{ for some } X \in \mathcal{C}\},$$

and the intersection of $\mathcal{C}$ is

$$\bigcap \mathcal{C} = \{x : x \in X \text{ for every } X \in \mathcal{C}\}.$$

If $\mathcal{C} = \{A, B\}$, then this definition reduces to our previous one for $A \cup B$ and $A \cap B$.

The Cartesian product $X \times Y$ of sets X, Y is the set of all ordered pairs (x, y) with $x \in X$ and $y \in Y$. If $X = Y$, we often write $X \times X = X^2$. Two ordered pairs (x_1, y_1), (x_2, y_2) in $X \times Y$ are equal if and only if $x_1 = x_2$ and $y_1 = y_2$. Thus, $(x, y) \neq (y, x)$ unless $x = y$. This contrasts with sets where $\{x, y\} = \{y, x\}$.

Example 1.6. If $X = \{1, 2, 3\}$ and $Y = \{4, 5\}$ then

$$X \times Y = \{(1, 4), (1, 5), (2, 4), (2, 5), (3, 4), (3, 5)\}.$$

Example 1.7. The Cartesian product of $\mathbb{R}$ with itself is the Cartesian plane $\mathbb{R}^2$ consisting of all points with coordinates (x, y) where $x, y \in \mathbb{R}$.

The Cartesian product of finitely many sets is defined analogously.

Definition 1.8. The Cartesian products of n sets X_1, $X_2, \ldots, X_n$ is the set of ordered n-tuples,

$$X_1 \times X_2 \times \cdots \times X_n = \{(x_1, x_2, \ldots, x_n) : x_i \in X_i \text{ for } i = 1, 2, \ldots, n\},$$

where $(x_1, x_2, \ldots, x_n) = (y_1, y_2, \ldots, y_n)$ if and only if $x_i = y_i$ for every $i = 1, 2, \ldots, n$.

1.2. Functions

A function $f : X \to Y$ between sets X, Y assigns to each $x \in X$ a unique element $f(x) \in Y$. Functions are also called maps, mappings, or transformations. The set X on which f is defined is called the domain of f and the set Y in which it takes its values is called the codomain. We write $f : x \mapsto f(x)$ to indicate that f is the function that maps x to $f(x)$.

Example 1.9. The identity function $\mathrm{id}_X : X \to X$ on a set X is the function $\mathrm{id}_X : x \mapsto x$ that maps every element to itself.

Example 1.10. Let $A \subset X$. The characteristic (or indicator) function of A,

$$\chi_A : X \to \{0, 1\},$$

is defined by

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

Specifying the function χ_A is equivalent to specifying the subset A.

Example 1.11. Let A, B be the sets in Example 1.4. We can define a function $f : A \to B$ by

$$f(2) = 7, \quad f(3) = 1, \quad f(5) = 11, \, f(7) = 3, \quad f(11) = 9,$$

and a function $g : B \to A$ by

$$g(1) = 3, \quad g(3) = 7, \quad g(5) = 2, \, g(7) = 2, \quad g(9) = 5, \quad g(11) = 11.$$

Example 1.12. The square function $f : \mathbb{N} \to \mathbb{N}$ is defined by

$$f(n) = n^2,$$

which we also write as $f : n \mapsto n^2$. The equation $g(n) = \sqrt{n}$, where $\sqrt{n}$ is the positive square root, defines a function $g : \mathbb{N} \to \mathbb{R}$, but $h(n) = \pm\sqrt{n}$ does not define a function since it doesn't specify a unique value for $h(n)$. Sometimes we use a convenient oxymoron and refer to h as a multi-valued function.

One way to specify a function is to explicitly list its values, as in Example 1.11. Another way is to give a definite rule, as in Example 1.12. If X is infinite and f is not given by a definite rule, then neither of these methods can be used to specify the function. Nevertheless, we suppose that a general function $f : X \to Y$ may be "defined" by picking for each $x \in X$ a corresponding value $f(x) \in Y$.

If $f : X \to Y$ and $U \subset X$, then we denote the restriction of f to U by $f|_U : U \to Y$, where $f|_U (x) = f(x)$ for $x \in U$.

In defining a function $f : X \to Y$, it is crucial to specify the domain X of elements on which it is defined. There is more ambiguity about the choice of codomain, however, since we can extend the codomain to any set $Z \supset Y$ and define a function $g : X \to Z$ by $g(x) = f(x)$. Strictly speaking, even though f and g have exactly the same values, they are different functions since they have different codomains. Usually, however, we will ignore this distinction and regard f and g as being the same function.

The graph of a function $f : X \to Y$ is the subset G_f of $X \times Y$ defined by

$$G_f = \{(x, y) \in X \times Y : x \in X \text{ and } y = f(x)\}.$$

For example, if $f : \mathbb{R} \to \mathbb{R}$, then the graph of f is the usual set of points (x, y) with $y = f(x)$ in the Cartesian plane $\mathbb{R}^2$. Since a function is defined at every point in its domain, there is some point $(x, y) \in G_f$ for every $x \in X$, and since the value of a function is uniquely defined, there is exactly one such point. In other words, for each $x \in X$ the "vertical line" $L_x = \{(x, y) \in X \times Y : y \in Y\}$ through x intersects the graph of a function $f : X \to Y$ in exactly one point: $L_x \cap G_f = (x, f(x))$.

Definition 1.13. The range, or image, of a function $f : X \to Y$ is the set of values

$$\operatorname{ran} f = \{y \in Y : y = f(x) \text{ for some } x \in X\}.$$

A function is onto if its range is all of Y; that is, if

$$\text{for every } y \in Y \text{ there exists } x \in X \text{ such that } y = f(x).$$

A function is one-to-one if it maps distinct elements of X to distinct elements of Y; that is, if

$$x_1, x_2 \in X \text{ and } x_1 \neq x_2 \text{ implies that } f(x_1) \neq f(x_2).$$

An onto function is also called a surjection, a one-to-one function an injection, and a one-to-one, onto function a bijection.

Example 1.14. The function $f : A \to B$ defined in Example 1.11 is one-to-one but not onto, since $5 \notin \operatorname{ran} f$, while the function $g : B \to A$ is onto but not one-to-one, since $g(5) = g(7)$.

1.3. Composition and inverses of functions

The successive application of mappings leads to the notion of the composition of functions.

Definition 1.15. The composition of functions $f : X \to Y$ and $g : Y \to Z$ is the function $g \circ f : X \to Z$ defined by

$$(g \circ f)(x) = g\left(f(x)\right).$$

The order of application of the functions in a composition is crucial and is read from from right to left. The composition $g \circ f$ can only be defined if the domain of g includes the range of f, and the existence of $g \circ f$ does not imply that $f \circ g$ even makes sense.

Example 1.16. Let X be the set of students in a class and $f : X \to \mathbb{N}$ the function that maps a student to her age. Let $g : \mathbb{N} \to \mathbb{N}$ be the function that adds up the digits in a number e.g., $g(1729) = 19$. If $x \in X$ is 23 years old, then $(g \circ f)(x) = 5$, but $(f \circ g)(x)$ makes no sense, since students in the class are not natural numbers.

Even if both $g \circ f$ and $f \circ g$ are defined, they are, in general, different functions.

Example 1.17. If $f : A \to B$ and $g : B \to A$ are the functions in Example 1.11, then $g \circ f : A \to A$ is given by

$$(g \circ f)(2) = 2, \quad (g \circ f)(3) = 3, \quad (g \circ f)(5) = 11,$$
$$(g \circ f)(7) = 7, \quad (g \circ f)(11) = 5.$$

and $f \circ g : B \to B$ is given by

$$(f \circ g)(1) = 1, \quad (f \circ g)(3) = 3, \quad (f \circ g)(5) = 7,$$
$$(f \circ g)(7) = 7, \quad (f \circ g)(9) = 11, \quad (f \circ g)(11) = 9.$$

A one-to-one, onto function $f : X \to Y$ has an inverse $f^{-1} : Y \to X$ defined by

$$f^{-1}(y) = x \text{ if and only if } f(x) = y.$$

Equivalently, $f^{-1} \circ f = \operatorname{id}_X$ and $f \circ f^{-1} = \operatorname{id}_Y$. A value $f^{-1}(y)$ is defined for every $y \in Y$ since f is onto, and it is unique since f is one-to-one. If $f : X \to Y$ is one-to-one but not onto, then one can still define an inverse function $f^{-1} : \operatorname{ran} f \to X$ whose domain in the range of f.

The use of the notation f^{-1} to denote the inverse function should not be confused with its use to denote the reciprocal function; it should be clear from the context which meaning is intended.

Example 1.18. If $f : \mathbb{R} \to \mathbb{R}$ is the function $f(x) = x^3$, which is one-to-one and onto, then the inverse function $f^{-1} : \mathbb{R} \to \mathbb{R}$ is given by

$$f^{-1}(x) = x^{1/3}.$$

On the other hand, the reciprocal function $g = 1/f$ is given by

$$g(x) = \frac{1}{x^3}, \qquad g : \mathbb{R} \setminus \{0\} \to \mathbb{R}.$$

The reciprocal function is not defined at $x = 0$ where $f(x) = 0$.

If $f : X \to Y$ and $A \subset X$, then we let

$$f(A) = \{y \in Y : y = f(x) \text{ for some } x \in A\}$$

denote the set of values of f on points in A. Similarly, if $B \subset Y$, we let

$$f^{-1}(B) = \{x \in X : f(x) \in B\}$$

denote the set of points in X whose values belong to B. Note that $f^{-1}(B)$ makes sense as a set even if the inverse function $f^{-1} : Y \to X$ does not exist.

Example 1.19. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$. If $A = (-2, 2)$, then $f(A) = [0, 4)$. If $B = (0, 4)$, then

$$f^{-1}(B) = (-2, 0) \cup (0, 2).$$

If $C = (-4, 0)$, then $f^{-1}(C) = \varnothing$.

Finally, we introduce operations on a set.

Definition 1.20. A binary operation on a set X is a function $f : X \times X \to X$.

We think of f as "combining" two elements of X to give another element of X. One can also consider higher-order operations, such as ternary operations $f : X \times X \times X \to X$, but will will only use binary operations.

Example 1.21. Addition $a : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ and multiplication $m : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ are binary operations on $\mathbb{N}$ where

$$a(x, y) = x + y, \qquad m(x, y) = xy.$$

1.4. Indexed sets

We say that a set X is indexed by a set I, or X is an indexed set, if there is an onto function $f : I \to X$. We then write

$$X = \{x_i : i \in I\}$$

where $x_i = f(i)$. For example,

$$\{1, 4, 9, 16, \dots\} = \{n^2 : n \in \mathbb{N}\}.$$

The set X itself is the range of the indexing function f, and it doesn't depend on how we index it. If f isn't one-to-one, then some elements are repeated, but this doesn't affect the definition of the set X. For example,

$$\{-1, 1\} = \{(-1)^n : n \in \mathbb{N}\} = \{(-1)^{n+1} : n \in \mathbb{N}\}.$$

If $\mathcal{C} = \{X_i : i \in I\}$ is an indexed collection of sets X_i, then we denote the union and intersection of the sets in $\mathcal{C}$ by

$$\bigcup_{i \in I} X_i = \{x : x \in X_i \text{ for some } i \in I\}, \qquad \bigcap_{i \in I} X_i = \{x : x \in X_i \text{ for every } i \in I\},$$

or similar notation.

Example 1.22. For $n \in \mathbb{N}$, define the intervals

$$A_n = [1/n, 1 - 1/n] = \{x \in \mathbb{R} : 1/n \le x \le 1 - 1/n\},$$
$$B_n = (-1/n, 1/n) = \{x \in \mathbb{R} : -1/n < x < 1/n\}).$$

Then

$$\bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n=1}^{\infty} A_n = (0, 1), \qquad \bigcap_{n \in \mathbb{N}} B_n = \bigcap_{n=1}^{\infty} B_n = \{0\}.$$

The general statement of De Morgan's laws for a collection of sets is as follows.

Proposition 1.23 (De Morgan). If $\{X_i \subset X : i \in I\}$ is a collection of subsets of a set X, then

$$\left(\bigcup_{i \in I} X_i \right)^c = \bigcap_{i \in I} X_i^c, \qquad \left(\bigcap_{i \in I} X_i \right)^c = \bigcup_{i \in I} X_i^c.$$

Proof. We have $x \notin \bigcup_{i \in I} X_i$ if and only if $x \notin X_i$ for every $i \in I$, which holds if and only if $x \in \bigcap_{i \in I} X_i^c$. Similarly, $x \notin \bigcap_{i \in I} X_i$ if and only if $x \notin X_i$ for some $i \in I$, which holds if and only if $x \in \bigcup_{i \in I} X_i^c$. $\qquad\square$

The following theorem summarizes how unions and intersections map under functions.

Theorem 1.24. Let $f : X \to Y$ be a function. If $\{Y_j \subset Y : j \in J\}$ is a collection of subsets of Y, then

$$f^{-1}\left(\bigcup_{j \in J} Y_j \right) = \bigcup_{j \in J} f^{-1}(Y_j), \qquad f^{-1}\left(\bigcap_{j \in J} Y_j \right) = \bigcap_{j \in J} f^{-1}(Y_j);$$

and if $\{X_i \subset X : i \in I\}$ is a collection of subsets of X, then

$$f\left(\bigcup_{i \in I} X_i \right) = \bigcup_{i \in I} f(X_i), \qquad f\left(\bigcap_{i \in I} X_i \right) \subset \bigcap_{i \in I} f(X_i).$$

Proof. We prove only the results for the inverse image of a union and the image of an intersection; the proof of the remaining two results is similar.

If $x \in f^{-1}\left(\bigcup_{j \in J} Y_j \right)$, then there exists $y \in \bigcup_{j \in J} Y_j$ such that $f(x) = y$. Then $y \in Y_j$ for some $j \in J$ and $x \in f^{-1}(Y_j)$, so $x \in \bigcup_{j \in J} f^{-1}(Y_j)$. It follows that

$$f^{-1}\left(\bigcup_{j \in J} Y_j \right) \subset \bigcup_{j \in J} f^{-1}(Y_j).$$

Conversely, if $x \in \bigcup_{j \in J} f^{-1}(Y_j)$, then $x \in f^{-1}(Y_j)$ for some $j \in J$, so $f(x) \in Y_j$ and $f(x) \in \bigcup_{j \in J} Y_j$, meaning that $x \in f^{-1}\left(\bigcup_{j \in J} Y_j\right)$. It follows that

$$\bigcup_{j \in J} f^{-1}(Y_j) \subset f^{-1}\left(\bigcup_{j \in J} Y_j\right),$$

which proves that the sets are equal.

If $y \in f\left(\bigcap_{i \in I} X_i\right)$, then there exists $x \in \bigcap_{i \in I} X_i$ such that $f(x) = y$. Then $x \in X_i$ and $y \in f(X_i)$ for every $i \in I$, meaning that $y \in \bigcap_{i \in I} f(X_i)$. It follows that

$$f\left(\bigcap_{i \in I} X_i\right) \subset \bigcap_{i \in I} f(X_i).$$

$\square$

The only case in which we don't always have equality is for the image of an intersection, and we may get strict inclusion here if f is not one-to-one.

Example 1.25. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$. Let $A = (-1, 0)$ and $B = (0, 1)$. Then $A \cap B = \varnothing$ and $f(A \cap B) = \varnothing$, but $f(A) = f(B) = (0, 1)$, so $f(A) \cap f(B) = (0, 1) \neq f(A \cap B)$.

Next, we generalize the Cartesian product of finitely many sets to the product of possibly infinitely many sets.

Definition 1.26. Let $\mathcal{C} = \{X_i : i \in I\}$ be an indexed collection of sets X_i. The Cartesian product of $\mathcal{C}$ is the set of functions that assign to each index $i \in I$ an element $x_i \in X_i$. That is,

$$\prod_{i \in I} X_i = \left\{ f : I \to \bigcup_{i \in I} X_i : f(i) \in X_i \text{ for every } i \in I \right\}.$$

For example, if $I = \{1, 2, \ldots, n\}$, then f defines an ordered n-tuple of elements $(x_1, x_2, \ldots, x_n)$ with $x_i = f(i) \in X_i$, so this definition is equivalent to our previous one.

If $X_i = X$ for every $i \in I$, then $\prod_{i \in I} X_i$ is simply the set of functions from I to X, and we also write it as

$$X^I = \{f : I \to X\}.$$

We can think of this set as the set of ordered I-tuples of elements of X.

Example 1.27. A sequence of real numbers $(x_1, x_2, x_3, \ldots, x_n, \ldots) \in \mathbb{R}^{\mathbb{N}}$ is a function $f : \mathbb{N} \to \mathbb{R}$. We study sequences and their convergence properties in Chapter 3.

Example 1.28. Let $\mathbf{2} = \{0, 1\}$ be a set with two elements. Then a subset $A \subset I$ can be identified with its characteristic function $\chi_A : I \to \mathbf{2}$ by: $i \in A$ if and only if $\chi_A(i) = 1$. Thus, $A \mapsto \chi_A$ is a one-to-one map from $\mathcal{P}(I)$ onto $\mathbf{2}^I$.

Before giving another example, we introduce some convenient notation.

Definition 1.29. Let

$$\Sigma = \{(s_1, s_2, s_3, \ldots, s_k, \ldots) : s_k = 0, 1\}$$

denote the set of all binary sequences; that is, sequences whose terms are either 0 or 1.

Example 1.30. Let $\mathbf{2} = \{0, 1\}$. Then $\Sigma = \mathbf{2}^{\mathbb{N}}$, where we identify a sequence $(s_1, s_2, \ldots s_k, \ldots)$ with the function $f : \mathbb{N} \to \mathbf{2}$ such that $s_k = f(k)$. We can also identify Σ and $\mathbf{2}^{\mathbb{N}}$ with $\mathcal{P}(\mathbb{N})$ as in Example 1.28. For example, the sequence $(1, 0, 1, 0, 1, \ldots)$ of alternating ones and zeros corresponds to the function $f : \mathbb{N} \to \mathbf{2}$ defined by

$$f(k) = \begin{cases} 1 & \text{if } k \text{ is odd,} \\ 0 & \text{if } k \text{ is even,} \end{cases}$$

and to the set $\{1, 3, 5, 7, \ldots\} \subset \mathbb{N}$ of odd natural numbers.

1.5. Relations

A binary relation R on sets X and Y is a definite relation between elements of X and elements of Y. We write xRy if $x \in X$ and $y \in Y$ are related. One can also define relations on more than two sets, but we shall consider only binary relations and refer to them simply as relations. If $X = Y$, then we call R a relation on X.

Example 1.31. Suppose that S is a set of students enrolled in a university and B is a set of books in a library. We might define a relation R on S and B by:

$$s \in S \text{ has read } b \in B.$$

In that case, sRb if and only if s has read b. Another, probably inequivalent, relation is:

$$s \in S \text{ has checked } b \in B \text{ out of the library.}$$

When used informally, relations may be ambiguous (did s read b if she only read the first page?), but in mathematical usage we always require that relations are definite, meaning that one and only one of the statements "these elements are related" or "these elements are not related" is true.

The graph G_R of a relation R on X and Y is the subset of $X \times Y$ defined by

$$G_R = \{(x, y) \in X \times Y : xRy\} \,.$$

This graph contains all of the information about which elements are related. Conversely, any subset $G \subset X \times Y$ defines a relation R by: xRy if and only if $(x, y) \in G$. Thus, a relation on X and Y may be (and often is) defined as subset of $X \times Y$. As for sets, it doesn't matter how a relation is defined, only what elements are related.

A function $f : X \to Y$ determines a relation F on X and Y by: xFy if and only if $y = f(x)$. Thus, functions are a special case of relations. The graph G_R of a general relation differs from the graph G_F of a function in two ways: there may be elements $x \in X$ such that $(x, y) \notin G_R$ for any $y \in Y$, and there may be $x \in X$ such that $(x, y) \in G_R$ for many $y \in Y$.

For example, in the case of the relation R in Example 1.31, there may be some students who haven't read any books, and there may be other students who have

read lots of books, in which case we don't have a well-defined function from students to books.

Two important types of relations are orders and equivalence relations, and we define them next.

1.5.1. Orders. A primary example of an order is the standard order $\leq$ on the natural (or real) numbers. This order is a linear or total order, meaning that two numbers are always comparable. Another example of an order is inclusion $\subset$ on the power set of some set; one set is "smaller" than another set if it is included in it. This order is a partial order (provided the original set has at least two elements), meaning that two subsets need not be comparable.

Example 1.32. Let $X = \{1, 2\}$. The collection of subsets of X is

$$\mathcal{P}(X) = \{\varnothing, A, B, X\}, \qquad A = \{1\}, \quad B = \{2\}.$$

We have $\varnothing \subset A \subset X$ and $\varnothing \subset B \subset X$, but $A \not\subset B$ and $B \not\subset A$, so A and B are not comparable under ordering by inclusion.

The general definition of an order is as follows.

Definition 1.33. An order $\preceq$ on a set X is a binary relation on X such that for every $x, y, z \in X$:

  (a) $x \preceq x$ (reflexivity);

  (b) if $x \preceq y$ and $y \preceq x$ then $x = y$ (antisymmetry);

  (c) if $x \preceq y$ and $y \preceq z$ then $x \preceq z$ (transitivity).

An order is a linear, or total, order if for every $x, y \in X$ either $x \preceq y$ or $y \preceq x$, otherwise it is a partial order.

If $\preceq$ is an order, then we also write $y \succeq x$ instead of $x \preceq y$, and we define a corresponding strict order $\prec$ by

$$x \prec y \text{ if } x \preceq y \text{ and } x \neq y.$$

There are many ways to order a given set (with two or more elements).

Example 1.34. Let X be a set. One way to partially order the subsets of X is by inclusion, as in Example 1.32. Another way is to say that $A \preceq B$ for $A, B \subset X$ if and only if $A \supset B$, meaning that A is "smaller" than B if A includes B. Then $\preceq$ in an order on $\mathcal{P}(X)$, called ordering by reverse inclusion.

1.5.2. Equivalence relations. Equivalence relations decompose a set into disjoint subsets, called equivalence classes. We begin with an example of an equivalence relation on $\mathbb{N}$.

Example 1.35. Fix $N \in \mathbb{N}$ and say that $m \sim n$ if

$$m \equiv n \pmod{N},$$

meaning that $m - n$ is divisible by N. Two numbers are related by $\sim$ if they have the same remainder when divided by N. Moreover, $\mathbb{N}$ is the union of N equivalence classes, consisting of numbers with remainders $0, 1, \ldots N - 1$ modulo N.

The definition of an equivalence relation differs from the definition of an order only by changing antisymmetry to symmetry, but order relations and equivalence relations have completely different properties.

Definition 1.36. An equivalence relation $\sim$ on a set X is a binary relation on X such that for every $x, y, z \in X$:

(a) $x \sim x$ (reflexivity);

(b) if $x \sim y$ then $y \sim x$ (symmetry);

(c) if $x \sim y$ and $y \sim z$ then $x \sim z$ (transitivity).

For each $x \in X$, the set of elements equivalent to x,

$$[x/\sim] = \{y \in X : x \sim y\},$$

is called the equivalence class of x with respect to $\sim$. When the equivalence relation is understood, we write the equivalence class $[x/\sim]$ simply as $[x]$. The set of equivalence classes of an equivalence relation $\sim$ on a set X is denoted by $X/\sim$. Note that each element of $X/\sim$ is a subset of X, so $X/\sim$ is a subset of the power set $\mathcal{P}(X)$ of X.

The following theorem is the basic result about equivalence relations. It says that an equivalence relation on a set partitions the set into disjoint equivalence classes.

Theorem 1.37. Let $\sim$ be an equivalence relation on a set X. Every equivalence class is non-empty, and X is the disjoint union of the equivalence classes of $\sim$.

Proof. If $x \in X$, then the symmetry of $\sim$ implies that $x \in [x]$. Therefore every equivalence class is non-empty and the union of the equivalence classes is X.

To prove that the union is disjoint, we show that for every $x, y \in X$ either $[x] \cap [y] = \varnothing$ (if $x \nsim y$) or $[x] = [y]$ (if $x \sim y$).

Suppose that $[x] \cap [y] \neq \varnothing$. Let $z \in [x] \cap [y]$ be an element in both equivalence classes. If $x_1 \in [x]$, then $x_1 \sim z$ and $z \sim y$, so $x_1 \sim y$ by the transitivity of $\sim$, and therefore $x_1 \in [y]$. It follows that $[x] \subset [y]$. A similar argument applied to $y_1 \in [y]$ implies that $[y] \subset [x]$, and therefore $[x] = [y]$. In particular, $y \in [x]$, so $x \sim y$. On the other hand, if $[x] \cap [y] = \varnothing$, then $y \notin [x]$ since $y \in [y]$, so $x \nsim y$. $\qquad\square$

There is a natural projection $\pi : X \to X/\sim$, given by $\pi(x) = [x]$, that maps each element of X to the equivalence class that contains it. Conversely, we can index the collection of equivalence classes

$$X/\sim = \{[a] : a \in A\}$$

by a subset A of X which contains exactly one element from each equivalence class. It is important to recognize, however, that such an indexing involves an arbitrary choice of a representative element from each equivalence class, and it is better to think in terms of the collection of equivalence classes, rather than a subset of elements.

Example 1.38. The equivalence classes of $\mathbb{N}$ relative to the equivalence relation $m \sim n$ if $m \equiv n \pmod 3$ are given by

$$I_0 = \{3, 6, 9, \dots\}, \quad I_1 = \{1, 4, 7, \dots\}, \quad I_2 = \{2, 5, 8, \dots\}.$$

The projection $\pi : \mathbb{N} \to \{I_0, I_1, I_2\}$ maps a number to its equivalence class e.g. $\pi(101) = I_2$. We can choose $\{1, 2, 3\}$ as a set of representative elements, in which case

$$I_0 = [3], \qquad I_1 = [1], \qquad I_2 = [2],$$

but any other set $A \subset \mathbb{N}$ of three numbers with remainders 0, 1, 2 (mod 3) will do. For example, if we choose $A = \{7, 15, 101\}$, then

$$I_0 = [15], \qquad I_1 = [7], \qquad I_2 = [101].$$

1.6. Countable and uncountable sets

One way to show that two sets have the same "size" is to pair off their elements. For example, if we can match up every left shoe in a closet with a right shoe, with no right shoes left over, then we know that we have the same number of left and right shoes. That is, we have the same number of left and right shoes if there is a one-to-one, onto map $f : L \to R$, or one-to-one correspondence, from the set L of left shoes to the set R of right shoes.

We refer to the "size" of a set as measured by one-to-one correspondences as its cardinality. This notion enables us to compare the cardinality of both finite and infinite sets. In particular, we can use it to distinguish between "smaller" countably infinite sets, such as the integers or rational numbers, and "larger" uncountably infinite sets, such as the real numbers.

Definition 1.39. Two sets X, Y have equal cardinality, written $X \approx Y$, if there is a one-to-one, onto map $f : X \to Y$. The cardinality of X is less than or equal to the cardinality of Y, written $X \lesssim Y$, if there is a one-to-one (but not necessarily onto) map $g : X \to Y$.

If $X \approx Y$, then we also say that X, Y have the same cardinality. We don't define the notion of a "cardinal number" here, only the relation between sets of "equal cardinality."

Note that $\approx$ is an equivalence relation on any collection of sets. In particular, it is transitive because if $X \approx Y$ and $Y \approx Z$, then there are one-to-one and onto maps $f : X \to Y$ and $g : Y \to Z$, so $g \circ f : X \to Z$ is one-to-one and onto, and $X \approx Z$. We may therefore divide any collection of sets into equivalence classes of sets with equal cardinality.

It follows immediately from the definition that $\lesssim$ is reflexive and transitive. Furthermore, as stated in the following Schröder-Bernstein theorem, if $X \lesssim Y$ and $Y \lesssim X$, then $X \approx Y$. This result allows us to prove that two sets have equal cardinality by constructing one-to-one maps that need not be onto. The statement of the theorem is intuitively obvious but the proof, while elementary, is surprisingly involved and can be omitted without loss of continuity. (We will only use the theorem once, in the proof of Theorem 5.67.)

Theorem 1.40 (* Schröder-Bernstein). If X, Y are sets such that there are one-to-one maps $f : X \to Y$ and $g : Y \to X$, then there is a one-to-one, onto map $h : X \to Y$.

Proof. We divide X into three disjoint subsets X_X, X_Y, X_∞ with different mapping properties as follows.

Consider a point $x_1 \in X$. If x_1 is not in the range of g, then we say $x_1 \in X_X$. Otherwise there exists $y_1 \in Y$ such that $g(y_1) = x_1$, and y_1 is unique since g is one-to-one. If y_1 is not in the range of f, then we say $x_1 \in X_Y$. Otherwise there exists a unique $x_2 \in X$ such that $f(x_2) = y_1$. Continuing in this way, we generate a sequence of points

$$x_1, y_1, x_2, y_2, \ldots, x_n, y_n, x_{n+1}, \ldots$$

with $x_n \in X$, $y_n \in Y$ and

$$g(y_n) = x_n, \qquad f(x_{n+1}) = y_n.$$

We assign the starting point x_1 to a subset in the following way: (a) $x_1 \in X_X$ if the sequence terminates at some $x_n \in X$ that isn't in the range of g; (b) $x_1 \in X_Y$ if the sequence terminates at some $y_n \in Y$ that isn't in the range of f; (c) $x_1 \in X_\infty$ if the sequence never terminates.

Similarly, if $y_1 \in Y$, then we generate a sequence of points

$$y_1, x_1, y_2, x_2, \ldots, y_n, x_n, y_{n+1}, \ldots$$

with $x_n \in X$, $y_n \in Y$ by

$$f(x_n) = y_n, \qquad g(y_{n+1}) = x_n,$$

and we assign y_1 to a subset Y_X, Y_Y, or Y_∞ of Y as follows: (a) $y_1 \in Y_X$ if the sequence terminates at some $x_n \in X$ that isn't in the range of g; (b) $y_1 \in Y_Y$ if the sequence terminates at some $y_n \in Y$ that isn't in the range of f; (c) $y_1 \in Y_\infty$ if the sequence never terminates.

We claim that $f : X_X \to Y_X$ is one-to-one and onto. First, if $x \in X_X$, then $f(x) \in Y_X$ because the the sequence generated by $f(x)$ coincides with the sequence generated by x after its first term, so both sequences terminate at a point in X. Second, if $y \in Y_X$, then there is $x \in X$ such that $f(x) = y$, otherwise the sequence would terminate at $y \in Y$, meaning that $y \in Y_Y$. Furthermore, we must have $x \in X_X$ because the sequence generated by x is a continuation of the sequence generated by y and therefore also terminates at a point in X. Finally, f is one-to-one on X_X since f is one-to-one on X.

The same argument applied to $g : Y_Y \to X_Y$ implies that g is one-to-one and onto, so $g^{-1} : X_Y \to Y_Y$ is one-to-one and onto.

Finally, similar arguments show that $f : X_\infty \to Y_\infty$ is one-to-one and onto: If $x \in X_\infty$, then the sequence generated by $f(x) \in Y$ doesn't terminate, so $f(x) \in Y_\infty$; and every $y \in Y_\infty$ is the image of a point $x \in X$ which, like y, generates a sequence that does not terminate, so $x \in X_\infty$.

It then follows that $h : X \to Y$ defined by

$$h(x) = \begin{cases} f(x) & \text{if } x \in X_X \\ g^{-1}(x) & \text{if } x \in X_Y \\ f(x) & \text{if } x \in X_\infty \end{cases}$$

is a one-to-one, onto map from X to Y. $\qquad \square$

We can use the cardinality relation to describe the "size" of a set by comparing it with standard sets.

Definition 1.41. A set X is:

(1) Finite if it is the empty set or $X \approx \{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$;

(2) Countably infinite (or denumerable) if $X \approx \mathbb{N}$;

(3) Infinite if it is not finite;

(4) Countable if it is finite or countably infinite;

(5) Uncountable if it is not countable.

We'll take for granted some intuitively obvious facts which follow from the definitions. For example, a finite, non-empty set is in one-to-one correspondence with $\{1, 2, \ldots, n\}$ for a unique natural number $n \in \mathbb{N}$ (the number of elements in the set), a countably infinite set is not finite, and a subset of a countable set is countable.

According to Definition 1.41, we may divide sets into disjoint classes of finite, countably infinite, and uncountable sets. We also distinguish between finite and infinite sets, and countable and uncountable sets. We will show below, in Theorem 2.19, that the set of real numbers is uncountable, and we refer to its cardinality as the cardinality of the continuum.

Definition 1.42. A set X has the cardinality of the continuum if $X \approx \mathbb{R}$.

One has to be careful in extrapolating properties of finite sets to infinite sets.

Example 1.43. The set of squares

$$S = \{1, 4, 9, 16, \ldots, n^2, \ldots\}$$

is countably infinite since $f : \mathbb{N} \to S$ defined by $f(n) = n^2$ is one-to-one and onto. It may appear surprising at first that the set $\mathbb{N}$ can be in one-to-one correspondence with an apparently "smaller" proper subset S, since this doesn't happen for finite sets. In fact, assuming the axiom of choice, one can show that a set is infinite if and only if it has the same cardinality as a proper subset. Dedekind (1888) used this property to give a definition infinite sets that did not depend on the natural numbers $\mathbb{N}$.

Next, we prove some results about countable sets. The following proposition states a useful necessary and sufficient condition for a set to be countable.

Proposition 1.44. A non-empty set X is countable if and only if there is an onto map $f : \mathbb{N} \to X$.

Proof. If X is countably infinite, then there is a one-to-one, onto map $f : \mathbb{N} \to X$. If X is finite and non-empty, then for some $n \in \mathbb{N}$ there is a one-to-one, onto map $g : \{1, 2, \ldots, n\} \to X$. Choose any $x \in X$ and define the onto map $f : \mathbb{N} \to X$ by

$$f(k) = \begin{cases} g(k) & \text{if } k = 1, 2, \ldots, n, \\ x & \text{if } k = n+1, n+2, \ldots. \end{cases}$$

Conversely, suppose that such an onto map exists. We define a one-to-one, onto map g recursively by omitting repeated values of f. Explicitly, let $g(1) = f(1)$. Suppose that $n \geq 1$ and we have chosen n distinct g-values $g(1), g(2), \ldots, g(n)$. Let

$$A_n = \{k \in \mathbb{N} : f(k) \neq g(j) \text{ for every } j = 1, 2, \ldots, n\}$$

denote the set of natural numbers whose f-values are not already included among the g-values. If $A_n = \varnothing$, then $g : \{1, 2, \ldots, n\} \to X$ is one-to-one and onto, and X is finite. Otherwise, let $k_n = \min A_n$, and define $g(n+1) = f(k_n)$, which is distinct from all of the previous g-values. Either this process terminates, and X is finite, or we go through all the f-values and obtain a one-to-one, onto map $g : \mathbb{N} \to X$, and X is countably infinite. $\qquad \square$

If X is a countable set, then we refer to an onto function $f : \mathbb{N} \to X$ as an *enumeration* of X, and write $X = \{x_n : n \in \mathbb{N}\}$, where $x_n = f(n)$.

Proposition 1.45. *The Cartesian product $\mathbb{N} \times \mathbb{N}$ is countably infinite.*

Proof. Define a linear order $\prec$ on ordered pairs of natural numbers as follows:

$$(m, n) \prec (m', n') \quad \text{if either } m + n < m' + n' \text{ or } m + n = m' + n' \text{ and } n < n'.$$

That is, we arrange $\mathbb{N} \times \mathbb{N}$ in a table

$$
\begin{array}{ccccc}
(1,1) & (1,2) & (1,3) & (1,4) & \ldots \\
(2,1) & (2,2) & (2,3) & (2,4) & \ldots \\
(3,1) & (3,2) & (3,3) & (3,4) & \ldots \\
(4,1) & (4,2) & (4,3) & (4,4) & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{array}
$$

and list it along successive diagonals from bottom-left to top-right as

$$(1,1), \ (2,1), \ (1,2), \ (3,1), \ (2,2), \ (1,3), \ (4,1), \ (3,2), \ (2,3), \ (1,4), \ \ldots.$$

We define $f : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ by setting $f(n)$ equal to the nth pair in this order; for example, $f(7) = (4,1)$. Then f is one-to-one and onto, so $\mathbb{N} \times \mathbb{N}$ is countably infinite. $\qquad \square$

Theorem 1.46. *A countable union of countable sets is countable.*

Proof. Let $\{X_n : n \in \mathbb{N}\}$ be a countable collection of countable sets. From Proposition 1.44, there is an onto map $f_n : \mathbb{N} \to X_n$. We define

$$g : \mathbb{N} \times \mathbb{N} \to \bigcup_{n \in \mathbb{N}} X_n$$

by $g(n, k) = f_n(k)$. Then g is also onto. From Proposition 1.45, there is a one-to-one, onto map $h : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$, and it follows that

$$g \circ h : \mathbb{N} \to \bigcup_{n \in \mathbb{N}} X_n$$

is onto, so Proposition 1.44 implies that the union of the X_n is countable. $\qquad \square$

The next theorem gives a fundamental example of an uncountable set, namely the set of all subsets of natural numbers. The proof uses a "diagonal" argument due to Cantor (1891), which is of frequent use in analysis. Recall from Definition 1.1 that the power set of a set is the collection of all its subsets.

Theorem 1.47. The power set $\mathcal{P}(\mathbb{N})$ of $\mathbb{N}$ is uncountable.

Proof. Let $\mathcal{C} \subset \mathcal{P}(\mathbb{N})$ be a countable collection of subsets of $\mathbb{N}$

$$\mathcal{C} = \{A_n \subset \mathbb{N} : n \in \mathbb{N}\}.$$

Define a subset $A \subset \mathbb{N}$ by

$$A = \{n \in \mathbb{N} : n \notin A_n\}.$$

Then $A \neq A_n$ for every $n \in \mathbb{N}$ since either $n \in A$ and $n \notin A_n$ or $n \notin A$ and $n \in A_n$. Thus, $A \notin \mathcal{C}$. It follows that no countable collection of subsets of $\mathbb{N}$ includes all of the subsets of $\mathbb{N}$, so $\mathcal{P}(\mathbb{N})$ is uncountable. $\qquad\square$

This theorem has an immediate corollary for the set Σ of binary sequences defined in Definition 1.29.

Corollary 1.48. The set Σ of binary sequences has the same cardinality as $\mathcal{P}(\mathbb{N})$ and is uncountable.

Proof. By Example 1.30, the set Σ is in one-to-one correspondence with $\mathcal{P}(\mathbb{N})$, which is uncountable. $\qquad\square$

It is instructive to write the diagonal argument in terms of binary sequences. Suppose that $S = \{\mathbf{s}_n \in \Sigma : n \in \mathbb{N}\}$ is a countable set of binary sequences that begins, for example, as follows

$$\begin{aligned}
\mathbf{s}_1 &= 0\,0\,1\,1\,0\,1\dots \\
\mathbf{s}_2 &= 1\,1\,0\,0\,1\,0\dots \\
\mathbf{s}_3 &= 1\,1\,0\,1\,1\,0\dots \\
\mathbf{s}_4 &= 0\,1\,1\,0\,0\,0\dots \\
\mathbf{s}_5 &= 1\,0\,0\,1\,1\,1\dots \\
\mathbf{s}_6 &= 1\,0\,0\,1\,0\,0\dots \\
&\;\;\vdots
\end{aligned}$$

Then we get a sequence $\mathbf{s} \notin S$ by going down the diagonal and switching the values from 0 to 1 or from 1 to 0. For the previous sequences, this gives

$$\mathbf{s} = 1\,0\,1\,1\,0\,1\dots.$$

We will show in Theorem 5.67 below that Σ and $\mathcal{P}(\mathbb{N})$ are also in one-to-one correspondence with $\mathbb{R}$, so both have the cardinality of the continuum.

A similar diagonal argument to the one used in Theorem 1.47 shows that for every set X the cardinality of the power set $\mathcal{P}(X)$ is strictly greater than the cardinality of X. In particular, the cardinality of $\mathcal{P}(\mathcal{P}(\mathbb{N}))$ is strictly greater than the cardinality of $\mathcal{P}(\mathbb{N})$, the cardinality of $\mathcal{P}(\mathcal{P}(\mathcal{P}(\mathbb{N})))$ is strictly greater than

the cardinality of $\mathcal{P}(\mathcal{P}(\mathbb{N}))$, and so on. Thus, there are many other uncountable cardinalities apart from the cardinality of the continuum.

Cantor (1878) raised the question of whether or not there are any sets whose cardinality lies strictly between that of $\mathbb{N}$ and $\mathcal{P}(\mathbb{N})$. The statement that there are no such sets is called the continuum hypothesis, which may be formulated as follows.

Hypothesis 1.49 (Continuum). If $\mathcal{C} \subset \mathcal{P}(\mathbb{N})$ is infinite, then either $\mathcal{C} \approx \mathbb{N}$ or $\mathcal{C} \approx \mathcal{P}(\mathbb{N})$.

The work of Gödel (1940) and Cohen (1963) established the remarkable result that the continuum hypothesis cannot be proved or disproved from the standard axioms of set theory (assuming, as we believe to be the case, that these axioms are consistent). This result illustrates a fundamental and unavoidable incompleteness in the ability of any finite system of axioms to capture the properties of any mathematical structure that is rich enough to include the natural numbers.

Numbers

God created the integers and the rest is the work of man. (Leopold Kronecker, in an after-dinner speech at a conference, Berlin, 1886)

"God created the integers and the rest is the work of man." This maxim spoken by the algebraist Kronecker reveals more about his past as a banker who grew rich through monetary speculation than about his philosophical insight. There is hardly any doubt that, from a psychological and, for the writer, ontological point of view, the geometric continuum is the primordial entity. If one has any consciousness at all, it is consciousness of time and space; geometric continuity is in some way inseparably bound to conscious thought. (René Thom, 1986)

In this chapter, we describe the properties of the basic number systems. We briefly discuss the integers and rational numbers, and then consider the real numbers in more detail.

The real numbers form a complete number system which includes the rational numbers as a dense subset. We will summarize the properties of the real numbers in a list of intuitively reasonable axioms, which we assume in everything that follows. These axioms are of three types: (a) algebraic; (b) ordering; (c) completeness. The completeness of the real numbers is what distinguishes them from the rationals numbers and is the essential property for analysis.

The rational numbers may be constructed from the natural numbers as pairs of integers, and there are several ways to construct the real numbers from the rational numbers. For example, Dedekind used cuts of the rationals, while Cantor used equivalence classes of Cauchy sequences of rational numbers. The real numbers that are constructed in either way satisfy the axioms given in this chapter. These constructions show that the real numbers are as well-founded as the natural numbers (at least, if we take set theory for granted), but they don't lead to any new properties of the real numbers, and we won't describe them here.

2.1. Integers

> Why then is this view [the induction principle] imposed upon us with such an irresistible weight of evidence? It is because it is only the affirmation of the power of the mind which knows it can conceive of the indefinite repetition of the same act, when that act is once possible. (Poincaré, 1902)

The set of natural numbers, or positive integers, is

$$\mathbb{N} = \{1, 2, 3, \dots\}.$$

We add and multiply natural numbers in the usual way. (The formal algebraic properties of addition and multiplication on $\mathbb{N}$ follow from the ones stated below for $\mathbb{R}$.)

An essential property of the natural numbers is the following induction principle, which expresses the idea that we can reach every natural number by counting upwards from one.

Axiom 2.1. Suppose that $A \subset \mathbb{N}$ is a set of natural numbers such that: (a) $1 \in A$; (b) $n \in A$ implies $(n + 1) \in A$. Then $A = \mathbb{N}$.

This principle, together with appropriate algebraic properties, is enough to completely characterize the natural numbers. For example, one standard set of axioms is the Peano axioms, first stated by Dedekind [**3**], but we won't describe them in detail here.

As an illustration of how induction can be used, we prove the following result for the sum of the first n squares, written in summation notation as

$$\sum_{k=1}^{n} k^2 = 1^2 + 2^2 + 3^2 + \cdots + n^2.$$

Proposition 2.2. For every $n \in \mathbb{N}$,

$$\sum_{k=1}^{n} k^2 = \frac{1}{6} n(n + 1)(2n + 1).$$

Proof. Let A be the set of $n \in \mathbb{N}$ for which this identity holds. It holds for $n = 1$, so $1 \in A$. Suppose the identity holds for some $n \in \mathbb{N}$. Then

$$
\begin{aligned}
\sum_{k=1}^{n+1} k^2 &= \sum_{k=1}^{n} k^2 + (n + 1)^2 \\
&= \frac{1}{6} n(n + 1)(2n + 1) + (n + 1)^2 \\
&= \frac{1}{6}(n + 1)\left(2n^2 + 7n + 6\right) \\
&= \frac{1}{6}(n + 1)(n + 2)(2n + 3).
\end{aligned}
$$

It follows that the identity holds when n is replaced by $n + 1$. Thus $n \in A$ implies that $(n + 1) \in A$, so $A = \mathbb{N}$, and the proposition follows by induction. $\square$

Note that the right hand side of the identity in Proposition 2.2 is always an integer, as it must be, since one of n, $n + 1$ is divisible by 2 and one of n, $n + 1$, $2n + 1$ is divisible by 3.

Equations for the sum of the first n cubes,

$$\sum_{k=1}^{n} k^3 = \frac{1}{4} n^2 (n + 1)^2,$$

and other powers can be proved by induction in a similar way. Another example of a result that can be proved by induction is the Euler-Binet formula in Proposition 3.9 for the terms in the Fibonacci sequence.

One defect of such a proof by induction is that although it verifies the result, it does not explain where the original hypothesis comes from. A separate argument is often required to come up with a plausible hypothesis. For example, it is reasonable to guess that the sum of the first n squares might be a cubic polynomial in n. The possible values of the coefficients can then be found by evaluating the first few sums, after which the general result may be verified by induction.

The set of integers consists of the natural numbers, their negatives (or additive inverses), and zero (the additive identity):

$$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}.$$

We can add, subtract, and multiply integers in the usual way. In algebraic terminology, $(\mathbb{Z}, +, \cdot)$ is a commutative ring with identity.

Like the natural numbers $\mathbb{N}$, the integers $\mathbb{Z}$ are countably infinite.

Proposition 2.3. The set of integers $\mathbb{Z}$ is countably infinite.

Proof. The function $f : \mathbb{N} \to \mathbb{Z}$ defined by $f(1) = 0$, and

$$f(2n) = n, \quad f(2n + 1) = -n \qquad \text{for } n \geq 1,$$

is one-to-one and onto. $\qquad \square$

The function in the previous proof corresponds to listing the integers as

$$0, \ 1, \ -1, \ 2, \ -2, \ 3, \ -3, \ 4, \ -4, \ 5, \ -5, \ \dots.$$

Alternatively, but less directly, we can prove Proposition 2.3 by writing

$$\mathbb{Z} = -\mathbb{N} \cup \{0\} \cup \mathbb{N}$$

as a countable union of countable sets and applying Theorem 1.46.

2.2. Rational numbers

A rational number is a ratio of integers. We denote the set of rational numbers by

$$\mathbb{Q} = \left\{ \frac{p}{q} : p, q \in \mathbb{Z} \text{ and } q \neq 0 \right\}$$

where we may cancel common factors from the numerator and denominator, meaning that

$$\frac{p_1}{q_1} = \frac{p_2}{q_2} \qquad \text{if and only if } p_1 q_2 = p_2 q_1.$$

We can add, subtract, multiply, and divide (except by 0) rational numbers in the usual way. In algebraic terminology, $(\mathbb{Q}, +, \cdot)$ a field. We state the field axioms explicitly for $\mathbb{R}$ in Axiom 2.6 below.

We can construct $\mathbb{Q}$ from $\mathbb{Z}$ as the collection of equivalence classes in $\mathbb{Z} \times \mathbb{Z} \setminus \{0\}$ with respect to the equivalence relation $(p_1, q_1) \sim (p_2, q_2)$ if $p_1 q_2 = p_2 q_1$. The usual sums and products of rational numbers are well-defined on these equivalence classes.

The rational numbers are linearly ordered by their standard order, and this order is compatible with the algebraic structure of $\mathbb{Q}$. Thus, $(\mathbb{Q}, +, \cdot, <)$ is an ordered field. Moreover, this order is dense, meaning that if $r_1, r_2 \in \mathbb{Q}$ and $r_1 < r_2$, then there exists a rational number $r \in \mathbb{Q}$ between them with $r_1 < r < r_2$. For example, we can take

$$r = \frac{1}{2}\left(r_1 + r_2\right).$$

The fact that the rational numbers are densely ordered might suggest that they contain all the numbers we need. But this is not the case: they have a lot of "gaps," which are filled by the irrational real numbers.

The following theorem shows that $\sqrt{2}$ is irrational. In particular, the length of the hypotenuse of a right-angled triangle with sides of length one is not rational. Thus, the rational numbers are inadequate even for Euclidean geometry; they are yet more inadequate for analysis.

The irrationality of $\sqrt{2}$ was discovered by the Pythagoreans of Ancient Greece in the 5th century BC, perhaps by Hippasus of Metapontum. According to one legend, the Pythagoreans celebrated the discovery by sacrificing one hundred oxen. According to another legend, Hippasus showed the proof to Pythagoras on a boat, while they were having a lesson. Pythagoras believed that, like the musical scales, everything in the universe could be reduced to ratios of integers and threw Hippasus overboard to drown.

Theorem 2.4. There is no rational number $x \in \mathbb{Q}$ such that $x^2 = 2$.

Proof. Suppose for contradiction that $x^2 = 2$ and $x = p/q$ where $p, q \in \mathbb{N}$. By canceling common factors, we can assume p and q are relatively prime (that is, the only integers that divide both p and q are ± 1). Since $x^2 = 2$, we have

$$p^2 = 2q^2,$$

which implies that p^2 is even. Since the square of an odd number is odd, $p = 2r$ must be even. Therefore

$$2r^2 = q^2,$$

which implies that $q = 2s$ is even. Hence p and q have a common factor of 2, which contradicts the initial assumption. $\qquad\qquad\square$

Theorem 2.4 may raise the question of whether there is a real number $x \in \mathbb{R}$ such that $x^2 = 2$. As we will see in Example 7.47 below, there is.

A similar proof, using the prime factorization of integers, shows that $\sqrt{n}$ is irrational for every $n \in \mathbb{N}$ that isn't a perfect square. Two other examples of irrational numbers are π and e. We will prove in Theorem 4.52 that e is irrational, but a proof of the irrationality of π is harder.

The following result may appear somewhat surprising at first, but it is another indication that there are not "enough" rationals.

Theorem 2.5. The rational numbers are countably infinite.

Proof. The rational numbers are not finite since, for example, they contain the countably infinite set of integers as a subset, so we just have to show that $\mathbb{Q}$ is countable.

Let $\mathbb{Q}^+ = \{x \in \mathbb{Q} : x > 0\}$ denote the set of positive rational numbers, and define the onto (but not one-to-one) map

$$g : \mathbb{N} \times \mathbb{N} \to \mathbb{Q}^+, \qquad g(p, q) = \frac{p}{q}.$$

Let $h : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ be a one-to-one, onto map, as obtained in Proposition 1.45, and define $f : \mathbb{N} \to \mathbb{Q}^+$ by $f = g \circ h$. Then $f : \mathbb{N} \to \mathbb{Q}^+$ is onto, and Proposition 1.44 implies that $\mathbb{Q}^+$ is countable. It follows that $\mathbb{Q} = \mathbb{Q}^- \cup \{0\} \cup \mathbb{Q}^+$, where $\mathbb{Q}^- \approx \mathbb{Q}^+$ denotes the set of negative rational numbers, is countable. □

Alternatively, we can write

$$\mathbb{Q} = \bigcup_{q \in \mathbb{N}} \{p/q : p \in \mathbb{Z}\}$$

as a countable union of countable sets, and use Theorem 1.46. As we prove in Theorem 2.19, the real numbers are uncountable, so there are many "more" irrational numbers than rational numbers.

2.3. Real numbers: algebraic properties

The algebraic properties of $\mathbb{R}$ are summarized in the following axioms, which state that $(\mathbb{R}, +, \cdot)$ is a field.

Axiom 2.6. There exist binary operations

$$a, m : \mathbb{R} \times \mathbb{R} \to \mathbb{R},$$

written $a(x, y) = x + y$ and $m(x, y) = x \cdot y = xy$, and elements $0, 1 \in \mathbb{R}$ such that for all $x, y, z \in \mathbb{R}$:

(a) $x + 0 = x$ (existence of an additive identity 0);

(b) for every $x \in \mathbb{R}$ there exists $y \in \mathbb{R}$ such that $x + y = 0$ (existence of an additive inverse $y = -x$);

(c) $x + (y + z) = (x + y) + z$ (addition is associative);

(d) $x + y = y + x$ (addition is commutative);

(e) $x1 = x$ (existence of a multiplicative identity 1);

(f) for every $x \in \mathbb{R} \setminus \{0\}$, there exists $y \in \mathbb{R}$ such that $xy = 1$ (existence of a multiplicative inverse $y = x^{-1}$);

(g) $x(yz) = (xy)z$ (multiplication is associative);

(h) $xy = yx$ (multiplication is commutative);

(i) $(x + y)z = xz + yz$ (multiplication is distributive over addition).

Axioms (a)–(d) say that $\mathbb{R}$ is a commutative group with respect to addition; axioms (e)–(h) say that $\mathbb{R} \setminus \{0\}$ is a commutative group with respect to multiplication; and axiom (i) says that addition and multiplication are compatible, in the sense that they satisfy a distributive law.

All of the usual algebraic properties of addition, subtraction (subtracting x means adding $-x$), multiplication, and division (dividing by x means multiplying by x^{-1}) follow from these axioms, although we will not derive them in detail. The natural number $n \in \mathbb{N}$ is obtained by adding one to itself n times, the integer $-n$ is its additive inverse, and $p/q = pq^{-1}$, where p, q are integers with $q \neq 0$ is a rational number. Thus, $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$.

2.4. Real numbers: ordering properties

The real numbers have a natural order relation that is compatible with their algebraic structure. We visualize the ordered real numbers as the real line, with smaller numbers to the left and larger numbers to the right.

Axiom 2.7. There is a strict linear order $<$ on $\mathbb{R}$ such that for all $x, y, z \in \mathbb{R}$:

(a) either $x < y$, $x = y$, or $x > y$;

(b) if $x < y$ then $x + z < y + z$;

(c) if $x < y$ and $z > 0$, then $xz < yz$.

For any $a, b \in \mathbb{R}$ with $a \leq b$, we define the open intervals

$$(-\infty, b) = \{x \in \mathbb{R} : x < b\},$$
$$(a, b) = \{x \in \mathbb{R} : a < x < b\},$$
$$(a, \infty) = \{x \in \mathbb{R} : a < x\},$$

the closed intervals

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

and the half-open intervals

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

All standard properties of inequalities follow from Axiom 2.6 and Axiom 2.7. For example: if $x < y$ and $z < 0$, then $xz > yz$, meaning that the direction of an inequality is reversed when it is multiplied by a negative number; and $x^2 > 0$ for every $x \neq 0$. In future, when we write an inequality such as $x < y$, we will implicitly require that $x, y \in \mathbb{R}$.

Real numbers satisfy many inequalities. A simple, but fundamental, example is the following.

Proposition 2.8. If $x, y \in \mathbb{R}$, then

$$xy \leq \frac{1}{2}\left(x^2 + y^2\right),$$

with equality if and only if $x = y$.

Proof. We have
$$0 \le (x - y)^2 = x^2 - 2xy + y^2,$$
with equality if and only if $x = y$, so $2xy \le x^2 + y^2$. $\qquad\qquad\square$

On writing $x = \sqrt{a}$, $y = \sqrt{b}$, where $a, b \ge 0$, in the result of Proposition 2.8, we get that
$$\sqrt{ab} \le \frac{a + b}{2},$$
which says that the geometric mean of two nonnegative numbers is less than or equal to their arithmetic mean, with equality if and only if the numbers are equal. A geometric interpretation of this inequality is that the square-root of the area of a rectangle is less than or equal to one-quarter of its perimeter, with equality if and only if the rectangle is a square. Thus, a square encloses the largest area among all rectangles of a given perimeter, which is a simple form of an isoperimetric inequality.

The arithmetic-geometric mean inequality generalizes to more than two numbers: If $n \in \mathbb{N}$ and $a_1, a_2, \ldots, a_n \ge 0$ are nonnegative real numbers, then
$$(a_1 a_2 \ldots a_n)^{1/n} \le \frac{a_1 + a_2 + \cdots + a_n}{n},$$
with equality if and only if all of the a_k are equal. For a proof, see e.g., Steele [**13**].

2.5. The supremum and infimum

Next, we use the ordering properties of $\mathbb{R}$ to define the supremum and infimum of a set of real numbers. These concepts are of central importance in analysis. In particular, in the next section we use them to state the completeness property of $\mathbb{R}$.

First, we define upper and lower bounds.

Definition 2.9. A set $A \subset \mathbb{R}$ of real numbers is bounded from above if there exists a real number $M \in \mathbb{R}$, called an upper bound of A, such that $x \le M$ for every $x \in A$. Similarly, A is bounded from below if there exists $m \in \mathbb{R}$, called a lower bound of A, such that $x \ge m$ for every $x \in A$. A set is bounded if it is bounded both from above and below.

Equivalently, a set A is bounded if $A \subset I$ for some bounded interval $I = [m, M]$.

Example 2.10. The interval $(0, 1)$ is bounded from above by every $M \ge 1$ and from below by every $m \le 0$. The interval $(-\infty, 0)$ is bounded from above by every $M \ge 0$, but it not bounded from below. The set of integers $\mathbb{Z}$ is not bounded from above or below.

If $A \subset \mathbb{R}$, we define $-A \subset \mathbb{R}$ by
$$-A = \{y \in \mathbb{R} : y = -x \text{ for some } x \in A\}.$$
For example, if $A = (0, \infty)$ consists of the positive real numbers, then $-A = (-\infty, 0)$ consists of the negative real numbers. A number m is a lower bound of

A if and only if $M = -m$ is an upper bound of $-A$. Thus, every result for upper bounds has a corresponding result for lower bounds, and we will often consider only upper bounds.

Definition 2.11. Suppose that $A \subset \mathbb{R}$ is a set of real numbers. If $M \in \mathbb{R}$ is an upper bound of A such that $M \leq M'$ for every upper bound M' of A, then M is called the least upper bound or supremum of A, denoted

$$M = \sup A.$$

If $m \in \mathbb{R}$ is a lower bound of A such that $m \geq m'$ for every lower bound m' of A, then m is called the greatest lower bound or infimum of A, denoted

$$m = \inf A.$$

If $A = \{x_i : i \in I\}$ is an indexed subset of $\mathbb{R}$, we also write

$$\sup A = \sup_{i \in I} x_i, \qquad \inf A = \inf_{i \in I} x_i.$$

As an immediate consequence of the definition, we note that the supremum (or infimum) of a set is unique if one exists: If M, M' are suprema of A, then $M \leq M'$ since M' is an upper bound of A and M is a least upper bound; similarly, $M' \leq M$, so $M = M'$. Furthermore, the supremum of a nonempty set A is always greater than or equal to its infimum if both exist. To see this, choose any $x \in A$. Since $\inf A$ is a lower bound and $\sup A$ is an upper bound of A, we have $\inf A \leq x \leq \sup A$.

If $\sup A \in A$, then we also denote it by $\max A$ and refer to it as the maximum of A; and if $\inf A \in A$, then we also denote it by $\min A$ and refer to it as the minimum of A. As the following examples illustrate, $\sup A$ and $\inf A$ may or may not belong to A, so the concepts of supremum and infimum must be clearly distinguished from those of maximum and minimum.

Example 2.12. Every finite set of real numbers

$$A = \{x_1, x_2, \ldots, x_n\}$$

is bounded. Its supremum is the greatest element,

$$\sup A = \max\{x_1, x_2, \ldots, x_n\},$$

and its infimum is the smallest element,

$$\inf A = \min\{x_1, x_2, \ldots, x_n\}.$$

Both the supremum and infimum of a finite set belong to the set.

Example 2.13. If $A = (0, 1)$, then every $M \geq 1$ is an upper bound of A. The least upper bound is $M = 1$, so

$$\sup(0, 1) = 1.$$

Similarly, every $m \leq 0$ is a lower bound of A, so

$$\inf(0, 1) = 0.$$

In this case, neither $\sup A$ nor $\inf A$ belong to A. The set $R = (0, 1) \cap \mathbb{Q}$ of rational numbers in $(0, 1)$, the closed interval $B = [0, 1]$, and the half-open interval $C = (0, 1]$ all have the same supremum and infimum as A. Neither $\sup R$ nor $\inf R$ belong to R, while both $\sup B$ and $\inf B$ belong to B, and only $\sup C$ belongs to C.

Example 2.14. Let

$$A = \left\{ \frac{1}{n} : n \in \mathbb{N} \right\}$$

be the set of reciprocals of the natural numbers. Then $\sup A = 1$, which belongs to A, and $\inf A = 0$, which does not belong to A.

A set must be bounded from above to have a supremum (or bounded from below to have an infimum), but the following notation for unbounded sets is convenient. We introduce a system of extended real numbers

$$\overline{\mathbb{R}} = \{-\infty\} \cup \mathbb{R} \cup \{\infty\}$$

which includes two new elements denoted $-\infty$ and ∞, ordered so that $-\infty < x < \infty$ for every $x \in \mathbb{R}$.

Definition 2.15. If a set $A \subset \mathbb{R}$ is not bounded from above, then $\sup A = \infty$, and if A is not bounded from below, then $\inf A = -\infty$.

For example, $\sup \mathbb{N} = \infty$ and $\inf \mathbb{R} = -\infty$. We also define $\sup \varnothing = -\infty$ and $\inf \varnothing = \infty$, since — by a strict interpretation of logic — every real number is both an upper and a lower bound of the empty set. With these conventions, every set of real numbers has a supremum and an infimum in $\overline{\mathbb{R}}$. Moreover, we may define the supremum and infimum of sets of extended real numbers in an obvious way; for example, $\sup A = \infty$ if $\infty \in A$ and $\inf A = -\infty$ if $-\infty \in A$.

While $\overline{\mathbb{R}}$ is linearly ordered, we cannot make it into a field however we extend addition and multiplication from $\mathbb{R}$ to $\overline{\mathbb{R}}$. Expressions such as $\infty - \infty$ or $0 \cdot \infty$ are inherently ambiguous. To avoid any possible confusion, we will give explicit definitions in terms of $\mathbb{R}$ alone for every expression that involves $\pm\infty$. Moreover, when we say that $\sup A$ or $\inf A$ exists, we will always mean that it exists as a real number, not as an extended real number. To emphasize this meaning, we will sometimes say that the supremum or infimum "exists as a finite real number."

2.6. Real numbers: completeness

The rational numbers $\mathbb{Q}$ and real numbers $\mathbb{R}$ have similar algebraic and order properties (they are both densely ordered fields). The crucial property that distinguishes $\mathbb{R}$ from $\mathbb{Q}$ is its completeness. There are two main ways to define the completeness of $\mathbb{R}$. The first, which we describe here, is based on the order properties of $\mathbb{R}$ and the existence of suprema. The second, which we describe in Chapter 3, is based on the metric properties of $\mathbb{R}$ and the convergence of Cauchy sequences.

We begin with an example that illustrates the difference between $\mathbb{Q}$ and $\mathbb{R}$.

Example 2.16. Define $A \subset \mathbb{Q}$ by

$$A = \left\{ x \in \mathbb{Q} : x^2 < 2 \right\}.$$

Then A is bounded from above by every $M \in \mathbb{Q}^+$ such that $M^2 > 2$. Nevertheless, A has no supremum in $\mathbb{Q}$ because $\sqrt{2}$ is irrational: for every upper bound $M \in \mathbb{Q}$ there exists $M' \in \mathbb{Q}$ such that $\sqrt{2} < M' < M$, so M isn't a least upper bound of A in $\mathbb{Q}$. On the other hand, A has a supremum in $\mathbb{R}$, namely $\sup A = \sqrt{2}$.

The following axiomatic property of the real numbers is called Dedekind completeness. Dedekind (1872) showed that the real numbers are characterized by the condition that they are a complete ordered field (that is, by Axiom 2.6, Axiom 2.7, and Axiom 2.17).

Axiom 2.17. Every nonempty set of real numbers that is bounded from above has a supremum.

Since $\inf A = -\sup(-A)$ and A is bounded from below if and only if $-A$ is bounded from above, it follows that every nonempty set of real numbers that is bounded from below has an infimum. The restriction to nonempty sets in Axiom 2.17 is necessary, since the empty set is bounded from above, but its supremum does not exist.

As a first application of this axiom, we prove that $\mathbb{R}$ has the Archimedean property, meaning that no real number is greater than every natural number.

Theorem 2.18. If $x \in \mathbb{R}$, then there exists $n \in \mathbb{N}$ such that $x < n$.

Proof. Suppose, for contradiction, that there exists $x \in \mathbb{R}$ such that $x > n$ for every $n \in \mathbb{N}$. Then x is an upper bound of $\mathbb{N}$, so $\mathbb{N}$ has a supremum $M = \sup \mathbb{N} \in \mathbb{R}$. Since $n \leq M$ for every $n \in \mathbb{N}$, we have $n - 1 \leq M - 1$ for every $n \in \mathbb{N}$, which implies that $n \leq M - 1$ for every $n \in \mathbb{N}$. But then $M - 1$ is an upper bound of $\mathbb{N}$, which contradicts the assumption that M is a least upper bound. $\square$

By taking reciprocals, we also get from this theorem that for every $\epsilon > 0$ there exists $n \in \mathbb{N}$ such that $0 < 1/n < \epsilon$.

These results say roughly that there are no infinite or infinitesimal real numbers. This property is consistent with our intuitive picture of a real line $\mathbb{R}$ that does not "extend past the natural numbers," where the natural numbers are obtained by counting upwards from 1. Robinson (1961) introduced extensions of the real numbers, called non-standard real numbers, which form non-Archimedean ordered fields with both infinite and infinitesimal elements, but they do not satisfy Axiom 2.17.

The following proof of the uncountability of $\mathbb{R}$ is based on its completeness and is Cantor's original proof (1874). The idea is to show that given any countable set of real numbers, there are additional real numbers in the "gaps" between them.

Theorem 2.19. The set of real numbers is uncountable.

Proof. Suppose that
$$S = \{x_1, x_2, x_3, \ldots, x_n, \ldots\}$$
is a countably infinite set of distinct real numbers. We will prove that there is a real number $x \in \mathbb{R}$ that does not belong to S.

If x_1 is the largest element of S, then no real number greater than x_1 belongs to S. Otherwise, we select recursively from S an increasing sequence of real numbers a_k and a decreasing sequence b_k as follows. Let $a_1 = x_1$ and choose $b_1 = x_{n_1}$ where n_1 is the smallest integer such that $x_{n_1} > a_1$. Then $x_n \notin (a_1, b_1)$ for all $1 \leq n \leq n_1$. If $x_n \notin (a_1, b_1)$ for all $n \in \mathbb{N}$, then no real number in (a_1, b_1) belongs to S, and we are done e.g., take $x = (a_1 + b_1)/2$. Otherwise, choose $a_2 = x_{m_2}$ where

$m_2 > n_1$ is the smallest integer such that $a_1 < x_{m_2} < b_1$. Then $x_n \notin (a_2, b_1)$ for all $1 \leq n \leq m_2$. If $x_n \notin (a_2, b_1)$ for all $n \in \mathbb{N}$, we are done. Otherwise, choose $b_2 = x_{n_2}$ where $n_2 > m_2$ is the smallest integer such that $a_2 < x_{n_2} < b_1$.

Continuing in this way, we either stop after finitely many steps and get an interval that is not included in S, or we get subsets $\{a_1, a_2, \dots\}$ and $\{b_1, b_2, \dots\}$ of $\{x_1, x_2, \dots\}$ such that

$$a_1 < a_2 < \cdots < a_k < \cdots < b_k < \cdots < b_2 < b_1.$$

It follows from the construction that for each $n \in \mathbb{N}$, we have $x_n \notin (a_k, b_k)$ when k is sufficiently large. Let

$$a = \sup_{k \in \mathbb{N}} a_k, \qquad \inf_{k \in \mathbb{N}} b_k = b,$$

which exist by the completeness of $\mathbb{R}$. Then $a \leq b$ (see Proposition 2.22 below) and $x \notin S$ if $a \leq x \leq b$, which proves the result. $\qquad\square$

This theorem shows that $\mathbb{R}$ is uncountable, but it doesn't show that $\mathbb{R}$ has the same cardinality as the power set $\mathcal{P}(\mathbb{N})$ of the natural numbers, whose uncountability was proved in Theorem 1.47. In Theorem 5.67, we show that $\mathbb{R}$ has the same cardinality as $\mathcal{P}(\mathbb{N})$; this provides a second proof that $\mathbb{R}$ is uncountable and shows that $\mathcal{P}(\mathbb{N})$ has the cardinality of the continuum.

2.7. Properties of the supremum and infimum

In this section, we collect some properties of the supremum and infimum for later use. This section can be referred back to as needed.

First, we state an equivalent way to characterize the supremum and infimum, which is an immediate consequence of Definition 2.11.

Proposition 2.20. If $A \subset \mathbb{R}$, then $M = \sup A$ if and only if: (a) M is an upper bound of A; (b) for every $M' < M$ there exists $x \in A$ such that $x > M'$. Similarly, $m = \inf A$ if and only if: (a) m is a lower bound of A; (b) for every $m' > m$ there exists $x \in A$ such that $x < m'$.

We frequently use this proposition as follows: (a) if M is an upper bound of A, then $\sup A \leq M$; (b) if A is nonempty and bounded from above, then for every $\epsilon > 0$, there exists $x \in A$ such that $x > \sup A - \epsilon$. Similarly: (a) if m is a lower bound of A, then $m \leq \inf A$; (b) if A is nonempty and bounded from below, then for every $\epsilon > 0$, there exists $x \in A$ such that $x < \inf A + \epsilon$.

Making a set smaller decreases its supremum and increases its infimum. In the following inequalities, we allow the sup and inf to be extended real numbers.

Proposition 2.21. Suppose that A, B are subsets of $\mathbb{R}$ such that $A \subset B$. Then $\sup A \leq \sup B$, and $\inf A \geq \inf B$.

Proof. The result is immediate if $B = \varnothing$, when $A = \varnothing$, so we may assume that B is nonempty. If B is not bounded from above, then $\sup B = \infty$, so $\sup A \leq \sup B$. If B bounded from above, then $\sup B$ is an upper bound of B. Since $A \subset B$, it follows that $\sup B$ is an upper bound of A, so $\sup A \leq \sup B$. Similarly, either $\inf B = -\infty$ or $\inf B$ is a lower bound of A, so $\inf A \geq \inf B$. $\qquad\square$

The next proposition states that if every element in one set is less than or equal to every element in another set, then the sup of the first set is less than or equal to the inf of the second set.

Proposition 2.22. Suppose that A, B are nonempty sets of real numbers such that $x \leq y$ for all $x \in A$ and $y \in B$. Then $\sup A \leq \inf B$.

Proof. Fix $y \in B$. Since $x \leq y$ for all $x \in A$, it follows that y is an upper bound of A, so $\sup A$ is finite and $\sup A \leq y$. Hence, $\sup A$ is a lower bound of B, so $\inf B$ is finite and $\sup A \leq \inf B$. $\square$

If $A \subset \mathbb{R}$ and $c \in \mathbb{R}$, then we define

$$cA = \{y \in \mathbb{R} : y = cx \text{ for some } x \in A\}.$$

Multiplication of a set by a positive number multiplies its sup and inf; multiplication by a negative number also exchanges its sup and inf.

Proposition 2.23. If $c \geq 0$, then

$$\sup cA = c \sup A, \qquad \inf cA = c \inf A.$$

If $c < 0$, then

$$\sup cA = c \inf A, \qquad \inf cA = c \sup A.$$

Proof. The result is obvious if $c = 0$. If $c > 0$, then $cx \leq M$ if and only if $x \leq M/c$, which shows that M is an upper bound of cA if and only if M/c is an upper bound of A, so $\sup cA = c \sup A$. If $c < 0$, then then $cx \leq M$ if and only if $x \geq M/c$, so M is an upper bound of cA if and only if M/c is a lower bound of A, so $\sup cA = c \inf A$. The remaining results follow similarly. $\square$

If $A, B \subset \mathbb{R}$, then we define

$$A + B = \{z \in \mathbb{R} : z = x + y \text{ for some } x \in A, y \in B\},$$
$$A - B = \{z \in \mathbb{R} : z = x - y \text{ for some } x \in A, y \in B\}.$$

Proposition 2.24. If A, B are nonempty sets, then

$$\sup(A + B) = \sup A + \sup B, \qquad \inf(A + B) = \inf A + \inf B,$$
$$\sup(A - B) = \sup A - \inf B, \qquad \inf(A - B) = \inf A - \sup B.$$

Proof. The set $A + B$ is bounded from above if and only if A and B are bounded from above, so $\sup(A + B)$ exists if and only if both $\sup A$ and $\sup B$ exist. In that case, if $x \in A$ and $y \in B$, then

$$x + y \leq \sup A + \sup B,$$

so $\sup A + \sup B$ is an upper bound of $A + B$, and therefore

$$\sup(A + B) \leq \sup A + \sup B.$$

To get the inequality in the opposite direction, suppose that $\epsilon > 0$. Then there exist $x \in A$ and $y \in B$ such that

$$x > \sup A - \frac{\epsilon}{2}, \qquad y > \sup B - \frac{\epsilon}{2}.$$

It follows that
$$x + y > \sup A + \sup B - \epsilon$$
for every $\epsilon > 0$, which implies that
$$\sup(A + B) \geq \sup A + \sup B.$$
Thus, $\sup(A + B) = \sup A + \sup B$. It follows from this result and Proposition 2.23 that
$$\sup(A - B) = \sup A + \sup(-B) = \sup A - \inf B.$$

The proof of the results for $\inf(A + B)$ and $\inf(A - B)$ is similar, or we can apply the results for the supremum to $-A$ and $-B$. $\qquad\square$

Finally, we prove that taking the supremum over a pair of indices gives the same result as taking successive suprema over each index separately.

Proposition 2.25. Suppose that
$$\{x_{ij} : i \in I, j \in J\}$$
is a doubly-indexed set of real numbers. Then
$$\sup_{(i,j) \in I \times J} x_{ij} = \sup_{i \in I} \left(\sup_{j \in J} x_{ij} \right).$$

Proof. For each $a \in I$, we have $\{a\} \times J \subset I \times J$, so
$$\sup_{j \in J} x_{aj} \leq \sup_{(i,j) \in I \times J} x_{ij}.$$
Taking the supremum of this inequality over $a \in I$, and replacing 'a' by 'i', we get that
$$\sup_{i \in I} \left(\sup_{j \in J} x_{ij} \right) \leq \sup_{(i,j) \in I \times J} x_{ij}.$$

To prove the reverse inequality, first note that if
$$\sup_{(i,j) \in I \times J} x_{ij}$$
is finite, then given $\epsilon > 0$ there exists $a \in I$, $b \in J$ such that
$$x_{ab} > \sup_{(i,j) \in I \times J} x_{ij} - \epsilon.$$
It follows that
$$\sup_{j \in J} x_{aj} > \sup_{(i,j) \in I \times J} x_{ij} - \epsilon,$$
and therefore that
$$\sup_{i \in I} \left(\sup_{j \in J} x_{ij} \right) > \sup_{(i,j) \in I \times J} x_{ij} - \epsilon.$$
Since $\epsilon > 0$ is arbitrary, we have
$$\sup_{i \in I} \left(\sup_{j \in J} x_{ij} \right) \geq \sup_{(i,j) \in I \times J} x_{ij}.$$

Similarly, if
$$\sup_{(i,j) \in I \times J} x_{ij} = \infty,$$

then given $M \in \mathbb{R}$ there exists $a \in I$, $b \in J$ such that $x_{ab} > M$, and it follows that

$$\sup_{i \in I} \left(\sup_{j \in J} x_{ij} \right) > M.$$

Since M is arbitrary, we have

$$\sup_{i \in I} \left(\sup_{j \in J} x_{ij} \right) = \infty,$$

which completes the proof.   □

2.8. Density of the Rationals

An essential property of the natural numbers is the following induction principle, which expresses the idea that we can reach every natural number by counting upwards from one.

Axiom 2.26. Suppose that $A \subset \mathbb{N}$ is a set of natural numbers such that: (a) $1 \in A$; (b) $n \in A$ implies $(n + 1) \in A$. Then $A = \mathbb{N}$.

Before stating the next theorem, we give a formal definition of the maximal and minimal elements of a set.

Definition 2.27. Suppose that $A \subset \mathbb{R}$ is a set of real numbers. A maximal element of A is a number $M = \max A$ such that $M \in A$ and $M \geq x$ for every $x \in A$. A minimal element of A is a number $m = \min A$ such that $m \in A$ and $m \leq x$ for every $x \in A$.

It follows immediately from the definition that if A has a maximal or minimal element, then $\sup A = \max A$ or $\inf A = \min A$. However, $\sup A$ or $\inf A$ may exist, even when $\max A$ or $\min A$ don't exist, if they don't belong to A.

The following result states a fundamental property of the natural numbers called the well-ordering property.

Theorem 2.28. Every nonempty subset of natural numbers contains a minimal element.

Proof. We will use induction to show that if a set $S \subset \mathbb{N}$ contains no minimal element, then S is the empty set, which proves the result.

Let
$$A = \{n \in \mathbb{N} : 1, 2, \ldots, n \notin S\}.$$
If $1 \notin A$, then $1 \in S$, so 1 a minimal element of S since $1 \leq n$ for every $n \in \mathbb{N}$, which contradicts our assumption on S. Hence, $1 \in A$.

Suppose that $n \in A$ for some $n \in \mathbb{N}$, meaning that $1, 2, \ldots, n \notin S$. If $n+1 \notin A$, then $n + 1 \in S$, so $n + 1$ is a minimal element of S. This contradiction shows that $n + 1 \in A$. It follows by induction that $A = \mathbb{N}$, which implies that $S = \varnothing$. $\square$

One can also prove that the well-ordering property of $\mathbb{N}$ implies the induction axiom, so the well-ordering of $\mathbb{N}$ is equivalent to induction.

The integers $\mathbb{Z}$ are not well-ordered (for example, $\mathbb{Z}$ itself has no minimal element), but a similar property holds for sets of integers that are bounded from below.

Corollary 2.29. If a nonempty set $S \subset \mathbb{Z}$ is bounded from below in $\mathbb{R}$, then S has a minimal element.

Proof. Since S is bounded from below in $\mathbb{R}$, the Archimedean property of $\mathbb{R}$ implies that S is bounded from below by some integer $a \in \mathbb{Z}$. Then
$$B = \{n - a + 1 : n \in S\} \subset \mathbb{N},$$
so Theorem 2.28 implies that B has a minimal element $b \in B$. It follows that $m = b + a - 1 \in S$ is a minimal element of S. $\square$

We can now prove the following (intuitively obvious) lemma, which we will use to prove the density of $\mathbb{Q}$ in $\mathbb{R}$.

Lemma 2.30. If $x, y \in \mathbb{R}$ satisfy $y - x > 1$, then there exists $m \in \mathbb{Z}$ such that $x < m < y$.

Proof. We construct m as the minimal integer strictly greater than x. Let
$$S = \{n \in \mathbb{Z} : x < n\}.$$
Then $S \subset \mathbb{Z}$ is nonempty (by the Archimedean property) and bounded from below (by x), so it has a minimal element $m \in S$ with $m - 1 \notin S$. It follows that $x < m$ and $x \geq m - 1$, so $x < m \leq x + 1 < y$, which proves the result. $\qquad\square$

Finally, we prove the density of the rational numbers in the real numbers, meaning that there is a rational number strictly between any pair of distinct real numbers (rational or irrational), however close together those real numbers may be.

Theorem 2.31. If $x, y \in \mathbb{R}$ and $x < y$, then there exists $r \in \mathbb{Q}$ such that $x < r < y$.

Proof. Let $\epsilon = y - x > 0$. By the Archimedean property, there exists $n \in \mathbb{N}$ such that $0 < 1/n < \epsilon$, which implies that $ny - nx > 1$. By Lemma 2.30, there exists $m \in \mathbb{Z}$ such that $nx < m < ny$, which proves the result with $r = m/n$. $\qquad\square$

Repeated application of this theorem shows that there are, in fact, infinitely many rational numbers between any pair of distinct real numbers. One can also use the result to prove that the irrational numbers $\mathbb{R} \setminus \mathbb{Q}$ are dense in $\mathbb{R}$.

Sequences

In this chapter, we discuss sequences. We say what it means for a sequence to converge, and define the limit of a convergent sequence. We begin with some preliminary results about the absolute value, which can be used to define a distance function, or metric, on $\mathbb{R}$. In turn, convergence is defined in terms of this metric.

3.1. The absolute value

Definition 3.1. The absolute value of $x \in \mathbb{R}$ is defined by

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

Some basic properties of the absolute value are the following.

Proposition 3.2. For all $x, y \in \mathbb{R}$:

(a) $|x| \geq 0$ and $|x| = 0$ if and only if $x = 0$;

(b) $|-x| = |x|$;

(c) $|x + y| \leq |x| + |y|$ (triangle inequality);

(d) $|xy| = |x|\,|y|$;

Proof. Parts (a), (b) follow immediately from the definition. Part (c) remains valid if we change the signs of both x and y or exchange x and y. Therefore we can assume that $x \geq 0$ and $|x| \geq |y|$ without loss of generality, in which case $x + y \geq 0$. If $y \geq 0$, corresponding to the case when x and y have the same sign, then

$$|x + y| = x + y = |x| + |y|.$$

If $y < 0$, corresponding to the case when x and y have opposite signs and $x + y > 0$, then

$$|x + y| = x + y = |x| - |y| < |x| + |y|,$$

which proves (c). Part (d) remains valid if we change x to $-x$ or y to $-y$, so we can assume that $x, y \geq 0$ without loss of generality. Then $xy \geq 0$ and $|xy| = xy = |x||y|$. $\qquad\square$

One useful consequence of the triangle inequality is the following reverse triangle inequality.

Proposition 3.3. If $x, y \in \mathbb{R}$, then

$$||x| - |y|| \leq |x - y|.$$

Proof. By the triangle inequality,

$$|x| = |x - y + y| \leq |x - y| + |y|$$

so $|x| - |y| \leq |x - y|$. Similarly, exchanging x and y, we get $|y| - |x| \leq |x - y|$, which proves the result. $\qquad\square$

We can give an equivalent condition for the boundedness of a set by using the absolute value instead of upper and lower bounds as in Definition 2.9.

Proposition 3.4. A set $A \subset \mathbb{R}$ is bounded if and only if there exists a real number $M \geq 0$ such that

$$|x| \leq M \text{ for every } x \in A.$$

Proof. If the condition in the proposition holds, then M is an upper bound of A and $-M$ is a lower bound, so A is bounded. Conversely, if A is bounded from above by M' and from below by m', then $|x| \leq M$ for every $x \in A$ where $M = \max\{|m'|, |M'|\}$. $\qquad\square$

A third way to say that a set is bounded is in terms of its diameter.

Definition 3.5. Let $A \subset \mathbb{R}$. The diameter of A is

$$\operatorname{diam} A = \sup \left\{ |x - y| : x, y \in A \right\}.$$

Then a set is bounded if and only if its diameter is finite.

Example 3.6. If $A = (-a, a)$, then $\operatorname{diam} A = 2a$, and A is bounded. If $A = (-\infty, a)$, then $\operatorname{diam} A = \infty$, and A is unbounded.

3.2. Sequences

A sequence (x_n) of real numbers is an ordered list of numbers $x_n \in \mathbb{R}$, called the terms of the sequence, indexed by the natural numbers $n \in \mathbb{N}$. We often indicate a sequence by listing the first few terms, especially if they have an obvious pattern. Of course, no finite list of terms is, on its own, sufficient to define a sequence.
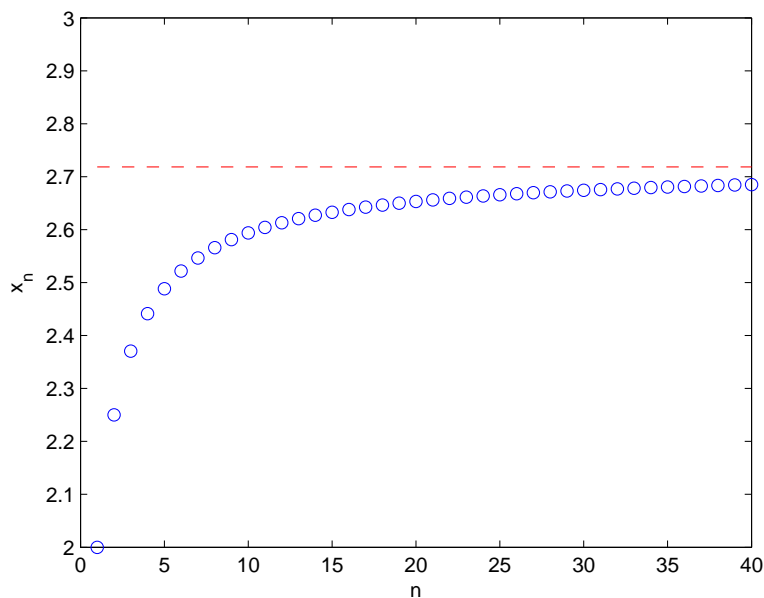
Figure 1. A plot of the first 40 terms in the sequence $x_n = (1 + 1/n)^n$, illustrating that it is monotone increasing and converges to $e \approx 2.718$, whose value is indicated by the dashed line.

Example 3.7. Here are some sequences:

$$1,\ 8,\ 27, 64,\ \ldots, \qquad x_n = n^3,$$
$$1,\ \frac{1}{2},\ \frac{1}{3},\ \frac{1}{4},\ \ldots \qquad x_n = \frac{1}{n};$$
$$1,\ -1,\ 1,\ -1,\ \ldots \qquad x_n = (-1)^{n+1},$$
$$(1+1),\ \left(1+\frac{1}{2}\right)^2,\ \left(1+\frac{1}{3}\right)^3,\ \ldots \qquad x_n = \left(1+\frac{1}{n}\right)^n.$$

Note that unlike sets, where elements are not repeated, the terms in a sequence may be repeated.

The formal definition of a sequence is as a function on $\mathbb{N}$, which is equivalent to its definition as a list.

Definition 3.8. A sequence (x_n) of real numbers is a function $f : \mathbb{N} \to \mathbb{R}$, where $x_n = f(n)$.

We can consider sequences of many different types of objects (for example, sequences of functions) but for now we only consider sequences of real numbers, and we will refer to them as sequences for short. A useful way to visualize a sequence (x_n) is to plot the graph of $x_n \in \mathbb{R}$ versus $n \in \mathbb{N}$. (See Figure 1 for an example.)

If we want to indicate the range of the index $n \in \mathbb{N}$ explicitly, we write the sequence as $(x_n)_{n=1}^{\infty}$. Sometimes it is convenient to start numbering a sequence from a different integer, such as $n = 0$ instead of $n = 1$. In that case, a sequence $(x_n)_{n=0}^{\infty}$ is a function $f : \mathbb{N}_0 \to \mathbb{R}$ where $x_n = f(n)$ and $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$, and similarly for other starting points.

Every function $f : \mathbb{N} \to \mathbb{R}$ defines a sequence, corresponding to an arbitrary choice of a real number $x_n \in \mathbb{R}$ for each $n \in \mathbb{N}$. Some sequences can be defined explicitly by giving an expression for the nth terms, as in Example 3.7; others can be defined recursively. That is, we specify the value of the initial term (or terms) in the sequence, and define x_n as a function of the previous terms $(x_1, x_2, \dots, x_{n-1})$.

A well-known example of a recursive sequence is the Fibonacci sequence (F_n)

$$1, \ 1, \ 2, \ 3, \ 5, \ 8, \ 13, \ \dots,$$

which is defined by $F_1 = F_2 = 1$ and

$$F_n = F_{n-1} + F_{n-2} \qquad \text{for } n \geq 3.$$

That is, we add the two preceding terms to get the next term. In general, we cannot expect to solve a recursion relation to get an explicit expression for the nth term in a sequence, but the recursion relation for the Fibonacci sequence is linear with constant coefficients, and it can be solved to give an expression for F_n called the Euler-Binet formula.

Proposition 3.9 (Euler-Binet formula). The nth term in the Fibonacci sequence is given by

$$F_n = \frac{1}{\sqrt{5}} \left[\phi^n - \left(-\frac{1}{\phi} \right)^n \right], \qquad \phi = \frac{1 + \sqrt{5}}{2}.$$

Proof. The terms in the Fibonacci sequence are uniquely determined by the linear difference equation

$$F_n - F_{n-1} - F_{n-2} = 0, \qquad n \geq 3,$$

with the initial conditions

$$F_1 = 1, \qquad F_2 = 1.$$

We see that $F_n = r^n$ is a solution of the difference equation if r satisfies

$$r^2 - r - 1 = 0,$$

which gives

$$r = \phi \text{ or } -\frac{1}{\phi}, \qquad \phi = \frac{1 + \sqrt{5}}{2} \approx 1.61803.$$

By linearity,

$$F_n = A\phi^n + B \left(-\frac{1}{\phi} \right)^n$$

is a solution of the difference equation for arbitrary constants A, B. This solution satisfies the initial conditions $F_1 = F_2 = 1$ if

$$A = \frac{1}{\sqrt{5}}, \qquad B = -\frac{1}{\sqrt{5}},$$

which proves the result. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \square$

Alternatively, once we know the answer, we can prove Proposition 3.9 by induction. The details are left as an exercise. Note that although the right-hand side of the equation for F_n involves the irrational number $\sqrt{5}$, its value is an integer for every $n \in \mathbb{N}$.

The number ϕ appearing in Proposition 3.9 is called the golden ratio. It has the property that subtracting 1 from it gives its reciprocal, or

$$\phi - 1 = \frac{1}{\phi}.$$

Geometrically, this property means that the removal of a square from a rectangle whose sides are in the ratio ϕ leaves a rectangle whose sides are in the same ratio. The number ϕ was originally defined in Euclid's Elements as the division of a line in "extreme and mean ratio," and Ancient Greek architects arguably used rectangles with this proportion in the Parthenon and other buildings. During the Renaissance, ϕ was referred to as the "divine proportion." The first use of the term "golden section" appears to be by Martin Ohm, brother of the physicist Georg Ohm, in a book published in 1835.

