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 limits of functions, continuity, differentiability, and sequences and series of functions, but not Riemann integration A background in sequences and series of real numbers and some elementary point set topology of the real numbers is assumed, although some of this material is briefly reviewed.

Contents

Chapter 1

The Real Numbers

In this chapter, we review some properties of the real numbers $\mathbb{R}$ and its subsets. We don't give proofs for most of the results stated here.

1.1. Completeness of $\mathbb{R}$

Intuitively, unlike the rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$ form a continuum with no 'gaps.' There are two main ways to state this completeness, one in terms of the existence of suprema and the other in terms of the convergence of Cauchy sequences.

1.1.1. Suprema and infima.

Definition 1.1. Let $A \subset \mathbb{R}$ be a set of real numbers. A real number $M \in \mathbb{R}$ is an upper bound of A if $x \leq M$ for every $x \in A$, and $m \in \mathbb{R}$ is a lower bound of A if $x \geq m$ for every $x \in A$. A set is bounded from above if it has an upper bound, bounded from below if it has a lower bound, and bounded if it has both an upper and a lower bound

An equivalent condition for A to be bounded is that there exists $R \in \mathbb{R}$ such that $|x| \leq R$ for every $x \in A$.

Example 1.2. The set of natural numbers

$$\mathbb{N} = \{1, 2, 3, 4, \dots\}$$

is bounded from below by any $m \in \mathbb{R}$ with $m \leq 1$. It is not bounded from above, so $\mathbb{N}$ is unbounded.

Definition 1.3. 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 supremum or least upper bound 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 infimum or greatest lower bound of A, denoted

$$m = \inf A.$$

The supremum or infimum of a set may or may not belong to the set. If $\sup A \in A$ does belong to A, then we also denote it by $\max A$ and refer to it as the maximum of A; if $\inf A \in A$ then we also denote it by $\min A$ and refer to it as the minimum of A.

Example 1.4. 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 1.5. 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.

Example 1.6. For $A = (0, 1)$, we have

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

In this case, neither $\sup A$ nor $\inf A$ belongs to A. The closed interval $B = [0, 1]$, and the half-open interval $C = (0, 1]$ have the same supremum and infimum as A. Both $\sup B$ and $\inf B$ belong to B, while only $\sup C$ belongs to C.

The completeness of $\mathbb{R}$ may be expressed in terms of the existence of suprema.

Theorem 1.7. Every nonempty set of real numbers that is bounded from above has a supremum.

Since $\inf A = -\sup(-A)$, it follows immediately that every nonempty set of real numbers that is bounded from below has an infimum.

Example 1.8. The supremum of the set of real numbers

$$A = \left\{ x \in \mathbb{R} : x < \sqrt{2} \right\}$$

is $\sup A = \sqrt{2}$. By contrast, since $\sqrt{2}$ is irrational, the set of rational numbers

$$B = \left\{ x \in \mathbb{Q} : x < \sqrt{2} \right\}$$

has no supremum in $\mathbb{Q}$. (If $M \in \mathbb{Q}$ is an