3.3. Convergence and limits

Roughly speaking, a sequence (x_n) converges to a limit x if its terms x_n get arbitrarily close to x for all sufficiently large n.

Definition 3.10. A sequence (x_n) of real numbers converges to a limit $x \in \mathbb{R}$, written

$$x = \lim_{n \to \infty} x_n, \quad \text{or} \quad x_n \to x \text{ as } n \to \infty,$$

if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$|x_n - x| < \epsilon \qquad \text{for all } n > N.$$

A sequence converges if it converges to some limit $x \in \mathbb{R}$, otherwise it diverges.

Although we don't show it explicitly in the definition, N is allowed to depend on ϵ. Typically, the smaller we choose ϵ, the larger we have to make N. One way to view a proof of convergence is as a game: If I give you an $\epsilon > 0$, you have to come up with an N that "works." Also note that $x_n \to x$ as $n \to \infty$ means the same thing as $|x_n - x| \to 0$ as $n \to \infty$.

It may appear obvious that a limit is unique if one exists, but this fact requires proof.

Proposition 3.11. If a sequence converges, then its limit is unique.

Proof. Suppose that (x_n) is a sequence such that $x_n \to x$ and $x_n \to x'$ as $n \to \infty$. Let $\epsilon > 0$. Then there exist $N, N' \in \mathbb{N}$ such that

$$|x_n - x| < \frac{\epsilon}{2} \qquad \text{for all } n > N,$$
$$|x_n - x'| < \frac{\epsilon}{2} \qquad \text{for all } n > N'.$$

Choose any $n > \max\{N, N'\}$. Then, by the triangle inequality,

$$|x - x'| \leq |x - x_n| + |x_n - x'| < \frac{\epsilon}{2} + \frac{\epsilon}{2} < \epsilon.$$

Since this inequality holds for all $\epsilon > 0$, we must have $|x - x'| = 0$ (otherwise the inequality would be false for $\epsilon = |x - x'|/2 > 0$), so $x = x'$. $\square$

The following notation for sequences that "diverge to infinity" is convenient.

Definition 3.12. If (x_n) is a sequence then

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

or $x_n \to \infty$ as $n \to \infty$, if for every $M \in \mathbb{R}$ there exists $N \in \mathbb{R}$ such that

$$x_n > M \qquad \text{for all } n > N.$$

Also

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

or $x_n \to -\infty$ as $n \to \infty$, if for every $M \in \mathbb{R}$ there exists $N \in \mathbb{R}$ such that

$$x_n < M \qquad \text{for all } n > N.$$

That is, $x_n \to \infty$ as $n \to \infty$ means the terms of the sequence (x_n) get arbitrarily large and positive for all sufficiently large n, while $x_n \to -\infty$ as $n \to \infty$ means that the terms get arbitrarily large and negative for all sufficiently large n. The notation $x_n \to \pm\infty$ does *not* mean that the sequence converges.

To illustrate these definitions, we discuss the convergence of the sequences in Example 3.7.

Example 3.13. The terms in the sequence

$$1, \ 8, \ 27, \ 64, \ \dots, \qquad x_n = n^3$$

eventually exceed any real number, so $n^3 \to \infty$ as $n \to \infty$ and this sequence does not converge. Explicitly, let $M \in \mathbb{R}$ be given, and choose $N \in \mathbb{N}$ such that $N > M^{1/3}$. (If $-\infty < M < 1$, we can choose $N = 1$.) Then for all $n > N$, we have $n^3 > N^3 > M$, which proves the result.

Example 3.14. The terms in the sequence

$$1, \ \frac{1}{2}, \ \frac{1}{3}, \ \frac{1}{4}, \ \dots \qquad x_n = \frac{1}{n}$$

get closer to 0 as n gets larger, and the sequence converges to 0:

$$\lim_{n \to \infty} \frac{1}{n} = 0.$$

To prove this limit, let $\epsilon > 0$ be given. Choose $N \in \mathbb{N}$ such that $N > 1/\epsilon$. (Such a number exists by the Archimedean property of $\mathbb{R}$ stated in Theorem 2.18.) Then for all $n > N$

$$\left| \frac{1}{n} - 0 \right| = \frac{1}{n} < \frac{1}{N} < \epsilon,$$

which proves that $1/n \to 0$ as $n \to \infty$.

Example 3.15. The terms in the sequence

$$1, \ -1, \ 1, \ -1, \ \ldots \qquad x_n = (-1)^{n+1},$$

oscillate back and forth infinitely often between 1 and -1, but they do not approach any fixed limit, so the sequence does not converge. To show this explicitly, note that for every $x \in \mathbb{R}$ we have either $|x - 1| \geq 1$ or $|x + 1| \geq 1$. It follows that there is no $N \in \mathbb{N}$ such that $|x_n - x| < 1$ for all $n > N$. Thus, Definition 3.10 fails if $\epsilon = 1$ however we choose $x \in \mathbb{R}$, and the sequence does not converge.

Example 3.16. The convergence of the sequence

$$(1 + 1), \ \left(1 + \frac{1}{2}\right)^2, \ \left(1 + \frac{1}{3}\right)^3, \ldots \qquad x_n = \left(1 + \frac{1}{n}\right)^n,$$

illustrated in Figure 1, is less obvious. Its terms are given by

$$x_n = a_n^n, \qquad a_n = 1 + \frac{1}{n}.$$

As n increases, we take larger powers of numbers that get closer to one. If $a > 1$ is any fixed real number, then $a^n \to \infty$ as $n \to \infty$ so the sequence (a^n) does not converge (see Proposition 3.31 below for a detailed proof). On the other hand, if $a = 1$, then $1^n = 1$ for all $n \in \mathbb{N}$ so the sequence (1^n) converges to 1. Thus, there are two competing factors in the sequence with increasing n: $a_n \to 1$ but $n \to \infty$. It is not immediately obvious which of these factors "wins."

In fact, they are in balance. As we prove in Proposition 3.32 below, the sequence converges with

$$\lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n = e,$$

where $2 < e < 3$. This limit can be taken as the definition of $e \approx 2.71828$.

For comparison, one can also show that

$$\lim_{n \to \infty} \left(1 + \frac{1}{n^2}\right)^n = 1, \qquad \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^{n^2} = \infty.$$

In the first case, the rapid approach of $a_n = 1 + 1/n^2$ to 1 "beats" the slower growth in the exponent n, while in the second case, the rapid growth in the exponent n^2 "beats" the slower approach of $a_n = 1 + 1/n$ to 1.

An important property of a sequence is whether or not it is bounded.

Definition 3.17. A sequence (x_n) of real numbers is bounded from above if there exists $M \in \mathbb{R}$ such that $x_n \leq M$ for all $n \in \mathbb{N}$, and bounded from below if there exists $m \in \mathbb{R}$ such that $x_n \geq m$ for all $n \in \mathbb{N}$. A sequence is bounded if it is bounded from above and below, otherwise it is unbounded.

An equivalent condition for a sequence (x_n) to be bounded is that there exists $M \geq 0$ such that

$$|x_n| \leq M \text{ for all } n \in \mathbb{N}.$$

Example 3.18. The sequence (n^3) is bounded from below but not from above, while the sequences $(1/n)$ and $\big((-1)^{n+1}\big)$ are bounded. The sequence

$$1, \ -2, \ 3, \ -4, \ 5, \ -6, \ \ldots \qquad x_n = (-1)^{n+1}n$$

is not bounded from below or above.

We then have the following property of convergent sequences.

Proposition 3.19. A convergent sequence is bounded.

Proof. Let (x_n) be a convergent sequence with limit x. There exists $N \in \mathbb{N}$ such that

$$|x_n - x| < 1 \qquad \text{for all } n > N.$$

The triangle inequality implies that

$$|x_n| \le |x_n - x| + |x| < 1 + |x| \qquad \text{for all } n > N.$$

Defining

$$M = \max \{|x_1|, |x_2|, \ldots, |x_N|, 1 + |x|\},$$

we see that $|x_n| \le M$ for all $n \in \mathbb{N}$, so (x_n) is bounded. $\qquad\square$

Thus, boundedness is a necessary condition for convergence, and every unbounded sequence diverges; for example, the unbounded sequence in Example 3.13 diverges. On the other hand, boundedness is not a sufficient condition for convergence; for example, the bounded sequence in Example 3.15 diverges.

The boundedness, or convergence, of a sequence $(x_n)_{n=1}^{\infty}$ depends only on the behavior of the infinite "tails" $(x_n)_{n=N}^{\infty}$ of the sequence, where N is arbitrarily large. Equivalently, the sequence $(x_n)_{n=1}^{\infty}$ and the shifted sequences $(x_{n+N})_{n=1}^{\infty}$ have the same convergence properties and limits for every $N \in \mathbb{N}$. As a result, changing a finite number of terms in a sequence doesn't alter its boundedness or convergence, nor does it alter the limit of a convergent sequence. In particular, the existence of a limit gives no information about how quickly a sequence converges to its limit.

Example 3.20. Changing the first hundred terms of the sequence $(1/n)$ from $1/n$ to n, we get the sequence

$$1, \ 2, \ 3, \ \ldots, \ 99, \ 100, \ \frac{1}{101}, \ \frac{1}{102}, \ \frac{1}{103}, \ \ldots,$$

which is still bounded (although by 100 instead of by 1) and still convergent to 0. Similarly, changing the first billion terms in the sequence doesn't change its boundedness or convergence.

We introduce some convenient terminology to describe the behavior of "tails" of a sequence,

Definition 3.21. Let $P(x)$ denote a property of real numbers $x \in \mathbb{R}$. If (x_n) is a real sequence, then $P(x_n)$ holds eventually if there exists $N \in \mathbb{N}$ such that $P(x_n)$ holds for all $n > N$; and $P(x_n)$ holds infinitely often if for every $N \in \mathbb{N}$ there exists $n > N$ such that $P(x_n)$ holds.

For example, (x_n) is bounded if there exists $M \ge 0$ such that $|x_n| \le M$ eventually; and (x_n) does not converge to $x \in \mathbb{R}$ if there exists $\epsilon_0 > 0$ such that $|x_n - x| \ge \epsilon_0$ infinitely often.

Note that if a property P holds infinitely often according to Definition 3.21, then it does indeed hold infinitely often: If $N = 1$, then there exists $n_1 > 1$ such that $P(x_{n_1})$ holds; if $N = n_1$, then there exists $n_2 > n_1$ such that $P(x_{n_2})$ holds; then there exists $n_3 > n_2$ such that $P(x_{n_3})$ holds, and so on.

3.4. Properties of limits

In this section, we prove some order and algebraic properties of limits of sequences.

3.4.1. Monotonicity. Limits of convergent sequences preserve (non-strict) inequalities.

Theorem 3.22. If (x_n) and (y_n) are convergent sequences and $x_n \leq y_n$ for all $n \in \mathbb{N}$, then
$$\lim_{n\to\infty} x_n \leq \lim_{n\to\infty} y_n.$$

Proof. Suppose that $x_n \to x$ and $y_n \to y$ as $n \to \infty$. Then for every $\epsilon > 0$ there exists $P, Q \in \mathbb{N}$ such that
$$|x - x_n| < \frac{\epsilon}{2} \qquad \text{for all } n > P,$$
$$|y - y_n| < \frac{\epsilon}{2} \qquad \text{for all } n > Q.$$
Choosing $n > \max\{P, Q\}$, we have
$$x = x_n + x - x_n < y_n + \frac{\epsilon}{2} = y + y_n - y + \frac{\epsilon}{2} < y + \epsilon.$$
Since $x < y + \epsilon$ for every $\epsilon > 0$, it follows that $x \leq y$. $\qquad\square$

This result, of course, remains valid if the inequality $x_n \leq y_n$ holds only for all sufficiently large n. Limits need not preserve strict inequalities. For example, $1/n > 0$ for all $n \in \mathbb{N}$ but $\lim_{n\to\infty} 1/n = 0$.

It follows immediately that if (x_n) is a convergent sequence with $m \leq x_n \leq M$ for all $n \in \mathbb{N}$, then
$$m \leq \lim_{n\to\infty} x_n \leq M.$$

The following "squeeze" or "sandwich" theorem is often useful in proving the convergence of a sequence by bounding it between two simpler convergent sequences with equal limits.

Theorem 3.23 (Sandwich). Suppose that (x_n) and (y_n) are convergent sequences of real numbers with the same limit L. If (z_n) is a sequence such that
$$x_n \leq z_n \leq y_n \qquad \text{for all } n \in \mathbb{N},$$
then (z_n) also converges to L.

Proof. Let $\epsilon > 0$ be given, and choose $P, Q \in \mathbb{N}$ such that
$$|x_n - L| < \epsilon \quad \text{for all } n > P, \qquad |y_n - L| < \epsilon \quad \text{for all } n > Q.$$
If $N = \max\{P, Q\}$, then for all $n > N$
$$-\epsilon < x_n - L \leq z_n - L \leq y_n - L < \epsilon,$$

which implies that $|z_n - L| < \epsilon$. This prove the result. $\qquad\square$

It is essential here that (x_n) and (y_n) have the same limit.

Example 3.24. If $x_n = -1$, $y_n = 1$, and $z_n = (-1)^{n+1}$, then $x_n \leq z_n \leq y_n$ for all $n \in \mathbb{N}$, the sequence (x_n) converges to -1 and (y_n) converges 1, but (z_n) does not converge.

As once consequence, we show that we can take absolute values inside limits.

Corollary 3.25. If $x_n \to x$ as $n \to \infty$, then $|x_n| \to |x|$ as $n \to \infty$.

Proof. By the reverse triangle inequality,

$$0 \leq \bigl|\,|x_n| - |x|\,\bigr| \leq |x_n - x|,$$

and the result follows from Theorem 3.23. $\qquad\square$

3.4.2. Linearity. Limits respect addition and multiplication. In proving the following theorem, we need to show that the sequences converge, not just get an expressions for their limits.

Theorem 3.26. Suppose that (x_n) and (y_n) are convergent real sequences and $c \in \mathbb{R}$. Then the sequences (cx_n), $(x_n + y_n)$, and $(x_n y_n)$ converge, and

$$\lim_{n\to\infty} cx_n = c \lim_{n\to\infty} x_n,$$
$$\lim_{n\to\infty} (x_n + y_n) = \lim_{n\to\infty} x_n + \lim_{n\to\infty} y_n,$$
$$\lim_{n\to\infty} (x_n y_n) = \left(\lim_{n\to\infty} x_n \right) \left(\lim_{n\to\infty} y_n \right).$$

Proof. We let

$$x = \lim_{n\to\infty} x_n, \qquad y = \lim_{n\to\infty} y_n.$$

The first statement is immediate if $c = 0$. Otherwise, let $\epsilon > 0$ be given, and choose $N \in \mathbb{N}$ such that

$$|x_n - x| < \frac{\epsilon}{|c|} \qquad \text{for all } n > N.$$

Then

$$|cx_n - cx| < \epsilon \qquad \text{for all } n > N,$$

which proves that (cx_n) converges to cx.

For the second statement, let $\epsilon > 0$ be given, and choose $P, Q \in \mathbb{N}$ such that

$$|x_n - x| < \frac{\epsilon}{2} \quad \text{for all } n > P, \qquad |y_n - y| < \frac{\epsilon}{2} \quad \text{for all } n > Q.$$

Let $N = \max\{P, Q\}$. Then for all $n > N$, we have

$$|(x_n + y_n) - (x + y)| \leq |x_n - x| + |y_n - y| < \epsilon,$$

which proves that $(x_n + y_n)$ converges to $x + y$.

For the third statement, note that since (x_n) and (y_n) converge, they are bounded and there exists $M > 0$ such that

$$|x_n|, |y_n| \leq M \qquad \text{for all } n \in \mathbb{N}$$

and $|x|, |y| \leq M$. Given $\epsilon > 0$, choose $P, Q \in \mathbb{N}$ such that

$$|x_n - x| < \frac{\epsilon}{2M} \quad \text{for all } n > P, \qquad |y_n - y| < \frac{\epsilon}{2M} \quad \text{for all } n > Q,$$

and let $N = \max\{P, Q\}$. Then for all $n > N$,

$$\begin{aligned}
|x_n y_n - xy| &= |(x_n - x)y_n + x(y_n - y)| \\
&\leq |x_n - x| \, |y_n| + |x| \, |y_n - y| \\
&\leq M \left(|x_n - x| + |y_n - y|\right) \\
&< \epsilon,
\end{aligned}$$

which proves that $(x_n y_n)$ converges to xy. $\qquad\square$

Note that the convergence of $(x_n + y_n)$ does not imply the convergence of (x_n) and (y_n) separately; for example, take $x_n = n$ and $y_n = -n$. If, however, (x_n) converges then (y_n) converges if and only if $(x_n + y_n)$ converges.

3.5. Monotone sequences

Monotone sequences have particularly simple convergence properties.

Definition 3.27. A sequence (x_n) of real numbers is increasing if $x_{n+1} \geq x_n$ for all $n \in \mathbb{N}$, decreasing if $x_{n+1} \leq x_n$ for all $n \in \mathbb{N}$, and monotone if it is increasing or decreasing. A sequence is strictly increasing if $x_{n+1} > x_n$, strictly decreasing if $x_{n+1} < x_n$, and strictly monotone if it is strictly increasing or strictly decreasing.

We don't require a monotone sequence to be strictly monotone, but this usage isn't universal. In some places, "increasing" or "decreasing" is used to mean "strictly increasing" or "strictly decreasing." In that case, what we call an increasing sequence is called a nondecreasing sequence and a decreasing sequence is called nonincreasing sequence. We'll use the more easily understood direct terminology.

Example 3.28. The sequence

$$1, \ 2, \ 2, \ 3, \ 3, \ 3, \ 4, \ 4, \ 4, \ 4, \ 5, \ \dots$$

is monotone increasing but not strictly monotone increasing; the sequence (n^3) is strictly monotone increasing; the sequence $(1/n)$ is strictly monotone decreasing; and the sequence $((-1)^{n+1})$ is not monotone.

Bounded monotone sequences always converge, and unbounded monotone sequences diverge to $\pm\infty$.

Theorem 3.29. A monotone sequence of real numbers converges if and only if it is bounded. If (x_n) is monotone increasing and bounded, then

$$\lim_{n \to \infty} x_n = \sup\{x_n : n \in \mathbb{N}\},$$

and if (x_n) is monotone decreasing and bounded, then

$$\lim_{n \to \infty} x_n = \inf\{x_n : n \in \mathbb{N}\}.$$

Furthermore, if (x_n) is monotone increasing and unbounded, then

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

and if (x_n) is monotone decreasing and unbounded, then

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

Proof. If the sequence converges, then by Proposition 3.19 it is bounded.

Conversely, suppose that (x_n) is a bounded, monotone increasing sequence. The set of terms $\{x_n : n \in \mathbb{N}\}$ is bounded from above, so by Axiom 2.17 it has a supremum

$$x = \sup\{x_n : n \in \mathbb{N}\}.$$

Let $\epsilon > 0$. From the definition of the supremum, there exists an $N \in \mathbb{N}$ such that $x_N > x - \epsilon$. Since the sequence is increasing, we have $x_n \geq x_N$ for all $n > N$, and therefore $x - \epsilon < x_n \leq x$. It follows that

$$|x_n - x| < \epsilon \qquad \text{for all } n > N,$$

which proves that $x_n \to x$ as $n \to \infty$.

If (x_n) is an unbounded monotone increasing sequence, then it is not bounded from above, since it is bounded from below by its first term x_1. Hence, for every $M \in \mathbb{R}$ there exists $N \in \mathbb{N}$ such that $x_N > M$. Since the sequence is increasing, we have $x_n \geq x_N > M$ for all $n > N$, which proves that $x_n \to \infty$ as $n \to \infty$.

The result for a monotone decreasing sequence (x_n) follows similarly, or by applying the previous result to the monotone increasing sequence $(-x_n)$. □

The fact that every bounded monotone sequence has a limit is another way to express the completeness of $\mathbb{R}$. For example, this is not true in $\mathbb{Q}$: an increasing sequence of rational numbers that converges to $\sqrt{2}$ is bounded from above in $\mathbb{Q}$ (for example, by 2) but has no limit in $\mathbb{Q}$.

We sometimes use the notation $x_n \uparrow x$ to indicate that (x_n) is a monotone increasing sequence that converges to x, and $x_n \downarrow x$ to indicate that (x_n) is a monotone decreasing sequence that converges to x, with a similar notation for monotone sequences that diverge to $\pm\infty$. For example, $1/n \downarrow 0$ and $n^3 \uparrow \infty$ as $n \to \infty$.

The following propositions give some examples of monotone sequences. In the proofs, we use the binomial theorem, which we state without proof.

Theorem 3.30 (Binomial). If $x, y \in \mathbb{R}$ and $n \in \mathbb{N}$, then

$$(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k, \qquad \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Here, $n! = 1 \cdot 2 \cdot 3 \cdots \cdot n$ and, by convention, $0! = 1$. The binomial coefficients

$$\binom{n}{k} = \frac{n \cdot (n-1) \cdot (n-2) \cdots \cdot (n-k+1)}{1 \cdot 2 \cdot 3 \cdots \cdot k},$$

read "n choose k," give the number of ways of choosing k objects from n objects, order not counting. For example,

$$(x + y)^2 = x^2 + 2xy + y^2,$$
$$(x + y)^3 = x^3 + 3x^2 y + 3xy^2 + y^3,$$
$$(x + y)^4 = x^4 + 4x^3 y + 6x^2 y^2 + 4xy^3 + y^4.$$

We also recall the sum of a geometric series: if $a \neq 1$, then

$$\sum_{k=0}^{n} a^k = \frac{1 - a^{n+1}}{1 - a}.$$

Proposition 3.31. The geometric sequence $(a^n)_{n=0}^{\infty}$,

$$1, \ a, \ a^2, \ a^3, \ \ldots,$$

is strictly monotone decreasing if $0 < a < 1$, with

$$\lim_{n \to \infty} a^n = 0,$$

and strictly monotone increasing if $1 < a < \infty$, with

$$\lim_{n \to \infty} a^n = \infty.$$

Proof. If $0 < a < 1$, then $0 < a^{n+1} = a \cdot a^n < a^n$, so the sequence (a^n) is strictly monotone decreasing and bounded from below by zero. Therefore by Theorem 3.29 it has a limit $x \in \mathbb{R}$. Theorem 3.26 implies that

$$x = \lim_{n \to \infty} a^{n+1} = \lim_{n \to \infty} a \cdot a^n = a \lim_{n \to \infty} a^n = ax.$$

Since $a \neq 1$, it follows that $x = 0$.

If $a > 1$, then $a^{n+1} = a \cdot a^n > a^n$, so (a^n) is strictly increasing. Let $a = 1 + \delta$ where $\delta > 0$. By the binomial theorem, we have

$$a^n = (1 + \delta)^n$$
$$= \sum_{k=0}^{n} \binom{n}{k} \delta^k$$
$$= 1 + n\delta + \frac{1}{2}n(n - 1)\delta^2 + \cdots + \delta^n$$
$$> 1 + n\delta.$$

Given $M \geq 0$, choose $N \in \mathbb{N}$ such that $N > M/\delta$. Then for all $n > N$, we have

$$a^n > 1 + n\delta > 1 + N\delta > M,$$

so $a_n \to \infty$ as $n \to \infty$. $\square$

The next proposition proves the existence of the limit for e in Example 3.16.

Proposition 3.32. The sequence (x_n) with

$$x_n = \left(1 + \frac{1}{n}\right)^n$$

is strictly monotone increasing and converges to a limit $2 < e < 3$.

Proof. By the binomial theorem,

$$\left(1 + \frac{1}{n}\right)^n = \sum_{k=0}^{n} \binom{n}{k} \frac{1}{n^k}$$

$$= 1 + n \cdot \frac{1}{n} + \frac{n(n-1)}{2!} \cdot \frac{1}{n^2} + \frac{n(n-1)(n-2)}{3!} \cdot \frac{1}{n^3}$$

$$+ \cdots + \frac{n(n-1)(n-2)\ldots 2 \cdot 1}{n!} \cdot \frac{1}{n^n}$$

$$= 2 + \frac{1}{2!}\left(1 - \frac{1}{n}\right) + \frac{1}{3!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)$$

$$+ \cdots + \frac{1}{n!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\ldots \frac{2}{n} \cdot \frac{1}{n}.$$

Each of the terms in the sum on the right hand side is a positive increasing function of n, and the number of terms increases with n. Therefore (x_n) is a strictly increasing sequence, and $x_n > 2$ for every $n \geq 2$. Moreover, since $0 \leq (1 - k/n) < 1$ for $1 \leq k \leq n$, we have

$$\left(1 + \frac{1}{n}\right)^n < 2 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!}.$$

Since $n! \geq 2^{n-1}$ for $n \geq 1$, it follows that

$$\left(1 + \frac{1}{n}\right)^n < 2 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^{n-1}} = 2 + \frac{1}{2}\left[\frac{1 - (1/2)^{n-1}}{1 - 1/2}\right] < 3,$$

so (x_n) is monotone increasing and bounded from above by a number strictly less than 3. By Theorem 3.29 the sequence converges to a limit $2 < e < 3$. □

3.6. The $\limsup$ **and** $\liminf$

The $\limsup$ and $\liminf$ allow us to reduce questions about the convergence and limits of general real sequences to ones about monotone sequences. They are somewhat subtle concepts, and after defining them we will consider a number of examples.

Unlike the limit, the $\limsup$ and $\liminf$ of every bounded sequence of real numbers exist. A sequence converges if and only if its $\limsup$ and $\liminf$ are equal, in which case its limit is their common value. Furthermore, a sequence is unbounded if and only if at least one of its $\limsup$ or $\liminf$ diverges to $\pm\infty$, and it diverges to $\pm\infty$ if and only if both its $\limsup$ and $\liminf$ diverge to $\pm\infty$.

In order to define the $\limsup$ and $\liminf$ of a sequence (x_n) of real numbers, we introduce the sequences (y_n) and (z_n) obtained by taking the supremum and infimum, respectively, of the "tails" of (x_n):

$$y_n = \sup\{x_k : k \geq n\}, \qquad z_n = \inf\{x_k : k \geq n\}.$$

As n increases, the supremum and infimum are taken over smaller sets, so (y_n) is monotone decreasing and (z_n) is monotone increasing. The limits of these sequences are the $\limsup$ and $\liminf$, respectively, of the original sequence.

Definition 3.33. Suppose that (x_n) is a sequence of real numbers. Then

$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} y_n, \qquad y_n = \sup\{x_k : k \geq n\},$$

$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty} z_n, \qquad z_n = \inf\{x_k : k \geq n\}.$$

Note that $\limsup x_n$ exists and is finite provided that each of the y_n is finite and (y_n) is bounded from below; similarly, $\liminf x_n$ exists and is finite provided that each of the z_n is finite and (z_n) is bounded from above. We may also write

$$\limsup_{n\to\infty} x_n = \inf_{n\in\mathbb{N}}\left(\sup_{k\geq n} x_k\right), \qquad \liminf_{n\to\infty} x_n = \sup_{n\in\mathbb{N}}\left(\inf_{k\geq n} x_k\right).$$

As for the limit of monotone sequences, it is convenient to allow the $\liminf$ or $\limsup$ to be $\pm\infty$, and we state explicitly what this means. We have $-\infty < y_n \leq \infty$ for every $n \in \mathbb{N}$, since the supremum of a non-empty set cannot be $-\infty$, but we may have $y_n \downarrow -\infty$; similarly, $-\infty \leq z_n < \infty$, but we may have $z_n \uparrow \infty$. These possibilities lead to the following cases.

Definition 3.34. Suppose that (x_n) is a sequence of real numbers and the sequences (y_n), (z_n) of possibly extended real numbers are given by Definition 3.33. Then

$$\limsup_{n\to\infty} x_n = \infty \qquad \text{if } y_n = \infty \text{ for every } n \in \mathbb{N},$$

$$\limsup_{n\to\infty} x_n = -\infty \qquad \text{if } y_n \downarrow -\infty \text{ as } n \to \infty,$$

$$\liminf_{n\to\infty} x_n = -\infty \qquad \text{if } z_n = -\infty \text{ for every } n \in \mathbb{N},$$

$$\liminf_{n\to\infty} x_n = \infty \qquad \text{if } z_n \uparrow \infty \text{ as } n \to \infty.$$

In all cases, we have $z_n \leq y_n$ for every $n \in \mathbb{N}$, with the usual ordering conventions for $\pm\infty$, and by taking the limit as $n \to \infty$, we get that

$$\liminf_{n\to\infty} x_n \leq \limsup_{n\to\infty} x_n.$$

We illustrate the definition of the $\limsup$ and $\liminf$ with a number of examples.

Example 3.35. Consider the bounded, increasing sequence

$$0, \ \frac{1}{2}, \ \frac{2}{3}, \ \frac{3}{4}, \ldots \qquad x_n = 1 - \frac{1}{n}.$$

Defining y_n and z_n as above, we have

$$y_n = \sup\left\{1 - \frac{1}{k} : k \geq n\right\} = 1, \quad z_n = \inf\left\{1 - \frac{1}{k} : k \geq n\right\} = 1 - \frac{1}{n},$$

and both $y_n \downarrow 1$ and $z_n \uparrow 1$ converge monotonically to the limit 1 of the original sequence. Thus,

$$\limsup_{n\to\infty} x_n = \liminf_{n\to\infty} x_n = \lim_{n\to\infty} x_n = 1.$$

Example 3.36. Consider the bounded, non-monotone sequence

$$1, \ -1, \ 1, \ -1, \ldots \qquad x_n = (-1)^{n+1}.$$
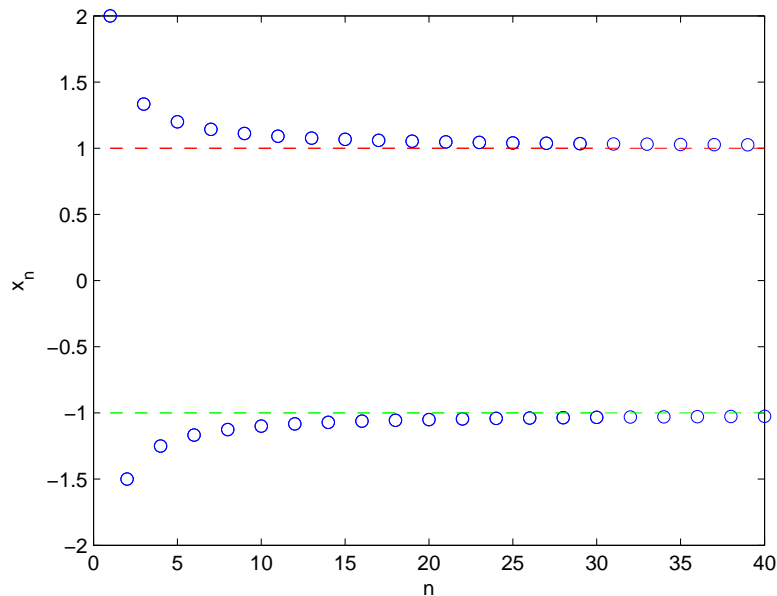
Figure 2. A plot of the first 40 terms in the sequence $x_n = (-1)^{n+1}(1+1/n)$ in Example 3.37. The dashed lines show $\limsup x_n = 1$ and $\liminf x_n = -1$.

Then

$$y_n = \sup \left\{ (-1)^{k+1} : k \geq n \right\} = 1, \quad z_n = \inf \left\{ (-1)^{k+1} : k \geq n \right\} = -1,$$

and $y_n \downarrow 1$, $z_n \uparrow -1$ converge to different limits. Thus,

$$\limsup_{n \to \infty} x_n = 1, \qquad \liminf_{n \to \infty} x_n = -1.$$

The original sequence does not converge, and $\lim x_n$ is undefined.

Example 3.37. The bounded, non-monotone sequence

$$2, \ -\frac{3}{2}, \ \frac{4}{3}, \ -\frac{5}{4}, \dots \qquad x_n = (-1)^{n+1}\left(1 + \frac{1}{n}\right)$$

is shown in Figure 2. We have

$$y_n = \sup \{x_k : k \geq n\} = \begin{cases} 1 + 1/n & \text{if } n \text{ is odd}, \\ 1 + 1/(n+1) & \text{if } n \text{ is even}, \end{cases}$$

$$z_n = \inf \{x_k : k \geq n\} = \begin{cases} -[1 + 1/(n+1)] & \text{if } n \text{ is odd}, \\ -[1 + 1/n] & \text{if } n \text{ is even}, \end{cases}$$

and it follows that

$$\limsup_{n \to \infty} x_n = 1, \qquad \liminf_{n \to \infty} x_n = -1.$$

The limit of the sequence does not exist. Note that infinitely many terms of the sequence are strictly greater than $\limsup x_n$, so $\limsup x_n$ does not bound any

"tail" of the sequence from above. However, every number strictly greater than $\limsup x_n$ eventually bounds the sequence from above. Similarly, $\liminf x_n$ does not bound any "tail" of the sequence from below, but every number strictly less than $\liminf x_n$ eventually bounds the sequence from below.

Example 3.38. Consider the unbounded, increasing sequence

$$1,\ 2,\ 3,\ 4,\ 5,\dots \qquad x_n = n.$$

We have

$$y_n = \sup\{x_k : k \geq n\} = \infty, \qquad z_n = \inf\{x_k : k \geq n\} = n,$$

so the lim sup, lim inf and lim all diverge to ∞,

$$\limsup_{n\to\infty} x_n = \liminf_{n\to\infty} x_n = \lim_{n\to\infty} x_n = \infty.$$

Example 3.39. Consider the unbounded, non-monotone sequence

$$1,\ -2,\ 3,\ -4,\ 5,\dots \qquad x_n = \begin{cases} n & \text{if } n \text{ is odd,} \\ -n & \text{if } n \text{ is even.} \end{cases}$$

We have $y_n = \infty$ and $z_n = -\infty$ for every $n \in \mathbb{N}$, and

$$\limsup_{n\to\infty} x_n = \infty, \qquad \liminf_{n\to\infty} x_n = -\infty.$$

The sequence oscillates and does not diverge to either ∞ or $-\infty$, so $\lim x_n$ is undefined even as an extended real number.

Example 3.40. Consider the unbounded non-monotone sequence

$$1,\ \frac{1}{2},\ 3,\ \frac{1}{4},\ 5,\dots \qquad x_n = \begin{cases} n & \text{if } n \text{ is odd.} \\ 1/n & \text{if } n \text{ is even,} \end{cases}$$

Then $y_n = \infty$ and $z_n = 0$, so

$$\limsup_{n\to\infty} x_n = \infty, \qquad \liminf_{n\to\infty} x_n = 0.$$

As noted above, the lim sup of a sequence needn't bound any tail of the sequence, but the sequence is eventually bounded from above by every number that is strictly greater than the lim sup, and the sequence is greater infinitely often than every number that is strictly less than the lim sup. This property gives an alternative characterization of the lim sup, one that we often use in practice.

Theorem 3.41. Let (x_n) be a real sequence. Then

$$y = \limsup_{n\to\infty} x_n$$

if and only if $-\infty \leq y \leq \infty$ satisfies one of the following conditions.

(1) $-\infty < y < \infty$ and for every $\epsilon > 0$: (a) there exists $N \in \mathbb{N}$ such that $x_n < y + \epsilon$ for all $n > N$; (b) for every $N \in \mathbb{N}$ there exists $n > N$ such that $x_n > y - \epsilon$.

(2) $y = \infty$ and for every $M \in \mathbb{R}$, there exists $n \in \mathbb{N}$ such that $x_n > M$, i.e., (x_n) is not bounded from above.

(3) $y = -\infty$ and for every $m \in \mathbb{R}$ there exists $N \in \mathbb{N}$ such that $x_n < m$ for all $n > N$, i.e., $x_n \to -\infty$ as $n \to \infty$.

Similarly,

$$z = \liminf_{n\to\infty} x_n$$

if and only if $-\infty \leq z \leq \infty$ satisfies one of the following conditions.

(1) $-\infty < z < \infty$ and for every $\epsilon > 0$: (a) there exists $N \in \mathbb{N}$ such that $x_n > z - \epsilon$ for all $n > N$; (b) for every $N \in \mathbb{N}$ there exists $n > N$ such that $x_n < z + \epsilon$.

(2) $z = -\infty$ and for every $m \in \mathbb{R}$, there exists $n \in \mathbb{N}$ such that $x_n < m$, i.e., (x_n) is not bounded from below.

(3) $z = \infty$ and for every $M \in \mathbb{R}$ there exists $N \in \mathbb{N}$ such that $x_n > M$ for all $n > N$, i.e., $x_n \to \infty$ as $n \to \infty$.

Proof. We prove the result for $\limsup$. The result for $\liminf$ follows by applying this result to the sequence $(-x_n)$.

First, suppose that $y = \limsup x_n$ and $-\infty < y < \infty$. Then (x_n) is bounded from above and

$$y_n = \sup\{x_k : k \geq n\} \downarrow y \qquad \text{as } n \to \infty.$$

Therefore, for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $y_N < y + \epsilon$. Since $x_n \leq y_N$ for all $n > N$, this proves (1a). To prove (1b), let $\epsilon > 0$ and suppose that $N \in \mathbb{N}$ is arbitrary. Since $y_N \geq y$ is the supremum of $\{x_n : n \geq N\}$, there exists $n \geq N$ such that $x_n > y_N - \epsilon \geq y - \epsilon$, which proves (1b).

Conversely, suppose that $-\infty < y < \infty$ satisfies condition (1) for the $\limsup$. Then, given any $\epsilon > 0$, (1a) implies that there exists $N \in \mathbb{N}$ such that

$$y_n = \sup\{x_k : k \geq n\} \leq y + \epsilon \qquad \text{for all } n > N,$$

and (1b) implies that $y_n > y - \epsilon$ for all $n \in \mathbb{N}$. Hence, $|y_n - y| < \epsilon$ for all $n > N$, so $y_n \to y$ as $n \to \infty$, which means that $y = \limsup x_n$.

We leave the verification of the equivalence for $y = \pm\infty$ as an exercise. $\qquad \square$

Next we give a necessary and sufficient condition for the convergence of a sequence in terms of its $\liminf$ and $\limsup$.

Theorem 3.42. A sequence (x_n) of real numbers converges if and only if

$$\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n = x$$

are finite and equal, in which case

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

Furthermore, the sequence diverges to ∞ if and only if

$$\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n = \infty$$

and diverges to $-\infty$ if and only if

$$\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n = -\infty$$

Proof. First suppose that
$$\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n = x$$
for some $x \in \mathbb{R}$. Then $y_n \downarrow x$ and $z_n \uparrow x$ as $n \to \infty$ where
$$y_n = \sup\{x_k : k \geq n\}, \qquad z_n = \inf\{x_k : k \geq n\}.$$
Since $z_n \leq x_n \leq y_n$, the "sandwich" theorem implies that
$$\lim_{n\to\infty} x_n = x.$$

Conversely, suppose that the sequence (x_n) converges to a limit $x \in \mathbb{R}$. Then for every $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that
$$x - \epsilon < x_n < x + \epsilon \qquad \text{for all } n > N.$$
It follows that
$$x - \epsilon \leq z_n \leq y_n \leq x + \epsilon \qquad \text{for all } n > N.$$
Therefore $y_n, z_n \to x$ as $n \to \infty$, so $\limsup x_n = \liminf x_n = x$.

The sequence (x_n) diverges to ∞ if and only if $\liminf x_n = \infty$, and then $\limsup x_n = \infty$, since $\liminf x_n \leq \limsup x_n$. Similarly, (x_n) diverges to $-\infty$ if and only if $\limsup x_n = -\infty$, and then $\liminf x_n = -\infty$. $\qquad\square$

If $\liminf x_n \neq \limsup x_n$, then we say that the sequence (x_n) oscillates. The difference
$$\limsup x_n - \liminf x_n$$
provides a measure of the size of the oscillations in the sequence as $n \to \infty$.

Every sequence has a finite or infinite $\limsup$, but not every sequence has a limit (even if we include sequences that diverge to $\pm\infty$). The following corollary gives a convenient way to prove the convergence of a sequence without having to refer to the limit before it is known to exist.

Corollary 3.43. Let (x_n) be a sequence of real numbers. Then (x_n) converges with $\lim_{n\to\infty} x_n = x$ if and only if $\limsup_{n\to\infty} |x_n - x| = 0$.

Proof. If $\lim_{n\to\infty} x_n = x$, then $\lim_{n\to\infty} |x_n - x| = 0$, so
$$\limsup_{n\to\infty} |x_n - x| = \lim_{n\to\infty} |x_n - x| = 0.$$

Conversely, if $\limsup_{n\to\infty} |x_n - x| = 0$, then
$$0 \leq \liminf_{n\to\infty} |x_n - x| \leq \limsup_{n\to\infty} |x_n - x| = 0,$$
so $\liminf_{n\to\infty} x_n |x_n - x| = \limsup_{n\to\infty} x_n |x_n - x| = 0$. Theorem 3.42 implies that $\lim_{n\to\infty} |x_n - x| = 0$, or $\lim_{n\to\infty} x_n = x$. $\qquad\square$

Note that the condition $\liminf_{n\to\infty} |x_n - x| = 0$ doesn't tell us anything about the convergence of (x_n).

Example 3.44. Let $x_n = 1 + (-1)^n$. Then (x_n) oscillates between 0 and 2, and
$$\liminf_{n\to\infty} x_n = 0, \qquad \limsup_{n\to\infty} x_n = 2.$$
The sequence is non-negative and its $\liminf$ is 0, but the sequence does not converge.

3.7. Cauchy sequences

> Cauchy has become unbearable. Every Monday, broadcasting the known
> facts he has learned over the week as a discovery. I believe there is no
> historical precedent for such a talent writing so much awful rubbish. This
> is why I have relegated him to the rank below us. (Jacobi in a letter to
> Dirichlet, 1841)

The Cauchy condition is a necessary and sufficient condition for the convergence
of a real sequence that depends only on the terms of the sequence and not on its
limit. Furthermore, the completeness of $\mathbb{R}$ can be defined by the convergence of
Cauchy sequences, instead of by the existence of suprema. This approach defines
completeness in terms of the distance properties of $\mathbb{R}$ rather than its order properties
and generalizes to other metric spaces that don't have a natural ordering.

Roughly speaking, a Cauchy sequence is a sequence whose terms eventually get
arbitrarily close together.

Definition 3.45. A sequence (x_n) of real numbers is a Cauchy sequence if for
every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$|x_m - x_n| < \epsilon \qquad \text{for all } m, n > N.$$

Theorem 3.46. A sequence of real numbers converges if and only if it is a Cauchy
sequence.

Proof. First suppose that (x_n) converges to a limit $x \in \mathbb{R}$. Then for every $\epsilon > 0$
there exists $N \in \mathbb{N}$ such that

$$|x_n - x| < \frac{\epsilon}{2} \qquad \text{for all } n > N.$$

It follows that if $m, n > N$, then

$$|x_m - x_n| \le |x_m - x| + |x - x_n| < \epsilon,$$

which implies that (x_n) is Cauchy. (This direction doesn't use the completeness of
$\mathbb{R}$; for example, it holds equally well for sequence of rational numbers that converge
in $\mathbb{Q}$.)

Conversely, suppose that (x_n) is Cauchy. Then there is $N_1 \in \mathbb{N}$ such that

$$|x_m - x_n| < 1 \qquad \text{for all } m, n > N_1.$$

It follows that if $n > N_1$, then

$$|x_n| \le |x_n - x_{N_1+1}| + |x_{N_1+1}| \le 1 + |x_{N_1+1}|.$$

Hence the sequence is bounded with

$$|x_n| \le \max\left\{|x_1|, |x_2|, \ldots, |x_{N_1}|, 1 + |x_{N_1+1}|\right\}.$$

Since the sequence is bounded, its lim sup and lim inf exist. We claim they are
equal. Given $\epsilon > 0$, choose $N \in \mathbb{N}$ such that the Cauchy condition in Definition 3.45
holds. Then

$$x_n - \epsilon < x_m < x_n + \epsilon \qquad \text{for all } m \ge n > N.$$

It follows that for all $n > N$ we have

$$x_n - \epsilon \le \inf\{x_m : m \ge n\}, \quad \sup\{x_m : m \ge n\} \le x_n + \epsilon,$$

which implies that

$$\sup\{x_m : m \geq n\} - \epsilon \leq \inf\{x_m : m \geq n\} + \epsilon.$$

Taking the limit as $n \to \infty$, we get that

$$\limsup_{n\to\infty} x_n - \epsilon \leq \liminf_{n\to\infty} x_n + \epsilon,$$

and since $\epsilon > 0$ is arbitrary, we have

$$\limsup_{n\to\infty} x_n \leq \liminf_{n\to\infty} x_n.$$

It follows that $\limsup x_n = \liminf x_n$, so Theorem 3.42 implies that the sequence converges. $\square$

3.8. Subsequences

A subsequence of a sequence (x_n)

$$x_1, \; x_2, \; x_3, \; \ldots, \; x_n, \; \ldots$$

is a sequence (x_{n_k}) of the form

$$x_{n_1}, \; x_{n_2}, \; x_{n_3}, \; \ldots, \; x_{n_k}, \; \ldots$$

where $n_1 < n_2 < n_3 < \cdots < n_k < \ldots$.

Example 3.47. A subsequence of the sequence $(1/n)$,

$$1, \; \frac{1}{2}, \; \frac{1}{3}, \; \frac{1}{4}, \; \frac{1}{5}, \; \ldots.$$

is the sequence $(1/k^2)$

$$1, \; \frac{1}{4}, \; \frac{1}{9}, \; \frac{1}{16}, \; \frac{1}{25}, \; \ldots.$$

Here, $n_k = k^2$. On the other hand, the sequences

$$1, \; 1, \; \frac{1}{2}, \; \frac{1}{3}, \; \frac{1}{4}, \; \frac{1}{5}, \; \ldots, \qquad\qquad \frac{1}{2}, \; 1, \; \frac{1}{3}, \; \frac{1}{4}, \; \frac{1}{5}, \; \ldots$$

aren't subsequences of $(1/n)$ since n_k is not a strictly increasing function of k in either case.

The standard short-hand notation for subsequences used above is convenient but not entirely consistent, and the notion of a subsequence is a bit more involved than it might appear at first sight. To explain it in more detail, we give the formal definition of a subsequence as a function on $\mathbb{N}$.

Definition 3.48. Let (x_n) be a sequence, where $x_n = f(n)$ and $f : \mathbb{N} \to \mathbb{R}$. A sequence (y_k), where $y_k = g(k)$ and $g : \mathbb{N} \to \mathbb{R}$, is a subsequence of (x_n) if there is a strictly increasing function $\phi : \mathbb{N} \to \mathbb{N}$ such that $g = f \circ \phi$. In that case, we write $\phi(k) = n_k$ and $y_k = x_{n_k}$.

Example 3.49. In Example 3.47, the sequence $(1/n)$ corresponds to the function $f(n) = 1/n$ and the subsequence $(1/k^2)$ corresponds to $g(k) = 1/k^2$. Here, $g = f \circ \phi$ with $\phi(k) = k^2$.

Note that since the indices in a subsequence form a strictly increasing sequence of integers (n_k), it follows that $n_k \to \infty$ as $k \to \infty$.

Proposition 3.50. Every subsequence of a convergent sequence converges to the limit of the sequence.

Proof. Suppose that (x_n) is a convergent sequence with $\lim_{n\to\infty} x_n = x$ and (x_{n_k}) is a subsequence. Let $\epsilon > 0$. There exists $N \in \mathbb{N}$ such that $|x_n - x| < \epsilon$ for all $n > N$. Since $n_k \to \infty$ as $k \to \infty$, there exists $K \in \mathbb{N}$ such that $n_k > N$ if $k > K$. Then $k > K$ implies that $|x_{n_k} - x| < \epsilon$, so $\lim_{k\to\infty} x_{n_k} = x$. $\square$

A useful criterion for the divergence of a sequence follows immediately from this result and the uniqueness of limits.

Corollary 3.51. If a sequence has subsequences that converge to different limits, then the sequence diverges.

Example 3.52. The sequence $((-1)^{n+1})$,

$$1, \ -1, \ 1, \ -1, \ 1, \ \ldots,$$

has subsequences (1) and (-1) that converge to different limits, so it diverges.

In general, we define the limit set of a sequence to be the set of all limits of its convergent subsequences.

Definition 3.53. The limit set of a sequence (x_n) is the set

$$\{x \in \mathbb{R} : \text{there is a subsequence } (x_{n_k}) \text{ such that } x_{n_k} \to x \text{ as } k \to \infty\}$$

of limits of all of its convergent subsequences.

The limit set of a convergent sequence consists of a single point, namely its limit.

Example 3.54. The limit set of the divergent sequence $((-1)^{n+1})$,

$$1, \ -1, \ 1, \ -1, \ 1, \ \ldots,$$

contains two points, and is $\{-1, 1\}$.

Example 3.55. Let $\{r_n : n \in \mathbb{N}\}$ be an enumeration of the rational numbers in $[0, 1]$. Every $x \in [0, 1]$ is a limit of a subsequence (r_{n_k}). To obtain such a subsequence recursively, choose $n_1 = 1$, and for each $k \geq 2$ choose a rational number r_{n_k} such that $|x - r_{n_k}| < 1/k$ and $n_k > n_{k-1}$. This is always possible since the rational numbers are dense in $[0, 1]$ and every interval contains infinitely many terms of the sequence. Conversely, if $r_{n_k} \to x$, then $0 \leq x \leq 1$ since $0 \leq r_{n_k} \leq 1$. Thus, the limit set of (r_n) is the interval $[0, 1]$.

Finally, we state a characterization of the $\limsup$ and $\liminf$ of a sequence in terms of of its limit set, where we use the usual conventions about $\pm\infty$. We leave the proof as an exercise.

Theorem 3.56. Suppose that (x_n) is sequence of real numbers with limit set S. Then

$$\limsup_{n\to\infty} x_n = \sup S, \qquad \liminf_{n\to\infty} x_n = \inf S.$$

3.9. The Bolzano-Weierstrass theorem

The Bolzano-Weierstrass theorem is a fundamental compactness result. It allows us to deduce the convergence of a subsequence from the boundedness of a sequence without having to know anything specific about the limit. In this respect, it is analogous to the result that a monotone increasing sequence converges if it is bounded from above, and it also provides another way of expressing the completeness of $\mathbb{R}$.

Theorem 3.57 (Bolzano-Weierstrass). Every bounded sequence of real numbers has a convergent subsequence.

Proof. Suppose that (x_n) is a bounded sequence of real numbers. Let

$$M = \sup_{n \in \mathbb{N}} x_n, \qquad m = \inf_{n \in \mathbb{N}} x_n,$$

and define the closed interval $I_0 = [m, M]$.

Divide $I_0 = L_0 \cup R_0$ in half into two closed intervals, where

$$L_0 = [m, (m + M)/2], \qquad R_0 = [(m + M)/2, M].$$

At least one of the intervals L_0, R_0 contains infinitely many terms of the sequence, meaning that $x_n \in L_0$ or $x_n \in R_0$ for infinitely many $n \in \mathbb{N}$ (even if the terms themselves are repeated).

Choose I_1 to be one of the intervals L_0, R_0 that contains infinitely many terms and choose $n_1 \in \mathbb{N}$ such that $x_{n_1} \in I_1$. Divide $I_1 = L_1 \cup R_1$ in half into two closed intervals. One or both of the intervals L_1, R_1 contains infinitely many terms of the sequence. Choose I_2 to be one of these intervals and choose $n_2 > n_1$ such that $x_{n_2} \in I_2$. This is always possible because I_2 contains infinitely many terms of the sequence. Divide I_2 in half, pick a closed half-interval I_3 that contains infinitely many terms, and choose $n_3 > n_2$ such that $x_{n_3} \in I_3$. Continuing in this way, we get a nested sequence of intervals $I_1 \supset I_2 \supset I_3 \supset \ldots I_k \supset \ldots$ of length $|I_k| = 2^{-k}(M - m)$, together with a subsequence (x_{n_k}) such that $x_{n_k} \in I_k$.

Let $\epsilon > 0$ be given. Since $|I_k| \to 0$ as $k \to \infty$, there exists $K \in \mathbb{N}$ such that $|I_k| < \epsilon$ for all $k > K$. Furthermore, since $x_{n_k} \in I_K$ for all $k > K$ we have $|x_{n_j} - x_{n_k}| < \epsilon$ for all $j, k > K$. This proves that (x_{n_k}) is a Cauchy sequence, and therefore it converges by Theorem 3.46. $\qquad \square$

The subsequence obtained in the proof of this theorem is not unique. In particular, if the sequence does not converge, then for some $k \in \mathbb{N}$ both the left and right intervals L_k and R_k contain infinitely many terms of the sequence. In that case, we can obtain convergent subsequences with different limits, depending on our choice of L_k or R_k. This loss of uniqueness is a typical feature of compactness arguments.

We can, however, use the Bolzano-Weierstrass theorem to give a criterion for the convergence of a sequence in terms of the convergence of its subsequences. It states that if every convergent subsequence of a bounded sequence has the same limit, then the entire sequence converges to that limit.

Theorem 3.58. If (x_n) is a bounded sequence of real numbers such that every convergent subsequence has the same limit x, then (x_n) converges to x.

Proof. We will prove that if a bounded sequence (x_n) does not converge to x, then it has a convergent subsequence whose limit is not equal to x.

If (x_n) does not converges to x then there exists $\epsilon_0 > 0$ such that $|x_n - x| \geq \epsilon_0$ for infinitely many $n \in \mathbb{N}$. We can therefore find a subsequence (x_{n_k}) such that

$$|x_{n_k} - x| \geq \epsilon_0$$

for every $k \in \mathbb{N}$. The subsequence (x_{n_k}) is bounded, since (x_n) is bounded, so by the Bolzano-Weierstrass theorem, it has a convergent subsequence $(x_{n_{k_j}})$. If

$$\lim_{j \to \infty} x_{n_{k_j}} = y,$$

then $|x - y| \geq \epsilon_0$, so $x \neq y$, which proves the result. $\qquad\square$

Series

Divergent series are the devil, and it is a shame to base on them any demonstration whatsoever. (Niels Henrik Abel, 1826)

This series is divergent, therefore we may be able to do something with it. (Oliver Heaviside, quoted by Kline)

In this chapter, we apply our results for sequences to series, or infinite sums. The convergence and sum of an infinite series is defined in terms of its sequence of finite partial sums.

4.1. Convergence of series

A finite sum of real numbers is well-defined by the algebraic properties of $\mathbb{R}$, but in order to make sense of an infinite series, we need to consider its convergence. We say that a series converges if its sequence of partial sums converges, and in that case we define the sum of the series to be the limit of its partial sums.

Definition 4.1. Let (a_n) be a sequence of real numbers. The series

$$\sum_{n=1}^{\infty} a_n$$

converges to a sum $S \in \mathbb{R}$ if the sequence (S_n) of partial sums

$$S_n = \sum_{k=1}^{n} a_k$$

converges to S as $n \to \infty$. Otherwise, the series diverges.

If a series converges to S, we write

$$S = \sum_{n=1}^{\infty} a_n.$$

We also say a series diverges to $\pm\infty$ if its sequence of partial sums does. As for sequences, we may start a series at other values of n than $n = 1$ without changing its convergence properties. It is sometimes convenient to omit the limits on a series when they aren't important, and write it as $\sum a_n$.

Example 4.2. If $|a| < 1$, then the geometric series with ratio a converges and its sum is

$$\sum_{n=0}^{\infty} a^n = \frac{1}{1-a}.$$

This series is simple enough that we can compute its partial sums explicitly,

$$S_n = \sum_{k=0}^{n} a^k = \frac{1 - a^{n+1}}{1 - a}.$$

As shown in Proposition 3.31, if $|a| < 1$, then $a^n \to 0$ as $n \to \infty$, so that $S_n \to 1/(1-a)$, which proves the result.

The geometric series diverges to ∞ if $a \geq 1$, and diverges in an oscillatory fashion if $a \leq -1$. The following examples consider the cases $a = \pm 1$ in more detail.

Example 4.3. The series

$$\sum_{n=1}^{\infty} 1 = 1 + 1 + 1 + \dots$$

diverges to ∞, since its nth partial sum is $S_n = n$.

Example 4.4. The series

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

diverges, since its partial sums

$$S_n = \begin{cases} 1 & \text{if } n \text{ is odd,} \\ 0 & \text{if } n \text{ is even,} \end{cases}$$

oscillate between 0 and 1.

This series illustrates the dangers of blindly applying algebraic rules for finite sums to series. For example, one might argue that

$$S = (1 - 1) + (1 - 1) + (1 - 1) + \dots = 0 + 0 + 0 + \dots = 0,$$

or that

$$S = 1 + (-1 + 1) + (-1 + 1) + \dots = 1 + 0 + 0 + \dots = 1,$$

or that

$$1 - S = 1 - (1 - 1 + 1 - 1 + \dots) = 1 - 1 + 1 - 1 + 1 - \dots = S,$$

so $2S = 1$ or $S = 1/2$.

The Italian mathematician and priest Luigi Grandi (1710) suggested that these results were evidence in favor of the existence of God, since they showed that it was possible to create something out of nothing.

Telescoping series of the form

$$\sum_{n=1}^{\infty} (a_n - a_{n+1})$$

are another class of series whose partial sums

$$S_n = a_1 - a_{n+1}$$

can be computed explicitly and then used to study their convergence. We give one example.

Example 4.5. The series

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \frac{1}{4 \cdot 5} + \cdots$$

converges to 1. To show this, we observe that

$$\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1},$$

so

$$\begin{aligned}
\sum_{k=1}^{n} \frac{1}{k(k+1)} &= \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k+1} \right) \\
&= \frac{1}{1} - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \frac{1}{3} - \frac{1}{4} + \cdots + \frac{1}{n} - \frac{1}{n+1} \\
&= 1 - \frac{1}{n+1},
\end{aligned}$$

and it follows that

$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)} = 1.$$

A condition for the convergence of series with positive terms follows immediately from the condition for the convergence of monotone sequences.

Proposition 4.6. A series $\sum a_n$ with positive terms $a_n \geq 0$ converges if and only if its partial sums

$$\sum_{k=1}^{n} a_k \leq M$$

are bounded from above, otherwise it diverges to ∞.

Proof. The partial sums $S_n = \sum_{k=1}^{n} a_k$ of such a series form a monotone increasing sequence, and the result follows immediately from Theorem 3.29 □

Although we have only defined sums of convergent series, divergent series are not necessarily meaningless. For example, the Cesàro sum C of a series $\sum a_n$ is defined by

$$C = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} S_n, \qquad S_n = a_1 + a_2 + \cdots + a_n.$$

That is, we average the first n partial sums the series, and let $n \to \infty$. One can prove that if a series converges to S, then its Cesàro sum exists and is equal to S, but a series may be Cesàro summable even if it is divergent.

Example 4.7. For the series $\sum (-1)^{n+1}$ in Example 4.4, we find that

$$\frac{1}{n} \sum_{k=1}^{n} S_k = \begin{cases} 1/2 + 1/(2n) & \text{if } n \text{ is odd}, \\ 1/2 & \text{if } n \text{ is even}, \end{cases}$$

since the S_n's alternate between 0 and 1. It follows the Cesàro sum of the series is $C = 1/2$. This is, in fact, what Grandi believed to be the "true" sum of the series.

Cesàro summation is important in the theory of Fourier series. There are also many other ways to sum a divergent series or assign a meaning to it (for example, as an asymptotic series), but we won't discuss them further here.

4.2. The Cauchy condition

The following Cauchy condition for the convergence of series is an immediate consequence of the Cauchy condition for the sequence of partial sums.

Theorem 4.8 (Cauchy condition)**.** The series

$$\sum_{n=1}^{\infty} a_n$$

converges if and only for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$\left| \sum_{k=m+1}^{n} a_k \right| = |a_{m+1} + a_{m+2} + \cdots + a_n| < \epsilon \qquad \text{for all } n > m > N.$$

Proof. The series converges if and only if the sequence (S_n) of partial sums is Cauchy, meaning that for every $\epsilon > 0$ there exists N such that

$$|S_n - S_m| = \left| \sum_{k=m+1}^{n} a_k \right| < \epsilon \qquad \text{for all } n > m > N,$$

which proves the result. $\qquad \square$

A special case of this theorem is a necessary condition for the convergence of a series, namely that its terms approach zero. This condition is the first thing to check when considering whether or not a given series converges.

Theorem 4.9. If the series

$$\sum_{n=1}^{\infty} a_n$$

converges, then

$$\lim_{n \to \infty} a_n = 0.$$

Proof. If the series converges, then it is Cauchy. Taking $m = n - 1$ in the Cauchy condition in Theorem 4.8, we find that for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $|a_n| < \epsilon$ for all $n > N$, which proves that $a_n \to 0$ as $n \to \infty$. $\qquad \square$

Example 4.10. The geometric series $\sum a^n$ converges if $|a| < 1$ and in that case $a^n \to 0$ as $n \to \infty$. If $|a| \geq 1$, then $a^n \not\to 0$ as $n \to \infty$, which implies that the series diverges.

The condition that the terms of a series approach zero is not, however, sufficient to imply convergence. The following series is a fundamental example.

Example 4.11. The harmonic series

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots$$

diverges, even though $1/n \to 0$ as $n \to \infty$. To see this, we collect the terms in successive groups of powers of two,

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \left(\frac{1}{9} + \frac{1}{10} + \dots + \frac{1}{16}\right) + \dots$$

$$> 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) + \left(\frac{1}{16} + \frac{1}{16} + \dots + \frac{1}{16}\right) + \dots$$

$$> 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \dots.$$

In general, for every $n \geq 1$, we have

$$\sum_{k=1}^{2^{n+1}} \frac{1}{k} = 1 + \frac{1}{2} + \sum_{j=1}^{n} \sum_{k=2^j+1}^{2^{j+1}} \frac{1}{k}$$

$$> 1 + \frac{1}{2} + \sum_{j=1}^{n} \sum_{k=2^j+1}^{2^{j+1}} \frac{1}{2^{j+1}}$$

$$> 1 + \frac{1}{2} + \sum_{j=1}^{n} \frac{1}{2}$$

$$> \frac{n}{2} + \frac{3}{2},$$

so the series diverges. We can similarly obtain an upper bound for the partial sums,

$$\sum_{k=1}^{2^{n+1}} \frac{1}{k} < 1 + \frac{1}{2} + \sum_{j=1}^{n} \sum_{k=2^j+1}^{2^{j+1}} \frac{1}{2^j} < n + \frac{3}{2}.$$

These inequalities are rather crude, but they show that the series diverges at a logarithmic rate, since the sum of 2^n terms is of the order n. This rate of divergence is very slow. It takes 12367 terms for the partial sums of harmonic series to exceed 10, and more than 1.5×10^{43} terms for the partial sums to exceed 100.

A more refined argument, using integration, shows that

$$\lim_{n \to \infty} \left[\sum_{k=1}^{n} \frac{1}{k} - \log n \right] = \gamma$$

where $\gamma \approx 0.5772$ is the Euler constant. (See Example 12.45.)

4.3. Absolutely convergent series

There is an important distinction between absolutely and conditionally convergent series.

Definition 4.12. The series

$$\sum_{n=1}^{\infty} a_n$$

converges absolutely if

$$\sum_{n=1}^{\infty} |a_n| \ \text{converges},$$

and converges conditionally if

$$\sum_{n=1}^{\infty} a_n \ \text{converges, but} \ \sum_{n=1}^{\infty} |a_n| \ \text{diverges}.$$

We will show in Proposition 4.17 below that every absolutely convergent series converges. For series with positive terms, there is no difference between convergence and absolute convergence. Also note from Proposition 4.6 that $\sum a_n$ converges absolutely if and only if the partial sums $\sum_{k=1}^{n} |a_k|$ are bounded from above.

Example 4.13. The geometric series $\sum a^n$ is absolutely convergent if $|a| < 1$.

Example 4.14. The alternating harmonic series,

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots$$

is not absolutely convergent since, as shown in Example 4.11, the harmonic series diverges. It follows from Theorem 4.30 below that the alternating harmonic series converges, so it is a conditionally convergent series. Its convergence is made possible by the cancelation between terms of opposite signs.

As we show next, the convergence of an absolutely convergent series follows from the Cauchy condition. Moreover, the series of positive and negative terms in an absolutely convergent series converge separately. First, we introduce some convenient notation.

Definition 4.15. The positive and negative parts of a real number $a \in \mathbb{R}$ are given by

$$a^+ = \begin{cases} a & \text{if } a > 0, \\ 0 & \text{if } a \leq 0, \end{cases} \qquad a^- = \begin{cases} 0 & \text{if } a \geq 0, \\ |a| & \text{if } a < 0. \end{cases}$$

It follows, in particular, that

$$0 \leq a^+, a^- \leq |a|, \qquad a = a^+ - a^-, \qquad |a| = a^+ + a^-.$$

We may then split a series of real numbers into its positive and negative parts.

Example 4.16. Consider the alternating harmonic series

$$\sum_{n=1}^{\infty} a_n = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \dots$$

Its positive and negative parts are given by

$$\sum_{n=1}^{\infty} a_n^+ = 1 + 0 + \frac{1}{3} + 0 + \frac{1}{5} + 0 + \dots,$$

$$\sum_{n=1}^{\infty} a_n^- = 0 + \frac{1}{2} + 0 + \frac{1}{4} + 0 + \frac{1}{6} + \dots.$$

Both of these series diverge to infinity, since the harmonic series diverges and

$$\sum_{n=1}^{\infty} a_n^+ > \sum_{n=1}^{\infty} a_n^- = \frac{1}{2} \sum_{n=1}^{\infty} \frac{1}{n}.$$

Proposition 4.17. An absolutely convergent series converges. Moreover,

$$\sum_{n=1}^{\infty} a_n$$

converges absolutely if and only if the series

$$\sum_{n=1}^{\infty} a_n^+, \qquad \sum_{n=1}^{\infty} a_n^-$$

of positive and negative terms both converge. Furthermore, in that case

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} a_n^+ - \sum_{n=1}^{\infty} a_n^-, \qquad \sum_{n=1}^{\infty} |a_n| = \sum_{n=1}^{\infty} a_n^+ + \sum_{n=1}^{\infty} a_n^-.$$

Proof. If $\sum a_n$ is absolutely convergent, then $\sum |a_n|$ is convergent, so it satisfies the Cauchy condition. Since

$$\left| \sum_{k=m+1}^{n} a_k \right| \leq \sum_{k=m+1}^{n} |a_k|,$$

the series $\sum a_n$ also satisfies the Cauchy condition, and therefore it converges.

For the second part, note that

$$0 \leq \sum_{k=m+1}^{n} |a_k| = \sum_{k=m+1}^{n} a_k^+ + \sum_{k=m+1}^{n} a_k^-,$$

$$0 \leq \sum_{k=m+1}^{n} a_k^+ \leq \sum_{k=m+1}^{n} |a_k|,$$

$$0 \leq \sum_{k=m+1}^{n} a_k^- \leq \sum_{k=m+1}^{n} |a_k|,$$

which shows that $\sum |a_n|$ is Cauchy if and only if both $\sum a_n^+$, $\sum a_n^-$ are Cauchy. It follows that $\sum |a_n|$ converges if and only if both $\sum a_n^+$, $\sum a_n^-$ converge. In that

case, we have

$$\sum_{n=1}^{\infty} a_n = \lim_{n\to\infty} \sum_{k=1}^{n} a_k$$

$$= \lim_{n\to\infty} \left(\sum_{k=1}^{n} a_k^+ - \sum_{k=1}^{n} a_k^- \right)$$

$$= \lim_{n\to\infty} \sum_{k=1}^{n} a_k^+ - \lim_{n\to\infty} \sum_{k=1}^{n} a_k^-$$

$$= \sum_{n=1}^{\infty} a_n^+ - \sum_{n=1}^{\infty} a_n^-,$$

and similarly for $\sum |a_n|$, which proves the proposition. $\qquad\square$

It is worth noting that this result depends crucially on the completeness of $\mathbb{R}$.

Example 4.18. Suppose that $a_n^+, a_n^- \in \mathbb{Q}^+$ are positive rational numbers such that

$$\sum_{n=1}^{\infty} a_n^+ = \sqrt{2}, \qquad \sum_{n=1}^{\infty} a_n^- = 2 - \sqrt{2},$$

and let $a_n = a_n^+ - a_n^-$. Then

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} a_n^+ - \sum_{n=1}^{\infty} a_n^- = 2\sqrt{2} - 2 \notin \mathbb{Q},$$

$$\sum_{n=1}^{\infty} |a_n| = \sum_{n=1}^{\infty} a_n^+ + \sum_{n=1}^{\infty} a_n^- = 2 \in \mathbb{Q}.$$

Thus, the series converges absolutely in $\mathbb{Q}$, but it doesn't converge in $\mathbb{Q}$.

4.4. The comparison test

One of the most useful ways of showing that a series is absolutely convergent is to compare it with a simpler series whose convergence is already known.

Theorem 4.19 (Comparison test)**.** Suppose that $b_n \geq 0$ and

$$\sum_{n=1}^{\infty} b_n$$

converges. If $|a_n| \leq b_n$, then

$$\sum_{n=1}^{\infty} a_n$$

converges absolutely.

Proof. Since $\sum b_n$ converges it satisfies the Cauchy condition, and since

$$\sum_{k=m+1}^{n} |a_k| \leq \sum_{k=m+1}^{n} b_k$$

the series $\sum |a_n|$ also satisfies the Cauchy condition. Therefore $\sum a_n$ converges absolutely. $\qquad\square$

Example 4.20. The series

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots.$$

converges by comparison with the telescoping series in Example 4.5. We have

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = 1 + \sum_{n=1}^{\infty} \frac{1}{(n+1)^2}$$

and

$$0 \le \frac{1}{(n+1)^2} < \frac{1}{n(n+1)}.$$

We also get the explicit upper bound

$$\sum_{n=1}^{\infty} \frac{1}{n^2} < 1 + \sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 2.$$

In fact, the sum is

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

Mengoli (1644) posed the problem of finding this sum, which was solved by Euler (1735). The evaluation of the sum is known as the Basel problem, perhaps after Euler's birthplace in Switzerland.

Example 4.21. The series in Example 4.20 is a special case of the following series, called the *p*-series,

$$\sum_{n=1}^{\infty} \frac{1}{n^p},$$

where $0 < p < \infty$. It follows by comparison with the harmonic series in Example 4.11 that the *p*-series diverges for $p \le 1$. (If it converged for some $p \le 1$, then the harmonic series would also converge.) On the other hand, the *p*-series converges for every $1 < p < \infty$. To show this, note that

$$\frac{1}{2^p} + \frac{1}{3^p} < \frac{2}{2^p}, \qquad \frac{1}{4^p} + \frac{1}{5^p} + \frac{1}{6^p} + \frac{1}{7^p} < \frac{4}{4^p},$$

and so on, which implies that

$$\sum_{n=1}^{2^N-1} \frac{1}{n^p} < 1 + \frac{1}{2^{p-1}} + \frac{1}{4^{p-1}} + \frac{1}{8^{p-1}} + \cdots + \frac{1}{2^{(N-1)(p-1)}} < \frac{1}{1 - 2^{1-p}}.$$

Thus, the partial sums are bounded from above, so the series converges by Proposition 4.6. An alternative proof of the convergence of the *p*-series for $p > 1$ and divergence for $0 < p \le 1$, using the integral test, is given in Example 12.44.

4.5. * The Riemann ζ-function

Example 4.21 justifies the following definition.

Definition 4.22. The Riemann ζ-function is defined for $1 < s < \infty$ by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

For instance, as stated in Example 4.20, we have $\zeta(2) = \pi^2/6$. In fact, Euler (1755) discovered a general formula for the value $\zeta(2n)$ of the ζ-function at even natural numbers,

$$\zeta(2n) = (-1)^{n+1} \frac{(2\pi)^{2n} B_{2n}}{2(2n)!}, \qquad n = 1, 2, 3, \ldots,$$

where the coefficients B_{2n} are the Bernoulli numbers (see Example 10.19). In particular,

$$\zeta(4) = \frac{\pi^4}{90}, \quad \zeta(6) = \frac{\pi^6}{945}, \quad \zeta(8) = \frac{\pi^8}{9450}, \quad \zeta(10) = \frac{\pi^{10}}{93555}.$$

On the other hand, the values of the ζ-function at odd natural numbers are harder to study. For instance,

$$\zeta(3) = \sum_{n=1}^{\infty} \frac{1}{n^3} = 1.2020569\ldots$$

is called Apéry's constant. It was proved to be irrational by Apéry (1979) but a simple explicit expression for $\zeta(3)$ is not known (and likely doesn't exist).

The Riemann ζ-function is intimately connected with number theory and the distribution of primes. Every positive integer n has a unique factorization

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \ldots p_k^{\alpha_k},$$

where the p_j are primes and the exponents α_j are positive integers. Using the binomial expansion in Example 4.2, we have

$$\left(1 - \frac{1}{p^s}\right)^{-1} = 1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \frac{1}{p^{3s}} + \frac{1}{p^{4s}} + \ldots.$$

By expanding the products and rearranging the resulting sums, one can see that

$$\zeta(s) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1},$$

where the product is taken over all primes p, since every possible prime factorization of a positive integer appears exactly once in the sum on the right-hand side. The infinite product here is defined as a limit of finite products,

$$\prod_p \left(1 - \frac{1}{p^s}\right)^{-1} = \lim_{N \to \infty} \prod_{p \leq N} \left(1 - \frac{1}{p^s}\right)^{-1}.$$

Using complex analysis, one can show that the ζ-function may be extended in a unique way to an analytic (i.e., differentiable) function of a complex variable $s = \sigma + it \in \mathbb{C}$

$$\zeta : \mathbb{C} \setminus \{1\} \to \mathbb{C},$$

where $\sigma = \Re s$ is the real part of s and $t = \Im s$ is the imaginary part. The ζ-function has a singularity at $s = 1$, called a simple pole, where it goes to infinity like $1/(1-s)$, and is equal to zero at the negative even integers $s = -2, -4, \ldots, -2n, \ldots$. These zeros are called the trivial zeros of the ζ-function. Riemann (1859) made the following conjecture.

Hypothesis 4.23 (Riemann hypothesis)**.** Except for the trivial zeros, the only zeros of the Riemann ζ-function occur on the line $\Re s = 1/2$.

If true, the Riemann hypothesis has significant consequences for the distribution of primes (and many other things); roughly speaking, it implies that the prime numbers are "randomly distributed" among the natural numbers (with density $1/\log n$ near a large integer $n \in \mathbb{N}$). Despite enormous efforts, this conjecture has neither been proved nor disproved, and it remains one of the most significant open problems in mathematics (perhaps *the* most significant open problem).

4.6. The ratio and root tests

In this section, we describe the ratio and root tests, which provide explicit sufficient conditions for the absolute convergence of a series that can be compared with a geometric series. These tests are particularly useful in studying power series, but they aren't effective in determining the convergence or divergence of series whose terms do not approach zero at a geometric rate.

Theorem 4.24 (Ratio test)**.** Suppose that (a_n) is a sequence of nonzero real numbers such that the limit

$$r = \lim_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right|$$

exists or diverges to infinity. Then the series

$$\sum_{n=1}^{\infty} a_n$$

converges absolutely if $0 \le r < 1$ and diverges if $1 < r \le \infty$.

Proof. If $r < 1$, choose s such that $r < s < 1$. Then there exists $N \in \mathbb{N}$ such that

$$\left| \frac{a_{n+1}}{a_n} \right| < s \qquad \text{for all } n > N.$$

It follows that

$$|a_n| \le M s^n \qquad \text{for all } n > N$$

where M is a suitable constant. Therefore $\sum a_n$ converges absolutely by comparison with the convergent geometric series $\sum M s^n$.

If $r > 1$, choose s such that $r > s > 1$. There exists $N \in \mathbb{N}$ such that

$$\left| \frac{a_{n+1}}{a_n} \right| > s \qquad \text{for all } n > N,$$

so that $|a_n| \geq M s^n$ for all $n > N$ and some $M > 0$. It follows that (a_n) does not approach 0 as $n \to \infty$, so the series diverges. $\square$

Example 4.25. Let $a \in \mathbb{R}$, and consider the series

$$\sum_{n=1}^{\infty} na^n = a + 2a^2 + 3a^3 + \dots.$$

Then

$$\lim_{n \to \infty} \left| \frac{(n+1)a^{n+1}}{na^n} \right| = |a| \lim_{n \to \infty} \left(1 + \frac{1}{n} \right) = |a|.$$

By the ratio test, the series converges if $|a| < 1$ and diverges if $|a| > 1$; the series also diverges if $|a| = 1$. The convergence of the series for $|a| < 1$ is explained by the fact that the geometric decay of the factor a^n is more rapid than the algebraic growth of the coefficient n.

Example 4.26. Let $p > 0$ and consider the p-series

$$\sum_{n=1}^{\infty} \frac{1}{n^p}.$$

Then

$$\lim_{n \to \infty} \left[\frac{1/(n+1)^p}{1/n^p} \right] = \lim_{n \to \infty} \left[\frac{1}{(1+1/n)^p} \right] = 1,$$

so the ratio test is inconclusive. In this case, the series diverges if $0 < p \leq 1$ and converges if $p > 1$, which shows that either possibility may occur when the limit in the ratio test is 1.

The root test provides a criterion for convergence of a series that is closely related to the ratio test, but it doesn't require that the limit of the ratios of successive terms exists.

Theorem 4.27 (Root test). Suppose that (a_n) is a sequence of real numbers and let

$$r = \limsup_{n \to \infty} |a_n|^{1/n}.$$

Then the series

$$\sum_{n=1}^{\infty} a_n$$

converges absolutely if $0 \leq r < 1$ and diverges if $1 < r \leq \infty$.

Proof. First suppose $0 \leq r < 1$. If $0 < r < 1$, choose s such that $r < s < 1$, and let

$$t = \frac{r}{s}, \qquad r < t < 1.$$

If $r = 0$, choose any $0 < t < 1$. Since $t > \limsup |a_n|^{1/n}$, Theorem 3.41 implies that there exists $N \in \mathbb{N}$ such that

$$|a_n|^{1/n} < t \qquad \text{for all } n > N.$$

Therefore $|a_n| < t^n$ for all $n > N$, where $t < 1$, so it follows that the series converges by comparison with the convergent geometric series $\sum t^n$.

Next suppose $1 < r \le \infty$. If $1 < r < \infty$, choose s such that $1 < s < r$, and let

$$t = \frac{r}{s}, \qquad 1 < t < r.$$

If $r = \infty$, choose any $1 < t < \infty$. Since $t < \limsup |a_n|^{1/n}$, Theorem 3.41 implies that

$$|a_n|^{1/n} > t \qquad \text{for infinitely many } n \in \mathbb{N}.$$

Therefore $|a_n| > t^n$ for infinitely many $n \in \mathbb{N}$, where $t > 1$, so (a_n) does not approach zero as $n \to \infty$, and the series diverges. □

The root test may succeed where the ratio test fails.

Example 4.28. Consider the geometric series with ratio $1/2$,

$$\sum_{n=1}^{\infty} a_n = \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \frac{1}{2^5} + \dots, \qquad a_n = \frac{1}{2^n}.$$

Then (of course) both the ratio and root test imply convergence since

$$\lim_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right| = \limsup_{n \to \infty} |a_n|^{1/n} = \frac{1}{2} < 1.$$

Now consider the series obtained by switching successive odd and even terms

$$\sum_{n=1}^{\infty} b_n = \frac{1}{2^2} + \frac{1}{2} + \frac{1}{2^4} + \frac{1}{2^3} + \frac{1}{2^6} + \dots, \qquad b_n = \begin{cases} 1/2^{n+1} & \text{if } n \text{ is odd,} \\ 1/2^{n-1} & \text{if } n \text{ is even} \end{cases}$$

For this series,

$$\left| \frac{b_{n+1}}{b_n} \right| = \begin{cases} 2 & \text{if } n \text{ is odd,} \\ 1/8 & \text{if } n \text{ is even,} \end{cases}$$

and the ratio test doesn't apply, since the required limit does not exist. (The series still converges at a geometric rate, however, because the the decrease in the terms by a factor of $1/8$ for even n dominates the increase by a factor of 2 for odd n.) On the other hand

$$\limsup_{n \to \infty} |b_n|^{1/n} = \frac{1}{2},$$

so the ratio test still works. In fact, as we discuss in Section 4.8, since the series is absolutely convergent, every rearrangement of it converges to the same sum.

4.7. Alternating series

An alternating series is one in which successive terms have opposite signs. If the terms in an alternating series have decreasing absolute values and converge to zero, then the series converges however slowly its terms approach zero. This allows us to prove the convergence of some series which aren't absolutely convergent.

Example 4.29. The alternating harmonic series from Example 4.14 is

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \dots.$$

The behavior of its partial sums is shown in Figure 1, which illustrates the idea of the convergence proof for alternating series.
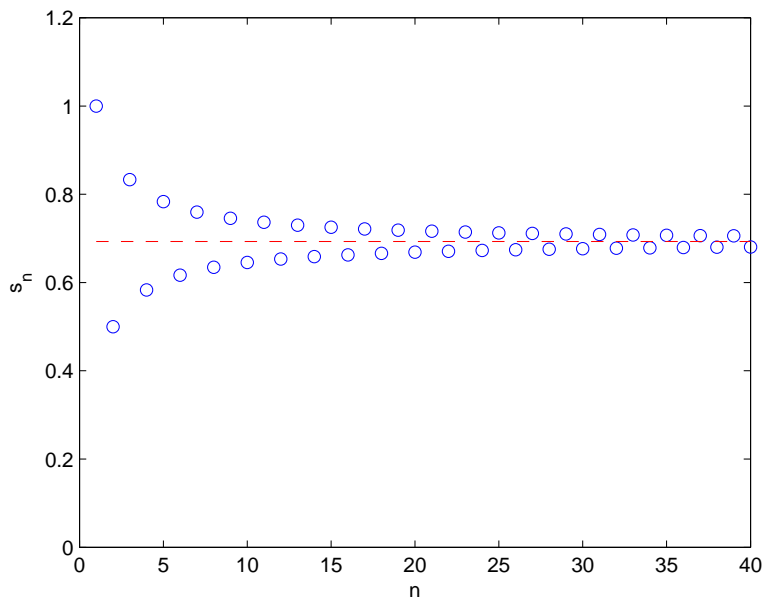
Figure 1. A plot of the first 40 partial sums S_n of the alternating harmonic series in Example 4.14. The odd partial sums decrease and the even partial sums increase to the sum of the series $\log 2 \approx 0.6931$, which is indicated by the dashed line.

Theorem 4.30 (Alternating series). Suppose that (a_n) is a decreasing sequence of nonnegative real numbers, meaning that $0 \leq a_{n+1} \leq a_n$, such that $a_n \to 0$ as $n \to \infty$. Then the alternating series

$$\sum_{n=1}^{\infty}(-1)^{n+1}a_n = a_1 - a_2 + a_3 - a_4 + a_5 - \dots$$

converges.

Proof. Let

$$S_n = \sum_{k=1}^{n}(-1)^{k+1}a_k$$

denote the nth partial sum. If $n = 2m - 1$ is odd, then

$$S_{2m-1} = S_{2m-3} - a_{2m-2} + a_{2m-1} \leq S_{2m-3},$$

since (a_n) is decreasing, and

$$S_{2m-1} = (a_1 - a_2) + (a_3 - a_4) + \dots + (a_{2m-3} - a_{2m-2}) + a_{2m-1} \geq 0.$$

Thus, the sequence (S_{2m-1}) of odd partial sums is decreasing and bounded from below by 0, so $S_{2m-1} \downarrow S^+$ as $m \to \infty$ for some $S^+ \geq 0$.

Similarly, if $n = 2m$ is even, then

$$S_{2m} = S_{2m-2} + a_{2m-1} - a_{2m} \geq S_{2m-2},$$

and
$$S_{2m} = a_1 - (a_2 - a_3) - (a_4 - a_5) - \cdots - (a_{2m-1} - a_{2m}) \le a_1.$$
Thus, (S_{2m}) is increasing and bounded from above by a_1, so $S_{2m} \uparrow S^- \le a_1$ as $m \to \infty$.

Finally, note that
$$\lim_{m \to \infty} (S_{2m-1} - S_{2m}) = \lim_{m \to \infty} a_{2m} = 0,$$
so $S^+ = S^-$, which implies that the series converges to their common value. $\qquad \square$

The proof also shows that the sum $S_{2m} \le S \le S_{2n-1}$ is bounded from below and above by all even and odd partial sums, respectively, and that the error $|S_n - S|$ is less than the first term a_{n+1} in the series that is neglected.

Example 4.31. The alternating p-series
$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^p}$$
converges for every $p > 0$. The convergence is absolute for $p > 1$ and conditional for $0 < p \le 1$.

4.8. Rearrangements

A rearrangement of a series is a series that consists of the same terms in a different order. The convergence of rearranged series may initially appear to be unconnected with absolute convergence, but absolutely convergent series are exactly those series whose sums remain the same under every rearrangement of their terms. On the other hand, a conditionally convergent series can be rearranged to give any sum we please, or to diverge.

Example 4.32. A rearrangement of the alternating harmonic series in Example 4.14 is
$$1 - \frac{1}{2} - \frac{1}{4} + \frac{1}{3} - \frac{1}{6} - \frac{1}{8} + \frac{1}{5} - \frac{1}{10} - \frac{1}{12} + \dots,$$
where we put two negative even terms between each of the positive odd terms. The behavior of its partial sums is shown in Figure 2. As proved in Example 12.47, this series converges to one-half of the sum of the alternating harmonic series. The sum of the alternating harmonic series can change under rearrangement because it is conditionally convergent.

Note also that both the positive and negative parts of the alternating harmonic series diverge to infinity, since
$$1 + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \dots > \frac{1}{2} + \frac{1}{4} + \frac{1}{6} + \frac{1}{8} + \dots$$
$$> \frac{1}{2} \left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots \right),$$
$$\frac{1}{2} + \frac{1}{4} + \frac{1}{6} + \frac{1}{8} + \dots = \frac{1}{2} \left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots \right),$$
and the harmonic series diverges. This is what allows us to change the sum by rearranging the series.
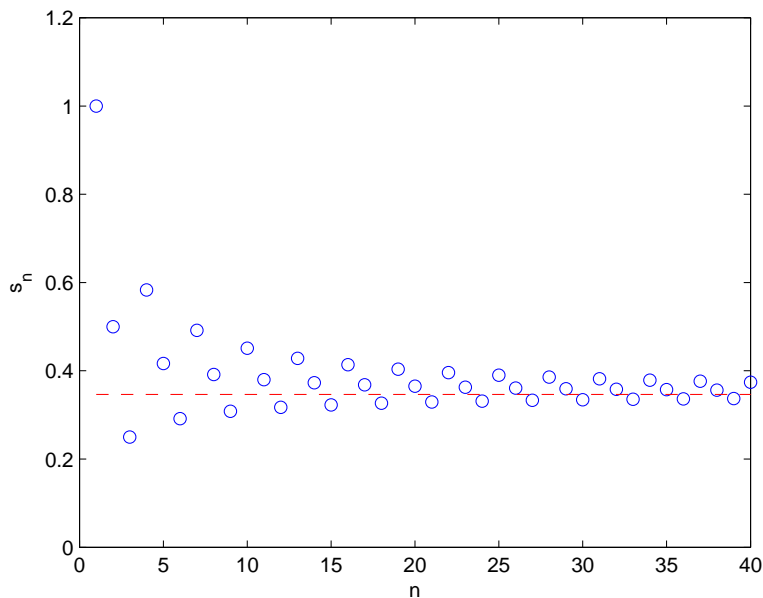
Figure 2. A plot of the first 40 partial sums S_n of the rearranged alternating harmonic series in Example 4.32. The series converges to half the sum of the alternating harmonic series, $\frac{1}{2} \log 2 \approx 0.3466$. Compare this picture with Figure 1.

The formal definition of a rearrangement is as follows.

Definition 4.33. A series

$$\sum_{m=1}^{\infty} b_m$$

is a rearrangement of a series

$$\sum_{n=1}^{\infty} a_n$$

if there is a one-to-one, onto function $f : \mathbb{N} \to \mathbb{N}$ such that $b_m = a_{f(m)}$.

If $\sum b_m$ is a rearrangement of $\sum a_n$ with $n = f(m)$, then $\sum a_n$ is a rearrangement of $\sum b_m$, with $m = f^{-1}(n)$.

Theorem 4.34. If a series is absolutely convergent, then every rearrangement of the series converges to the same sum.

Proof. First, suppose that

$$\sum_{n=1}^{\infty} a_n$$

is a convergent series with $a_n \geq 0$, and let

$$\sum_{m=1}^{\infty} b_m, \qquad b_m = a_{f(m)}$$

be a rearrangement.

Given $\epsilon > 0$, choose $N \in \mathbb{N}$ such that

$$0 \leq \sum_{k=1}^{\infty} a_k - \sum_{k=1}^{N} a_k < \epsilon.$$

Since $f : \mathbb{N} \to \mathbb{N}$ is one-to-one and onto, there exists $M \in \mathbb{N}$ such that

$$\{1, 2, \ldots, N\} \subset f^{-1}\left(\{1, 2, \ldots, M\}\right),$$

meaning that all of the terms $a_1, a_2, \ldots, a_N$ are included among the $b_1, b_2, \ldots, b_M$. For example, we can take $M = \max\{m \in \mathbb{N} : 1 \leq f(m) \leq N\}$; this maximum is well-defined since there are finitely many such m (in fact, N of them).

If $m > M$, then

$$\sum_{k=1}^{N} a_k \leq \sum_{j=1}^{m} b_j \leq \sum_{k=1}^{\infty} a_k$$

since the b_j's include all the a_k's in the left sum, all the b_j's are included among the a_k's in the right sum, and $a_k, b_j \geq 0$. It follows that

$$0 \leq \sum_{k=1}^{\infty} a_k - \sum_{j=1}^{m} b_j < \epsilon,$$

for all $m > M$, which proves that

$$\sum_{j=1}^{\infty} b_j = \sum_{k=1}^{\infty} a_k.$$

If $\sum a_n$ is a general absolutely convergent series, then from Proposition 4.17 the positive and negative parts of the series

$$\sum_{n=1}^{\infty} a_n^+, \qquad \sum_{n=1}^{\infty} a_n^-$$

converge. If $\sum b_m$ is a rearrangement of $\sum a_n$, then $\sum b_m^+$ and $\sum b_m^-$ are rearrangements of $\sum a_n^+$ and $\sum a_n^-$, respectively. It follows from what we've just proved that they converge and

$$\sum_{m=1}^{\infty} b_m^+ = \sum_{n=1}^{\infty} a_n^+, \qquad \sum_{m=1}^{\infty} b_m^- = \sum_{n=1}^{\infty} a_n^-.$$

Proposition 4.17 then implies that $\sum b_m$ is absolutely convergent and

$$\sum_{m=1}^{\infty} b_m = \sum_{m=1}^{\infty} b_m^+ - \sum_{m=1}^{\infty} b_m^- = \sum_{n=1}^{\infty} a_n^+ - \sum_{n=1}^{\infty} a_n^- = \sum_{n=1}^{\infty} a_n,$$

which proves the result. □
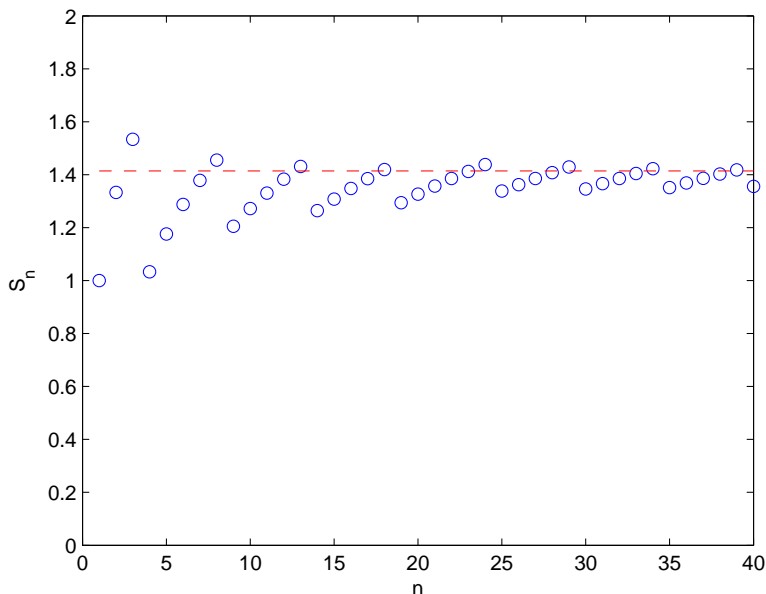
Figure 3. A plot of the first 40 partial sums S_n of the rearranged alternating harmonic series described in Example 4.35, which converges to $\sqrt{2}$.

Conditionally convergent series behave completely differently from absolutely convergent series under rearrangement. As Riemann observed, they can be rearranged to give any sum we want, or to diverge. Before giving the proof, we illustrate the idea with an example.

Example 4.35. Suppose we want to rearrange the alternating harmonic series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \dots.$$

so that its sum is $\sqrt{2} \approx 1.4142$. We choose positive terms until we get a partial sum that is greater than $\sqrt{2}$, which gives $1 + 1/3 + 1/5$; followed by negative terms until we get a sum less than $\sqrt{2}$, which gives $1 + 1/3 + 1/5 - 1/2$; followed by positive terms until we get a sum greater than $\sqrt{2}$, which gives

$$1 + \frac{1}{3} + \frac{1}{5} - \frac{1}{2} + \frac{1}{7} + \frac{1}{9} + \frac{1}{11} + \frac{1}{13};$$

followed by another negative term $-1/4$ to get a sum less than $\sqrt{2}$; and so on. The first 40 partial sums of the resulting series are shown in Figure 3.

Theorem 4.36. If a series is conditionally convergent, then it has rearrangements that converge to an arbitrary real number and rearrangements that diverge to ∞ or $-\infty$.

Proof. Suppose that $\sum a_n$ is conditionally convergent. Since the series converges, $a_n \to 0$ as $n \to \infty$. If both the positive part $\sum a_n^+$ and negative part $\sum a_n^-$ of the

series converge, then the series converges absolutely; and if only one part diverges, then the series diverges (to ∞ if $\sum a_n^+$ diverges, or $-\infty$ if $\sum a_n^-$ diverges). Therefore both $\sum a_n^+$ and $\sum a_n^-$ diverge. This means that we can make sums of successive positive or negative terms in the series as large as we wish.

Suppose $S \in \mathbb{R}$. Starting from the beginning of the series, we choose successive positive or zero terms in the series until their partial sum is greater than or equal to S. Then we choose successive strictly negative terms, starting again from the beginning of the series, until the partial sum of all the terms is strictly less than S. After that, we choose successive positive or zero terms until the partial sum is greater than or equal S, followed by negative terms until the partial sum is strictly less than S, and so on. The partial sums are greater than S by at most the value of the last positive term retained, and are less than S by at most the value of the last negative term retained. Since $a_n \to 0$ as $n \to \infty$, it follows that the rearranged series converges to S.

A similar argument shows that we can rearrange a conditional convergent series to diverge to ∞ or $-\infty$, and that we can rearrange the series so that it diverges in a finite or infinite oscillatory fashion. $\qquad\square$

The previous results indicate that conditionally convergent series behave in many ways more like divergent series than absolutely convergent series.

4.9. The Cauchy product

In this section, we prove a result about the product of absolutely convergent series that is useful in multiplying power series. It is convenient to begin numbering the terms of the series at $n = 0$.

Definition 4.37. The Cauchy product of the series

$$\sum_{n=0}^{\infty} a_n, \qquad \sum_{n=0}^{\infty} b_n$$

is the series

$$\sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} a_k b_{n-k} \right).$$

The Cauchy product arises formally by term-by-term multiplication and rearrangement:

$$(a_0 + a_1 + a_2 + a_3 + \dots)(b_0 + b_1 + b_2 + b_3 + \dots)$$
$$= a_0 b_0 + a_0 b_1 + a_0 b_2 + a_0 b_3 + \dots + a_1 b_0 + a_1 b_1 + a_1 b_2 + \dots$$
$$+ a_2 b_0 + a_2 b_1 + \dots + a_3 b_0 + \dots$$
$$= a_0 b_0 + (a_0 b_1 + a_1 b_0) + (a_0 b_2 + a_1 b_1 + a_2 b_0)$$
$$+ (a_0 b_3 + a_1 b_2 + a_2 b_1 + a_3 b_0) + \dots.$$

In general, writing $m = n - k$, we have formally that

$$\left(\sum_{n=0}^{\infty} a_n \right) \left(\sum_{n=0}^{\infty} b_n \right) = \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} a_k b_m = \sum_{n=0}^{\infty} \sum_{k=0}^{n} a_k b_{n-k}.$$

There are no convergence issues about the individual terms in the Cauchy product, since $\sum_{k=0}^{n} a_k b_{n-k}$ is a finite sum.

Theorem 4.38 (Cauchy product). If the series

$$\sum_{n=0}^{\infty} a_n, \qquad \sum_{n=0}^{\infty} b_n$$

are absolutely convergent, then the Cauchy product is absolutely convergent and

$$\sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} a_k b_{n-k} \right) = \left(\sum_{n=0}^{\infty} a_n \right) \left(\sum_{n=0}^{\infty} b_n \right).$$

Proof. For every $N \in \mathbb{N}$, we have

$$\sum_{n=0}^{N} \left| \sum_{k=0}^{n} a_k b_{n-k} \right| \leq \sum_{n=0}^{N} \left(\sum_{k=0}^{n} |a_k||b_{n-k}| \right)$$

$$\leq \left(\sum_{k=0}^{N} |a_k| \right) \left(\sum_{m=0}^{N} |b_m| \right)$$

$$\leq \left(\sum_{n=0}^{\infty} |a_n| \right) \left(\sum_{n=0}^{\infty} |b_n| \right).$$

Thus, the Cauchy product is absolutely convergent, since the partial sums of its absolute values are bounded from above.

Since the series for the Cauchy product is absolutely convergent, any rearrangement of it converges to the same sum. In particular, the subsequence of partial sums given by

$$\left(\sum_{n=0}^{N} a_n \right) \left(\sum_{n=0}^{N} b_n \right) = \sum_{n=0}^{N} \sum_{m=0}^{N} a_n b_m$$

corresponds to a rearrangement of the Cauchy product, so

$$\sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} a_k b_{n-k} \right) = \lim_{N \to \infty} \left(\sum_{n=0}^{N} a_n \right) \left(\sum_{n=0}^{N} b_n \right) = \left(\sum_{n=0}^{\infty} a_n \right) \left(\sum_{n=0}^{\infty} b_n \right).$$

$\square$

In fact, as we discuss in the next section, since the series of term-by-term products of absolutely convergent series converges absolutely, every rearrangement of the product series — not just the one in the Cauchy product — converges to the product of the sums.

4.10. * Double series

A double series is a series of the form

$$\sum_{m,n=1}^{\infty} a_{mn},$$

where the terms a_{mn} are indexed by a pair of natural numbers $m, n \in \mathbb{R}$. More formally, the terms are defined by a function $f : \mathbb{N} \times \mathbb{N} \to \mathbb{R}$ where $a_{mn} = f(m, n)$. In this section, we consider sums of double series; this material is not used later on.

There are many different ways to define the sum of a double series since, unlike the index set $\mathbb{N}$ of a series, the index set $\mathbb{N} \times \mathbb{N}$ of a double series does not come with a natural linear order. Our main interest here is in giving a definition of the sum of a double series that does not depend on the order of its terms. As for series, this unordered sum exists if and only if the double series is absolutely convergent.

If $F \subset \mathbb{N} \times \mathbb{N}$ is a finite subset of pairs of natural numbers, then we denote by

$$\sum_F a_{mn} = \sum_{(m,n) \in F} a_{mn}$$

the partial sum of all terms a_{mn} whose indices (m, n) belong to F.

Definition 4.39. The unordered sum of nonnegative real numbers $a_{mn} \geq 0$ is

$$\sum_{\mathbb{N} \times \mathbb{N}} a_{mn} = \sup_{F \in \mathcal{F}} \left\{ \sum_F a_{mn} \right\}$$

where the supremum is taken over the collection $\mathcal{F}$ of all finite subsets $F \subset \mathbb{N} \times \mathbb{N}$. The unordered sum converges if this supremum is finite and diverges to ∞ if this supremum is ∞.

In other words, the unordered sum of a double series of nonnegative terms is the supremum of the set of all finite partial sums of the series. Note that this supremum exists if and only if the finite partial sums of the series are bounded from above.

Example 4.40. The unordered sum

$$\sum_{(m,n) \in \mathbb{N} \times \mathbb{N}} \frac{1}{(m+n)^p} = \frac{1}{(1+1)^p} + \frac{1}{(1+2)^p} + \frac{1}{(2+1)^p} + \frac{1}{(1+3)^p} + \frac{1}{(2+2)^p} + \dots$$

converges if $p > 2$ and diverges if $p \leq 2$. To see this, first note that if

$$T = \{(m, n) \in \mathbb{N} \times \mathbb{N} : 2 \leq m + n \leq N\}$$

is a "triangular" set of indices, then

$$\sum_T \frac{1}{(m+n)^p} = \sum_{m=1}^{N-1} \sum_{n=1}^{N-m} \frac{1}{(m+n)^p}$$

$$= \sum_{k=2}^{N} \sum_{n=1}^{k} \frac{1}{k^p}$$

$$= \sum_{k=2}^{N} \frac{1}{k^{p-1}}.$$

It follows that

$$\sum_{\mathbb{N} \times \mathbb{N}} \frac{1}{(m+n)^p} \geq \sum_{k=2}^{N} \frac{1}{k^{p-1}}$$

for every $N \in \mathbb{N}$, so the double series diverges if $p \leq 2$ since the $(p-1)$-series diverges. Moreover, if $p > 2$ and F is a finite subset of $\mathbb{N} \times \mathbb{N}$, then there exists a triangular set T such that $F \subset T$, so that

$$\sum_F \frac{1}{(m+n)^p} \leq \sum_T \frac{1}{(m+n)^p} < \sum_{k=2}^{\infty} \frac{1}{k^{p-1}}.$$

It follows that the unordered sum converges if $p > 2$, with

$$\sum_{\mathbb{N} \times \mathbb{N}} \frac{1}{(m+n)^p} = \sum_{k=2}^{\infty} \frac{1}{k^{p-1}}.$$

Note that this double p-series converges only if its terms approach zero at a faster rate than the terms in a single p-series (of degree greater than 2 in (m,n) rather than degree greater than 1 in n).

We define a general unordered sum of real numbers by summing its positive and negative terms separately. (Recall the notation in Definition 4.15 for the positive and negative parts of a real number.)

Definition 4.41. The unordered sum of a double series of real numbers is

$$\sum_{\mathbb{N} \times \mathbb{N}} a_{mn} = \sum_{\mathbb{N} \times \mathbb{N}} a_{mn}^+ - \sum_{\mathbb{N} \times \mathbb{N}} a_{mn}^-$$

where $a_{mn} = a_{mn}^+ - a_{mn}^-$ is the decomposition of a_{mn} into its positive and negative parts. The unordered sum converges if both $\sum a_{mn}^+$ and $\sum a_{mn}^-$ are finite; diverges to ∞ if $\sum a_{mn}^+ = \infty$ and $\sum a_{mn}^-$ is finite; diverges to $-\infty$ if $\sum a_{mn}^+$ is finite and $\sum a_{mn}^- = \infty$; and is undefined if both $\sum a_{mn}^+$ and $\sum a_{mn}^-$ diverge to ∞.

This definition does not require us to order the index set $\mathbb{N} \times \mathbb{N}$ in any way; in fact, the same definition applies to any series of real numbers

$$\sum_{i \in I} a_i$$

whose terms are indexed by an arbitrary set I. A sum over a set of indices will always denote an unordered sum.

Definition 4.42. A double series

$$\sum_{m,n=1}^{\infty} a_{mn}$$

of real numbers converges absolutely if the unordered sum

$$\sum_{\mathbb{N} \times \mathbb{N}} |a_{mn}| < \infty$$

converges.

The following result is a straightforward consequence of the definitions and the completeness of $\mathbb{R}$.

Proposition 4.43. An unordered sum $\sum a_{mn}$ of real numbers converges if and only if it converges absolutely, and in that case

$$\sum_{\mathbb{N} \times \mathbb{N}} |a_{mn}| = \sum_{\mathbb{N} \times \mathbb{N}} a_{mn}^+ + \sum_{\mathbb{N} \times \mathbb{N}} a_{mn}^-.$$

Proof. First, suppose that the unordered sum

$$\sum_{\mathbb{N} \times \mathbb{N}} a_{mn}$$

converges absolutely. If $F \subset \mathbb{N} \times \mathbb{N}$ is any finite subset, then

$$\sum_{F} a_{mn}^\pm \leq \sum_{F} |a_{mn}| \leq \sum_{\mathbb{N} \times \mathbb{N}} |a_{mn}|.$$

It follows that the finite partial sums of the positive and negative terms of the series are bounded from above, so the unordered sum converges.

Conversely, suppose that the unordered sum converges. If $F \subset \mathbb{N} \times \mathbb{N}$ is any finite subset, then

$$\sum_{F} |a_{mn}| = \sum_{F} a_{mn}^+ + \sum_{F} a_{mn}^- \leq \sum_{\mathbb{N} \times \mathbb{N}} a_{mn}^+ + \sum_{\mathbb{N} \times \mathbb{N}} a_{mn}^-.$$

It follows that the finite partial sums of the absolute values of the terms in the series are bounded from above, so the unordered sum converges absolutely. Furthermore,

$$\sum_{\mathbb{N} \times \mathbb{N}} |a_{mn}| \leq \sum_{\mathbb{N} \times \mathbb{N}} a_{mn}^+ + \sum_{\mathbb{N} \times \mathbb{N}} a_{mn}^-.$$

To prove the inequality in the other direction, let $\epsilon > 0$. There exist finite sets $F_+, F_- \subset \mathbb{N} \times \mathbb{N}$ such that

$$\sum_{F_+} a_{mn}^+ > \sum_{\mathbb{N} \times \mathbb{N}} a_{mn}^+ - \frac{\epsilon}{2}, \qquad \sum_{F_-} a_{mn}^- > \sum_{\mathbb{N} \times \mathbb{N}} a_{mn}^- - \frac{\epsilon}{2}.$$

Let $F = F_+ \cup F_-$. Then, since $a_{mn}^\pm \geq 0$, we have

$$\sum_{F} |a_{mn}| = \sum_{F} a_{mn}^+ + \sum_{F} a_{mn}^-$$

$$\geq \sum_{F_+} a_{mn}^+ + \sum_{F_-} a_{mn}^-$$

$$> \sum_{\mathbb{N} \times \mathbb{N}} a_{mn}^+ + \sum_{\mathbb{N} \times \mathbb{N}} a_{mn}^- - \epsilon.$$

Since $\epsilon > 0$ is arbitrary, it follows that

$$\sum_{\mathbb{N} \times \mathbb{N}} |a_{mn}| \geq \sum_{\mathbb{N} \times \mathbb{N}} a_{mn}^+ + \sum_{\mathbb{N} \times \mathbb{N}} a_{mn}^-,$$

which completes the proof. $\qquad \square$

Next, we define rearrangements of a double series into single series and show that every rearrangement of a convergent unordered sum converges to the same sum.

Definition 4.44. A rearrangement of a double series

$$\sum_{m,n=1}^{\infty} a_{mn}$$

is a series of the form

$$\sum_{k=1}^{\infty} b_k, \qquad b_k = a_{\sigma(k)}$$

where $\sigma : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ is a one-to-one, onto map.

Example 4.45. The rearrangement corresponding to the map $f : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ defined in the proof of Proposition 1.45 is given by

$$\sum_{m,n=1}^{\infty} a_{mn} = a_{11} + a_{21} + a_{12} + a_{31} + a_{22} + a_{13} + a_{41} + a_{32} + a_{23} + a_{14} + \ldots.$$

Theorem 4.46. If the unordered sum of a double series of real numbers converges, then every rearrangement of the double series into a single series converges to the unordered sum.

Proof. Suppose that the unordered sum $\sum a_{mn}$ converges with

$$\sum_{\mathbb{N} \times \mathbb{N}} a_{mn}^+ = S_+, \qquad \sum_{\mathbb{N} \times \mathbb{N}} a_{mn}^- = S_-,$$

and let

$$\sum_{k=1}^{\infty} b_k$$

be a rearrangement of the double series corresponding to a map $\sigma : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$. For $k \in \mathbb{N}$, let

$$F_k = \{(m,n) \in \mathbb{N} \times \mathbb{N} : (m,n) = \sigma(j) \text{ for some } 1 \le j \le k\},$$

so that

$$\sum_{j=1}^{k} b_j = \sum_{F_k} a_{mn} = \sum_{F_k} a_{mn}^+ - \sum_{F_k} a_{mn}^-.$$

Given $\epsilon > 0$, choose finite sets

$$F_+, F_- \subset \mathbb{N} \times \mathbb{N}$$

such that

$$S_+ - \epsilon < \sum_{F_+} a_{mn}^+ \le S_+, \qquad S_- - \epsilon < \sum_{F_-} a_{mn}^- \le S_-,$$

and define $N \in \mathbb{N}$ by

$$N = \max\{j \in \mathbb{N} : \sigma(j) \in F_+ \cup F_-\}.$$

If $k \ge N$, then $F_k \supset F_+ \cup F_-$ and, since $a_{mn}^{\pm} \ge 0$,

$$\sum_{F_+} a_{mn}^+ \le \sum_{F_k} a_{mn}^+ \le S_+, \qquad \sum_{F_-} a_{mn}^- \le \sum_{F_k} a_{mn}^- \le S_-.$$

It follows that

$$S_+ - \epsilon < \sum_{F_k} a_{mn}^+ \leq S_+, \qquad S_- - \epsilon < \sum_{F_k} a_{mn}^- \leq S_-,$$

which implies that

$$\left| \sum_{j=1}^{k} b_j - (S_+ - S_-) \right| < \epsilon.$$

This inequality proves that the rearrangement $\sum b_k$ converges to the unordered sum $S_+ - S_-$ of the double series. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

The rearrangement of a double series into a single series is one natural way to interpret a double series in terms of single series. Another way is to use iterated sums of single series.

Given a double series $\sum a_{mn}$, one can define two iterated sums, obtained by summing first over one index followed by the other:

$$\sum_{m=1}^{\infty} \left(\sum_{n=1}^{\infty} a_{mn} \right) = \lim_{M \to \infty} \sum_{m=1}^{M} \left(\lim_{N \to \infty} \sum_{n=1}^{N} a_{mn} \right);$$

$$\sum_{n=1}^{\infty} \left(\sum_{m=1}^{\infty} a_{mn} \right) = \lim_{N \to \infty} \sum_{n=1}^{N} \left(\lim_{M \to \infty} \sum_{m=1}^{M} a_{mn} \right).$$

As the following example shows, these iterated sums may not equal, even if both of them converge.

Example 4.47. Define a_{mn} by

$$a_{mn} = \begin{cases} 1 & \text{if } n = m + 1 \\ -1 & \text{if } m = n + 1 \\ 0 & \text{otherwise} \end{cases}$$

Then, by writing out the terms a_{mn} in a table, one can see that

$$\sum_{n=1}^{\infty} a_{mn} = \begin{cases} 1 & \text{if } m = 1 \\ 0 & \text{otherwise} \end{cases}, \qquad \sum_{m=1}^{\infty} a_{mn} = \begin{cases} -1 & \text{if } n = 1 \\ 0 & \text{otherwise} \end{cases},$$

so that

$$\sum_{m=1}^{\infty} \left(\sum_{n=1}^{\infty} a_{mn} \right) = 1, \qquad \sum_{n=1}^{\infty} \left(\sum_{m=1}^{\infty} a_{mn} \right) = -1.$$

Note that for this series

$$\sum_{\mathbb{N} \times \mathbb{N}} |a_{mn}| = \infty,$$

so it is not absolutely convergent. Furthermore, both of the sums

$$\sum_{\mathbb{N} \times \mathbb{N}} a_{mn}^+, \qquad \sum_{\mathbb{N} \times \mathbb{N}} a_{mn}^-$$

diverge to ∞, so the unordered sum is not well-defined.

The following basic result, which is a special case of Fubini's theorem, guarantees that both of the iterated sums of an absolutely convergent double series exist and are equal to the unordered sum. It also gives a criterion for the absolute convergence of a double series in terms of the convergence of its iterated sums. The key point of the proof is that a double sum over a "rectangular" set of indices is equal to an iterated sum. We can then estimate sums over arbitrary finite subsets of indices in terms of sums over rectangles.

Theorem 4.48 (Fubini for double series)**.** A double series of real numbers $\sum a_{mn}$ converges absolutely if and only if either one of the iterated series

$$\sum_{m=1}^{\infty} \left(\sum_{n=1}^{\infty} |a_{mn}| \right), \qquad \sum_{n=1}^{\infty} \left(\sum_{m=1}^{\infty} |a_{mn}| \right)$$

converges. In that case, both iterated series converge to the unordered sum of the double series:

$$\sum_{\mathbb{N} \times \mathbb{N}} a_{mn} = \sum_{m=1}^{\infty} \left(\sum_{n=1}^{\infty} a_{mn} \right) = \sum_{n=1}^{\infty} \left(\sum_{m=1}^{\infty} a_{mn} \right).$$

Proof. First suppose that one of the iterated sums of the absolute values exists. Without loss of generality, we suppose that

$$\sum_{m=1}^{\infty} \left(\sum_{n=1}^{\infty} |a_{mn}| \right) < \infty.$$

Let $F \subset \mathbb{N} \times \mathbb{N}$ be a finite subset. Choose $M, N \in \mathbb{N}$ such that $m \leq M$ and $n \leq N$ for all $(m, n) \in F$. Then $F \subset R$ where the rectangle R is given by $R = \{1, 2, \ldots, M\} \times \{1, 2, \ldots, N\}$, so that

$$\sum_{F} |a_{mn}| \leq \sum_{R} |a_{mn}| = \sum_{m=1}^{M} \left(\sum_{n=1}^{N} |a_{mn}| \right) \leq \sum_{m=1}^{\infty} \left(\sum_{n=1}^{\infty} |a_{mn}| \right).$$

Thus, the finite partial sums of $\sum |a_{mn}|$ are bounded from above, so the unordered sum converges absolutely.

Conversely, suppose that the unordered sum converges absolutely. Then, using Proposition 2.25 and the fact that the supremum of partial sums of non-negative terms over rectangles in $\mathbb{N} \times \mathbb{N}$ is equal to the supremum over all finite subsets, we get that

$$\sum_{m=1}^{\infty} \left(\sum_{n=1}^{\infty} |a_{mn}| \right) = \sup_{M \in \mathbb{N}} \left\{ \sum_{m=1}^{M} \sup_{N \in \mathbb{N}} \left\{ \sum_{n=1}^{N} |a_{mn}| \right\} \right\}$$

$$= \sup_{(M,N) \in \mathbb{N} \times \mathbb{N}} \left\{ \sum_{m=1}^{M} \sum_{n=1}^{N} |a_{mn}| \right\}$$

$$= \sum_{\mathbb{N} \times \mathbb{N}} |a_{mn}|.$$

Thus, the iterated sums converge to the unordered sum. Moreover, we have similarly that

$$\sum_{m=1}^{\infty} \left(\sum_{n=1}^{\infty} a_{mn}^{+} \right) = \sum_{\mathbb{N} \times \mathbb{N}} a_{ij}^{+}, \qquad \sum_{m=1}^{\infty} \left(\sum_{n=1}^{\infty} a_{mn}^{-} \right) = \sum_{\mathbb{N} \times \mathbb{N}} a_{ij}^{-},$$

which implies that

$$\sum_{m=1}^{\infty} \left(\sum_{n=1}^{\infty} a_{mn} \right) = \sum_{m=1}^{\infty} \left(\sum_{n=1}^{\infty} a_{mn}^{+} \right) - \sum_{m=1}^{\infty} \left(\sum_{n=1}^{\infty} a_{mn}^{-} \right) = \sum_{\mathbb{N} \times \mathbb{N}} a_{mn}.$$

$$\square$$

The preceding results show that the sum of an absolutely convergent double series is unambiguous; unordered sums, sums of rearrangements into single series, and iterated sums all converge to the same value. On the other hand, the sum of a conditionally convergent double series depends on how it is defined, e.g., on how one chooses to rearrange the double series into a single series. We conclude by describing one other way to define double sums.

Example 4.49. A natural way to generalize the sum of a single series to a double series, going back to Cauchy (1821) and Pringsheim (1897), is to say that

$$\sum_{m,n=1}^{\infty} a_{mn} = S$$

if for every $\epsilon > 0$ there exists $M, N \in \mathbb{N}$ such that

$$\left| \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} - S \right| < \epsilon$$

for all $m > M$ and $n > N$. We write this definition briefly as

$$\sum_{m,n=1}^{\infty} a_{mn} = \lim_{M,N \to \infty} \left[\sum_{m=1}^{M} \sum_{n=1}^{N} a_{mn} \right].$$

That is, we sum over larger and larger "rectangles" of indices in the $\mathbb{N} \times \mathbb{N}$-plane.

An absolutely convergent series converges in this sense to the same sum, but some conditionally convergent series also converge. For example, using this definition of the sum, we have

$$\sum_{m,n=1}^{\infty} \frac{(-1)^{m+n}}{mn} = \lim_{M,N \to \infty} \left[\sum_{m=1}^{M} \sum_{n=1}^{N} \frac{(-1)^{m+n}}{mn} \right]$$

$$= \lim_{M \to \infty} \left[\sum_{m=1}^{M} \frac{(-1)^m}{m} \right] \lim_{N \to \infty} \left[\sum_{n=1}^{N} \frac{(-1)^n}{n} \right]$$

$$= \left[\sum_{n=1}^{\infty} \frac{(-1)^n}{n} \right]^2$$

$$= (\log 2)^2,$$

but the series is not absolutely convergent, since the sums of both its positive and negative terms diverge to ∞.

This definition of a sum is not, however, as easy to use as the unordered sum of an absolutely convergent series. For example, it does not satisfy Fubini's theorem (although one can show that if the sum of a double series exists in this sense *and* the iterated sums also exist, then they are equal [**16**]).

4.11. * The irrationality of e

In this section, we use series to prove that e is an irrational number. In Proposition 3.32, we defined $e \approx 2.71828\ldots$ as the limit

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

We first obtain an alternative expression for e as the sum of a series.

Proposition 4.50.

$$e = \sum_{n=0}^{\infty} \frac{1}{n!}.$$

Proof. Using the binomial theorem, as in the proof of Proposition 3.32, we find that

$$\left(1 + \frac{1}{n}\right)^n = 2 + \frac{1}{2!}\left(1 - \frac{1}{n}\right) + \frac{1}{3!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)$$
$$+ \cdots + \frac{1}{k!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\ldots\left(1 - \frac{k-1}{n}\right)$$
$$+ \cdots + \frac{1}{n!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\ldots\frac{2}{n}\cdot\frac{1}{n}$$
$$< 2 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \cdots + \frac{1}{n!}.$$

Taking the limit of this inequality as $n \to \infty$, we get that

$$e \leq \sum_{n=0}^{\infty} \frac{1}{n!}.$$

To get the reverse inequality, we observe that for every $2 \leq k \leq n$,

$$\left(1 + \frac{1}{n}\right)^n \geq 2 + \frac{1}{2!}\left(1 - \frac{1}{n}\right) + \frac{1}{3!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)$$
$$+ \cdots + \frac{1}{k!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\ldots\left(1 - \frac{k-1}{n}\right).$$

Fixing k and taking the limit as $n \to \infty$, we get that

$$e \geq \sum_{j=0}^{k} \frac{1}{j!}.$$

Then, taking the limit as $k \to \infty$, we find that

$$e \geq \sum_{n=0}^{\infty} \frac{1}{n!},$$

which proves the result. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

This series for e is very rapidly convergent. The next proposition gives an explicit error estimate.

Proposition 4.51. For every $n \in \mathbb{N}$,

$$0 < e - \sum_{k=0}^{n} \frac{1}{k!} < \frac{1}{n \cdot n!}.$$

Proof. The lower bound is obvious. For the upper bound, we estimate the tail of the series by a geometric series:

$$e - \sum_{k=0}^{n} \frac{1}{k!} = \sum_{k=n+1}^{\infty} \frac{1}{k!}$$

$$= \frac{1}{n!} \left(\frac{1}{n+1} + \frac{1}{(n+1)(n+2)} + \cdots + \frac{1}{(n+1)(n+2)\ldots(n+k)} + \cdots \right)$$

$$< \frac{1}{n!} \left(\frac{1}{n+1} + \frac{1}{(n+1)^2} + \cdots + \frac{1}{(n+1)^k} + \cdots \right)$$

$$< \frac{1}{n!} \cdot \frac{1}{n}.$$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

Theorem 4.52. The number e is irrational.

Proof. Suppose for contradiction that $e = p/q$ for some $p, q \in \mathbb{N}$. From Proposition 4.51,

$$0 < \frac{p}{q} - \sum_{k=0}^{n} \frac{1}{k!} < \frac{1}{n \cdot n!}$$

for every $n \in \mathbb{N}$. Multiplying this inequality by $n!$, we get

$$0 < \frac{p \cdot n!}{q} - \sum_{k=0}^{n} \frac{n!}{k!} < \frac{1}{n}.$$

The middle term is an integer if $n \geq q$, which is impossible since it lies strictly between 0 and $1/n$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

A real number is said to be algebraic if it is a root of a polynomial with integer coefficients, otherwise it is transcendental. For example, every rational number $x = p/q$, where $p, q \in \mathbb{Z}$ and $q \neq 0$, is algebraic, since it is the solution of $qx - p = 0$; and $\sqrt{2}$ is an irrational algebraic number, since it is a solution of $x^2 - 2 = 0$. It's possible to prove that e is not only irrational but transcendental, but the proof is harder. Two other explicit examples of transcendental numbers are π and $2^{\sqrt{2}}$.

There is a long history of the study of irrational and transcendental numbers. Euler (1737) proved the irrationality of e, and Lambert (1761) proved the irrationality of π. The first proof of the existence of a transcendental number was given by Liouville (1844), who showed that

$$\sum_{n=1}^{\infty} \frac{1}{10^{n!}} = 0.1100010000000000000000000100\ldots$$

is transcendental. The transcendence of e was proved by Hermite (1873), the transcendence of π by Lindemann (1882), and the transcendence of $2^{\sqrt{2}}$ independently by Gelfond and Schneider (1934).

Cantor (1878) observed that the set of algebraic numbers is countably infinite, since there are countable many polynomials with integer coefficients, each such polynomial has finitely many roots, and the countable union of countable sets is countable. This argument proved the existence of uncountably many transcendental numbers without exhibiting any explicit examples, which was a remarkable result given the difficulties mathematicians had encountered (and still encounter) in proving that specific numbers are transcendental.

There remain many unsolved problems in this area. For example, it's not known if the Euler constant

$$\gamma = \lim_{n \to \infty} \left(\sum_{k=1}^{n} \frac{1}{k} - \log n \right)$$

from Example 4.11 is rational or irrational.

Topology of the Real Numbers

In this chapter, we define some topological properties of the real numbers $\mathbb{R}$ and its subsets.

5.1. Open sets

Open sets are among the most important subsets of $\mathbb{R}$. A collection of open sets is called a topology, and any property (such as convergence, compactness, or continuity) that can be defined entirely in terms of open sets is called a topological property.

Definition 5.1. A set $G \subset \mathbb{R}$ is open if for every $x \in G$ there exists a $\delta > 0$ such that $G \supset (x - \delta, x + \delta)$.

The entire set of real numbers $\mathbb{R}$ is obviously open, and the empty set $\varnothing$ is open since it satisfies the definition vacuously (there is no $x \in \varnothing$).

Example 5.2. The open interval $I = (0, 1)$ is open. If $x \in I$, then

$$I \supset (x - \delta, x + \delta), \qquad \delta = \min\left(\frac{x}{2}, \frac{1-x}{2}\right) > 0.$$

Similarly, every finite or infinite open interval (a, b), $(-\infty, b)$, or (a, ∞) is open.

Example 5.3. The half-open interval $J = (0, 1]$ isn't open, since $1 \in J$ and $(1 - \delta, 1 + \delta)$ isn't a subset of J for any $\delta > 0$, however small.

The next proposition states a characteristic property of open sets.

Proposition 5.4. An arbitrary union of open sets is open, and a finite intersection of open sets is open.

Proof. Suppose that $\{A_i \subset \mathbb{R} : i \in I\}$ is an arbitrary collection of open sets. If $x \in \bigcup_{i \in I} A_i$, then $x \in A_i$ for some $i \in I$. Since A_i is open, there is $\delta > 0$ such that $A_i \supset (x - \delta, x + \delta)$, and therefore

$$\bigcup_{i \in I} A_i \supset (x - \delta, x + \delta),$$

which proves that $\bigcup_{i \in I} A_i$ is open.

Suppose that $\{A_i \subset \mathbb{R} : i = 1, 2, \ldots, n\}$ is a finite collection of open sets. If $x \in \bigcap_{i=1}^n A_i$, then $x \in A_i$ for every $1 \leq i \leq n$. Since A_i is open, there is $\delta_i > 0$ such that $A_i \supset (x - \delta_i, x + \delta_i)$. Let

$$\delta = \min\{\delta_1, \delta_2, \ldots, \delta_n\} > 0.$$

Then we see that

$$\bigcap_{i=1}^n A_i \supset (x - \delta, x + \delta),$$

which proves that $\bigcap_{i=1}^n A_i$ is open. $\qquad\qquad\square$

The previous proof fails for an infinite intersection of open sets, since we may have $\delta_i > 0$ for every $i \in \mathbb{N}$ but $\inf\{\delta_i : i \in \mathbb{N}\} = 0$.

Example 5.5. The interval

$$I_n = \left(-\frac{1}{n}, \frac{1}{n}\right)$$

is open for every $n \in \mathbb{N}$, but

$$\bigcap_{n=1}^{\infty} I_n = \{0\}$$

is not open.

In fact, every open set in $\mathbb{R}$ is a countable union of disjoint open intervals, but we won't prove it here.

5.1.1. Neighborhoods. Next, we introduce the notion of the neighborhood of a point, which often gives clearer, but equivalent, descriptions of topological concepts than ones that use open intervals.

Definition 5.6. A set $U \subset \mathbb{R}$ is a neighborhood of a point $x \in \mathbb{R}$ if

$$U \supset (x - \delta, x + \delta)$$

for some $\delta > 0$. The open interval $(x - \delta, x + \delta)$ is called a δ-neighborhood of x.

A neighborhood of x needn't be an open interval about x, it just has to contain one. Some people require than a neighborhood is also an open set, but we don't; we'll specify that a neighborhood is open if it's needed.

Example 5.7. If $a < x < b$, then the closed interval $[a, b]$ is a neighborhood of x, since it contains the interval $(x - \delta, x + \delta)$ for sufficiently small $\delta > 0$. On the other hand, $[a, b]$ is not a neighborhood of the endpoints a, b since no open interval about a or b is contained in $[a, b]$.

We can restate the definition of open sets in terms of neighborhoods as follows.

Definition 5.8. A set $G \subset \mathbb{R}$ is open if every $x \in G$ has a neighborhood U such that $G \supset U$.

In particular, an open set is itself a neighborhood of each of its points.

We can restate Definition 3.10 for the limit of a sequence in terms of neighborhoods as follows.

Proposition 5.9. A sequence (x_n) of real numbers converges to a limit $x \in \mathbb{R}$ if and only if for every neighborhood U of x there exists $N \in \mathbb{N}$ such that $x_n \in U$ for all $n > N$.

Proof. First suppose the condition in the proposition holds. Given $\epsilon > 0$, let $U = (x - \epsilon, x + \epsilon)$ be an ϵ-neighborhood of x. Then there exists $N \in \mathbb{N}$ such that $x_n \in U$ for all $n > N$, which means that $|x_n - x| < \epsilon$. Thus, $x_n \to x$ as $n \to \infty$.

Conversely, suppose that $x_n \to x$ as $n \to \infty$, and let U be a neighborhood of x. Then there exists $\epsilon > 0$ such that $U \supset (x - \epsilon, x + \epsilon)$. Choose $N \in \mathbb{N}$ such that $|x_n - x| < \epsilon$ for all $n > N$. Then $x_n \in U$ for all $n > N$, which proves the condition. $\qquad\square$

5.1.2. Relatively open sets. We define relatively open sets by restricting open sets in $\mathbb{R}$ to a subset.

Definition 5.10. If $A \subset \mathbb{R}$ then $B \subset A$ is relatively open in A, or open in A, if $B = A \cap G$ where G is open in $\mathbb{R}$.

Example 5.11. Let $A = [0, 1]$. Then the half-open intervals $(a, 1]$ and $[0, b)$ are open in A for every $0 \le a < 1$ and $0 < b \le 1$, since

$$(a, 1] = [0, 1] \cap (a, 2), \qquad [0, b) = [0, 1] \cap (-1, b)$$

and $(a, 2)$, $(-1, b)$ are open in $\mathbb{R}$. By contrast, neither $(a, 1]$ nor $[0, b)$ is open in $\mathbb{R}$.

The neighborhood definition of open sets generalizes to relatively open sets. First, we define relative neighborhoods in the obvious way.

Definition 5.12. If $A \subset \mathbb{R}$ then a relative neighborhood in A of a point $x \in A$ is a set $V = A \cap U$ where U is a neighborhood of x in $\mathbb{R}$.

As we show next, a set is relatively open if and only if it contains a relative neighborhood of every point.

Proposition 5.13. A set $B \subset A$ is relatively open in A if and only if every $x \in B$ has a relative neighborhood V in A such that $B \supset V$.

Proof. Assume that B is open in A. Then $B = A \cap G$ where G is open in $\mathbb{R}$. If $x \in B$, then $x \in G$, and since G is open, there is a neighborhood U of x in $\mathbb{R}$ such that $G \supset U$. Then $V = A \cap U$ is a relative neighborhood of x with $B \supset V$.

Conversely, assume that every point $x \in B$ has a relative neighborhood $V_x = A \cap U_x$ in A such that $V_x \subset B$, where U_x is a neighborhood of x in $\mathbb{R}$. Since U_x is a neighborhood of x, it contains an open neighborhood $G_x \subset U_x$. We claim that that $B = A \cap G$ where

$$G = \bigcup_{x \in B} G_x.$$

It then follows that G is open, since it's a union of open sets, and therefore $B = A \cap G$ is relatively open in A.

To prove the claim, we show that $B \subset A \cap G$ and $B \supset A \cap G$. First, $B \subset A \cap G$ since $x \in A \cap G_x \subset A \cap G$ for every $x \in B$. Second, $A \cap G_x \subset A \cap U_x \subset B$ for every $x \in B$. Taking the union over $x \in B$, we get that $A \cap G \subset B$. $\square$

5.2. Closed sets

Sets are not doors. (Attributed to James Munkres.)

Closed sets are defined topologically as complements of open sets.

Definition 5.14. A set $F \subset \mathbb{R}$ is closed if $F^c = \{x \in \mathbb{R} : x \notin F\}$ is open.

Example 5.15. The closed interval $I = [0, 1]$ is closed since

$$I^c = (-\infty, 0) \cup (1, \infty)$$

is a union of open intervals, and therefore it's open. Similarly, every finite or infinite closed interval $[a, b]$, $(-\infty, b]$, or $[a, \infty)$ is closed.

The empty set $\varnothing$ and $\mathbb{R}$ are both open and closed; they're the only such sets. Most subsets of $\mathbb{R}$ are neither open nor closed (so, unlike doors, "not open" doesn't mean "closed" and "not closed" doesn't mean "open").

Example 5.16. The half-open interval $I = (0, 1]$ isn't open because it doesn't contain any neighborhood of the right endpoint $1 \in I$. Its complement

$$I^c = (\infty, 0] \cup (1, \infty)$$

isn't open either, since it doesn't contain any neighborhood of $0 \in I^c$. Thus, I isn't closed either.

Example 5.17. The set of rational numbers $\mathbb{Q} \subset \mathbb{R}$ is neither open nor closed. It isn't open because every neighborhood of a rational number contains irrational numbers, and its complement isn't open because every neighborhood of an irrational number contains rational numbers.

Closed sets can also be characterized in terms of sequences.

Proposition 5.18. A set $F \subset \mathbb{R}$ is closed if and only if the limit of every convergent sequence in F belongs to F.

Proof. First suppose that F is closed and (x_n) is a convergent sequence of points $x_n \in F$ such that $x_n \to x$. Then every neighborhood of x contains points $x_n \in F$. It follows that $x \notin F^c$, since F^c is open and every $y \in F^c$ has a neighborhood $U \subset F^c$ that contains no points in F. Therefore, $x \in F$.

Conversely, suppose that the limit of every convergent sequence of points in F belongs to F. Let $x \in F^c$. Then x must have a neighborhood $U \subset F^c$; otherwise for every $n \in \mathbb{N}$ there exists $x_n \in F$ such that $x_n \in (x - 1/n, x + 1/n)$, so $x = \lim x_n$, and x is the limit of a sequence in F. Thus, F^c is open and F is closed. $\square$

Example 5.19. To verify that the closed interval $[0, 1]$ is closed from Proposition 5.18, suppose that (x_n) is a convergent sequence in $[0, 1]$. Then $0 \leq x_n \leq 1$ for all $n \in \mathbb{N}$, and since limits preserve (non-strict) inequalities, we have

$$0 \leq \lim_{n \to \infty} x_n \leq 1,$$

meaning that the limit belongs to $[0, 1]$. On the other hand, the half-open interval $I = (0, 1]$ isn't closed since, for example, $(1/n)$ is a convergent sequence in I whose limit 0 doesn't belong to I.

Closed sets have complementary properties to those of open sets stated in Proposition 5.4.

Proposition 5.20. An arbitrary intersection of closed sets is closed, and a finite union of closed sets is closed.

Proof. If $\{F_i : i \in I\}$ is an arbitrary collection of closed sets, then every F_i^c is open. By De Morgan's laws in Proposition 1.23, we have

$$\left(\bigcap_{i \in I} F_i \right)^c = \bigcup_{i \in I} F_i^c,$$

which is open by Proposition 5.4. Thus $\bigcap_{i \in I} F_i$ is closed. Similarly, the complement of a finite union of closed sets is open, since

$$\left(\bigcup_{i=1}^{n} F_i \right)^c = \bigcap_{i=1}^{n} F_i^c,$$

so a finite union of closed sets is closed. □

The union of infinitely many closed sets needn't be closed.

Example 5.21. If I_n is the closed interval

$$I_n = \left[\frac{1}{n}, 1 - \frac{1}{n} \right],$$

then the union of the I_n is an open interval

$$\bigcup_{n=1}^{\infty} I_n = (0, 1).$$

If A is a subset of $\mathbb{R}$, it is useful to consider different ways in which a point $x \in \mathbb{R}$ can belong to A or be "close" to A.

Definition 5.22. Let $A \subset \mathbb{R}$ be a subset of $\mathbb{R}$. Then $x \in \mathbb{R}$ is:

(1) an interior point of A if there exists $\delta > 0$ such that $A \supset (x - \delta, x + \delta)$;

(2) an isolated point of A if $x \in A$ and there exists $\delta > 0$ such that x is the only point in A that belongs to the interval $(x - \delta, x + \delta)$;

(3) a boundary point of A if for every $\delta > 0$ the interval $(x - \delta, x + \delta)$ contains points in A and points not in A;

(4) an accumulation point of A if for every $\delta > 0$ the interval $(x - \delta, x + \delta)$ contains a point in A that is distinct from x.

When the set A is understood from the context, we refer, for example, to an "interior point."

Interior and isolated points of a set belong to the set, whereas boundary and accumulation points may or may not belong to the set. In the definition of a boundary point x, we allow the possibility that x itself is a point in A belonging to $(x - \delta, x + \delta)$, but in the definition of an accumulation point, we consider only points in A belonging to $(x - \delta, x + \delta)$ that are distinct from x. Thus an isolated point is a boundary point, but it isn't an accumulation point. Accumulation points are also called cluster points or limit points.

We illustrate these definitions with a number of examples.

Example 5.23. Let $I = (a, b)$ be an open interval and $J = [a, b]$ a closed interval. Then the set of interior points of I or J is (a, b), and the set of boundary points consists of the two endpoints $\{a, b\}$. The set of accumulation points of I or J is the closed interval $[a, b]$ and I, J have no isolated points. Thus, I, J have the same interior, isolated, boundary and accumulation points, but J contains its boundary points and all of its accumulation points, while I does not.

Example 5.24. Let $a < c < b$ and suppose that

$$A = (a, c) \cup (c, b)$$

is an open interval punctured at c. Then the set of interior points is A, the set of boundary points is $\{a, b, c\}$, the set of accumulation points is the closed interval $[a, b]$, and there are no isolated points.

Example 5.25. Let

$$A = \left\{ \frac{1}{n} : n \in \mathbb{N} \right\}.$$

Then every point of A is an isolated point, since a sufficiently small interval about $1/n$ doesn't contain $1/m$ for any integer $m \neq n$, and A has no interior points. The set of boundary points of A is $A \cup \{0\}$. The point $0 \notin A$ is the only accumulation point of A, since every open interval about 0 contains $1/n$ for sufficiently large n.

Example 5.26. The set $\mathbb{N}$ of natural numbers has no interior or accumulation points. Every point of $\mathbb{N}$ is both a boundary point and an isolated point.

Example 5.27. The set $\mathbb{Q}$ of rational numbers has no interior or isolated points, and every real number is both a boundary and accumulation point of $\mathbb{Q}$.

Example 5.28. The Cantor set C defined in Section 5.5 below has no interior points and no isolated points. The set of accumulation points and the set of boundary points of C is equal to C.

The following proposition gives a sequential definition of an accumulation point.

Proposition 5.29. A point $x \in \mathbb{R}$ is an accumulation point of $A \subset \mathbb{R}$ if and only if there is a sequence (x_n) in A with $x_n \neq x$ for every $n \in \mathbb{N}$ such that $x_n \to x$ as $n \to \infty$.

Proof. Suppose $x \in \mathbb{R}$ is an accumulation point of A. Definition 5.22 implies that for every $n \in \mathbb{N}$ there exists $x_n \in A \setminus \{x\}$ such that $x_n \in (x - 1/n, x + 1/n)$. It follows that $x_n \to x$ as $n \to \infty$.

Conversely, if x is the limit of a sequence (x_n) in A with $x_n \neq x$, and U is a neighborhood of x, then $x_n \in U \setminus \{x\}$ for sufficiently large $n \in \mathbb{N}$, which proves that x is an accumulation point of A. □

Example 5.30. If

$$A = \left\{ \frac{1}{n} : n \in \mathbb{N} \right\},$$

then 0 is an accumulation point of A, since $(1/n)$ is a sequence in A such that $1/n \to 0$ as $n \to \infty$. On the other hand, 1 is not an accumulation point of A since the only sequences in A that converges to 1 are the ones whose terms eventually equal 1, and the terms are required to be distinct from 1.

We can also characterize open and closed sets in terms of their interior and accumulation points.

Proposition 5.31. A set $A \subset \mathbb{R}$ is:

(1) open if and only if every point of A is an interior point;

(2) closed if and only if every accumulation point belongs to A.

Proof. If A is open, then it is an immediate consequence of the definitions that every point in A is an interior point. Conversely, if every point $x \in A$ is an interior point, then there is an open neighborhood $U_x \subset A$ of x, so

$$A = \bigcup_{x \in A} U_x$$

is a union of open sets, and therefore A is open.

If A is closed and x is an accumulation point, then Proposition 5.29 and Proposition 5.18 imply that $x \in A$. Conversely, if every accumulation point of A belongs to A, then every $x \in A^c$ has a neighborhood with no points in A, so A^c is open and A is closed. □

5.3. Compact sets

The significance of compact sets is not as immediately apparent as the significance of open sets, but the notion of compactness plays a central role in analysis. One indication of its importance already appears in the Bolzano-Weierstrass theorem (Theorem 3.57).

Compact sets may be characterized in many different ways, and we will give the two most important definitions. One is based on sequences (every sequence has a convergent subsequence), and the other is based on open sets (every open cover has a finite subcover).

We will prove that a subset of $\mathbb{R}$ is compact if and only if it is closed and bounded. For example, every closed, bounded interval $[a, b]$ is compact. There are, however, many other compact subsets of $\mathbb{R}$. In Section 5.5 we describe a particularly interesting example called the Cantor set.

We emphasize that although the compact sets in $\mathbb{R}$ are exactly the closed and bounded sets, this isn't their fundamental definition; rather it's an explicit description of what compact sets look like in $\mathbb{R}$. In more general spaces than $\mathbb{R}$, closed and

bounded sets need not be compact, and it's the properties defining compactness that are the crucial ones. Chapter 13 has further explanation.

5.3.1. Sequential definition. Intuitively, a compact set confines every sequence of points in the set so much that the sequence must accumulate at some point of the set. This implies that a subsequence converges to an accumulation point and leads to the following definition.

Definition 5.32. A set $K \subset \mathbb{R}$ is sequentially compact if every sequence in K has a convergent subsequence whose limit belongs to K.

Note that we require that the subsequence converges to a point in K, not to a point outside K.

We usually abbreviate "sequentially compact" to "compact," but sometimes we need to distinguish explicitly between the sequential definition of compactness given above and the topological definition given in Definition 5.52 below.

Example 5.33. The open interval $I = (0, 1)$ is not compact. The sequence $(1/n)$ in I converges to 0, so every subsequence also converges to $0 \notin I$. Therefore, $(1/n)$ has no convergent subsequence whose limit belongs to I.

Example 5.34. The set $A = \mathbb{Q} \cap [0, 1]$ of rational numbers in $[0, 1]$ is not compact. If (r_n) is a sequence of rational numbers $0 \leq r_n \leq 1$ that converges to $1/\sqrt{2}$, then every subsequence also converges to $1/\sqrt{2} \notin A$, so (r_n) has no subsequence that converges to a point in A.

Example 5.35. The set $\mathbb{N}$ is closed, but it is not compact. The sequence (n) in $\mathbb{N}$ has no convergent subsequence since every subsequence diverges to infinity.

As these examples illustrate, a compact set must be closed and bounded. Conversely, the Bolzano-Weierstrass theorem implies that that every closed, bounded subset of $\mathbb{R}$ is compact. This fact may be taken as an alternative statement of the theorem.

Theorem 5.36 (Bolzano-Weierstrass)**.** A subset of $\mathbb{R}$ is sequentially compact if and only if it is closed and bounded.

Proof. First, assume that $K \subset \mathbb{R}$ is sequentially compact. Let (x_n) be a sequence in K that converges to $x \in \mathbb{R}$. Then every subsequence of K also converges to x, so the compactness of K implies that $x \in K$. It follows from Proposition 5.18 that K is closed. Next, suppose for contradiction that K is unbounded. Then there is a sequence (x_n) in K such that $|x_n| \to \infty$ as $n \to \infty$. Every subsequence of (x_n) is also unbounded and therefore diverges, so (x_n) has no convergent subsequence. This contradicts the assumption that K is sequentially compact, so K is bounded.

Conversely, assume that $K \subset \mathbb{R}$ is closed and bounded. Let (x_n) be a sequence in K. Then (x_n) is bounded since K is bounded, and Theorem 3.57 implies that (x_n) has a convergent subsequence. Since K is closed the limit of this subsequence belongs to K, so K is sequentially compact. $\square$

Example 5.37. Every closed, bounded interval $[a, b]$ is compact.

Example 5.38. Let
$$A = \left\{ \frac{1}{n} : n \in \mathbb{N} \right\}.$$
Then A is not compact, since it isn't closed. However, the set $K = A \cup \{0\}$ is closed and bounded, so it is compact.

Example 5.39. The Cantor set defined in Section 5.5 is compact.

For later use, we prove a useful property of compact sets in $\mathbb{R}$ which follows from Theorem 5.36.

Proposition 5.40. If $K \subset \mathbb{R}$ is compact, then K has a maximum and minimum.

Proof. Since K is compact it is bounded and therefore it has a (finite) supremum $M = \sup K$. From the definition of the supremum, for every $n \in \mathbb{N}$ there exists $x_n \in K$ such that
$$M - \frac{1}{n} < x_n \le M.$$
It follows from the 'squeeze' theorem that $x_n \to M$ as $n \to \infty$. Since K is closed, $M \in K$, which proves that K has a maximum. A similar argument shows that $m = \inf K$ belongs to K, so K has a minimum. $\qquad \square$

Example 5.41. The bounded closed interval $[0,1]$ is compact and its maximum 1 and minimum 0 belong to the set, while the open interval $(0,1)$ is not compact and its supremum 1 and infimum 0 do not belong to the set. The unbounded, closed interval $[0, \infty)$ is not compact, and it has no maximum.

Example 5.42. The set A in Example 5.38 is not compact and its infimum 0 does not belong to the set, but the compact set K has 0 as a minimum value.

Compact sets have the following nonempty intersection property.

Theorem 5.43. Let $\{K_n : n \in \mathbb{N}\}$ be a decreasing sequence of nonempty compact sets of real numbers, meaning that
$$K_1 \supset K_2 \supset \cdots \supset K_n \supset K_{n+1} \supset \dots,$$
and $K_n \ne \varnothing$. Then
$$\bigcap_{n=1}^{\infty} K_n \ne \varnothing.$$
Moreover, if $\operatorname{diam} K_n \to 0$ as $n \to \infty$, then the intersection consists of a single point.

Proof. For each $n \in \mathbb{N}$, choose $x_n \in K_n$. Since (x_n) is a sequence in the compact set K_1, it has a convergent subsequence (x_{n_k}) with $x_{n_k} \to x$ as $k \to \infty$. Then $x_{n_k} \in K_n$ for all k sufficiently large that $n_k \ge n$. Since a "tail" of the subsequence belongs to K_n and K_n is closed, we have $x \in K_n$ for every $n \in \mathbb{N}$. Hence, $x \in \bigcap K_n$, and the intersection is nonempty.

If $x, y \in \bigcap K_n$, then $x, y \in K_n$ for every $n \in \mathbb{N}$, so $|x - y| \le \operatorname{diam} K_n$. If $\operatorname{diam} K_n \to 0$ as $n \to \infty$, then $|x - y| = 0$, so $x = y$ and $\bigcap K_n$ consists of a single point. $\qquad \square$

We refer to a decreasing sequence of sets as a nested sequence. In the case when each $K_n = [a_n, b_n]$ is a compact interval, the preceding result is called the nested interval theorem.

Example 5.44. The nested compact intervals $[0, 1 + 1/n]$ have nonempty intersection $[0, 1]$. Here, $\operatorname{diam}[0, 1 + 1/n] \to 1$ as $n \to \infty$, and the intersection consists of an interval. The nested compact intervals $[0, 1/n]$ have nonempty intersection $\{0\}$, which consists of a single point since $\operatorname{diam}[0, 1/n] \to 0$ as $n \to \infty$. On the other hand, the nested half-open intervals $(0, 1/n]$ have empty intersection, as do the nested unbounded, closed intervals $[n, \infty)$. In particular, Theorem 5.43 doesn't hold if we replace "compact" by "closed."

Example 5.45. Define a nested sequence $A_1 \supset A_2 \supset \ldots$ of non-compact sets by

$$A_n = \left\{ \frac{1}{k} : k = n, n + 1, n + 2, \ldots \right\},$$

so $A_1 = A$ where A is the set considered in Example 5.38. Then

$$\bigcap_{n=1}^{\infty} A_n = \varnothing.$$

If we add 0 to the A_n to make them compact and define $K_n = A_n \cup \{0\}$, then the intersection

$$\bigcap_{n=1}^{\infty} K_n = \{0\}$$

is nonempty.

5.3.2. Topological definition. To give a topological definition of compactness in terms of open sets, we introduce the notion of an open cover of a set.

Definition 5.46. Let $A \subset \mathbb{R}$. A cover of A is a collection of sets $\{A_i \subset \mathbb{R} : i \in I\}$ whose union contains A,

$$\bigcup_{i \in I} A_i \supset A.$$

An open cover of A is a cover such that A_i is open for every $i \in I$.

Example 5.47. Let $A_i = (1/i, 2)$. Then $\mathcal{C} = \{A_i : i \in \mathbb{N}\}$ is an open cover of $(0, 1]$, since

$$\bigcup_{i=1}^{\infty} \left(\frac{1}{i}, 2 \right) = (0, 2) \supset (0, 1].$$

On the other hand, $\mathcal{C}$ is not a cover of $[0, 1]$ since its union does not contain 0. If, for any $\delta > 0$, we add the interval $B = (-\delta, \delta)$ to $\mathcal{C}$, then

$$\bigcup_{i=1}^{\infty} \left(\frac{1}{i}, 2 \right) \cup B = (-\delta, 2) \supset [0, 1],$$

so $\mathcal{C}' = \mathcal{C} \cup \{B\}$ is an open cover of $[0, 1]$.

Example 5.48. If $A_i = (i - 1, i + 1)$, then $\{A_i : i \in \mathbb{Z}\}$ is an open cover of $\mathbb{R}$. On the other hand, if $B_i = (i, i + 1)$, then $\{B_i : i \in \mathbb{Z}\}$ is not open cover of $\mathbb{R}$, since its union doesn't contain any of the integers. Finally, if $C_i = [i, i + 1)$, then $\{C_i : i \in \mathbb{Z}\}$ is a cover of $\mathbb{R}$ by disjoint, half-open intervals, but it isn't an open cover. Thus, to get an open cover, we need the intervals to "overlap".

Example 5.49. Let $\{r_i : i \in \mathbb{N}\}$ be an enumeration of the rational numbers $r_i \in [0, 1]$, and fix $\epsilon > 0$. Define $A_i = (r_i - \epsilon, r_i + \epsilon)$. Then $\{A_i : i \in \mathbb{N}\}$ is an open cover of $[0, 1]$ since every irrational number $x \in [0, 1]$ can be approximated to within ϵ by some rational number. Similarly, if $I = [0, 1] \setminus \mathbb{Q}$ denotes the set of irrational numbers in $[0, 1]$, then $\{(x - \epsilon, x + \epsilon) : x \in I\}$ is an open cover of $[0, 1]$. In this case, the cover consists of uncountably many sets.

Next, we define subcovers.

Definition 5.50. Suppose that $\mathcal{C} = \{A_i \subset \mathbb{R} : i \in I\}$ is a cover of $A \subset \mathbb{R}$. A subcover $\mathcal{S}$ of $\mathcal{C}$ is a sub-collection $\mathcal{S} \subset \mathcal{C}$ that covers A, meaning that

$$\mathcal{S} = \{A_{i_k} \in \mathcal{C} : k \in J\}, \qquad \bigcup_{k \in J} A_{i_k} \supset A.$$

A finite subcover is a subcover $\{A_{i_1}, A_{i_2}, \ldots, A_{i_n}\}$ that consists of finitely many sets.

Example 5.51. Consider the cover $\mathcal{C} = \{A_i : i \in \mathbb{N}\}$ of $(0, 1]$ in Example 5.47, where $A_i = (1/i, 2)$. Then $\{A_{2j} : j \in \mathbb{N}\}$ is a subcover. There is, however, no finite subcover $\{A_{i_1}, A_{i_2}, \ldots, A_{i_n}\}$ since if $N = \max\{i_1, i_2, \ldots, i_n\}$ then

$$\bigcup_{k=1}^{n} A_{k_i} = \left(\frac{1}{N}, 2 \right),$$

which does not contain the points $x \in (0, 1)$ with $0 < x \le 1/N$. On the other hand, the cover $\mathcal{C}' = \mathcal{C} \cup \{(-\delta, \delta)\}$ of $[0, 1]$ does have a finite subcover. For example, if $N \in \mathbb{N}$ is such that $1/N < \delta$, then

$$\left\{ \left(\frac{1}{N}, 2 \right), (-\delta, \delta) \right\}$$

is a finite subcover of $[0, 1]$ consisting of two sets (whose union is the same as the original cover).

Having introduced this terminology, we give a topological definition of compact sets.

Definition 5.52. A set $K \subset \mathbb{R}$ is compact if every open cover of K has a finite subcover.

First, we illustrate Definition 5.52 with several examples.

Example 5.53. The collection of open intervals

$$\{A_i : i \in \mathbb{N}\}, \qquad A_i = (i - 1, i + 1)$$

is an open cover of the natural numbers $\mathbb{N}$, since

$$\bigcup_{i=1}^{\infty} A_i = (0, \infty) \supset \mathbb{N}.$$

However, no finite sub-collection

$$\{A_{i_1}, A_{i_2}, \ldots, A_{i_n}\}$$

covers $\mathbb{N}$, since if $N = \max\{i_1, i_2, \ldots, i_n\}$, then

$$\bigcup_{k=1}^{n} A_{i_k} \subset (0, N+1)$$

so its union does not contain sufficiently large integers with $n \geq N + 1$. Thus, $\mathbb{N}$ is not compact.

Example 5.54. Consider the open intervals

$$A_i = \left(\frac{1}{2^i} - \frac{1}{2^{i+1}}, \frac{1}{2^i} + \frac{1}{2^{i+1}} \right),$$

which get smaller as they get closer to 0. Then $\{A_i : i = 0, 1, 2, \ldots\}$ is an open cover of the open interval $(0, 1)$; in fact

$$\bigcup_{i=0}^{\infty} A_i = \left(0, \frac{3}{2} \right) \supset (0, 1).$$

However, no finite sub-collection

$$\{A_{i_1}, A_{i_2}, \ldots, A_{i_n}\}$$

of intervals covers $(0, 1)$, since if $N = \max\{i_1, i_2, \ldots, i_n\}$, then

$$\bigcup_{k=1}^{n} A_{i_k} \subset \left(\frac{1}{2^N} - \frac{1}{2^{N+1}}, \frac{3}{2} \right),$$

so it does not contain the points in $(0, 1)$ that are sufficiently close to 0. Thus, $(0, 1)$ is not compact. Example 5.51 gives another example of an open cover of $(0, 1)$ with no finite subcover.

Example 5.55. The collection of open intervals $\{A_0, A_1, A_2, \ldots\}$ in Example 5.54 isn't an open cover of the closed interval $[0, 1]$ since 0 doesn't belong to their union. We can get an open cover $\{A_0, A_1, A_2, \ldots, B\}$ of $[0, 1]$ by adding to the A_i an open interval $B = (-\delta, \delta)$, where $\delta > 0$ is arbitrarily small. In that case, if we choose $n \in \mathbb{N}$ sufficiently large that

$$\frac{1}{2^n} - \frac{1}{2^{n+1}} < \delta,$$

then $\{A_0, A_1, A_2, \ldots, A_n, B\}$ is a finite subcover of $[0, 1]$ since

$$\bigcup_{i=0}^{n} A_i \cup B = \left(-\delta, \frac{3}{2} \right) \supset [0, 1].$$

Points sufficiently close to 0 belong to B, while points further away belong to A_i for some $0 \leq i \leq n$. The open cover of $[0, 1]$ in Example 5.51 is similar.

As the previous example suggests, and as follows from the next theorem, every open cover of $[0, 1]$ has a finite subcover, and $[0, 1]$ is compact.

Theorem 5.56 (Heine-Borel)**.** A subset of $\mathbb{R}$ is compact if and only if it is closed and bounded.

Proof. The most important direction of the theorem is that a closed, bounded set is compact.

First, we prove that a closed, bounded interval $K = [a, b]$ is compact. Suppose that $\mathcal{C} = \{A_i : i \in I\}$ is an open cover of $[a, b]$, and let

$$B = \{x \in [a, b] : [a, x] \text{ has a finite subcover } \mathcal{S} \subset \mathcal{C}\}.$$

We claim that $\sup B = b$. The idea of the proof is that any open cover of $[a, x]$ must cover a larger interval since the open set that contains x extends past x.

Since $\mathcal{C}$ covers $[a, b]$, there exists a set $A_i \in \mathcal{C}$ with $a \in A_i$, so $[a, a] = \{a\}$ has a subcover consisting of a single set, and $a \in B$. Thus, B is non-empty and bounded from above by b, so $c = \sup B \le b$ exists. Assume for contradiction that $c < b$. Then $[a, c]$ has a finite subcover

$$\{A_{i_1}, A_{i_2}, \dots, A_{i_n}\},$$

with $c \in A_{i_k}$ for some $1 \le k \le n$. Since A_{i_k} is open and $a \le c < b$, there exists $\delta > 0$ such that $[c, c + \delta) \subset A_{i_k} \cap [a, b]$. Then $\{A_{i_1}, A_{i_2}, \dots, A_{i_n}\}$ is a finite subcover of $[a, x]$ for $c < x < c + \delta$, contradicting the definition of c, so $\sup B = b$. Moreover, the following argument shows that, in fact, $b = \max B$.

Since $\mathcal{C}$ covers $[a, b]$, there is an open set $A_{i_0} \in \mathcal{C}$ such that $b \in A_{i_0}$. Then $(b - \delta, b + \delta) \subset A_{i_0}$ for some $\delta > 0$, and since $\sup B = b$ there exists $c \in B$ such that $b - \delta < c \le b$. Let $\{A_{i_1}, \dots, A_{i_n}\}$ be a finite subcover of $[a, c]$. Then $\{A_{i_0}, A_{i_1}, \dots, A_{i_n}\}$ is a finite subcover of $[a, b]$, which proves that $[a, b]$ is compact.

Now suppose that $K \subset \mathbb{R}$ is a closed, bounded set, and let $\mathcal{C} = \{A_i : i \in I\}$ be an open cover of K. Since K is bounded, $K \subset [a, b]$ for some closed bounded interval $[a, b]$, and, since K is closed, $\mathcal{C}' = \mathcal{C} \cup \{K^c\}$ is an open cover of $[a, b]$. From what we have just proved, $[a, b]$ has a finite subcover that is included in $\mathcal{C}'$. Omitting K^c from this subcover, if necessary, we get a finite subcover of K that is included in the original cover $\mathcal{C}$.

To prove the converse, suppose that $K \subset \mathbb{R}$ is compact. Let $A_i = (-i, i)$. Then

$$\bigcup_{i=1}^{\infty} A_i = \mathbb{R} \supset K,$$

so $\{A_i : i \in \mathbb{N}\}$ is an open cover of K, which has a finite subcover $\{A_{i_1}, A_{i_2}, \dots, A_{i_n}\}$. Let $N = \max\{i_1, i_2, \dots, i_n\}$. Then

$$K \subset \bigcup_{k=1}^{n} A_{i_k} = (-N, N),$$

so K is bounded.

To prove that K is closed, we prove that K^c is open. Suppose that $x \in K^c$. For $i \in \mathbb{N}$, let

$$A_i = \left[x - \frac{1}{i}, x + \frac{1}{i}\right]^c = \left(-\infty, x - \frac{1}{i}\right) \cup \left(x + \frac{1}{i}, \infty\right).$$

Then $\{A_i : i \in \mathbb{N}\}$ is an open cover of K, since

$$\bigcup_{i=1}^{\infty} A_i = (-\infty, x) \cup (x, \infty) \supset K.$$

Since K is compact, there is a finite subcover $\{A_{i_1}, A_{i_2}, \ldots, A_{i_n}\}$. Let $N = \max\{i_1, i_2, \ldots, i_n\}$. Then

$$K \subset \bigcup_{k=1}^{n} A_{i_k} = \left(-\infty, x - \frac{1}{N}\right) \cup \left(x + \frac{1}{N}, \infty\right),$$

which implies that $(x - 1/N, x + 1/N) \subset K^c$. This proves that K^c is open and K is closed. □

The following corollary is an immediate consequence of what we have proved.

Corollary 5.57. A subset of $\mathbb{R}$ is compact if and only if it is sequentially compact.

Proof. By Theorem 5.36 and Theorem 5.56, a subset of $\mathbb{R}$ is compact or sequentially compact if and only if it is closed and bounded. □

Corollary 5.57 generalizes to an arbitrary metric space, where a set is compact if and only if it is sequentially compact, although a different proof is required. By contrast, Theorem 5.36 and Theorem 5.56 do not hold in an arbitrary metric space, where a closed, bounded set need not be compact.

5.4. Connected sets

A connected set is, roughly speaking, a set that cannot be divided into "separated" parts. The formal definition is as follows.

Definition 5.58. A set of real numbers $A \subset \mathbb{R}$ is disconnected if there are disjoint open sets $U, V \subset \mathbb{R}$ such that $A \cap U$ and $A \cap V$ are nonempty and

$$A = (A \cap U) \cup (A \cap V).$$

A set is connected if it not disconnected.

The condition $A = (A \cap U) \cup (A \cap V)$ is equivalent to $U \cup V \supset A$. If A is disconnected as in the definition, then we say that the open sets U, V separate A.

It is easy to give examples of disconnected sets. As the following examples illustrate, any set of real numbers that is "missing" a point is disconnected.

Example 5.59. The set $\{0, 1\}$ consisting of two points is disconnected. For example, let $U = (-1/2, 1/2)$ and $V = (1/2, 3/2)$. Then U, V are open and $U \cap V = \varnothing$. Furthermore, $A \cap U = \{0\}$ and $A \cap V = \{1\}$ are nonempty, and $A = (A \cap U) \cup (A \cap V)$. Similarly, the union of half-open intervals $[0, 1/2) \cup (1/2, 1]$ is disconnected.

Example 5.60. The set $\mathbb{R} \setminus \{0\}$ is disconnected since $\mathbb{R} \setminus \{0\} = (-\infty, 0) \cup (0, \infty)$.

Example 5.61. The set $\mathbb{Q}$ of rational numbers is disconnected. For example, let $U = (-\infty, \sqrt{2})$ and $V = (\sqrt{2}, \infty)$. Then U, V are disjoint open sets, $\mathbb{Q} \cap U$ and $\mathbb{Q} \cap V$ are nonempty, and $U \cup V = \mathbb{R} \setminus \{\sqrt{2}\} \supset \mathbb{Q}$.

In general, it is harder to prove that a set is connected than disconnected, because one has to show that there is no way to separate it by open sets. However, the ordering properties of $\mathbb{R}$ enable us to characterize its connected sets: they are exactly the intervals.

First, we give a precise definition of an interval.

Definition 5.62. A set of real numbers $I \subset \mathbb{R}$ is an interval if $x, y \in I$ and $x < y$ implies that $z \in I$ for every $x < z < y$.

That is, an interval is a set with the property that it contains all the points between any two points in the set.

We claim that, according to this definition, an interval contains all the points that lie between its infimum and supremum. The infimum and supremum may be finite or infinite, and they may or may not belong to the interval. Depending on which of these these possibilities occur, we see that an interval is any open, closed, half-open, bounded, or unbounded interval of the form

$$\varnothing, \quad (a, b), \quad [a, b], \quad [a, b), \quad (a, b], \quad (a, \infty), \quad [a, \infty), \quad (-\infty, b), \quad (-\infty, b], \quad \mathbb{R},$$

where $a, b \in \mathbb{R}$ and $a \leq b$. If $a = b$, then $[a, a] = \{a\}$ is an interval that consists of a single point, which — like the empty set — satisfies the definition vacuously. Thus, Definition 5.62 is consistent with the usual definition of an interval.

To prove the previous claim, suppose that I is an interval and let $a = \inf I$, $b = \sup I$ where $-\infty \leq a, b \leq \infty$. If $a > b$, then $I = \varnothing$, and if $a = b$, then I consists of a single point $\{a\}$. Otherwise, $-\infty \leq a < b \leq \infty$. In that case, the definition of the infimum and supremum implies that for every $a', b' \in \mathbb{R}$ with $a < a' < b' < b$, there exist $x, y \in I$ such that $a \leq x < a'$ and $b' < y \leq b$. Since I is an interval, it follows that $I \supset [x, y] \supset [a', b']$, and since $a' > a$, $b' < b$ are arbitrary, it follows that $I \supset (a, b)$. Moreover, since $a = \inf I$ and $b = \sup I$, the interval I cannot contain any points $x \in \mathbb{R}$ such that $x < a$ or $x > b$.

The slightly tricky part of the following theorem is the proof that every interval is connected.

Theorem 5.63. A set of real numbers is connected if and only if it is an interval.

Proof. First, suppose that $A \subset \mathbb{R}$ is not an interval. Then there are $a, b \in A$ and $c \notin A$ such that $a < c < b$. If $U = (-\infty, c)$ and $V = (c, \infty)$, then $a \in A \cap U$, $b \in A \cap V$, and $A = (A \cap U) \cup (A \cap V)$, so A is disconnected. It follows that every connected set is an interval.

To prove the converse, suppose that $I \subset \mathbb{R}$ is not connected. We will show that I is not an interval. Let U, V be open sets that separate I. Choose $a \in I \cap U$ and $b \in I \cap V$, where we can assume without loss of generality that $a < b$. Let

$$c = \sup \left(U \cap [a, b] \right).$$

We will prove that $a < c < b$ and $c \notin I$, meaning that I is not an interval. If $a \leq x < b$ and $x \in U$, then $U \supset [x, x + \delta)$ for some $\delta > 0$, so $x \neq \sup \left(U \cap [a, b] \right)$. Thus, $c \neq a$ and if $a < c < b$, then $c \notin U$. If $a < y \leq b$ and $y \in V$, then $V \supset (y - \delta, y]$ for some $\delta > 0$, and therefore $(y - \delta, y]$ is disjoint from U, which implies that $y \neq \sup \left(U \cap [a, b] \right)$. It follows that $c \neq b$ and $c \notin U \cap V$, so $a < c < b$ and $c \notin I$, which completes the proof. □

Figure 1. An illustration of the removal of middle-thirds from an interval in the construction of the Cantor set. The figure shows the interval $[0, 1]$ and the first four sets F_1, F_2, F_3, F_4, going from top to bottom.

5.5. * The Cantor set

One of the most interesting examples of a compact set is the Cantor set, which is obtained by "removing middle-thirds" from closed intervals in $[0, 1]$, as illustrated in Figure 1.

We define a nested sequence (F_n) of sets $F_n \subset [0, 1]$ as follows. First, we remove the middle-third from $[0, 1]$ to get $F_1 = [0, 1] \setminus (1/3, 2/3)$, or

$$F_1 = I_0 \cup I_1, \qquad I_0 = \left[0, \frac{1}{3}\right], \quad I_1 = \left[\frac{2}{3}, 1\right].$$

Next, we remove middle-thirds from I_0 and I_1, which splits $I_0 \setminus (1/9, 2/9)$ into $I_{00} \cup I_{01}$ and $I_1 \setminus (7/9, 8/9)$ into $I_{10} \cup I_{11}$, to get

$$F_2 = I_{00} \cup I_{01} \cup I_{10} \cup I_{11},$$

$$I_{00} = \left[0, \frac{1}{9}\right], \quad I_{01} = \left[\frac{2}{9}, \frac{1}{3}\right], \qquad I_{10} = \left[\frac{2}{3}, \frac{7}{9}\right], \quad I_{11} = \left[\frac{8}{9}, 1\right].$$

Then we remove middle-thirds from I_{00}, I_{01}, I_{10}, and I_{11} to get

$$F_3 = I_{000} \cup I_{001} \cup I_{010} \cup I_{011} \cup I_{100} \cup I_{101} \cup I_{110} \cup I_{111},$$

$$I_{000} = \left[0, \frac{1}{27}\right], \quad I_{010} = \left[\frac{2}{27}, \frac{1}{9}\right], \qquad I_{010} = \left[\frac{2}{9}, \frac{7}{27}\right], \quad I_{011} = \left[\frac{8}{27}, \frac{1}{3}\right],$$

$$I_{100} = \left[\frac{2}{3}, \frac{19}{27}\right], \quad I_{101} = \left[\frac{20}{27}, \frac{7}{9}\right], \qquad I_{110} = \left[\frac{8}{9}, \frac{25}{27}\right], \quad I_{111} = \left[\frac{26}{27}, 1\right].$$

Continuing in this way, we get at the nth stage a set of the form

$$F_n = \bigcup_{\mathbf{s} \in \Sigma_n} I_\mathbf{s},$$

where $\Sigma_n = \{(s_1, s_2, \ldots, s_n) : s_n = 0, 1\}$ is the set of binary n-tuples. Furthermore, each $I_\mathbf{s} = [a_\mathbf{s}, b_\mathbf{s}]$ is a closed interval, and if $\mathbf{s} = (s_1, s_2, \ldots, s_n)$, then

$$a_\mathbf{s} = \sum_{k=1}^{n} \frac{2s_k}{3^k}, \qquad b_\mathbf{s} = a_\mathbf{s} + \frac{1}{3^n}.$$

In other words, the left endpoints $a_\mathbf{s}$ are the points in $[0, 1]$ that have a finite base three expansion consisting entirely of 0's and 2's.

We can verify this formula for the endpoints $a_\mathbf{s}$, $b_\mathbf{s}$ of the intervals in F_n by induction. It holds when $n = 1$. Assume that it holds for some $n \in \mathbb{N}$. If we remove the middle-third of length $1/3^{n+1}$ from the interval $[a_\mathbf{s}, b_\mathbf{s}]$ of length $1/3^n$ with $\mathbf{s} \in \Sigma_n$, then we get the original left endpoint, which may be written as

$$a_{\mathbf{s}'} = \sum_{k=1}^{n+1} \frac{2s_k'}{3^k},$$

where $\mathbf{s}' = (s_1', \ldots, s_{n+1}') \in \Sigma_{n+1}$ is given by $s_k' = s_k$ for $k = 1, \ldots, n$ and $s_{n+1}' = 0$. We also get a new left endpoint $a_{\mathbf{s}'} = a_\mathbf{s} + 2/3^{n+1}$, which may be written as

$$a_{\mathbf{s}'} = \sum_{k=1}^{n+1} \frac{2s_k'}{3^k},$$

where $s_k' = s_k$ for $k = 1, \ldots, n$ and $s_{n+1}' = 1$. Moreover, $b_{\mathbf{s}'} = a_{\mathbf{s}'} + 1/3^{n+1}$, which proves that the formula for the endpoints holds for $n + 1$.

Definition 5.64. The Cantor set C is the intersection

$$C = \bigcap_{n=1}^{\infty} F_n,$$

of the nested sequence of sets (F_n) defined above.

The compactness of C follows immediately from its definition.

Theorem 5.65. The Cantor set is compact.

Proof. The Cantor set C is bounded, since it is a subset of $[0, 1]$. All the sets F_n are closed, because they are a finite union of closed intervals, so, from Proposition 5.20, their intersection is closed. It follows that C is closed and bounded, and Theorem 5.36 implies that it is compact. $\square$

The Cantor set C is clearly nonempty since the endpoints $a_\mathbf{s}$, $b_\mathbf{s}$ of $I_\mathbf{s}$ are contained in F_n for every finite binary sequence $\mathbf{s}$ and every $n \in \mathbb{N}$. These endpoints form a countably infinite set. What may be initially surprising is that there are uncountably many other points in C that are not endpoints. For example, $1/4$ has the infinite base three expansion $1/4 = 0.020202\ldots$, so it is not one of the endpoints, but, as we will show, it belongs to C because it has a base three expansion consisting entirely of 0's and 2's.

Let Σ be the set of binary sequences in Definition 1.29. The idea of the next theorem is that each binary sequence picks out a unique point of the Cantor set by telling us whether to choose the left or the right interval at each stage of the "middle-thirds" construction. For example, $1/4$ corresponds to the sequence $(0, 1, 0, 1, 0, 1, \dots)$, and we get it by alternately choosing left and right intervals.

Theorem 5.66. *The Cantor set has the same cardinality as Σ.*

Proof. We use the same notation as above. Let $\mathbf{s} = (s_1, s_2, \dots, s_k, \dots) \in \Sigma$, and define $\mathbf{s}_n = (s_1, s_2, \dots, s_n) \in \Sigma_n$. Then $(I_{\mathbf{s}_n})$ is a nested sequence of intervals such that $\operatorname{diam} I_{\mathbf{s}_n} = 1/3^n \to 0$ as $n \to \infty$. Since each $I_{\mathbf{s}_n}$ is a compact interval, Theorem 5.43 implies that there is a unique point

$$x \in \bigcap_{n=1}^{\infty} I_{\mathbf{s}_n} \subset C.$$

Thus, $\mathbf{s} \mapsto x$ defines a function $f : \Sigma \to C$. Furthermore, this function is one-to-one: if two sequences differ in the nth place, say, then the corresponding points in C belong to different intervals $I_{\mathbf{s}_n}$ at the nth stage of the construction, and therefore the points are different since the intervals are disjoint.

Conversely, if $x \in C$, then $x \in F_n$ for every $n \in \mathbb{N}$ and there is a unique $\mathbf{s}_n \in \Sigma_n$ such that $x \in I_{\mathbf{s}_n}$. The intervals $(I_{\mathbf{s}_n})$ are nested, so there is a unique sequence $\mathbf{s} = (s_1, s_2, \dots, s_k, \dots) \in \Sigma$, such that $\mathbf{s}_n = (s_1, s_2, \dots, s_n)$. It follows that $f : \Sigma \to C$ is onto, which proves the result. $\square$

The argument also shows that $x \in C$ if and only if it is a limit of left endpoints $a_{\mathbf{s}}$, meaning that

$$x = \sum_{k=1}^{\infty} \frac{2s_k}{3^k}, \qquad s_k = 0, 1.$$

In other words, $x \in C$ if and only if it has a base 3 expansion consisting entirely of 0's and 2's. Note that this condition does not exclude 1, which corresponds to the sequence $(1, 1, 1, 1, \dots)$ or "always pick the right interval," and

$$1 = 0.2222 \cdots = \sum_{k=1}^{\infty} \frac{2}{3^k}.$$

We may use Theorem 5.66, together with the Schröder-Bernstein theorem, to prove that Σ, $\mathcal{P}(\mathbb{N})$ and $\mathbb{R}$ have the same uncountable cardinality of the continuum. It follows, in particular, that the Cantor set has the same cardinality as $\mathbb{R}$, even though it appears, at first sight, to be a very sparse subset.

Theorem 5.67. *The set $\mathbb{R}$ of real numbers has the same cardinality as $\mathcal{P}(\mathbb{N})$.*

Proof. The inclusion map $f : C \to \mathbb{R}$, where $f(x) = x$, is one-to-one, so $C \lesssim \mathbb{R}$. From Theorem 5.66 and Corollary 1.48, we have $C \approx \Sigma \approx \mathcal{P}(\mathbb{N})$, so $\mathcal{P}(\mathbb{N}) \lesssim \mathbb{R}$.

Conversely, the map from real numbers to their Dedekind cuts, given by

$$g : \mathbb{R} \to \mathcal{P}(\mathbb{Q}), \qquad g : x \mapsto \{r \in \mathbb{Q} : r < x\},$$

is one-to-one, so $\mathbb{R} \lesssim \mathcal{P}(\mathbb{Q})$. Since $\mathbb{Q}$ is countably infinite, $\mathcal{P}(\mathbb{N}) \approx \mathcal{P}(\mathbb{Q})$, so $\mathbb{R} \lesssim \mathcal{P}(\mathbb{N})$. The conclusion then follows from Theorem 1.40. $\square$

Another proof of this theorem, which doesn't require the Schröder-Bernstein theorem, can be given by associating binary sequences in Σ with binary expansions of real numbers in $[0, 1]$:

$$h : (s_1, s_2, \ldots, s_k, \ldots) \mapsto \sum_{k=1}^{\infty} \frac{s_k}{2^k}.$$

Some real numbers, however, have two distinct binary expansion; e.g.,

$$\frac{1}{2} = 0.10000\ldots = 0.01111\ldots.$$

There are only countably many such numbers, so they do not affect the cardinality of $[0, 1]$, but they complicate the explicit construction of a one-to-one, onto map $f : \Sigma \to \mathbb{R}$ by this approach. An alternative method is to represent real numbers by continued fractions instead of binary expansions, but we won't describe these proofs in more detail here.

Limits of Functions

In this chapter, we define limits of functions and describe their properties.

6.1. Limits

We begin with the ϵ-δ definition of the limit of a function.

Definition 6.1. Let $f : A \to \mathbb{R}$, where $A \subset \mathbb{R}$, and suppose that $c \in \mathbb{R}$ is an accumulation point of A. Then

$$\lim_{x \to c} f(x) = L$$

if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$0 < |x - c| < \delta \text{ and } x \in A \text{ implies that } |f(x) - L| < \epsilon.$$

We also denote limits by the 'arrow' notation $f(x) \to L$ as $x \to c$, and often leave it to be implicitly understood that $x \in A$ is restricted to the domain of f. Note that it follows directly from the definition that

$$\lim_{x \to c} f(x) = L \quad \text{if and only if} \quad \lim_{x \to c} |f(x) - L| = 0.$$

In defining a limit as $x \to c$, we do not consider what happens when $x = c$, and a function needn't be defined at c for its limit to exist. This is the case, for example, when we define the derivative of a function as a limit of its difference quotients. Moreover, even if a function is defined at c and its limit as $x \to c$ exists, the value of the function need not equal the limit. In fact, the condition that $\lim_{x \to c} f(x) = f(c)$ defines the continuity of f at c. We study continuous functions in Chapter 7.

Example 6.2. Let $A = [0, \infty) \setminus \{9\}$ and define $f : A \to \mathbb{R}$ by

$$f(x) = \frac{x - 9}{\sqrt{x} - 3}.$$

We claim that
$$\lim_{x \to 9} f(x) = 6.$$
To prove this, let $\epsilon > 0$ be given. If $x \in A$, then $\sqrt{x} - 3 \neq 0$, and dividing this factor into the numerator we get $f(x) = \sqrt{x} + 3$. It follows that
$$|f(x) - 6| = \left|\sqrt{x} - 3\right| = \left|\frac{x - 9}{\sqrt{x} + 3}\right| \leq \frac{1}{3}|x - 9|.$$
Thus, if $\delta = 3\epsilon$, then $x \in A$ and $|x - 9| < \delta$ implies that $|f(x) - 6| < \epsilon$.

Like the limits of sequences, limits of functions are unique.

Proposition 6.3. The limit of a function is unique if it exists.

Proof. Suppose that $f : A \to \mathbb{R}$ and $c \in \mathbb{R}$ is an accumulation point of $A \subset \mathbb{R}$. Assume that
$$\lim_{x \to c} f(x) = L_1, \qquad \lim_{x \to c} f(x) = L_2$$
where $L_1, L_2 \in \mathbb{R}$. For every $\epsilon > 0$ there exist $\delta_1, \delta_2 > 0$ such that
$$0 < |x - c| < \delta_1 \text{ and } x \in A \text{ implies that } |f(x) - L_1| < \epsilon/2,$$
$$0 < |x - c| < \delta_2 \text{ and } x \in A \text{ implies that } |f(x) - L_2| < \epsilon/2.$$
Let $\delta = \min(\delta_1, \delta_2) > 0$. Then, since c is an accumulation point of A, there exists $x \in A$ such that $0 < |x - c| < \delta$. It follows that
$$|L_1 - L_2| \leq |L_1 - f(x)| + |f(x) - L_2| < \epsilon.$$
Since this holds for arbitrary $\epsilon > 0$, we must have $L_1 = L_2$. $\qquad\square$

Note that in this proof we used the requirement in the definition of a limit that c is an accumulation point of A. The limit definition would be vacuous if it was applied to a non-accumulation point, and in that case every $L \in \mathbb{R}$ would be a limit.

We can rephrase the ϵ-δ definition of limits in terms of neighborhoods. Recall from Definition 5.6 that a set $V \subset \mathbb{R}$ is a neighborhood of $c \in \mathbb{R}$ if $V \supset (c - \delta, c + \delta)$ for some $\delta > 0$, and $(c - \delta, c + \delta)$ is called a δ-neighborhood of c.

Definition 6.4. A set $U \subset \mathbb{R}$ is a punctured (or deleted) neighborhood of $c \in \mathbb{R}$ if $U \supset (c - \delta, c) \cup (c, c + \delta)$ for some $\delta > 0$. The set $(c - \delta, c) \cup (c, c + \delta)$ is called a punctured (or deleted) δ-neighborhood of c.

That is, a punctured neighborhood of c is a neighborhood of c with the point c itself removed.

Definition 6.5. Let $f : A \to \mathbb{R}$, where $A \subset \mathbb{R}$, and suppose that $c \in \mathbb{R}$ is an accumulation point of A. Then
$$\lim_{x \to c} f(x) = L$$
if and only if for every neighborhood V of L, there is a punctured neighborhood U of c such that
$$x \in A \cap U \text{ implies that } f(x) \in V.$$

This is essentially a rewording of the ϵ-δ definition. If Definition 6.1 holds and V is a neighborhood of L, then V contains an ϵ-neighborhood of L, so there is a punctured δ-neighborhood U of c such that f maps $U \cap A$ into V, which verifies Definition 6.5. Conversely, if Definition 6.5 holds and $\epsilon > 0$, then $V = (L - \epsilon, L + \epsilon)$ is a neighborhood of L, so there is a punctured neighborhood U of c such that f maps $U \cap A$ into V, and U contains a punctured δ-neighborhood of c, which verifies Definition 6.1.

The next theorem gives an equivalent sequential characterization of the limit.

Theorem 6.6. Let $f : A \to \mathbb{R}$, where $A \subset \mathbb{R}$, and suppose that $c \in \mathbb{R}$ is an accumulation point of A. Then

$$\lim_{x \to c} f(x) = L$$

if and only if

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

for every sequence (x_n) in A with $x_n \neq c$ for all $n \in \mathbb{N}$ such that

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

Proof. First assume that the limit exists and is equal to L. Suppose that (x_n) is any sequence in A with $x_n \neq c$ that converges to c, and let $\epsilon > 0$ be given. From Definition 6.1, there exists $\delta > 0$ such that $|f(x) - L| < \epsilon$ whenever $0 < |x - c| < \delta$, and since $x_n \to c$ there exists $N \in \mathbb{N}$ such that $0 < |x_n - c| < \delta$ for all $n > N$. It follows that $|f(x_n) - L| < \epsilon$ whenever $n > N$, so $f(x_n) \to L$ as $n \to \infty$.

To prove the converse, assume that the limit does not exist or is not equal to L. Then there is an $\epsilon_0 > 0$ such that for every $\delta > 0$ there is a point $x \in A$ with $0 < |x - c| < \delta$ but $|f(x) - L| \geq \epsilon_0$. Therefore, for every $n \in \mathbb{N}$ there is an $x_n \in A$ such that

$$0 < |x_n - c| < \frac{1}{n}, \qquad |f(x_n) - L| \geq \epsilon_0.$$

It follows that $x_n \neq c$ and $x_n \to c$, but $f(x_n) \not\to L$, so the sequential condition does not hold. This proves the result. $\qquad \square$

A non-existence proof for a limit directly from Definition 6.1 is often awkward. (One has to show that for every $L \in \mathbb{R}$ there exists $\epsilon_0 > 0$ such that for every $\delta > 0$ there exists $x \in A$ with $0 < |x - c| < \delta$ and $|f(x) - L| \geq \epsilon_0$.) The previous theorem gives a convenient way to show that a limit of a function does not exist.

Corollary 6.7. Suppose that $f : A \to \mathbb{R}$ and $c \in \mathbb{R}$ is an accumulation point of A. Then $\lim_{x \to c} f(x)$ does not exist if either of the following conditions holds:

(1) There are sequences (x_n), (y_n) in A with $x_n, y_n \neq c$ such that

$$\lim_{n \to \infty} x_n = \lim_{n \to \infty} y_n = c, \quad \text{but} \quad \lim_{n \to \infty} f(x_n) \neq \lim_{n \to \infty} f(y_n).$$

(2) There is a sequence (x_n) in A with $x_n \neq c$ such that $\lim_{n \to \infty} x_n = c$ but the sequence $(f(x_n))$ diverges.
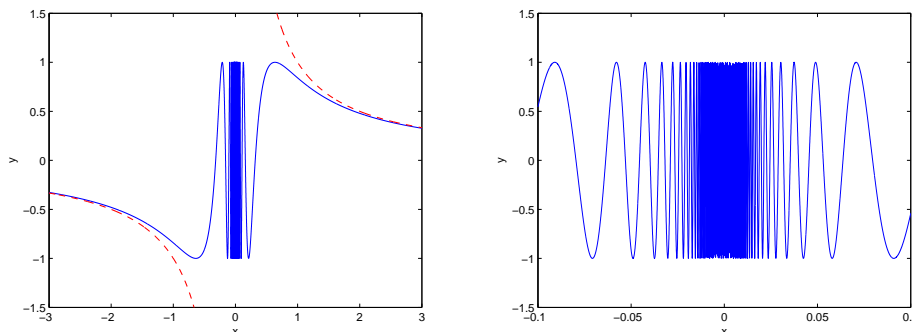
Figure 1. A plot of the function $y = \sin(1/x)$, with the hyperbola $y = 1/x$ shown in red, and a detail near the origin.

Example 6.8. Define the sign function $\operatorname{sgn} : \mathbb{R} \to \mathbb{R}$ by

$$\operatorname{sgn} x = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -1 & \text{if } x < 0, \end{cases}$$

Then the limit

$$\lim_{x \to 0} \operatorname{sgn} x$$

doesn't exist. To prove this, note that $(1/n)$ is a non-zero sequence such that $1/n \to 0$ and $\operatorname{sgn}(1/n) \to 1$ as $n \to \infty$, while $(-1/n)$ is a non-zero sequence such that $-1/n \to 0$ and $\operatorname{sgn}(-1/n) \to -1$ as $n \to \infty$. Since the sequences of sgn-values have different limits, Corollary 6.7 implies that the limit does not exist.

Example 6.9. The limit

$$\lim_{x \to 0} \frac{1}{x},$$

corresponding to the function $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ given by $f(x) = 1/x$, doesn't exist. For example, if (x_n) is the non-zero sequence given by $x_n = 1/n$, then $1/n \to 0$ but the sequence of values (n) diverges to ∞.

Example 6.10. The limit

$$\lim_{x \to 0} \sin\left(\frac{1}{x}\right),$$

corresponding to the function $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ given by $f(x) = \sin(1/x)$, doesn't exist. (See Figure 1.) For example, the non-zero sequences (x_n), (y_n) defined by

$$x_n = \frac{1}{2\pi n}, \qquad y_n = \frac{1}{2\pi n + \pi/2}$$

both converge to zero as $n \to \infty$, but the limits

$$\lim_{n \to \infty} f(x_n) = 0, \qquad \lim_{n \to \infty} f(y_n) = 1$$

are different.

Like sequences, functions must satisfy a boundedness condition if their limit is to exist. Before stating this condition, we define the supremum and infimum of a function, which are the supremum or infimum of its range.

Definition 6.11. If $f : A \to \mathbb{R}$ is a real-valued function, then

$$\sup_A f = \sup \left\{ f(x) : x \in A \right\}, \qquad \inf_A f = \inf \left\{ f(x) : x \in A \right\}.$$

A function is bounded if its range is bounded.

Definition 6.12. If $f : A \to \mathbb{R}$, then f is bounded from above if $\sup_A f$ is finite, bounded from below if $\inf_A f$ is finite, and bounded if both are finite. A function that is not bounded is said to be unbounded.

Example 6.13. If $f : [0, 2] \to \mathbb{R}$ is defined by $f(x) = x^2$, then

$$\sup_{[0,2]} f = 4, \qquad \inf_{[0,2]} f = 0,$$

so f is bounded.

Example 6.14. If $f : (0, 1] \to \mathbb{R}$ is defined by $f(x) = 1/x$, then

$$\sup_{(0,1]} f = \infty, \qquad \inf_{(0,1]} f = 1,$$

so f is bounded from below, not bounded from above, and unbounded. Note that if we extend f to a function $g : [0, 1] \to \mathbb{R}$ by defining, for example,

$$g(x) = \begin{cases} 1/x & \text{if } 0 < x \leq 1, \\ 0 & \text{if } x = 0, \end{cases}$$

then g is still unbounded on $[0, 1]$.

Equivalently, a function $f : A \to \mathbb{R}$ is bounded if $\sup_A |f|$ is finite, meaning that there exists $M \geq 0$ such that

$$|f(x)| \leq M \text{ for every } x \in A.$$

If $B \subset A$, then we say that f is bounded from above on B if $\sup_B f$ is finite, with similar terminology for bounded from below on B, and bounded on B.

Example 6.15. The function $f : (0, 1] \to \mathbb{R}$ defined by $f(x) = 1/x$ is unbounded, but it is bounded on every interval $[\delta, 1]$ with $0 < \delta < 1$. The function $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = x^2$ is unbounded, but it is bounded on every finite interval $[a, b]$.

We also introduce a notion of being bounded near a point.

Definition 6.16. Suppose that $f : A \to \mathbb{R}$ and c is an accumulation point of A. Then f is locally bounded at c if there is a neighborhood U of c such that f is bounded on $A \cap U$.

Example 6.17. The function $f : (0, 1] \to \mathbb{R}$ defined by $f(x) = 1/x$ is locally bounded at every $0 < c \leq 1$, but it is not locally bounded at 0.

Proposition 6.18. Suppose that $f : A \to \mathbb{R}$ and c is an accumulation point of A. If $\lim_{x \to c} f(x)$ exists, then f is locally bounded at c.

Proof. Let $\lim_{x \to c} f(x) = L$. Taking $\epsilon = 1$ in the definition of the limit, we get that there exists a $\delta > 0$ such that

$$0 < |x - c| < \delta \text{ and } x \in A \text{ implies that } |f(x) - L| < 1.$$

Let $U = (c - \delta, c + \delta)$. If $x \in A \cap U$ and $x \neq c$, then

$$|f(x)| \leq |f(x) - L| + |L| < 1 + |L|,$$

so f is bounded on $A \cap U$. (If $c \in A$, then $|f| \leq \max\{1 + |L|, |f(c)|\}$ on $A \cap U$.) $\quad \square$

As for sequences, boundedness is a necessary but not sufficient condition for the existence of a limit.

Example 6.19. The limit

$$\lim_{x \to 0} \frac{1}{x},$$

considered in Example 6.9 doesn't exist because the function $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ given by $f(x) = 1/x$ is not locally bounded at 0.

Example 6.20. The function $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ defined by

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

is bounded, but $\lim_{x \to 0} f(x)$ doesn't exist.

6.2. Left, right, and infinite limits

We can define other kinds of limits in an obvious way. We list some of them here and give examples, whose proofs are left as an exercise. All these definitions can be combined in various ways and have obvious equivalent sequential characterizations.

Definition 6.21 (Right and left limits). Let $f : A \to \mathbb{R}$, where $A \subset \mathbb{R}$. If $c \in \mathbb{R}$ is an accumulation point of $\{x \in A : x > c\}$, then f has the right limit

$$\lim_{x \to c^+} f(x) = L,$$

if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$c < x < c + \delta \text{ and } x \in A \text{ implies that } |f(x) - L| < \epsilon.$$

If $c \in \mathbb{R}$ is an accumulation point of $\{x \in A : x < c\}$, then f has the left limit

$$\lim_{x \to c^-} f(x) = L,$$

if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$c - \delta < x < c \text{ and } x \in A \text{ implies that } |f(x) - L| < \epsilon.$$

Equivalently, the right limit of f is the limit of the restriction $f|_{A^+}$ of f to the set $A^+ = \{x \in A : x > c\}$,

$$\lim_{x \to c^+} f(x) = \lim_{x \to c} f|_{A^+}(x),$$

and analogously for the left limit.

Example 6.22. For the sign function in Example 6.8, we have
$$\lim_{x\to 0^+} \operatorname{sgn} x = 1, \qquad \lim_{x\to 0^-} \operatorname{sgn} x = -1,$$
although the corresponding limit does not exist.

The existence and equality of the left and right limits implies the existence of the limit.

Proposition 6.23. Suppose that $f : A \to \mathbb{R}$, where $A \subset \mathbb{R}$, and $c \in \mathbb{R}$ is an accumulation point of both $\{x \in A : x > c\}$ and $\{x \in A : x < c\}$. Then
$$\lim_{x\to c} f(x) = L$$
if and only if
$$\lim_{x\to c^+} f(x) = \lim_{x\to c^-} f(x) = L.$$

Proof. It follows immediately from the definitions that the existence of the limit implies the existence of the left and right limits with the same value. Conversely, if both left and right limits exists and are equal to L, then given $\epsilon > 0$, there exist $\delta_1 > 0$ and $\delta_2 > 0$ such that
$$c - \delta_1 < x < c \text{ and } x \in A \text{ implies that } |f(x) - L| < \epsilon,$$
$$c < x < c + \delta_2 \text{ and } x \in A \text{ implies that } |f(x) - L| < \epsilon.$$
Choosing $\delta = \min(\delta_1, \delta_2) > 0$, we get that
$$|x - c| < \delta \text{ and } x \in A \text{ implies that } |f(x) - L| < \epsilon,$$
which show that the limit exists. $\qquad\square$

Next we introduce some convenient definitions for various kinds of limits involving infinity. We emphasize that ∞ and $-\infty$ are not real numbers (what is $\sin\infty$, for example?) and all these definition have precise translations into statements that involve only real numbers.

Definition 6.24 (Limits as $x \to \pm\infty$). Let $f : A \to \mathbb{R}$, where $A \subset \mathbb{R}$. If A is not bounded from above, then
$$\lim_{x\to\infty} f(x) = L$$
if for every $\epsilon > 0$ there exists an $M \in \mathbb{R}$ such that
$$x > M \text{ and } x \in A \text{ implies that } |f(x) - L| < \epsilon.$$
If A is not bounded from below, then
$$\lim_{x\to-\infty} f(x) = L$$
if for every $\epsilon > 0$ there exists an $m \in \mathbb{R}$ such that
$$x < m \text{ and } x \in A \text{ implies that } |f(x) - L| < \epsilon.$$

Sometimes we write $+\infty$ instead of ∞ to indicate that it denotes arbitrarily large, positive values, while $-\infty$ denotes arbitrarily large, negative values.

It follows from the definitions that
$$\lim_{x\to\infty} f(x) = \lim_{t\to 0^+} f\left(\frac{1}{t}\right), \qquad \lim_{x\to-\infty} f(x) = \lim_{t\to 0^-} f\left(\frac{1}{t}\right),$$

and it is often useful to convert one of these limits into the other.

Example 6.25. We have

$$\lim_{x \to \infty} \frac{x}{\sqrt{1+x^2}} = 1, \qquad \lim_{x \to -\infty} \frac{x}{\sqrt{1+x^2}} = -1.$$

Definition 6.26 (Divergence to $\pm\infty$). Let $f : A \to \mathbb{R}$, where $A \subset \mathbb{R}$, and suppose that $c \in \mathbb{R}$ is an accumulation point of A. Then

$$\lim_{x \to c} f(x) = \infty$$

if for every $M \in \mathbb{R}$ there exists a $\delta > 0$ such that

$$0 < |x - c| < \delta \text{ and } x \in A \text{ implies that } f(x) > M,$$

and

$$\lim_{x \to c} f(x) = -\infty$$

if for every $m \in \mathbb{R}$ there exists a $\delta > 0$ such that

$$0 < |x - c| < \delta \text{ and } x \in A \text{ implies that } f(x) < m.$$

The notation $\lim_{x \to c} f(x) = \pm\infty$ is simply shorthand for the property stated in this definition; it does not mean that the limit exists, and we say that f diverges to $\pm\infty$.

Example 6.27. We have

$$\lim_{x \to 0} \frac{1}{x^2} = \infty, \qquad \lim_{x \to \infty} \frac{1}{x^2} = 0.$$

Example 6.28. We have

$$\lim_{x \to 0^+} \frac{1}{x} = \infty, \qquad \lim_{x \to 0^-} \frac{1}{x} = -\infty.$$

How would you define these statements precisely? Note that

$$\lim_{x \to 0} \frac{1}{x} \neq \pm\infty,$$

since $1/x$ takes arbitrarily large positive (if $x > 0$) and negative (if $x < 0$) values in every two-sided neighborhood of 0.

Example 6.29. None of the limits

$$\lim_{x \to 0^+} \frac{1}{x} \sin\left(\frac{1}{x}\right), \qquad \lim_{x \to 0^-} \frac{1}{x} \sin\left(\frac{1}{x}\right), \qquad \lim_{x \to 0} \frac{1}{x} \sin\left(\frac{1}{x}\right)$$

is ∞ or $-\infty$, since $(1/x) \sin(1/x)$ oscillates between arbitrarily large positive and negative values in every one-sided or two-sided neighborhood of 0.

Example 6.30. We have

$$\lim_{x \to \infty} \left(\frac{1}{x} - x^3\right) = -\infty, \qquad \lim_{x \to -\infty} \left(\frac{1}{x} - x^3\right) = \infty.$$

How would you define these statements precisely and prove them?

6.3. Properties of limits

The properties of limits of functions follow from the corresponding properties of sequences and the sequential characterization of the limit in Theorem 6.6. We can also prove them directly from the ϵ-δ definition of the limit.

6.3.1. Order properties. As for limits of sequences, limits of functions preserve (non-strict) inequalities.

Theorem 6.31. Suppose that $f, g : A \to \mathbb{R}$ and c is an accumulation point of A. If

$$f(x) \leq g(x) \qquad \text{for all } x \in A,$$

and $\lim_{x \to c} f(x)$, $\lim_{x \to c} g(x)$ exist, then

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

Proof. Let

$$\lim_{x \to c} f(x) = L, \qquad \lim_{x \to c} g(x) = M.$$

Suppose for contradiction that $L > M$, and let

$$\epsilon = \frac{1}{2}(L - M) > 0.$$

From the definition of the limit, there exist $\delta_1, \delta_2 > 0$ such that

$$|f(x) - L| < \epsilon \qquad \text{if } x \in A \text{ and } 0 < |x - c| < \delta_1,$$
$$|g(x) - M| < \epsilon \qquad \text{if } x \in A \text{ and } 0 < |x - c| < \delta_2.$$

Let $\delta = \min(\delta_1, \delta_2)$. Since c is an accumulation point of A, there exists $x \in A$ such that $0 < |x - a| < \delta$, and it follows that

$$f(x) - g(x) = [f(x) - L] + L - M + [M - g(x)]$$
$$> L - M - 2\epsilon$$
$$> 0,$$

which contradicts the assumption that $f(x) \leq g(x)$. $\qquad \square$

Finally, we state a useful "sandwich" or "squeeze" criterion for the existence of a limit.

Theorem 6.32. Suppose that $f, g, h : A \to \mathbb{R}$ and c is an accumulation point of A. If

$$f(x) \leq g(x) \leq h(x) \qquad \text{for all } x \in A$$

and

$$\lim_{x \to c} f(x) = \lim_{x \to c} h(x) = L,$$

then the limit of $g(x)$ as $x \to c$ exists and

$$\lim_{x \to c} g(x) = L.$$

We leave the proof as an exercise. We often use this result, without comment, in the following way: If

$$0 \leq f(x) \leq g(x) \quad \text{or} \quad |f(x)| \leq g(x)$$

and $g(x) \to 0$ as $x \to c$, then $f(x) \to 0$ as $x \to c$.

It is essential for the bounding functions f, h in Theorem 6.32 to have the same limit.

Example 6.33. We have

$$-1 \leq \sin\left(\frac{1}{x}\right) \leq 1 \qquad \text{for all } x \neq 0$$

and

$$\lim_{x \to 0}(-1) = -1, \qquad \lim_{x \to 0} 1 = 1,$$

but

$$\lim_{x \to 0} \sin\left(\frac{1}{x}\right) \qquad \text{does not exist.}$$

6.3.2. Algebraic properties. Limits of functions respect algebraic operations.

Theorem 6.34. Suppose that $f, g : A \to \mathbb{R}$, c is an accumulation point of A, and the limits

$$\lim_{x \to c} f(x) = L, \qquad \lim_{x \to c} g(x) = M$$

exist. Then

$$\lim_{x \to c} kf(x) = kL \qquad \text{for every } k \in \mathbb{R},$$
$$\lim_{x \to c} [f(x) + g(x)] = L + M,$$
$$\lim_{x \to c} [f(x)g(x)] = LM,$$
$$\lim_{x \to c} \frac{f(x)}{g(x)} = \frac{L}{M} \qquad \text{if } M \neq 0.$$

Proof. We prove the results for sums and products from the definition of the limit, and leave the remaining proofs as an exercise. All of the results also follow from the corresponding results for sequences.

First, we consider the limit of $f + g$. Given $\epsilon > 0$, choose δ_1, δ_2 such that

$$0 < |x - c| < \delta_1 \text{ and } x \in A \text{ implies that } |f(x) - L| < \epsilon/2,$$
$$0 < |x - c| < \delta_2 \text{ and } x \in A \text{ implies that } |g(x) - M| < \epsilon/2,$$

and let $\delta = \min(\delta_1, \delta_2) > 0$. Then $0 < |x - c| < \delta$ implies that

$$|f(x) + g(x) - (L + M)| \leq |f(x) - L| + |g(x) - M| < \epsilon,$$

which proves that $\lim(f + g) = \lim f + \lim g$.

To prove the result for the limit of the product, first note that from the local boundedness of functions with a limit (Proposition 6.18) there exists $\delta_0 > 0$ and

$K > 0$ such that $|g(x)| \leq K$ for all $x \in A$ with $0 < |x - c| < \delta_0$. Choose $\delta_1, \delta_2 > 0$ such that

$$0 < |x - c| < \delta_1 \text{ and } x \in A \text{ implies that } |f(x) - L| < \epsilon/(2K),$$
$$0 < |x - c| < \delta_2 \text{ and } x \in A \text{ implies that } |g(x) - M| < \epsilon/(2|L| + 1).$$

Let $\delta = \min(\delta_0, \delta_1, \delta_2) > 0$. Then for $0 < |x - c| < \delta$ and $x \in A$,

$$\begin{aligned}
|f(x)g(x) - LM| &= |(f(x) - L)\, g(x) + L\,(g(x) - M)| \\
&\leq |f(x) - L|\, |g(x)| + |L|\, |g(x) - M| \\
&< \frac{\epsilon}{2K} \cdot K + |L| \cdot \frac{\epsilon}{2|L| + 1} \\
&< \epsilon,
\end{aligned}$$

which proves that $\lim(fg) = \lim f \lim g$. $\qquad\square$

Continuous Functions

In this chapter, we define continuous functions and study their properties.

7.1. Continuity

Continuous functions are functions that take nearby values at nearby points.

Definition 7.1. Let $f : A \to \mathbb{R}$, where $A \subset \mathbb{R}$, and suppose that $c \in A$. Then f is continuous at c if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|x - c| < \delta \text{ and } x \in A \text{ implies that } |f(x) - f(c)| < \epsilon.$$

A function $f : A \to \mathbb{R}$ is continuous if it is continuous at every point of A, and it is continuous on $B \subset A$ if it is continuous at every point in B.

The definition of continuity at a point may be stated in terms of neighborhoods as follows.

Definition 7.2. A function $f : A \to \mathbb{R}$, where $A \subset \mathbb{R}$, is continuous at $c \in A$ if for every neighborhood V of $f(c)$ there is a neighborhood U of c such that

$$x \in A \cap U \text{ implies that } f(x) \in V.$$

The ϵ-δ definition corresponds to the case when V is an ϵ-neighborhood of $f(c)$ and U is a δ-neighborhood of c.

Note that c must belong to the domain A of f in order to define the continuity of f at c. If c is an isolated point of A, then the continuity condition holds automatically since, for sufficiently small $\delta > 0$, the only point $x \in A$ with $|x - c| < \delta$ is $x = c$, and then $0 = |f(x) - f(c)| < \epsilon$. Thus, a function is continuous at every isolated point of its domain, and isolated points are not of much interest.

If $c \in A$ is an accumulation point of A, then the continuity of f at c is equivalent to the condition that

$$\lim_{x \to c} f(x) = f(c),$$

meaning that the limit of f as $x \to c$ exists and is equal to the value of f at c.

Example 7.3. If $f : (a, b) \to \mathbb{R}$ is defined on an open interval, then f is continuous on (a, b) if and only if

$$\lim_{x \to c} f(x) = f(c) \qquad \text{for every } a < c < b$$

since every point of (a, b) is an accumulation point.

Example 7.4. If $f : [a, b] \to \mathbb{R}$ is defined on a closed, bounded interval, then f is continuous on $[a, b]$ if and only if

$$\lim_{x \to c} f(x) = f(c) \qquad \text{for every } a < c < b,$$
$$\lim_{x \to a^+} f(x) = f(a), \qquad \lim_{x \to b^-} f(x) = f(b).$$

Example 7.5. Suppose that

$$A = \left\{ 0, 1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}, \dots \right\}$$

and $f : A \to \mathbb{R}$ is defined by

$$f(0) = y_0, \qquad f\left(\frac{1}{n}\right) = y_n$$

for some values $y_0, y_n \in \mathbb{R}$. Then $1/n$ is an isolated point of A for every $n \in \mathbb{N}$, so f is continuous at $1/n$ for every choice of y_n. The remaining point $0 \in A$ is an accumulation point of A, and the condition for f to be continuous at 0 is that

$$\lim_{n \to \infty} y_n = y_0.$$

As for limits, we can give an equivalent sequential definition of continuity, which follows immediately from Theorem 6.6.

Theorem 7.6. If $f : A \to \mathbb{R}$ and $c \in A$ is an accumulation point of A, then f is continuous at c if and only if

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

for every sequence (x_n) in A such that $x_n \to c$ as $n \to \infty$.

In particular, f is discontinuous at $c \in A$ if there is sequence (x_n) in the domain A of f such that $x_n \to c$ but $f(x_n) \not\to f(c)$.

Let's consider some examples of continuous and discontinuous functions to illustrate the definition.

Example 7.7. The function $f : [0, \infty) \to \mathbb{R}$ defined by $f(x) = \sqrt{x}$ is continuous on $[0, \infty)$. To prove that f is continuous at $c > 0$, we note that for $0 \le x < \infty$,

$$|f(x) - f(c)| = \left| \sqrt{x} - \sqrt{c} \right| = \left| \frac{x - c}{\sqrt{x} + \sqrt{c}} \right| \le \frac{1}{\sqrt{c}} |x - c|,$$

so given $\epsilon > 0$, we can choose $\delta = \sqrt{c}\epsilon > 0$ in the definition of continuity. To prove that f is continuous at 0, we note that if $0 \le x < \delta$ where $\delta = \epsilon^2 > 0$, then

$$|f(x) - f(0)| = \sqrt{x} < \epsilon.$$

Example 7.8. The function $\sin : \mathbb{R} \to \mathbb{R}$ is continuous on $\mathbb{R}$. To prove this, we use the trigonometric identity for the difference of sines and the inequality $|\sin x| \leq |x|$:

$$
\begin{aligned}
|\sin x - \sin c| &= \left| 2 \cos \left(\frac{x+c}{2} \right) \sin \left(\frac{x-c}{2} \right) \right| \\
&\leq 2 \left| \sin \left(\frac{x-c}{2} \right) \right| \\
&\leq |x - c|.
\end{aligned}
$$

It follows that we can take $\delta = \epsilon$ in the definition of continuity for every $c \in \mathbb{R}$.

Example 7.9. The sign function $\mathrm{sgn} : \mathbb{R} \to \mathbb{R}$, defined by

$$
\mathrm{sgn}\, x = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -1 & \text{if } x < 0, \end{cases}
$$

is not continuous at 0 since $\lim_{x \to 0} \mathrm{sgn}\, x$ does not exist (see Example 6.8). The left and right limits of sgn at 0,

$$
\lim_{x \to 0^-} f(x) = -1, \qquad \lim_{x \to 0^+} f(x) = 1,
$$

do exist, but they are unequal. We say that f has a jump discontinuity at 0.

Example 7.10. The function $f : \mathbb{R} \to \mathbb{R}$ defined by

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

is not continuous at 0 since $\lim_{x \to 0} f(x)$ does not exist (see Example 6.9). The left and right limits of f at 0 do not exist either, and we say that f has an essential discontinuity at 0.

Example 7.11. The function $f : \mathbb{R} \to \mathbb{R}$ defined by

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

is continuous at $c \neq 0$ (see Example 7.21 below) but discontinuous at 0 because $\lim_{x \to 0} f(x)$ does not exist (see Example 6.10).

Example 7.12. The function $f : \mathbb{R} \to \mathbb{R}$ defined by

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

is continuous at every point of $\mathbb{R}$. (See Figure 1.) The continuity at $c \neq 0$ is proved in Example 7.22 below. To prove continuity at 0, note that for $x \neq 0$,

$$
|f(x) - f(0)| = |x \sin(1/x)| \leq |x|,
$$

so $f(x) \to f(0)$ as $x \to 0$. If we had defined $f(0)$ to be any value other than 0, then f would not be continuous at 0. In that case, f would have a removable discontinuity at 0.
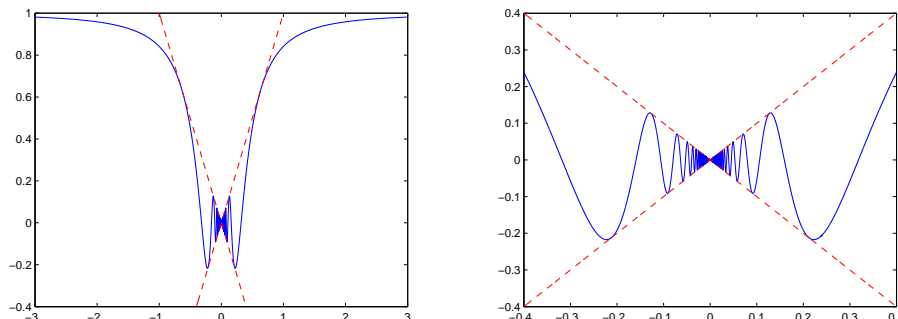
Figure 1. A plot of the function $y = x \sin(1/x)$ and a detail near the origin with the lines $y = \pm x$ shown in red.

Example 7.13. The Dirichlet function $f : \mathbb{R} \to \mathbb{R}$ defined by

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

is discontinuous at every $c \in \mathbb{R}$. If $c \notin \mathbb{Q}$, choose a sequence (x_n) of rational numbers such that $x_n \to c$ (possible since $\mathbb{Q}$ is dense in $\mathbb{R}$). Then $x_n \to c$ and $f(x_n) \to 1$ but $f(c) = 0$. If $c \in \mathbb{Q}$, choose a sequence (x_n) of irrational numbers such that $x_n \to c$; for example if $c = p/q$, we can take

$$x_n = \frac{p}{q} + \frac{\sqrt{2}}{n},$$

since $x_n \in \mathbb{Q}$ would imply that $\sqrt{2} \in \mathbb{Q}$. Then $x_n \to c$ and $f(x_n) \to 0$ but $f(c) = 1$. Alternatively, by taking a rational sequence (x_n) and an irrational sequence $(\tilde{x}_n)$ that converge to c, we can see that $\lim_{x \to c} f(x)$ does not exist for any $c \in \mathbb{R}$.

Example 7.14. The Thomae function $f : \mathbb{R} \to \mathbb{R}$ is defined by

$$f(x) = \begin{cases} 1/q & \text{if } x = p/q \in \mathbb{Q} \text{ where } p \text{ and } q > 0 \text{ are relatively prime,} \\ 0 & \text{if } x \notin \mathbb{Q} \text{ or } x = 0. \end{cases}$$

Figure 2 shows the graph of f on $[0, 1]$. The Thomae function is continuous at 0 and every irrational number and discontinuous at every nonzero rational number.

To prove this claim, first suppose that $x = p/q \in \mathbb{Q} \setminus \{0\}$ is rational and nonzero. Then $f(x) = 1/q > 0$, but for every $\delta > 0$, the interval $(x - \delta, x + \delta)$ contains irrational points y such that $f(y) = 0$ and $|f(x) - f(y)| = 1/q$. The definition of continuity therefore fails if $0 < \epsilon \le 1/q$, and f is discontinuous at x.

Second, suppose that $x \notin \mathbb{Q}$ is irrational. Given $\epsilon > 0$, choose $n \in \mathbb{N}$ such that $1/n < \epsilon$. There are finitely many rational numbers $r = p/q$ in the interval $(x - 1, x + 1)$ with p, q relatively prime and $1 \le q \le n$; we list them as $\{r_1, r_2, \ldots, r_m\}$. Choose

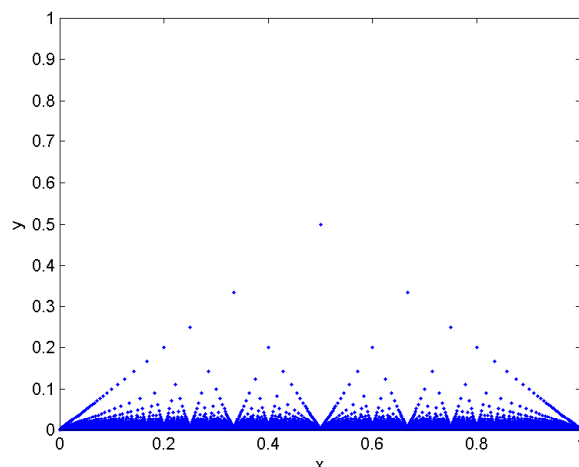$$\delta = \min\{|x - r_k| : k = 1, 2, \ldots, n\}$$

Figure 2. A plot of the Thomae function in Example 7.14 on $[0, 1]$

to be the distance of x to the closest such rational number. Then $\delta > 0$ since $x \notin \mathbb{Q}$. Furthermore, if $|x - y| < \delta$, then either y is irrational and $f(y) = 0$, or $y = p/q$ in lowest terms with $q > n$ and $f(y) = 1/q < 1/n < \epsilon$. In either case, $|f(x) - f(y)| = |f(y)| < \epsilon$, which proves that f is continuous at $x \notin \mathbb{Q}$.

The continuity of f at 0 follows immediately from the inequality $0 \le f(x) \le |x|$ for all $x \in \mathbb{R}$.

We give a rough classification of discontinuities of a function $f : A \to \mathbb{R}$ at an accumulation point $c \in A$ as follows.

(1) *Removable discontinuity*: $\lim_{x \to c} f(x) = L$ exists but $L \ne f(c)$, in which case we can make f continuous at c by redefining $f(c) = L$ (see Example 7.12).

(2) *Jump discontinuity*: $\lim_{x \to c} f(x)$ doesn't exist, but both the left and right limits $\lim_{x \to c^-} f(x)$, $\lim_{x \to c^+} f(x)$ exist and are different (see Example 7.9).

(3) *Essential discontinuity*: $\lim_{x \to c} f(x)$ doesn't exist and at least one of the left or right limits $\lim_{x \to c^-} f(x)$, $\lim_{x \to c^+} f(x)$ doesn't exist (see Examples 7.10, 7.11, 7.13).

7.2. Properties of continuous functions

The basic properties of continuous functions follow from those of limits.

Theorem 7.15. If $f, g : A \to \mathbb{R}$ are continuous at $c \in A$ and $k \in \mathbb{R}$, then kf, $f + g$, and fg are continuous at c. Moreover, if $g(c) \ne 0$ then f/g is continuous at c.

Proof. This result follows immediately Theorem 6.34. $\qquad \square$

A polynomial function is a function $P : \mathbb{R} \to \mathbb{R}$ of the form

$$P(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

where $a_0, a_1, a_2, \ldots, a_n$ are real coefficients. A rational function R is a ratio of polynomials P, Q

$$R(x) = \frac{P(x)}{Q(x)}.$$

The domain of R is the set of points in $\mathbb{R}$ such that $Q \neq 0$.

Corollary 7.16. Every polynomial function is continuous on $\mathbb{R}$ and every rational function is continuous on its domain.

Proof. The constant function $f(x) = 1$ and the identity function $g(x) = x$ are continuous on $\mathbb{R}$. Repeated application of Theorem 7.15 for scalar multiples, sums, and products implies that every polynomial is continuous on $\mathbb{R}$. It also follows that a rational function $R = P/Q$ is continuous at every point where $Q \neq 0$.   □

Example 7.17. The function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \frac{x + 3x^3 + 5x^5}{1 + x^2 + x^4}$$

is continuous on $\mathbb{R}$ since it is a rational function whose denominator never vanishes.

In addition to forming sums, products and quotients, another way to build up more complicated functions from simpler functions is by composition. We recall that the composition $g \circ f$ of functions f, g is defined by $(g \circ f)(x) = g\,(f(x))$. The next theorem states that the composition of continuous functions is continuous; note carefully the points at which we assume f and g are continuous.

Theorem 7.18. Let $f : A \to \mathbb{R}$ and $g : B \to \mathbb{R}$ where $f(A) \subset B$. If f is continuous at $c \in A$ and g is continuous at $f(c) \in B$, then $g \circ f : A \to \mathbb{R}$ is continuous at c.

Proof. Let $\epsilon > 0$ be given. Since g is continuous at $f(c)$, there exists $\eta > 0$ such that

$$|y - f(c)| < \eta \text{ and } y \in B \text{ implies that } |g(y) - g\,(f(c))| < \epsilon.$$

Next, since f is continuous at c, there exists $\delta > 0$ such that

$$|x - c| < \delta \text{ and } x \in A \text{ implies that } |f(x) - f(c)| < \eta.$$

Combing these inequalities, we get that

$$|x - c| < \delta \text{ and } x \in A \text{ implies that } |g\,(f(x)) - g\,(f(c))| < \epsilon,$$

which proves that $g \circ f$ is continuous at c.   □

Corollary 7.19. Let $f : A \to \mathbb{R}$ and $g : B \to \mathbb{R}$ where $f(A) \subset B$. If f is continuous on A and g is continuous on $f(A)$, then $g \circ f$ is continuous on A.

Example 7.20. The function

$$f(x) = \begin{cases} 1/\sin x & \text{if } x \neq n\pi \text{ for } n \in \mathbb{Z}, \\ 0 & \text{if } x = n\pi \text{ for } n \in \mathbb{Z} \end{cases}$$

is continuous on $\mathbb{R} \setminus \{n\pi : n \in \mathbb{Z}\}$, since it is the composition of $x \mapsto \sin x$, which is continuous on $\mathbb{R}$, and $y \mapsto 1/y$, which is continuous on $\mathbb{R} \setminus \{0\}$, and $\sin x \neq 0$ when $x \neq n\pi$. It is discontinuous at $x = n\pi$ because it is not locally bounded at those points.

Example 7.21. The function

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

is continuous on $\mathbb{R} \setminus \{0\}$, since it is the composition of $x \mapsto 1/x$, which is continuous on $\mathbb{R} \setminus \{0\}$, and $y \mapsto \sin y$, which is continuous on $\mathbb{R}$.

Example 7.22. The function

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

is continuous on $\mathbb{R} \setminus \{0\}$ since it is a product of functions that are continuous on $\mathbb{R} \setminus \{0\}$. As shown in Example 7.12, f is also continuous at 0, so f is continuous on $\mathbb{R}$.

7.3. Uniform continuity

Uniform continuity is a subtle but powerful strengthening of continuity.

Definition 7.23. Let $f : A \to \mathbb{R}$, where $A \subset \mathbb{R}$. Then f is uniformly continuous on A if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|x - y| < \delta \text{ and } x, y \in A \text{ implies that } |f(x) - f(y)| < \epsilon.$$

The key point of this definition is that δ depends only on ϵ, not on x, y. A uniformly continuous function on A is continuous at every point of A, but the converse is not true.

To explain this point in more detail, note that if a function f is continuous on A, then given $\epsilon > 0$ and $c \in A$, there exists $\delta(\epsilon, c) > 0$ such that

$$|x - c| < \delta(\epsilon, c) \text{ and } x \in A \text{ implies that } |f(x) - f(c)| < \epsilon.$$

If for some $\epsilon_0 > 0$ we have

$$\inf_{c \in A} \delta(\epsilon_0, c) = 0$$

however we choose $\delta(\epsilon_0, c) > 0$ in the definition of continuity, then no $\delta_0(\epsilon_0) > 0$ depending only on ϵ_0 works simultaneously for every $c \in A$. In that case, the function is continuous on A but not uniformly continuous.

Before giving some examples, we state a sequential condition for uniform continuity to fail.

Proposition 7.24. A function $f : A \to \mathbb{R}$ is not uniformly continuous on A if and only if there exists $\epsilon_0 > 0$ and sequences (x_n), (y_n) in A such that

$$\lim_{n \to \infty} |x_n - y_n| = 0 \text{ and } |f(x_n) - f(y_n)| \geq \epsilon_0 \text{ for all } n \in \mathbb{N}.$$

Proof. If f is not uniformly continuous, then there exists $\epsilon_0 > 0$ such that for every $\delta > 0$ there are points $x, y \in A$ with $|x - y| < \delta$ and $|f(x) - f(y)| \geq \epsilon_0$. Choosing $x_n, y_n \in A$ to be any such points for $\delta = 1/n$, we get the required sequences.

Conversely, if the sequential condition holds, then for every $\delta > 0$ there exists $n \in \mathbb{N}$ such that $|x_n - y_n| < \delta$ and $|f(x_n) - f(y_n)| \geq \epsilon_0$. It follows that the uniform

continuity condition in Definition 7.23 cannot hold for any $\delta > 0$ if $\epsilon = \epsilon_0$, so f is not uniformly continuous.   □

Example 7.25. Example 7.8 shows that the sine function is uniformly continuous on $\mathbb{R}$, since we can take $\delta = \epsilon$ for every $x, y \in \mathbb{R}$.

Example 7.26. Define $f : [0, 1] \to \mathbb{R}$ by $f(x) = x^2$. Then f is uniformly continuous on $[0, 1]$. To prove this, note that for all $x, y \in [0, 1]$ we have

$$\left| x^2 - y^2 \right| = |x + y| \, |x - y| \leq 2|x - y|,$$

so we can take $\delta = \epsilon/2$ in the definition of uniform continuity. Similarly, $f(x) = x^2$ is uniformly continuous on any bounded set.

Example 7.27. The function $f(x) = x^2$ is continuous but not uniformly continuous on $\mathbb{R}$. We have already proved that f is continuous on $\mathbb{R}$ (it's a polynomial). To prove that f is not uniformly continuous, let

$$x_n = n, \qquad y_n = n + \frac{1}{n}.$$

Then

$$\lim_{n \to \infty} |x_n - y_n| = \lim_{n \to \infty} \frac{1}{n} = 0,$$

but

$$|f(x_n) - f(y_n)| = \left(n + \frac{1}{n} \right)^2 - n^2 = 2 + \frac{1}{n^2} \geq 2 \qquad \text{for every } n \in \mathbb{N}.$$

It follows from Proposition 7.24 that f is not uniformly continuous on $\mathbb{R}$. The problem here is that in order to prove the continuity of f at c, given $\epsilon > 0$ we need to make $\delta(\epsilon, c)$ smaller as c gets larger, and $\delta(\epsilon, c) \to 0$ as $c \to \infty$.

Example 7.28. The function $f : (0, 1] \to \mathbb{R}$ defined by

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

is continuous but not uniformly continuous on $(0, 1]$. It is continuous on $(0, 1]$ since it's a rational function whose denominator x is nonzero in $(0, 1]$. To prove that f is not uniformly continuous, we define $x_n, y_n \in (0, 1]$ for $n \in \mathbb{N}$ by

$$x_n = \frac{1}{n}, \qquad y_n = \frac{1}{n + 1}.$$

Then $|x_n - y_n| \to 0$ as $n \to \infty$, but

$$|f(x_n) - f(y_n)| = (n + 1) - n = 1 \qquad \text{for every } n \in \mathbb{N}.$$

It follows from Proposition 7.24 that f is not uniformly continuous on $(0, 1]$. The problem here is that given $\epsilon > 0$, we need to make $\delta(\epsilon, c)$ smaller as c gets closer to 0, and $\delta(\epsilon, c) \to 0$ as $c \to 0^+$.

The non-uniformly continuous functions in the last two examples were unbounded. However, even bounded continuous functions can fail to be uniformly continuous if they oscillate arbitrarily quickly.

Example 7.29. Define $f : (0,1] \to \mathbb{R}$ by

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

Then f is continuous on $(0,1]$ but it isn't uniformly continuous on $(0,1]$. To prove this, define $x_n, y_n \in (0,1]$ for $n \in \mathbb{N}$ by

$$x_n = \frac{1}{2n\pi}, \qquad y_n = \frac{1}{2n\pi + \pi/2}.$$

Then $|x_n - y_n| \to 0$ as $n \to \infty$, but

$$|f(x_n) - f(y_n)| = \sin\left(2n\pi + \frac{\pi}{2}\right) - \sin 2n\pi = 1 \qquad \text{for all } n \in \mathbb{N}.$$

It isn't a coincidence that these examples of non-uniformly continuous functions have domains that are either unbounded or not closed. We will prove in Section 7.5 that a continuous function on a compact set is uniformly continuous.

7.4. Continuous functions and open sets

Let $f : A \to \mathbb{R}$ be a function. Recall that if $B \subset A$, then the image of B under f is the set

$$f(B) = \{y \in \mathbb{R} : y = f(x) \text{ for some } x \in B\},$$

and if $C \subset \mathbb{R}$, then the inverse image, or preimage, of C under f is the set

$$f^{-1}(C) = \{x \in A : f(x) \in C\}.$$

The next example illustrates how open sets behave under continuous functions.

Example 7.30. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$, and consider the open interval $I = (1,4)$. Then both $f(I) = (1,16)$ and $f^{-1}(I) = (-2,-1) \cup (1,2)$ are open. There are two intervals in the inverse image of I because f is two-to-one on $f^{-1}(I)$. On the other hand, if $J = (-1,1)$, then

$$f(J) = [0,1), \qquad f^{-1}(J) = (-1,1),$$

so the inverse image of the open interval J is open, but the image is not.

Thus, a continuous function needn't map open sets to open sets. As we will show, however, the inverse image of an open set under a continuous function is always open. This property is the topological definition of a continuous function; it is a global definition in the sense that it is equivalent to the continuity of the function at every point of its domain.

Recall from Section 5.1 that a subset B of a set $A \subset \mathbb{R}$ is relatively open in A, or open in A, if $B = A \cap U$ where U is open in $\mathbb{R}$. Moreover, as stated in Proposition 5.13, B is relatively open in A if and only if every point $x \in B$ has a relative neighborhood $C = A \cap V$ such that $C \subset B$, where V is a neighborhood of x in $\mathbb{R}$.

Theorem 7.31. A function $f : A \to \mathbb{R}$ is continuous on A if and only if $f^{-1}(V)$ is open in A for every set V that is open in $\mathbb{R}$.

Proof. First assume that f is continuous on A, and suppose that $c \in f^{-1}(V)$. Then $f(c) \in V$ and since V is open it contains an ϵ-neighborhood

$$V_\epsilon\left(f(c)\right) = \left(f(c) - \epsilon, f(c) + \epsilon\right)$$

of $f(c)$. Since f is continuous at c, there is a δ-neighborhood

$$U_\delta(c) = (c - \delta, c + \delta)$$

of c such that

$$f\left(A \cap U_\delta(c)\right) \subset V_\epsilon\left(f(c)\right).$$

This statement just says that if $|x - c| < \delta$ and $x \in A$, then $|f(x) - f(c)| < \epsilon$. It follows that

$$A \cap U_\delta(c) \subset f^{-1}(V),$$

meaning that $f^{-1}(V)$ contains a relative neighborhood of c. Therefore $f^{-1}(V)$ is relatively open in A.

Conversely, assume that $f^{-1}(V)$ is open in A for every open V in $\mathbb{R}$, and let $c \in A$. Then the preimage of the ϵ-neighborhood $(f(c) - \epsilon, f(c) + \epsilon)$ is open in A, so it contains a relative δ-neighborhood $A \cap (c - \delta, c + \delta)$. It follows that $|f(x) - f(c)| < \epsilon$ if $|x - c| < \delta$ and $x \in A$, which means that f is continuous at c. $\qquad\square$

As one illustration of how we can use this result, we prove that continuous functions map intervals to intervals. (See Definition 5.62 for what we mean by an interval.) In view of Theorem 5.63, this is a special case of the fact that continuous functions map connected sets to connected sets (see Theorem 13.82).

Theorem 7.32. Suppose that $f : I \to \mathbb{R}$ is continuous and $I \subset \mathbb{R}$ is an interval. Then $f(I)$ is an interval.

Proof. Suppose that $f(I)$ is not a interval. Then, by Theorem 5.63, $f(I)$ is disconnected, and there exist nonempty, disjoint open sets U, V such that $U \cup V \supset f(I)$. Since f is continuous, $f^{-1}(U)$, $f^{-1}(V)$ are open from Theorem 7.31. Furthermore, $f^{-1}(U)$, $f^{-1}(V)$ are nonempty and disjoint, and $f^{-1}(U) \cup f^{-1}(V) = I$. It follows that I is disconnected and therefore I is not an interval from Theorem 5.63. This shows that if I is an interval, then $f(I)$ is also an interval. $\qquad\square$

We can also define open functions, or open mappings. Although they form an important class of functions, they aren't as fundamental as continuous functions.

Definition 7.33. An open mapping on a set $A \subset \mathbb{R}$ is a function $f : A \to \mathbb{R}$ such that $f(B)$ is open in $\mathbb{R}$ for every set $B \subset A$ that is open in A.

A continuous function needn't be open, but if $f : A \to \mathbb{R}$ is continuous and one-to-one, then $f^{-1} : f(A) \to \mathbb{R}$ is open.

Example 7.34. Example 7.30 shows that the square function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is not an open mapping on $\mathbb{R}$. On the other hand $f : [0, \infty) \to \mathbb{R}$ is open because it is one-to-one with a continuous inverse $f^{-1} : [0, \infty) \to \mathbb{R}$ given by $f^{-1}(x) = \sqrt{x}$.

7.5. Continuous functions on compact sets

Continuous functions on compact sets have especially nice properties. For example, they are bounded and attain their maximum and minimum values, and they are uniformly continuous. Since a closed, bounded interval is compact, these results apply, in particular, to continuous functions $f : [a, b] \to \mathbb{R}$.

First, we prove that the continuous image of a compact set in $\mathbb{R}$ is compact. This is a special case of the fact that continuous functions map compact sets to compact sets (see Theorem 13.82).

Theorem 7.35. If $K \subset \mathbb{R}$ is compact and $f : K \to \mathbb{R}$ is continuous, then $f(K)$ is compact.

Proof. We will give two proofs, one using sequences and the other using open covers.

We show that $f(K)$ is sequentially compact. Let (y_n) be a sequence in $f(K)$. Then $y_n = f(x_n)$ for some $x_n \in K$. Since K is compact, the sequence (x_n) has a convergent subsequence (x_{n_i}) such that

$$\lim_{i \to \infty} x_{n_i} = x$$

where $x \in K$. Since f is continuous on K,

$$\lim_{i \to \infty} f(x_{n_i}) = f(x).$$

Writing $y = f(x)$, we have $y \in f(K)$ and

$$\lim_{i \to \infty} y_{n_i} = y.$$

Therefore every sequence (y_n) in $f(K)$ has a convergent subsequence whose limit belongs to $f(K)$, so $f(K)$ is compact.

As an alternative proof, we show that $f(K)$ has the Heine-Borel property. Suppose that $\{V_i : i \in I\}$ is an open cover of $f(K)$. Since f is continuous, Theorem 7.31 implies that $f^{-1}(V_i)$ is open in K, so $\{f^{-1}(V_i) : i \in I\}$ is an open cover of K. Since K is compact, there is a finite subcover

$$\left\{ f^{-1}(V_{i_1}), f^{-1}(V_{i_2}), \ldots, f^{-1}(V_{i_N}) \right\}$$

of K, and it follows that

$$\{V_{i_1}, V_{i_2}, \ldots, V_{i_N}\}$$

is a finite subcover of the original open cover of $f(K)$. This proves that $f(K)$ is compact. $\qquad\square$

Note that compactness is essential here; it is not true, in general, that a continuous function maps closed sets to closed sets.

Example 7.36. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \frac{1}{1 + x^2}.$$

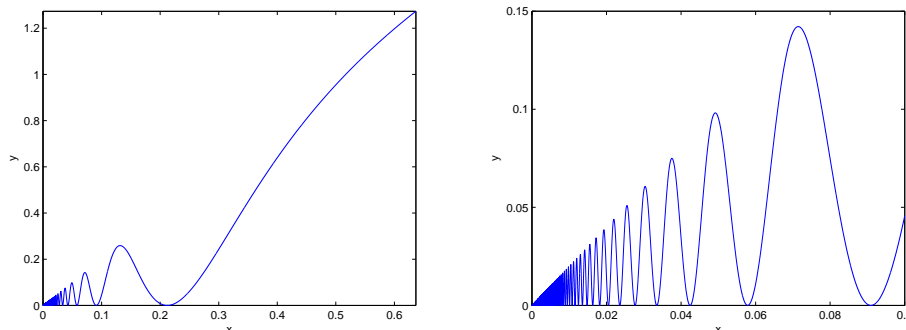Then $[0, \infty)$ is closed but $f([0, \infty)) = (0, 1]$ is not.

Figure 3. A plot of the function $y = x + x\sin(1/x)$ on $[0, 2/\pi]$ and a detail near the origin.

The following result is one of the most important property of continuous functions on compact sets.

Theorem 7.37 (Weierstrass extreme value). *If* $f : K \to \mathbb{R}$ *is continuous and* $K \subset \mathbb{R}$ *is compact, then* f *is bounded on* K *and* f *attains its maximum and minimum values on* K.

Proof. The image $f(K)$ is compact from Theorem 7.35. Proposition 5.40 implies that $f(K)$ is bounded and the maximum M and minimum m belong to $f(K)$. Therefore there are points $x, y \in K$ such that $f(x) = M$, $f(y) = m$, and f attains its maximum and minimum on K. □

Example 7.38. Define $f : [0, 1] \to \mathbb{R}$ by

$$f(x) = \begin{cases} 1/x & \text{if } 0 < x \le 1, \\ 0 & \text{if } x = 0. \end{cases}$$

Then f is unbounded on $[0, 1]$ and has no maximum value (f does, however, have a minimum value of 0 attained at $x = 0$). In this example, $[0, 1]$ is compact but f is discontinuous at 0, which shows that a discontinuous function on a compact set needn't be bounded.

Example 7.39. Define $f : (0, 1] \to \mathbb{R}$ by $f(x) = 1/x$. Then f is unbounded on $(0, 1]$ with no maximum value (f does, however, have a minimum value of 1 attained at $x = 1$). In this example, f is continuous but the half-open interval $(0, 1]$ isn't compact, which shows that a continuous function on a non-compact set needn't be bounded.

Example 7.40. Define $f : (0, 1) \to \mathbb{R}$ by $f(x) = x$. Then

$$\inf_{x \in (0,1)} f(x) = 0, \qquad \sup_{x \in (0,1)} f(x) = 1$$

but $f(x) \ne 0$, $f(x) \ne 1$ for any $0 < x < 1$. Thus, even if a continuous function on a non-compact set is bounded, it needn't attain its supremum or infimum.

Example 7.41. Define $f : [0, 2/\pi] \to \mathbb{R}$ by

$$f(x) = \begin{cases} x + x\sin(1/x) & \text{if } 0 < x \leq 2/\pi, \\ 0 & \text{if } x = 0. \end{cases}$$

(See Figure 3.) Then f is continuous on the compact interval $[0, 2/\pi]$, so by Theorem 7.37 it attains its maximum and minimum. For $0 \leq x \leq 2/\pi$, we have $0 \leq f(x) \leq 1/\pi$ since $|\sin 1/x| \leq 1$. Thus, the minimum value of f is 0, attained at $x = 0$. It is also attained at infinitely many other interior points in the interval,

$$x_n = \frac{1}{2n\pi + 3\pi/2}, \qquad n = 0, 1, 2, 3, \ldots,$$

where $\sin(1/x_n) = -1$. The maximum value of f is $1/\pi$, attained at $x = 2/\pi$.

Finally, we prove that continuous functions on compact sets are uniformly continuous

Theorem 7.42. If $f : K \to \mathbb{R}$ is continuous and $K \subset \mathbb{R}$ is compact, then f is uniformly continuous on K.

Proof. Suppose for contradiction that f is not uniformly continuous on K. Then from Proposition 7.24 there exists $\epsilon_0 > 0$ and sequences (x_n), (y_n) in K such that

$$\lim_{n \to \infty} |x_n - y_n| = 0 \text{ and } |f(x_n) - f(y_n)| \geq \epsilon_0 \text{ for every } n \in \mathbb{N}.$$

Since K is compact, there is a convergent subsequence (x_{n_i}) of (x_n) such that

$$\lim_{i \to \infty} x_{n_i} = x \in K.$$

Moreover, since $(x_n - y_n) \to 0$ as $n \to \infty$, it follows that

$$\lim_{i \to \infty} y_{n_i} = \lim_{i \to \infty} [x_{n_i} - (x_{n_i} - y_{n_i})] = \lim_{i \to \infty} x_{n_i} - \lim_{i \to \infty} (x_{n_i} - y_{n_i}) = x,$$

so (y_{n_i}) also converges to x. Then, since f is continuous on K,

$$\lim_{i \to \infty} |f(x_{n_i}) - f(y_{n_i})| = \left| \lim_{i \to \infty} f(x_{n_i}) - \lim_{i \to \infty} f(y_{n_i}) \right| = |f(x) - f(x)| = 0,$$

but this contradicts the non-uniform continuity condition

$$|f(x_{n_i}) - f(y_{n_i})| \geq \epsilon_0.$$

Therefore f is uniformly continuous. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Example 7.43. The function $f : [0, 2/\pi] \to \mathbb{R}$ defined in Example 7.41 is uniformly continuous on $[0, 2/\pi]$ since it is is continuous and $[0, 2/\pi]$ is compact.

7.6. The intermediate value theorem

The intermediate value theorem states that a continuous function on an interval takes on all values between any two of its values. We first prove a special case.

Theorem 7.44 (Intermediate value)**.** Suppose that $f : [a, b] \to \mathbb{R}$ is a continuous function on a closed, bounded interval. If $f(a) < 0$ and $f(b) > 0$, or $f(a) > 0$ and $f(b) < 0$, then there is a point $a < c < b$ such that $f(c) = 0$.

Proof. Assume for definiteness that $f(a) < 0$ and $f(b) > 0$. (If $f(a) > 0$ and $f(b) < 0$, consider $-f$ instead of f.) The set

$$E = \{x \in [a, b] : f(x) < 0\}$$

is nonempty, since $a \in E$, and E is bounded from above by b. Let

$$c = \sup E \in [a, b],$$

which exists by the completeness of $\mathbb{R}$. We claim that $f(c) = 0$.

Suppose for contradiction that $f(c) \neq 0$. Since f is continuous at c, there exists $\delta > 0$ such that

$$|x - c| < \delta \text{ and } x \in [a, b] \text{ implies that } |f(x) - f(c)| < \frac{1}{2}|f(c)|.$$

If $f(c) < 0$, then $c \neq b$ and

$$f(x) = f(c) + f(x) - f(c) < f(c) - \frac{1}{2}f(c)$$

for all $x \in [a, b]$ such that $|x - c| < \delta$, so $f(x) < \frac{1}{2}f(c) < 0$. It follows that there are points $x \in E$ with $x > c$, which contradicts the fact that c is an upper bound of E.

If $f(c) > 0$, then $c \neq a$ and

$$f(x) = f(c) + f(x) - f(c) > f(c) - \frac{1}{2}f(c)$$

for all $x \in [a, b]$ such that $|x - c| < \delta$, so $f(x) > \frac{1}{2}f(c) > 0$. It follows that there exists $\eta > 0$ such that $c - \eta \geq a$ and

$$f(x) > 0 \text{ for } c - \eta \leq x \leq c.$$

In that case, $c - \eta < c$ is an upper bound for E, since c is an upper bound and $f(x) > 0$ for $c - \eta \leq x \leq c$, which contradicts the fact that c is the least upper bound. This proves that $f(c) = 0$. Finally, $c \neq a, b$ since f is nonzero at the endpoints, so $a < c < b$. $\square$

We give some examples to show that all of the hypotheses in this theorem are necessary.

Example 7.45. Let $K = [-2, -1] \cup [1, 2]$ and define $f : K \to \mathbb{R}$ by

$$f(x) = \begin{cases} -1 & \text{if } -2 \leq x \leq -1 \\ 1 & \text{if } 1 \leq x \leq 2 \end{cases}$$

Then $f(-2) < 0$ and $f(2) > 0$, but f doesn't vanish at any point in its domain. Thus, in general, Theorem 7.44 fails if the domain of f is not a connected interval $[a, b]$.

Example 7.46. Define $f : [-1, 1] \to \mathbb{R}$ by

$$f(x) = \begin{cases} -1 & \text{if } -1 \leq x < 0 \\ 1 & \text{if } 0 \leq x \leq 1 \end{cases}$$

Then $f(-1) < 0$ and $f(1) > 0$, but f doesn't vanish at any point in its domain. Here, f is defined on an interval but it is discontinuous at 0. Thus, in general, Theorem 7.44 fails for discontinuous functions.

As one immediate consequence of the intermediate value theorem, we show that the real numbers contain the square root of 2.

Example 7.47. Define the continuous function $f : [1, 2] \to \mathbb{R}$ by

$$f(x) = x^2 - 2.$$

Then $f(1) < 0$ and $f(2) > 0$, so Theorem 7.44 implies that there exists $1 < c < 2$ such that $c^2 = 2$. Moreover, since $x^2 - 2$ is strictly increasing on $[0, \infty)$, there is a unique such positive number, and we have proved the existence of $\sqrt{2}$.

We can get more accurate approximations to $\sqrt{2}$ by repeatedly bisecting the interval $[1, 2]$. For example $f(3/2) = 1/4 > 0$ so $1 < \sqrt{2} < 3/2$, and $f(5/4) < 0$ so $5/4 < \sqrt{2} < 3/2$, and so on. This bisection method is a simple, but useful, algorithm for computing numerical approximations of solutions of $f(x) = 0$ where f is a continuous function.

Note that we used the existence of a supremum in the proof of Theorem 7.44. If we restrict $f(x) = x^2 - 2$ to rational numbers, $f : A \to \mathbb{Q}$ where $A = [1, 2] \cap \mathbb{Q}$, then f is continuous on A, $f(1) < 0$ and $f(2) > 0$, but $f(c) \neq 0$ for any $c \in A$ since $\sqrt{2}$ is irrational. This shows that the completeness of $\mathbb{R}$ is essential for Theorem 7.44 to hold. (Thus, in a sense, the theorem actually describes the completeness of the continuum $\mathbb{R}$ rather than the continuity of f!)

The general statement of the Intermediate Value Theorem follows immediately from this special case.

Theorem 7.48 (Intermediate value theorem). Suppose that $f : [a, b] \to \mathbb{R}$ is a continuous function on a closed, bounded interval. Then for every d strictly between $f(a)$ and $f(b)$ there is a point $a < c < b$ such that $f(c) = d$.

Proof. Suppose, for definiteness, that $f(a) < f(b)$ and $f(a) < d < f(b)$. (If $f(a) > f(b)$ and $f(b) < d < f(a)$, apply the same proof to $-f$, and if $f(a) = f(b)$ there is nothing to prove.) Let $g(x) = f(x) - d$. Then $g(a) < 0$ and $g(b) > 0$, so Theorem 7.44 implies that $g(c) = 0$ for some $a < c < b$, meaning that $f(c) = d$. $\square$

As one consequence of our previous results, we prove that a continuous function maps compact intervals to compact intervals.

Theorem 7.49. Suppose that $f : [a, b] \to \mathbb{R}$ is a continuous function on a closed, bounded interval. Then $f([a, b]) = [m, M]$ is a closed, bounded interval.

Proof. Theorem 7.37 implies that $m \leq f(x) \leq M$ for all $x \in [a, b]$, where m and M are the maximum and minimum values of f on $[a, b]$, so $f([a, b]) \subset [m, M]$. Moreover, there are points $c, d \in [a, b]$ such that $f(c) = m$, $f(d) = M$.

Let $J = [c, d]$ if $c \leq d$ or $J = [d, c]$ if $d < c$. Then $J \subset [a, b]$, and Theorem 7.48 implies that f takes on all values in $[m, M]$ on J. It follows that $f([a, b]) \supset [m, M]$, so $f([a, b]) = [m, M]$. $\square$

First we give an example to illustrate the theorem.

Example 7.50. Define $f : [-1, 1] \to \mathbb{R}$ by

$$f(x) = x - x^3.$$

Then, using calculus to compute the maximum and minimum of f, we find that

$$f\left([-1,1]\right) = [-M, M], \qquad M = \frac{2}{3\sqrt{3}}.$$

This example illustrates that $f\left([a,b]\right) \neq [f(a), f(b)]$ unless f is increasing.

Next we give some examples to show that the continuity of f and the connectedness and compactness of the interval $[a,b]$ are essential for Theorem 7.49 to hold.

Example 7.51. Let $\mathrm{sgn} : [-1,1] \to \mathbb{R}$ be the sign function defined in Example 6.8. Then f is a discontinuous function on a compact interval $[-1,1]$, but the range $f([-1,1]) = \{-1, 0, 1\}$ consists of three isolated points and is not an interval.

Example 7.52. In Example 7.45, the function $f : K \to \mathbb{R}$ is continuous on a compact set K but $f(K) = \{-1, 1\}$ consists of two isolated points and is not an interval.

Example 7.53. The continuous function $f : \mathbb{R} \to \mathbb{R}$ in Example 7.36 maps the unbounded, closed interval $[0, \infty)$ to the half-open interval $(0, 1]$.

The last example shows that a continuous function may map a closed but unbounded interval to an interval which isn't closed (or open). Nevertheless, as shown in Theorem 7.32, a continuous function always maps intervals to intervals, although the intervals may be open, closed, half-open, bounded, or unbounded.

7.7. Monotonic functions

Monotonic functions have continuity properties that are not shared by general functions.

Definition 7.54. Let $I \subset \mathbb{R}$ be an interval. A function $f : I \to \mathbb{R}$ is increasing if

$$f(x_1) \leq f(x_2) \qquad \text{if } x_1, x_2 \in I \text{ and } x_1 < x_2,$$

strictly increasing if

$$f(x_1) < f(x_2) \qquad \text{if } x_1, x_2 \in I \text{ and } x_1 < x_2,$$

decreasing if

$$f(x_1) \geq f(x_2) \qquad \text{if } x_1, x_2 \in I \text{ and } x_1 < x_2,$$

and strictly decreasing if

$$f(x_1) > f(x_2) \qquad \text{if } x_1, x_2 \in I \text{ and } x_1 < x_2.$$

An increasing or decreasing function is called a monotonic function, and a strictly increasing or strictly decreasing function is called a strictly monotonic function.

A commonly used alternative (and, unfortunately, incompatible) terminology is "nondecreasing" for "increasing," "increasing" for "strictly increasing," "nonincreasing" for "decreasing," and "decreasing" for "strictly decreasing." According to our terminology, a constant function is both increasing and decreasing. Monotonic functions are also referred to as monotone functions.

Theorem 7.55. If $f : I \to \mathbb{R}$ is monotonic on an interval I, then the left and right limits of f,

$$\lim_{x \to c^-} f(x), \qquad \lim_{x \to c^+} f(x),$$

exist at every interior point c of I.

Proof. Assume for definiteness that f is increasing. (If f is decreasing, we can apply the same argument to $-f$ which is increasing). We will prove that

$$\lim_{x \to c^-} f(x) = \sup E, \qquad E = \{f(x) \in \mathbb{R} : x \in I \text{ and } x < c\}.$$

The set E is nonempty since c in an interior point of I, so there exists $x \in I$ with $x < c$, and E bounded from above by $f(c)$ since f is increasing. It follows that $L = \sup E \in \mathbb{R}$ exists. (Note that L may be strictly less than $f(c)$!)

Suppose that $\epsilon > 0$ is given. Since L is a least upper bound of E, there exists $y_0 \in E$ such that $L - \epsilon < y_0 \leq L$, and therefore $x_0 \in I$ with $x_0 < c$ such that $f(x_0) = y_0$. Let $\delta = c - x_0 > 0$. If $c - \delta < x < c$, then $x_0 < x < c$ and therefore $f(x_0) \leq f(x) \leq L$ since f is increasing and L is an upper bound of E. It follows that

$$L - \epsilon < f(x) \leq L \qquad \text{if } c - \delta < x < c,$$

which proves that $\lim_{x \to c^-} f(x) = L$.

A similar argument, or the same argument applied to $g(x) = -f(-x)$, shows that

$$\lim_{x \to c^+} f(x) = \inf \{f(x) \in \mathbb{R} : x \in I \text{ and } x > c\}.$$

We leave the details as an exercise. $\qquad \square$

Similarly, if $I = (a, b]$ has right-endpoint $b \in I$ and f is monotonic on I, then the left limit $\lim_{x \to b^-} f(x)$ exists, although it may not equal $f(b)$, and if $a \in I$ is a left-endpoint, then the right limit $\lim_{x \to a^+} f(x)$ exists, although it may not equal $f(a)$.

Corollary 7.56. Every discontinuity of a monotonic function $f : I \to \mathbb{R}$ at an interior point of the interval I is a jump discontinuity.

Proof. If c is an interior point of I, then the left and right limits of f at c exist by the previous theorem. Moreover, assuming for definiteness that f is increasing, we have

$$f(x) \leq f(c) \leq f(y) \qquad \text{for all } x, y \in I \text{ with } x < c < y,$$

and since limits preserve inequalities

$$\lim_{x \to c^-} f(x) \leq f(c) \leq \lim_{x \to c^+} f(x).$$

If the left and right limits are equal, then the limit exists and is equal to the left and right limits, so

$$\lim_{x \to c} f(x) = f(c),$$

meaning that f is continuous at c. In particular, a monotonic function cannot have a removable discontinuity at an interior point of its domain (although it can have one at an endpoint of a closed interval). If the left and right limits are not equal,

then f has a jump discontinuity at c, so f cannot have an essential discontinuity either. $\square$

One can show that a monotonic function has, at most, a countable number of discontinuities, and it may have a countably infinite number, but we omit the proof. By contrast, the non-monotonic Dirichlet function has uncountably many discontinuities at every point of $\mathbb{R}$.

Differentiable Functions

A differentiable function is a function that can be approximated locally by a linear function.

8.1. The derivative

Definition 8.1. Suppose that $f : (a, b) \to \mathbb{R}$ and $a < c < b$. Then f is differentiable at c with derivative $f'(c)$ if

$$\lim_{h \to 0} \left[\frac{f(c + h) - f(c)}{h} \right] = f'(c).$$

The domain of f' is the set of points $c \in (a, b)$ for which this limit exists. If the limit exists for every $c \in (a, b)$ then we say that f is differentiable on (a, b).

Graphically, this definition says that the derivative of f at c is the slope of the tangent line to $y = f(x)$ at c, which is the limit as $h \to 0$ of the slopes of the lines through $(c, f(c))$ and $(c + h, f(c + h))$.

We can also write

$$f'(c) = \lim_{x \to c} \left[\frac{f(x) - f(c)}{x - c} \right],$$

since if $x = c + h$, the conditions $0 < |x - c| < \delta$ and $0 < |h| < \delta$ in the definitions of the limits are equivalent. The ratio

$$\frac{f(x) - f(c)}{x - c}$$

is undefined $(0/0)$ at $x = c$, but it doesn't have to be defined in order for the limit as $x \to c$ to exist.

Like continuity, differentiability is a local property. That is, the differentiability of a function f at c and the value of the derivative, if it exists, depend only the values of f in a arbitrarily small neighborhood of c. In particular if $f : A \to \mathbb{R}$

where $A \subset \mathbb{R}$, then we can define the differentiability of f at any interior point $c \in A$ since there is an open interval $(a, b) \subset A$ with $c \in (a, b)$.

8.1.1. Examples of derivatives. Let us give a number of examples that illustrate differentiable and non-differentiable functions.

Example 8.2. The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is differentiable on $\mathbb{R}$ with derivative $f'(x) = 2x$ since

$$\lim_{h \to 0} \left[\frac{(c + h)^2 - c^2}{h} \right] = \lim_{h \to 0} \frac{h(2c + h)}{h} = \lim_{h \to 0} (2c + h) = 2c.$$

Note that in computing the derivative, we first cancel by h, which is valid since $h \neq 0$ in the definition of the limit, and then set $h = 0$ to evaluate the limit. This procedure would be inconsistent if we didn't use limits.

Example 8.3. The function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x^2 & \text{if } x > 0, \\ 0 & \text{if } x \leq 0. \end{cases}$$

is differentiable on $\mathbb{R}$ with derivative

$$f'(x) = \begin{cases} 2x & \text{if } x > 0, \\ 0 & \text{if } x \leq 0. \end{cases}$$

For $x > 0$, the derivative is $f'(x) = 2x$ as above, and for $x < 0$, we have $f'(x) = 0$. For 0, we consider the limit

$$\lim_{h \to 0} \left[\frac{f(h) - f(0)}{h} \right] = \lim_{h \to 0} \frac{f(h)}{h}.$$

The right limit is

$$\lim_{h \to 0^+} \frac{f(h)}{h} = \lim_{h \to 0} h = 0,$$

and the left limit is

$$\lim_{h \to 0^-} \frac{f(h)}{h} = 0.$$

Since the left and right limits exist and are equal, the limit also exists, and f is differentiable at 0 with $f'(0) = 0$.

Next, we consider some examples of non-differentiability at discontinuities, corners, and cusps.

Example 8.4. The function $f : \mathbb{R} \to \mathbb{R}$ defined by

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

is differentiable at $x \neq 0$ with derivative $f'(x) = -1/x^2$ since

$$\lim_{h \to 0} \left[\frac{f(c+h) - f(c)}{h} \right] = \lim_{h \to 0} \left[\frac{1/(c+h) - 1/c}{h} \right]$$
$$= \lim_{h \to 0} \left[\frac{c - (c+h)}{hc(c+h)} \right]$$
$$= -\lim_{h \to 0} \frac{1}{c(c+h)}$$
$$= -\frac{1}{c^2}.$$

However, f is not differentiable at 0 since the limit

$$\lim_{h \to 0} \left[\frac{f(h) - f(0)}{h} \right] = \lim_{h \to 0} \left[\frac{1/h - 0}{h} \right] = \lim_{h \to 0} \frac{1}{h^2}$$

does not exist.

Example 8.5. The sign function $f(x) = \operatorname{sgn} x$, defined in Example 6.8, is differentiable at $x \neq 0$ with $f'(x) = 0$, since in that case $f(x+h) - f(x) = 0$ for all sufficiently small h. The sign function is not differentiable at 0 since

$$\lim_{h \to 0} \left[\frac{\operatorname{sgn} h - \operatorname{sgn} 0}{h} \right] = \lim_{h \to 0} \frac{\operatorname{sgn} h}{h}$$

and

$$\frac{\operatorname{sgn} h}{h} = \begin{cases} 1/h & \text{if } h > 0 \\ -1/h & \text{if } h < 0 \end{cases}$$

is unbounded in every neighborhood of 0, so its limit does not exist.

Example 8.6. The absolute value function $f(x) = |x|$ is differentiable at $x \neq 0$ with derivative $f'(x) = \operatorname{sgn} x$. It is not differentiable at 0, however, since

$$\lim_{h \to 0} \frac{f(h) - f(0)}{h} = \lim_{h \to 0} \frac{|h|}{h} = \lim_{h \to 0} \operatorname{sgn} h$$

does not exist. (The right limit is 1 and the left limit is -1.)

Example 8.7. The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = |x|^{1/2}$ is differentiable at $x \neq 0$ with

$$f'(x) = \frac{\operatorname{sgn} x}{2|x|^{1/2}}.$$

If $c > 0$, then using the difference of two square to rationalize the numerator, we get

$$\lim_{h \to 0} \left[\frac{f(c+h) - f(c)}{h} \right] = \lim_{h \to 0} \frac{(c+h)^{1/2} - c^{1/2}}{h}$$
$$= \lim_{h \to 0} \frac{(c+h) - c}{h \left[(c+h)^{1/2} + c^{1/2} \right]}$$
$$= \lim_{h \to 0} \frac{1}{(c+h)^{1/2} + c^{1/2}}$$
$$= \frac{1}{2c^{1/2}}.$$

If $c < 0$, we get the analogous result with a negative sign. However, f is not differentiable at 0, since

$$\lim_{h \to 0^+} \frac{f(h) - f(0)}{h} = \lim_{h \to 0^+} \frac{1}{h^{1/2}}$$

does not exist.

Example 8.8. The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^{1/3}$ is differentiable at $x \neq 0$ with

$$f'(x) = \frac{1}{3x^{2/3}}.$$

To prove this result, we use the identity for the difference of cubes,

$$a^3 - b^3 = (a - b)(a^2 + ab + b^2),$$

and get for $c \neq 0$ that

$$\lim_{h \to 0} \left[\frac{f(c + h) - f(c)}{h} \right] = \lim_{h \to 0} \frac{(c + h)^{1/3} - c^{1/3}}{h}$$

$$= \lim_{h \to 0} \frac{(c + h) - c}{h \left[(c + h)^{2/3} + (c + h)^{1/3} c^{1/3} + c^{2/3} \right]}$$

$$= \lim_{h \to 0} \frac{1}{(c + h)^{2/3} + (c + h)^{1/3} c^{1/3} + c^{2/3}}$$

$$= \frac{1}{3c^{2/3}}.$$

However, f is not differentiable at 0, since

$$\lim_{h \to 0} \frac{f(h) - f(0)}{h} = \lim_{h \to 0} \frac{1}{h^{2/3}}$$

does not exist.

Finally, we consider some examples of highly oscillatory functions.

Example 8.9. Define $f : \mathbb{R} \to \mathbb{R}$ by

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

It follows from the product and chain rules proved below that f is differentiable at $x \neq 0$ with derivative

$$f'(x) = \sin \frac{1}{x} - \frac{1}{x} \cos \frac{1}{x}.$$

However, f is not differentiable at 0, since

$$\lim_{h \to 0} \frac{f(h) - f(0)}{h} = \lim_{h \to 0} \sin \frac{1}{h},$$

which does not exist.

Example 8.10. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} x^2 \sin(1/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$
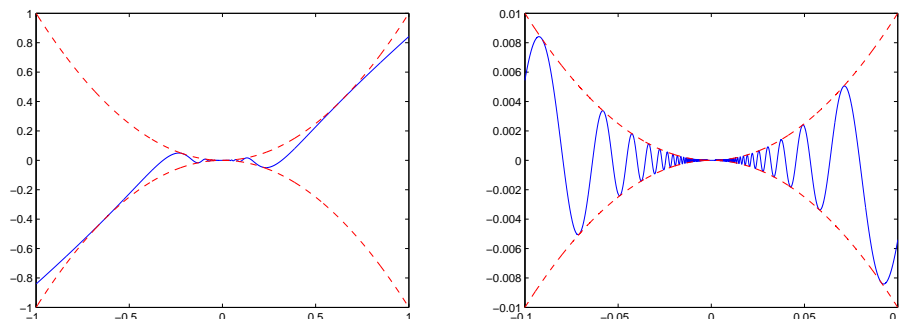
Figure 1. A plot of the function $y = x^2 \sin(1/x)$ and a detail near the origin with the parabolas $y = \pm x^2$ shown in red.

Then f is differentiable on $\mathbb{R}$. (See Figure 1.) It follows from the product and chain rules proved below that f is differentiable at $x \neq 0$ with derivative

$$f'(x) = 2x \sin \frac{1}{x} - \cos \frac{1}{x}.$$

Moreover, f is differentiable at 0 with $f'(0) = 0$, since

$$\lim_{h \to 0} \frac{f(h) - f(0)}{h} = \lim_{h \to 0} h \sin \frac{1}{h} = 0.$$

In this example, $\lim_{x \to 0} f'(x)$ does not exist, so although f is differentiable on $\mathbb{R}$, its derivative f' is not continuous at 0.

8.1.2. Derivatives as linear approximations. Another way to view Definition 8.1 is to write

$$f(c + h) = f(c) + f'(c)h + r(h)$$

as the sum of a linear (or, strictly speaking, affine) approximation $f(c) + f'(c)h$ of $f(c + h)$ and a remainder $r(h)$. In general, the remainder also depends on c, but we don't show this explicitly since we're regarding c as fixed.

As we prove in the following proposition, the differentiability of f at c is equivalent to the condition

$$\lim_{h \to 0} \frac{r(h)}{h} = 0.$$

That is, the remainder $r(h)$ approaches 0 faster than h, so the linear terms in h provide a leading order approximation to $f(c + h)$ when h is small. We also write this condition on the remainder as

$$r(h) = o(h) \qquad \text{as } h \to 0,$$

pronounced "r is little-oh of h as $h \to 0$."

Graphically, this condition means that the graph of f near c is close the line through the point $(c, f(c))$ with slope $f'(c)$. Analytically, it means that the function

$$h \mapsto f(c + h) - f(c)$$

is approximated near c by the linear function

$$h \mapsto f'(c)h.$$

Thus, $f'(c)$ may be interpreted as a scaling factor by which a differentiable function f shrinks or stretches lengths near c.

If $|f'(c)| < 1$, then f shrinks the length of a small interval about c by (approximately) this factor; if $|f'(c)| > 1$, then f stretches the length of an interval by (approximately) this factor; if $f'(c) > 0$, then f preserves the orientation of the interval, meaning that it maps the left endpoint to the left endpoint of the image and the right endpoint to the right endpoints; if $f'(c) < 0$, then f reverses the orientation of the interval, meaning that it maps the left endpoint to the right endpoint of the image and visa-versa.

We can use this description as a definition of the derivative.

Proposition 8.11. Suppose that $f : (a, b) \to \mathbb{R}$. Then f is differentiable at $c \in (a, b)$ if and only if there exists a constant $A \in \mathbb{R}$ and a function $r : (a - c, b - c) \to \mathbb{R}$ such that

$$f(c + h) = f(c) + Ah + r(h), \qquad \lim_{h \to 0} \frac{r(h)}{h} = 0.$$

In that case, $A = f'(c)$.

Proof. First suppose that f is differentiable at c according to Definition 8.1, and define

$$r(h) = f(c + h) - f(c) - f'(c)h.$$

Then

$$\lim_{h \to 0} \frac{r(h)}{h} = \lim_{h \to 0} \left[\frac{f(c + h) - f(c)}{h} - f'(c) \right] = 0,$$

so the condition in the proposition holds with $A = f'(c)$.

Conversely, suppose that $f(c + h) = f(c) + Ah + r(h)$ where $r(h)/h \to 0$ as $h \to 0$. Then

$$\lim_{h \to 0} \left[\frac{f(c + h) - f(c)}{h} \right] = \lim_{h \to 0} \left[A + \frac{r(h)}{h} \right] = A,$$

so f is differentiable at c with $f'(c) = A$. $\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Example 8.12. For Example 8.2 with $f(x) = x^2$, we get

$$(c + h)^2 = c^2 + 2ch + h^2,$$

and $r(h) = h^2$, which goes to zero at a quadratic rate as $h \to 0$.

Example 8.13. For Example 8.4 with $f(x) = 1/x$, we get

$$\frac{1}{c + h} = \frac{1}{c} - \frac{1}{c^2}h + r(h),$$

for $c \neq 0$, where the quadratically small remainder is

$$r(h) = \frac{h^2}{c^2(c + h)}.$$

8.1.3. Left and right derivatives. For the most part, we will use derivatives that are defined only at the interior points of the domain of a function. Sometimes, however, it is convenient to use one-sided left or right derivatives that are defined at the endpoint of an interval.

Definition 8.14. Suppose that $f : [a, b] \to \mathbb{R}$. Then f is right-differentiable at $a \leq c < b$ with right derivative $f'(c^+)$ if

$$\lim_{h \to 0^+} \left[\frac{f(c + h) - f(c)}{h} \right] = f'(c^+)$$

exists, and f is left-differentiable at $a < c \leq b$ with left derivative $f'(c^-)$ if

$$\lim_{h \to 0^-} \left[\frac{f(c + h) - f(c)}{h} \right] = \lim_{h \to 0^+} \left[\frac{f(c) - f(c - h)}{h} \right] = f'(c^-).$$

A function is differentiable at $a < c < b$ if and only if the left and right derivatives at c both exist and are equal.

Example 8.15. If $f : [0, 1] \to \mathbb{R}$ is defined by $f(x) = x^2$, then

$$f'(0^+) = 0, \qquad f'(1^-) = 2.$$

These left and right derivatives remain the same if f is extended to a function defined on a larger domain, say

$$f(x) = \begin{cases} x^2 & \text{if } 0 \leq x \leq 1, \\ 1 & \text{if } x > 1, \\ 1/x & \text{if } x < 0. \end{cases}$$

For this extended function we have $f'(1^+) = 0$, which is not equal to $f'(1^-)$, and $f'(0^-)$ does not exist, so the extended function is not differentiable at either 0 or 1.

Example 8.16. The absolute value function $f(x) = |x|$ in Example 8.6 is left and right differentiable at 0 with left and right derivatives

$$f'(0^+) = 1, \qquad f'(0^-) = -1.$$

These are not equal, and f is not differentiable at 0.

8.2. Properties of the derivative

In this section, we prove some basic properties of differentiable functions.

8.2.1. Differentiability and continuity. First we discuss the relation between differentiability and continuity.

Theorem 8.17. If $f : (a, b) \to \mathbb{R}$ is differentiable at at $c \in (a, b)$, then f is continuous at c.

Proof. If f is differentiable at c, then

$$\lim_{h \to 0} f(c + h) - f(c) = \lim_{h \to 0} \left[\frac{f(c + h) - f(c)}{h} \cdot h \right]$$
$$= \lim_{h \to 0} \left[\frac{f(c + h) - f(c)}{h} \right] \cdot \lim_{h \to 0} h$$
$$= f'(c) \cdot 0$$
$$= 0,$$

which implies that f is continuous at c. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

For example, the sign function in Example 8.5 has a jump discontinuity at 0 so it cannot be differentiable at 0. The converse does not hold, and a continuous function needn't be differentiable. The functions in Examples 8.6, 8.8, 8.9 are continuous but not differentiable at 0. Example 9.24 describes a function that is continuous on $\mathbb{R}$ but not differentiable anywhere.

In Example 8.10, the function is differentiable on $\mathbb{R}$, but the derivative f' is not continuous at 0. Thus, while a function f has to be continuous to be differentiable, if f is differentiable its derivative f' need not be continuous. This leads to the following definition.

Definition 8.18. A function $f : (a, b) \to \mathbb{R}$ is continuously differentiable on (a, b), written $f \in C^1(a, b)$, if it is differentiable on (a, b) and $f' : (a, b) \to \mathbb{R}$ is continuous.

For example, the function $f(x) = x^2$ with derivative $f'(x) = 2x$ is continuously differentiable on $\mathbb{R}$, whereas the function in Example 8.10 is not continuously differentiable at 0. As this example illustrates, functions that are differentiable but not continuously differentiable may behave in rather pathological ways. On the other hand, the behavior of continuously differentiable functions, whose graphs have continuously varying tangent lines, is more-or-less consistent with what one expects.

8.2.2. Algebraic properties of the derivative. A fundamental property of the derivative is that it is a linear operation. In addition, we have the following product and quotient rules.

Theorem 8.19. If $f, g : (a, b) \to \mathbb{R}$ are differentiable at $c \in (a, b)$ and $k \in \mathbb{R}$, then kf, $f + g$, and fg are differentiable at c with

$$(kf)'(c) = kf'(c), \quad (f + g)'(c) = f'(c) + g'(c), \quad (fg)'(c) = f'(c)g(c) + f(c)g'(c).$$

Furthermore, if $g(c) \neq 0$, then f/g is differentiable at c with

$$\left(\frac{f}{g} \right)'(c) = \frac{f'(c)g(c) - f(c)g'(c)}{g^2(c)}.$$

Proof. The first two properties follow immediately from the linearity of limits stated in Theorem 6.34. For the product rule, we write

$$(fg)'(c) = \lim_{h \to 0} \left[\frac{f(c+h)g(c+h) - f(c)g(c)}{h} \right]$$

$$= \lim_{h \to 0} \left[\frac{(f(c+h) - f(c))\, g(c+h) + f(c)\, (g(c+h) - g(c))}{h} \right]$$

$$= \lim_{h \to 0} \left[\frac{f(c+h) - f(c)}{h} \right] \lim_{h \to 0} g(c+h) + f(c) \lim_{h \to 0} \left[\frac{g(c+h) - g(c)}{h} \right]$$

$$= f'(c)g(c) + f(c)g'(c),$$

where we have used the properties of limits in Theorem 6.34 and Theorem 8.19, which implies that g is continuous at c. The quotient rule follows by a similar argument, or by combining the product rule with the chain rule, which implies that $(1/g)' = -g'/g^2$. (See Example 8.22 below.) □

Example 8.20. We have $1' = 0$ and $x' = 1$. Repeated application of the product rule implies that x^n is differentiable on $\mathbb{R}$ for every $n \in \mathbb{N}$ with

$$(x^n)' = nx^{n-1}.$$

Alternatively, we can prove this result by induction: The formula holds for $n = 1$. Assuming that it holds for some $n \in \mathbb{N}$, we get from the product rule that

$$(x^{n+1})' = (x \cdot x^n)' = 1 \cdot x^n + x \cdot nx^{n-1} = (n+1)x^n,$$

and the result follows. It also follows by linearity that every polynomial function is differentiable on $\mathbb{R}$, and from the quotient rule that every rational function is differentiable at every point where its denominator is nonzero. The derivatives are given by their usual formulae.

8.3. The chain rule

The chain rule states that the composition of differentiable functions is differentiable. The result is quite natural if one thinks in terms of derivatives as linear maps. If f is differentiable at c, it scales lengths by a factor $f'(c)$, and if g is differentiable at $f(c)$, it scales lengths by a factor $g'(f(c))$. Thus, the composition $g \circ f$ scales lengths at c by a factor $g'(f(c)) \cdot f'(c)$. Equivalently, the derivative of a composition is the composition of the derivatives (regarded as linear maps).

We will prove the chain rule by showing that the composition of remainder terms in the linear approximations of f and g leads to a similar remainder term in the linear approximation of $g \circ f$. The argument is complicated by the fact that we have to evaluate the remainder of g at a point that depends on the remainder of f, but this complication should not obscure the simplicity of the final result.

Theorem 8.21 (Chain rule). Let $f : A \to \mathbb{R}$ and $g : B \to \mathbb{R}$ where $A \subset \mathbb{R}$ and $f(A) \subset B$, and suppose that c is an interior point of A and $f(c)$ is an interior point of B. If f is differentiable at c and g is differentiable at $f(c)$, then $g \circ f : A \to \mathbb{R}$ is differentiable at c and

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

Proof. Since f is differentiable at c, there is a function $r(h)$ such that

$$f(c+h) = f(c) + f'(c)h + r(h), \qquad \lim_{h \to 0} \frac{r(h)}{h} = 0,$$

and since g is differentiable at $f(c)$, there is a function $s(k)$ such that

$$g\left(f(c) + k\right) = g\left(f(c)\right) + g'\left(f(c)\right)k + s(k), \qquad \lim_{k \to 0} \frac{s(k)}{k} = 0.$$

It follows that

$$\begin{aligned}
(g \circ f)(c+h) &= g\left(f(c) + f'(c)h + r(h)\right) \\
&= g\left(f(c)\right) + g'\left(f(c)\right) \cdot \left(f'(c)h + r(h)\right) + s\left(f'(c)h + r(h)\right) \\
&= g\left(f(c)\right) + g'\left(f(c)\right)f'(c) \cdot h + t(h)
\end{aligned}$$

where

$$t(h) = g'\left(f(c)\right) \cdot r(h) + s\left(\phi(h)\right), \qquad \phi(h) = f'(c)h + r(h).$$

Since $r(h)/h \to 0$ as $h \to 0$, we have

$$\lim_{h \to 0} \frac{t(h)}{h} = \lim_{h \to 0} \frac{s\left(\phi(h)\right)}{h}.$$

We claim that this limit exists and is zero, and then it follows from Proposition 8.11 that $g \circ f$ is differentiable at c with

$$(g \circ f)'(c) = g'\left(f(c)\right)f'(c).$$

To prove the claim, we use the facts that

$$\frac{\phi(h)}{h} \to f'(c) \quad \text{as } h \to 0, \qquad \frac{s(k)}{k} \to 0 \quad \text{as } k \to 0.$$

Roughly speaking, we have $\phi(h) \sim f'(c)h$ when h is small and therefore

$$\frac{s\left(\phi(h)\right)}{h} \sim \frac{s\left(f'(c)h\right)}{h} \to 0 \qquad \text{as } h \to 0.$$

In detail, let $\epsilon > 0$ be given. We want to show that there exists $\delta > 0$ such that

$$\left| \frac{s\left(\phi(h)\right)}{h} \right| < \epsilon \qquad \text{if } 0 < |h| < \delta.$$

First, choose $\delta_1 > 0$ such that

$$\left| \frac{r(h)}{h} \right| < |f'(c)| + 1 \qquad \text{if } 0 < |h| < \delta_1.$$

If $0 < |h| < \delta_1$, then

$$\begin{aligned}
|\phi(h)| &\leq |f'(c)| \, |h| + |r(h)| \\
&< |f'(c)| \, |h| + (|f'(c)| + 1)|h| \\
&< (2|f'(c)| + 1) \, |h|.
\end{aligned}$$

Next, choose $\eta > 0$ so that

$$\left| \frac{s(k)}{k} \right| < \frac{\epsilon}{2|f'(c)| + 1} \qquad \text{if } 0 < |k| < \eta.$$

(We include a "1" in the denominator on the right-hand side to avoid a division by zero if $f'(c) = 0$.) Finally, define $\delta_2 > 0$ by

$$\delta_2 = \frac{\eta}{2|f'(c)| + 1},$$

and let $\delta = \min(\delta_1, \delta_2) > 0$.

If $0 < |h| < \delta$ and $\phi(h) \neq 0$, then $0 < |\phi(h)| < \eta$, so

$$|s\,(\phi(h))| \leq \frac{\epsilon|\phi(h)|}{2|f'(c)| + 1} < \epsilon|h|.$$

If $\phi(h) = 0$, then $s(\phi(h)) = 0$, so the inequality holds in that case also. This proves that

$$\lim_{h \to 0} \frac{s\,(\phi(h))}{h} = 0.$$

$\square$

Example 8.22. Suppose that f is differentiable at c and $f(c) \neq 0$. Then $g(y) = 1/y$ is differentiable at $f(c)$, with $g'(y) = -1/y^2$ (see Example 8.4). It follows that the reciprocal function $1/f = g \circ f$ is differentiable at c with

$$\left(\frac{1}{f}\right)'(c) = g'(f(c))f'(c) = -\frac{f'(c)}{f(c)^2}.$$

The chain rule gives an expression for the derivative of an inverse function. In terms of linear approximations, it states that if f scales lengths at c by a nonzero factor $f'(c)$, then f^{-1} scales lengths at $f(c)$ by the factor $1/f'(c)$.

Proposition 8.23. Suppose that $f : A \to \mathbb{R}$ is a one-to-one function on $A \subset \mathbb{R}$ with inverse $f^{-1} : B \to \mathbb{R}$ where $B = f(A)$. Assume that f is differentiable at an interior point $c \in A$ and f^{-1} is differentiable at $f(c)$, where $f(c)$ is an interior point of B. Then $f'(c) \neq 0$ and

$$(f^{-1})'\,(f(c)) = \frac{1}{f'(c)}.$$

Proof. The definition of the inverse implies that

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

Since f is differentiable at c and f^{-1} is differentiable at $f(c)$, the chain rule implies that

$$\left(f^{-1}\right)'\,(f(c))\,f'(c) = 1.$$

Dividing this equation by $f'(c) \neq 0$, we get the result. Moreover, it follows that f^{-1} cannot be differentiable at $f(c)$ if $f'(c) = 0$. $\square$

Alternatively, setting $d = f(c)$, we can write the result as

$$(f^{-1})'(d) = \frac{1}{f'\,(f^{-1}(d))}.$$

Proposition 8.23 is not entirely satisfactory because it assumes the existence and differentiability of an inverse function. We will return to this question in Section 8.7 below, but we end this section with some examples that illustrate the

necessity of the condition $f'(c) \neq 0$ for the existence and differentiability of the inverse.

Example 8.24. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$. Then $f'(0) = 0$ and f is not invertible on any neighborhood of the origin, since it is non-monotone and not one-to-one. On the other hand, if $f : (0, \infty) \to (0, \infty)$ is defined by $f(x) = x^2$, then $f'(x) = 2x \neq 0$ and the inverse function $f^{-1} : (0, \infty) \to (0, \infty)$ is given by

$$f^{-1}(y) = \sqrt{y}.$$

The formula for the inverse of the derivative gives

$$(f^{-1})'(x^2) = \frac{1}{f'(x)} = \frac{1}{2x},$$

or, writing $x = f^{-1}(y)$,

$$(f^{-1})'(y) = \frac{1}{2\sqrt{y}},$$

in agreement with Example 8.7.

Example 8.25. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^3$. Then f is strictly increasing, one-to-one, and onto. The inverse function $f^{-1} : \mathbb{R} \to \mathbb{R}$ is given by

$$f^{-1}(y) = y^{1/3}.$$

Then $f'(0) = 0$ and f^{-1} is not differentiable at $f(0) = 0$. On the other hand, f^{-1} is differentiable at non-zero points of $\mathbb{R}$, with

$$(f^{-1})'(x^3) = \frac{1}{f'(x)} = \frac{1}{3x^2},$$

or, writing $x = y^{1/3}$,

$$(f^{-1})'(y) = \frac{1}{3y^{2/3}},$$

in agreement with Example 8.8.

8.4. Extreme values

One of the most useful applications of the derivative is in locating the maxima and minima of functions.

Definition 8.26. Suppose that $f : A \to \mathbb{R}$. Then f has a global (or absolute) maximum at $c \in A$ if

$$f(x) \leq f(c) \qquad \text{for all } x \in A,$$

and f has a local (or relative) maximum at $c \in A$ if there is a neighborhood U of c such that

$$f(x) \leq f(c) \qquad \text{for all } x \in A \cap U.$$

Similarly, f has a global (or absolute) minimum at $c \in A$ if

$$f(x) \geq f(c) \qquad \text{for all } x \in A,$$

and f has a local (or relative) minimum at $c \in A$ if there is a neighborhood U of c such that

$$f(x) \geq f(c) \qquad \text{for all } x \in A \cap U.$$

If f has a (local or global) maximum or minimum at $c \in A$, then f is said to have a (local or global) extreme value at c.

Theorem 7.37 states that a continuous function on a compact set has a global maximum and minimum but does not say how to find them. The following fundamental result goes back to Fermat.

Theorem 8.27. If $f : A \subset \mathbb{R} \to \mathbb{R}$ has a local extreme value at an interior point $c \in A$ and f is differentiable at c, then $f'(c) = 0$.

Proof. If f has a local maximum at c, then $f(x) \leq f(c)$ for all x in a δ-neighborhood $(c - \delta, c + \delta)$ of c, so

$$\frac{f(c + h) - f(c)}{h} \leq 0 \qquad \text{for all } 0 < h < \delta,$$

which implies that

$$f'(c) = \lim_{h \to 0^+} \left[\frac{f(c + h) - f(c)}{h} \right] \leq 0.$$

Moreover,

$$\frac{f(c + h) - f(c)}{h} \geq 0 \qquad \text{for all } -\delta < h < 0,$$

which implies that

$$f'(c) = \lim_{h \to 0^-} \left[\frac{f(c + h) - f(c)}{h} \right] \geq 0.$$

It follows that $f'(c) = 0$. If f has a local minimum at c, then the signs in these inequalities are reversed, and we also conclude that $f'(c) = 0$. $\qquad \square$

For this result to hold, it is crucial that c is an interior point, since we look at the sign of the difference quotient of f on both sides of c. At an endpoint, we get the following inequality condition on the derivative. (Draw a graph!)

Proposition 8.28. Let $f : [a, b] \to \mathbb{R}$. If the right derivative of f exists at a, then: $f'(a^+) \leq 0$ if f has a local maximum at a; and $f'(a^+) \geq 0$ if f has a local minimum at a. Similarly, if the left derivative of f exists at b, then: $f'(b^-) \geq 0$ if f has a local maximum at b; and $f'(b^-) \leq 0$ if f has a local minimum at b.

Proof. If the right derivative of f exists at a, and f has a local maximum at a, then there exists $\delta > 0$ such that $f(x) \leq f(a)$ for $a \leq x < a + \delta$, so

$$f'(a^+) = \lim_{h \to 0^+} \left[\frac{f(a + h) - f(a)}{h} \right] \leq 0.$$

Similarly, if the left derivative of f exists at b, and f has a local maximum at b, then $f(x) \leq f(b)$ for $b - \delta < x \leq b$, so $f'(b^-) \geq 0$. The signs are reversed for local minima at the endpoints. $\qquad \square$

In searching for extreme values of a function, it is convenient to introduce the following classification of points in the domain of the function.

Definition 8.29. Suppose that $f : A \subset \mathbb{R} \to \mathbb{R}$. An interior point $c \in A$ such that f is not differentiable at c or $f'(c) = 0$ is called a critical point of f. An interior point where $f'(c) = 0$ is called a stationary point of f.

Theorem 8.27 limits the search for maxima or minima of a function f on A to the following points.

(1) Boundary points of A.
(2) Critical points of f:
    (a) interior points where f is not differentiable;
    (b) stationary points where $f'(c) = 0$.

Additional tests are required to determine which of these points gives local or global extreme values of f. In particular, a function need not attain an extreme value at a critical point.

Example 8.30. If $f : [-1, 1] \to \mathbb{R}$ is the function

$$f(x) = \begin{cases} x & \text{if } -1 \leq x \leq 0, \\ 2x & \text{if } 0 < x \leq 1, \end{cases}$$

then $x = 0$ is a critical point since f is not differentiable at 0, but f does not attain a local extreme value at 0. The global maximum and minimum of f are attained at the endpoints $x = 1$ and $x = -1$, respectively, and f has no other local extreme values.

Example 8.31. If $f : [-1, 1] \to \mathbb{R}$ is the function $f(x) = x^3$, then $x = 0$ is a critical point since $f'(0) = 0$, but f does not attain a local extreme value at 0. The global maximum and minimum of f are attained at the endpoints $x = 1$ and $x = -1$, respectively, and f has no other local extreme values.

8.5. The mean value theorem

The mean value theorem is a key result that connects the global behavior of a function $f : [a, b] \to \mathbb{R}$, described by the difference $f(b) - f(a)$, to its local behavior, described by the derivative $f' : (a, b) \to \mathbb{R}$. We begin by proving a special case.

Theorem 8.32 (Rolle)**.** Suppose that $f : [a, b] \to \mathbb{R}$ is continuous on the closed, bounded interval $[a, b]$, differentiable on the open interval (a, b), and $f(a) = f(b)$. Then there exists $a < c < b$ such that $f'(c) = 0$.

Proof. By the Weierstrass extreme value theorem, Theorem 7.37, f attains its global maximum and minimum values on $[a, b]$. If these are both attained at the endpoints, then f is constant, and $f'(c) = 0$ for every $a < c < b$. Otherwise, f attains at least one of its global maximum or minimum values at an interior point $a < c < b$. Theorem 8.27 implies that $f'(c) = 0$. $\qquad\square$

Note that we require continuity on the closed interval $[a, b]$ but differentiability only on the open interval (a, b). This proof is deceptively simple, but the result is not trivial. It relies on the extreme value theorem, which in turn relies on the completeness of $\mathbb{R}$. The theorem would not be true if we restricted attention to functions defined on the rationals $\mathbb{Q}$.

The mean value theorem is an immediate consequence of Rolle's theorem: for a general function f with $f(a) \neq f(b)$, we subtract off a linear function to make the values of the resulting function equal at the endpoints.

Theorem 8.33 (Mean value). Suppose that $f : [a, b] \to \mathbb{R}$ is continuous on the closed, bounded interval $[a, b]$ and differentiable on the open interval (a, b). Then there exists $a < c < b$ such that
$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Proof. The function $g : [a, b] \to \mathbb{R}$ defined by
$$g(x) = f(x) - f(a) - \left[\frac{f(b) - f(a)}{b - a} \right] (x - a)$$
is continuous on $[a, b]$ and differentiable on (a, b) with
$$g'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}.$$
Moreover, $g(a) = g(b) = 0$. Rolle's Theorem implies that there exists $a < c < b$ such that $g'(c) = 0$, which proves the result. $\qquad\square$

Graphically, this result says that there is point $a < c < b$ at which the slope of the tangent line to the graph $y = f(x)$ is equal to the slope of the chord between the endpoints $(a, f(a))$ and $(b, f(b))$.

As a first application, we prove a converse to the obvious fact that the derivative of a constant functions is zero.

Theorem 8.34. If $f : (a, b) \to \mathbb{R}$ is differentiable on (a, b) and $f'(x) = 0$ for every $a < x < b$, then f is constant on (a, b).

Proof. Fix $x_0 \in (a, b)$. The mean value theorem implies that for all $x \in (a, b)$ with $x \neq x_0$
$$f'(c) = \frac{f(x) - f(x_0)}{x - x_0}$$
for some c between x_0 and x. Since $f'(c) = 0$, it follows that $f(x) = f(x_0)$ for all $x \in (a, b)$, meaning that f is constant on (a, b). $\qquad\square$

Corollary 8.35. If $f, g : (a, b) \to \mathbb{R}$ are differentiable on (a, b) and $f'(x) = g'(x)$ for every $a < x < b$, then $f(x) = g(x) + C$ for some constant C.

Proof. This follows from the previous theorem since $(f - g)' = 0$. $\qquad\square$

We can also use the mean value theorem to relate the monotonicity of a differentiable function with the sign of its derivative. (See Definition 7.54 for our terminology for increasing and decreasing functions.)

Theorem 8.36. Suppose that $f : (a, b) \to \mathbb{R}$ is differentiable on (a, b). Then f is increasing if and only if $f'(x) \geq 0$ for every $a < x < b$, and decreasing if and only if $f'(x) \leq 0$ for every $a < x < b$. Furthermore, if $f'(x) > 0$ for every $a < x < b$ then f is strictly increasing, and if $f'(x) < 0$ for every $a < x < b$ then f is strictly decreasing.

Proof. If f is increasing and $a < x < b$, then

$$\frac{f(x + h) - f(x)}{h} \geq 0$$

for all sufficiently small h (positive or negative), so

$$f'(x) = \lim_{h \to 0} \left[\frac{f(x + h) - f(x)}{h} \right] \geq 0.$$

Conversely if $f' \geq 0$ and $a < x < y < b$, then by the mean value theorem there exists $x < c < y$ such that

$$\frac{f(y) - f(x)}{y - x} = f'(c) \geq 0,$$

which implies that $f(x) \leq f(y)$, so f is increasing. Moreover, if $f'(c) > 0$, we get $f(x) < f(y)$, so f is strictly increasing.

The results for a decreasing function f follow in a similar way, or we can apply of the previous results to the increasing function $-f$. $\square$

Note that although $f' > 0$ implies that f is strictly increasing, f is strictly increasing does not imply that $f' > 0$.

Example 8.37. The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3$ is strictly increasing on $\mathbb{R}$, but $f'(0) = 0$.

If f is continuously differentiable and $f'(c) > 0$, then $f'(x) > 0$ for all x in a neighborhood of c and Theorem 8.36 implies that f is strictly increasing near c. This conclusion may fail if f is not continuously differentiable at c.

Example 8.38. Define $f : \mathbb{R} \to \mathbb{R}$ by

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

Then f is differentiable on $\mathbb{R}$ with

$$f'(x) = \begin{cases} 1/2 - \cos(1/x) + 2x \sin(1/x) & \text{if } x \neq 0, \\ 1/2 & \text{if } x = 0. \end{cases}$$

Every neighborhood of 0 includes intervals where $f' < 0$ or $f' > 0$, in which f is strictly decreasing or strictly increasing, respectively. Thus, despite the fact that $f'(0) > 0$, the function f is not strictly increasing in any neighborhood of 0. As a result, no local inverse of the function f exists on any neighborhood of 0.

8.6. Taylor's theorem

If $f : (a, b) \to \mathbb{R}$ is differentiable on (a, b) and $f' : (a, b) \to \mathbb{R}$ is differentiable, then we define the second derivative $f'' : (a, b) \to \mathbb{R}$ of f as the derivative of f'. We define higher-order derivatives similarly. If f has derivatives $f^{(n)} : (a, b) \to \mathbb{R}$ of all orders $n \in \mathbb{N}$, then we say that f is infinitely differentiable on (a, b).

Taylor's theorem gives an approximation for an $(n + 1)$-times differentiable function in terms of its Taylor polynomial of degree n.

Definition 8.39. Let $f : (a, b) \to \mathbb{R}$ and suppose that f has n derivatives

$$f', f'', \ldots f^{(n)} : (a, b) \to \mathbb{R}$$

on (a, b). The Taylor polynomial of degree n of f at $a < c < b$ is

$$P_n(x) = f(c) + f'(c)(x - c) + \frac{1}{2!}f''(c)(x - c)^2 + \cdots + \frac{1}{n!}f^{(n)}(c)(x - c)^n.$$

Equivalently,

$$P_n(x) = \sum_{k=0}^{n} a_k(x - c)^k, \qquad a_k = \frac{1}{k!}f^{(k)}(c).$$

We call a_k the kth Taylor coefficient of f at c. The computation of the Taylor polynomials in the following examples are left as an exercise.

Example 8.40. If $P(x)$ is a polynomial of degree n, then $P_n(x) = P(x)$.

Example 8.41. The Taylor polynomial of degree n of e^x at $x = 0$ is

$$P_n(x) = 1 + x + \frac{1}{2!}x^2 \cdots + \frac{1}{n!}x^n.$$

Example 8.42. The Taylor polynomial of degree $2n$ of $\cos x$ at $x = 0$ is

$$P_{2n}(x) = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \cdots + (-1)^n \frac{1}{(2n)!}x^{2n}.$$

We also have $P_{2n+1} = P_{2n}$ since the Tayor coefficients of odd order are zero.

Example 8.43. The Taylor polynomial of degree $2n + 1$ of $\sin x$ at $x = 0$ is

$$P_{2n+1}(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \cdots + (-1)^n \frac{1}{(2n + 1)!}x^{2n+1}.$$

We also have $P_{2n+2} = P_{2n+1}$.

Example 8.44. The Taylor polynomial of degree n of $1/x$ at $x = 1$ is

$$P_n(x) = 1 - (x - 1) + (x - 1)^2 - \cdots + (-1)^n(x - 1)^n.$$

Example 8.45. The Taylor polynomial of degree n of $\log x$ at $x = 1$ is

$$P_n(x) = (x - 1) - \frac{1}{2}(x - 1)^2 + \frac{1}{3}(x - 1)^3 - \cdots + (-1)^{n+1}(x - 1)^n.$$

We write

$$f(x) = P_n(x) + R_n(x).$$

where R_n is the error, or remainder, between f and its Taylor polynomial P_n. The next theorem is one version of Taylor's theorem, which gives an expression for the remainder due to Lagrange. It can be regarded as a generalization of the mean value theorem, which corresponds to the case $n = 1$. The idea of the proof is to subtract a suitable polynomial from the function and apply Rolle's theorem, just as we proved the mean value theorem by subtracting a suitable linear function.

Theorem 8.46 (Taylor with Lagrange Remainder). Suppose that $f : (a, b) \to \mathbb{R}$ has $n + 1$ derivatives on (a, b) and let $a < c < b$. For every $a < x < b$, there exists ξ between c and x such that

$$f(x) = f(c) + f'(c)(x - c) + \frac{1}{2!}f''(c)(x - c)^2 + \cdots + \frac{1}{n!}f^{(n)}(c)(x - c)^n + R_n(x)$$

where

$$R_n(x) = \frac{1}{(n + 1)!}f^{(n+1)}(\xi)(x - c)^{n+1}.$$

Proof. Fix $x, c \in (a, b)$. For $t \in (a, b)$, let

$$g(t) = f(x) - f(t) - f'(t)(x - t) - \frac{1}{2!}f''(t)(x - t)^2 - \cdots - \frac{1}{n!}f^{(n)}(t)(x - t)^n.$$

Then $g(x) = 0$ and

$$g'(t) = -\frac{1}{n!}f^{(n+1)}(t)(x - t)^n.$$

Define

$$h(t) = g(t) - \left(\frac{x - t}{x - c}\right)^{n+1} g(c).$$

Then $h(c) = h(x) = 0$, so by Rolle's theorem, there exists a point ξ between c and x such that $h'(\xi) = 0$, which implies that

$$g'(\xi) + (n + 1)\frac{(x - \xi)^n}{(x - c)^{n+1}}g(c) = 0.$$

It follows from the expression for g' that

$$\frac{1}{n!}f^{(n+1)}(\xi)(x - \xi)^n = (n + 1)\frac{(x - \xi)^n}{(x - c)^{n+1}}g(c),$$

and using the expression for g in this equation, we get the result. $\qquad\square$

Note that the remainder term

$$R_n(x) = \frac{1}{(n + 1)!}f^{(n+1)}(\xi)(x - c)^{n+1}$$

has the same form as the $(n + 1)$th-term in the Taylor polynomial of f, except that the derivative is evaluated at a (typically unknown) intermediate point ξ between c and x, instead of at c.

Example 8.47. Let us prove that

$$\lim_{x \to 0}\left(\frac{1 - \cos x}{x^2}\right) = \frac{1}{2}.$$

By Taylor's theorem,

$$\cos x = 1 - \frac{1}{2}x^2 + \frac{1}{4!}(\cos \xi)x^4$$

for some ξ between 0 and x. It follows that for $x \neq 0$,

$$\frac{1 - \cos x}{x^2} - \frac{1}{2} = -\frac{1}{4!}(\cos \xi)x^2.$$

Since $|\cos \xi| \leq 1$, we get

$$\left|\frac{1 - \cos x}{x^2} - \frac{1}{2}\right| \leq \frac{1}{4!}x^2,$$

which implies that
$$\lim_{x \to 0} \left| \frac{1 - \cos x}{x^2} - \frac{1}{2} \right| = 0.$$
Note that as well as proving the limit, Taylor's theorem gives an explicit upper bound for the difference between $(1 - \cos x)/x^2$ and its limit $1/2$. For example,
$$\left| \frac{1 - \cos(0.1)}{(0.1)^2} - \frac{1}{2} \right| \leq \frac{1}{2400}.$$
Numerically, we have
$$\frac{1}{2} - \frac{1 - \cos(0.1)}{(0.1)^2} \approx 0.00041653, \qquad \frac{1}{2400} \approx 0.00041667.$$

In Section 12.7, we derive an alternative expression for the remainder R_n as an integral.

8.7. * The inverse function theorem

The inverse function theorem gives a sufficient condition for a differentiable function f to be locally invertible at a point c with differentiable inverse: namely, that f is *continuously* differentiable at c and $f'(c) \neq 0$. Example 8.24 shows that one cannot expect the inverse of a differentiable function f to exist locally at c if $f'(c) = 0$, while Example 8.38 shows that the condition $f'(c) \neq 0$ is not, on its own, sufficient to imply the existence of a local inverse.

Before stating the theorem, we give a precise definition of local invertibility.

Definition 8.48. A function $f : A \to \mathbb{R}$ is locally invertible at an interior point $c \in A$ if there exist open neighborhoods U of c and V of $f(c)$ such that $f|_U : U \to V$ is one-to-one and onto, in which case f has a local inverse $(f|_U)^{-1} : V \to U$.

The following examples illustrate the definition.

Example 8.49. If $f : \mathbb{R} \to \mathbb{R}$ is the square function $f(x) = x^2$, then a local inverse at $c = 2$ with $U = (1, 3)$ and $V = (1, 9)$ is defined by
$$(f|_U)^{-1}(y) = \sqrt{y}.$$
Similarly, a local inverse at $c = -2$ with $U = (-3, -1)$ and $V = (1, 9)$ is defined by
$$(f|_U)^{-1}(y) = -\sqrt{y}.$$
In defining a local inverse at c, we require that it maps an open neighborhood V of $f(c)$ onto an open neighborhood U of c; that is, we want $(f|_U)^{-1}(y)$ to be "close" to c when y is "close" to $f(c)$, not some more distant point that f also maps "close" to $f(c)$. Thus, the one-to-one, onto function g defined by
$$g : (1, 9) \to (-2, -1) \cup [2, 3), \qquad g(y) = \begin{cases} -\sqrt{y} & \text{if } 1 < y < 4 \\ \sqrt{y} & \text{if } 4 \leq y < 9 \end{cases}$$
is not a local inverse of f at $c = 2$ in the sense of Definition 8.48, even though $g(f(2)) = 2$ and both compositions
$$f \circ g : (1, 9) \to (1, 9), \qquad g \circ f : (-2, -1) \cup [2, 3) \to (-2, -1) \cup [2, 3)$$
are identity maps, since $U = (-2, -1) \cup [2, 3)$ is not a neighborhood of 2.

Example 8.50. The function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} \cos(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

is locally invertible at every $c \in \mathbb{R}$ with $c \neq 0$ or $c \neq 1/(n\pi)$ for some $n \in \mathbb{Z}$.

Theorem 8.51 (Inverse function)**.** Suppose that $f : A \subset \mathbb{R} \to \mathbb{R}$ and $c \in A$ is an interior point of A. If f is differentiable in a neighborhood of c, $f'(c) \neq 0$, and f' is continuous at c, then there are open neighborhoods U of c and V of $f(c)$ such that f has a local inverse $(f|_U)^{-1} : V \to U$. Furthermore, the local inverse function is differentiable at $f(c)$ with derivative

$$[(f|_U)^{-1}]'(f(c)) = \frac{1}{f'(c)}.$$

.

Proof. Suppose, for definiteness, that $f'(c) > 0$ (otherwise, consider $-f$). By the continuity of f', there exists an open interval $U = (a, b)$ containing c on which $f' > 0$. It follows from Theorem 8.36 that f is strictly increasing on U. Writing

$$V = f(U) = (f(a), f(b)),$$

we see that $f|_U : U \to V$ is one-to-one and onto, so f has a local inverse on V, which proves the first part of the theorem.

It remains to prove that the local inverse $(f|_U)^{-1}$, which we denote by f^{-1} for short, is differentiable. First, since f is differentiable at c, we have

$$f(c + h) = f(c) + f'(c)h + r(h)$$

where the remainder r satisfies

$$\lim_{h \to 0} \frac{r(h)}{h} = 0.$$

Since $f'(c) > 0$, there exists $\delta > 0$ such that

$$|r(h)| \leq \frac{1}{2} f'(c)|h| \qquad \text{for } |h| < \delta.$$

It follows from the differentiability of f that, if $|h| < \delta$,

$$\begin{aligned} f'(c)|h| &= |f(c + h) - f(c) - r(h)| \\ &\leq |f(c + h) - f(c)| + |r(h)| \\ &\leq |f(c + h) - f(c)| + \frac{1}{2} f'(c)|h|. \end{aligned}$$

Absorbing the term proportional to $|h|$ on the right hand side of this inequality into the left hand side and writing

$$f(c + h) = f(c) + k,$$

we find that

$$\frac{1}{2} f'(c)|h| \leq |k| \qquad \text{for } |h| < \delta.$$

Choosing $\delta > 0$ small enough that $(c - \delta, c + \delta) \subset U$, we can express h in terms of k as

$$h = f^{-1}(f(c) + k) - f^{-1}(f(c)).$$

Using this expression in the expansion of f evaluated at $c + h$,

$$f(c + h) = f(c) + f'(c)h + r(h),$$

we get that

$$f(c) + k = f(c) + f'(c) \left[f^{-1}(f(c) + k) - f^{-1}(f(c)) \right] + r(h).$$

Simplifying and rearranging this equation, we obtain the corresponding expansion for f^{-1} evaluated at $f(c) + k$,

$$f^{-1}(f(c) + k) = f^{-1}(f(c)) + \frac{1}{f'(c)} k + s(k),$$

where the remainder s is given by

$$s(k) = -\frac{1}{f'(c)} r(h) = -\frac{1}{f'(c)} r \left(f^{-1}(f(c) + k) - f^{-1}(f(c)) \right).$$

Since $f'(c)|h|/2 \leq |k|$, it follows that

$$\frac{|s(k)|}{|k|} \leq \frac{2}{f'(c)^2} \frac{|r(h)|}{|h|}.$$

Therefore, by the "sandwich" theorem and the fact that $h \to 0$ as $k \to 0$,

$$\lim_{k \to 0} \frac{|s(k)|}{|k|} = 0.$$

This result proves that f^{-1} is differentiable at $f(c)$ with

$$\left[f^{-1} \left(f(c) \right) \right]' = \frac{1}{f'(c)}.$$

The expression for the derivative of the inverse also follows from Proposition 8.23, but only once we know that f^{-1} is dfferentiable at $f(c)$. $\qquad\square$

One can show that Theorem 8.51 remains true under the weaker hypothesis that the derivative exists and is nonzero in an open neighborhood of c, but in practise, we almost always apply the theorem to continuously differentiable functions.

The inverse function theorem generalizes to functions of several variables, $f : A \subset \mathbb{R}^n \to \mathbb{R}^n$, with a suitable generalization of the derivative of f at c as the linear map $f'(c) : \mathbb{R}^n \to \mathbb{R}^n$ that approximates f near c. A different proof of the existence of a local inverse is required in that case, since one cannot use monotonicity arguments.

As an example of the application of the inverse function theorem, we consider a simple problem from bifurcation theory.

Example 8.52. Consider the transcendental equation

$$y = x - k \left(e^x - 1 \right)$$

where $k \in \mathbb{R}$ is a constant parameter. Suppose that we want to solve for $x \in \mathbb{R}$ given $y \in \mathbb{R}$. If $y = 0$, then an obvious solution is $x = 0$. The inverse function theorem applied to the continuously differentiable function $f(x; k) = x - k(e^x - 1)$ implies that there are neighborhoods U, V of 0 (depending on k) such that the equation has a unique solution $x \in U$ for every $y \in V$ provided that the derivative
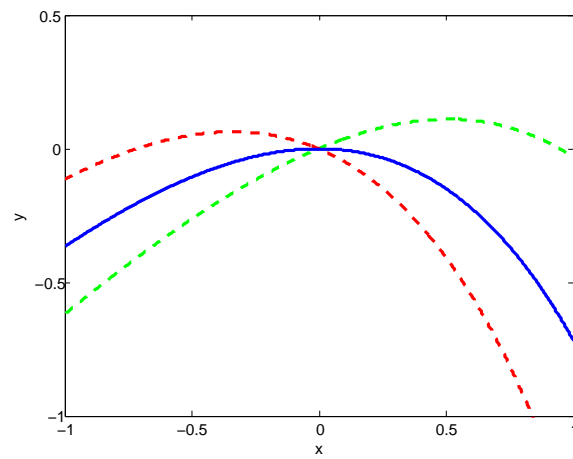
Figure 2. Graph of $y = f(x; k)$ for the function in Example 8.52: (a) $k = 0.5$ (green); (b) $k = 1$ (blue); (c) $k = 1.5$ (red). When y is sufficiently close to zero, there is a unique solution for x in some neighborhood of zero unless $k = 1$.

of f with respect to x at 0, given by $f_x(0; k) = 1 - k$ is non-zero i.e., provided that $k \neq 1$ (see Figure 2).
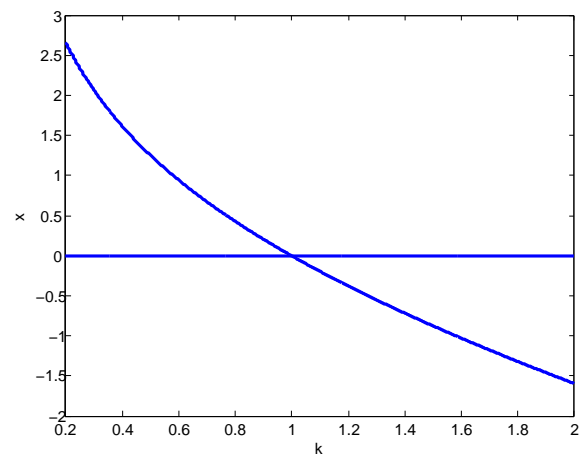
Figure 3. Plot of the solutions for x of the nonlinear equation $x = k(e^x - 1)$ as a function of the parameter k (see Example 8.52). The point $(x, k) = (0, 1)$ where the two solution branches cross is called a bifurcation point.

Alternatively, we can fix a value of y, say $y = 0$, and ask how the solutions of the corresponding equation for x,

$$x - k\,(e^x - 1) = 0,$$

depend on the parameter k. Figure 2 plots the solutions for x as a function of k for $0.2 \leq k \leq 2$. The equation has two different solutions for x unless $k = 1$. The branch of nonzero solutions crosses the branch of zero solution at the point $(x, k) = (0, 1)$, called a bifurcation point. The implicit function theorem, which is a generalization of the inverse function theorem, implies that a necessary condition for a solution (x_0, k_0) of the equation $f(x; k) = 0$ to be a bifurcation point, meaning that the equation fails to have a unique solution branch $x = g(k)$ in some neighborhood of (x_0, k_0), is that $f_x(x_0; k_0) = 0$.

8.8. * L'Hôpital's rule

In this section, we prove a rule (much beloved by calculus students) for the evaluation of inderminate limits of the form $0/0$ or ∞/∞. Our proof uses the following generalization of the mean value theorem.

Theorem 8.53 (Cauchy mean value)**.** Suppose that $f, g : [a, b] \to \mathbb{R}$ are continuous on the closed, bounded interval $[a, b]$ and differentiable on the open interval (a, b). Then there exists $a < c < b$ such that

$$f'(c) \left[g(b) - g(a) \right] = \left[f(b) - f(a) \right] g'(c).$$

Proof. The function $h : [a, b] \to \mathbb{R}$ defined by

$$h(x) = \left[f(x) - f(a) \right] \left[g(b) - g(a) \right] - \left[f(b) - f(a) \right] \left[g(x) - g(a) \right]$$

is continuous on $[a, b]$ and differentiable on (a, b) with

$$h'(x) = f'(x) \left[g(b) - g(a) \right] - \left[f(b) - f(a) \right] g'(x).$$

Moreover, $h(a) = h(b) = 0$. Rolle's Theorem implies that there exists $a < c < b$ such that $h'(c) = 0$, which proves the result. $\qquad\square$

If $g(x) = x$, then this theorem reduces to the usual mean value theorem (Theorem 8.33). Next, we state one form of l'Hôpital's rule.

Theorem 8.54 (l'Hôpital's rule: $0/0$)**.** Suppose that $f, g : (a, b) \to \mathbb{R}$ are differentiable functions on a bounded open interval (a, b) such that $g'(x) \neq 0$ for $x \in (a, b)$ and

$$\lim_{x \to a^+} f(x) = 0, \qquad \lim_{x \to a^+} g(x) = 0.$$

Then

$$\lim_{x \to a^+} \frac{f'(x)}{g'(x)} = L \quad \text{implies that} \quad \lim_{x \to a^+} \frac{f(x)}{g(x)} = L.$$

Proof. We may extend $f, g : [a, b) \to \mathbb{R}$ to continuous functions on $[a, b)$ by defining $f(a) = g(a) = 0$. If $a < x < b$, then by the mean value theorem, there exists $a < c < x$ such that

$$g(x) = g(x) - g(a) = g'(c)(x - a) \neq 0,$$

so $g \neq 0$ on (a, b). Moreover, by the Cauchy mean value theorem (Theorem 8.53), there exists $a < c < x$ such that

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f'(c)}{g'(c)}.$$

Since $c \to a^+$ as $x \to a^+$, the result follows. (In fact, since $a < c < x$, the δ that "works" for f'/g' also "works" for f/g.) $\qquad\square$

Example 8.55. Using l'Hôpital's rule twice (verify that all of the hypotheses are satisfied!), we find that

$$\lim_{x \to 0^+} \frac{1 - \cos x}{x^2} = \lim_{x \to 0^+} \frac{\sin x}{2x} = \lim_{x \to 0^+} \frac{\cos x}{2} = \frac{1}{2}.$$

Analogous results and proofs apply to left limits ($x \to a^-$), two-sided limits ($x \to a$), and infinite limits ($x \to \infty$ or $x \to -\infty$). Alternatively, one can reduce these limits to the left limit considered in Theorem 8.54.

For example, suppose that $f, g : (a, \infty) \to \mathbb{R}$ are differentiable, $g' \neq 0$, and $f(x) \to 0$, $g(x) \to 0$ as $x \to \infty$. Assuming that $a > 0$ without loss of generality, we define $F, G : (0, 1/a) \to \mathbb{R}$ by

$$F(t) = f\left(\frac{1}{t}\right), \qquad G(t) = g\left(\frac{1}{t}\right).$$

The chain rule implies that

$$F'(t) = -\frac{1}{t^2} f'\left(\frac{1}{t}\right), \qquad G'(t) = -\frac{1}{t^2} g'\left(\frac{1}{t}\right).$$

Replacing limits as $x \to \infty$ by equivalent limits as $t \to 0^+$ and applying Theorem 8.54 to F, G, all of whose hypothesis are satisfied if the limit of $f'(x)/g'(x)$ as $x \to \infty$ exists, we get

$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = \lim_{t \to 0^+} \frac{F(t)}{G(t)} = \lim_{t \to 0^+} \frac{F'(t)}{G'(t)} = \lim_{x \to \infty} \frac{f'(x)}{g'(x)}.$$

A less straightforward generalization is to the case when g and possibly f have infinite limits as $x \to a^+$. In that case, we cannot simply extend f and g by continuity to the point a. Instead, we introduce two points $a < x < y < b$ and consider the limits $x \to a^+$ followed by $y \to a^+$.

Theorem 8.56 (l'Hôpital's rule: ∞/∞). Suppose that $f, g : (a, b) \to \mathbb{R}$ are differentiable functions on a bounded open interval (a, b) such that $g'(x) \neq 0$ for $x \in (a, b)$ and

$$\lim_{x \to a^+} |g(x)| = \infty.$$

Then

$$\lim_{x \to a^+} \frac{f'(x)}{g'(x)} = L \quad \text{implies that} \quad \lim_{x \to a^+} \frac{f(x)}{g(x)} = L.$$

Proof. Since $|g(x)| \to \infty$ as $x \to a^+$, we have $g \neq 0$ near a, and we may assume without loss of generality that $g \neq 0$ on (a, b). If $a < x < y < b$, then the mean value theorem implies that $g(x) - g(y) \neq 0$, since $g' \neq 0$, and the Cauchy mean value theorem implies that there exists $x < c < y$ such that

$$\frac{f(x) - f(y)}{g(x) - g(y)} = \frac{f'(c)}{g'(c)}.$$

We may therefore write

$$\frac{f(x)}{g(x)} = \left[\frac{f(x)-f(y)}{g(x)-g(y)}\right]\left[\frac{g(x)-g(y)}{g(x)}\right] + \frac{f(y)}{g(x)}$$

$$= \frac{f'(c)}{g'(c)}\left[1 - \frac{g(y)}{g(x)}\right] + \frac{f(y)}{g(x)}.$$

It follows that

$$\left|\frac{f(x)}{g(x)} - L\right| \le \left|\frac{f'(c)}{g'(c)} - L\right| + \left|\frac{f'(c)}{g'(c)}\right|\left|\frac{g(y)}{g(x)}\right| + \left|\frac{f(y)}{g(x)}\right|.$$

Given $\epsilon > 0$, choose $\delta > 0$ such that

$$\left|\frac{f'(c)}{g'(c)} - L\right| < \epsilon \qquad \text{for } a < c < a + \delta.$$

Then, since $a < c < y$, we have for all $a < x < y < a + \delta$ that

$$\left|\frac{f(x)}{g(x)} - L\right| < \epsilon + (|L| + \epsilon)\left|\frac{g(y)}{g(x)}\right| + \left|\frac{f(y)}{g(x)}\right|.$$

Fixing y, taking the lim sup of this inequality as $x \to a^+$, and using the assumption that $|g(x)| \to \infty$, we find that

$$\limsup_{x \to a^+}\left|\frac{f(x)}{g(x)} - L\right| \le \epsilon.$$

Since $\epsilon > 0$ is arbitrary, we have

$$\limsup_{x \to a^+}\left|\frac{f(x)}{g(x)} - L\right| = 0,$$

which proves the result.

Alternatively, instead of using the lim sup, we can verify the limit explicitly by an "$\epsilon/3$"-argument. Given $\epsilon > 0$, choose $\eta > 0$ such that

$$\left|\frac{f'(c)}{g'(c)} - L\right| < \frac{\epsilon}{3} \qquad \text{for } a < c < a + \eta,$$

choose $a < y < a + \eta$, and let $\delta_1 = y - a > 0$. Next, choose $\delta_2 > 0$ such that

$$|g(x)| > \frac{3}{\epsilon}\left(|L| + \frac{\epsilon}{3}\right)|g(y)| \qquad \text{for } a < x < a + \delta_2,$$

and choose $\delta_3 > 0$ such that

$$|g(x)| > \frac{3}{\epsilon}|f(y)| \qquad \text{for } a < x < a + \delta_3.$$

Let $\delta = \min(\delta_1, \delta_2, \delta_3) > 0$. Then for $a < x < a + \delta$, we have

$$\left|\frac{f(x)}{g(x)} - L\right| \le \left|\frac{f'(c)}{g'(c)} - L\right| + \left|\frac{f'(c)}{g'(c)}\right|\left|\frac{g(y)}{g(x)}\right| + \left|\frac{f(y)}{g(x)}\right| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3},$$

which proves the result.   □

We often use this result when both $f(x)$ and $g(x)$ diverge to infinity as $x \to a^+$, but no assumption on the behavior of $f(x)$ is required.

As for the previous theorem, analogous results and proofs apply to other limits $(x \to a^-, x \to a, \text{ or } x \to \pm\infty)$. There are also versions of l'Hôpital's rule that imply the divergence of $f(x)/g(x)$ to $\pm\infty$, but we consider here only the case of a finite limit L.

Example 8.57. Since $e^x \to \infty$ as $x \to \infty$, we get by l'Hôpital's rule that

$$\lim_{x\to\infty} \frac{x}{e^x} = \lim_{x\to\infty} \frac{1}{e^x} = 0.$$

Similarly, since $x \to \infty$ as as $x \to \infty$, we get by l'Hôpital's rule that

$$\lim_{x\to\infty} \frac{\log x}{x} = \lim_{x\to\infty} \frac{1/x}{1} = 0.$$

That is, e^x grows faster than x and $\log x$ grows slower than x as $x \to \infty$. We also write these limits using "little oh" notation as $x = o(e^x)$ and $\log x = o(x)$ as $x \to \infty$.

Finally, we note that one cannot use l'Hôpital's rule "in reverse" to deduce that f'/g' has a limit if f/g has a limit.

Example 8.58. Let $f(x) = x + \sin x$ and $g(x) = x$. Then $f(x), g(x) \to \infty$ as $x \to \infty$ and

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = \lim_{x\to\infty} \left(1 + \frac{\sin x}{x}\right) = 1,$$

but the limit

$$\lim_{x\to\infty} \frac{f'(x)}{g'(x)} = \lim_{x\to\infty} (1 + \cos x)$$

does not exist.

Sequences and Series of Functions

In this chapter, we define and study the convergence of sequences and series of functions. There are many different ways to define the convergence of a sequence of functions, and different definitions lead to inequivalent types of convergence. We consider here two basic types: pointwise and uniform convergence.

9.1. Pointwise convergence

Pointwise convergence defines the convergence of functions in terms of the convergence of their values at each point of their domain.

Definition 9.1. Suppose that (f_n) is a sequence of functions $f_n : A \to \mathbb{R}$ and $f : A \to \mathbb{R}$. Then $f_n \to f$ pointwise on A if $f_n(x) \to f(x)$ as $n \to \infty$ for every $x \in A$.

We say that the sequence (f_n) converges pointwise if it converges pointwise to some function f, in which case

$$f(x) = \lim_{n \to \infty} f_n(x).$$

Pointwise convergence is, perhaps, the most obvious way to define the convergence of functions, and it is one of the most important. Nevertheless, as the following examples illustrate, it is not as well-behaved as one might initially expect.

Example 9.2. Suppose that $f_n : (0, 1) \to \mathbb{R}$ is defined by

$$f_n(x) = \frac{n}{nx + 1}.$$

Then, since $x \neq 0$,

$$\lim_{n \to \infty} f_n(x) = \lim_{n \to \infty} \frac{1}{x + 1/n} = \frac{1}{x},$$

so $f_n \to f$ pointwise where $f : (0, 1) \to \mathbb{R}$ is given by

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

We have $|f_n(x)| < n$ for all $x \in (0, 1)$, so each f_n is bounded on $(0, 1)$, but the pointwise limit f is not. Thus, pointwise convergence does not, in general, preserve boundedness.

Example 9.3. Suppose that $f_n : [0, 1] \to \mathbb{R}$ is defined by $f_n(x) = x^n$. If $0 \leq x < 1$, then $x^n \to 0$ as $n \to \infty$, while if $x = 1$, then $x^n \to 1$ as $n \to \infty$. So $f_n \to f$ pointwise where

$$f(x) = \begin{cases} 0 & \text{if } 0 \leq x < 1, \\ 1 & \text{if } x = 1. \end{cases}$$

Although each f_n is continuous on $[0, 1]$, the pointwise limit f is not (it is discontinuous at 1). Thus, pointwise convergence does not, in general, preserve continuity.

Example 9.4. Define $f_n : [0, 1] \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} 2n^2 x & \text{if } 0 \leq x \leq 1/(2n) \\ 2n^2(1/n - x) & \text{if } 1/(2n) < x < 1/n, \\ 0 & 1/n \leq x \leq 1. \end{cases}$$

If $0 < x \leq 1$, then $f_n(x) = 0$ for all $n \geq 1/x$, so $f_n(x) \to 0$ as $n \to \infty$; and if $x = 0$, then $f_n(x) = 0$ for all n, so $f_n(x) \to 0$ also. It follows that $f_n \to 0$ pointwise on $[0, 1]$. This is the case even though $\max f_n = n \to \infty$ as $n \to \infty$. Thus, a pointwise convergent sequence (f_n) of functions need not be uniformly bounded (that is, bounded independently of n), even if it converges to zero.

Example 9.5. Define $f_n : \mathbb{R} \to \mathbb{R}$ by

$$f_n(x) = \frac{\sin nx}{n}.$$

Then $f_n \to 0$ pointwise on $\mathbb{R}$. The sequence (f_n') of derivatives $f_n'(x) = \cos nx$ does not converge pointwise on $\mathbb{R}$; for example,

$$f_n'(\pi) = (-1)^n$$

does not converge as $n \to \infty$. Thus, in general, one cannot differentiate a pointwise convergent sequence. This behavior isn't limited to pointwise convergent sequences, and happens because the derivative of a small, rapidly oscillating function can be large.

Example 9.6. Define $f_n : \mathbb{R} \to \mathbb{R}$ by

$$f_n(x) = \frac{x^2}{\sqrt{x^2 + 1/n}}.$$

If $x \neq 0$, then

$$\lim_{n \to \infty} \frac{x^2}{\sqrt{x^2 + 1/n}} = \frac{x^2}{|x|} = |x|$$

while $f_n(0) = 0$ for all $n \in \mathbb{N}$, so $f_n \to |x|$ pointwise on $\mathbb{R}$. Moreover,

$$f_n'(x) = \frac{x^3 + 2x/n}{(x^2 + 1/n)^{3/2}} \to \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$$

The pointwise limit $|x|$ isn't differentiable at 0 even though all of the f_n are differentiable on $\mathbb{R}$ and the derivatives f_n' converge pointwise on $\mathbb{R}$. (The f_n's "round off" the corner in the absolute value function.)

Example 9.7. Define $f_n : \mathbb{R} \to \mathbb{R}$ by

$$f_n(x) = \left(1 + \frac{x}{n}\right)^n.$$

Then, by the limit formula for the exponential, $f_n \to e^x$ pointwise on $\mathbb{R}$.

9.2. Uniform convergence

In this section, we introduce a stronger notion of convergence of functions than pointwise convergence, called uniform convergence. The difference between pointwise convergence and uniform convergence is analogous to the difference between continuity and uniform continuity.

Definition 9.8. Suppose that (f_n) is a sequence of functions $f_n : A \to \mathbb{R}$ and $f : A \to \mathbb{R}$. Then $f_n \to f$ uniformly on A if, for every $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$n > N \text{ implies that } |f_n(x) - f(x)| < \epsilon \text{ for all } x \in A.$$

When the domain A of the functions is understood, we will often say $f_n \to f$ uniformly instead of uniformly on A.

The crucial point in this definition is that N depends only on ϵ and not on $x \in A$, whereas for a pointwise convergent sequence N may depend on both ϵ and x. A uniformly convergent sequence is always pointwise convergent (to the same limit), but the converse is not true. If a sequence converges pointwise, it may happen that for some $\epsilon > 0$ one needs to choose arbitrarily large N's for different points $x \in A$, meaning that the sequences of values converge arbitrarily slowly on A. In that case a pointwise convergent sequence of functions is not uniformly convergent.

Example 9.9. The sequence $f_n(x) = x^n$ in Example 9.3 converges pointwise on $[0, 1]$ but not uniformly on $[0, 1]$. For $0 \le x < 1$, we have

$$|f_n(x) - f(x)| = x^n.$$

If $0 < \epsilon < 1$, we cannot make $x^n < \epsilon$ for all $0 \le x < 1$ however large we choose n. The problem is that x^n converges to 0 at an arbitrarily slow rate for x sufficiently close to 1. There is no difficulty in the rate of convergence at 1 itself, since $f_n(1) = 1$ for every $n \in \mathbb{N}$. As we will show, the uniform limit of continuous functions is continuous, so since the pointwise limit of the continuous functions f_n is discontinuous, the sequence cannot converge uniformly on $[0, 1]$. The sequence does, however, converge uniformly to 0 on $[0, b]$ for every $0 \le b < 1$; given $\epsilon > 0$, we take N large enough that $b^N < \epsilon$.

Example 9.10. The pointwise convergent sequence in Example 9.4 does not converge uniformly. If it did, it would have to converge to the pointwise limit 0, but

$$\left| f_n \left(\frac{1}{2n} \right) \right| = n,$$

so for no $\epsilon > 0$ does there exist an $N \in \mathbb{N}$ such that $|f_n(x) - 0| < \epsilon$ for all $x \in A$ and $n > N$, since this inequality fails for $n \geq \epsilon$ if $x = 1/(2n)$.

Example 9.11. The functions in Example 9.5 converge uniformly to 0 on $\mathbb{R}$, since

$$|f_n(x)| = \frac{|\sin nx|}{n} \leq \frac{1}{n},$$

so $|f_n(x) - 0| < \epsilon$ for all $x \in \mathbb{R}$ if $n > 1/\epsilon$.

9.3. Cauchy condition for uniform convergence

The Cauchy condition in Definition 3.45 provides a necessary and sufficient condition for a sequence of real numbers to converge. There is an analogous uniform Cauchy condition that provides a necessary and sufficient condition for a sequence of functions to converge uniformly.

Definition 9.12. A sequence (f_n) of functions $f_n : A \to \mathbb{R}$ is uniformly Cauchy on A if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$m, n > N \text{ implies that } |f_m(x) - f_n(x)| < \epsilon \text{ for all } x \in A.$$

The key part of the following proof is the argument to show that a pointwise convergent, uniformly Cauchy sequence converges uniformly.

Theorem 9.13. A sequence (f_n) of functions $f_n : A \to \mathbb{R}$ converges uniformly on A if and only if it is uniformly Cauchy on A.

Proof. Suppose that (f_n) converges uniformly to f on A. Then, given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$|f_n(x) - f(x)| < \frac{\epsilon}{2} \qquad \text{for all } x \in A \text{ if } n > N.$$

It follows that if $m, n > N$ then

$$|f_m(x) - f_n(x)| \leq |f_m(x) - f(x)| + |f(x) - f_n(x)| < \epsilon \qquad \text{for all } x \in A,$$

which shows that (f_n) is uniformly Cauchy.

Conversely, suppose that (f_n) is uniformly Cauchy. Then for each $x \in A$, the real sequence $(f_n(x))$ is Cauchy, so it converges by the completeness of $\mathbb{R}$. We define $f : A \to \mathbb{R}$ by

$$f(x) = \lim_{n \to \infty} f_n(x),$$

and then $f_n \to f$ pointwise.

To prove that $f_n \to f$ uniformly, let $\epsilon > 0$. Since (f_n) is uniformly Cauchy, we can choose $N \in \mathbb{N}$ (depending only on ϵ) such that

$$|f_m(x) - f_n(x)| < \frac{\epsilon}{2} \qquad \text{for all } x \in A \text{ if } m, n > N.$$

Let $n > N$ and $x \in A$. Then for every $m > N$ we have

$$|f_n(x) - f(x)| \leq |f_n(x) - f_m(x)| + |f_m(x) - f(x)| < \frac{\epsilon}{2} + |f_m(x) - f(x)|.$$

Since $f_m(x) \to f(x)$ as $m \to \infty$, we can choose $m > N$ (depending on x, but it doesn't matter since m doesn't appear in the final result) such that

$$|f_m(x) - f(x)| < \frac{\epsilon}{2}.$$

It follows that if $n > N$, then

$$|f_n(x) - f(x)| < \epsilon \qquad \text{for all } x \in A,$$

which proves that $f_n \to f$ uniformly.

Alternatively, we can take the limit as $m \to \infty$ in the uniform Cauchy condition to get for all $x \in A$ and $n > N$ that

$$|f(x) - f_n(x)| = \lim_{m \to \infty} |f_m(x) - f_n(x)| \leq \frac{\epsilon}{2} < \epsilon.$$

$\square$

9.4. Properties of uniform convergence

In this section we prove that, unlike pointwise convergence, uniform convergence preserves boundedness and continuity. Uniform convergence does not preserve differentiability any better than pointwise convergence. Nevertheless, we give a result that allows us to differentiate a convergent sequence; the key assumption is that the derivatives converge uniformly.

9.4.1. Boundedness. First, we consider the uniform convergence of bounded functions.

Theorem 9.14. Suppose that $f_n : A \to \mathbb{R}$ is bounded on A for every $n \in \mathbb{N}$ and $f_n \to f$ uniformly on A. Then $f : A \to \mathbb{R}$ is bounded on A.

Proof. Taking $\epsilon = 1$ in the definition of the uniform convergence, we find that there exists $N \in \mathbb{N}$ such that

$$|f_n(x) - f(x)| < 1 \qquad \text{for all } x \in A \text{ if } n > N.$$

Choose some $n > N$. Then, since f_n is bounded, there is a constant $M \geq 0$ such that

$$|f_n(x)| \leq M \qquad \text{for all } x \in A.$$

It follows that

$$|f(x)| \leq |f(x) - f_n(x)| + |f_n(x)| < 1 + M \qquad \text{for all } x \in A,$$

meaning that f is bounded on A. $\square$

In particular, it follows that if a sequence of bounded functions converges pointwise to an unbounded function, then the convergence is not uniform.

Example 9.15. The sequence of functions $f_n : (0,1) \to \mathbb{R}$ in Example 9.2, defined by

$$f_n(x) = \frac{n}{nx+1},$$

cannot converge uniformly on $(0,1)$, since each f_n is bounded on $(0,1)$, but the pointwise limit $f(x) = 1/x$ is not. The sequence (f_n) does, however, converge uniformly to f on every interval $[a,1)$ with $0 < a < 1$. To prove this, we estimate for $a \le x < 1$ that

$$|f_n(x) - f(x)| = \left| \frac{n}{nx+1} - \frac{1}{x} \right| = \frac{1}{x(nx+1)} < \frac{1}{nx^2} \le \frac{1}{na^2}.$$

Thus, given $\epsilon > 0$ choose $N = 1/(a^2\epsilon)$, and then

$$|f_n(x) - f(x)| < \epsilon \qquad \text{for all } x \in [a,1) \text{ if } n > N,$$

which proves that $f_n \to f$ uniformly on $[a,1)$. Note that

$$|f(x)| \le \frac{1}{a} \qquad \text{for all } x \in [a,1)$$

so the uniform limit f is bounded on $[a,1)$, as Theorem 9.14 requires.

9.4.2. Continuity. One of the most important properties of uniform convergence is that it preserves continuity. We use an "$\epsilon/3$" argument to get the continuity of the uniform limit f from the continuity of the f_n.

Theorem 9.16. If a sequence (f_n) of continuous functions $f_n : A \to \mathbb{R}$ converges uniformly on $A \subset \mathbb{R}$ to $f : A \to \mathbb{R}$, then f is continuous on A.

Proof. Suppose that $c \in A$ and let $\epsilon > 0$. Then, for every $n \in \mathbb{N}$,

$$|f(x) - f(c)| \le |f(x) - f_n(x)| + |f_n(x) - f_n(c)| + |f_n(c) - f(c)|.$$

By the uniform convergence of (f_n), we can choose $n \in \mathbb{N}$ such that

$$|f_n(x) - f(x)| < \frac{\epsilon}{3} \qquad \text{for all } x \in A,$$

and for such an n it follows that

$$|f(x) - f(c)| < |f_n(x) - f_n(c)| + \frac{2\epsilon}{3}.$$

(Here, we use the fact that f_n is close to f at both x and c, where x is an an arbitrary point in a neighborhood of c; this is where we use the uniform convergence in a crucial way.)

Since f_n is continuous on A, there exists $\delta > 0$ such that

$$|f_n(x) - f_n(c)| < \frac{\epsilon}{3} \qquad \text{if } |x - c| < \delta \text{ and } x \in A,$$

which implies that

$$|f(x) - f(c)| < \epsilon \qquad \text{if } |x - c| < \delta \text{ and } x \in A.$$

This proves that f is continuous. $\qquad\qquad\square$

This result can be interpreted as justifying an "exchange in the order of limits"

$$\lim_{n \to \infty} \lim_{x \to c} f_n(x) = \lim_{x \to c} \lim_{n \to \infty} f_n(x).$$

Such exchanges of limits always require some sort of condition for their validity — in this case, the uniform convergence of f_n to f is sufficient, but pointwise convergence is not.

It follows from Theorem 9.16 that if a sequence of continuous functions converges pointwise to a discontinuous function, as in Example 9.3, then the convergence is not uniform. The converse is not true, however, and the pointwise limit of a sequence of continuous functions may be continuous even if the convergence is not uniform, as in Example 9.4.

9.4.3. Differentiability. The uniform convergence of differentiable functions does not, in general, imply anything about the convergence of their derivatives or the differentiability of their limit. As noted above, this is because the values of two functions may be close together while the values of their derivatives are far apart (if, for example, one function varies slowly while the other oscillates rapidly, as in Example 9.5). Thus, we have to impose strong conditions on a sequence of functions and their derivatives if we hope to prove that $f_n \to f$ implies $f'_n \to f'$.

The following example shows that the limit of the derivatives need not equal the derivative of the limit even if a sequence of differentiable functions converges uniformly and their derivatives converge pointwise.

Example 9.17. Consider the sequence (f_n) of functions $f_n : \mathbb{R} \to \mathbb{R}$ defined by

$$f_n(x) = \frac{x}{1 + nx^2}.$$

Then $f_n \to 0$ uniformly on $\mathbb{R}$. To see this, we write

$$|f_n(x)| = \frac{1}{\sqrt{n}} \left(\frac{\sqrt{n}|x|}{1 + nx^2} \right) = \frac{1}{\sqrt{n}} \left(\frac{t}{1 + t^2} \right)$$

where $t = \sqrt{n}|x|$. We have

$$\frac{t}{1 + t^2} \le \frac{1}{2} \qquad \text{for all } t \in \mathbb{R},$$

since $(1 - t)^2 \ge 0$, which implies that $2t \le 1 + t^2$. Using this inequality, we get

$$|f_n(x)| \le \frac{1}{2\sqrt{n}} \qquad \text{for all } x \in \mathbb{R}.$$

Hence, given $\epsilon > 0$, choose $N = 1/(4\epsilon^2)$. Then

$$|f_n(x)| < \epsilon \qquad \text{for all } x \in \mathbb{R} \text{ if } n > N,$$

which proves that (f_n) converges uniformly to 0 on $\mathbb{R}$. (Alternatively, we could get the same result by using calculus to compute the maximum value of $|f_n|$ on $\mathbb{R}$.)

Each f_n is differentiable with

$$f'_n(x) = \frac{1 - nx^2}{(1 + nx^2)^2}.$$

It follows that $f_n' \to g$ pointwise as $n \to \infty$ where

$$g(x) = \begin{cases} 0 & \text{if } x \neq 0, \\ 1 & \text{if } x = 0. \end{cases}$$

The convergence is not uniform since g is discontinuous at 0. Thus, $f_n \to 0$ uniformly, but $f_n'(0) \to 1$, so the limit of the derivatives is not the derivative of the limit.

However, we do get a useful result if we strengthen the assumptions and require that the derivatives converge uniformly, not just pointwise. The proof involves a slightly tricky application of the mean value theorem.

Theorem 9.18. Suppose that (f_n) is a sequence of differentiable functions $f_n : (a, b) \to \mathbb{R}$ such that $f_n \to f$ pointwise and $f_n' \to g$ uniformly for some $f, g : (a, b) \to \mathbb{R}$. Then f is differentiable on (a, b) and $f' = g$.

Proof. Let $c \in (a, b)$, and let $\epsilon > 0$ be given. To prove that $f'(c) = g(c)$, we estimate the difference quotient of f in terms of the difference quotients of the f_n:

$$\left| \frac{f(x) - f(c)}{x - c} - g(c) \right| \leq \left| \frac{f(x) - f(c)}{x - c} - \frac{f_n(x) - f_n(c)}{x - c} \right|$$
$$+ \left| \frac{f_n(x) - f_n(c)}{x - c} - f_n'(c) \right| + |f_n'(c) - g(c)|$$

where $x \in (a, b)$ and $x \neq c$. We want to make each of the terms on the right-hand side of the inequality less than $\epsilon/3$. This is straightforward for the second term (since f_n is differentiable) and the third term (since $f_n' \to g$). To estimate the first term, we approximate f by f_m, use the mean value theorem, and let $m \to \infty$.

Since $f_m - f_n$ is differentiable, the mean value theorem implies that there exists ξ between c and x such that

$$\frac{f_m(x) - f_m(c)}{x - c} - \frac{f_n(x) - f_n(c)}{x - c} = \frac{(f_m - f_n)(x) - (f_m - f_n)(c)}{x - c}$$
$$= f_m'(\xi) - f_n'(\xi).$$

Since (f_n') converges uniformly, it is uniformly Cauchy by Theorem 9.13. Therefore there exists $N_1 \in \mathbb{N}$ such that

$$|f_m'(\xi) - f_n'(\xi)| < \frac{\epsilon}{3} \qquad \text{for all } \xi \in (a, b) \text{ if } m, n > N_1,$$

which implies that

$$\left| \frac{f_m(x) - f_m(c)}{x - c} - \frac{f_n(x) - f_n(c)}{x - c} \right| < \frac{\epsilon}{3}.$$

Taking the limit of this equation as $m \to \infty$, and using the pointwise convergence of (f_m) to f, we get that

$$\left| \frac{f(x) - f(c)}{x - c} - \frac{f_n(x) - f_n(c)}{x - c} \right| \leq \frac{\epsilon}{3} \qquad \text{for } n > N_1.$$

Next, since (f_n') converges to g, there exists $N_2 \in \mathbb{N}$ such that

$$|f_n'(c) - g(c)| < \frac{\epsilon}{3} \qquad \text{for all } n > N_2.$$

Choose some $n > \max(N_1, N_2)$. Then the differentiability of f_n implies that there exists $\delta > 0$ such that

$$\left| \frac{f_n(x) - f_n(c)}{x - c} - f'_n(c) \right| < \frac{\epsilon}{3} \qquad \text{if } 0 < |x - c| < \delta.$$

Putting these inequalities together, we get that

$$\left| \frac{f(x) - f(c)}{x - c} - g(c) \right| < \epsilon \qquad \text{if } 0 < |x - c| < \delta,$$

which proves that f is differentiable at c with $f'(c) = g(c)$. $\qquad\qquad\square$

Like Theorem 9.16, Theorem 9.18 can be interpreted as giving sufficient conditions for an exchange in the order of limits:

$$\lim_{n \to \infty} \lim_{x \to c} \left[\frac{f_n(x) - f_n(c)}{x - c} \right] = \lim_{x \to c} \lim_{n \to \infty} \left[\frac{f_n(x) - f_n(c)}{x - c} \right].$$

It is worth noting that in Theorem 9.18 the derivatives f'_n are not assumed to be continuous. If they are continuous, then one can use Riemann integration and the fundamental theorem of calculus to give a simpler proof (see Theorem 12.21).

9.5. Series

The convergence of a series is defined in terms of the convergence of its sequence of partial sums, and any result about sequences is easily translated into a corresponding result about series.

Definition 9.19. Suppose that (f_n) is a sequence of functions $f_n : A \to \mathbb{R}$. Let (S_n) be the sequence of partial sums $S_n : A \to \mathbb{R}$, defined by

$$S_n(x) = \sum_{k=1}^{n} f_k(x).$$

Then the series

$$S(x) = \sum_{n=1}^{\infty} f_n(x)$$

converges pointwise to $S : A \to \mathbb{R}$ on A if $S_n \to S$ as $n \to \infty$ pointwise on A, and uniformly to S on A if $S_n \to S$ uniformly on A.

We illustrate the definition with a series whose partial sums we can compute explicitly.

Example 9.20. The geometric series

$$\sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \dots$$

has partial sums

$$S_n(x) = \sum_{k=0}^{n} x^k = \frac{1 - x^{n+1}}{1 - x}.$$

Thus, $S_n(x) \to 1/(1-x)$ as $n \to \infty$ if $|x| < 1$ and diverges if $|x| \geq 1$, meaning that

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \qquad \text{pointwise on } (-1, 1).$$

Since $1/(1-x)$ is unbounded on $(-1, 1)$, Theorem 9.14 implies that the convergence cannot be uniform.

The series does, however, converge uniformly on $[-\rho, \rho]$ for every $0 \leq \rho < 1$. To prove this, we estimate for $|x| \leq \rho$ that

$$\left| S_n(x) - \frac{1}{1-x} \right| = \frac{|x|^{n+1}}{1-x} \leq \frac{\rho^{n+1}}{1-\rho}.$$

Since $\rho^{n+1}/(1-\rho) \to 0$ as $n \to \infty$, given any $\epsilon > 0$ there exists $N \in \mathbb{N}$, depending only on ϵ and ρ, such that

$$0 \leq \frac{\rho^{n+1}}{1-\rho} < \epsilon \qquad \text{for all } n > N.$$

It follows that

$$\left| \sum_{k=0}^{n} x^k - \frac{1}{1-x} \right| < \epsilon \qquad \text{for all } x \in [-\rho, \rho] \text{ and all } n > N,$$

which proves that the series converges uniformly on $[-\rho, \rho]$.

The Cauchy condition for the uniform convergence of sequences immediately gives a corresponding Cauchy condition for the uniform convergence of series.

Theorem 9.21. Let (f_n) be a sequence of functions $f_n : A \to \mathbb{R}$. The series

$$\sum_{n=1}^{\infty} f_n$$

converges uniformly on A if and only if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$\left| \sum_{k=m+1}^{n} f_k(x) \right| < \epsilon \qquad \text{for all } x \in A \text{ and all } n > m > N.$$

Proof. Let

$$S_n(x) = \sum_{k=1}^{n} f_k(x) = f_1(x) + f_2(x) + \cdots + f_n(x).$$

From Theorem 9.13 the sequence (S_n), and therefore the series $\sum f_n$, converges uniformly if and only if for every $\epsilon > 0$ there exists N such that

$$|S_n(x) - S_m(x)| < \epsilon \qquad \text{for all } x \in A \text{ and all } n, m > N.$$

Assuming $n > m$ without loss of generality, we have

$$S_n(x) - S_m(x) = f_{m+1}(x) + f_{m+2}(x) + \cdots + f_n(x) = \sum_{k=m+1}^{n} f_k(x),$$

so the result follows. $\qquad\qquad \square$

The following simple criterion for the uniform convergence of a series is very useful. The name comes from the letter traditionally used to denote the constants, or "majorants," that bound the functions in the series.

Theorem 9.22 (Weierstrass M-test)**.** Let (f_n) be a sequence of functions $f_n : A \to \mathbb{R}$, and suppose that for every $n \in \mathbb{N}$ there exists a constant $M_n \geq 0$ such that

$$|f_n(x)| \leq M_n \quad \text{for all } x \in A, \qquad \sum_{n=1}^{\infty} M_n < \infty.$$

Then

$$\sum_{n=1}^{\infty} f_n(x).$$

converges uniformly on A.

Proof. The result follows immediately from the observation that $\sum f_n$ is uniformly Cauchy if $\sum M_n$ is Cauchy.

In detail, let $\epsilon > 0$ be given. The Cauchy condition for the convergence of a real series implies that there exists $N \in \mathbb{N}$ such that

$$\sum_{k=m+1}^{n} M_k < \epsilon \qquad \text{for all } n > m > N.$$

Then for all $x \in A$ and all $n > m > N$, we have

$$\left| \sum_{k=m+1}^{n} f_k(x) \right| \leq \sum_{k=m+1}^{n} |f_k(x)|$$
$$\leq \sum_{k=m+1}^{n} M_k$$
$$< \epsilon.$$

Thus, $\sum f_n$ satisfies the uniform Cauchy condition in Theorem 9.21, so it converges uniformly. $\square$

Example 9.23. Returning to Example 9.20, we consider the geometric series

$$\sum_{n=0}^{\infty} x^n.$$

If $|x| \leq \rho$ where $0 \leq \rho < 1$, then

$$|x^n| \leq \rho^n, \qquad \sum_{n=0}^{\infty} \rho^n < 1.$$

The M-test, with $M_n = \rho^n$, implies that the series converges uniformly on $[-\rho, \rho]$.

Example 9.24. The series

$$f(x) = \sum_{n=1}^{\infty} \frac{1}{2^n} \cos\left(3^n x\right)$$
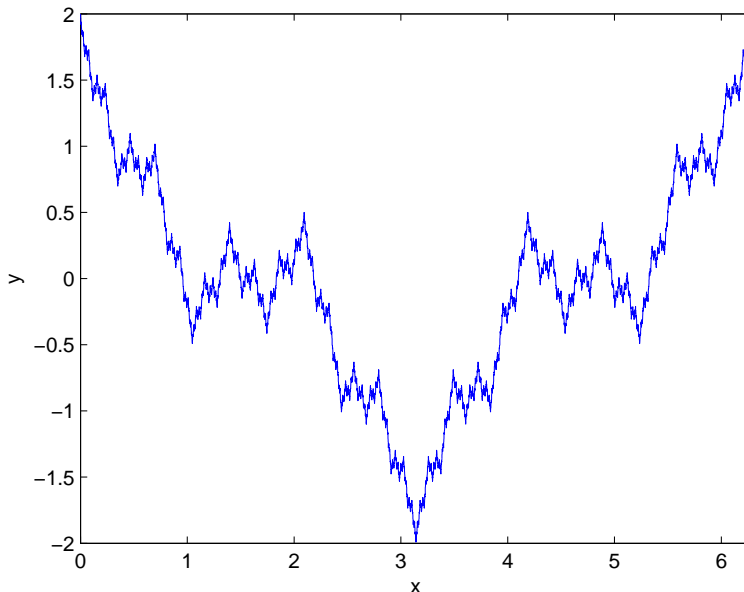
Figure 1. Graph of the Weierstrass continuous, nowhere differentiable function $y = \sum_{n=0}^{\infty} 2^{-n} \cos(3^n x)$ on one period $[0, 2\pi]$.

converges uniformly on $\mathbb{R}$ by the M-test since

$$\left| \frac{1}{2^n} \cos\left(3^n x\right) \right| \le \frac{1}{2^n}, \qquad \sum_{n=1}^{\infty} \frac{1}{2^n} = 1.$$

It then follows from Theorem 9.16 that f is continuous on $\mathbb{R}$. (See Figure 1.)

Taking the formal term-by-term derivative of the series for f, we get a series whose coefficients grow with n,

$$-\sum_{n=1}^{\infty} \left(\frac{3}{2}\right)^n \sin\left(3^n x\right),$$

so we might expect that there are difficulties in differentiating f. As Figure 2 illustrates, the function doesn't look smooth at any length-scale. Weierstrass (1872) proved that f is not differentiable at any point of $\mathbb{R}$. Bolzano (1830) had also constructed a continuous, nowhere differentiable function, but his results weren't published until 1922. Subsequently, Tagaki (1903) constructed a similar function to the Weierstrass function whose nowhere-differentiability is easier to prove. Such functions were considered to be highly counter-intuitive and pathological at the time Weierstrass discovered them, and they weren't well-received by many prominent mathematicians.
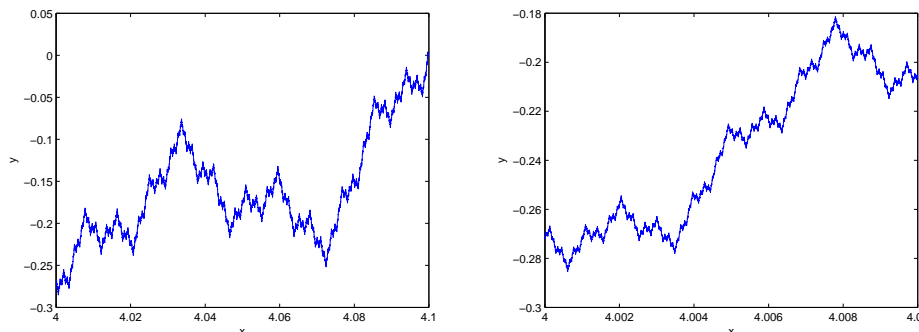
Figure 2. Details of the Weierstrass function for $4 \leq x \leq 4.1$ (left) and $4 \leq x \leq 4.01$ (right) showing its self-similar, fractal behavior under rescalings.

If the Weierstrass M-test applies to a series of functions to prove uniform convergence, then it also implies that the series converges absolutely, meaning that

$$\sum_{n=1}^{\infty} |f_n(x)| < \infty \qquad \text{for every } x \in A.$$

Thus, the M-test is not applicable to series that converge uniformly but not absolutely.

Absolute convergence of a series is completely different from uniform convergence, and the two concepts shouldn't be confused. Absolute convergence on A is a pointwise condition for each $x \in A$, while uniform convergence is a global condition that involves all points $x \in A$ simultaneously. We illustrate the difference with a rather trivial example.

Example 9.25. Let $f_n : \mathbb{R} \to \mathbb{R}$ be the constant function

$$f_n(x) = \frac{(-1)^{n+1}}{n}.$$

Then $\sum f_n$ converges on $\mathbb{R}$ to the constant function $f(x) = c$, where

$$c = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$$

is the sum of the alternating harmonic series ($c = \log 2$). The convergence of $\sum f_n$ is uniform on $\mathbb{R}$ since the terms in the series do not depend on x, but the convergence isn't absolute at any $x \in \mathbb{R}$ since the harmonic series

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

diverges to infinity.

Power Series

In discussing power series it is good to recall a nursery rhyme:
> "There was a little girl
> Who had a little curl
> Right in the middle of her forehead
> When she was good
> She was very, very good
> But when she was bad
> She was horrid."

(Robert Strichartz [**14**])

Power series are one of the most useful type of series in analysis. For example, we can use them to define transcendental functions such as the exponential and trigonometric functions (as well as many other less familiar functions).

10.1. Introduction

A power series (centered at 0) is a series of the form

$$\sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n + \ldots.$$

where the constants a_n are some coefficients. If all but finitely many of the a_n are zero, then the power series is a polynomial function, but if infinitely many of the a_n are nonzero, then we need to consider the convergence of the power series.

The basic facts are these: Every power series has a radius of convergence $0 \leq R \leq \infty$, which depends on the coefficients a_n. The power series converges absolutely in $|x| < R$ and diverges in $|x| > R$. Moreover, the convergence is uniform on every interval $|x| < \rho$ where $0 \leq \rho < R$. If $R > 0$, then the sum of the power series is infinitely differentiable in $|x| < R$, and its derivatives are given by differentiating the original power series term-by-term.

Power series work just as well for complex numbers as real numbers, and are in fact best viewed from that perspective. We will consider here only real-valued power series, although many of the results extend immediately to complex-valued power series.

Definition 10.1. Let $(a_n)_{n=0}^{\infty}$ be a sequence of real numbers and $c \in \mathbb{R}$. The power series centered at c with coefficients a_n is the series

$$\sum_{n=0}^{\infty} a_n(x-c)^n.$$

Example 10.2. The following are power series centered at 0:

$$\sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + x^4 + \dots,$$

$$\sum_{n=0}^{\infty} \frac{1}{n!}x^n = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{24}x^4 + \dots,$$

$$\sum_{n=0}^{\infty} (n!)x^n = 1 + x + 2x^2 + 6x^3 + 24x^4 + \dots,$$

$$\sum_{n=0}^{\infty} (-1)^n x^{2^n} = x - x^2 + x^4 - x^8 + \dots.$$

An example of a power series centered at 1 is

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}(x-1)^n = (x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3 - \frac{1}{4}(x-1)^4 + \dots.$$

The power series in Definition 10.1 is a formal, algebraic expression, since we haven't said anything yet about its convergence. By changing variables $(x-c) \mapsto x$, we can assume without loss of generality that a power series is centered at 0, and we will do so whenever it's convenient.

10.2. Radius of convergence

First, we prove that every power series has a radius of convergence.

Theorem 10.3. Let

$$\sum_{n=0}^{\infty} a_n(x-c)^n$$

be a power series. There is a non-negative, extended real number $0 \le R \le \infty$ such that the series converges absolutely for $0 \le |x-c| < R$ and diverges for $|x-c| > R$. Furthermore, if $0 \le \rho < R$, then the power series converges uniformly on the interval $|x-c| \le \rho$, and the sum of the series is continuous in $|x-c| < R$.

Proof. We assume without loss of generality that $c = 0$. Suppose the power series

$$\sum_{n=0}^{\infty} a_n x_0^n$$

converges for some $x_0 \in \mathbb{R}$ with $x_0 \neq 0$. Then its terms converge to zero, so they are bounded and there exists $M \geq 0$ such that

$$|a_n x_0^n| \leq M \qquad \text{for } n = 0, 1, 2, \ldots.$$

If $|x| < |x_0|$, then

$$|a_n x^n| = |a_n x_0^n| \left| \frac{x}{x_0} \right|^n \leq M r^n, \qquad r = \left| \frac{x}{x_0} \right| < 1.$$

Comparing the power series with the convergent geometric series $\sum M r^n$, we see that $\sum a_n x^n$ is absolutely convergent. Thus, if the power series converges for some $x_0 \in \mathbb{R}$, then it converges absolutely for every $x \in \mathbb{R}$ with $|x| < |x_0|$.

Let

$$R = \sup \left\{ |x| \geq 0 : \sum a_n x^n \text{ converges} \right\}.$$

If $R = 0$, then the series converges only for $x = 0$. If $R > 0$, then the series converges absolutely for every $x \in \mathbb{R}$ with $|x| < R$, since it converges for some $x_0 \in \mathbb{R}$ with $|x| < |x_0| < R$. Moreover, the definition of R implies that the series diverges for every $x \in \mathbb{R}$ with $|x| > R$. If $R = \infty$, then the series converges for all $x \in \mathbb{R}$.

Finally, let $0 \leq \rho < R$ and suppose $|x| \leq \rho$. Choose $\sigma > 0$ such that $\rho < \sigma < R$. Then $\sum |a_n \sigma^n|$ converges, so $|a_n \sigma^n| \leq M$, and therefore

$$|a_n x^n| = |a_n \sigma^n| \left| \frac{x}{\sigma} \right|^n \leq |a_n \sigma^n| \left| \frac{\rho}{\sigma} \right|^n \leq M r^n,$$

where $r = \rho/\sigma < 1$. Since $\sum M r^n < \infty$, the M-test (Theorem 9.22) implies that the series converges uniformly on $|x| \leq \rho$, and then it follows from Theorem 9.16 that the sum is continuous on $|x| \leq \rho$. Since this holds for every $0 \leq \rho < R$, the sum is continuous in $|x| < R$. $\qquad \square$

The following definition therefore makes sense for every power series.

Definition 10.4. If the power series

$$\sum_{n=0}^{\infty} a_n (x - c)^n$$

converges for $|x - c| < R$ and diverges for $|x - c| > R$, then $0 \leq R \leq \infty$ is called the radius of convergence of the power series.

Theorem 10.3 does not say what happens at the endpoints $x = c \pm R$, and in general the power series may converge or diverge there. We refer to the set of all points where the power series converges as its interval of convergence, which is one of

$$(c - R, c + R), \quad (c - R, c + R], \quad [c - R, c + R), \quad [c - R, c + R].$$

We won't discuss here any general theorems about the convergence of power series at the endpoints (e.g., the Abel theorem). Also note that a power series need not converge uniformly on $|x - c| < R$.

Theorem 10.3 does not give an explicit expression for the radius of convergence of a power series in terms of its coefficients. The ratio test gives a simple, but useful,

way to compute the radius of convergence, although it doesn't apply to every power series.

Theorem 10.5. Suppose that $a_n \neq 0$ for all sufficiently large n and the limit

$$R = \lim_{n \to \infty} \left| \frac{a_n}{a_{n+1}} \right|$$

exists or diverges to infinity. Then the power series

$$\sum_{n=0}^{\infty} a_n (x - c)^n$$

has radius of convergence R.

Proof. Let

$$r = \lim_{n \to \infty} \left| \frac{a_{n+1}(x - c)^{n+1}}{a_n (x - c)^n} \right| = |x - c| \lim_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right|.$$

By the ratio test, the power series converges if $0 \leq r < 1$, or $|x - c| < R$, and diverges if $1 < r \leq \infty$, or $|x - c| > R$, which proves the result. □

The root test gives an expression for the radius of convergence of a general power series.

Theorem 10.6 (Hadamard)**.** The radius of convergence R of the power series

$$\sum_{n=0}^{\infty} a_n (x - c)^n$$

is given by

$$R = \frac{1}{\limsup_{n \to \infty} |a_n|^{1/n}}$$

where $R = 0$ if the $\limsup$ diverges to ∞, and $R = \infty$ if the $\limsup$ is 0.

Proof. Let

$$r = \limsup_{n \to \infty} |a_n(x - c)^n|^{1/n} = |x - c| \limsup_{n \to \infty} |a_n|^{1/n}.$$

By the root test, the series converges if $0 \leq r < 1$, or $|x - c| < R$, and diverges if $1 < r \leq \infty$, or $|x - c| > R$, which proves the result. □

This theorem provides an alternate proof of Theorem 10.3 from the root test; in fact, our proof of Theorem 10.3 is more-or-less a proof of the root test.

10.3. Examples of power series

We consider a number of examples of power series and their radii of convergence.

Example 10.7. The geometric series

$$\sum_{n=0}^{\infty} x^n = 1 + x + x^2 + \dots$$

has radius of convergence

$$R = \lim_{n \to \infty} \frac{1}{1} = 1.$$

It converges to

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n \qquad \text{for } |x| < 1,$$

and diverges for $|x| > 1$. At $x = 1$, the series becomes

$$1 + 1 + 1 + 1 + \dots$$

and at $x = -1$ it becomes

$$1 - 1 + 1 - 1 + 1 - \dots,$$

so the series diverges at both endpoints $x = \pm 1$. Thus, the interval of convergence of the power series is $(-1, 1)$. The series converges uniformly on $[-\rho, \rho]$ for every $0 \le \rho < 1$ but does not converge uniformly on $(-1, 1)$ (see Example 9.20). Note that although the function $1/(1-x)$ is well-defined for all $x \ne 1$, the power series only converges when $|x| < 1$.

Example 10.8. The series

$$\sum_{n=1}^{\infty} \frac{1}{n} x^n = x + \frac{1}{2} x^2 + \frac{1}{3} x^3 + \frac{1}{4} x^4 + \dots$$

has radius of convergence

$$R = \lim_{n \to \infty} \frac{1/n}{1/(n+1)} = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right) = 1.$$

At $x = 1$, the series becomes the harmonic series

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots,$$

which diverges, and at $x = -1$ it is minus the alternating harmonic series

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n} = -1 + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \dots,$$

which converges, but not absolutely. Thus the interval of convergence of the power series is $[-1, 1)$. The series converges uniformly on $[-\rho, \rho]$ for every $0 \le \rho < 1$ but does not converge uniformly on $(-1, 1)$.

Example 10.9. The power series

$$\sum_{n=0}^{\infty} \frac{1}{n!} x^n = 1 + x + \frac{1}{2!} x + \frac{1}{3!} x^3 + \dots$$

has radius of convergence

$$R = \lim_{n \to \infty} \frac{1/n!}{1/(n+1)!} = \lim_{n \to \infty} \frac{(n+1)!}{n!} = \lim_{n \to \infty} (n+1) = \infty,$$

so it converges for all $x \in \mathbb{R}$. The sum is the exponential function

$$e^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n.$$

In fact, this power series may be used to define the exponential function. (See Section 10.6.)

Example 10.10. The power series

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} = 1 - \frac{1}{2!} x^2 + \frac{1}{4!} x^4 + \dots$$

has radius of convergence $R = \infty$, and it converges for all $x \in \mathbb{R}$. Its sum $\cos x$ provides an analytic definition of the cosine function.

Example 10.11. The power series

$$\sum_{n=0\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} = x - \frac{1}{3!} x^3 + \frac{1}{5!} x^5 + \dots$$

has radius of convergence $R = \infty$, and it converges for all $x \in \mathbb{R}$. Its sum $\sin x$ provides an analytic definition of the sine function.

Example 10.12. The power series

$$\sum_{n=0}^{\infty} (n!) x^n = 1 + x + (2!)x + (3!)x^3 + (4!)x^4 + \dots$$

has radius of convergence

$$R = \lim_{n \to \infty} \frac{n!}{(n+1)!} = \lim_{n \to \infty} \frac{1}{n+1} = 0,$$

so it converges only for $x = 0$. If $x \neq 0$, its terms grow larger once $n > 1/|x|$ and $|(n!)x^n| \to \infty$ as $n \to \infty$.

Example 10.13. The series

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} (x-1)^n = (x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3 - \dots$$

has radius of convergence

$$R = \lim_{n \to \infty} \left| \frac{(-1)^{n+1}/n}{(-1)^{n+2}/(n+1)} \right| = \lim_{n \to \infty} \frac{n}{n+1} = \lim_{n \to \infty} \frac{1}{1+1/n} = 1,$$

so it converges if $|x-1| < 1$ and diverges if $|x-1| > 1$. At the endpoint $x = 2$, the power series becomes the alternating harmonic series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots,$$

which converges. At the endpoint $x = 0$, the power series becomes the harmonic series

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots,$$

which diverges. Thus, the interval of convergence is $(0, 2]$.
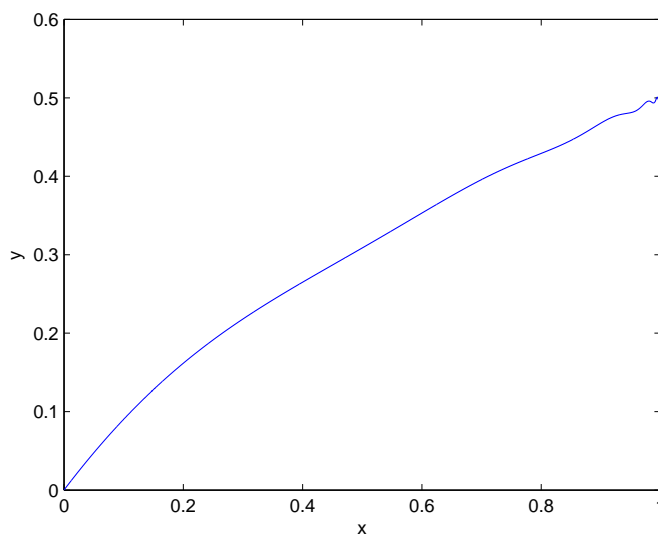
Figure 1. Graph of the lacunary power series $y = \sum_{n=0}^{\infty} (-1)^n x^{2^n}$ on $[0, 1)$. It appears relatively well-behaved; however, the small oscillations visible near $x = 1$ are not a numerical artifact.

Example 10.14. The power series

$$\sum_{n=0}^{\infty} (-1)^n x^{2^n} = x - x^2 + x^4 - x^8 + x^{16} - x^{32} + \dots$$

with

$$a_n = \begin{cases} (-1)^k & \text{if } n = 2^k, \\ 0 & \text{if } n \neq 2^k, \end{cases}$$

has radius of convergence $R = 1$. To prove this, note that the series converges for $|x| < 1$ by comparison with the convergent geometric series $\sum |x|^n$, since

$$|a_n x^n| = \begin{cases} |x|^n & \text{if } n = 2^k \\ 0 & \text{if } n \neq 2^k \end{cases} \leq |x|^n.$$

If $|x| > 1$, then the terms do not approach 0 as $n \to \infty$, so the series diverges. Alternatively, we have

$$|a_n|^{1/n} = \begin{cases} 1 & \text{if } n = 2^k, \\ 0 & \text{if } n \neq 2^k, \end{cases}$$

so

$$\limsup_{n \to \infty} |a_n|^{1/n} = 1$$

and the Hadamard formula (Theorem 10.6) gives $R = 1$. The series does not converge at either endpoint $x = \pm 1$, so its interval of convergence is $(-1, 1)$.

In this series, there are successively longer gaps (or "lacuna") between the powers with non-zero coefficients. Such series are called lacunary power series, and
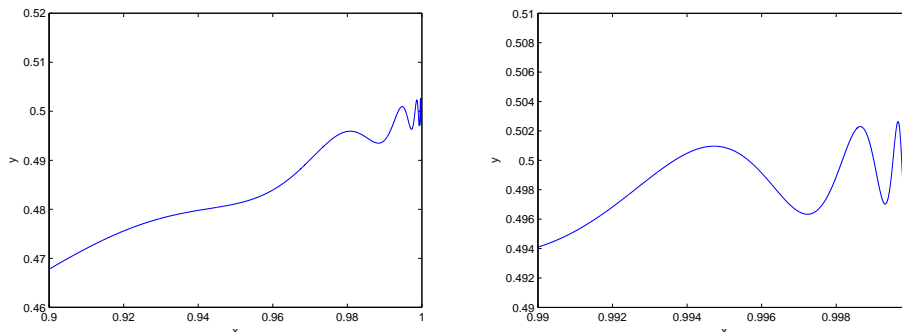
Figure 2. Details of the lacunary power series $\sum_{n=0}^{\infty}(-1)^n x^{2^n}$ near $x = 1$, showing its oscillatory behavior and the nonexistence of a limit as $x \to 1^-$.

they have many interesting properties. For example, although the series does not converge at $x = 1$, one can ask if

$$\lim_{x \to 1^-} \left[\sum_{n=0}^{\infty} (-1)^n x^{2^n} \right]$$

exists. In a plot of this sum on $[0, 1)$, shown in Figure 1, the function appears relatively well-behaved near $x = 1$. However, Hardy (1907) proved that the function has infinitely many, very small oscillations as $x \to 1^-$, as illustrated in Figure 2, and the limit does not exist. Subsequent results by Hardy and Littlewood (1926) showed, under suitable assumptions on the growth of the "gaps" between non-zero coefficients, that if the limit of a lacunary power series as $x \to 1^-$ exists, then the series must converge at $x = 1$. Since the lacunary power series considered here does not converge at 1, its limit as $x \to 1^-$ cannot exist. For further discussion of lacunary power series, see [**4**].

10.4. Algebraic operations on power series

We can add, multiply, and divide power series in a standard way. For simplicity, we consider power series centered at 0.

Proposition 10.15. If $R, S > 0$ and the functions

$$f(x) = \sum_{n=0}^{\infty} a_n x^n \quad \text{in } |x| < R, \qquad g(x) = \sum_{n=0}^{\infty} b_n x^n \qquad \text{in } |x| < S$$

are sums of convergent power series, then

$$(f + g)(x) = \sum_{n=0}^{\infty} (a_n + b_n) x^n \qquad \text{in } |x| < T,$$

$$(fg)(x) = \sum_{n=0}^{\infty} c_n x^n \qquad\qquad \text{in } |x| < T,$$

where $T = \min(R, S)$ and

$$c_n = \sum_{k=0}^{n} a_{n-k} b_k.$$

Proof. The power series expansion of $f + g$ follows immediately from the linearity of limits. The power series expansion of fg follows from the Cauchy product (Theorem 4.38), since power series converge absolutely inside their intervals of convergence, and

$$\left(\sum_{n=0}^{\infty} a_n x^n \right) \left(\sum_{n=0}^{\infty} b_n x^n \right) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} a_{n-k} x^{n-k} \cdot b_k x^k \right) = \sum_{n=0}^{\infty} c_n x^n.$$

$\square$

It may happen that the radius of convergence of the power series for $f + g$ or fg is larger than the radius of convergence of the power series for f, g. For example, if $g = -f$, then the radius of convergence of the power series for $f + g = 0$ is ∞ whatever the radius of convergence of the power series for f.

The reciprocal of a convergent power series that is nonzero at its center also has a power series expansion.

Proposition 10.16. If $R > 0$ and

$$f(x) = \sum_{n=0}^{\infty} a_n x^n \qquad \text{in } |x| < R,$$

is the sum of a power series with $a_0 \neq 0$, then there exists $S > 0$ such that

$$\frac{1}{f(x)} = \sum_{n=0}^{\infty} b_n x^n \qquad \text{in } |x| < S.$$

The coefficients b_n are determined recursively by

$$b_0 = \frac{1}{a_0}, \qquad b_n = -\frac{1}{a_0} \sum_{k=0}^{n-1} a_{n-k} b_k, \qquad \text{for } n \geq 1.$$

Proof. First, we look for a formal power series expansion (i.e., without regard to its convergence)

$$g(x) = \sum_{n=0}^{\infty} b_n x^n$$

such that the formal Cauchy product fg is equal to 1. This condition is satisfied if

$$\left(\sum_{n=0}^{\infty} a_n x^n \right) \left(\sum_{n=0}^{\infty} b_n x^n \right) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} a_{n-k} b_k \right) x^n = 1.$$

Matching the coefficients of x^n, we find that

$$a_0 b_0 = 1, \qquad a_0 b_n + \sum_{k=0}^{n-1} a_{n-k} b_k = 0 \qquad \text{for } n \geq 1,$$

which gives the stated recursion relation.

To complete the proof, we need to show that the formal power series for g has a nonzero radius of convergence. In that case, Proposition 10.15 shows that $fg = 1$ inside the common interval of convergence of f and g, so $1/f = g$ has a power series expansion. We assume without loss of generality that $a_0 = 1$; otherwise replace f by f/a_0.

The power series for f converges absolutely and uniformly on compact sets inside its interval of convergence, so the function

$$\sum_{n=1}^{\infty} |a_n| \, |x|^n$$

is continuous in $|x| < R$ and vanishes at $x = 0$. It follows that there exists $\delta > 0$ such that

$$\sum_{n=1}^{\infty} |a_n| \, |x|^n \leq 1 \qquad \text{for } |x| \leq \delta.$$

Then $f(x) \neq 0$ for $|x| < \delta$, since

$$|f(x)| \geq 1 - \sum_{n=1}^{\infty} |a_n| \, |x|^n > 0,$$

so $1/f(x)$ is well defined.

We claim that

$$|b_n| \leq \frac{1}{\delta^n} \qquad \text{for } n = 0, 1, 2, \ldots.$$

The proof is by induction. Since $b_0 = 1$, this inequality is true for $n = 0$. If $n \geq 1$ and the inequality holds for b_k with $0 \leq k \leq n - 1$, then by taking the absolute value of the recursion relation for b_n, we get

$$|b_n| \leq \sum_{k=1}^{n} |a_k| |b_{n-k}| \leq \sum_{k=1}^{n} \frac{|a_k|}{\delta^{n-k}} \leq \frac{1}{\delta^n} \sum_{k=1}^{\infty} |a_k| \delta^k \leq \frac{1}{\delta^n},$$

so the inequality holds for b_k with $0 \leq k \leq n$, and the claim follows.

We then get that

$$\limsup_{n \to \infty} |b_n|^{1/n} \leq \frac{1}{\delta},$$

so the Hadamard formula in Theorem 10.6 implies that the radius of convergence of $\sum b_n x^n$ is greater than or equal to $\delta > 0$, which completes the proof. $\qquad \square$

An immediate consequence of these results for products and reciprocals of power series is that quotients of convergent power series are given by convergent power series, provided that the denominator is nonzero.

Proposition 10.17. If $R, S > 0$ and

$$f(x) = \sum_{n=0}^{\infty} a_n x^n \quad \text{in } |x| < R, \qquad g(x) = \sum_{n=0}^{\infty} b_n x^n \quad \text{in } |x| < S$$

are the sums of power series with $b_0 \neq 0$, then there exists $T > 0$ and coefficients c_n such that

$$\frac{f(x)}{g(x)} = \sum_{n=0}^{\infty} c_n x^n \qquad \text{in } |x| < T.$$

The previous results do not give an explicit expression for the coefficients in the power series expansion of f/g or a sharp estimate for its radius of convergence. Using complex analysis, one can show that radius of convergence of the power series for f/g centered at 0 is equal to the distance from the origin of the nearest singularity of f/g in the complex plane. We will not discuss complex analysis here, but we consider two examples.

Example 10.18. Replacing x by $-x^2$ in the geometric power series from Example 10.7, we get the following power series centered at 0

$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + \cdots = \sum_{n=0}^{\infty} (-1)^{n+1} x^{2n},$$

which has radius of convergence $R = 1$. From the point of view of real functions, it may appear strange that the radius of convergence is 1, since the function $1/(1+x^2)$ is well-defined on $\mathbb{R}$, has continuous derivatives of all orders, and has power series expansions with nonzero radius of convergence centered at every $c \in \mathbb{R}$. However, when $1/(1+z^2)$ is regarded as a function of a complex variable $z \in \mathbb{C}$, one sees that it has singularities at $z = \pm i$, where the denominator vanishes, and $|\pm i| = 1$, which explains why $R = 1$.

Example 10.19. The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(0) = 1$ and

$$f(x) = \frac{e^x - 1}{x}, \qquad \text{for } x \neq 0$$

has the power series expansion

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

with infinite radius of convergence. The reciprocal function $g : \mathbb{R} \to \mathbb{R}$ of f is given by $g(0) = 1$ and

$$g(x) = \frac{x}{e^x - 1}, \qquad \text{for } x \neq 0.$$

Proposition 10.16 implies that

$$g(x) = \sum_{n=0}^{\infty} b_n x^n$$

has a convergent power series expansion at 0, with $b_0 = 1$ and

$$b_n = -\sum_{k=0}^{n-1} \frac{b_k}{(n-k+1)!} \qquad \text{for } n \geq 1.$$

The numbers $B_n = n! b_n$ are called Bernoulli numbers. They may be defined as the coefficients in the power series expansion

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} \frac{B_n}{n!} x^n.$$

The function $x/(e^x - 1)$ is called the generating function of the Bernoulli numbers, where we adopt the convention that $x/(e^x - 1) = 1$ at $x = 0$.

A number of properties of the Bernoulli numbers follow from their generating function. First, we observe that

$$\frac{x}{e^x - 1} + \frac{1}{2}x = \frac{1}{2}x \left(\frac{e^{x/2} + e^{-x/2}}{e^{x/2} - e^{-x/2}} \right)$$

is an even function of x. It follows that

$$B_0 = 1, \qquad B_1 = -\frac{1}{2},$$

and $B_n = 0$ for all odd $n \geq 3$. Thus, the power series expansion of $x/(e^x - 1)$ has the form

$$\frac{x}{e^x - 1} = 1 - \frac{1}{2}x + \sum_{n=1}^{\infty} \frac{B_{2n}}{(2n)!} x^{2n}.$$

The recursion formula for b_n can be written in terms of B_n as

$$\sum_{k=0}^{n} \binom{n+1}{k} B_k = 0,$$

which implies that the Bernoulli numbers are rational. For example, one finds that

$$B_2 = \frac{1}{6}, \qquad B_4 = -\frac{1}{30}, \quad B_6 = \frac{1}{42}, \quad B_8 = -\frac{1}{30}, \quad B_{10} = \frac{5}{66}, \quad B_{12} = -\frac{691}{2730}.$$

As the sudden appearance of the large irregular prime number 691 in the numerator of B_{12} suggests, there is no simple pattern for the values of B_{2n}, although they continue to alternate in sign.[1] The Bernoulli numbers have many surprising connections with number theory and other areas of mathematics; for example, as noted in Section 4.5, they give the values of the Riemann zeta function at even natural numbers.

Using complex analysis, one can show that the radius of convergence of the power series for $z/(e^z - 1)$ at $z = 0$ is equal to 2π, since the closest zeros of the denominator $e^z - 1$ to the origin in the complex plane occur at $z = \pm 2\pi i$, where $|z| = 2\pi$. Given this fact, the Hadamard formula (Theorem 10.6) implies that

$$\limsup_{n \to \infty} \left| \frac{B_n}{n!} \right|^{1/n} = \frac{1}{2\pi},$$

which shows that at least some of the Bernoulli numbers B_n grow very rapidly (factorially) as $n \to \infty$.

Finally, we remark that we have proved that algebraic operations on convergent power series lead to convergent power series. If one is interested only in the formal algebraic properties of power series, and not their convergence, one can introduce a purely algebraic structure called the ring of formal power series (over the field $\mathbb{R}$) in a variable x,

$$\mathbb{R}[[x]] = \left\{ \sum_{n=0}^{\infty} a_n x^n : a_n \in \mathbb{R} \right\},$$

---

[1] A prime number p is said to be irregular if it divides the numerator of B_{2n}, expressed in lowest terms, for some $2 \leq 2n \leq p - 3$; otherwise it is regular. The smallest irregular prime number is 37, which divides the numerator of $B_{32} = -7709321041217/5100$, since $7709321041217 = 37 \cdot 683 \cdot 305065927$. There are infinitely many irregular primes, and it is conjectured that there are infinitely many regular primes. A proof of this conjecture is, however, an open problem.

with sums and products on $\mathbb{R}[[x]]$ defined in the obvious way:

$$\sum_{n=0}^{\infty} a_n x^n + \sum_{n=0}^{\infty} b_n x^n = \sum_{n=0}^{\infty} (a_n + b_n) x^n,$$

$$\left(\sum_{n=0}^{\infty} a_n x^n \right) \left(\sum_{n=0}^{\infty} b_n x^n \right) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} a_{n-k} b_k \right) x^n.$$

10.5. Differentiation of power series

We saw in Section 9.4.3 that, in general, one cannot differentiate a uniformly convergent sequence or series. We can, however, differentiate power series, and they behaves as nicely as one can imagine in this respect. The sum of a power series

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + \dots$$

is infinitely differentiable inside its interval of convergence, and its derivative

$$f'(x) = a_1 + 2a_2 x + 3a_3 x^2 + 4a_4 x^3 + \dots$$

is given by term-by-term differentiation. To prove this result, we first show that the term-by-term derivative of a power series has the same radius of convergence as the original power series. The idea is that the geometrical decay of the terms of the power series inside its radius of convergence dominates the algebraic growth of the factor n that comes from taking the derivative.

Theorem 10.20. Suppose that the power series

$$\sum_{n=0}^{\infty} a_n (x - c)^n$$

has radius of convergence R. Then the power series

$$\sum_{n=1}^{\infty} n a_n (x - c)^{n-1}$$

also has radius of convergence R.

Proof. Assume without loss of generality that $c = 0$, and suppose $|x| < R$. Choose ρ such that $|x| < \rho < R$, and let

$$r = \frac{|x|}{\rho}, \qquad 0 < r < 1.$$

To estimate the terms in the differentiated power series by the terms in the original series, we rewrite their absolute values as follows:

$$\left| n a_n x^{n-1} \right| = \frac{n}{\rho} \left(\frac{|x|}{\rho} \right)^{n-1} |a_n \rho^n| = \frac{n r^{n-1}}{\rho} |a_n \rho^n|.$$

The ratio test shows that the series $\sum n r^{n-1}$ converges, since

$$\lim_{n \to \infty} \left[\frac{(n+1) r^n}{n r^{n-1}} \right] = \lim_{n \to \infty} \left[\left(1 + \frac{1}{n} \right) r \right] = r < 1,$$

so the sequence (nr^{n-1}) is bounded, by M say. It follows that

$$\left|na_n x^{n-1}\right| \le \frac{M}{\rho}|a_n\rho^n| \qquad \text{for all } n \in \mathbb{N}.$$

The series $\sum |a_n\rho^n|$ converges, since $\rho < R$, so the comparison test implies that $\sum na_n x^{n-1}$ converges absolutely.

Conversely, suppose $|x| > R$. Then $\sum |a_n x^n|$ diverges (since $\sum a_n x^n$ diverges) and

$$\left|na_n x^{n-1}\right| \ge \frac{1}{|x|}\,|a_n x^n|$$

for $n \ge 1$, so the comparison test implies that $\sum na_n x^{n-1}$ diverges. Thus the series have the same radius of convergence. $\qquad \square$

Theorem 10.21. Suppose that the power series

$$f(x) = \sum_{n=0}^{\infty} a_n(x - c)^n \qquad \text{for } |x - c| < R$$

has radius of convergence $R > 0$ and sum f. Then f is differentiable in $|x - c| < R$ and

$$f'(x) = \sum_{n=1}^{\infty} na_n(x - c)^{n-1} \qquad \text{for } |x - c| < R.$$

Proof. The term-by-term differentiated power series converges in $|x - c| < R$ by Theorem 10.20. We denote its sum by

$$g(x) = \sum_{n=1}^{\infty} na_n(x - c)^{n-1}.$$

Let $0 < \rho < R$. Then, by Theorem 10.3, the power series for f and g both converge uniformly in $|x - c| < \rho$. Applying Theorem 9.18 to their partial sums, we conclude that f is differentiable in $|x - c| < \rho$ and $f' = g$. Since this holds for every $0 \le \rho < R$, it follows that f is differentiable in $|x - c| < R$ and $f' = g$, which proves the result. $\qquad \square$

Repeated application of Theorem 10.21 implies that the sum of a power series is infinitely differentiable inside its interval of convergence and its derivatives are given by term-by-term differentiation of the power series. Furthermore, we can get an expression for the coefficients a_n in terms of the function f; they are simply the Taylor coefficients of f at c.

Theorem 10.22. If the power series

$$f(x) = \sum_{n=0}^{\infty} a_n(x - c)^n$$

has radius of convergence $R > 0$, then f is infinitely differentiable in $|x - c| < R$ and

$$a_n = \frac{f^{(n)}(c)}{n!}.$$

Proof. We assume $c = 0$ without loss of generality. Applying Theorem 10.22 to the power series

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_n x^n + \ldots$$

k times, we find that f has derivatives of every order in $|x| < R$, and

$$f'(x) = a_1 + 2a_2 x + 3a_3 x^2 + \cdots + na_n x^{n-1} + \ldots,$$
$$f''(x) = 2a_2 + (3 \cdot 2)a_3 x + \cdots + n(n-1)a_n x^{n-2} + \ldots,$$
$$f'''(x) = (3 \cdot 2 \cdot 1)a_3 + \cdots + n(n-1)(n-2)a_n x^{n-3} + \ldots,$$
$$\vdots$$
$$f^{(k)}(x) = (k!)a_k + \cdots + \frac{n!}{(n-k)!}x^{n-k} + \ldots,$$

where all of these power series have radius of convergence R. Setting $x = 0$ in these series, we get

$$a_0 = f(0), \quad a_1 = f'(0), \quad \ldots \quad a_k = \frac{f^{(k)}(0)}{k!}, \quad \ldots$$

which proves the result (after replacing 0 by c). □

One consequence of this result is that power series with different coefficients cannot converge to the same sum.

Corollary 10.23. If two power series

$$\sum_{n=0}^{\infty} a_n (x - c)^n, \qquad \sum_{n=0}^{\infty} b_n (x - c)^n$$

have nonzero-radius of convergence and are equal in some neighborhood of 0, then $a_n = b_n$ for every $n = 0, 1, 2, \ldots$.

Proof. If the common sum in $|x - c| < \delta$ is $f(x)$, we have

$$a_n = \frac{f^{(n)}(c)}{n!}, \qquad b_n = \frac{f^{(n)}(c)}{n!},$$

since the derivatives of f at c are determined by the values of f in an arbitrarily small open interval about c, so the coefficients are equal. □

10.6. The exponential function

We showed in Example 10.9 that the power series

$$E(x) = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots + \frac{1}{n!}x^n + \ldots.$$

has radius of convergence ∞. It therefore defines an infinitely differentiable function $E : \mathbb{R} \to \mathbb{R}$.

Term-by-term differentiation of the power series, which is justified by Theorem 10.21, implies that

$$E'(x) = 1 + x + \frac{1}{2!}x^2 + \cdots + \frac{1}{(n-1)!}x^{(n-1)} + \ldots,$$

so $E' = E$. Moreover $E(0) = 1$. As we show below, there is a unique function with these properties, which are shared by the exponential function e^x. Thus, this power series provides an analytical definition of $e^x = E(x)$. All of the other familiar properties of the exponential follow from its power-series definition, and we will prove a few of them here

First, we show that $e^x e^y = e^{x+y}$. For the moment, we continue to write the function e^x as $E(x)$ to emphasise that we use nothing beyond its power series definition.

Proposition 10.24. For every $x, y \in \mathbb{R}$,

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

Proof. We have

$$E(x) = \sum_{j=0}^{\infty} \frac{x^j}{j!}, \qquad E(y) = \sum_{k=0}^{\infty} \frac{y^k}{k!}.$$

Multiplying these series term-by-term and rearranging the sum as a Cauchy product, which is justified by Theorem 4.38, we get

$$E(x)E(y) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{x^j y^k}{j!\, k!}$$

$$= \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{x^{n-k} y^k}{(n-k)!\, k!}.$$

From the binomial theorem,

$$\sum_{k=0}^{n} \frac{x^{n-k} y^k}{(n-k)!\, k!} = \frac{1}{n!} \sum_{k=0}^{n} \frac{n!}{(n-k)!\, k!} x^{n-k} y^k = \frac{1}{n!} \left(x + y \right)^n.$$

Hence,

$$E(x)E(y) = \sum_{n=0}^{\infty} \frac{(x+y)^n}{n!} = E(x + y),$$

which proves the result. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

In particular, since $E(0) = 1$, it follows that

$$E(-x) = \frac{1}{E(x)}.$$

We have $E(x) > 0$ for all $x \geq 0$, since all of the terms in its power series are positive, so $E(x) > 0$ for all $x \in \mathbb{R}$.

Next, we prove that the exponential is characterized by the properties $E' = E$ and $E(0) = 1$. This is a simple uniqueness result for an initial value problem for a linear ordinary differential equation.

Proposition 10.25. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a differentiable function such that

$$f' = f, \qquad f(0) = 1.$$

Then $f = E$.

Proof. Suppose that $f' = f$. Using the equation $E' = E$, the fact that E is nonzero on $\mathbb{R}$, and the quotient rule, we get

$$\left(\frac{f}{E}\right)' = \frac{fE' - Ef'}{E^2} = \frac{fE - Ef}{E^2} = 0.$$

It follows from Theorem 8.34 that f/E is constant on $\mathbb{R}$. Since $f(0) = E(0) = 1$, we have $f/E = 1$, which implies that $f = E$. $\square$

In view of this result, we now write $E(x) = e^x$. The following proposition, which we use below in Section 10.7.2, shows that e^x grows faster than any power of x as $x \to \infty$.

Proposition 10.26. Suppose that n is a non-negative integer. Then

$$\lim_{x \to \infty} \frac{x^n}{e^x} = 0.$$

Proof. The terms in the power series of e^x are positive for $x > 0$, so for every $k \in \mathbb{N}$

$$e^x = \sum_{j=0}^{\infty} \frac{x^j}{j!} > \frac{x^k}{k!} \qquad \text{for all } x > 0.$$

Taking $k = n + 1$, we get for $x > 0$ that

$$0 < \frac{x^n}{e^x} < \frac{x^n}{x^{(n+1)}/(n+1)!} = \frac{(n+1)!}{x}.$$

Since $1/x \to 0$ as $x \to \infty$, the result follows. $\square$

The logarithm $\log : (0, \infty) \to \mathbb{R}$ can be defined as the inverse of the exponential function $\exp : \mathbb{R} \to (0, \infty)$, which is strictly increasing on $\mathbb{R}$ since its derivative is strictly positive. Having the logarithm and the exponential, we can define the power function for all exponents $p \in \mathbb{R}$ by

$$x^p = e^{p \log x}, \qquad x > 0.$$

Other transcendental functions, such as the trigonometric functions, can be defined in terms of their power series, and these can be used to prove their usual properties. We will not carry all this out in detail; we just want to emphasize that, once we have developed the theory of power series, we can define all of the functions arising in elementary calculus from the first principles of analysis.

10.7. * Smooth versus analytic functions

The power series theorem, Theorem 10.22, looks similar to Taylor's theorem, Theorem 8.46, but there is a fundamental difference. Taylor's theorem gives an expression for the error between a function and its Taylor polynomials. No question of convergence is involved. On the other hand, Theorem 10.22 asserts the convergence of an infinite power series to a function f. The coefficients of the Taylor polynomials and the power series are the same in both cases, but Taylor's theorem approximates f by its Taylor polynomials $P_n(x)$ of degree n at c in the limit $x \to c$ with n fixed, while the power series theorem approximates f by $P_n(x)$ in the limit $n \to \infty$ with x fixed.

10.7.1. Taylor's theorem and power series. To explain the difference between Taylor's theorem and power series in more detail, we introduce an important distinction between smooth and analytic functions: smooth functions have continuous derivatives of all orders, while analytic functions are sums of power series.

Definition 10.27. Let $k \in \mathbb{N}$. A function $f : (a, b) \to \mathbb{R}$ is C^k on (a, b), written $f \in C^k(a, b)$, if it has continuous derivatives $f^{(j)} : (a, b) \to \mathbb{R}$ of orders $1 \leq j \leq k$. A function f is smooth (or C^∞, or infinitely differentiable) on (a, b), written $f \in C^\infty(a, b)$, if it has continuous derivatives of all orders on (a, b).

In fact, if f has derivatives of all orders, then they are automatically continuous, since the differentiability of $f^{(k)}$ implies its continuity; on the other hand, the existence of k derivatives of f does not imply the continuity of $f^{(k)}$. The statement "f is smooth" is sometimes used rather loosely to mean "f has as many continuous derivatives as we want," but we will use it to mean that f is C^∞.

Definition 10.28. A function $f : (a, b) \to \mathbb{R}$ is analytic on (a, b) if for every $c \in (a, b)$ the function f is the sum in a neighborhood of c of a power series centered at c with nonzero radius of convergence.

Strictly speaking, this is the definition of a real analytic function, and analytic functions are complex functions that are sums of power series. Since we consider only real functions here, we abbreviate "real analytic" to "analytic."

Theorem 10.22 implies that an analytic function is smooth: If f is analytic on (a, b) and $c \in (a, b)$, then there is an $R > 0$ and coefficients (a_n) such that

$$f(x) = \sum_{n=0}^\infty a_n (x - c)^n \qquad \text{for } |x - c| < R.$$

Then Theorem 10.22 implies that f has derivatives of all orders in $|x - c| < R$, and since $c \in (a, b)$ is arbitrary, f has derivatives of all orders in (a, b). Moreover, it follows that the coefficients a_n in the power series expansion of f at c are given by Taylor's formula.

What is less obvious is that a smooth function need not be analytic. If f is smooth, then we can define its Taylor coefficients $a_n = f^{(n)}(c)/n!$ at c for every $n \geq 0$, and write down the corresponding Taylor series $\sum a_n (x - c)^n$. The problem is that the Taylor series may have zero radius of convergence if the derivatives of f grow too rapidly as $n \to \infty$, in which case it diverges for every $x \neq c$, or the Taylor series may converge, but not to f.

10.7.2. A smooth, non-analytic function. In this section, we give an example of a smooth function that is not the sum of its Taylor series.

It follows from Proposition 10.26 that if

$$p(x) = \sum_{k=0}^n a_k x^k$$

is any polynomial function, then

$$\lim_{x \to \infty} \frac{p(x)}{e^x} = \sum_{k=0}^n a_k \lim_{x \to \infty} \frac{x^k}{e^x} = 0.$$
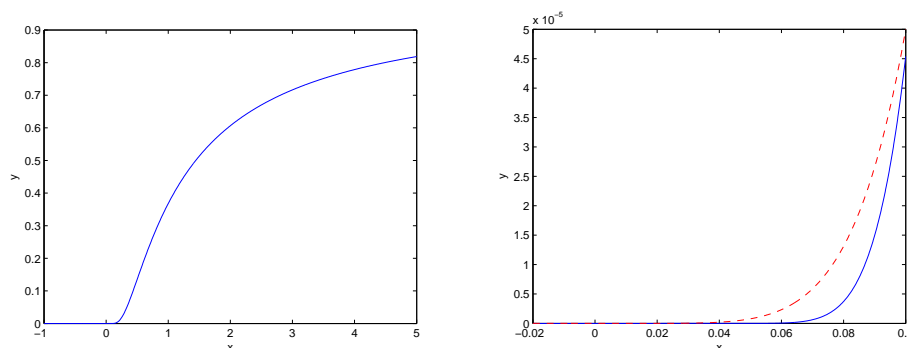
Figure 3. Left: Plot $y = \phi(x)$ of the smooth, non-analytic function defined in Proposition 10.29. Right: A detail of the function near $x = 0$. The dotted line is the power-function $y = x^6/50$. The graph of ϕ near 0 is "flatter' than the graph of the power-function, illustrating that $\phi(x)$ goes to zero faster than any power of x as $x \to 0$.

We will use this limit to exhibit a non-zero function that approaches zero faster than every power of x as $x \to 0$. As a result, all of its derivatives at 0 vanish, even though the function itself does not vanish in any neighborhood of 0. (See Figure 3.)

Proposition 10.29. Define $\phi : \mathbb{R} \to \mathbb{R}$ by

$$\phi(x) = \begin{cases} \exp(-1/x) & \text{if } x > 0, \\ 0 & \text{if } x \leq 0. \end{cases}$$

Then ϕ has derivatives of all orders on $\mathbb{R}$ and

$$\phi^{(n)}(0) = 0 \qquad \text{for all } n \geq 0.$$

Proof. The infinite differentiability of $\phi(x)$ at $x \neq 0$ follows from the chain rule. Moreover, its nth derivative has the form

$$\phi^{(n)}(x) = \begin{cases} p_n(1/x) \exp(-1/x) & \text{if } x > 0, \\ 0 & \text{if } x < 0, \end{cases}$$

where $p_n(1/x)$ is a polynomial of degree $2n$ in $1/x$. This follows, for example, by induction, since differentiation of $\phi^{(n)}$ shows that p_n satisfies the recursion relation

$$p_{n+1}(z) = z^2 \left[p_n(z) - p_n'(z) \right], \qquad p_0(z) = 1.$$

Thus, we just have to show that ϕ has derivatives of all orders at 0, and that these derivatives are equal to zero.

First, consider $\phi'(0)$. The left derivative $\phi'(0^-)$ of ϕ at 0 is 0 since $\phi(0) = 0$ and $\phi(h) = 0$ for all $h < 0$. To find the right derivative, we write $1/h = x$ and use

Proposition 10.26, which gives

$$\begin{aligned}
\phi'(0^+) &= \lim_{h \to 0^+} \left[\frac{\phi(h) - \phi(0)}{h} \right] \\
&= \lim_{h \to 0^+} \frac{\exp(-1/h)}{h} \\
&= \lim_{x \to \infty} \frac{x}{e^x} \\
&= 0.
\end{aligned}$$

Since both the left and right derivatives equal zero, we have $\phi'(0) = 0$.

To show that all the derivatives of ϕ at 0 exist and are zero, we use a proof by induction. Suppose that $\phi^{(n)}(0) = 0$, which we have verified for $n = 1$. The left derivative $\phi^{(n+1)}(0^-)$ is clearly zero, so we just need to prove that the right derivative is zero. Using the form of $\phi^{(n)}(h)$ for $h > 0$ and Proposition 10.26, we get that

$$\begin{aligned}
\phi^{(n+1)}(0^+) &= \lim_{h \to 0^+} \left[\frac{\phi^{(n)}(h) - \phi^{(n)}(0)}{h} \right] \\
&= \lim_{h \to 0^+} \frac{p_n(1/h) \exp(-1/h)}{h} \\
&= \lim_{x \to \infty} \frac{x p_n(x)}{e^x} \\
&= 0,
\end{aligned}$$

which proves the result.   □

Corollary 10.30. The function $\phi : \mathbb{R} \to \mathbb{R}$ defined by

$$\phi(x) = \begin{cases} \exp(-1/x) & \text{if } x > 0, \\ 0 & \text{if } x \leq 0, \end{cases}$$

is smooth but not analytic on $\mathbb{R}$.

Proof. From Proposition 10.29, the function ϕ is smooth, and the nth Taylor coefficient of ϕ at 0 is $a_n = 0$. The Taylor series of ϕ at 0 therefore converges to 0, so its sum is not equal to ϕ in any neighborhood of 0, meaning that ϕ is not analytic at 0.   □

The fact that the Taylor polynomial of ϕ at 0 is zero for every degree $n \in \mathbb{N}$ does not contradict Taylor's theorem, which says that for for every $n \in \mathbb{N}$ and $x > 0$ there exists $0 < \xi < x$ such that

$$\phi(x) = \frac{\phi^{(n)}(\xi)}{n!} x^n.$$

Since the derivatives of ϕ are bounded, it follows that there is a constant C_n, depending on n, such that

$$|\phi(x)| \leq C_n x^n \qquad \text{for all } 0 < x < \infty.$$

Thus, $\phi(x) \to 0$ as $x \to 0$ faster than any power of x. But this inequality does not imply that $\phi(x) = 0$ for $x > 0$ since C_n grows rapidly as n increases, and $C_n x^n \not\to 0$ as $n \to \infty$ for any $x > 0$, however small.

We can construct other smooth, non-analytic functions from ϕ.

Example 10.31. The function

$$\psi(x) = \begin{cases} \exp(-1/x^2) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$$

is infinitely differentiable on $\mathbb{R}$, since $\psi(x) = \phi(x^2)$ is a composition of smooth functions.

The function in the next example is useful in many parts of analysis. Before giving the example, we introduce some terminology.

Definition 10.32. A function $f : \mathbb{R} \to \mathbb{R}$ has compact support if there exists $R \geq 0$ such that $f(x) = 0$ for all $x \in \mathbb{R}$ with $|x| \geq R$.

It isn't hard to construct continuous functions with compact support; one example that vanishes for $|x| \geq 1$ is the piecewise-linear, triangular (or 'tent') function

$$f(x) = \begin{cases} 1 - |x| & \text{if } |x| < 1, \\ 0 & \text{if } |x| \geq 1. \end{cases}$$

By matching left and right derivatives of piecewise-polynomial functions, we can similarly construct C^1 or C^k functions with compact support. Using ϕ, however, we can construct a smooth (C^∞) function with compact support, which might seem unexpected at first sight.

Example 10.33. The function

$$\eta(x) = \begin{cases} \exp[-1/(1 - x^2)] & \text{if } |x| < 1, \\ 0 & \text{if } |x| \geq 1, \end{cases}$$

is infinitely differentiable on $\mathbb{R}$, since $\eta(x) = \phi(1 - x^2)$ is a composition of smooth functions. Moreover, it vanishes for $|x| \geq 1$, so it is a smooth function with compact support. Figure 4 shows its graph. This function is sometimes called a 'bump' function.

The function ϕ defined in Proposition 10.29 illustrates that knowing the values of a smooth function and all of its derivatives at one point does not tell us anything about the values of the function at nearby points. This behavior contrasts with, and highlights, the remarkable property of analytic functions that the values of an analytic function and all of its derivatives at a single point of an interval determine the function on the whole interval.

We make this principle of analytic continuation precise in the following proposition. The proof uses a common trick of going from a local result (equality of functions in a neighborhood of a point) to a global result (equality of functions on the whole of their connected domain) by proving that an appropriate subset is open, closed, and non-empty.
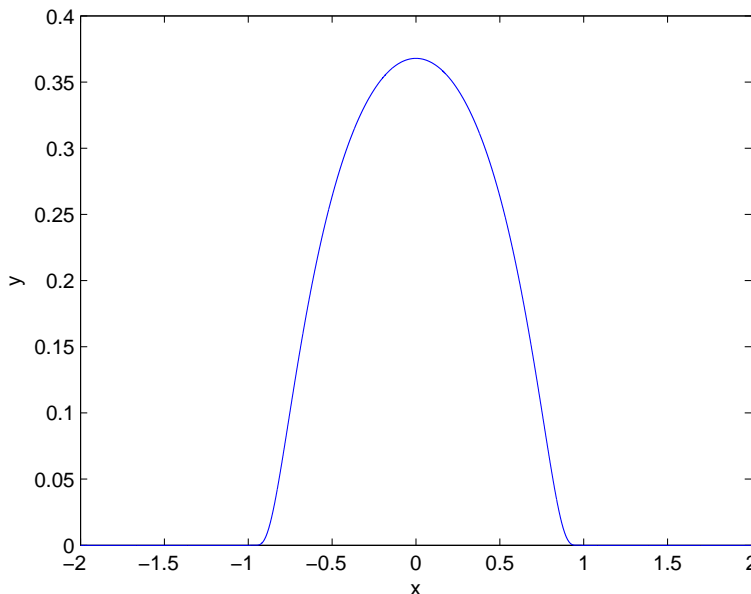
Figure 4. Plot of the smooth, compactly supported "bump" function defined in Example 10.33.

Proposition 10.34. Suppose that $f, g : (a, b) \to \mathbb{R}$ are analytic functions on an open interval (a, b). If $f^{(n)}(c) = g^{(n)}(c)$ for all $n \geq 0$ at some point $c \in (a, b)$, then $f = g$ on (a, b).

Proof. Let
$$E = \left\{ x \in (a, b) : f^{(n)}(x) = g^{(n)}(x) \text{ all } n \geq 0 \right\}.$$
The continuity of the derivatives $f^{(n)}$, $g^{(n)}$ implies that E is closed in (a, b): If $x_k \in E$ and $x_k \to x \in (a, b)$, then
$$f^{(n)}(x) = \lim_{k \to \infty} f^{(n)}(x_k) = \lim_{k \to \infty} g^{(n)}(x_k) = g^{(n)}(x),$$
so $x \in E$, and E is closed.

The analyticity of f, g implies that E is open in (a, b): If $x \in E$, then $f = g$ in some open interval $(x - r, x + r)$ with $r > 0$, since both functions have the same Taylor coefficients and convergent power series centered at x, so $f^{(n)} = g^{(n)}$ in $(x - r, x + r)$, meaning that $(x - r, x + r) \subset E$, and E is open.

From Theorem 5.63, the interval (a, b) is connected, meaning that the only subsets that are open and closed in (a, b) are the empty set and the entire interval. But $E \neq \varnothing$ since $c \in E$, so $E = (a, b)$, which proves the result. $\square$

It is worth noting the choice of the set E in the preceding proof. For example, the proof would not work if we try to use the set
$$\tilde{E} = \{ x \in (a, b) : f(x) = g(x) \}$$

instead of E. The continuity of f, g implies that $\tilde{E}$ is closed, but $\tilde{E}$ is not, in general, open.

One particular consequence of Proposition 10.34 is that a non-zero analytic function on $\mathbb{R}$ cannot have compact support, since an analytic function on $\mathbb{R}$ that is equal to zero on any interval $(a, b) \subset \mathbb{R}$ must equal zero on $\mathbb{R}$. Thus, the non-analyticity of the 'bump'-function η in Example 10.33 is essential.

The Riemann Integral

I know of some universities in England where the Lebesgue integral is *taught* in the first year of a mathematics degree instead of the Riemann integral, but I know of no universities in England where students *learn* the Lebesgue integral in the first year of a mathematics degree. (Approximate quotation attributed to T. W. Körner)

Let $f : [a, b] \to \mathbb{R}$ be a bounded (not necessarily continuous) function on a compact (closed, bounded) interval. We will define what it means for f to be Riemann integrable on $[a, b]$ and, in that case, define its Riemann integral $\int_a^b f$. The integral of f on $[a, b]$ is a real number whose geometrical interpretation is the signed area under the graph $y = f(x)$ for $a \le x \le b$. This number is also called the definite integral of f. By integrating f over an interval $[a, x]$ with varying right end-point, we get a function of x, called an indefinite integral of f.

The most important result about integration is the fundamental theorem of calculus, which states that integration and differentiation are inverse operations in an appropriately understood sense. Among other things, this connection enables us to compute many integrals explicitly. We will prove the fundamental theorem in the next chapter. In this chapter, we define the Riemann integral and prove some of its basic properties.

Integrability is a less restrictive condition on a function than differentiability. Generally speaking, integration makes functions smoother, while differentiation makes functions rougher. For example, the indefinite integral of every continuous function exists and is differentiable, whereas the derivative of a continuous function need not exist (and typically doesn't).

The Riemann integral is the simplest integral to define, and it allows one to integrate every continuous function as well as some not-too-badly discontinuous functions. There are, however, many other types of integrals, the most important of which is the Lebesgue integral. The Lebesgue integral allows one to integrate unbounded or highly discontinuous functions whose Riemann integrals do not exist,

and it has better mathematical properties than the Riemann integral. The definition of the Lebesgue integral is more involved, requiring the use of measure theory, and we will not discuss it here. In any event, the Riemann integral is adequate for many purposes, and even if one needs the Lebesgue integral, it is best to understand the Riemann integral first.

11.1. The supremum and infimum of functions

In this section we collect some results about the supremum and infimum of functions that we use to study Riemann integration. These results can be referred back to as needed.

From Definition 6.11, the supremum or infimum of a function is the supremum or infimum of its range, and results about the supremum or infimum of sets translate immediately to results about functions. There are, however, a few differences, which come from the fact that we often compare the values of functions at the same point, rather than all of their values simultaneously.

Inequalities and operations on functions are defined pointwise as usual; for example, if $f, g : A \to \mathbb{R}$, then $f \leq g$ means that $f(x) \leq g(x)$ for every $x \in A$, and $f + g : A \to \mathbb{R}$ is defined by $(f + g)(x) = f(x) + g(x)$.

Proposition 11.1. Suppose that $f, g : A \to \mathbb{R}$ and $f \leq g$. Then

$$\sup_A f \leq \sup_A g, \qquad \inf_A f \leq \inf_A g.$$

Proof. If $\sup g = \infty$, then $\sup f \leq \sup g$. Otherwise, if $f \leq g$ and g is bounded from above, then

$$f(x) \leq g(x) \leq \sup_A g \qquad \text{for every } x \in A.$$

Thus, f is bounded from above by $\sup_A g$, so $\sup_A f \leq \sup_A g$. Similarly, $-f \geq -g$ implies that $\sup_A(-f) \geq \sup_A(-g)$, so $\inf_A f \leq \inf_A g$. $\square$

Note that $f \leq g$ does not imply that $\sup_A f \leq \inf_A g$; to get that conclusion, we need to know that $f(x) \leq g(y)$ for all $x, y \in A$ and use Proposition 2.24.

Example 11.2. Define $f, g : [0, 1] \to \mathbb{R}$ by $f(x) = 2x$, $g(x) = 2x + 1$. Then $f < g$ and

$$\sup_{[0,1]} f = 2, \quad \inf_{[0,1]} f = 0, \qquad \sup_{[0,1]} g = 3, \quad \inf_{[0,1]} g = 1.$$

Thus, $\sup f > \inf g$ even though $f < g$.

As for sets, the supremum and infimum of functions do not, in general, preserve strict inequalities, and a function need not attain its supremum or infimum even if it exists.

Example 11.3. Define $f : [0, 1] \to \mathbb{R}$ by

$$f(x) = \begin{cases} x & \text{if } 0 \leq x < 1, \\ 0 & \text{if } x = 1. \end{cases}$$

Then $f < 1$ on $[0, 1]$ but $\sup_{[0,1]} f = 1$, and there is no point $x \in [0, 1]$ such that $f(x) = 1$.

Next, we consider the supremum and infimum of linear combinations of functions. Multiplication of a function by a positive constant multiplies the inf or sup, while multiplication by a negative constant switches the inf and sup,

Proposition 11.4. Suppose that $f : A \to \mathbb{R}$ is a bounded function and $c \in \mathbb{R}$. If $c \geq 0$, then

$$\sup_A cf = c \sup_A f, \qquad \inf_A cf = c \inf_A f.$$

If $c < 0$, then

$$\sup_A cf = c \inf_A f, \qquad \inf_A cf = c \sup_A f.$$

Proof. Apply Proposition 2.23 to the set $\{cf(x) : x \in A\} = c\{f(x) : x \in A\}$. $\qquad \square$

For sums of functions, we get an inequality.

Proposition 11.5. If $f, g : A \to \mathbb{R}$ are bounded functions, then

$$\sup_A (f + g) \leq \sup_A f + \sup_A g, \qquad \inf_A (f + g) \geq \inf_A f + \inf_A g.$$

Proof. Since $f(x) \leq \sup_A f$ and $g(x) \leq \sup_A g$ for every $x \in [a, b]$, we have

$$f(x) + g(x) \leq \sup_A f + \sup_A g.$$

Thus, $f + g$ is bounded from above by $\sup_A f + \sup_A g$, so

$$\sup_A (f + g) \leq \sup_A f + \sup_A g.$$

The proof for the infimum is analogous (or apply the result for the supremum to the functions $-f$, $-g$). $\qquad \square$

We may have strict inequality in Proposition 11.5 because f and g may take values close to their suprema (or infima) at different points.

Example 11.6. Define $f, g : [0, 1] \to \mathbb{R}$ by $f(x) = x$, $g(x) = 1 - x$. Then

$$\sup_{[0,1]} f = \sup_{[0,1]} g = \sup_{[0,1]} (f + g) = 1,$$

so $\sup(f + g) = 1$ but $\sup f + \sup g = 2$. Here, f attains its supremum at 1, while g attains its supremum at 0.

Finally, we prove some inequalities that involve the absolute value.

Proposition 11.7. If $f, g : A \to \mathbb{R}$ are bounded functions, then

$$\left| \sup_A f - \sup_A g \right| \leq \sup_A |f - g|, \qquad \left| \inf_A f - \inf_A g \right| \leq \sup_A |f - g|.$$

Proof. Since $f = f - g + g$ and $f - g \leq |f - g|$, we get from Proposition 11.5 and Proposition 11.1 that

$$\sup_A f \leq \sup_A (f - g) + \sup_A g \leq \sup_A |f - g| + \sup_A g,$$

so

$$\sup_A f - \sup_A g \leq \sup_A |f - g|.$$

Exchanging f and g in this inequality, we get

$$\sup_A g - \sup_A f \le \sup_A |f - g|,$$

which implies that

$$\left| \sup_A f - \sup_A g \right| \le \sup_A |f - g|.$$

Replacing f by $-f$ and g by $-g$ in this inequality, we get

$$\left| \inf_A f - \inf_A g \right| \le \sup_A |f - g|,$$

where we use the fact that $\sup(-f) = -\inf f$. $\qquad\square$

Proposition 11.8. If $f, g : A \to \mathbb{R}$ are bounded functions such that

$$|f(x) - f(y)| \le |g(x) - g(y)| \qquad \text{for all } x, y \in A,$$

then

$$\sup_A f - \inf_A f \le \sup_A g - \inf_A g.$$

Proof. The condition implies that for all $x, y \in A$, we have

$$f(x) - f(y) \le |g(x) - g(y)| = \max\left[g(x), g(y)\right] - \min\left[g(x), g(y)\right] \le \sup_A g - \inf_A g,$$

which implies that

$$\sup\{f(x) - f(y) : x, y \in A\} \le \sup_A g - \inf_A g.$$

From Proposition 2.24, we have

$$\sup\{f(x) - f(y) : x, y \in A\} = \sup_A f - \inf_A f,$$

so the result follows. $\qquad\square$

11.2. Definition of the integral

The definition of the integral is more involved than the definition of the derivative. The derivative is approximated by difference quotients, whereas the integral is approximated by upper and lower sums based on a partition of an interval.

We say that two intervals are almost disjoint if they are disjoint or intersect only at a common endpoint. For example, the intervals $[0, 1]$ and $[1, 3]$ are almost disjoint, whereas the intervals $[0, 2]$ and $[1, 3]$ are not.

Definition 11.9. Let I be a nonempty, compact interval. A partition of I is a finite collection $\{I_1, I_2, \ldots, I_n\}$ of almost disjoint, nonempty, compact subintervals whose union is I.

A partition of $[a, b]$ with subintervals $I_k = [x_{k-1}, x_k]$ is determined by the set of endpoints of the intervals

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b.$$

Abusing notation, we will denote a partition P either by its intervals

$$P = \{I_1, I_2, \ldots, I_n\}$$

or by the set of endpoints of the intervals

$$P = \{x_0, x_1, x_2, \ldots, x_{n-1}, x_n\}.$$

We'll adopt either notation as convenient; the context should make it clear which one is being used. There is always one more endpoint than interval.

Example 11.10. The set of intervals

$$\{[0, 1/5], [1/5, 1/4], [1/4, 1/3], [1/3, 1/2], [1/2, 1]\}$$

is a partition of $[0, 1]$. The corresponding set of endpoints is

$$\{0, 1/5, 1/4, 1/3, 1/2, 1\}.$$

We denote the length of an interval $I = [a, b]$ by

$$|I| = b - a.$$

Note that the sum of the lengths $|I_k| = x_k - x_{k-1}$ of the almost disjoint subintervals in a partition $\{I_1, I_2, \ldots, I_n\}$ of an interval I is equal to length of the whole interval. This is obvious geometrically; algebraically, it follows from the telescoping series

$$\begin{aligned}
\sum_{k=1}^{n} |I_k| &= \sum_{k=1}^{n} (x_k - x_{k-1}) \\
&= x_n - x_{n-1} + x_{n-1} - x_{n-2} + \cdots + x_2 - x_1 + x_1 - x_0 \\
&= x_n - x_0 \\
&= |I|.
\end{aligned}$$

Suppose that $f : [a, b] \to \mathbb{R}$ is a bounded function on the compact interval $I = [a, b]$ with

$$M = \sup_I f, \qquad m = \inf_I f.$$

If $P = \{I_1, I_2, \ldots, I_n\}$ is a partition of I, let

$$M_k = \sup_{I_k} f, \qquad m_k = \inf_{I_k} f.$$

These suprema and infima are well-defined, finite real numbers since f is bounded. Moreover,

$$m \le m_k \le M_k \le M.$$

If f is continuous on the interval I, then it is bounded and attains its maximum and minimum values on each subinterval, but a bounded discontinuous function need not attain its supremum or infimum.

We define the upper Riemann sum of f with respect to the partition P by

$$U(f; P) = \sum_{k=1}^{n} M_k |I_k| = \sum_{k=1}^{n} M_k (x_k - x_{k-1}),$$

and the lower Riemann sum of f with respect to the partition P by

$$L(f; P) = \sum_{k=1}^{n} m_k |I_k| = \sum_{k=1}^{n} m_k (x_k - x_{k-1}).$$

Geometrically, $U(f;P)$ is the sum of the signed areas of rectangles based on the intervals I_k that lie above the graph of f, and $L(f;P)$ is the sum of the signed areas of rectangles that lie below the graph of f. Note that

$$m(b-a) \leq L(f;P) \leq U(f;P) \leq M(b-a).$$

Let $\Pi(a,b)$, or Π for short, denote the collection of all partitions of $[a,b]$. We define the upper Riemann integral of f on $[a,b]$ by

$$U(f) = \inf_{P \in \Pi} U(f;P).$$

The set $\{U(f;P) : P \in \Pi\}$ of all upper Riemann sums of f is bounded from below by $m(b-a)$, so this infimum is well-defined and finite. Similarly, the set $\{L(f;P) : P \in \Pi\}$ of all lower Riemann sums is bounded from above by $M(b-a)$, and we define the lower Riemann integral of f on $[a,b]$ by

$$L(f) = \sup_{P \in \Pi} L(f;P).$$

These upper and lower sums and integrals depend on the interval $[a,b]$ as well as the function f, but to simplify the notation we won't show this explicitly. A commonly used alternative notation for the upper and lower integrals is

$$U(f) = \overline{\int_a^b} f, \qquad L(f) = \underline{\int_a^b} f.$$

Note the use of "lower-upper" and "upper-lower" approximations for the integrals: we take the infimum of the upper sums and the supremum of the lower sums. As we show in Proposition 11.22 below, we always have $L(f) \leq U(f)$, but in general the upper and lower integrals need not be equal. We define Riemann integrability by their equality.

Definition 11.11. A function $f : [a,b] \to \mathbb{R}$ is Riemann integrable on $[a,b]$ if it is bounded and its upper integral $U(f)$ and lower integral $L(f)$ are equal. In that case, the Riemann integral of f on $[a,b]$, denoted by

$$\int_a^b f(x)\, dx, \qquad \int_a^b f, \qquad \int_{[a,b]} f$$

or similar notations, is the common value of $U(f)$ and $L(f)$.

An unbounded function is not Riemann integrable. In the following, "integrable" will mean "Riemann integrable, and "integral" will mean "Riemann integral" unless stated explicitly otherwise.

11.2.1. Examples. Let us illustrate the definition of Riemann integrability with a number of examples.

Example 11.12. Define $f : [0,1] \to \mathbb{R}$ by

$$f(x) = \begin{cases} 1/x & \text{if } 0 < x \leq 1, \\ 0 & \text{if } x = 0. \end{cases}$$

Then

$$\int_0^1 \frac{1}{x}\, dx$$

isn't defined as a Riemann integral becuase f is unbounded. In fact, if

$$0 < x_1 < x_2 < \cdots < x_{n-1} < 1$$

is a partition of $[0, 1]$, then

$$\sup_{[0, x_1]} f = \infty,$$

so the upper Riemann sums of f are not well-defined.

An integral with an unbounded interval of integration, such as

$$\int_1^\infty \frac{1}{x}\, dx,$$

also isn't defined as a Riemann integral. In this case, a partition of $[1, \infty)$ into finitely many intervals contains at least one unbounded interval, so the corresponding Riemann sum is not well-defined. A partition of $[1, \infty)$ into bounded intervals (for example, $I_k = [k, k+1]$ with $k \in \mathbb{N}$) gives an infinite series rather than a finite Riemann sum, leading to questions of convergence.

One can interpret the integrals in this example as limits of Riemann integrals, or improper Riemann integrals,

$$\int_0^1 \frac{1}{x}\, dx = \lim_{\epsilon \to 0^+} \int_\epsilon^1 \frac{1}{x}\, dx, \qquad \int_1^\infty \frac{1}{x}\, dx = \lim_{r \to \infty} \int_1^r \frac{1}{x}\, dx,$$

but these are not proper Riemann integrals in the sense of Definition 11.11. Such improper Riemann integrals involve two limits — a limit of Riemann sums to define the Riemann integrals, followed by a limit of Riemann integrals. Both of the improper integrals in this example diverge to infinity. (See Section 12.4.)

Next, we consider some examples of bounded functions on compact intervals.

Example 11.13. The constant function $f(x) = 1$ on $[0, 1]$ is Riemann integrable, and

$$\int_0^1 1\, dx = 1.$$

To show this, let $P = \{I_1, I_2, \ldots, I_n\}$ be any partition of $[0, 1]$ with endpoints

$$\{0, x_1, x_2, \ldots, x_{n-1}, 1\}.$$

Since f is constant,

$$M_k = \sup_{I_k} f = 1, \quad m_k = \inf_{I_k} f = 1 \qquad \text{for } k = 1, \ldots, n,$$

and therefore

$$U(f; P) = L(f; P) = \sum_{k=1}^n (x_k - x_{k-1}) = x_n - x_0 = 1.$$

Geometrically, this equation is the obvious fact that the sum of the areas of the rectangles over (or, equivalently, under) the graph of a constant function is exactly equal to the area under the graph. Thus, every upper and lower sum of f on $[0, 1]$ is equal to 1, which implies that the upper and lower integrals

$$U(f) = \inf_{P \in \Pi} U(f; P) = \inf\{1\} = 1, \qquad L(f) = \sup_{P \in \Pi} L(f; P) = \sup\{1\} = 1$$

are equal, and the integral of f is 1.

More generally, the same argument shows that every constant function $f(x) = c$ is integrable and

$$\int_a^b c\,dx = c(b-a).$$

The following is an example of a discontinuous function that is Riemann integrable.

Example 11.14. The function

$$f(x) = \begin{cases} 0 & \text{if } 0 < x \leq 1 \\ 1 & \text{if } x = 0 \end{cases}$$

is Riemann integrable, and

$$\int_0^1 f\,dx = 0.$$

To show this, let $P = \{I_1, I_2, \ldots, I_n\}$ be a partition of $[0,1]$. Then, since $f(x) = 0$ for $x > 0$,

$$M_k = \sup_{I_k} f = 0, \quad m_k = \inf_{I_k} f = 0 \qquad \text{for } k = 2, \ldots, n.$$

The first interval in the partition is $I_1 = [0, x_1]$, where $0 < x_1 \leq 1$, and

$$M_1 = 1, \quad m_1 = 0,$$

since $f(0) = 1$ and $f(x) = 0$ for $0 < x \leq x_1$. It follows that

$$U(f; P) = x_1, \qquad L(f; P) = 0.$$

Thus, $L(f) = 0$ and

$$U(f) = \inf\{x_1 : 0 < x_1 \leq 1\} = 0,$$

so $U(f) = L(f) = 0$ are equal, and the integral of f is 0. In this example, the infimum of the upper Riemann sums is not attained and $U(f; P) > U(f)$ for every partition P.

A similar argument shows that a function $f : [a, b] \to \mathbb{R}$ that is zero except at finitely many points in $[a, b]$ is Riemann integrable with integral 0.

The next example is a bounded function on a compact interval whose Riemann integral doesn't exist.

Example 11.15. The Dirichlet function $f : [0, 1] \to \mathbb{R}$ is defined by

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

That is, f is one at every rational number and zero at every irrational number.

This function is not Riemann integrable. If $P = \{I_1, I_2, \ldots, I_n\}$ is a partition of $[0, 1]$, then

$$M_k = \sup_{I_k} f = 1, \qquad m_k = \inf_{I_k} = 0,$$

since every interval of non-zero length contains both rational and irrational numbers. It follows that

$$U(f; P) = 1, \qquad L(f; P) = 0$$

for every partition P of $[0, 1]$, so $U(f) = 1$ and $L(f) = 0$ are not equal.

The Dirichlet function is discontinuous at every point of $[0, 1]$, and the moral of the last example is that the Riemann integral of a highly discontinuous function need not exist. Nevertheless, some fairly discontinuous functions are still Riemann integrable.

Example 11.16. The Thomae function defined in Example 7.14 is Riemann integrable. The proof is left as an exercise.

Theorem 11.58 and Theorem 11.61 below give precise statements of the extent to which a Riemann integrable function can be discontinuous.

11.2.2. Refinements of partitions. As the previous examples illustrate, a direct verification of integrability from Definition 11.11 is unwieldy even for the simplest functions because we have to consider all possible partitions of the interval of integration. To give an effective analysis of Riemann integrability, we need to study how upper and lower sums behave under the refinement of partitions.

Definition 11.17. A partition $Q = \{J_1, J_2, \ldots, J_m\}$ is a refinement of a partition $P = \{I_1, I_2, \ldots, I_n\}$ if every interval I_k in P is an almost disjoint union of one or more intervals J_ℓ in Q.

Equivalently, if we represent partitions by their endpoints, then Q is a refinement of P if $Q \supset P$, meaning that every endpoint of P is an endpoint of Q. We don't require that every interval — or even any interval — in a partition has to be split into smaller intervals to obtain a refinement; for example, every partition is a refinement of itself.

Example 11.18. Consider the partitions of $[0, 1]$ with endpoints

$$P = \{0, 1/2, 1\}, \qquad Q = \{0, 1/3, 2/3, 1\}, \qquad R = \{0, 1/4, 1/2, 3/4, 1\}.$$

Thus, P, Q, and R partition $[0, 1]$ into intervals of equal length $1/2$, $1/3$, and $1/4$, respectively. Then Q is not a refinement of P but R is a refinement of P.

Given two partitions, neither one need be a refinement of the other. However, two partitions P, Q always have a common refinement; the smallest one is $R = P \cup Q$, meaning that the endpoints of R are exactly the endpoints of P or Q (or both).

Example 11.19. Let $P = \{0, 1/2, 1\}$ and $Q = \{0, 1/3, 2/3, 1\}$, as in Example 11.18. Then Q isn't a refinement of P and P isn't a refinement of Q. The partition $S = P \cup Q$, or

$$S = \{0, 1/3, 1/2, 2/3, 1\},$$

is a refinement of both P and Q. The partition S is not a refinement of R, but $T = R \cup S$, or

$$T = \{0, 1/4, 1/3, 1/2, 2/3, 3/4, 1\},$$

is a common refinement of all of the partitions $\{P, Q, R, S\}$.

As we show next, refining partitions decreases upper sums and increases lower sums. (The proof is easier to understand than it is to write out — draw a picture!)

Theorem 11.20. Suppose that $f : [a, b] \to \mathbb{R}$ is bounded, P is a partitions of $[a, b]$, and Q is refinement of P. Then

$$U(f; Q) \leq U(f; P), \qquad L(f; P) \leq L(f; Q).$$

Proof. Let

$$P = \{I_1, I_2, \ldots, I_n\}, \qquad Q = \{J_1, J_2, \ldots, J_m\}$$

be partitions of $[a, b]$, where Q is a refinement of P, so $m \geq n$. We list the intervals in increasing order of their endpoints. Define

$$M_k = \sup_{I_k} f, \quad m_k = \inf_{I_k} f, \qquad M'_\ell = \sup_{J_\ell} f, \quad m'_\ell = \inf_{J_\ell} f.$$

Since Q is a refinement of P, each interval I_k in P is an almost disjoint union of intervals in Q, which we can write as

$$I_k = \bigcup_{\ell = p_k}^{q_k} J_\ell$$

for some indices $p_k \leq q_k$. If $p_k < q_k$, then I_k is split into two or more smaller intervals in Q, and if $p_k = q_k$, then I_k belongs to both P and Q. Since the intervals are listed in order, we have

$$p_1 = 1, \qquad p_{k+1} = q_k + 1, \qquad q_n = m.$$

If $p_k \leq \ell \leq q_k$, then $J_\ell \subset I_k$, so

$$M'_\ell \leq M_k, \qquad m_k \geq m'_\ell \qquad \text{for } p_k \leq \ell \leq q_k.$$

Using the fact that the sum of the lengths of the J-intervals is the length of the corresponding I-interval, we get that

$$\sum_{\ell = p_k}^{q_k} M'_\ell |J_\ell| \leq \sum_{\ell = p_k}^{q_k} M_k |J_\ell| = M_k \sum_{\ell = p_k}^{q_k} |J_\ell| = M_k |I_k|.$$

It follows that

$$U(f; Q) = \sum_{\ell = 1}^{m} M'_\ell |J_\ell| = \sum_{k=1}^{n} \sum_{\ell = p_k}^{q_k} M'_\ell |J_\ell| \leq \sum_{k=1}^{n} M_k |I_k| = U(f; P).$$

Similarly,

$$\sum_{\ell = p_k}^{q_k} m'_\ell |J_\ell| \geq \sum_{\ell = p_k}^{q_k} m_k |J_\ell| = m_k |I_k|,$$

and

$$L(f; Q) = \sum_{k=1}^{n} \sum_{\ell = p_k}^{q_k} m'_\ell |J_\ell| \geq \sum_{k=1}^{n} m_k |I_k| = L(f; P),$$

which proves the result. □

It follows from this theorem that all lower sums are less than or equal to all upper sums, not just the lower and upper sums associated with the same partition.

Proposition 11.21. If $f : [a, b] \to \mathbb{R}$ is bounded and P, Q are partitions of $[a, b]$, then

$$L(f; P) \leq U(f; Q).$$

Proof. Let R be a common refinement of P and Q. Then, by Theorem 11.20,

$$L(f;P) \leq L(f;R), \qquad U(f;R) \leq U(f;Q).$$

It follows that

$$L(f;P) \leq L(f;R) \leq U(f;R) \leq U(f;Q).$$

$\square$

An immediate consequence of this result is that the lower integral is always less than or equal to the upper integral.

Proposition 11.22. If $f : [a,b] \to \mathbb{R}$ is bounded, then

$$L(f) \leq U(f).$$

Proof. Let

$$A = \{L(f;P) : P \in \Pi\}, \qquad B = \{U(f;P) : P \in \Pi\}.$$

From Proposition 11.21, $L \leq U$ for every $L \in A$ and $U \in B$, so Proposition 2.22 implies that $\sup A \leq \inf B$, or $L(f) \leq U(f)$. $\square$

11.3. The Cauchy criterion for integrability

The following theorem gives a criterion for integrability that is analogous to the Cauchy condition for the convergence of a sequence.

Theorem 11.23. A bounded function $f : [a,b] \to \mathbb{R}$ is Riemann integrable if and only if for every $\epsilon > 0$ there exists a partition P of $[a,b]$, which may depend on ϵ, such that

$$U(f;P) - L(f;P) < \epsilon.$$

Proof. First, suppose that the condition holds. Let $\epsilon > 0$ and choose a partition P that satisfies the condition. Then, since $U(f) \leq U(f;P)$ and $L(f;P) \leq L(f)$, we have

$$0 \leq U(f) - L(f) \leq U(f;P) - L(f;P) < \epsilon.$$

Since this inequality holds for every $\epsilon > 0$, we must have $U(f) - L(f) = 0$, and f is integrable.

Conversely, suppose that f is integrable. Given any $\epsilon > 0$, there are partitions Q, R such that

$$U(f;Q) < U(f) + \frac{\epsilon}{2}, \qquad L(f;R) > L(f) - \frac{\epsilon}{2}.$$

Let P be a common refinement of Q and R. Then, by Theorem 11.20,

$$U(f;P) - L(f;P) \leq U(f;Q) - L(f;R) < U(f) - L(f) + \epsilon.$$

Since $U(f) = L(f)$, the condition follows. $\square$

If $U(f;P) - L(f;P) < \epsilon$, then $U(f;Q) - L(f;Q) < \epsilon$ for every refinement Q of P, so the Cauchy condition means that a function is integrable if and only if its upper and lower sums get arbitrarily close together for all sufficiently refined partitions.

It is worth considering in more detail what the Cauchy condition in Theorem 11.23 implies about the behavior of a Riemann integrable function.

Definition 11.24. The oscillation of a bounded function f on a set A is

$$\operatorname*{osc}_{A} f = \sup_{A} f - \inf_{A} f.$$

If $f : [a, b] \to \mathbb{R}$ is bounded and $P = \{I_1, I_2, \ldots, I_n\}$ is a partition of $[a, b]$, then

$$U(f; P) - L(f; P) = \sum_{k=1}^{n} \sup_{I_k} f \cdot |I_k| - \sum_{k=1}^{n} \inf_{I_k} f \cdot |I_k| = \sum_{k=1}^{n} \operatorname*{osc}_{I_k} f \cdot |I_k|.$$

A function f is Riemann integrable if we can make $U(f; P) - L(f; P)$ as small as we wish. This is the case if we can find a sufficiently refined partition P such that the oscillation of f on most intervals is arbitrarily small, and the sum of the lengths of the remaining intervals (where the oscillation of f is large) is arbitrarily small. For example, the discontinuous function in Example 11.14 has zero oscillation on every interval except the first one, where the function has oscillation one, but the length of that interval can be made as small as we wish.

Thus, roughly speaking, a function is Riemann integrable if it oscillates by an arbitrary small amount except on a finite collection of intervals whose total length is arbitrarily small. Theorem 11.58 gives a precise statement.

One direct consequence of the Cauchy criterion is that a function is integrable if we can estimate its oscillation by the oscillation of an integrable function.

Proposition 11.25. Suppose that $f, g : [a, b] \to \mathbb{R}$ are bounded functions and g is integrable on $[a, b]$. If there exists a constant $C \geq 0$ such that

$$\operatorname*{osc}_{I} f \leq C \operatorname*{osc}_{I} g$$

on every interval $I \subset [a, b]$, then f is integrable.

Proof. If $P = \{I_1, I_2, \ldots, I_n\}$ is a partition of $[a, b]$, then

$$\begin{aligned}
U(f; P) - L(f; P) &= \sum_{k=1}^{n} \operatorname*{osc}_{I_k} f \cdot |I_k| \\
&\leq C \sum_{k=1}^{n} \operatorname*{osc}_{I_k} g \cdot |I_k| \\
&\leq C \left[U(g; P) - L(g; P) \right].
\end{aligned}$$

Thus, f satisfies the Cauchy criterion in Theorem 11.23 if g does, which proves that f is integrable if g is integrable. $\square$

We can also use the Cauchy criterion to give a sequential characterization of integrability.

Theorem 11.26. A bounded function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if there is a sequence (P_n) of partitions such that

$$\lim_{n \to \infty} [U(f; P_n) - L(f; P_n)] = 0.$$

In that case,

$$\int_a^b f = \lim_{n\to\infty} U(f;P_n) = \lim_{n\to\infty} L(f;P_n).$$

Proof. First, suppose that the condition holds. Then, given $\epsilon > 0$, there is an $n \in \mathbb{N}$ such that $U(f;P_n) - L(f;P_n) < \epsilon$, so Theorem 11.23 implies that f is integrable and $U(f) = L(f)$.

Furthermore, since $U(f) \leq U(f;P_n)$ and $L(f;P_n) \leq L(f)$, we have

$$0 \leq U(f;P_n) - U(f) = U(f;P_n) - L(f) \leq U(f;P_n) - L(f;P_n).$$

Since the limit of the right-hand side is zero, the 'squeeze' theorem implies that

$$\lim_{n\to\infty} U(f;P_n) = U(f) = \int_a^b f.$$

It also follows that

$$\lim_{n\to\infty} L(f;P_n) = \lim_{n\to\infty} U(f;P_n) - \lim_{n\to\infty} [U(f;P_n) - L(f;P_n)] = \int_a^b f.$$

Conversely, if f is integrable then, by Theorem 11.23, for every $n \in \mathbb{N}$ there exists a partition P_n such that

$$0 \leq U(f;P_n) - L(f;P_n) < \frac{1}{n},$$

and $U(f;P_n) - L(f;P_n) \to 0$ as $n \to \infty$. $\qquad\square$

Note that if the limits of $U(f;P_n)$ and $L(f;P_n)$ both exist and are equal, then

$$\lim_{n\to\infty} [U(f;P_n) - L(f;P_n)] = \lim_{n\to\infty} U(f;P_n) - \lim_{n\to\infty} L(f;P_n),$$

so the conditions of the theorem are satisfied. Conversely, the proof of the theorem shows that if the limit of $U(f;P_n) - L(f;P_n)$ is zero, then the limits of $U(f;P_n)$ and $L(f;P_n)$ both exist and are equal. This isn't true for general sequences, where one may have $\lim(a_n - b_n) = 0$ even though $\lim a_n$ and $\lim b_n$ don't exist.

Theorem 11.26 provides one way to prove the existence of an integral and, in some cases, evaluate it.

Example 11.27. Consider the function $f(x) = x^2$ on $[0,1]$. Let P_n be the partition of $[0,1]$ into n-intervals of equal length $1/n$ with endpoints $x_k = k/n$ for $k = 0,1,2,\ldots,n$. If $I_k = [(k-1)/n, k/n]$ is the kth interval, then

$$\sup_{I_k} f = x_k^2, \qquad \inf_{I_k} f = x_{k-1}^2$$

since f is increasing. Using the formula for the sum of squares

$$\sum_{k=1}^n k^2 = \frac{1}{6}n(n+1)(2n+1),$$

we get

$$U(f;P_n) = \sum_{k=1}^n x_k^2 \cdot \frac{1}{n} = \frac{1}{n^3} \sum_{k=1}^n k^2 = \frac{1}{6}\left(1 + \frac{1}{n}\right)\left(2 + \frac{1}{n}\right)$$
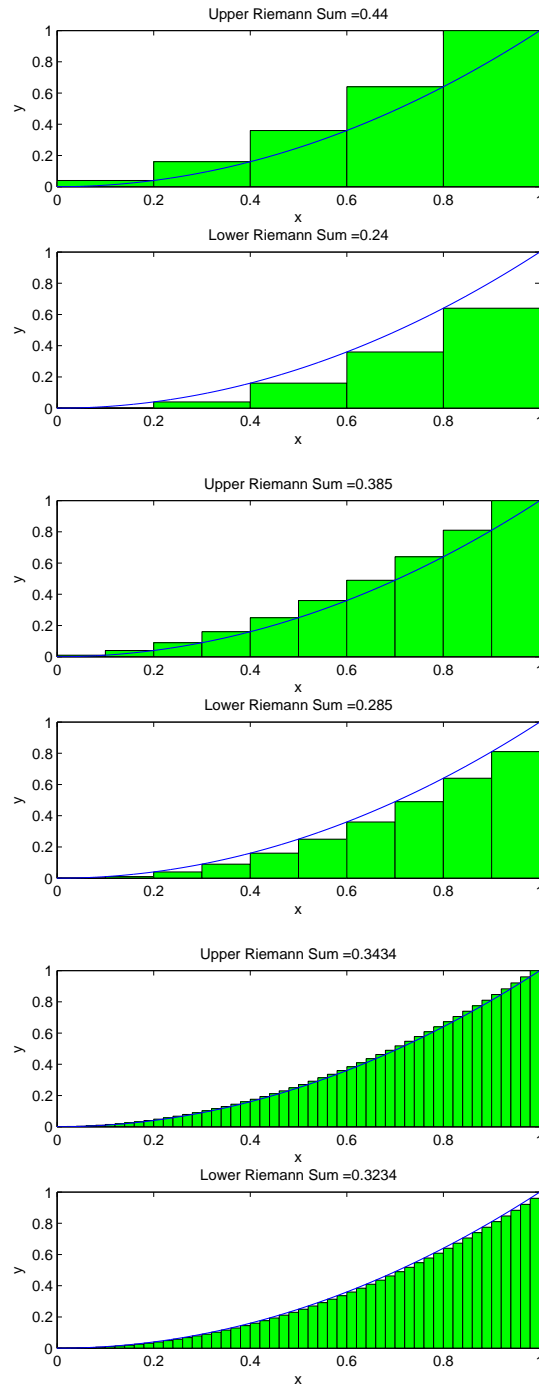
Figure 1. Upper and lower Riemann sums for Example 11.27 with $n = 5, 10, 50$ subintervals of equal length.

and

$$L(f; P_n) = \sum_{k=1}^{n} x_{k-1}^2 \cdot \frac{1}{n} = \frac{1}{n^3} \sum_{k=1}^{n-1} k^2 = \frac{1}{6} \left(1 - \frac{1}{n} \right) \left(2 - \frac{1}{n} \right).$$

(See Figure 11.27.) It follows that

$$\lim_{n \to \infty} U(f; P_n) = \lim_{n \to \infty} L(f; P_n) = \frac{1}{3},$$

and Theorem 11.26 implies that x^2 is integrable on $[0, 1]$ with

$$\int_0^1 x^2 \, dx = \frac{1}{3}.$$

The fundamental theorem of calculus, Theorem 12.1 below, provides a much easier way to evaluate this integral, but the Riemann sums provide the basic definition of the integral.

11.4. Continuous and monotonic functions

The Cauchy criterion leads to the following fundamental result that every continuous function is Riemann integrable. To prove this result, we use the fact that a continuous function oscillates by an arbitrarily small amount on every interval of a sufficiently refined partition.

Theorem 11.28. A continuous function $f : [a, b] \to \mathbb{R}$ on a compact interval is Riemann integrable.

Proof. A continuous function on a compact set is bounded, so we just need to verify the Cauchy condition in Theorem 11.23.

Let $\epsilon > 0$. A continuous function on a compact set is uniformly continuous, so there exists $\delta > 0$ such that

$$|f(x) - f(y)| < \frac{\epsilon}{b - a} \qquad \text{for all } x, y \in [a, b] \text{ such that } |x - y| < \delta.$$

Choose a partition $P = \{I_1, I_2, \dots, I_n\}$ of $[a, b]$ such that $|I_k| < \delta$ for every k; for example, we can take n intervals of equal length $(b - a)/n$ with $n > (b - a)/\delta$.

Since f is continuous, it attains its maximum and minimum values M_k and m_k on the compact interval I_k at points x_k and y_k in I_k. These points satisfy $|x_k - y_k| < \delta$, so

$$M_k - m_k = f(x_k) - f(y_k) < \frac{\epsilon}{b - a}.$$

The upper and lower sums of f therefore satisfy

$$U(f; P) - L(f; P) = \sum_{k=1}^{n} M_k |I_k| - \sum_{k=1}^{n} m_k |I_k|$$

$$= \sum_{k=1}^{n} (M_k - m_k) |I_k|$$

$$< \frac{\epsilon}{b - a} \sum_{k=1}^{n} |I_k|$$

$$< \epsilon,$$

and Theorem 11.23 implies that f is integrable. $\square$

Example 11.29. The function $f(x) = x^2$ on $[0, 1]$ considered in Example 11.27 is integrable since it is continuous.

Another class of integrable functions consists of monotonic (increasing or decreasing) functions.

Theorem 11.30. A monotonic function $f : [a, b] \to \mathbb{R}$ on a compact interval is Riemann integrable.

Proof. Suppose that f is monotonic increasing, meaning that $f(x) \le f(y)$ for $x \le y$. Let $P_n = \{I_1, I_2, \ldots, I_n\}$ be a partition of $[a, b]$ into n intervals $I_k = [x_{k-1}, x_k]$, of equal length $(b - a)/n$, with endpoints

$$x_k = a + (b - a)\frac{k}{n}, \qquad k = 0, 1, \ldots, n.$$

Since f is increasing,

$$M_k = \sup_{I_k} f = f(x_k), \qquad m_k = \inf_{I_k} f = f(x_{k-1}).$$

Hence, summing a telescoping series, we get

$$U(f; P_n) - L(U; P_n) = \sum_{k=1}^{n} (M_k - m_k)(x_k - x_{k-1})$$

$$= \frac{b - a}{n} \sum_{k=1}^{n} [f(x_k) - f(x_{k-1})]$$

$$= \frac{b - a}{n} [f(b) - f(a)].$$

It follows that $U(f; P_n) - L(U; P_n) \to 0$ as $n \to \infty$, and Theorem 11.26 implies that f is integrable.

The proof for a monotonic decreasing function f is similar, with

$$\sup_{I_k} f = f(x_{k-1}), \qquad \inf_{I_k} f = f(x_k),$$

or we can apply the result for increasing functions to $-f$ and use Theorem 11.32 below. $\square$
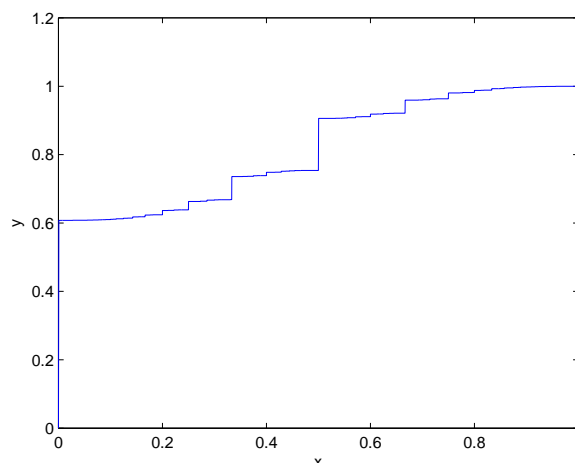
Figure 2. The graph of the monotonic function in Example 11.31 with a countably infinite, dense set of jump discontinuities.

Monotonic functions needn't be continuous, and they may be discontinuous at a countably infinite number of points.

Example 11.31. Let $\{q_k : k \in \mathbb{N}\}$ be an enumeration of the rational numbers in $[0, 1)$ and let (a_k) be a sequence of strictly positive real numbers such that

$$\sum_{k=1}^{\infty} a_k = 1.$$

Define $f : [0, 1] \to \mathbb{R}$ by

$$f(x) = \sum_{k \in Q(x)} a_k, \qquad Q(x) = \{k \in \mathbb{N} : q_k \in [0, x)\}.$$

for $x > 0$, and $f(0) = 0$. That is, $f(x)$ is obtained by summing the terms in the series whose indices k correspond to the rational numbers such that $0 \leq q_k < x$.

For $x = 1$, this sum includes all the terms in the series, so $f(1) = 1$. For every $0 < x < 1$, there are infinitely many terms in the sum, since the rationals are dense in $[0, x)$, and f is increasing, since the number of terms increases with x. By Theorem 11.30, f is Riemann integrable on $[0, 1]$. Although f is integrable, it has a countably infinite number of jump discontinuities at every rational number in $[0, 1)$, which are dense in $[0, 1]$, The function is continuous elsewhere (the proof is left as an exercise).

Figure 2 shows the graph of f corresponding to the enumeration

$$\{0, 1/2, 1/3, 2/3, 1/4, 3/4, 1/5, 2/5, 3/5, 4/5, 1/6, 5/6, 1/7, \dots\}$$

of the rational numbers in $[0, 1)$ and

$$a_k = \frac{6}{\pi^2} \frac{1}{k^2}.$$

What is its Riemann integral?

11.5. Linearity, monotonicity, and additivity

The integral has the following three basic properties.

(1) Linearity:
$$\int_a^b cf = c\int_a^b f, \qquad \int_a^b (f+g) = \int_a^b f + \int_a^b g.$$

(2) Monotonicity: if $f \le g$, then
$$\int_a^b f \le \int_a^b g.$$

(3) Additivity: if $a < c < b$, then
$$\int_a^c f + \int_c^b f = \int_a^b f.$$

These properties are analogous to the corresponding properties of sums (or convergent series):

$$\sum_{k=1}^n ca_k = c\sum_{k=1}^n a_k, \qquad \sum_{k=1}^n (a_k + b_k) = \sum_{k=1}^n a_k + \sum_{k=1}^n b_k;$$

$$\sum_{k=1}^n a_k \le \sum_{k=1}^n b_k \qquad \text{if } a_k \le b_k;$$

$$\sum_{k=1}^m a_k + \sum_{k=m+1}^n a_k = \sum_{k=1}^n a_k.$$

In this section, we prove these properties and derive a few of their consequences.

11.5.1. Linearity. We begin by proving the linearity. First we prove linearity with respect to scalar multiplication and then linearity with respect to sums.

Theorem 11.32. If $f : [a,b] \to \mathbb{R}$ is integrable and $c \in \mathbb{R}$, then cf is integrable and
$$\int_a^b cf = c\int_a^b f.$$

Proof. Suppose that $c \ge 0$. Then for any set $A \subset [a,b]$, we have
$$\sup_A cf = c\sup_A f, \qquad \inf_A cf = c\inf_A f,$$
so $U(cf; P) = cU(f; P)$ for every partition P. Taking the infimum over the set Π of all partitions of $[a,b]$, we get
$$U(cf) = \inf_{P \in \Pi} U(cf; P) = \inf_{P \in \Pi} cU(f; P) = c\inf_{P \in \Pi} U(f; P) = cU(f).$$
Similarly, $L(cf; P) = cL(f; P)$ and $L(cf) = cL(f)$. If f is integrable, then
$$U(cf) = cU(f) = cL(f) = L(cf),$$
which shows that cf is integrable and
$$\int_a^b cf = c\int_a^b f.$$

Now consider $-f$. Since

$$\sup_A(-f) = -\inf_A f, \qquad \inf_A(-f) = -\sup_A f,$$

we have

$$U(-f;P) = -L(f;P), \qquad L(-f;P) = -U(f;P).$$

Therefore

$$U(-f) = \inf_{P\in\Pi} U(-f;P) = \inf_{P\in\Pi} [-L(f;P)] = -\sup_{P\in\Pi} L(f;P) = -L(f),$$
$$L(-f) = \sup_{P\in\Pi} L(-f;P) = \sup_{P\in\Pi} [-U(f;P)] = -\inf_{P\in\Pi} U(f;P) = -U(f).$$

Hence, $-f$ is integrable if f is integrable and

$$\int_a^b (-f) = -\int_a^b f.$$

Finally, if $c < 0$, then $c = -|c|$, and a successive application of the previous results shows that cf is integrable with $\int_a^b cf = c\int_a^b f$. $\square$

Next, we prove the linearity of the integral with respect to sums. If f, g are bounded, then $f + g$ is bounded and

$$\sup_I(f + g) \leq \sup_I f + \sup_I g, \qquad \inf_I(f + g) \geq \inf_I f + \inf_I g.$$

It follows that

$$\operatorname*{osc}_I(f + g) \leq \operatorname*{osc}_I f + \operatorname*{osc}_I g,$$

so $f + g$ is integrable if f, g are integrable. In general, however, the upper (or lower) sum of $f + g$ needn't be the sum of the corresponding upper (or lower) sums of f and g. As a result, we don't get

$$\int_a^b (f + g) = \int_a^b f + \int_a^b g$$

simply by adding upper and lower sums. Instead, we prove this equality by estimating the upper and lower integrals of $f + g$ from above and below by those of f and g.

Theorem 11.33. If $f, g : [a, b] \to \mathbb{R}$ are integrable functions, then $f + g$ is integrable, and

$$\int_a^b (f + g) = \int_a^b f + \int_a^b g.$$

Proof. We first prove that if $f, g : [a, b] \to \mathbb{R}$ are bounded, but not necessarily integrable, then

$$U(f + g) \leq U(f) + U(g), \qquad L(f + g) \geq L(f) + L(g).$$

Suppose that $P = \{I_1, I_2, \ldots, I_n\}$ is a partition of $[a, b]$. Then

$$U(f + g; P) = \sum_{k=1}^{n} \sup_{I_k}(f + g) \cdot |I_k|$$

$$\leq \sum_{k=1}^{n} \sup_{I_k} f \cdot |I_k| + \sum_{k=1}^{n} \sup_{I_k} g \cdot |I_k|$$

$$\leq U(f; P) + U(g; P).$$

Let $\epsilon > 0$. Since the upper integral is the infimum of the upper sums, there are partitions Q, R such that

$$U(f; Q) < U(f) + \frac{\epsilon}{2}, \qquad U(g; R) < U(g) + \frac{\epsilon}{2},$$

and if P is a common refinement of Q and R, then

$$U(f; P) < U(f) + \frac{\epsilon}{2}, \qquad U(g; P) < U(g) + \frac{\epsilon}{2}.$$

It follows that

$$U(f + g) \leq U(f + g; P) \leq U(f; P) + U(g; P) < U(f) + U(g) + \epsilon.$$

Since this inequality holds for arbitrary $\epsilon > 0$, we must have $U(f+g) \leq U(f)+U(g)$.

Similarly, we have $L(f + g; P) \geq L(f; P) + L(g; P)$ for all partitions P, and for every $\epsilon > 0$, we get $L(f + g) > L(f) + L(g) - \epsilon$, so $L(f + g) \geq L(f) + L(g)$.

For integrable functions f and g, it follows that

$$U(f + g) \leq U(f) + U(g) = L(f) + L(g) \leq L(f + g).$$

Since $U(f + g) \geq L(f + g)$, we have $U(f + g) = L(f + g)$ and $f + g$ is integrable. Moreover, there is equality throughout the previous inequality, which proves the result. $\qquad\square$

Although the integral is linear, the upper and lower integrals of non-integrable functions are not, in general, linear.

Example 11.34. Define $f, g : [0, 1] \to \mathbb{R}$ by

$$f(x) = \begin{cases} 1 & \text{if } x \in [0, 1] \cap \mathbb{Q}, \\ 0 & \text{if } x \in [0, 1] \setminus \mathbb{Q}, \end{cases} \qquad g(x) = \begin{cases} 0 & \text{if } x \in [0, 1] \cap \mathbb{Q}, \\ 1 & \text{if } x \in [0, 1] \setminus \mathbb{Q}. \end{cases}$$

That is, f is the Dirichlet function and $g = 1 - f$. Then

$$U(f) = U(g) = 1, \qquad L(f) = L(g) = 0, \qquad U(f + g) = L(f + g) = 1,$$

so

$$U(f + g) < U(f) + U(g), \qquad L(f + g) > L(f) + L(g).$$

The product of integrable functions is also integrable, as is the quotient provided it remains bounded. Unlike the integral of the sum, however, there is no way to express the integral of the product $\int fg$ in terms of $\int f$ and $\int g$.

Theorem 11.35. If $f, g : [a, b] \to \mathbb{R}$ are integrable, then $fg : [a, b] \to \mathbb{R}$ is integrable. If, in addition, $g \neq 0$ and $1/g$ is bounded, then $f/g : [a, b] \to \mathbb{R}$ is integrable.

Proof. First, we show that the square of an integrable function is integrable. If f is integrable, then f is bounded, with $|f| \leq M$ for some $M \geq 0$. For all $x, y \in [a, b]$, we have

$$\left| f^2(x) - f^2(y) \right| = |f(x) + f(y)| \cdot |f(x) - f(y)| \leq 2M|f(x) - f(y)|.$$

Taking the supremum of this inequality over $x, y \in I \subset [a, b]$ and using Proposition 11.8, we get that

$$\sup_I (f^2) - \inf_I (f^2) \leq 2M \left[\sup_I f - \inf_I f \right].$$

meaning that

$$\operatorname*{osc}_I (f^2) \leq 2M \operatorname*{osc}_I f.$$

If follows from Proposition 11.25 that f^2 is integrable if f is integrable.

Since the integral is linear, we then see from the identity

$$fg = \frac{1}{4} \left[(f + g)^2 - (f - g)^2 \right]$$

that fg is integrable if f, g are integrable. We remark that the trick of representing a product as a difference of squares isn't a new one: the ancient Babylonian apparently used this identity, together with a table of squares, to compute products.

In a similar way, if $g \neq 0$ and $|1/g| \leq M$, then

$$\left| \frac{1}{g(x)} - \frac{1}{g(y)} \right| = \frac{|g(x) - g(y)|}{|g(x)g(y)|} \leq M^2 \left| g(x) - g(y) \right|.$$

Taking the supremum of this equation over $x, y \in I \subset [a, b]$, we get

$$\sup_I \left(\frac{1}{g} \right) - \inf_I \left(\frac{1}{g} \right) \leq M^2 \left[\sup_I g - \inf_I g \right],$$

meaning that $\operatorname{osc}_I (1/g) \leq M^2 \operatorname{osc}_I g$, and Proposition 11.25 implies that $1/g$ is integrable if g is integrable. Therefore $f/g = f \cdot (1/g)$ is integrable. $\square$

11.5.2. Monotonicity. Next, we prove the monotonicity of the integral.

Theorem 11.36. Suppose that $f, g : [a, b] \to \mathbb{R}$ are integrable and $f \leq g$. Then

$$\int_a^b f \leq \int_a^b g.$$

Proof. First suppose that $f \geq 0$ is integrable. Let P be the partition consisting of the single interval $[a, b]$. Then

$$L(f; P) = \inf_{[a,b]} f \cdot (b - a) \geq 0,$$

so

$$\int_a^b f \geq L(f; P) \geq 0.$$

If $f \geq g$, then $h = f - g \geq 0$, and the linearity of the integral implies that

$$\int_a^b f - \int_a^b g = \int_a^b h \geq 0,$$

which proves the theorem. $\square$

One immediate consequence of this theorem is the following simple, but useful, estimate for integrals.

Theorem 11.37. Suppose that $f : [a, b] \to \mathbb{R}$ is integrable and

$$M = \sup_{[a,b]} f, \qquad m = \inf_{[a,b]} f.$$

Then

$$m(b - a) \le \int_a^b f \le M(b - a).$$

Proof. Since $m \le f \le M$ on $[a, b]$, Theorem 11.36 implies that

$$\int_a^b m \le \int_a^b f \le \int_a^b M,$$

which gives the result. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

This estimate also follows from the definition of the integral in terms of upper and lower sums, but once we've established the monotonicity of the integral, we don't need to go back to the definition.

A further consequence is the intermediate value theorem for integrals, which states that a continuous function on a compact interval is equal to its average value at some point in the interval.

Theorem 11.38. If $f : [a, b] \to \mathbb{R}$ is continuous, then there exists $x \in [a, b]$ such that

$$f(x) = \frac{1}{b - a} \int_a^b f.$$

Proof. Since f is a continuous function on a compact interval, the extreme value theorem (Theorem 7.37) implies it attains its maximum value M and its minimum value m. From Theorem 11.37,

$$m \le \frac{1}{b - a} \int_a^b f \le M.$$

By the intermediate value theorem (Theorem 7.44), f takes on every value between m and M, and the result follows. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

As shown in the proof of Theorem 11.36, given linearity, monotonicity is equivalent to positivity,

$$\int_a^b f \ge 0 \qquad \text{if } f \ge 0.$$

We remark that even though the upper and lower integrals aren't linear, they are monotone.

Proposition 11.39. If $f, g : [a, b] \to \mathbb{R}$ are bounded functions and $f \le g$, then

$$U(f) \le U(g), \qquad L(f) \le L(g).$$

Proof. From Proposition 11.1, we have for every interval $I \subset [a, b]$ that

$$\sup_I f \leq \sup_I g, \qquad \inf_I f \leq \inf_I g.$$

It follows that for every partition P of $[a, b]$, we have

$$U(f; P) \leq U(g; P), \qquad L(f; P) \leq L(g; P).$$

Taking the infimum of the upper inequality and the supremum of the lower inequality over P, we get that $U(f) \leq U(g)$ and $L(f) \leq L(g)$. $\square$

We can estimate the absolute value of an integral by taking the absolute value under the integral sign. This is analogous to the corresponding property of sums:

$$\left| \sum_{k=1}^{n} a_n \right| \leq \sum_{k=1}^{n} |a_k|.$$

Theorem 11.40. If f is integrable, then $|f|$ is integrable and

$$\left| \int_a^b f \right| \leq \int_a^b |f|.$$

Proof. First, suppose that $|f|$ is integrable. Since

$$-|f| \leq f \leq |f|,$$

we get from Theorem 11.36 that

$$-\int_a^b |f| \leq \int_a^b f \leq \int_a^b |f|, \qquad \text{or} \qquad \left| \int_a^b f \right| \leq \int_a^b |f|.$$

To complete the proof, we need to show that $|f|$ is integrable if f is integrable. For $x, y \in [a, b]$, the reverse triangle inequality gives

$$\big|\, |f(x)| - |f(y)| \,\big| \leq |f(x) - f(y)|.$$

Using Proposition 11.8, we get that

$$\sup_I |f| - \inf_I |f| \leq \sup_I f - \inf_I f,$$

meaning that $\mathrm{osc}_I |f| \leq \mathrm{osc}_I f$. Proposition 11.25 then implies that $|f|$ is integrable if f is integrable. $\square$

In particular, we immediately get the following basic estimate for an integral.

Corollary 11.41. If $f : [a, b] \to \mathbb{R}$ is integrable and $M = \sup_{[a,b]} |f|$, then

$$\left| \int_a^b f \right| \leq M(b - a).$$

Finally, we prove a useful positivity result for the integral of continuous functions.

Proposition 11.42. If $f : [a, b] \to \mathbb{R}$ is a continuous function such that $f \geq 0$ and $\int_a^b f = 0$, then $f = 0$.

Proof. Suppose for contradiction that $f(c) > 0$ for some $a \leq c \leq b$. For definiteness, assume that $a < c < b$. (The proof is similar if c is an endpoint.) Then, since f is continuous, there exists $\delta > 0$ such that

$$|f(x) - f(c)| \leq \frac{f(c)}{2} \qquad \text{for } c - \delta \leq x \leq c + \delta,$$

where we choose δ small enough that $c - \delta > a$ and $c + \delta < b$. It follows that

$$f(x) = f(c) + f(x) - f(c) \geq f(c) - |f(x) - f(c)| \geq \frac{f(c)}{2}$$

for $c - \delta \leq x \leq c + \delta$. Using this inequality and the assumption that $f \geq 0$, we get

$$\int_a^b f = \int_a^{c-\delta} f + \int_{c-\delta}^{c+\delta} f + \int_{c+\delta}^b f \geq 0 + \frac{f(c)}{2} \cdot 2\delta + 0 > 0.$$

This contradiction proves the result. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

The assumption that $f \geq 0$ is, of course, required, otherwise the integral of the function may be zero due to cancelation.

Example 11.43. The function $f : [-1, 1] \to \mathbb{R}$ defined by $f(x) = x$ is continuous and nonzero, but $\int_{-1}^1 f = 0$.

Continuity is also required; for example, the discontinuous function in Example 11.14 is nonzero, but its integral is zero.

11.5.3. Additivity. Finally, we prove additivity. This property refers to additivity with respect to the interval of integration, rather than linearity with respect to the function being integrated.

Theorem 11.44. Suppose that $f : [a, b] \to \mathbb{R}$ and $a < c < b$. Then f is Riemann integrable on $[a, b]$ if and only if it is Riemann integrable on $[a, c]$ and $[c, b]$. Moreover, in that case,

$$\int_a^b f = \int_a^c f + \int_c^b f.$$

Proof. Suppose that f is integrable on $[a, b]$. Then, given $\epsilon > 0$, there is a partition P of $[a, b]$ such that $U(f; P) - L(f; P) < \epsilon$. Let $P' = P \cup \{c\}$ be the refinement of P obtained by adding c to the endpoints of P. (If $c \in P$, then $P' = P$.) Then $P' = Q \cup R$ where $Q = P' \cap [a, c]$ and $R = P' \cap [c, b]$ are partitions of $[a, c]$ and $[c, b]$ respectively. Moreover,

$$U(f; P') = U(f; Q) + U(f; R), \qquad L(f; P') = L(f; Q) + L(f; R).$$

It follows that

$$U(f; Q) - L(f; Q) = U(f; P') - L(f; P') - [U(f; R) - L(f; R)]$$
$$\leq U(f; P) - L(f; P) < \epsilon,$$

which proves that f is integrable on $[a, c]$. Exchanging Q and R, we get the proof for $[c, b]$.

Conversely, if f is integrable on $[a, c]$ and $[c, b]$, then there are partitions Q of $[a, c]$ and R of $[c, b]$ such that

$$U(f; Q) - L(f; Q) < \frac{\epsilon}{2}, \qquad U(f; R) - L(f; R) < \frac{\epsilon}{2}.$$

Let $P = Q \cup R$. Then

$$U(f; P) - L(f; P) = U(f; Q) - L(f; Q) + U(f; R) - L(f; R) < \epsilon,$$

which proves that f is integrable on $[a, b]$.

Finally, if f is integrable, then with the partitions P, Q, R as above, we have

$$\begin{aligned}
\int_a^b f &\leq U(f; P) = U(f; Q) + U(f; R) \\
&< L(f; Q) + L(f; R) + \epsilon \\
&< \int_a^c f + \int_c^b f + \epsilon.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\int_a^b f &\geq L(f; P) = L(f; Q) + L(f; R) \\
&> U(f; Q) + U(f; R) - \epsilon \\
&> \int_a^c f + \int_c^b f - \epsilon.
\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, we see that $\int_a^b f = \int_a^c f + \int_c^b f$. $\square$

We can extend the additivity property of the integral by defining an oriented Riemann integral.

Definition 11.45. If $f : [a, b] \to \mathbb{R}$ is integrable, where $a < b$, and $a \leq c \leq b$, then

$$\int_b^a f = -\int_a^b f, \qquad \int_c^c f = 0.$$

With this definition, the additivity property in Theorem 11.44 holds for all $a, b, c \in \mathbb{R}$ for which the oriented integrals exist. Moreover, if $|f| \leq M$, then the estimate in Corollary 11.41 becomes

$$\left| \int_a^b f \right| \leq M|b - a|$$

for all $a, b \in \mathbb{R}$ (even if $a \geq b$).

The oriented Riemann integral is a special case of the integral of a differential form. It assigns a value to the integral of a one-form $f\, dx$ on an oriented interval.

11.6. Further existence results

In this section, we prove several further useful conditions for the existences of the Riemann integral.

First, we show that changing the values of a function at finitely many points doesn't change its integrability of the value of its integral.

Proposition 11.46. Suppose that $f, g : [a, b] \to \mathbb{R}$ and $f(x) = g(x)$ except at finitely many points $x \in [a, b]$. Then f is integrable if and only if g is integrable, and in that case

$$\int_a^b f = \int_a^b g.$$

Proof. It is sufficient to prove the result for functions whose values differ at a single point, say $c \in [a, b]$. The general result then follows by repeated application of this result.

Since f, g differ at a single point, f is bounded if and only if g is bounded. If f, g are unbounded, then neither one is integrable. If f, g are bounded, we will show that f, g have the same upper and lower integrals. The reason is that their upper and lower sums differ by an arbitrarily small amount with respect to a partition that is sufficiently refined near the point where the functions differ.

Suppose that f, g are bounded with $|f|, |g| \leq M$ on $[a, b]$ for some $M > 0$. Let $\epsilon > 0$. Choose a partition P of $[a, b]$ such that

$$U(f; P) < U(f) + \frac{\epsilon}{2}.$$

Let $Q = \{I_1, \ldots, I_n\}$ be a refinement of P such that $|I_k| < \delta$ for $k = 1, \ldots, n$, where

$$\delta = \frac{\epsilon}{8M}.$$

Then g differs from f on at most two intervals in Q. (This could happen on two intervals if c is an endpoint of the partition.) On such an interval I_k we have

$$\left| \sup_{I_k} g - \sup_{I_k} f \right| \leq \sup_{I_k} |g| + \sup_{I_k} |f| \leq 2M,$$

and on the remaining intervals, $\sup_{I_k} g - \sup_{I_k} f = 0$. It follows that

$$|U(g; Q) - U(f; Q)| < 2M \cdot 2\delta < \frac{\epsilon}{2}.$$

Using the properties of upper integrals and refinements, we obtain that

$$U(g) \leq U(g; Q) < U(f; Q) + \frac{\epsilon}{2} \leq U(f; P) + \frac{\epsilon}{2} < U(f) + \epsilon.$$

Since this inequality holds for arbitrary $\epsilon > 0$, we get that $U(g) \leq U(f)$. Exchanging f and g, we see similarly that $U(f) \leq U(g)$, so $U(f) = U(g)$.

An analogous argument for lower sums (or an application of the result for upper sums to $-f$, $-g$) shows that $L(f) = L(g)$. Thus $U(f) = L(f)$ if and only if $U(g) = L(g)$, in which case $\int_a^b f = \int_a^b g$. $\qquad \square$

Example 11.47. The function f in Example 11.14 differs from the 0-function at one point. It is integrable and its integral is equal to 0.

The conclusion of Proposition 11.46 can fail if the functions differ at a countably infinite number of points. One reason is that we can turn a bounded function into an unbounded function by changing its values at an countably infinite number of points.

Example 11.48. Define $f : [0, 1] \to \mathbb{R}$ by

$$f(x) = \begin{cases} n & \text{if } x = 1/n \text{ for } n \in \mathbb{N}, \\ 0 & \text{otherwise.} \end{cases}$$

Then f is equal to the 0-function except on the countably infinite set $\{1/n : n \in \mathbb{N}\}$, but f is unbounded and therefore it's not Riemann integrable.

The result in Proposition 11.46 is still false, however, for bounded functions that differ at a countably infinite number of points.

Example 11.49. The Dirichlet function in Example 11.15 is bounded and differs from the 0-function on the countably infinite set of rationals, but it isn't Riemann integrable.

The Lebesgue integral is better behaved than the Riemann intgeral in this respect: two functions that are equal almost everywhere, meaning that they differ on a set of Lebesgue measure zero, have the same Lebesgue integrals. In particular, two functions that differ on a countable set have the same Lebesgue integrals (see Section 11.8).

The next proposition allows us to deduce the integrability of a bounded function on an interval from its integrability on slightly smaller intervals.

Proposition 11.50. Suppose that $f : [a, b] \to \mathbb{R}$ is bounded and integrable on $[a, r]$ for every $a < r < b$. Then f is integrable on $[a, b]$ and

$$\int_a^b f = \lim_{r \to b^-} \int_a^r f.$$

Proof. Since f is bounded, $|f| \leq M$ on $[a, b]$ for some $M > 0$. Given $\epsilon > 0$, let

$$r = b - \frac{\epsilon}{4M}$$

(where we assume ϵ is sufficiently small that $r > a$). Since f is integrable on $[a, r]$, there is a partition Q of $[a, r]$ such that

$$U(f; Q) - L(f; Q) < \frac{\epsilon}{2}.$$

Then $P = Q \cup \{b\}$ is a partition of $[a, b]$ whose last interval is $[r, b]$. The boundedness of f implies that

$$\sup_{[r,b]} f - \inf_{[r,b]} f \leq 2M.$$

Therefore

$$U(f; P) - L(f; P) = U(f; Q) - L(f; Q) + \left(\sup_{[r,b]} f - \inf_{[r,b]} f \right) \cdot (b - r)$$

$$< \frac{\epsilon}{2} + 2M \cdot (b - r) = \epsilon,$$

so f is integrable on $[a, b]$ by Theorem 11.23. Moreover, using the additivity of the integral, we get

$$\left| \int_a^b f - \int_a^r f \right| = \left| \int_r^b f \right| \leq M \cdot (b - r) \to 0 \qquad \text{as } r \to b^-.$$

$\square$

An obvious analogous result holds for the left endpoint.

Example 11.51. Define $f : [0, 1] \to \mathbb{R}$ by

$$f(x) = \begin{cases} \sin(1/x) & \text{if } 0 < x \leq 1, \\ 0 & \text{if } x = 0. \end{cases}$$

Then f is bounded on $[0, 1]$. Furthemore, f is continuous and therefore integrable on $[r, 1]$ for every $0 < r < 1$. It follows from Proposition 11.50 that f is integrable on $[0, 1]$.

The assumption in Proposition 11.50 that f is bounded on $[a, b]$ is essential.

Example 11.52. The function $f : [0, 1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1/x & \text{for } 0 < x \leq 1, \\ 0 & \text{for } x = 0, \end{cases}$$

is continuous and therefore integrable on $[r, 1]$ for every $0 < r < 1$, but it's unbounded and therefore not integrable on $[0, 1]$.

As a corollary of this result and the additivity of the integral, we prove a generalization of the integrability of continuous functions to piecewise continuous functions.

Theorem 11.53. If $f : [a, b] \to \mathbb{R}$ is a bounded function with finitely many discontinuities, then f is Riemann integrable.

Proof. By splitting the interval into subintervals with the discontinuities of f at an endpoint and using Theorem 11.44, we see that it is sufficient to prove the result if f is discontinuous only at one endpoint of $[a, b]$, say at b. In that case, f is continuous and therefore integrable on any smaller interval $[a, r]$ with $a < r < b$, and Proposition 11.50 implies that f is integrable on $[a, b]$. $\square$

Example 11.54. Define $f : [0, 2\pi] \to \mathbb{R}$ by

$$f(x) = \begin{cases} \sin\left(1/\sin x\right) & \text{if } x \neq 0, \pi, 2\pi, \\ 0 & \text{if } x = 0, \pi, 2\pi. \end{cases}$$

Then f is bounded and continuous except at $x = 0, \pi, 2\pi$, so it is integrable on $[0, 2\pi]$ (see Figure 3). This function doesn't have jump discontinuities, but Theorem 11.53 still applies.
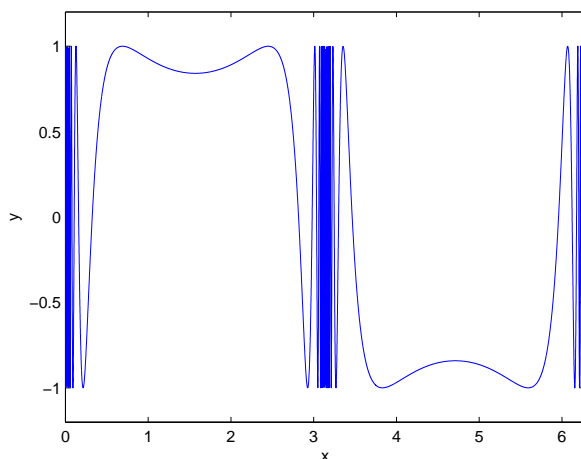
Figure 3. Graph of the Riemann integrable function $y = \sin(1/\sin x)$ in Example 11.54.
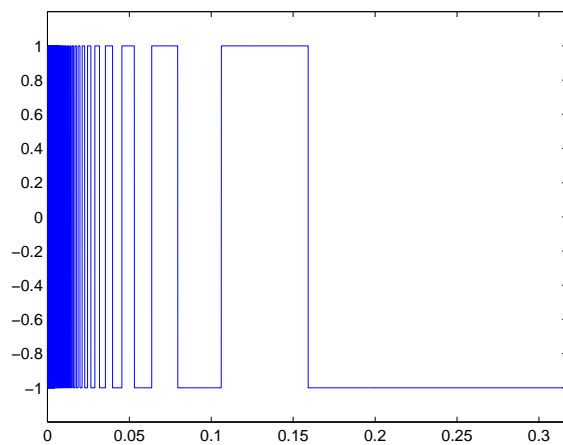
Figure 4. Graph of the Riemann integrable function $y = \text{sgn}(\sin(1/x))$ in Example 11.55.

Example 11.55. Define $f : [0, 1/\pi] \to \mathbb{R}$ by

$$f(x) = \begin{cases} \text{sgn}\left[\sin\left(1/x\right)\right] & \text{if } x \neq 1/n\pi \text{ for } n \in \mathbb{N}, \\ 0 & \text{if } x = 0 \text{ or } x \neq 1/n\pi \text{ for } n \in \mathbb{N}, \end{cases}$$

where sgn is the sign function,

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

Then f oscillates between 1 and -1 a countably infinite number of times as $x \to 0^+$ (see Figure 4). It has jump discontinuities at $x = 1/(n\pi)$ and an essential discontinuity at $x = 0$. Nevertheless, it is Riemann integrable. To see this, note that f is bounded on $[0, 1]$ and piecewise continuous with finitely many discontinuities on $[r, 1]$ for every $0 < r < 1$. Theorem 11.53 implies that f is Riemann integrable on $[r, 1]$, and then Theorem 11.50 implies that f is integrable on $[0, 1]$.

11.7. * Riemann sums

Instead of using upper and lower sums, we can give an equivalent definition of the Riemann integral as a limit of Riemann sums. This was, in fact, Riemann's original definition [**11**], which he gave in 1854 in his *Habilitationsschrift* (a kind of post-doctoral dissertation required of German academics), building on previous work of Cauchy who defined the integral for continuous functions.

It is interesting to note that the topic of Riemann's *Habilitationsschrift* was not integration theory, but Fourier series. Riemann introduced a definition of the integral along the way so that he could state his results more precisely. Many of the fundamental developments of rigorous real analysis in the nineteenth century were motivated by problems related to Fourier series and their convergence.

Upper and lower sums were introduced subsequently by Darboux, and they simplify the theory. We won't use Riemann sums here, but we will explain the equivalence of the definitions. We'll say, temporarily, that a function is Darboux integrable if it satisfies Definition 11.11.

To give Riemann's definition, we first define a tagged partition (P, C) of a compact interval $[a, b]$ to be a partition

$$P = \{I_1, I_2, \ldots, I_n\}$$

of the interval together with a set

$$C = \{c_1, c_2, \ldots, c_n\}$$

of points such that $c_k \in I_k$ for $k = 1, \ldots, n$. (We think of the point c_k as a "tag" attached to the interval I_k.)

If $f : [a, b] \to \mathbb{R}$, then we define the Riemann sum of f with respect to the tagged partition (P, C) by

$$S(f; P, C) = \sum_{k=1}^{n} f(c_k)|I_k|.$$

That is, instead of using the supremum or infimum of f on the kth interval in the sum, we evaluate f at a point in the interval. Roughly speaking, a function is Riemann integrable if its Riemann sums approach the same value as the partition is refined, independently of how we choose the points $c_k \in I_k$.

As a measure of the refinement of a partition $P = \{I_1, I_2, \ldots, I_n\}$, we define the mesh (or norm) of P to be the maximum length of its intervals,

$$\text{mesh}(P) = \max_{1 \le k \le n} |I_k| = \max_{1 \le k \le n} |x_k - x_{k-1}|.$$

Definition 11.56. A function $f : [a, b] \to \mathbb{R}$ is Riemann integrable on $[a, b]$ if there exists a number $R \in \mathbb{R}$ with the following property: For every $\epsilon > 0$ there is a $\delta > 0$ such that

$$|S(f; P, C) - R| < \epsilon$$

for every tagged partition (P, C) of $[a, b]$ with $\text{mesh}(P) < \delta$. In that case, $R = \int_a^b f$ is the Riemann integral of f on $[a, b]$.

Note that

$$L(f; P) \leq S(f; P, C) \leq U(f; P),$$

so the Riemann sums are "squeezed" between the upper and lower sums. The following theorem shows that the Darboux and Riemann definitions lead to the same notion of the integral, so it's a matter of convenience which definition we adopt as our starting point.

Theorem 11.57. A function is Riemann integrable (in the sense of Definition 11.56) if and only if it is Darboux integrable (in the sense of Definition 11.11). Furthermore, in that case, the Riemann and Darboux integrals of the function are equal.

Proof. First, suppose that $f : [a, b] \to \mathbb{R}$ is Riemann integrable with integral R. Then f is bounded on $[a, b]$; otherwise, it would be unbounded in some interval I_k of every partition P, and its Riemann sums with respect to P would be arbitrarily large for suitable points $c_k \in I_k$, so no $R \in \mathbb{R}$ could satisfy Definition 11.56.

Let $\epsilon > 0$. Since f is Riemann integrable, there is a partition $P = \{I_1, I_2, \ldots, I_n\}$ of $[a, b]$ such that

$$|S(f; P, C) - R| < \frac{\epsilon}{2}$$

for every set of points $C = \{c_k \in I_k : k = 1, \ldots, n\}$. If $M_k = \sup_{I_k} f$, then there exists $c_k \in I_k$ such that

$$M_k - \frac{\epsilon}{2(b-a)} < f(c_k).$$

It follows that

$$\sum_{k=1}^n M_k |I_k| - \frac{\epsilon}{2} < \sum_{k=1}^n f(c_k)|I_k|,$$

meaning that $U(f; P) - \epsilon/2 < S(f; P, C)$. Since $S(f; P, C) < R + \epsilon/2$, we get that

$$U(f) \leq U(f; P) < R + \epsilon.$$

Similarly, if $m_k = \inf_{I_k} f$, then there exists $c_k \in I_k$ such that

$$m_k + \frac{\epsilon}{2(b-a)} > f(c_k), \qquad \sum_{k=1}^n m_k |I_k| + \frac{\epsilon}{2} > \sum_{k=1}^n f(c_k)|I_k|,$$

and $L(f; P) + \epsilon/2 > S(f; P, C)$. Since $S(f; P, C) > R - \epsilon/2$, we get that

$$L(f) \geq L(f; P) > R - \epsilon.$$

These inequalities imply that

$$L(f) + \epsilon > R > U(f) - \epsilon$$

for every $\epsilon > 0$, and therefore $L(f) \geq R \geq U(f)$. Since $L(f) \leq U(f)$, we conclude that $L(f) = R = U(f)$, so f is Darboux integrable with integral R.

Conversely, suppose that f is Darboux integrable. The main point is to show that if $\epsilon > 0$, then $U(f; P) - L(f; P) < \epsilon$ not just for some partition but for every partition whose mesh is sufficiently small.

Let $\epsilon > 0$ be given. Since f is Darboux integrable. there exists a partition Q such that

$$U(f; Q) - L(f; Q) < \frac{\epsilon}{4}.$$

Suppose that Q contains m intervals and $|f| \leq M$ on $[a, b]$. We claim that if

$$\delta = \frac{\epsilon}{8mM},$$

then $U(f; P) - L(f; P) < \epsilon$ for every partition P with mesh$(P) < \delta$.

To prove this claim, suppose that $P = \{I_1, I_2, \ldots, I_n\}$ is a partition with mesh$(P) < \delta$. Let P' be the largest common refinement of P and Q, so that the endpoints of P' consist of the endpoints of P or Q. Since a, b are common endpoints of P and Q, there are at most $m - 1$ endpoints of Q that are distinct from endpoints of P. Therefore, at most $m - 1$ intervals in P contain additional endpoints of Q and are strictly refined in P', meaning that they are the union of two or more intervals in P'.

Now consider $U(f; P) - U(f; P')$. The terms that correspond to the same, unrefined intervals in P and P' cancel. If I_k is a strictly refined interval in P, then the corresponding terms in each of the sums $U(f; P)$ and $U(f; P')$ can be estimated by $M|I_k|$ and their difference by $2M|I_k|$. There are at most $m - 1$ such intervals and $|I_k| < \delta$, so it follows that

$$U(f; P) - U(f; P') < 2(m - 1)M\delta < \frac{\epsilon}{4}.$$

Since P' is a refinement of Q, we get

$$U(f; P) < U(f; P') + \frac{\epsilon}{4} \leq U(f; Q) + \frac{\epsilon}{4} < L(f; Q) + \frac{\epsilon}{2}.$$

It follows by a similar argument that

$$L(f; P') - L(f; P) < \frac{\epsilon}{4},$$

and

$$L(f; P) > L(f; P') - \frac{\epsilon}{4} \geq L(f; Q) - \frac{\epsilon}{4} > U(f; Q) - \frac{\epsilon}{2}.$$

Since $L(f; Q) \leq U(f; Q)$, we conclude from these inequalities that

$$U(f; P) - L(f; P) < \epsilon$$

for every partition P with mesh$(P) < \delta$.

If D denotes the Darboux integral of f, then we have

$$L(f; P) \leq D \leq U(f, P), \qquad L(f; P) \leq S(f; P, C) \leq U(f; P).$$

Since $U(f; P) - L(f; P) < \epsilon$ for every partition P with mesh$(P) < \delta$, it follows that

$$|S(f; P, C) - D| < \epsilon.$$

Thus, f is Riemann integrable with Riemann integral D. $\qquad\qquad\square$

Finally, we give a necessary and sufficient condition for Riemann integrability that was proved by Riemann himself (1854). (See [**5**] for further discussion.) To state the condition, we introduce some notation.

Let $f; [a, b] \to \mathbb{R}$ be a bounded function. If $P = \{I_1, I_2, \ldots, I_n\}$ is a partition of $[a, b]$ and $\epsilon > 0$, let $A_\epsilon(P) \subset \{1, \ldots, n\}$ be the set of indices k such that

$$\underset{I_k}{\operatorname{osc}} f = \sup_{I_k} f - \inf_{I_k} f \geq \epsilon \qquad \text{for } k \in A_\epsilon(P).$$

Similarly, let $B_\epsilon(P) \subset \{1, \ldots, n\}$ be the set of indices such that

$$\underset{I_k}{\operatorname{osc}} f < \epsilon \qquad \text{for } k \in B_\epsilon(P).$$

That is, the oscillation of f on I_k is "large" if $k \in A_\epsilon(P)$ and "small" if $k \in B_\epsilon(P)$. We denote the sum of the lengths of the intervals in P where the oscillation of f is "large" by

$$s_\epsilon(P) = \sum_{k \in A_\epsilon(P)} |I_k|.$$

Fixing $\epsilon > 0$, we say that $s_\epsilon(P) \to 0$ as $\operatorname{mesh}(P) \to 0$ if for every $\eta > 0$ there exists $\delta > 0$ such that $\operatorname{mesh}(P) < \delta$ implies that $s_\epsilon(P) < \eta$.

Theorem 11.58. A function is Riemann integrable if and only if $s_\epsilon(P) \to 0$ as $\operatorname{mesh}(P) \to 0$ for every $\epsilon > 0$.

Proof. Let $f : [a, b] \to \mathbb{R}$ be Riemann integrable with $|f| \leq M$ on $[a, b]$.

First, suppose that the condition holds, and let $\epsilon > 0$. If P is a partition of $[a, b]$, then, using the notation above for $A_\epsilon(P)$, $B_\epsilon(P)$ and the inequality

$$0 \leq \underset{I_k}{\operatorname{osc}} f \leq 2M,$$

we get that

$$U(f; P) - L(f; P) = \sum_{k=1}^{n} \underset{I_k}{\operatorname{osc}} f \cdot |I_k|$$

$$= \sum_{k \in A_\epsilon(P)} \underset{I_k}{\operatorname{osc}} f \cdot |I_k| + \sum_{k \in B_\epsilon(P)} \underset{I_k}{\operatorname{osc}} f \cdot |I_k|$$

$$\leq 2M \sum_{k \in A_\epsilon(P)} |I_k| + \epsilon \sum_{k \in B_\epsilon(P)} |I_k|$$

$$\leq 2M s_\epsilon(P) + \epsilon(b - a).$$

By assumption, there exists $\delta > 0$ such that $s_\epsilon(P) < \epsilon$ if $\operatorname{mesh}(P) < \delta$, in which case

$$U(f; P) - L(f; P) < \epsilon(2M + b - a).$$

The Cauchy criterion in Theorem 11.23 then implies that f is integrable.

Conversely, suppose that f is integrable, and let $\epsilon > 0$ be given. If P is a partition, we can bound $s_\epsilon(P)$ from above by the difference between the upper and lower sums as follows:

$$U(f; P) - L(f; P) \geq \sum_{k \in A_\epsilon(P)} \underset{I_k}{\operatorname{osc}} f \cdot |I_k| \geq \epsilon \sum_{k \in A_\epsilon(P)} |I_k| = \epsilon s_\epsilon(P).$$

Since f is integrable, for every $\eta > 0$ there exists $\delta > 0$ such that $\text{mesh}(P) < \delta$ implies that

$$U(f; P) - L(f; P) < \epsilon\eta.$$

Therefore, $\text{mesh}(P) < \delta$ implies that

$$s_\epsilon(P) \leq \frac{1}{\epsilon} \left[U(f; P) - L(f; P)\right] < \eta,$$

which proves the result. $\square$

This theorem has the drawback that the necessary and sufficient condition for Riemann integrability is somewhat complicated and, in general, isn't easy to verify. In the next section, we state a simpler necessary and sufficient condition for Riemann integrability.

11.8. * The Lebesgue criterion

Although the Dirichlet function in Example 11.15 is not Riemann integrable, it is Lebesgue integrable. Its Lebesgue integral is given by

$$\int_0^1 f = 1 \cdot |A| + 0 \cdot |B|$$

where $A = [0, 1] \cap \mathbb{Q}$ is the set of rational numbers in $[0, 1]$, $B = [0, 1] \setminus \mathbb{Q}$ is the set of irrational numbers, and $|E|$ denotes the Lebesgue measure of a set E. The Lebesgue measure of a subset of $\mathbb{R}$ is a generalization of the length of an interval which applies to more general sets. It turns out that $|A| = 0$ (as is true for any countable set of real numbers — see Example 11.60 below) and $|B| = 1$. Thus, the Lebesgue integral of the Dirichlet function is 0.

A necessary and sufficient condition for Riemann integrability can be given in terms of Lebesgue measure. To state this condition, we first define what it means for a set to have Lebesgue measure zero.

Definition 11.59. A set $E \subset \mathbb{R}$ has Lebesgue measure zero if for every $\epsilon > 0$ there is a countable collection of open intervals $\{(a_k, b_k) : k \in \mathbb{N}\}$ such that

$$E \subset \bigcup_{k=1}^\infty (a_k, b_k), \qquad \sum_{k=1}^\infty (b_k - a_k) < \epsilon.$$

The open intervals is this definition are not required to be disjoint, and they may "overlap."

Example 11.60. Every countable set $E = \{x_k \in \mathbb{R} : k \in \mathbb{N}\}$ has Lebesgue measure zero. To prove this, let $\epsilon > 0$ and for each $k \in \mathbb{N}$ define

$$a_k = x_k - \frac{\epsilon}{2^{k+2}}, \qquad b_k = x_k + \frac{\epsilon}{2^{k+2}}.$$

Then $E \subset \bigcup_{k=1}^\infty (a_k, b_k)$ since $x_k \in (a_k, b_k)$ and

$$\sum_{k=1}^\infty (b_k - a_k) = \sum_{k=1}^\infty \frac{\epsilon}{2^{k+1}} = \frac{\epsilon}{2} < \epsilon,$$

so the Lebesgue measure of E is equal to zero. (The '$\epsilon/2^k$' trick used here is a common one in measure theory.)

If $E = [0,1] \cap \mathbb{Q}$ consists of the rational numbers in $[0,1]$, then the set $G = \bigcup_{k=1}^{\infty}(a_k, b_k)$ described above encloses the dense set of rationals in a collection of open intervals the sum of whose lengths is arbitrarily small. This set isn't so easy to visualize. Roughly speaking, if ϵ is small and we look at a section of $[0,1]$ at a given magnification, then we see a few of the longer intervals in G with relatively large gaps between them. Magnifying one of these gaps, we see a few more intervals with large gaps between them, magnifying those gaps, we see a few more intervals, and so on. Thus, the set G has a fractal structure, meaning that it looks similar at all scales of magnification.

In general, we have the following result, due to Lebesgue, which we state without proof.

Theorem 11.61. A function $f : [a,b] \to \mathbb{R}$ is Riemann integrable if and only if it is bounded and the set of points at which it is discontinuous has Lebesgue measure zero.

For example, the set of discontinuities of the Riemann-integrable function in Example 11.14 consists of a single point $\{0\}$, which has Lebesgue measure zero. On the other hand, the set of discontinuities of the non-Riemann-integrable Dirichlet function in Example 11.15 is the entire interval $[0,1]$, and its set of discontinuities has Lebesgue measure one.

In particular, every bounded function with a countable set of discontinuities is Riemann integrable, since such a set has Lebesgue measure zero. Riemann integrability of a function does not, however, imply that the function has only countably many discontinuities.

Example 11.62. The Cantor set C in Example 5.64 has Lebesgue measure zero. To prove this, using the same notation as in Section 5.5, we note that for every $n \in \mathbb{N}$ the set $F_n \supset C$ consists of 2^n closed intervals $I_{\mathbf{s}}$ of length $|I_{\mathbf{s}}| = 3^{-n}$. For every $\epsilon > 0$ and $\mathbf{s} \in \Sigma_n$, there is an open interval $U_{\mathbf{s}}$ of slightly larger length $|U_{\mathbf{s}}| = 3^{-n} + \epsilon 2^{-n}$ that contains $I_{\mathbf{s}}$. Then $\{U_{\mathbf{s}} : \mathbf{s} \in \Sigma_n\}$ is a cover of C by open intervals, and

$$\sum_{s \in \Sigma_n} |U_{\mathbf{s}}| = \left(\frac{2}{3}\right)^n + \epsilon.$$

We can make the right-hand side as small as we wish by choosing n large enough and ϵ small enough, so C has Lebesgue measure zero.

Let $\chi_C : [0,1] \to \mathbb{R}$ be the characteristic function of the Cantor set,

$$\chi_C(x) = \begin{cases} 1 & \text{if } x \in C, \\ 0 & \text{otherwise.} \end{cases}$$

By partitioning $[0,1]$ into the closed intervals $\{\bar{U}_{\mathbf{s}} : \mathbf{s} \in \Sigma_n\}$ and the closures of the complementary intervals, we see similarly that the upper Riemann sums of χ_C can be made arbitrarily small, so χ_C is Riemann integrable on $[0,1]$ with zero integral. The Riemann integrability of the function χ_C also follows from Theorem 11.61.

It is, however, discontinuous at every point of C. Thus, χ_C is an example of a Riemann integrable function with uncountably many discontinuities.

Properties and Applications of the Integral

In the integral calculus I find much less interesting the parts that involve only substitutions, transformations, and the like, in short, the parts that involve the known skillfully applied mechanics of reducing integrals to algebraic, logarithmic, and circular functions, than I find the careful and profound study of transcendental functions that cannot be reduced to these functions. (Gauss, 1808)

12.1. The fundamental theorem of calculus

The fundamental theorem of calculus states that differentiation and integration are inverse operations in an appropriately understood sense. The theorem has two parts: in one direction, it says roughly that the integral of the derivative is the original function; in the other direction, it says that the derivative of the integral is the original function.

In more detail, the first part states that if $F : [a, b] \to \mathbb{R}$ is differentiable with integrable derivative, then

$$\int_a^b F'(x)\, dx = F(b) - F(a).$$

This result can be thought of as a continuous analog of the corresponding identity for sums of differences,

$$\sum_{k=1}^n (A_k - A_{k-1}) = A_n - A_0.$$

The second part states that if $f : [a, b] \to \mathbb{R}$ is continuous, then

$$\frac{d}{dx} \int_a^x f(t)\, dt = f(x).$$

This is a continuous analog of the corresponding identity for differences of sums,

$$\sum_{j=1}^{k} a_j - \sum_{j=1}^{k-1} a_j = a_k.$$

The proof of the fundamental theorem consists essentially of applying the identities for sums or differences to the appropriate Riemann sums or difference quotients and proving, under appropriate hypotheses, that they converge to the corresponding integrals or derivatives.

We'll split the statement and proof of the fundamental theorem into two parts. (The numbering of the parts as I and II is arbitrary.)

12.1.1. Fundamental theorem I. First we prove the statement about the integral of a derivative.

Theorem 12.1 (Fundamental theorem of calculus I). *If* $F : [a, b] \to \mathbb{R}$ *is continuous on* $[a, b]$ *and differentiable in* (a, b) *with* $F' = f$ *where* $f : [a, b] \to \mathbb{R}$ *is Riemann integrable, then*

$$\int_a^b f(x)\, dx = F(b) - F(a).$$

Proof. Let

$$P = \{x_0, x_1, x_2, \ldots, x_{n-1}, x_n\},$$

be a partition of $[a, b]$, with $x_0 = a$ and $x_n = b$. Then

$$F(b) - F(a) = \sum_{k=1}^{n} \left[F(x_k) - F(x_{k-1}) \right].$$

The function F is continuous on the closed interval $[x_{k-1}, x_k]$ and differentiable in the open interval (x_{k-1}, x_k) with $F' = f$. By the mean value theorem, there exists $x_{k-1} < c_k < x_k$ such that

$$F(x_k) - F(x_{k-1}) = f(c_k)(x_k - x_{k-1}).$$

Since f is Riemann integrable, it is bounded, and

$$m_k(x_k - x_{k-1}) \le F(x_k) - F(x_{k-1}) \le M_k(x_k - x_{k-1}),$$

where

$$M_k = \sup_{[x_{k-1}, x_k]} f, \qquad m_k = \inf_{[x_{k-1}, x_k]} f.$$

Hence, $L(f; P) \le F(b) - F(a) \le U(f; P)$ for every partition P of $[a, b]$, which implies that $L(f) \le F(b) - F(a) \le U(f)$. Since f is integrable, $L(f) = U(f) = \int_a^b f$ and therefore $F(b) - F(a) = \int_a^b f$. $\qquad\square$

In Theorem 12.1, we assume that F is continuous on the closed interval $[a, b]$ and differentiable in the open interval (a, b) where its usual two-sided derivative is defined and is equal to f. It isn't necessary to assume the existence of the right derivative of F at a or the left derivative at b, so the values of f at the endpoints are not necessarily determined by F. By Proposition 11.46, however, the integrability of f on $[a, b]$ and the value of its integral do not depend on these values, so the statement of the theorem makes sense. As a result, we'll sometimes

abuse terminology and say that "F' is integrable on $[a, b]$" even if it's only defined on (a, b).

Theorem 12.1 imposes the integrability of F' as a hypothesis. Every function F that is continuously differentiable on the closed interval $[a, b]$ satisfies this condition, but the theorem remains true even if F' is a discontinuous, Riemann integrable function.

Example 12.2. Define $F : [0, 1] \to \mathbb{R}$ by

$$F(x) = \begin{cases} x^2 \sin(1/x) & \text{if } 0 < x \leq 1, \\ 0 & \text{if } x = 0. \end{cases}$$

Then F is continuous on $[0, 1]$ and, by the product and chain rules, differentiable in $(0, 1]$. It is also differentiable — but not continuously differentiable — at 0, with $F'(0^+) = 0$. Thus,

$$F'(x) = \begin{cases} -\cos(1/x) + 2x \sin(1/x) & \text{if } 0 < x \leq 1, \\ 0 & \text{if } x = 0. \end{cases}$$

The derivative F' is bounded on $[0, 1]$ and discontinuous only at one point $(x = 0)$, so Theorem 11.53 implies that F' is integrable on $[0, 1]$. This verifies all of the hypotheses in Theorem 12.1, and we conclude that

$$\int_0^1 F'(x) \, dx = \sin 1.$$

There are, however, differentiable functions whose derivatives are unbounded or so discontinuous that they aren't Riemann integrable.

Example 12.3. Define $F : [0, 1] \to \mathbb{R}$ by $F(x) = \sqrt{x}$. Then F is continuous on $[0, 1]$ and differentiable in $(0, 1]$, with

$$F'(x) = \frac{1}{2\sqrt{x}} \qquad \text{for } 0 < x \leq 1.$$

This function is unbounded, so F' is not Riemann integrable on $[0, 1]$, however we define its value at 0, and Theorem 12.1 does not apply.

We can interpret the integral of F' on $[0, 1]$ as an improper Riemann integral (as is discussed further in Section 12.4). The function F is continuously differentiable on $[\epsilon, 1]$ for every $0 < \epsilon < 1$, so

$$\int_\epsilon^1 \frac{1}{2\sqrt{x}} \, dx = 1 - \sqrt{\epsilon}.$$

Thus, we get the improper integral

$$\lim_{\epsilon \to 0^+} \int_\epsilon^1 \frac{1}{2\sqrt{x}} \, dx = 1.$$

The construction of a function with a bounded, non-integrable derivative is more involved. It's not sufficient to give a function with a bounded derivative that is discontinuous at finitely many points, as in Example 12.2, because such a function is Riemann integrable. Rather, one has to construct a differentiable function whose

derivative is discontinuous on a set of nonzero Lebesgue measure. Abbott [**1**] gives an example.

Finally, we remark that Theorem 12.1 remains valid for the oriented Riemann integral, since exchanging a and b reverses the sign of both sides.

12.1.2. Fundamental theorem of calculus II. Next, we prove the other direction of the fundamental theorem. We will use the following result, of independent interest, which states that the average of a continuous function on an interval approaches the value of the function as the length of the interval shrinks to zero. The proof uses a common trick of taking a constant inside an average.

Theorem 12.4. Suppose that $f : [a, b] \to \mathbb{R}$ is integrable on $[a, b]$ and continuous at a. Then

$$\lim_{h \to 0^+} \frac{1}{h} \int_a^{a+h} f(x)\, dx = f(a).$$

Proof. If k is a constant, then we have

$$k = \frac{1}{h} \int_a^{a+h} k\, dx.$$

(That is, the average of a constant is equal to the constant.) We can therefore write

$$\frac{1}{h} \int_a^{a+h} f(x)\, dx - f(a) = \frac{1}{h} \int_a^{a+h} [f(x) - f(a)]\, dx.$$

Let $\epsilon > 0$. Since f is continuous at a, there exists $\delta > 0$ such that

$$|f(x) - f(a)| < \epsilon \quad \text{for} \quad a \le x < a + \delta.$$

It follows that if $0 < h < \delta$, then

$$\left| \frac{1}{h} \int_a^{a+h} f(x)\, dx - f(a) \right| \le \frac{1}{h} \cdot \sup_{a \le a \le a+h} |f(x) - f(a)| \cdot h \le \epsilon,$$

which proves the result. $\qquad\square$

A similar proof shows that if f is continuous at b, then

$$\lim_{h \to 0^+} \frac{1}{h} \int_{b-h}^b f = f(b),$$

and if f is continuous at $a < c < b$, then

$$\lim_{h \to 0^+} \frac{1}{2h} \int_{c-h}^{c+h} f = f(c).$$

More generally, if f is continuous at c and $\{I_h : h > 0\}$ is any collection of intervals with $c \in I_h$ and $|I_h| \to 0$ as $h \to 0^+$, then

$$\lim_{h \to 0^+} \frac{1}{|I_h|} \int_{I_h} f = f(c).$$

The assumption in Theorem 12.4 that f is continuous at the point about which we take the averages is essential.

Example 12.5. Let $f : \mathbb{R} \to \mathbb{R}$ be the sign function

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

Then

$$\lim_{h \to 0^+} \frac{1}{h} \int_0^h f(x)\,dx = 1, \qquad \lim_{h \to 0^+} \frac{1}{h} \int_{-h}^0 f(x)\,dx = -1,$$

and neither limit is equal to $f(0)$. In this example, the limit of the symmetric averages

$$\lim_{h \to 0^+} \frac{1}{2h} \int_{-h}^h f(x)\,dx = 0$$

is equal to $f(0)$, but this equality doesn't hold if we change $f(0)$ to a nonzero value (for example, if $f(x) = 1$ for $x \geq 0$ and $f(x) = -1$ for $x < 0$) since the limit of the symmetric averages is still 0.

The second part of the fundamental theorem follows from this result and the fact that the difference quotients of F are averages of f.

Theorem 12.6 (Fundamental theorem of calculus II)**.** Suppose that $f : [a, b] \to \mathbb{R}$ is integrable and $F : [a, b] \to \mathbb{R}$ is defined by

$$F(x) = \int_a^x f(t)\,dt.$$

Then F is continuous on $[a, b]$. Moreover, if f is continuous at $a \leq c \leq b$, then F is differentiable at c and $F'(c) = f(c)$.

Proof. First, note that Theorem 11.44 implies that f is integrable on $[a, x]$ for every $a \leq x \leq b$, so F is well-defined. Since f is Riemann integrable, it is bounded, and $|f| \leq M$ for some $M \geq 0$. It follows that

$$|F(x + h) - F(x)| = \left| \int_x^{x+h} f(t)\,dt \right| \leq M|h|,$$

which shows that F is continuous on $[a, b]$ (in fact, Lipschitz continuous).

Moreover, we have

$$\frac{F(c + h) - F(c)}{h} = \frac{1}{h} \int_c^{c+h} f(t)\,dt.$$

It follows from Theorem 12.4 that if f is continuous at c, then F is differentiable at c with

$$F'(c) = \lim_{h \to 0} \left[\frac{F(c + h) - F(c)}{h} \right] = \lim_{h \to 0} \frac{1}{h} \int_c^{c+h} f(t)\,dt = f(c),$$

where we use the appropriate right or left limit at an endpoint. $\square$

The assumption that f is continuous is needed to ensure that F is differentiable.

Example 12.7. If

$$f(x) = \begin{cases} 1 & \text{for } x \geq 0, \\ 0 & \text{for } x < 0, \end{cases}$$

then

$$F(x) = \int_0^x f(t)\, dt = \begin{cases} x & \text{for } x \geq 0, \\ 0 & \text{for } x < 0. \end{cases}$$

The function F is continuous but not differentiable at $x = 0$, where f is discontinuous, since the left and right derivatives of F at 0, given by $F'(0^-) = 0$ and $F'(0^+) = 1$, are different.

12.2. Consequences of the fundamental theorem

The first part of the fundamental theorem, Theorem 12.1, is the basic tool for the exact evaluation of integrals. It allows us to compute the integral of of a function f if we can find an antiderivative; that is, a function F such that $F' = f$. There is no systematic procedure for finding antiderivatives. Moreover, an antiderivative of an elementary function (constructed from power, trigonometric, and exponential functions and their inverses) need not be — and often isn't — expressible in terms of elementary functions. By contrast, the rules of differentiation provide a mechanical algorithm for the computation of the derivative of any function formed from elementary functions by algebraic operations and compositions.

Example 12.8. For $p = 0, 1, 2, \ldots$, we have

$$\frac{d}{dx}\left[\frac{1}{p+1} x^{p+1}\right] = x^p,$$

and it follows that

$$\int_0^1 x^p\, dx = \frac{1}{p+1}.$$

Example 12.9. We can use the fundamental theorem to evaluate certain limits of sums. For example,

$$\lim_{n \to \infty}\left[\frac{1}{n^{p+1}} \sum_{k=1}^n k^p\right] = \frac{1}{p+1},$$

since the sum on the left-hand side is the upper sum of x^p on a partition of $[0, 1]$ into n intervals of equal length. Example 11.27 illustrates this result explicitly for $p = 2$.

Two important general consequences of the first part of the fundamental theorem are integration by parts and substitution (or change of variable), which come from inverting the product rule and chain rule for derivatives, respectively.

Theorem 12.10 (Integration by parts)**.** Suppose that $f, g : [a, b] \to \mathbb{R}$ are continuous on $[a, b]$ and differentiable in (a, b), and f', g' are integrable on $[a, b]$. Then

$$\int_a^b fg'\, dx = f(b)g(b) - f(a)g(a) - \int_a^b f'g\, dx.$$

Proof. The function fg is continuous on $[a, b]$ and, by the product rule, differentiable in (a, b) with derivative

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

Since f, g, f', g' are integrable on $[a, b]$, Theorem 11.35 implies that fg', $f'g$, and $(fg)'$, are integrable. From Theorem 12.1, we get that

$$\int_a^b fg' \, dx + \int_a^b f'g \, dx = \int_a^b (fg)' \, dx = f(b)g(b) - f(a)g(a),$$

which proves the result. $\qquad\square$

Integration by parts says that we can move a derivative from one factor in an integral onto the other factor, with a change of sign and the appearance of a boundary term. The product rule for derivatives expresses the derivative of a product in terms of the derivatives of the factors. By contrast, integration by parts doesn't give an explicit expression for the integral of a product, it simply replaces one integral by another. This can sometimes be used transform an integral into an integral that is easier to evaluate, but the importance of integration by parts goes far beyond its use as an integration technique.

Example 12.11. For $n = 0, 1, 2, 3, \ldots$, let

$$I_n(x) = \int_0^x t^n e^{-t} \, dt.$$

If $n \geq 1$, then integration by parts with $f(t) = t^n$ and $g'(t) = e^{-t}$ gives

$$I_n(x) = -x^n e^{-x} + n \int_0^x t^{n-1} e^{-t} \, dt = -x^n e^{-x} + n I_{n-1}(x).$$

Also, by the fundamental theorem of calculus,

$$I_0(x) = \int_0^x e^{-t} \, dt = 1 - e^{-x}.$$

It then follows by induction that

$$I_n(x) = n! \left[1 - e^{-x} \sum_{k=0}^n \frac{x^k}{k!} \right].$$

Since $x^k e^{-x} \to 0$ as $x \to \infty$ for every $k = 0, 1, 2, \ldots$, we get the improper integral

$$\int_0^\infty t^n e^{-t} \, dt = \lim_{r \to \infty} \int_0^r t^n e^{-t} \, dt = n!$$

This formula suggests an extension of the factorial function to complex numbers $z \in \mathbb{C}$, called the Gamma function, which is defined for $\Re z > 0$ by the improper, complex-valued integral

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} \, dt.$$

In particular, $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$. The Gamma function is an important special function, which is studied further in complex analysis.

Next we consider the change of variable formula for integrals.

Theorem 12.12 (Change of variable)**.** Suppose that $g : I \to \mathbb{R}$ differentiable on an open interval I and g' is integrable on I. Let $J = g(I)$. If $f : J \to \mathbb{R}$ continuous, then for every $a, b \in I$,

$$\int_a^b f\left(g(x)\right) g'(x)\, dx = \int_{g(a)}^{g(b)} f(u)\, du.$$

Proof. For $x \in J$, let

$$F(x) = \int_{g(a)}^x f(u)\, du.$$

Since f is continuous, Theorem 12.6 implies that F is differentiable in J with $F' = f$. The chain rule implies that the composition $F \circ g : I \to \mathbb{R}$ is differentiable in I, with

$$(F \circ g)'(x) = f\left(g(x)\right) g'(x).$$

This derivative is integrable on $[a, b]$ since $f \circ g$ is continuous and g' is integrable. Theorem 12.1, the definition of F, and the additivity of the integral then imply that

$$
\begin{aligned}
\int_a^b f\left(g(x)\right) g'(x)\, dx &= \int_a^b (F \circ g)'\, dx \\
&= F\left(g(b)\right) - F\left(g(a)\right) \\
&= \int_{g(a)}^{g(b)} F'(u)\, du,
\end{aligned}
$$

which proves the result. $\square$

There is no assumption in this theorem that g is invertible, but we often use the theorem in that case. A continuous function maps an interval to an interval, and it is one-to-one if and only if it is strictly monotone. An increasing function preserves the orientation of the interval, while a decreasing function reverses it, in which case the integrals in the previous theorem are understood as oriented integrals.

This result can also be formulated in terms of non-oriented integrals. Suppose that $g : I \to J$ is one-to-one and onto from an interval $I = [a, b]$ to an interval $J = g(I) = [c, d]$ where $c = g(a)$, $d = g(b)$ if g is increasing, and $c = g(b)$, $d = g(a)$ if g is decreasing, then

$$\int_{g(I)} f(u)\, du = \int_I (f \circ g)(x)|g'(x)|\, dx.$$

In this identity, both integrals are over positively oriented intervals and we include an absolute value in the Jacobian factor $|g'(x)|$. If $g' \geq 0$, then this identity is the

same as the oriented form, while if $g' \leq 0$, then

$$\int_I (f \circ g)(x)|g'(x)| \, dx = \int_a^b (f \circ g)(x)[-g'(x)] \, dx$$

$$= -\int_{g(a)}^{g(b)} f(u) \, du$$

$$= \int_{g(b)}^{g(a)} f(u) \, du$$

$$= \int_{g(I)} f(u) \, du.$$

Example 12.13. For every $a > 0$, the increasing, differentiable function $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = x^3$ maps $(-a, a)$ one-to-one and onto $(-a^3, a^3)$ and preserves orientation. Thus, if $f : [-a, a] \to \mathbb{R}$ is continuous,

$$\int_{-a}^a f(x^3) \cdot 3x^2 \, dx = \int_{-a^3}^{a^3} f(u) \, du.$$

The decreasing, differentiable function $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = -x^3$ maps $(-a, a)$ one-to-one and onto $(-a^3, a^3)$ and reverses orientation. Thus,

$$\int_{-a}^a f(-x^3) \cdot (-3x^2) \, dx = \int_{a^3}^{-a^3} f(u) \, du = -\int_{-a^3}^{a^3} f(u) \, du.$$

The non-monotone, differentiable function $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = x^2$ maps $(-a, a)$ onto $[0, a^2)$. It is two-to-one, except at $x = 0$. The change of variables formula gives

$$\int_{-a}^a f(x^2) \cdot 2x \, dx = \int_{a^2}^{a^2} f(u) \, du = 0.$$

The contributions to the original integral from $[0, a]$ and $[-a, 0]$ cancel since the integrand is an odd function of x.

One consequence of the second part of the fundamental theorem, Theorem 12.6, is that every continuous function has an antiderivative, even if it can't be expressed explicitly in terms of elementary functions. This provides a way to define transcendental functions as integrals of elementary functions.

Example 12.14. One way to define the natural logarithm $\log : (0, \infty) \to \mathbb{R}$ in terms of algebraic functions is as the integral

$$\log x = \int_1^x \frac{1}{t} \, dt.$$

This integral is well-defined for every $0 < x < \infty$ since $1/t$ is continuous on the interval $[1, x]$ if $x > 1$, or $[x, 1]$ if $0 < x < 1$. The usual properties of the logarithm follow from this representation. We have $(\log x)' = 1/x$ by definition; and, for example, by making the substitution $s = xt$ in the second integral in the following equation, when $dt/t = ds/s$, we get

$$\log x + \log y = \int_1^x \frac{1}{t} \, dt + \int_1^y \frac{1}{t} \, dt = \int_1^x \frac{1}{t} \, dt + \int_x^{xy} \frac{1}{s} \, ds = \int_1^{xy} \frac{1}{t} \, dt = \log(xy).$$
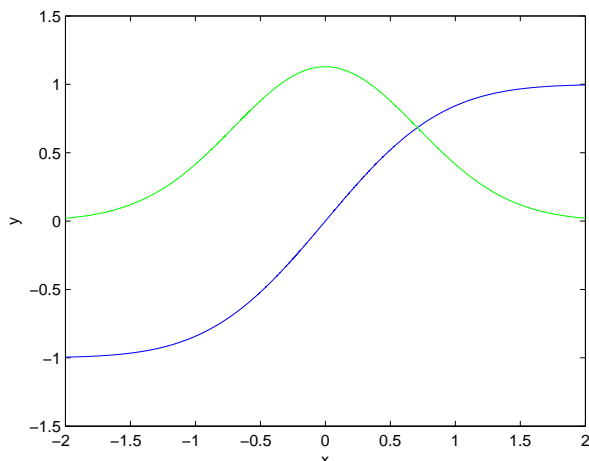
Figure 1. Graphs of the error function $y = F(x)$ (blue) and its derivative, the Gaussian function $y = f(x)$ (green), from Example 12.15.

We can also define many non-elementary functions as integrals.

Example 12.15. The error function

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} \, dt$$

is an anti-derivative on $\mathbb{R}$ of the Gaussian function

$$f(x) = \frac{2}{\sqrt{\pi}} e^{-x^2}.$$

The error function isn't expressible in terms of elementary functions. Nevertheless, it is defined as a limit of Riemann sums for the integral. Figure 1 shows the graphs of f and F. The name "error function" comes from the fact that the probability of a Gaussian random variable deviating by more than a given amount from its mean can be expressed in terms of F. Error functions also arise in other applications; for example, in modeling diffusion processes such as heat flow.

Example 12.16. The Fresnel sine function S is defined by

$$S(x) = \int_0^x \sin\left(\frac{\pi t^2}{2}\right) \, dt.$$

The function S is an antiderivative of $\sin(\pi t^2/2)$ on $\mathbb{R}$ (see Figure 2), but it can't be expressed in terms of elementary functions. Fresnel integrals arise, among other places, in analysing the diffraction of waves, such as light waves. From the perspective of complex analysis, they are closely related to the error function through the Euler formula $e^{i\theta} = \cos\theta + i\sin\theta$.

Discontinuous functions may or may not have an antiderivative and typically don't. Darboux proved that every function $f : (a, b) \to \mathbb{R}$ that is the derivative of a function $F : (a, b) \to \mathbb{R}$, where $F' = f$ at all points of (a, b), has the intermediate
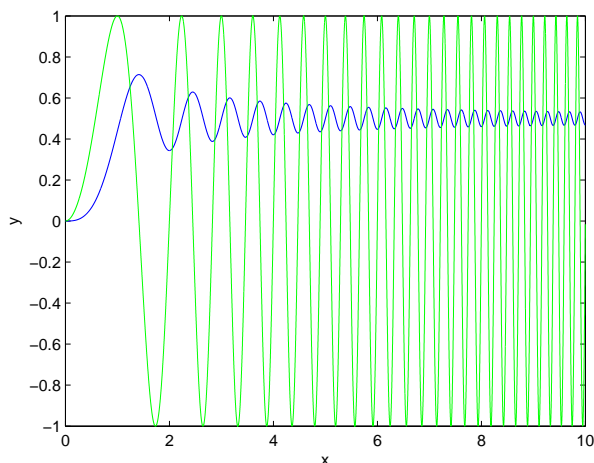
Figure 2. Graphs of the Fresnel integral $y = S(x)$ (blue) and its derivative $y = \sin(\pi x^2/2)$ (green) from Example 12.16.

value property. That is, for all c, d such that if $a < c < d < b$ and all y between $f(c)$ and $f(d)$, there exists an x between c and d such that $f(x) = y$. A continuous derivative has this property by the intermediate value theorem, but a discontinuous derivative also has it. Thus, discontinuous functions without the intermediate value property, such as ones with a jump discontinuity, don't have an antiderivative. For example, the function F in Example 12.7 is not an antiderivative of the step function f on $\mathbb{R}$ since it isn't differentiable at 0.

In dealing with functions that are not continuously differentiable, it turns out to be more useful to abandon the idea of a derivative that is defined pointwise everywhere (pointwise values of discontinuous functions are somewhat arbitrary) and introduce the notion of a weak derivative. We won't define or study weak derivatives here.

12.3. Integrals and sequences of functions

A fundamental question that arises throughout analysis is the validity of an exchange in the order of limits. Some sort of condition is always required.

In this section, we consider the question of when the convergence of a sequence of functions $f_n \to f$ implies the convergence of their integrals $\int f_n \to \int f$. Here, we exchange a limit of a sequence of functions with a limit of the Riemann sums that define their integrals. The two types of convergence we'll discuss are pointwise and uniform convergence, which are defined in Chapter 9.

As we show first, the Riemann integral is well-behaved with respect to uniform convergence. The drawback to uniform convergence is that it's a strong form of convergence, and we often want to use a weaker form, such as pointwise convergence, in which case the Riemann integral may not be suitable.

12.3.1. Uniform convergence. The uniform limit of continuous functions is continuous and therefore integrable. The next result shows, more generally, that the uniform limit of integrable functions is integrable. Furthermore, the limit of the integrals is the integral of the limit.

Theorem 12.17. Suppose that $f_n : [a, b] \to \mathbb{R}$ is Riemann integrable for each $n \in \mathbb{N}$ and $f_n \to f$ uniformly on $[a, b]$ as $n \to \infty$. Then $f : [a, b] \to \mathbb{R}$ is Riemann integrable on $[a, b]$ and

$$\int_a^b f = \lim_{n \to \infty} \int_a^b f_n.$$

Proof. The main statement we need to prove is that f is integrable. Let $\epsilon > 0$. Since $f_n \to f$ uniformly, there is an $N \in \mathbb{N}$ such that if $n > N$ then

$$f_n(x) - \frac{\epsilon}{b-a} < f(x) < f_n(x) + \frac{\epsilon}{b-a} \qquad \text{for all } a \leq x \leq b.$$

It follows from Proposition 11.39 that

$$L\left(f_n - \frac{\epsilon}{b-a}\right) \leq L(f), \qquad U(f) \leq U\left(f_n + \frac{\epsilon}{b-a}\right).$$

Since f_n is integrable and upper integrals are greater than lower integrals, we get that

$$\int_a^b f_n - \epsilon \leq L(f) \leq U(f) \leq \int_a^b f_n + \epsilon$$

for all $n > N$, which implies that

$$0 \leq U(f) - L(f) \leq 2\epsilon.$$

Since $\epsilon > 0$ is arbitrary, we conclude that $L(f) = U(f)$, so f is integrable. Moreover, it follows that for all $n > N$ we have

$$\left| \int_a^b f_n - \int_a^b f \right| \leq \epsilon,$$

which shows that $\int_a^b f_n \to \int_a^b f$ as $n \to \infty$. $\qquad\qquad\square$

Alternatively, once we know that the uniform limit of integrable functions is integrable, the convergence of the integrals follows directly from the estimate

$$\left| \int_a^b f_n - \int_a^b f \right| = \left| \int_a^b (f_n - f) \right| \leq \sup_{x \in [a,b]} |f_n(x) - f(x)| \cdot (b - a) \to 0 \qquad \text{as } n \to \infty.$$

Example 12.18. The function $f_n : [0, 1] \to \mathbb{R}$ defined by

$$f_n(x) = \frac{n + \cos x}{ne^x + \sin x}$$

converges uniformly on $[0, 1]$ to $f(x) = e^{-x}$ since, for $0 \leq x \leq 1$,

$$\left| \frac{n + \cos x}{ne^x + \sin x} - e^{-x} \right| = \left| \frac{\cos x - e^{-x} \sin x}{ne^x + \sin x} \right| \leq \frac{1}{n}.$$

It follows that

$$\lim_{n \to \infty} \int_0^1 \frac{n + \cos x}{ne^x + \sin x}\, dx = \int_0^1 e^{-x}\, dx = 1 - \frac{1}{e}.$$

Example 12.19. Every power series

$$f(x) = a_0 + a_1 x + a^2 x^2 + \cdots + a_n x^n + \ldots$$

with radius of convergence $R > 0$ converges uniformly on compact intervals inside the interval $|x| < R$, so we can integrate it term-by-term to get

$$\int_0^x f(t)\, dt = a_0 x + \frac{1}{2} a_1 x^2 + \frac{1}{3} a_2 x^3 + \cdots + \frac{1}{n+1} a_n x^{n+1} + \ldots \qquad \text{for } |x| < R.$$

Example 12.20. If we integrate the geometric series

$$\frac{1}{1-x} = 1 + x + x^2 + \cdots + x^n + \ldots \qquad \text{for } |x| < 1,$$

we get a power series for log,

$$\log\left(\frac{1}{1-x}\right) = x + \frac{1}{2}x^2 + \frac{1}{3}x^3 \cdots + \frac{1}{n}x^n + \ldots \qquad \text{for } |x| < 1.$$

For instance, taking $x = 1/2$, we get the rapidly convergent series

$$\log 2 = \sum_{n=1}^{\infty} \frac{1}{n 2^n}$$

for the irrational number $\log 2 \approx 0.6931$. This series was known and used by Euler. For comparison, the alternating harmonic series in Example 12.46 also converges to $\log 2$, but it does so extremely slowly and would be a poor choice for computing a numerical approximation.

Although we can integrate uniformly convergent sequences, we cannot in general differentiate them. In fact, it's often easier to prove results about the convergence of derivatives by using results about the convergence of integrals, together with the fundamental theorem of calculus. The following theorem provides sufficient conditions for $f_n \to f$ to imply that $f_n' \to f'$.

Theorem 12.21. Let $f_n : (a, b) \to \mathbb{R}$ be a sequence of differentiable functions whose derivatives $f_n' : (a, b) \to \mathbb{R}$ are integrable on (a, b). Suppose that $f_n \to f$ pointwise and $f_n' \to g$ uniformly on (a, b) as $n \to \infty$, where $g : (a, b) \to \mathbb{R}$ is continuous. Then $f : (a, b) \to \mathbb{R}$ is continuously differentiable on (a, b) and $f' = g$.

Proof. Choose some point $a < c < b$. Since f_n' is integrable, the fundamental theorem of calculus, Theorem 12.1, implies that

$$f_n(x) = f_n(c) + \int_c^x f_n' \qquad \text{for } a < x < b.$$

Since $f_n \to f$ pointwise and $f_n' \to g$ uniformly on $[a, x]$, we find that

$$f(x) = f(c) + \int_c^x g.$$

Since g is continuous, the other direction of the fundamental theorem, Theorem 12.6, implies that f is differentiable in (a, b) and $f' = g$. $\qquad \square$

In particular, this theorem shows that the limit of a uniformly convergent sequence of continuously differentiable functions whose derivatives converge uniformly is also continuously differentiable.

The key assumption in Theorem 12.21 is that the derivatives f_n' converge uniformly, not just pointwise; the result is false if we only assume pointwise convergence of the f_n'. In the proof of the theorem, we only use the assumption that $f_n(x)$ converges at a single point $x = c$. This assumption together with the assumption that $f_n' \to g$ uniformly implies that $f_n \to f$ uniformly, where

$$f(x) = \lim_{n \to \infty} f_n(c) + \int_c^x g.$$

Thus, the theorem remains true if we replace the assumption that $f_n \to f$ pointwise on (a, b) by the weaker assumption that $\lim_{n \to \infty} f_n(c)$ exists for some $c \in (a, b)$. This isn't an important change, however, because the restrictive assumption in the theorem is the uniform convergence of the derivatives f_n', not the pointwise (or uniform) convergence of the functions f_n.

The assumption that $g = \lim f_n'$ is continuous is needed to show the differentiability of f by the fundamental theorem, but the result remains true even if g isn't continuous. In that case, however, a different — and more complicated — proof is required, which is given in Theorem 9.18.

12.3.2. Pointwise convergence. On its own, the pointwise convergence of functions is never sufficient to imply convergence of their integrals.

Example 12.22. For $n \in \mathbb{N}$, define $f_n : [0, 1] \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} n & \text{if } 0 < x < 1/n, \\ 0 & \text{if } x = 0 \text{ or } 1/n \le x \le 1. \end{cases}$$

Then $f_n \to 0$ pointwise on $[0, 1]$ but

$$\int_0^1 f_n = 1$$

for every $n \in \mathbb{N}$. By slightly modifying these functions to

$$f_n(x) = \begin{cases} n^2 & \text{if } 0 < x < 1/n, \\ 0 & \text{if } x = 0 \text{ or } 1/n \le x \le 1, \end{cases}$$

we get a sequence that converges pointwise to 0 but whose integrals diverge to ∞. The fact that the f_n are discontinuous is not important; we could replace the step functions by continuous "tent" functions or smooth "bump" functions.

The behavior of the integral under pointwise convergence in the previous example is unavoidable whatever definition of the integral one uses. A more serious defect of the Riemann integral is that the pointwise limit of Riemann integrable functions needn't be Riemann integrable at all, even if it is bounded.

Example 12.23. Let $\{r_k : k \in \mathbb{N}\}$ be an enumeration of the rational numbers in $[0, 1]$ and define $f_n : [0, 1] \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} 1 & \text{if } x = r_k \text{ for some } 1 \le k \le n, \\ 0 & \text{otherwise.} \end{cases}$$

Then each f_n is Riemann integrable since it differs from the zero function at finitely many points. However, $f_n \to f$ pointwise on $[0, 1]$ to the Dirichlet function f, which is not Riemann integrable.

This is another place where the Lebesgue integral has better properties than the Riemann integral. The pointwise (or pointwise almost everywhere) limit of Lebesgue measurable functions is Lebesgue measurable. As Example 12.22 shows, we still need conditions to ensure the convergence of the integrals, but there are quite simple and general conditions for the Lebesgue integral (such as the monotone convergence and dominated convergence theorems).

12.4. Improper Riemann integrals

The Riemann integral is only defined for a bounded function on a compact interval (or a finite union of such intervals). Nevertheless, we frequently want to integrate unbounded functions or functions on an infinite interval. One way to interpret such integrals is as a limit of Riemann integrals; these limits are called improper Riemann integrals.

12.4.1. Improper integrals. First, we define the improper integral of a function that fails to be integrable at one endpoint of a bounded interval.

Definition 12.24. Suppose that $f : (a, b] \to \mathbb{R}$ is integrable on $[c, b]$ for every $a < c < b$. Then the improper integral of f on $[a, b]$ is

$$\int_a^b f = \lim_{\epsilon \to 0^+} \int_{a+\epsilon}^b f.$$

The improper integral converges if this limit exists (as a finite real number), otherwise it diverges. Similarly, if $f : [a, b) \to \mathbb{R}$ is integrable on $[a, c]$ for every $a < c < b$, then

$$\int_a^b f = \lim_{\epsilon \to 0^+} \int_a^{b-\epsilon} f.$$

We use the same notation to denote proper and improper integrals; it should be clear from the context which integrals are proper Riemann integrals (i.e., ones given by Definition 11.11) and which are improper. If f is Riemann integrable on $[a, b]$, then Proposition 11.50 shows that its improper and proper integrals agree, but an improper integral may exist even if f isn't integrable.

Example 12.25. If $p > 0$, then the integral

$$\int_0^1 \frac{1}{x^p} \, dx$$

isn't defined as a Riemann integral since $1/x^p$ is unbounded on $(0, 1]$. The corresponding improper integral is

$$\int_0^1 \frac{1}{x^p}\, dx = \lim_{\epsilon \to 0^+} \int_\epsilon^1 \frac{1}{x^p}\, dx.$$

For $p \neq 1$, we have

$$\int_\epsilon^1 \frac{1}{x^p}\, dx = \frac{1 - \epsilon^{1-p}}{1 - p},$$

so the improper integral converges if $0 < p < 1$, with

$$\int_0^1 \frac{1}{x^p}\, dx = \frac{1}{p - 1},$$

and diverges to ∞ if $p > 1$. The integral also diverges (more slowly) to ∞ if $p = 1$ since

$$\int_\epsilon^1 \frac{1}{x}\, dx = \log \frac{1}{\epsilon}.$$

Thus, we get a convergent improper integral if the integrand $1/x^p$ does not grow too rapidly as $x \to 0^+$ (slower than $1/x$).

We define improper integrals on an unbounded interval as limits of integrals on bounded intervals.

Definition 12.26. Suppose that $f : [a, \infty) \to \mathbb{R}$ is integrable on $[a, r]$ for every $r > a$. Then the improper integral of f on $[a, \infty)$ is

$$\int_a^\infty f = \lim_{r \to \infty} \int_a^r f.$$

Similarly, if $f : (-\infty, b] \to \mathbb{R}$ is integrable on $[r, b]$ for every $r < b$, then

$$\int_{-\infty}^b f = \lim_{r \to \infty} \int_{-r}^b f.$$

Let's consider the convergence of the integral of the power function in Example 12.25 at infinity rather than at zero.

Example 12.27. Suppose that $p > 0$. The improper integral

$$\int_1^\infty \frac{1}{x^p}\, dx = \lim_{r \to \infty} \int_1^r \frac{1}{x^p}\, dx = \lim_{r \to \infty} \left(\frac{r^{1-p} - 1}{1 - p} \right)$$

converges to $1/(p-1)$ if $p > 1$ and diverges to ∞ if $0 < p < 1$. It also diverges (more slowly) if $p = 1$ since

$$\int_1^\infty \frac{1}{x}\, dx = \lim_{r \to \infty} \int_1^r \frac{1}{x}\, dx = \lim_{r \to \infty} \log r = \infty.$$

Thus, we get a convergent improper integral if the integrand $1/x^p$ decays sufficiently rapidly as $x \to \infty$ (faster than $1/x$).

A divergent improper integral may diverge to ∞ (or $-\infty$) as in the previous examples, or — if the integrand changes sign — it may oscillate.

Example 12.28. Define $f : [0, \infty) \to \mathbb{R}$ by
$$f(x) = (-1)^n \qquad \text{for } n \leq x < n + 1 \text{ where } n = 0, 1, 2, \ldots.$$
Then $0 \leq \int_0^r f \leq 1$ and
$$\int_0^n f = \begin{cases} 1 & \text{if } n \text{ is an odd integer,} \\ 0 & \text{if } n \text{ is an even integer.} \end{cases}$$
Thus, the improper integral $\int_0^\infty f$ doesn't converge.

More general improper integrals may be defined as finite sums of improper integrals of the previous forms. For example, if $f : [a, b] \setminus \{c\} \to \mathbb{R}$ is integrable on closed intervals not including $a < c < b$, then
$$\int_a^b f = \lim_{\delta \to 0^+} \int_a^{c-\delta} f + \lim_{\epsilon \to 0^+} \int_{c+\epsilon}^b f;$$
and if $f : \mathbb{R} \to \mathbb{R}$ is integrable on every compact interval, then
$$\int_{-\infty}^\infty f = \lim_{s \to \infty} \int_{-s}^c f + \lim_{r \to \infty} \int_c^r f,$$
where we split the integral at an arbitrary point $c \in \mathbb{R}$. Note that each limit is required to exist separately.

Example 12.29. The improper Riemann integral
$$\int_0^\infty \frac{1}{x^p} \, dx = \lim_{\epsilon \to 0^+} \int_\epsilon^1 \frac{1}{x^p} \, dx + \lim_{r \to \infty} \int_1^r \frac{1}{x^p} \, dx$$
does not converge for any $p \in \mathbb{R}$, since the integral either diverges at 0 (if $p \geq 1$) or at infinity (if $p \leq 1$).

Example 12.30. If $f : [0, 1] \to \mathbb{R}$ is continuous and $0 < c < 1$, then we define as an improper integral
$$\int_0^1 \frac{f(x)}{|x - c|^{1/2}} \, dx = \lim_{\delta \to 0^+} \int_0^{c-\delta} \frac{f(x)}{|x - c|^{1/2}} \, dx + \lim_{\epsilon \to 0^+} \int_{c+\epsilon}^1 \frac{f(x)}{|x - c|^{1/2}} \, dx.$$

Example 12.31. Consider the following integral, called a Frullani integral,
$$I = \int_0^\infty \frac{f(ax) - f(bx)}{x} \, dx,$$
where $0 < a < b$ and $f : [0, \infty) \to \mathbb{R}$ is a continuous function whose limit as $x \to \infty$ exists. We write this limit as
$$f(\infty) = \lim_{x \to \infty} f(x).$$
We interpret the integral as an improper integral $I = I_1 + I_2$ where
$$I_1 = \lim_{\epsilon \to 0^+} \int_\epsilon^1 \frac{f(ax) - f(bx)}{x} \, dx, \qquad I_2 = \lim_{r \to \infty} \int_1^r \frac{f(ax) - f(bx)}{x} \, dx.$$
Consider I_1. After making the substitutions $s = ax$ and $t = bx$ and using the additivity property of the integral, we get that
$$I_1 = \lim_{\epsilon \to 0^+} \left(\int_{\epsilon a}^a \frac{f(s)}{s} \, ds - \int_{\epsilon b}^b \frac{f(t)}{t} \, dt \right) = \lim_{\epsilon \to 0^+} \left(\int_{\epsilon a}^{\epsilon b} \frac{f(t)}{t} \, dt \right) - \int_a^b \frac{f(t)}{t} \, dt.$$

To evaluate the limit, we write

$$\int_{\epsilon a}^{\epsilon b} \frac{f(t)}{t}\, dt = \int_{\epsilon a}^{\epsilon b} \frac{f(t) - f(0)}{t}\, dt + f(0) \int_{\epsilon a}^{\epsilon b} \frac{1}{t}\, dt$$

$$= \int_{\epsilon a}^{\epsilon b} \frac{f(t) - f(0)}{t}\, dt + f(0) \log\left(\frac{b}{a}\right).$$

Since f is continuous at 0 and $t \geq \epsilon a$ in the interval of integration of length $\epsilon(b-a)$, we have

$$\left| \int_{\epsilon a}^{\epsilon b} \frac{f(t) - f(0)}{t}\, dt \right| \leq \left(\frac{b-a}{a} \right) \cdot \max\{ |f(t) - f(0)| : \epsilon a \leq t \leq \epsilon b \} \to 0$$

as $\epsilon \to 0^+$. It follows that

$$I_1 = f(0) \log\left(\frac{b}{a}\right) - \int_a^b \frac{f(t)}{t}\, dt.$$

A similar argument gives

$$I_2 = -f(\infty) \log\left(\frac{b}{a}\right) + \int_a^b \frac{f(t)}{t}\, dt.$$

Adding these results, we conclude that

$$\int_0^\infty \frac{f(ax) - f(bx)}{x}\, dx = \{ f(0) - f(\infty) \} \log\left(\frac{b}{a}\right).$$

12.4.2. Absolutely convergent improper integrals. The convergence of improper integrals is analogous to the convergence of series. A series $\sum a_n$ converges absolutely if $\sum |a_n|$ converges, and conditionally if $\sum a_n$ converges but $\sum |a_n|$ diverges. We introduce a similar definition for improper integrals and provide a test for the absolute convergence of an improper integral that is analogous to the comparison test for series.

Definition 12.32. An improper integral $\int_a^b f$ is absolutely convergent if the improper integral $\int_a^b |f|$ converges, and conditionally convergent if $\int_a^b f$ converges but $\int_a^b |f|$ diverges.

As part of the next theorem, we prove that an absolutely convergent improper integral converges (similarly, an absolutely convergent series converges).

Theorem 12.33. Suppose that $f, g : I \to \mathbb{R}$ are defined on some finite or infinite interval I. If $|f| \leq g$ and the improper integral $\int_I g$ converges, then the improper integral $\int_I f$ converges absolutely. Moreover, an absolutely convergent improper integral converges.

Proof. To be specific, we suppose that $f, g : [a, \infty) \to \mathbb{R}$ are integrable on $[a, r]$ for $r > a$ and consider the improper integral

$$\int_a^\infty f = \lim_{r \to \infty} \int_a^r f.$$

A similar argument applies to other types of improper integrals.

First, suppose that $f \geq 0$. Then

$$\int_a^r f \leq \int_a^r g \leq \int_a^\infty g,$$

so $\int_a^r f$ is a monotonic increasing function of r that is bounded from above. Therefore it converges as $r \to \infty$.

In general, we decompose f into its positive and negative parts,

$$f = f_+ - f_-, \qquad |f| = f_+ + f_-,$$
$$f_+ = \max\{f, 0\}, \qquad f_- = \max\{-f, 0\}.$$

We have $0 \leq f_\pm \leq g$, so the improper integrals of $f_\pm$ converge by the previous argument, and therefore so does the improper integral of f:

$$\int_a^\infty f = \lim_{r \to \infty} \left(\int_a^r f_+ - \int_a^r f_- \right)$$
$$= \lim_{r \to \infty} \int_a^r f_+ - \lim_{r \to \infty} \int_a^r f_-$$
$$= \int_a^\infty f_+ - \int_a^\infty f_-.$$

Moreover, since $0 \leq f_\pm \leq |f|$, we see that $\int_a^\infty f_+$ and $\int_a^\infty f_-$ both converge if $\int_a^\infty |f|$ converges, and therefore so does $\int_a^\infty f$, so an absolutely convergent improper integral converges. $\qquad \square$

Example 12.34. Consider the limiting behavior of the error function $\mathrm{erf}(x)$ in Example 12.15 as $x \to \infty$, which is given by

$$\frac{2}{\sqrt{\pi}} \int_0^\infty e^{-x^2}\, dx = \frac{2}{\sqrt{\pi}} \lim_{r \to \infty} \int_0^r e^{-x^2}\, dx.$$

The convergence of this improper integral follows by comparison with e^{-x}, for example, since

$$0 \leq e^{-x^2} \leq e^{-x} \qquad \text{for } x \geq 1,$$

and

$$\int_1^\infty e^{-x}\, dx = \lim_{r \to \infty} \int_1^r e^{-x}\, dx = \lim_{r \to \infty} \left(e^{-1} - e^{-r} \right) = \frac{1}{e}.$$

This argument proves that the error function approaches a finite limit as $x \to \infty$, but it doesn't give the exact value, only an upper bound

$$\frac{2}{\sqrt{\pi}} \int_0^\infty e^{-x^2}\, dx \leq M, \qquad M = \frac{2}{\sqrt{\pi}} \left(\int_0^1 e^{-x^2}\, dx + \frac{1}{e} \right) \leq \frac{2}{\sqrt{\pi}} \left(1 + \frac{1}{e} \right).$$

One can evaluate this improper integral exactly, with the result that

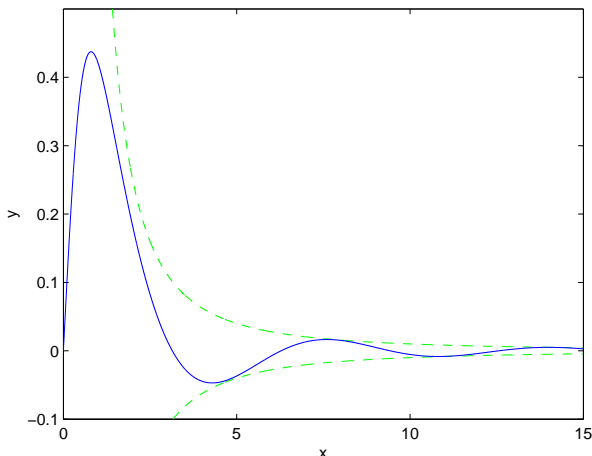$$\frac{2}{\sqrt{\pi}} \int_0^\infty e^{-x^2}\, dx = 1.$$

Figure 3. Graph of $y = (\sin x)/(1 + x^2)$ from Example 12.35. The dashed green lines are the graphs of $y = \pm 1/x^2$.

The standard trick to obtain this result (apparently introduced by Laplace) uses double integration, polar coordinates, and the substitution $u = r^2$:

$$\left(\int_0^\infty e^{-x^2} \, dx \right)^2 = \int_0^\infty \int_0^\infty e^{-x^2 - y^2} \, dx dy$$
$$= \int_0^{\pi/2} \left(\int_0^\infty e^{-r^2} \, r \, dr \right) d\theta$$
$$= \frac{\pi}{4} \int_0^\infty e^{-u} \, du = \frac{\pi}{4}.$$

This formal computation can be justified rigorously, but we won't do so here. There are also many other ways to obtain the same result.

Example 12.35. The improper integral

$$\int_0^\infty \frac{\sin x}{1 + x^2} \, dx = \lim_{r \to \infty} \int_0^r \frac{\sin x}{1 + x^2} \, dx$$

converges absolutely, since

$$\int_0^\infty \frac{\sin x}{1 + x^2} \, dx = \int_0^1 \frac{\sin x}{1 + x^2} \, dx + \int_1^\infty \frac{\sin x}{1 + x^2} \, dx$$

and (see Figure 3)

$$\left| \frac{\sin x}{1 + x^2} \right| \le \frac{1}{x^2} \quad \text{for } x \ge 1, \qquad \int_1^\infty \frac{1}{x^2} \, dx < \infty.$$

The value of this integral doesn't have an elementary expression, but by using contour integration from complex analysis one can show that

$$\int_0^\infty \frac{\sin x}{1 + x^2} \, dx = \frac{1}{2e} \operatorname{Ei}(1) - \frac{e}{2} \operatorname{Ei}(-1) \approx 0.6468,$$

where Ei is the exponential integral function defined in Example 12.41.

Improper integrals, and the principal value integrals discussed below, arise frequently in complex analysis, and many such integrals can be evaluated by contour integration.

Example 12.36. The improper integral

$$\int_0^\infty \frac{\sin x}{x}\, dx = \lim_{r\to\infty} \int_0^r \frac{\sin x}{x}\, dx = \frac{\pi}{2}$$

converges conditionally. We leave the proof as an exercise. Note that there is no difficulty at 0, since $\sin x/x \to 1$ as $x \to 0$, and comparison with the function $1/x$ doesn't imply absolute convergence at infinity because the improper integral $\int_1^\infty 1/x\, dx$ diverges. There are many ways to show that the exact value of the improper integral is $\pi/2$. The standard method uses contour integration.

Example 12.37. Consider the limiting behavior of the Fresnel sine function $S(x)$ in Example 12.16 as $x \to \infty$. The improper integral

$$\int_0^\infty \sin\left(\frac{\pi x^2}{2}\right) dx = \lim_{r\to\infty} \int_0^r \sin\left(\frac{\pi x^2}{2}\right) dx = \frac{1}{2}.$$

converges conditionally. This example may seem surprising since the integrand $\sin(\pi x^2/2)$ doesn't converge to 0 as $x \to \infty$. The explanation is that the integrand oscillates more rapidly with increasing x, leading to a more rapid cancelation between positive and negative values in the integral (see Figure 2). The exact value can be found by contour integration, again, which shows that

$$\int_0^\infty \sin\left(\frac{\pi x^2}{2}\right) dx = \frac{1}{\sqrt{2}} \int_0^\infty \exp\left(-\frac{\pi x^2}{2}\right) dx.$$

Evaluation of the resulting Gaussian integral gives $1/2$.

12.5. * Principal value integrals

Some integrals have a singularity that is too strong for them to converge as improper integrals but, due to cancelation between positive and negative parts of the integrand, they have a finite limit as a principal value integral. We begin with an example.

Example 12.38. Consider $f : [-1, 1] \setminus \{0\}$ defined by

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

The definition of the integral of f on $[-1, 1]$ as an improper integral is

$$\int_{-1}^{1} \frac{1}{x}\, dx = \lim_{\delta\to 0^+} \int_{-1}^{-\delta} \frac{1}{x}\, dx + \lim_{\epsilon\to 0^+} \int_{\epsilon}^{1} \frac{1}{x}\, dx$$
$$= \lim_{\delta\to 0^+} \log\delta - \lim_{\epsilon\to 0^+} \log\epsilon.$$

Neither limit exists, so the improper integral diverges. (Formally, we get $\infty - \infty$.) If, however, we take $\delta = \epsilon$ and combine the limits, we get a convergent principal value integral, which is defined by

$$\text{p.v.} \int_{-1}^{1} \frac{1}{x} \, dx = \lim_{\epsilon \to 0^+} \left(\int_{-1}^{-\epsilon} \frac{1}{x} \, dx + \int_{\epsilon}^{1} \frac{1}{x} \, dx \right) = \lim_{\epsilon \to 0^+} (\log \epsilon - \log \epsilon) = 0.$$

The value of 0 is what one might expect from the oddness of the integrand. A cancelation in the contributions from either side of the singularity is essential to obtain a finite limit.

The principal value integral of $1/x$ on a non-symmetric interval about 0 still exists but is non-zero. For example, if $b > 0$, then

$$\text{p.v.} \int_{-1}^{b} \frac{1}{x} \, dx = \lim_{\epsilon \to 0^+} \left(\int_{-1}^{-\epsilon} \frac{1}{x} \, dx + \int_{\epsilon}^{b} \frac{1}{x} \, dx \right) = \lim_{\epsilon \to 0^+} (\log \epsilon + \log b - \log \epsilon) = \log b.$$

The crucial feature of a principal value integral is that we remove a symmetric interval around a singular point, or infinity. The resulting cancelation in the integral of a non-integrable function that changes sign across the singularity may lead to a finite limit.

Definition 12.39. If $f : [a, b] \setminus \{c\} \to \mathbb{R}$ is integrable on closed intervals not including $a < c < b$, then the principal value integral of f on $[a, b]$ is

$$\text{p.v.} \int_{a}^{b} f = \lim_{\epsilon \to 0^+} \left(\int_{a}^{c-\epsilon} f + \int_{c+\epsilon}^{b} f \right).$$

If $f : \mathbb{R} \to \mathbb{R}$ is integrable on compact intervals, then the principal value integral of f on $\mathbb{R}$ is

$$\text{p.v.} \int_{-\infty}^{\infty} f = \lim_{r \to \infty} \int_{-r}^{r} f.$$

If the improper integral exists, then the principal value integral exists and is equal to the improper integral. As Example 12.38 shows, the principal value integral may exist even if the improper integral does not. Of course, a principal value integral may also diverge.

Example 12.40. Consider the principal value integral

$$\text{p.v.} \int_{-1}^{1} \frac{1}{x^2} \, dx = \lim_{\epsilon \to 0^+} \left(\int_{-1}^{-\epsilon} \frac{1}{x^2} \, dx + \int_{\epsilon}^{1} \frac{1}{x^2} \, dx \right)$$
$$= \lim_{\epsilon \to 0^+} \left(\frac{2}{\epsilon} - 2 \right) = \infty.$$

In this case, the function $1/x^2$ is positive and approaches ∞ on both sides of the singularity at $x = 0$, so there is no cancelation and the principal value integral diverges to ∞.

Principal value integrals arise frequently in complex analysis, harmonic analysis, and a variety of applications.
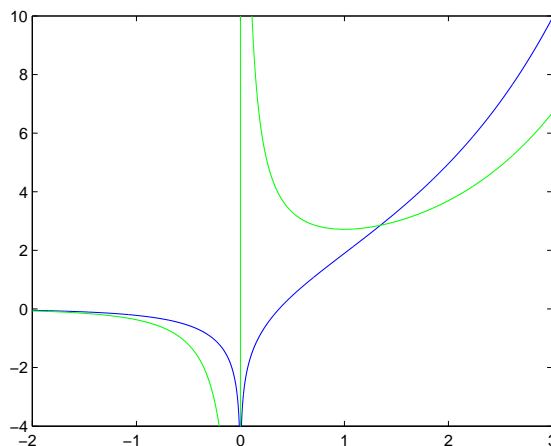
Figure 4. Graphs of the exponential integral $y = \text{Ei}(x)$ (blue) and its derivative $y = e^x/x$ (green) from Example 12.41.

Example 12.41. The exponential integral Ei is a non-elementary function defined by

$$\text{Ei}(x) = \int_{-\infty}^{x} \frac{e^t}{t}\, dt.$$

Its graph is shown in Figure 4. This integral has to be understood, in general, as an improper, principal value integral, and the function has a logarithmic singularity at $x = 0$.

If $x < 0$, then the integrand is continuous for $-\infty < t \leq x$, and the integral is interpreted as an improper integral,

$$\int_{\infty}^{x} \frac{e^t}{t}\, dt = \lim_{r \to \infty} \int_{-r}^{x} \frac{e^t}{t}\, dt.$$

This improper integral converges absolutely by comparison with e^t, since

$$\left| \frac{e^t}{t} \right| \leq e^t \qquad \text{for } -\infty < t \leq -1,$$

and

$$\int_{-\infty}^{-1} e^t\, dt = \lim_{r \to \infty} \int_{-r}^{-1} e^t\, dt = \lim_{r \to \infty} \left(e^{-r} - e^{-1} \right) = \frac{1}{e}.$$

If $x > 0$, then the integrand has a non-integrable singularity at $t = 0$, and we interpret it as a principal value integral. We write

$$\int_{-\infty}^{x} \frac{e^t}{t}\, dt = \int_{-\infty}^{-1} \frac{e^t}{t}\, dt + \int_{-1}^{x} \frac{e^t}{t}\, dt.$$

The first integral is interpreted as an improper integral as before. The second integral is interpreted as a principal value integral

$$\text{p.v.} \int_{-1}^{x} \frac{e^t}{t}\, dt = \lim_{\epsilon \to 0^+} \left(\int_{-1}^{-\epsilon} \frac{e^t}{t}\, dt + \int_{\epsilon}^{x} \frac{e^t}{t}\, dt \right).$$

This principal value integral converges, since

$$\text{p.v.} \int_{-1}^{x} \frac{e^t}{t}\, dt = \int_{-1}^{x} \frac{e^t - 1}{t}\, dt + \text{p.v.} \int_{-1}^{x} \frac{1}{t}\, dt = \int_{-1}^{x} \frac{e^t - 1}{t}\, dt + \log x.$$

The first integral makes sense as a Riemann integral since the integrand has a removable singularity at $t = 0$, with

$$\lim_{t \to 0} \left(\frac{e^t - 1}{t} \right) = 1,$$

so it extends to a continuous function on $[-1, x]$.

Finally, if $x = 0$, then the integrand is unbounded at the left endpoint $t = 0$. The corresponding improper integral diverges, and $\text{Ei}(0)$ is undefined.

The exponential integral arises in physical applications such as heat flow and radiative transfer. It is also related to the logarithmic integral

$$\text{li}(x) = \int_{0}^{x} \frac{dt}{\log t}$$

by $\text{li}(x) = \text{Ei}(\log x)$. The logarithmic integral is important in number theory, where it gives an asymptotic approximation for the number of primes less than x as $x \to \infty$.

Example 12.42. Let $f : \mathbb{R} \to \mathbb{R}$ and assume, for simplicity, that f has compact support, meaning that $f = 0$ outside a compact interval $[-r, r]$. If f is integrable, we define the Hilbert transform $Hf : \mathbb{R} \to \mathbb{R}$ of f by the principal value integral

$$Hf(x) = \frac{1}{\pi} \text{p.v.} \int_{-\infty}^{\infty} \frac{f(t)}{x - t}\, dt = \frac{1}{\pi} \lim_{\epsilon \to 0^+} \left(\int_{-\infty}^{x-\epsilon} \frac{f(t)}{x - t}\, dt + \int_{x+\epsilon}^{\infty} \frac{f(t)}{x - t}\, dt \right).$$

Here, x plays the role of a parameter in the integral with respect to t. We use a principal value because the integrand may have a non-integrable singularity at $t = x$. Since f has compact support, the intervals of integration are bounded and there is no issue with the convergence of the integrals at infinity.

For example, suppose that f is the step function

$$f(x) = \begin{cases} 1 & \text{for } 0 \le x \le 1, \\ 0 & \text{for } x < 0 \text{ or } x > 1. \end{cases}$$

If $x < 0$ or $x > 1$, then $t \ne x$ for $0 \le t \le 1$, and we get a proper Riemann integral

$$Hf(x) = \frac{1}{\pi} \int_{0}^{1} \frac{1}{x - t}\, dt = \frac{1}{\pi} \log \left| \frac{x}{x - 1} \right|.$$

If $0 < x < 1$, then we get a principal value integral

$$Hf(x) = \frac{1}{\pi} \lim_{\epsilon \to 0^+} \left(\int_0^{x-\epsilon} \frac{1}{x-t} \, dt + \frac{1}{\pi} \int_{x+\epsilon}^1 \frac{1}{x-t} \, dt \right)$$

$$= \frac{1}{\pi} \lim_{\epsilon \to 0^+} \left[\log\left(\frac{x}{\epsilon}\right) + \log\left(\frac{\epsilon}{1-x}\right) \right]$$

$$= \frac{1}{\pi} \log\left(\frac{x}{1-x}\right)$$

Thus, for $x \neq 0, 1$ we have

$$Hf(x) = \frac{1}{\pi} \log\left| \frac{x}{x-1} \right|.$$

The principal value integral with respect to t diverges if $x = 0, 1$ because $f(t)$ has a jump discontinuity at the point where $t = x$. Consequently the values $Hf(0)$, $Hf(1)$ of the Hilbert transform of the step function are undefined.

12.6. The integral test for series

An a further application of the improper integral, we prove a useful test for the convergence or divergence of a monotone decreasing, positive series. The idea is to interpret the series as an upper or lower sum of an integral.

Theorem 12.43 (Integral test). Suppose that $f : [1, \infty) \to \mathbb{R}$ is a positive decreasing function (i.e., $0 \leq f(x) \leq f(y)$ for $x \geq y$). Let $a_n = f(n)$. Then the series

$$\sum_{n=1}^{\infty} a_n$$

converges if and only if the improper integral

$$\int_1^{\infty} f(x) \, dx$$

converges. Furthermore, the limit

$$D = \lim_{n \to \infty} \left[\sum_{k=1}^{n} a_k - \int_1^n f(x) \, dx \right]$$

exists, and $0 \leq D \leq a_1$.

Proof. Let

$$S_n = \sum_{k=1}^{n} a_k, \qquad T_n = \int_1^n f(x) \, dx.$$

The integral T_n exists since f is monotone, and the sequences (S_n), (T_n) are increasing since f is positive.

Let

$$P_n = \{[1, 2], [2, 3], \ldots, [n-1, n]\}$$

be the partition of $[1, n]$ into $n - 1$ intervals of length 1. Since f is decreasing,

$$\sup_{[k, k+1]} f = a_k, \qquad \inf_{[k, k+1]} f = a_{k+1},$$

and the upper and lower sums of f on P_n are given by

$$U(f; P_n) = \sum_{k=1}^{n-1} a_k, \qquad L(f; P_n) = \sum_{k=1}^{n-1} a_{k+1}.$$

Since the integral of f on $[1, n]$ is bounded by its upper and lower sums, we get that

$$S_n - a_1 \le T_n \le S_{n-1}.$$

This inequality shows that (T_n) is bounded from above by S if $S_n \uparrow S$, and (S_n) is bounded from above by $T + a_1$ if $T_n \uparrow T$, so (S_n) converges if and only if (T_n) converges, which proves the first part of the theorem.

Let $D_n = S_n - T_n$. Then the inequality shows that $a_n \le D_n \le a_1$; in particular, (D_n) is bounded from below by zero. Moreover, since f is decreasing,

$$D_n - D_{n+1} = \int_n^{n+1} f(x)\,dx - a_{n+1} \ge f(n+1) \cdot 1 - a_{n+1} = 0,$$

so (D_n) is decreasing. Therefore $D_n \downarrow D$ where $0 \le D \le a_1$, which proves the second part of the theorem. $\qquad \square$

A basic application of this result is to the p-series.

Example 12.44. Applying Theorem 12.43 to the function $f(x) = 1/x^p$ and using Example 12.27, we find that

$$\sum_{n=1}^{\infty} \frac{1}{n^p}$$

converges if $p > 1$ and diverges if $0 < p \le 1$.

Theorem 12.43 is also useful for divergent series, since it tells us how quickly their partial sums diverge. We remark that one can obtain similar, but more accurate, asymptotic approximations than the one in theorem for the behavior of the partial sums in terms of integrals, called the Euler-MacLaurin summation formulae.

Example 12.45. Applying the second part of Theorem 12.43 to the function $f(x) = 1/x$, we find that

$$\lim_{n \to \infty} \left[\sum_{k=1}^{n} \frac{1}{k} - \log n \right] = \gamma$$

where the limit $0 \le \gamma < 1$ is the Euler constant.

Example 12.46. We can use the result of Example 12.45 to compute the sum A of the alternating harmonic series

$$A = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \dots.$$

The partial sum of the first $2m$ terms is given by

$$A_{2m} = \sum_{k=1}^{m} \frac{1}{2k-1} - \sum_{k=1}^{m} \frac{1}{2k}$$

$$= \sum_{k=1}^{2m} \frac{1}{k} - 2\sum_{k=1}^{m} \frac{1}{2k}$$

$$= \sum_{k=1}^{2m} \frac{1}{k} - \sum_{k=1}^{m} \frac{1}{k}.$$

Here, we rewrite a sum of the odd terms in the harmonic series as the difference between the harmonic series and its even terms, then use the fact that a sum of the even terms in the harmonic series is one-half the sum of the series. It follows that

$$\lim_{m\to\infty} A_{2m} = \lim_{m\to\infty} \left\{ \sum_{k=1}^{2m} \frac{1}{k} - \log 2m - \left[\sum_{k=1}^{m} \frac{1}{k} - \log m \right] + \log 2m - \log m \right\}.$$

Since $\log 2m - \log m = \log 2$, we get that

$$\lim_{m\to\infty} A_{2m} = \gamma - \gamma + \log 2 = \log 2.$$

The odd partial sums also converge to $\log 2$ since

$$A_{2m+1} = A_{2m} + \frac{1}{2m+1} \to \log 2 \qquad \text{as } m \to \infty.$$

Thus,

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = \log 2.$$

Example 12.47. A similar calculation to the previous one can be used to to compute the sum S of the rearrangement of the alternating harmonic series

$$S = 1 - \frac{1}{2} - \frac{1}{4} + \frac{1}{3} - \frac{1}{6} - \frac{1}{8} + \frac{1}{5} - \frac{1}{10} - \frac{1}{12} + \dots$$

discussed in Example 4.32. The partial sum S_{3m} of the series may be written in terms of partial sums of the harmonic series as

$$S_{3m} = 1 - \frac{1}{2} - \frac{1}{4} + \dots + \frac{1}{2m-1} - \frac{1}{4m-2} - \frac{1}{4m}$$

$$= \sum_{k=1}^{m} \frac{1}{2k-1} - \sum_{k=1}^{2m} \frac{1}{2k}$$

$$= \sum_{k=1}^{2m} \frac{1}{k} - \sum_{k=1}^{m} \frac{1}{2k} - \sum_{k=1}^{2m} \frac{1}{2k}$$

$$= \sum_{k=1}^{2m} \frac{1}{k} - \frac{1}{2}\sum_{k=1}^{m} \frac{1}{k} - \frac{1}{2}\sum_{k=1}^{2m} \frac{1}{k}.$$

It follows that

$$\lim_{m\to\infty} S_{3m} = \lim_{m\to\infty} \left\{ \sum_{k=1}^{2m} \frac{1}{k} - \log 2m - \frac{1}{2}\left[\sum_{k=1}^{m} \frac{1}{k} - \log m \right] - \frac{1}{2}\left[\sum_{k=1}^{2m} \frac{1}{k} - \log 2m \right] \right.$$
$$\left. + \log 2m - \frac{1}{2}\log - \frac{1}{2}\log 2m \right\}.$$

Since $\log 2m - \frac{1}{2}\log m - \frac{1}{2}\log 2m = \frac{1}{2}\log 2$, we get that that

$$\lim_{m\to\infty} S_{3m} = \gamma - \frac{1}{2}\gamma - \frac{1}{2}\gamma + \frac{1}{2}\log 2 = \frac{1}{2}\log 2.$$

Finally, since

$$\lim_{m\to\infty} (S_{3m+1} - S_{3m}) = \lim_{m\to\infty} (S_{3m+2} - S_{3m}) = 0,$$

we conclude that the whole series converges to $S = \frac{1}{2}\log 2$.

12.7. Taylor's theorem with integral remainder

In Theorem 8.46, we gave an expression for the error between a function and its Taylor polynomial of order n in terms of the Lagrange remainder, which involves a pointwise value of the derivative of order $n+1$ evaluated at some intermediate point. In the next theorem, we give an alternative expression for the error in terms of an integral of the derivative.

Theorem 12.48 (Taylor with integral remainder)**.** Suppose that $f : (a,b) \to \mathbb{R}$ has $n+1$ derivatives on (a,b) and f^{n+1} is Riemann integrable on every subinterval of (a,b). Let $a < c < b$. Then for every $a < x < b$,

$$f(x) = f(c) + f'(c)(x-c) + \frac{1}{2!}f''(c)(x-c)^2 + \cdots + \frac{1}{n!}f^{(n)}(c)(x-c)^n + R_n(x)$$

where

$$R_n(x) = \frac{1}{n!}\int_c^x f^{(n+1)}(t)(x-t)^n \, dt.$$

Proof. We use proof by induction. The formula is true for $n = 0$, since the fundamental theorem of calculus (Theorem 12.1) implies that

$$f(x) = f(c) + \int_c^x f'(t)\,dt = f(c) + R_0(x).$$

Assume that the formula is true for some $n \in \mathbb{N}_0$ and f^{n+2} is Riemann integrable. Then, since

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

an integration by parts with respect to t (Theorem 12.10) implies that

$$R_n(x) = -\left[\frac{1}{(n+1)!}f^{(n+1)}(t)(x-t)^{n+1} \right]_c^x + \frac{1}{(n+1)!}\int_c^x f^{(n+2)}(t)(x-t)^{n+1}\,dt$$
$$= \frac{1}{(n+1)!}f^{(n+1)}(c)(x-c)^{n+1} + R_{n+1}(x).$$

Use of this equation in the formula for n gives the formula for $n + 1$, which proves the result. $\qquad\square$

By making the change of variable

$$t = c + s(x - c),$$

we can also write the remainder as

$$R_n(x) = \frac{1}{n!} \left[\int_0^1 f^{(n+1)} \left(c + s(x - c) \right) (1 - s)^n \, ds \right] (x - c)^{n+1}.$$

In particular, if $|f^{(n+1)}(x)| \leq M$ for $a < x < b$, then

$$|R_n(x)| \leq \frac{1}{n!} M \left[\int_0^1 (1 - s)^n \, ds \right] |x - c|^{n+1}$$

$$\leq \frac{M}{(n + 1)!} |x - c|^{n+1},$$

which agrees with what one gets from the Lagrange remainder.

Thus, the integral form of the remainder is as effective as the Lagrange form in estimating its size from a uniform bound on the derivative. The integral form requires slightly stronger assumptions than the Lagrange form, since we need to assume that the derivative of order $n+1$ is integrable, but its proof is straightforward once we have the integral. Moreover, the integral form generalizes to vector-valued functions $f : (a, b) \to \mathbb{R}^n$, while the Lagrange form does not.

Metric, Normed, and Topological Spaces

A metric space is a set X that has a notion of the distance $d(x, y)$ between every pair of points $x, y \in X$. A fundamental example is $\mathbb{R}$ with the absolute-value metric $d(x, y) = |x - y|$, and nearly all of the concepts we discuss below for metric spaces are natural generalizations of the corresponding concepts for $\mathbb{R}$.

A special type of metric space that is particularly important in analysis is a normed space, which is a vector space whose metric is derived from a norm. On the other hand, every metric space is a special type of topological space, which is a set with the notion of an open set but not necessarily a distance.

The concepts of metric, normed, and topological spaces clarify our previous discussion of the analysis of real functions, and they provide the foundation for wide-ranging developments in analysis. The aim of this chapter is to introduce these spaces and give some examples, but their theory is too extensive to describe here in any detail.

13.1. Metric spaces

A metric on a set is a function that satisfies the minimal properties we might expect of a distance.

Definition 13.1. A metric d on a set X is a function $d : X \times X \to \mathbb{R}$ such that for all $x, y, z \in X$:

(1) $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$ (positivity);

(2) $d(x, y) = d(y, x)$ (symmetry);

(3) $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality).

A metric space (X, d) is a set X with a metric d defined on X.

In general, many different metrics can be defined on the same set X, but if the metric on X is clear from the context, we refer to X as a metric space.

Subspaces of a metric space are subsets whose metric is obtained by restricting the metric on the whole space.

Definition 13.2. Let (X, d) be a metric space. A metric subspace (A, d_A) of (X, d) consists of a subset $A \subset X$ whose metric $d_A : A \times A \to \mathbb{R}$ is is the restriction of d to A; that is, $d_A(x, y) = d(x, y)$ for all $x, y \in A$.

We can often formulate intrinsic properties of a subset $A \subset X$ of a metric space X in terms of properties of the corresponding metric subspace (A, d_A).

When it is clear that we are discussing metric spaces, we refer to a metric subspace as a subspace, but metric subspaces should not be confused with other types of subspaces (for example, vector subspaces of a vector space).

13.1.1. Examples. In the following examples of metric spaces, the verification of the properties of a metric is mostly straightforward and is left as an exercise.

Example 13.3. A rather trivial example of a metric on any set X is the discrete metric

$$d(x, y) = \begin{cases} 0 & \text{if } x = y, \\ 1 & \text{if } x \neq y. \end{cases}$$

This metric is nevertheless useful in illustrating the definitions and providing counter-examples.

Example 13.4. Define $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ by

$$d(x, y) = |x - y|.$$

Then d is a metric on $\mathbb{R}$. The natural numbers $\mathbb{N}$ and the rational numbers $\mathbb{Q}$ with the absolute-value metric are metric subspaces of $\mathbb{R}$, as is any other subset $A \subset \mathbb{R}$.

Example 13.5. Define $d : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ by

$$d(x, y) = |x_1 - y_1| + |x_2 - y_2| \qquad x = (x_1, x_2), \quad y = (y_1, y_2).$$

Then d is a metric on $\mathbb{R}^2$, called the ℓ^1 metric. (Here, "ℓ^1" is pronounced "ell-one.") For example, writing $z = (z_1, z_2)$, we have

$$\begin{aligned} d(x, y) &= |x_1 - z_1 + z_1 - y_1| + |x_2 - z_2 + z_2 - y_2| \\ &\leq |x_1 - z_1| + |z_1 - y_1| + |x_2 - z_2| + |z_2 - y_2| \\ &\leq d(x, z) + d(z, y), \end{aligned}$$

so d satisfies the triangle inequality. This metric is sometimes referred to informally as the "taxicab" metric, since it's the distance one would travel by taxi on a rectangular grid of streets.

Example 13.6. Define $d : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ by

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \qquad x = (x_1, x_2), \quad y = (y_1, y_2).$$

Then d is a metric on $\mathbb{R}^2$, called the Euclidean, or ℓ^2, metric. It corresponds to the usual notion of distance between points in the plane. The triangle inequality is geometrically obvious but an analytical proof is non-trivial (see Theorem 13.26 below).
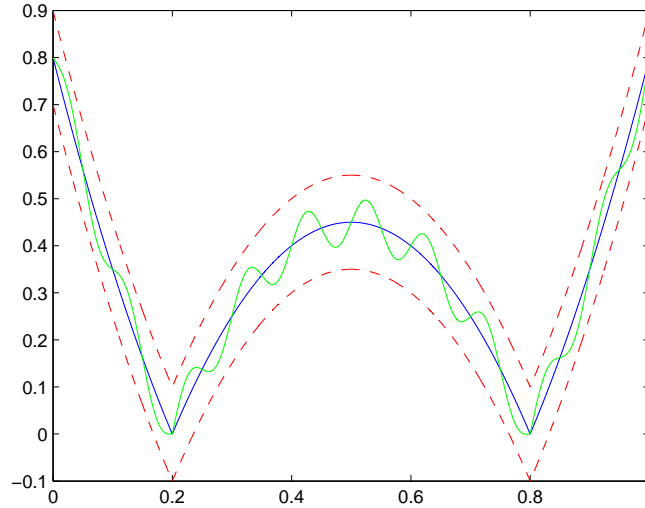
Figure 1. The graph of a function $f \in C([0,1])$ is in blue. A function whose distance from f with respect to the sup-norm is less than 0.1 has a graph that lies inside the dotted red lines $y = f(x) \pm 0.1$ e.g., the green graph.

Example 13.7. Define $d : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ by

$$d(x,y) = \max\left(|x_1 - y_1|, |x_2 - y_2|\right) \qquad x = (x_1, x_2), \quad y = (y_1, y_2).$$

Then d is a metric on $\mathbb{R}^2$, called the ℓ^∞, or maximum, metric.

Example 13.8. Define $d : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ for $x = (x_1, x_2)$, $y = (y_1, y_2)$ as follows: if $(x_1, x_2) \neq k(y_1, y_2)$ for $k \in \mathbb{R}$, then

$$d(x,y) = \sqrt{x_1^2 + x_2^2} + \sqrt{y_1^2 + y_2^2};$$

and if $(x_1, x_2) = k(y_1, y_2)$ for some $k \in \mathbb{R}$, then

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}.$$

That is, $d(x,y)$ is the sum of the Euclidean distances of x and y from the origin, unless x and y lie on the same line through the origin, in which case it is the Euclidean distance from x to y. Then d defines a metric on $\mathbb{R}^2$.

In England, d is sometimes called the "British Rail" metric, because all the train lines radiate from London (located at 0). To take a train from town x to town y, one has to take a train from x to 0 and then take a train from 0 to y, unless x and y are on the same line, when one can take a direct train.

Example 13.9. Let $C(K)$ denote the set of continuous functions $f : K \to \mathbb{R}$, where $K \subset \mathbb{R}$ is compact; for example, $K = [a, b]$ is a closed, bounded interval. If $f, g \in C(K)$ define

$$d(f,g) = \sup_{x \in K} |f(x) - g(x)| = \|f - g\|_\infty, \qquad \|f\|_\infty = \sup_{x \in K} |f(x)|.$$
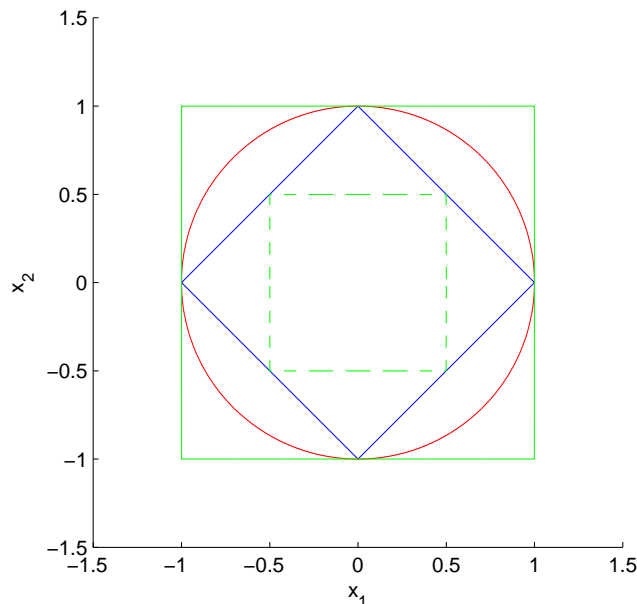
Figure 2. The unit balls $B_1(0)$ on $\mathbb{R}^2$ for different metrics: they are the interior of a diamond (ℓ^1-norm), a circle (ℓ^2-norm), or a square (ℓ^∞-norm). The ℓ^∞-ball of radius $1/2$ is also indicated by the dashed line.

The function $d : C(K) \times C(K) \to \mathbb{R}$ is well-defined, since a continuous function on a compact set is bounded, and d is a metric on $C(K)$. Two functions are close with respect to this metric if their values are close at every point $x \in K$. (See Figure 1.) We refer to $\|f\|_\infty$ as the sup-norm of f. Section 13.6 has further discussion.

13.1.2. Open and closed balls. A ball in a metric space is analogous to an interval in $\mathbb{R}$.

Definition 13.10. Let (X, d) be a metric space. The open ball $B_r(x)$ of radius $r > 0$ and center $x \in X$ is the set of points whose distance from x is less than r,

$$B_r(x) = \{y \in X : d(x, y) < r\}.$$

The closed ball $\bar{B}_r(x)$ of radius $r > 0$ and center $x \in X$ as the set of points whose distance from x is less than or equal to r,

$$\bar{B}_r(x) = \{y \in X : d(x, y) \le r\}.$$

The term "ball" is used to denote a "solid ball," rather than the "sphere" of points whose distance from the center x is equal to r.

Example 13.11. Consider $\mathbb{R}$ with its standard absolute-value metric, defined in Example 13.4. Then the open ball $B_r(x) = \{y \in \mathbb{R} : |x - y| < r\}$ is the open interval of radius r centered at x, and the closed ball $\bar{B}_r(x) = \{y \in \mathbb{R} : |x - y| \le r\}$ is the closed interval of radius r centered at x.

Example 13.12. For $\mathbb{R}^2$ with the Euclidean metric defined in Example 13.6, the ball $B_r(x)$ is an open disc of radius r centered at x. For the ℓ^1-metric in Example 13.5, the ball is a diamond of diameter $2r$, and for the ℓ^∞-metric in Example 13.7, it is a square of side $2r$. The unit ball $B_1(0)$ for each of these metrics is illustrated in Figure 2.

Example 13.13. Consider the space $C(K)$ of continuous functions $f : K \to \mathbb{R}$ on a compact set $K \subset \mathbb{R}$ with the sup-norm metric defined in Example 13.9. The ball $B_r(f)$ consists of all continuous functions $g : K \to \mathbb{R}$ whose values are within r of the values of f at every $x \in K$. For example, for the function f shown in Figure 1 with $r = 0.1$, the open ball $B_r(f)$ consists of all continuous functions g whose graphs lie between the red lines.

One has to be a little careful with the notion of balls in a general metric space, because they don't always behave the way their name suggests.

Example 13.14. Let X be a set with the discrete metric given in Example 13.3. Then $B_r(x) = \{x\}$ consists of a single point if $0 \le r < 1$ and $B_r(x) = X$ is the whole space if $r \ge 1$. (See also Example 13.44.)

An another example, what are the open balls for the metric in Example 13.8?

A set in a metric space is bounded if it is contained in a ball of finite radius.

Definition 13.15. Let (X, d) be a metric space. A set $A \subset X$ is bounded if there exist $x \in X$ and $0 \le R < \infty$ such that $d(x, y) \le R$ for all $y \in A$, meaning that $A \subset B_R(x)$.

Unlike $\mathbb{R}$, or a vector space, a general metric space has no distinguished origin, but the center point of the ball is not important in this definition of a bounded set. The triangle inequality implies that $d(y, z) < R + d(x, y)$ if $d(x, z) < R$, so

$$B_R(x) \subset B_{R'}(y) \qquad \text{for } R' = R + d(x, y).$$

Thus, if Definition 13.15 holds for some $x \in X$, then it holds for every $x \in X$.

We can say equivalently that $A \subset X$ is bounded if the metric subspace (A, d_A) is bounded.

Example 13.16. Let X be a set with the discrete metric given in Example 13.3. Then X is bounded since $X = B_r(x)$ if $r > 1$ and $x \in X$.

Example 13.17. A subset $A \subset \mathbb{R}$ is bounded with respect to the absolute-value metric if $A \subset (-R, R)$ for some $0 < R < \infty$.

Example 13.18. Let $C(K)$ be the space of continuous functions $f : K \to \mathbb{R}$ on a compact set defined in Example 13.9. The set $\mathcal{F} \subset C(K)$ of all continuous functions $f : K \to \mathbb{R}$ such that $|f(x)| \le 1$ for every $x \in K$ is bounded, since $d(f, 0) = \|f\|_\infty \le 1$ for all $f \in \mathcal{F}$. The set of constant functions $\{f : f(x) = c \text{ for all } x \in K\}$ isn't bounded, since $\|f\|_\infty = |c|$ may be arbitrarily large.

We define the diameter of a set in an analogous way to Definition 3.5 for subsets of $\mathbb{R}$.

Definition 13.19. Let (X, d) be a metric space and $A \subset X$. The diameter $0 \leq \operatorname{diam} A \leq \infty$ of A is

$$\operatorname{diam} A = \sup \left\{ d(x, y) : x, y \in A \right\}.$$

It follows from the definitions that A is bounded if and only if $\operatorname{diam} A < \infty$.

The notions of an upper bound, lower bound, supremum, and infimum in $\mathbb{R}$ depend on its order properties. Unlike properties of $\mathbb{R}$ based on the absolute value, they do not generalize to an arbitrary metric space, which isn't equipped with an order relation.

13.2. Normed spaces

In general, there are no algebraic operations defined on a metric space, only a distance function. Most of the spaces that arise in analysis are vector, or linear, spaces, and the metrics on them are usually derived from a norm, which gives the "length" of a vector.

We assume that the reader is familiar with the basic theory of vector spaces, and we consider only real vector spaces.

Definition 13.20. A normed vector space $(X, \| \cdot \|)$ is a vector space X together with a function $\| \cdot \| : X \to \mathbb{R}$, called a norm on X, such that for all $x, y \in X$ and $k \in \mathbb{R}$:

(1) $0 \leq \|x\| < \infty$ and $\|x\| = 0$ if and only if $x = 0$;

(2) $\|kx\| = |k| \|x\|$;

(3) $\|x + y\| \leq \|x\| + \|y\|$.

The properties in Definition 13.20 are natural ones to require of a length. The length of x is 0 if and only if x is the 0-vector; multiplying a vector by a scalar k multiplies its length by $|k|$; and the length of the "hypoteneuse" $x + y$ is less than or equal to the sum of the lengths of the "sides" x, y. Because of this last interpretation, property (3) is called the triangle inequality. We also refer to a normed vector space as a normed space for short.

Proposition 13.21. If $(X, \| \cdot \|)$ is a normed vector space, then $d : X \times X \to \mathbb{R}$ defined by $d(x, y) = \|x - y\|$ is a metric on X.

Proof. The metric-properties of d follow directly from the properties of a norm in Definition 13.20. The positivity is immediate. Also, we have

$$d(x, y) = \|x - y\| = \| - (x - y)\| = \|y - x\| = d(y, x),$$
$$d(x, y) = \|x - z + z - y\| \leq \|x - z\| + \|z - y\| = d(x, z) + d(y, z),$$

which proves the symmetry of d and the triangle inequality. $\qquad \square$

If X is a normed vector space, then we always use the metric associated with its norm, unless stated specifically otherwise.

A metric associated with a norm has the additional properties that for all $x, y, z \in X$ and $k \in \mathbb{R}$

$$d(x + z, y + z) = d(x, y), \qquad d(kx, ky) = |k| d(x, y),$$

which are called translation invariance and homogeneity, respectively. These properties imply that the open balls $B_r(x)$ in a normed space are rescaled, translated versions of the unit ball $B_1(0)$.

Example 13.22. The set of real numbers $\mathbb{R}$ with the absolute-value norm $|\cdot|$ is a one-dimensional normed vector space.

Example 13.23. The discrete metric in Example 13.3 on $\mathbb{R}$, and the metric in Example 13.8 on $\mathbb{R}^2$ are not derived from a norm. (Why?)

Example 13.24. The space $\mathbb{R}^2$ with any of the norms defined for $x = (x_1, x_2)$ by

$$\|x\|_1 = |x_1| + |x_2|, \qquad \|x\|_2 = \sqrt{x_1^2 + x_2^2}, \qquad \|x\|_\infty = \max\left(|x_1|, |x_2|\right)$$

is a normed vector space. The corresponding metrics are the "taxicab" metric in Example 13.5, the Euclidean metric in Example 13.6, and the maximum metric in Example 13.7, respectively.

The norms in Example 13.24 are special cases of a fundamental family of ℓ^p-norms on $\mathbb{R}^n$. All of the ℓ^p-norms reduce to the absolute-value norm if $n = 1$, but they are different if $n \geq 2$.

Definition 13.25. For $1 \leq p < \infty$, the ℓ^p-norm $\|\cdot\|_p : \mathbb{R}^n \to \mathbb{R}$ is defined for $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ by

$$\|x\|_p = \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{1/p}.$$

The ℓ^2-norm is called the Euclidean norm. For $p = \infty$, the ℓ^∞-norm $\|\cdot\|_\infty : \mathbb{R}^n \to \mathbb{R}$ is defined by

$$\|x\|_\infty = \max\left(|x_1|, |x_2|, \ldots, |x_n|\right).$$

The notation for the ℓ^∞-norm is explained by the fact that

$$\|x\|_\infty = \lim_{p \to \infty} \|x\|_p.$$

Moreover, consistent with its name, the ℓ^p-norm is a norm.

Theorem 13.26. Let $1 \leq p \leq \infty$. The space $\mathbb{R}^n$ with the ℓ^p-norm is a normed vector space.

Proof. The space $\mathbb{R}^n$ is an n-dimensional vector space, so we just need to verify the properties of the norm.

The positivity and homogeneity of the ℓ^p-norm follow immediately from its definition. We verify the triangle inequality here only for the cases $p = 1, \infty$.

Let $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ be points in $\mathbb{R}^n$. For $p = 1$, we have

$$\begin{aligned}
\|x + y\|_1 &= |x_1 + y_1| + |x_2 + y_2| + \cdots + |x_n + y_n| \\
&\leq |x_1| + |y_1| + |x_2| + |y_2| + \cdots + |x_n| + |y_n| \\
&\leq \|x\|_1 + \|y\|_1.
\end{aligned}$$

For $p = \infty$, we have

$$\|x + y\|_\infty = \max\left(|x_1 + y_1|, |x_2 + y_2|, \ldots, |x_n + y_n|\right)$$
$$\leq \max\left(|x_1| + |y_1|, |x_2| + |y_2|, \ldots, |x_n| + |y_n|\right)$$
$$\leq \max\left(|x_1|, |x_2|, \ldots, |x_n|\right) + \max\left(|y_1|, |y_2|, \ldots, |y_n|\right)$$
$$\leq \|x\|_\infty + \|y\|_\infty.$$

The proof of the triangle inequality for $1 < p < \infty$ is more difficult and is given in Section 13.7.   □

We can use Definition 13.25 to define $\|x\|_p$ for any $0 < p \leq \infty$. However, if $0 < p < 1$, then $\|\cdot\|_p$ doesn't satisfy the triangle inequality, so it is not a norm. This explains the restriction $1 \leq p \leq \infty$.

Although the ℓ^p-norms are numerically different for different values of p, they are equivalent in the following sense (see Corollary 13.29).

Definition 13.27. Let X be a vector space. Two norms $\|\cdot\|_a$, $\|\cdot\|_b$ on X are equivalent if there exist strictly positive constants $M \geq m > 0$ such that

$$m\|x\|_a \leq \|x\|_b \leq M\|x\|_a \qquad \text{for all } x \in X.$$

Geometrically, two norms are equivalent if and only if an open ball with respect to either one of the norms contains an open ball with respect to the other. Equivalent norms define the same open sets, convergent sequences, and continuous functions, so there are no topological differences between them.

The next theorem shows that every ℓ^p-norm is equivalent to the ℓ^∞-norm. (See Figure 2.)

Theorem 13.28. Suppose that $1 \leq p < \infty$. Then, for every $x \in \mathbb{R}^n$,

$$\|x\|_\infty \leq \|x\|_p \leq n^{1/p}\|x\|_\infty.$$

Proof. Let $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$. Then for each $1 \leq i \leq n$, we have

$$|x_i| \leq \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{1/p} = \|x\|_p,$$

which implies that

$$\|x\|_\infty = \max\left\{|x_i| : 1 \leq i \leq n\right\} \leq \|x\|_p.$$

On the other hand, since $|x_i| \leq \|x\|_\infty$ for every $1 \leq i \leq n$, we have

$$\|x\|_p \leq \left(n\|x\|_\infty^p\right)^{1/p} = n^{1/p}\|x\|_\infty,$$

which proves the result   □

As an immediate consequence, we get the equivalence of the ℓ^p-norms.

Corollary 13.29. The ℓ^p and ℓ^q norms on $\mathbb{R}^n$ are equivalent for every $1 \leq p, q \leq \infty$.

Proof. We have $n^{-1/q}\|x\|_q \leq \|x\|_\infty \leq \|x\|_p \leq n^{1/p}\|x\|_\infty \leq n^{1/p}\|x\|_q$.   □

With more work, one can prove that that $\|x\|_q \leq \|x\|_p$ for $1 \leq p \leq q \leq \infty$, meaning that the unit ball with respect to the ℓ^q-norm contains the unit ball with respect to the ℓ^p-norm.

13.3. Open and closed sets

There are natural definitions of open and closed sets in a metric space, analogous to the definitions in $\mathbb{R}$. Many of the properties of such sets in $\mathbb{R}$ carry over immediately to general metric spaces.

Definition 13.30. Let X be a metric space. A set $G \subset X$ is open if for every $x \in G$ there exists $r > 0$ such that $B_r(x) \subset G$. A subset $F \subset X$ is closed if $F^c = X \setminus F$ is open.

We can rephrase this definition more geometrically in terms of neighborhoods.

Definition 13.31. Let X be a metric space. A set $U \subset X$ is a neighborhood of $x \in X$ if $B_r(x) \subset U$ for some $r > 0$.

Definition 13.30 then states that a subset of a metric space is open if and only if every point in the set has a neighborhood that is contained in the set. In particular, a set is open if and only if it is a neighborhood of every point in the set.

Example 13.32. If d is the discrete metric on a set X in Example 13.3, then every subset $A \subset X$ is open, since for every $x \in A$ we have $B_{1/2}(x) = \{x\} \subset A$. Every set is also closed, since its complement is open.

Example 13.33. Consider $\mathbb{R}^2$ with the Euclidean norm (or any other ℓ^p-norm). If $f : \mathbb{R} \to \mathbb{R}$ is a continuous function, then
$$E = \big\{ (x_1, x_2) \in \mathbb{R}^2 : x_2 < f(x_1) \big\}$$
is an open subset of $\mathbb{R}^2$. If f is discontinuous, then E needn't be open. We leave the proofs as an exercise.

Example 13.34. If (X, d) is a metric space and $A \subset X$, then $B \subset A$ is open in the metric subspace (A, d_A) if and only if $B = A \cap G$ where G is an open subset of X. This is consistent with our previous definition of relatively open sets in $A \subset \mathbb{R}$.

Open sets with respect to one metric on a set need not be open with respect to another metric. For example, every subset of $\mathbb{R}$ with the discrete metric is open, but this is not true of $\mathbb{R}$ with the absolute-value metric.

Consistent with our terminology, open balls are open and closed balls are closed.

Proposition 13.35. Let X be a metric space. If $x \in X$ and $r > 0$, then the open ball $B_r(x)$ is open and the closed ball $\bar{B}_r(x)$ is closed.

Proof. Suppose that $y \in B_r(x)$ where $r > 0$, and let $\epsilon = r - d(x, y) > 0$. The triangle inequality implies that $B_\epsilon(y) \subset B_r(x)$, which proves that $B_r(x)$ is open. Similarly, if $y \in \bar{B}_r(x)^c$ and $\epsilon = d(x, y) - r > 0$, then the triangle inequality implies that $B_\epsilon(y) \subset \bar{B}_r(x)^c$, which proves that $\bar{B}_r(x)^c$ is open and $\bar{B}_r(x)$ is closed. $\square$

The next theorem summarizes three basic properties of open sets.

Theorem 13.36. Let X be a metric space.

(1) The empty set $\varnothing$ and the whole set X are open.
(2) An arbitrary union of open sets is open.

(3) A finite intersection of open sets is open.

Proof. Property (1) follows immediately from Definition 13.30. (The empty set satisfies the definition vacuously: since it has no points, every point has a neighborhood that is contained in the set.)

To prove (2), let $\{G_i \subset X : i \in I\}$ be an arbitrary collection of open sets. If

$$x \in \bigcup_{i \in I} G_i,$$

then $x \in G_i$ for some $i \in I$. Since G_i is open, there exists $r > 0$ such that $B_r(x) \subset G_i$, and then

$$B_r(x) \subset \bigcup_{i \in I} G_i.$$

Thus, the union $\bigcup G_i$ is open.

The prove (3), let $\{G_1, G_2, \ldots, G_n\}$ be a finite collection of open sets. If

$$x \in \bigcap_{i=1}^{n} G_i,$$

then $x \in G_i$ for every $1 \le i \le n$. Since G_i is open, there exists $r_i > 0$ such that $B_{r_i}(x) \subset G_i$. Let $r = \min(r_1, r_2, \ldots, r_n) > 0$. Then $B_r(x) \subset B_{r_i}(x) \subset G_i$ for every $1 \le i \le n$, which implies that

$$B_r(x) \subset \bigcap_{i=1}^{n} G_i.$$

Thus, the finite intersection $\bigcap G_i$ is open. $\square$

The previous proof fails if we consider the intersection of infinitely many open sets $\{G_i : i \in I\}$ because we may have $\inf\{r_i : i \in I\} = 0$ even though $r_i > 0$ for every $i \in I$.

The properties of closed sets follow by taking complements of the corresponding properties of open sets and using De Morgan's laws, exactly as in the proof of Proposition 5.20.

Theorem 13.37. Let X be a metric space.

(1) The empty set $\varnothing$ and the whole set X are closed.

(2) An arbitrary intersection of closed sets is closed.

(3) A finite union of closed sets is closed.

The following relationships of points to sets are entirely analogous to the ones in Definition 5.22 for $\mathbb{R}$.

Definition 13.38. Let X be a metric space and $A \subset X$.

(1) A point $x \in A$ is an interior point of A if $B_r(x) \subset A$ for some $r > 0$.

(2) A point $x \in A$ is an isolated point of A if $B_r(x) \cap A = \{x\}$ for some $r > 0$, meaning that x is the only point of A that belongs to $B_r(x)$.

(3) A point $x \in X$ is a boundary point of A if, for every $r > 0$, the ball $B_r(x)$ contains a point in A and a point not in A.

(4) A point $x \in X$ is an accumulation point of A if, for every $r > 0$, the ball $B_r(x)$ contains a point $y \in A$ such that $y \neq x$.

A set is open if and only if every point is an interior point and closed if and only if every accumulation point belongs to the set.

We define the interior, boundary, and closure of a set as follows.

Definition 13.39. Let A be a subset of a metric space. The interior A° of A is the set of interior points of A. The boundary ∂A of A is the set of boundary points. The closure of A is $\bar{A} = A \cup \partial A$.

It follows that $x \in \bar{A}$ if and only if the ball $B_r(x)$ contains some point in A for every $r > 0$. The next proposition gives equivalent topological definitions.

Proposition 13.40. Let X be a metric space and $A \subset X$. The interior of A is the largest open set contained in A,

$$A^\circ = \bigcup \{G \subset A : G \text{ is open in } X\},$$

the closure of A is the smallest closed set that contains A,

$$\bar{A} = \bigcap \{F \supset A : F \text{ is closed in } X\},$$

and the boundary of A is their set-theoretic difference,

$$\partial A = \bar{A} \setminus A^\circ.$$

Proof. Let A_1 denote the set of interior points of A, as in Definition 13.38, and $A_2 = \bigcup \{G \subset A : G \text{ is open}\}$. If $x \in A_1$, then there is an open neighborhood $G \subset A$ of x, so $G \subset A_2$ and $x \in A_2$. It follows that $A_1 \subset A_2$. To get the opposite inclusion, note that A_2 is open by Theorem 13.36. Thus, if $x \in A_2$, then $A_2 \subset A$ is a neighborhood of x, so $x \in A_1$ and $A_2 \subset A_1$. Therefore $A_1 = A_2$, which proves the result for the interior.

Next, Definition 13.38 and the previous result imply that

$$(\bar{A})^c = (A^c)^\circ = \bigcup \{G \subset A^c : G \text{ is open}\}.$$

Using De Morgan's laws, and writing $G^c = F$, we get that

$$\bar{A} = \bigcup \{G \subset A^c : G \text{ is open}\}^c = \bigcap \{F \supset A : F \text{ is closed}\},$$

which proves the result for the closure.

Finally, if $x \in \partial A$, then $x \in \bar{A} = A \cup \partial A$, and no neighborhood of x is contained in A, so $x \notin A^\circ$. It follows that $x \in \bar{A} \setminus A^\circ$ and $\partial A \subset \bar{A} \setminus A^\circ$. Conversely, if $x \in \bar{A} \setminus A^\circ$, then every neighborhood of x contains points in A, since $x \in \bar{A}$, and every neighborhood contains points in A^c, since $x \notin A^\circ$. It follows that $x \in \partial A$ and $\bar{A} \setminus A^\circ \subset \partial A$, which completes the proof. $\square$

It follows from Theorem 13.36, and Theorem 13.37 that the interior A° is open, the closure $\bar{A}$ is closed, and the boundary ∂A is closed. Furthermore, A is open if and only if $A = A^\circ$, and A is closed if and only if $A = \bar{A}$.

Let us illustrate these definitions with some examples, whose verification we leave as an exercise.

Example 13.41. Consider $\mathbb{R}$ with the absolute-value metric. If $I = (a, b)$ and $J = [a, b]$, then $I^\circ = J^\circ = (a, b)$, $\bar{I} = \bar{J} = [a, b]$, and $\partial I = \partial J = \{a, b\}$. Note that $I = I^\circ$, meaning that I is open, and $J = \bar{J}$, meaning that J is closed. If $A = \{1/n : n \in \mathbb{N}\}$, then $A^\circ = \varnothing$ and $\bar{A} = \partial A = A \cup \{0\}$. Thus, A is neither open ($A \neq A^\circ$) nor closed ($A \neq \bar{A}$). If $\mathbb{Q}$ is the set of rational numbers, then $\mathbb{Q}^\circ = \varnothing$ and $\bar{\mathbb{Q}} = \partial \mathbb{Q} = \mathbb{R}$. Thus, $\mathbb{Q}$ is neither open nor closed. Since $\bar{\mathbb{Q}} = \mathbb{R}$, we say that $\mathbb{Q}$ is dense in $\mathbb{R}$.

Example 13.42. Let A be the unit open ball in $\mathbb{R}^2$ with the Euclidean metric,

$$A = \left\{ (x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1 \right\}.$$

Then $A^\circ = A$, the closure of A is the closed unit ball

$$\bar{A} = \left\{ (x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1 \right\},$$

and the boundary of A is the unit circle

$$\partial A = \left\{ (x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1 \right\},$$

Example 13.43. Let A be the unit open ball with the x-axis deleted in $\mathbb{R}^2$ with the Euclidean metric,

$$A = \left\{ (x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1, \, y \neq 0 \right\}.$$

Then $A^\circ = A$, the closure of A is the closed unit ball

$$\bar{A} = \left\{ (x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1 \right\},$$

and the boundary of A consists of the unit circle and the x-axis,

$$\partial A = \left\{ (x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1 \right\} \cup \left\{ (x, 0) \in \mathbb{R}^2 : |x| \leq 1 \right\}.$$

Example 13.44. Suppose that X is a set containing at least two elements with the discrete metric defined in Example 13.3. If $x \in X$, then the unit open ball is $B_1(x) = \{x\}$, and it is equal to its closure $\overline{B_r(x)} = \{x\}$. On the other hand, the closed unit ball is $\bar{B}_1(x) = X$. Thus, in a general metric space, the closure of an open ball of radius $r > 0$ need not be the closed ball of radius r.

13.4. Completeness, compactness, and continuity

A sequence (x_n) in a set X is a function $f : \mathbb{N} \to X$, where $x_n = f(n)$ is the nth term in the sequence.

Definition 13.45. Let (X, d) be a metric space. A sequence (x_n) in X converges to $x \in X$, written $x_n \to x$ as $n \to \infty$ or

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

if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$n > N \text{ implies that } d(x_n, x) < \epsilon.$$

That is, $x_n \to x$ in X if $d(x_n, x) \to 0$ in $\mathbb{R}$. Equivalently, $x_n \to x$ as $n \to \infty$ if for every neighborhood U of x there exists $N \in \mathbb{N}$ such that $x_n \in U$ for all $n > N$.

Example 13.46. If d is the discrete metric on a set X, then a sequence (x_n) converges in (X, d) if and only if it is eventually constant. That is, there exists $x \in X$ and $N \in \mathbb{N}$ such that $x_n = x$ for all $n > N$; and, in that case, the sequence converges to x.

Example 13.47. For $\mathbb{R}$ with its standard absolute-value metric, Definition 13.45 is the definition of the convergence of a real sequence.

As for subsets of $\mathbb{R}$, we can give a sequential characterization of closed sets in a metric space.

Theorem 13.48. A subset $F \subset X$ of a metric space X is closed if and only if the limit of every convergent sequence in F belongs to F.

Proof. First suppose that F is closed, meaning that F^c is open. If (x_n) be a sequence in F and $x \in F^c$, then there is a neighborhood $U \subset F^c$ of x which contains no terms in the sequence, so (x_n) cannot converge to x. Thus, the limit of every convergent sequence in F belongs to F.

Conversely, suppose that F is not closed. Then F^c is not open, and there exists a point $x \in F^c$ such that every neighborhood of x contains points in F. Choose $x_n \in F$ such that $x_n \in B_{1/n}(x)$. Then (x_n) is a sequence in F whose limit x does not belong to F, which proves the result. □

We define the completeness of metric spaces in terms of Cauchy sequences.

Definition 13.49. Let (X, d) be a metric space. A sequence (x_n) in X is a Cauchy sequence for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$m, n > N \text{ implies that } d(x_m, x_n) < \epsilon.$$

Every convergent sequence is Cauchy: if $x_n \to x$ then given $\epsilon > 0$ there exists N such that $d(x_n, x) < \epsilon/2$ for all $n > N$, and then for all $m, n > N$ we have

$$d(x_m, x_n) \le d(x_m, x) + d(x_n, x) < \epsilon.$$

Complete spaces are ones in which the converse is also true.

Definition 13.50. A metric space is complete if every Cauchy sequence converges.

Example 13.51. If d is the discrete metric on a set X, then (X, d) is a complete metric space since every Cauchy sequence is eventually constant.

Example 13.52. The space $(\mathbb{R}, |\cdot|)$ is complete, but the metric subspace $(\mathbb{Q}, |\cdot|)$ is not complete.

In a complete space, we have the following simple criterion for the completeness of a subspace.

Proposition 13.53. A subspace (A, d_A) of a complete metric space (X, d) is complete if and only if A is closed in X.

Proof. If A is a closed subspace of a complete space X and (x_n) is a Cauchy sequence in A, then (x_n) is a Cauchy sequence in X, so it converges to $x \in X$. Since A is closed, $x \in A$, which shows that A is complete.

Conversely, if A is not closed, then by Proposition 13.48 there is a convergent sequence in A whose limit does not belong to A. Since it converges, the sequence is Cauchy, but it doesn't have a limit in A, so A is not complete. $\square$

The most important complete metric spaces in analysis are the complete normed spaces, or Banach spaces.

Definition 13.54. A Banach space is a complete normed vector space.

For example, $\mathbb{R}$ with the absolute-value norm is a one-dimensional Banach space. Furthermore, it follows from the completeness of $\mathbb{R}$ that every finite-dimensional normed vector space over $\mathbb{R}$ is complete. We prove this for the ℓ^p-norms given in Definition 13.25.

Theorem 13.55. Let $1 \le p \le \infty$. The vector space $\mathbb{R}^n$ with the ℓ^p-norm is a Banach space.

Proof. Suppose that $(x_k)_{k=1}^\infty$ is a sequence of points
$$x_k = (x_{1,k}, x_{2,k}, \ldots, x_{n,k})$$
in $\mathbb{R}^n$ that is Cauchy with respect to the ℓ^p-norm. From Theorem 13.28,
$$|x_{i,j} - x_{i,k}| \le \|x_j - x_k\|_p,$$
so each coordinate sequence $(x_{i,k})_{k=1}^\infty$ is Cauchy in $\mathbb{R}$. The completeness of $\mathbb{R}$ implies that $x_{i,k} \to x_i$ as $k \to \infty$ for some $x_i \in \mathbb{R}$. Let $x = (x_1, x_2, \ldots, x_n)$. Then, from Theorem 13.28 again,
$$\|x_k - x\|_p \le C \max \{|x_{i,k} - x_i| : i = 1, 2, \ldots, n\},$$
where $C = n^{1/p}$ if $1 \le p < \infty$ or $C = 1$ if $p = \infty$. Given $\epsilon > 0$, choose $N_i \in \mathbb{N}$ such that $|x_{i,k} - x_i| < \epsilon/C$ for all $k > N_i$, and let $N = \max\{N_1, N_2, \ldots, N_n\}$. Then $k > N$ implies that $\|x_k - x\|_p < \epsilon$, which proves that $x_k \to x$ with respect to the ℓ^p-norm. Thus, $(\mathbb{R}^n, \|\cdot\|_p)$ is complete. $\square$

The Bolzano-Weierstrass property provides a sequential definition of compactness in a general metric space.

Definition 13.56. A subset $K \subset X$ of a metric space X is sequentially compact, or compact for short, if every sequence in K has a convergent subsequence whose limit belongs to K.

Explicitly, this condition means that if (x_n) is a sequence of points $x_n \in K$ then there is a subsequence (x_{n_k}) such that $x_{n_k} \to x$ as $k \to \infty$ where $x \in K$. Compactness is an intrinsic property of a subset: $K \subset X$ is compact if and only if the metric subspace (K, d_K) is compact.

Although this definition is similar to the one for compact sets in $\mathbb{R}$, there is a significant difference between compact sets in a general metric space and in $\mathbb{R}$. Every compact subset of a metric space is closed and bounded, as in $\mathbb{R}$, but it is not always true that a closed, bounded set is compact.

First, as the following example illustrates, a set must be complete, not just closed, to be compact. (A closed subset of $\mathbb{R}$ is complete because $\mathbb{R}$ is complete.)

Example 13.57. Consider the metric space $\mathbb{Q}$ with the absolute value norm. The set $[0,2] \cap \mathbb{Q}$ is a closed, bounded subspace, but it is not compact since a sequence of rational numbers that converges in $\mathbb{R}$ to an irrational number such as $\sqrt{2}$ has no convergent subsequence in $\mathbb{Q}$.

Second, completeness and boundedness is not enough, in general, to imply compactness.

Example 13.58. Consider $\mathbb{N}$, or any other infinite set, with the discrete metric,

$$d(m,n) = \begin{cases} 0 & \text{if } m = n, \\ 1 & \text{if } m \neq n. \end{cases}$$

Then $\mathbb{N}$ is complete and bounded with respect to this metric. However, it is not compact since $x_n = n$ is a sequence with no convergent subsequence, as is clear from Example 13.46.

The correct generalization to an arbitrary metric space of the characterization of compact sets in $\mathbb{R}$ as closed and bounded replaces "closed" with "complete" and "bounded" with "totally bounded," which is defined as follows.

Definition 13.59. Let (X,d) be a metric space. A subset $A \subset X$ is totally bounded if for every $\epsilon > 0$ there exists a finite set $\{x_1, x_2, \ldots, x_n\}$ of points in X such that

$$A \subset \bigcup_{i=1}^{n} B_\epsilon(x_i).$$

The proof of the following result is then completely analogous to the proof of the Bolzano-Weierstrass theorem in Theorem 3.57 for $\mathbb{R}$.

Theorem 13.60. A subset $K \subset X$ of a metric space X is sequentially compact if and only if it is is complete and totally bounded.

The definition of the continuity of functions between metric spaces parallels the definitions for real functions.

Definition 13.61. Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f : X \to Y$ is continuous at $c \in X$ if for every $\epsilon > 0$ there exists $\delta > 0$ such that

$$d_X(x,c) < \delta \text{ implies that } d_Y(f(x), f(c)) < \epsilon.$$

The function is continuous on X if it is continuous at every point of X.

Example 13.62. A function $f : \mathbb{R}^2 \to \mathbb{R}$, where $\mathbb{R}^2$ is equipped with the Euclidean norm $\| \cdot \|$ and $\mathbb{R}$ with the absolute value norm $| \cdot |$, is continuous at $c \in \mathbb{R}^2$ if

$$\|x - c\| < \delta \text{ implies that } \|f(x) - f(c)\| < \epsilon$$

Explicitly, if $x = (x_1, x_2)$, $c = (c_1, c_2)$ and

$$f(x) = (f_1(x_1, x_2), f_2(x_1, x_2)),$$

this condition reads:

$$\sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2} < \delta$$

implies that

$$|f(x_1, x_2) - f_1(c_1, c_2)| < \epsilon.$$

Example 13.63. A function $f : \mathbb{R} \to \mathbb{R}^2$, where $\mathbb{R}^2$ is equipped with the Euclidean norm $\| \cdot \|$ and $\mathbb{R}$ with the absolute value norm $| \cdot |$, is continuous at $c \in \mathbb{R}^2$ if

$$|x - c| < \delta \text{ implies that } \|f(x) - f(c)\| < \epsilon$$

Explicitly, if

$$f(x) = (f_1(x), f_2(x)),$$

where $f_1, f_2 : \mathbb{R} \to \mathbb{R}$, this condition reads: $|x - c| < \delta$ implies that

$$\sqrt{[f_1(x) - f_1(c)]^2 + [f_2(x) - f_2(c)]^2} < \epsilon.$$

The previous examples generalize in a natural way to define the continuity of an m-component vector-valued function of n variables, $f : \mathbb{R}^n \to \mathbb{R}^m$. The definition looks complicated if it is written out explicitly, but it is much clearer if it is expressed in terms or metrics or norms.

Example 13.64. Define $F : C([0, 1]) \to \mathbb{R}$ by

$$F(f) = f(0),$$

where $C([0, 1])$ is the space of continuous functions $f : [0, 1] \to \mathbb{R}$ equipped with the sup-norm described in Example 13.9, and $\mathbb{R}$ has the absolute value norm. That is, F evaluates a function $f(x)$ at $x = 0$. Thus F is a function acting on functions, and its values are scalars; such a function, which maps functions to scalars, is called a functional. Then F is continuous, since $\|f - g\|_\infty < \epsilon$ implies that $|f(0) - g(0)| < \epsilon$. (That is, we take $\delta = \epsilon$).

We also have a sequential characterization of continuity in a metric space.

Theorem 13.65. Let X and Y be metric spaces. A function $f : X \to Y$ is continuous at $c \in X$ if and only if $f(x_n) \to f(c)$ as $n \to \infty$ for every sequence (x_n) in X such that $x_n \to c$ as $n \to \infty$,

We define uniform continuity similarly to the real case.

Definition 13.66. Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f : X \to Y$ is uniformly continuous on X if for every $\epsilon > 0$ there exists $\delta > 0$ such that

$$d_X(x, y) < \delta \text{ implies that } d_Y(f(x), f(y)) < \epsilon.$$

The proofs of the following theorems are identically to the proofs we gave for functions $f : A \subset \mathbb{R} \to \mathbb{R}$. First, a function on a metric space is continuous if and only if the inverse images of open sets are open.

Theorem 13.67. A function $f : X \to Y$ between metric spaces X and Y is continuous on X if and only if $f^{-1}(V)$ is open in X for every open set V in Y.

Second, the continuous image of a compact set is compact.

Theorem 13.68. Let $f : K \to Y$ be a continuous function from a compact metric space K to a metric space Y. Then $f(K)$ is a compact subspace of Y.

Third, a continuous function on a compact set is uniformly continuous.

Theorem 13.69. If $f : K \to Y$ is a continuous function on a compact set K, then f is uniformly continuous.

13.5. Topological spaces

A collection of subsets of a set X with the properties of the open sets in a metric space given in Theorem 13.36 is called a topology on X, and a set with such a collection of open sets is called a topological space.

Definition 13.70. Let X be a set. A collection $\mathcal{T} \subset \mathcal{P}(X)$ of subsets of X is a topology on X if it satisfies the following conditions.

(1) The empty set $\varnothing$ and the whole set X belong to $\mathcal{T}$.

(2) The union of an arbitrary collection of sets in $\mathcal{T}$ belongs to $\mathcal{T}$.

(3) The intersection of a finite collection of sets in $\mathcal{T}$ belongs to $\mathcal{T}$.

A set $G \subset X$ is open with respect to $\mathcal{T}$ if $G \in \mathcal{T}$, and a set $F \subset X$ is closed with respect to $\mathcal{T}$ if $F^c \in \mathcal{T}$. A topological space $(X, \mathcal{T})$ is a set X together with a topology $\mathcal{T}$ on X.

We can put different topologies on a set with two or more elements. If the topology on X is clear from the context, then we simply refer to X as a topological space and we don't specify the topology when we refer to open or closed sets.

Every metric space with the open sets in Definition 13.30 is a topological space; the resulting collection of open sets is called the metric topology of the metric space. There are, however, topological spaces whose topology is not derived from any metric on the space.

Example 13.71. Let X be any set. Then $\mathcal{T} = \mathcal{P}(X)$ is a topology on X, called the discrete topology. In this topology, every set is both open and closed. This topology is the metric topology associated with the discrete metric on X in Example 13.3.

Example 13.72. Let X be any set. Then $\mathcal{T} = \{\varnothing, X\}$ is a topology on X, called the trivial topology. The empty set and the whole set are both open and closed, and no other subsets of X are either open or closed.

If X has at least two elements, then this topology is different from the discrete topology in the previous example, and it is not derived from a metric. To see this, suppose that $x, y \in X$ and $x \neq y$. If $d : X \times X \to \mathbb{R}$ is a metric on X, then $d(x, y) = r > 0$ and $B_r(x)$ is a nonempty open set in the metric topology that doesn't contain y, so $B_r(x) \notin \mathcal{T}$.

The previous example illustrates a separation property of metric topologies that need not be satisfied by non-metric topologies.

Definition 13.73. A topological space $(X, \mathcal{T})$ is Hausdorff if for every $x, y \in X$ with $x \neq y$ there exist open sets $U, V \in \mathcal{T}$ such that $x \in U$, $y \in V$ and $U \cap V = \varnothing$.

That is, a topological space is Hausdorff if distinct points have disjoint neighborhoods. In that case, we also say that the topology is Hausdorff. Nearly all topological spaces that arise in analysis are Hausdorff, including, in particular, metric spaces.

Proposition 13.74. Every metric topology is Hausdorff.

Proof. Let (X, d) be a metric space. If $x, y \in X$ and $x \neq y$, then $d(x, y) = r > 0$, and $B_{r/2}(x)$, $B_{r/2}(y)$ are disjoint open neighborhoods of x, y.   □

Compact sets are defined topologically as sets with the Heine-Borel property.

Definition 13.75. Let X be a topological space. A set $K \subset X$ is compact if every open cover of K has a finite subcover. That is, if $\{G_i : i \in I\}$ is a collection of open sets such that
$$K \subset \bigcup_{i \in I} G_i,$$
then there is a finite subcollection $\{G_{i_1}, G_{i_2}, \ldots, G_{i_n}\}$ such that
$$K \subset \bigcup_{k=1}^{n} G_{i_k}.$$

The Heine-Borel and Bolzano-Weierstrass properties are equivalent in every metric space.

Theorem 13.76. A metric space is compact if and only if it sequentially compact.

We won't prove this result here, but we remark that it does not hold for general topological spaces.

Finally, we give the topological definitions of convergence, continuity, and connectedness which are essentially the same as the corresponding statements for $\mathbb{R}$. We also show that continuous maps preserve compactness and connectedness.

The definition of the convergence of a sequence is identical to the statement in Proposition 5.9 for $\mathbb{R}$.

Definition 13.77. Let X be a topological space. A sequence (x_n) in X converges to $x \in X$ if for every neighborhood U of x there exists $N \in \mathbb{N}$ such that $x_n \in U$ for every $n > N$.

The following definition of continuity in a topological space corresponds to Definition 7.2 for $\mathbb{R}$ (with the relative absolute-value topology on the domain A of f) and Theorem 7.31.

Definition 13.78. Let $f : X \to Y$ be a function between topological spaces X, Y. Then f is continuous at $x \in X$ if for every neighborhood $V \subset Y$ of $f(x)$, there exists a neighborhood $U \subset X$ of x such that $f(U) \subset V$. The function f is continuous on X if $f^{-1}(V)$ is open in X for every open set $V \subset Y$.

These definitions are equivalent to the corresponding "ϵ-δ" definitions in a metric space, but they make sense in a general topological space because they refer only to neighborhoods and open sets. We illustrate them with two simple examples.

Example 13.79. If X is a set with the discrete topology in Example 13.71, then a sequence converges to $x \in X$ if an only if its terms are eventually equal to x, since $\{x\}$ is a neighborhood of x. Every function $f : X \to Y$ is continuous with respect to the discrete topology on X, since every subset of X is open. On the other hand, if Y has the discrete topology, then $f : X \to Y$ is continuous if and only if $f^{-1}(\{y\})$ is open in X for every $y \in Y$.

Example 13.80. Let X be a set with the trivial topology in Example 13.72. Then every sequence converges to every point $x \in X$, since the only neighborhood of x is X itself. As this example illustrates, non-Hausdorff topologies have the unpleasant feature that limits need not be unique, which is one reason why they rarely arise in analysis. If Y has the trivial topology, then every function $X \to Y$ is continuous, since $f^{-1}(\varnothing) = \varnothing$ and $f^{-1}(Y) = X$ are open in X. On the other hand, if X has the trivial topology and Y is Hausdorff, then the only continuous functions $f : X \to Y$ are the constant functions.

Our last definition of a connected topological space corresponds to Definition 5.58 for connected sets of real numbers (with the relative topology).

Definition 13.81. A topological space X is disconnected if there exist nonempty, disjoint open sets U, V such that $X = U \cup V$. A topological space is connected if it is not disconnected.

The following proof that continuous functions map compact sets to compact sets and connected sets is the same as the proofs given in Theorem 7.35 and Theorem 7.32 for sets of real numbers. Note that a continuous function maps compact or connected sets in the opposite direction to open or closed sets, whose inverse image is open or closed.

Theorem 13.82. Suppose that $f : X \to Y$ is a continuous map between topological spaces X and Y. Then $f(X)$ is compact if X is compact, and $f(X)$ is connected if X is connected.

Proof. For the first part, suppose that X is compact. If $\{V_i : i \in I\}$ is an open cover of $f(X)$, then since f is continuous $\{f^{-1}(V_i) : i \in I\}$ is an open cover of X, and since X is compact there is a finite subcover

$$\left\{ f^{-1}(V_{i_1}), f^{-1}(V_{i_2}), \ldots, f^{-1}(V_{i_n}) \right\}.$$

It follows that

$$\{V_{i_1}, V_{i_2}, \ldots, V_{i_n}\}$$

is a finite subcover of the original open cover of $f(X)$, which proves that $f(X)$ is compact.

For the second part, suppose that $f(X)$ is disconnected. Then there exist nonempty, disjoint open sets U, V in Y such that $U \cup V \supset f(X)$. Since f is continuous, $f^{-1}(U)$, $f^{-1}(V)$ are open, nonempty, disjoint sets such that

$$X = f^{-1}(U) \cup f^{-1}(V),$$

so X is disconnected. It follows that $f(X)$ is connected if X is connected. $\square$

13.6. * Function spaces

There are many function spaces, and their study is a central topic in analysis. We discuss only one important example here: the space of continuous functions on a compact set equipped with the sup norm. We repeat its definition from Example 13.9.

Definition 13.83. Let $K \subset \mathbb{R}$ be a compact set. The space $C(K)$ consists of the continuous functions $f : K \to \mathbb{R}$. Addition and scalar multiplication of functions is defined pointwise in the usual way: if $f, g \in C(K)$ and $k \in \mathbb{R}$, then

$$(f + g)(x) = f(x) + g(x), \qquad (kf)(x) = k\,(f(x))\,.$$

The sup-norm of a function $f \in C(K)$ is defined by

$$\|f\|_\infty = \sup_{x \in K} |f(x)|.$$

Since a continuous function on a compact set attains its maximum and minimum value, for $f \in C(K)$ we can also write

$$\|f\|_\infty = \max_{x \in K} |f(x)|.$$

Thus, the sup-norm on $C(K)$ is analogous to the ℓ^∞-norm on $\mathbb{R}^n$. In fact, if $K = \{1, 2, \ldots, n\}$ is a finite set with the discrete topology, then it is identical to the ℓ^∞-norm.

Our previous results on continuous functions on a compact set can be formulated concisely in terms of this space. The following characterization of uniform convergence in terms of the sup-norm is easily seen to be equivalent to Definition 9.8.

Definition 13.84. A sequence (f_n) of functions $f_n : K \to \mathbb{R}$ converges uniformly on K to a function $f : K \to \mathbb{R}$ if

$$\lim_{n \to \infty} \|f_n - f\|_\infty = 0.$$

Similarly, we can rephrase Definition 9.12 for a uniformly Cauchy sequence in terms of the sup-norm.

Definition 13.85. A sequence (f_n) of functions $f_n : K \to \mathbb{R}$ is uniformly Cauchy on K if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$m, n > N \text{ implies that } \|f_m - f_n\|_\infty < \epsilon.$$

Thus, the uniform convergence of a sequence of functions is defined in exactly the same way as the convergence of a sequence of real numbers with the absolute value $|\cdot|$ replaced by the sup-norm $\|\cdot\|$. Moreover, like $\mathbb{R}$, the space $C(K)$ is complete.

Theorem 13.86. The space $C(K)$ with the sup-norm $\|\cdot\|_\infty$ is a Banach space.

Proof. From Theorem 7.15, the sum of continuous functions and the scalar multiple of a continuous function are continuous, so $C(K)$ is closed under addition and scalar multiplication. The algebraic vector-space properties for $C(K)$ follow immediately from those of $\mathbb{R}$.

From Theorem 7.37, a continuous function on a compact set is bounded, so $\|\cdot\|_\infty$ is well-defined on $C(K)$. The sup-norm is clearly non-negative, and $\|f\|_\infty = 0$ implies that $f(x) = 0$ for every $x \in K$, meaning that $f = 0$ is the zero function.

We also have for all $f, g \in C(K)$ and $k \in \mathbb{R}$ that

$$\|kf\|_\infty = \sup_{x \in K} |kf(x)| = |k| \sup_{x \in K} |f(x)| = |k| \, \|f\|_\infty,$$

$$\|f + g\|_\infty = \sup_{x \in K} |f(x) + g(x)|$$

$$\leq \sup_{x \in K} \{|f(x)| + |g(x)|\}$$

$$\leq \sup_{x \in K} |f(x)| + \sup_{x \in K} |g(x)|$$

$$\leq \|f\|_\infty + \|g\|_\infty,$$

which verifies the properties of a norm.

Finally, Theorem 9.13 implies that a uniformly Cauchy sequence converges uniformly so $C(K)$ is complete. $\qquad\square$

For comparison with the sup-norm, we consider a different norm on $C([a,b])$ called the one-norm, which is analogous to the ℓ^1-norm on $\mathbb{R}^n$.

Definition 13.87. If $f : [a,b] \to \mathbb{R}$ is a Riemann integrable function, then the one-norm of f is

$$\|f\|_1 = \int_a^b |f(x)| \, dx.$$

Theorem 13.88. The space $C([a,b])$ of continuous functions $f : [a,b] \to \mathbb{R}$ with the one-norm $\| \cdot \|_1$ is a normed space.

Proof. As shown in Theorem 13.86, $C([a,b])$ is a vector space. Every continuous function is Riemann integrable on a compact interval, so $\| \cdot \|_1 : C([a,b]) \to \mathbb{R}$ is well-defined, and we just have to verify that it satisfies the properties of a norm.

Since $|f| \geq 0$, we have $\|f\|_1 = \int_a^b |f| \geq 0$. Furthermore, since f is continuous, Proposition 11.42 shows that $\|f\|_1 = 0$ implies that $f = 0$, which verifies the positivity. If $k \in \mathbb{R}$, then

$$\|kf\|_1 = \int_a^b |kf| = |k| \int_a^b |f| = |k| \, \|f\|_1,$$

which verifies the homogeneity. Finally, the triangle inequality is satisfied since

$$\|f + g\|_1 = \int_a^b |f + g| \leq \int_a^b |f| + |g| = \int_a^b |f| + \int_a^b |g| = \|f\|_1 + \|g\|_1.$$

$\qquad\square$

Although $C([a,b])$ equipped with the one-norm $\| \cdot \|_1$ is a normed space, it is not complete, and therefore it is not a Banach space. The following example gives a non-convergent Cauchy sequence in this space.

Example 13.89. Define the continuous functions $f_n : [0,1] \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} 0 & \text{if } 0 \leq x \leq 1/2, \\ n(x - 1/2) & \text{if } 1/2 < x < 1/2 + 1/n, \\ 1 & \text{if } 1/2 + 1/n \leq x \leq 1. \end{cases}$$

If $n > m$, we have

$$\|f_n - f_m\|_1 = \int_{1/2}^{1/2+1/m} |f_n - f_m| \le \frac{1}{m},$$

since $|f_n - f_n| \le 1$. Thus, $\|f_n - f_m\|_1 < \epsilon$ for all $m, n > 1/\epsilon$, so (f_n) is a Cauchy sequence with respect to the one-norm.

We claim that if $\|f - f_n\|_1 \to 0$ as $n \to \infty$ where $f \in C([0,1])$, then f would have to be

$$f(x) = \begin{cases} 0 & \text{if } 0 \le x \le 1/2, \\ 1 & \text{if } 1/2 < x \le 1, \end{cases}$$

which is discontinuous at $1/2$, so (f_n) does not have a limit in $(C([0,1]), \|\cdot\|_1)$.

To prove the claim, note that if $\|f - f_n\|_1 \to 0$, then $\int_0^{1/2} |f| = 0$ since

$$\int_0^{1/2} |f| = \int_0^{1/2} |f - f_n| \le \int_0^1 |f - f_n| \to 0,$$

and Proposition 11.42 implies that $f(x) = 0$ for $0 \le x \le 1/2$. Similarly, for every $0 < \epsilon < 1/2$, we get that $\int_{1/2+\epsilon}^1 |f - 1| = 0$, so $f(x) = 1$ for $1/2 < x \le 1$.

The sequence (f_n) is not uniformly Cauchy since $\|f_n - f_m\|_\infty \to 1$ as $n \to \infty$ for every $m \in \mathbb{N}$, so this example does not contradict the completeness of $(C([0,1]), \|\cdot\|_\infty)$.

The ℓ^∞-norm and the ℓ^1-norm on the finite-dimensional space $\mathbb{R}^n$ are equivalent, but the sup-norm and the one-norm on $C([a,b])$ are not. In one direction, we have

$$\int_a^b |f| \le (b-a) \cdot \sup_{[a,b]} |f|,$$

so $\|f\|_1 \le (b-a)\|f\|_\infty$, and $\|f - f_n\|_\infty \to 0$ implies that $\|f - f_n\|_1 \to 0$. As the following example shows, the converse is not true. There is no constant M such that $\|f\|_\infty \le M\|f\|_1$ for all $f \in C([a,b])$, and $\|f - f_n\|_1 \to 0$ does not imply that $\|f - f_n\|_\infty \to 0$.

Example 13.90. For $n \in \mathbb{N}$, define the continuous function $f_n : [0,1] \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} 1 - nx & \text{if } 0 \le x \le 1/n, \\ 0 & \text{if } 1/n < x \le 1. \end{cases}$$

Then $\|f_n\|_\infty = 1$ for every $n \in \mathbb{N}$, but

$$\|f_n\|_1 = \int_0^{1/n} (1 - nx)\, dx = \left[x - \frac{1}{2}nx^2 \right]_0^{1/n} = \frac{1}{2n},$$

so $\|f_n\|_1 \to 0$ as $n \to \infty$.

Thus, unlike the finite-dimensional vector space $\mathbb{R}^n$, an infinite-dimensional vector space such as $C([a,b])$ has many inequivalent norms and many inequivalent notions of convergence.

The incompleteness of $C([a,b])$ with respect to the one-norm suggests that we use the larger space $R([a,b])$ of Riemann integrable functions on $[a,b]$, which

includes some discontinuous functions. A slight complication arises from the fact that if f is Riemann integrable and $\int_a^b |f| = 0$, then it does not follows that $f = 0$, so $\|f\|_1 = 0$ does not imply that $f = 0$. Thus, $\|\cdot\|_1$ is not, strictly speaking, a norm on $R([a, b])$. We can, however, get a normed space of equivalence classes of Riemann integrable functions, by defining $f, g \in R([a, b])$ to be equivalent if $\int_a^b |f - g| = 0$. For instance, the function in Example 11.14 is equivalent to the zero-function.

A much more fundamental defect of the space of (equivalence classes of) Riemann integrable functions with the one-norm is that it is still not complete. To get a space that is complete with respect to the one-norm, we have to use the space $L^1([a, b])$ of (equivalence classes of) Lebesgue integrable functions on $[a, b]$. This is another reason for the superiority of the Lebesgue integral over the Riemann integral: it leads function spaces that are complete with respect to integral norms.

The inclusion of the smaller incomplete space $C([a, b])$ of continuous functions in the larger complete space $L^1([a, b])$ of Lebesgue integrable functions is analogous to the inclusion of the incomplete space $\mathbb{Q}$ of rational numbers in the complete space $\mathbb{R}$ of real numbers.

13.7. * The Minkowski inequality

Inequalities are essential to analysis. Their proofs, however, may require considerable ingenuity, and there are often many different ways to prove the same inequality. In this section, we complete the proof that the ℓ^p-spaces are normed spaces by proving the triangle inequality given in Definition 13.25. This inequality is called the Minkowski inequality, and it's one of the most important inequalities in mathematics.

The simplest case is for the Euclidean norm with $p = 2$. We begin by proving the following fundamental Cauchy-Schwartz inequality.

Theorem 13.91 (Cauchy-Schwartz inequality)**.** If $(x_1, x_2, \ldots, x_n)$, $(y_1, y_2, \ldots, y_n)$ are points in $\mathbb{R}^n$, then

$$\left| \sum_{i=1}^n x_i y_i \right| \leq \left(\sum_{i=1}^n x_i^2 \right)^{1/2} \left(\sum_{i=1}^n y_i^2 \right)^{1/2}.$$

Proof. Since $|\sum x_i y_i| \leq \sum |x_i| |y_i|$, it is sufficient to prove the inequality for $x_i, y_i \geq 0$. Furthermore, the inequality is obvious if $x = 0$ or $y = 0$, so we assume that at least one x_i and one y_i is nonzero.

For every $\alpha, \beta \in \mathbb{R}$, we have

$$0 \leq \sum_{i=1}^n \left(\alpha x_i - \beta y_i \right)^2.$$

Expanding the square on the right-hand side and rearranging the terms, we get that

$$2\alpha\beta \sum_{i=1}^n x_i y_i \leq \alpha^2 \sum_{i=1}^n x_i^2 + \beta^2 \sum_{i=1}^n y_i^2.$$

We choose $\alpha, \beta > 0$ to "balance" the terms on the right-hand side,

$$\alpha = \left(\sum_{i=1}^{n} y_i^2\right)^{1/2}, \qquad \beta = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}.$$

Then division of the resulting inequality by $2\alpha\beta$ proves the theorem. $\qquad\square$

The Minkowski inequality for $p = 2$ is an immediate consequence of the Cauchy-Schwartz inequality.

Corollary 13.92 (Minkowski inequality). If $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ are points in $\mathbb{R}^n$, then

$$\left[\sum_{i=1}^{n}(x_i + y_i)^2\right]^{1/2} \leq \left(\sum_{i=1}^{n} x_i^2\right)^{1/2} + \left(\sum_{i=1}^{n} y_i^2\right)^{1/2}.$$

Proof. Expanding the square in the following equation and using the Cauchy-Schwartz inequality, we get

$$\sum_{i=1}^{n}(x_i + y_i)^2 = \sum_{i=1}^{n} x_i^2 + 2\sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n} y_i^2$$

$$\leq \sum_{i=1}^{n} x_i^2 + 2\left(\sum_{i=1}^{n} x_i^2\right)^{1/2}\left(\sum_{i=1}^{n} y_i^2\right)^{1/2} + \sum_{i=1}^{n} y_i^2$$

$$\leq \left[\left(\sum_{i=1}^{n} x_i^2\right)^{1/2} + \left(\sum_{i=1}^{n} y_i^2\right)^{1/2}\right]^2.$$

Taking the square root of this inequality, we obtain the result. $\qquad\square$

To prove the Minkowski inequality for general $1 < p < \infty$, we first define the Hölder conjugate p' of p and prove Young's inequality.

Definition 13.93. If $1 < p < \infty$, then the Hölder conjugate $1 < p' < \infty$ of p is the number such that

$$\frac{1}{p} + \frac{1}{p'} = 1.$$

If $p = 1$, then $p' = \infty$; and if $p = \infty$ then $p' = 1$.

The Hölder conjugate of $1 < p < \infty$ is given explicitly by

$$p' = \frac{p}{p-1}.$$

Note that if $1 < p < 2$, then $2 < p' < \infty$; and if $2 < p < \infty$, then $1 < p' < 2$. The number 2 is its own Hölder conjugate. Furthermore, if p' is the Hölder conjugate of p, then p is the Hölder conjugate of p'.

Theorem 13.94 (Young's inequality). Suppose that $1 < p < \infty$ and $1 < p' < \infty$ is its Hölder conjugate. If $a, b \geq 0$ are nonnegative real numbers, then

$$ab \leq \frac{a^p}{p} + \frac{b^{p'}}{p'}.$$

Moreover, there is equality if and only if $a^p = b^{p'}$.

Proof. There are several ways to prove this inequality. We give a proof based on calculus.

The result is trivial if $a = 0$ or $b = 0$, so suppose that $a, b > 0$. We write

$$\frac{a^p}{p} + \frac{b^{p'}}{p'} - ab = b^{p'} \left(\frac{1}{p} \frac{a^p}{b^{p'}} + \frac{1}{p'} - \frac{a}{b^{p'-1}} \right).$$

The definition of p' implies that $p'/p = p' - 1$, so that

$$\frac{a^p}{b^{p'}} = \left(\frac{a}{b^{p'/p}} \right)^p = \left(\frac{a}{b^{p'-1}} \right)^p.$$

Therefore, we have

$$\frac{a^p}{p} + \frac{b^{p'}}{p'} - ab = b^{p'} f(t), \qquad f(t) = \frac{t^p}{p} + \frac{1}{p'} - t, \qquad t = \frac{a}{b^{p'-1}}.$$

The derivative of f is

$$f'(t) = t^{p-1} - 1.$$

Thus, for $p > 1$, we have $f'(t) < 0$ if $0 < t < 1$, and Theorem 8.36 implies that $f(t)$ is strictly decreasing; moreover, $f'(t) > 0$ if $1 < t < \infty$, so $f(t)$ is strictly increasing. It follows that f has a strict global minimum on $(0, \infty)$ at $t = 1$. Since

$$f(1) = \frac{1}{p} + \frac{1}{p'} - 1 = 0,$$

we conclude that $f(t) \geq 0$ for all $0 < t < \infty$, with equality if and only if $t = 1$. Furthermore, $t = 1$ if and only if $a = b^{p'-1}$ or $a^p = b^{p'}$. It follows that

$$\frac{a^p}{p} + \frac{b^{p'}}{p'} - ab \geq 0$$

for all $a, b \geq 0$, with equality if and only $a^p = b^{p'}$, which proves the result. $\qquad \square$

For $p = 2$, Young's inequality reduces to the more easily proved inequality in Proposition 2.8.

Before continuing, we give a scaling argument which explains the appearance of the Hölder conjugate in Young's inequality. Suppose we look for an inequality of the form

$$ab \leq Ma^p + Na^q \qquad \text{for all } a, b \geq 0$$

for some exponents p, q and some constants M, N. Any inequality that holds for all positive real numbers must remain true under rescalings. Rescaling $a \mapsto \lambda a$, $b \mapsto \mu b$ in the inequality (where $\lambda, \mu > 0$) and dividing by $\lambda \mu$, we find that it becomes

$$ab \leq \frac{\lambda^{p-1}}{\mu} Ma^p + \frac{\mu^{q-1}}{\lambda} Nb^q.$$

We take $\mu = \lambda^{p-1}$ to make the first scaling factor equal to one, and then the inequality becomes

$$ab \leq Ma^p + \lambda^r Nb^q, \qquad r = (p-1)(q-1) - 1.$$

If the exponent r of λ is non-zero, then we can violate the inequality by taking λ sufficiently small (if $r > 0$) or sufficiently large (if $r < 0$), since it is clearly

impossible to bound ab by a^p for all $b \in \mathbb{R}$. Thus, the inequality can only hold if $r = 0$, which implies that $q = p'$.

This argument does not, of course, prove the inequality, but it shows that the only possible exponents for which an inequality of this form can hold must satisfy $q = p'$. Theorem 13.94 proves that such an inequality does in fact hold in that case provided $1 < p < \infty$.

Next, we use Young's inequality to deduce Hólder's inequality, which is a generalization of the Cauchy-Schwartz inequality for $p \neq 2$.

Theorem 13.95 (Hölder's inequality)**.** Suppose that $1 < p < \infty$ and $1 < p' < \infty$ is its Hölder conjugate. If $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ are points in $\mathbb{R}^n$, then

$$\left| \sum_{i=1}^{n} x_i y_i \right| \leq \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \left(\sum_{i=1}^{n} |y_i|^{p'} \right)^{1/p'} .$$

Proof. We assume without loss of generality that x_i, y_i are nonnegative and $x, y \neq 0$. Let $\alpha, \beta > 0$. Then applying Young's inequality in Theorem 13.94 with $a = \alpha x_i$, $b = \beta y_i$ and summing over i, we get

$$\alpha \beta \sum_{i=1}^{n} x_i y_i \leq \frac{\alpha^p}{p} \sum_{i=1}^{n} x_i^p + \frac{\beta^{p'}}{p'} \sum_{i=1}^{n} y_i^{p'}.$$

Then, choosing

$$\alpha = \left(\sum_{i=1}^{n} y_i^{p'} \right)^{1/p}, \qquad \beta = \left(\sum_{i=1}^{n} x_i^p \right)^{1/p'}$$

to "balance" the terms on the right-hand side, dividing by $\alpha\beta$, and using the fact that $1/p + 1/p' = 1$, we get Hölder's inequality. $\qquad \square$

Minkowski's inequality follows from Hölder's inequality.

Theorem 13.96 (Minkowski's inequality)**.** Suppose that $1 < p < \infty$ and $1 < p' < \infty$ is its Hölder conjugate. If $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ are points in $\mathbb{R}^n$, then

$$\left(\sum_{i=1}^{n} |x_i + y_i|^p \right)^{1/p} \leq \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} + \left(\sum_{i=1}^{n} |y_i|^p \right)^{1/p} .$$

Proof. We assume without loss of generality that x_i, y_i are nonnegative and $x, y \neq 0$. We split the sum on the left-hand side as follows:

$$\sum_{i=1}^{n} |x_i + y_i|^p = \sum_{i=1}^{n} |x_i + y_i| \, |x_i + y_i|^{p-1}$$

$$\leq \sum_{i=1}^{n} |x_i| \, |x_i + y_i|^{p-1} + \sum_{i=1}^{n} |y_i| \, |x_i + y_i|^{p-1} .$$

By Hölder's inequality, we have

$$\sum_{i=1}^{n} |x_i| \, |x_i + y_i|^{p-1} \leq \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \left(\sum_{i=1}^{n} |x_i + y_i|^{(p-1)p'} \right)^{1/p'} ,$$

and using the fact that $p' = p/(p-1)$, we get

$$\sum_{i=1}^{n} |x_i|\; |x_i + y_i|^{p-1} \leq \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |x_i + y_i|^p\right)^{1-1/p}.$$

Similarly,

$$\sum_{i=1}^{n} |y_i|\; |x_i + y_i|^{p-1} \leq \left(\sum_{i=1}^{n} |y_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |x_i + y_i|^p\right)^{1-1/p}.$$

Combining these inequalities, we obtain

$$\sum_{i=1}^{n} |x_i + y_i|^p \leq \left[\left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} + \left(\sum_{i=1}^{n} |y_i|^p\right)^{1/p}\right] \left(\sum_{i=1}^{n} |x_i + y_i|^p\right)^{1-1/p}.$$

Fianlly, dividing this inequality by $\left(\sum |x_i + y_i|^p\right)^{1-1/p}$, we get the result. $\qquad\square$

Bibliography

[1] S. Abbott, *Understanding Analysis*, Springer-Verlag, New York, 2001.

[2] T. Apostol, *Mathematical Analysis*, Addison-Wesley, 1974.

[3] R. Dedekind, *Was sind und was sollen die Zahlen?*, 1888. Tranbslation: *What are Numbers and What Should They Be?*, H. Pogorzelski, W. Ryan, W. Snyder, Research Institute for Mathematics, (1995).

[4] P. Duren, *Invitation to Classical Analysis*, AMS, 2012.

[5] W. Dunham, *The Calculus Gallery*, Princeton University Press, 2005.

[6] T, W. Körner, *A Companion to Analysis*, AMS, 2004.

[7] J. E. Marsden and M. J. Hoffman, *Elementary Classical Analysis*, Macmillan, 1993.

[8] F. A. Medvedev, *Scenes from the History of Real Functions*, Birkhäuser, Basel, 1991.

[9] V. H. Moll, *Numbers and Functions*, AMS, Providence, 2012.

[10] Y. Moschovakis, *Notes on Set Theory*, 2nd ed., Springer, 2006.

[11] B. Riemann, *Collected Works*, Translated by R. Baker, C. Christenson, and H. Orde, Kendrick Press, 2004.

[12] K. A. Ross, *Elementary Analysis*, Springer, 2010.

[13] J. M. Steele, *The Cauchy-Schwarz Master Class*, Cambridge University Press, 2004.

[14] R. Strichartz, *The Way of Analysis*, 2000.

[15] W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, 1976.

[16] E. T Whittaker and G. N. Watson, *A Course of Modern Analysis*, Cambridge University Press, 1927.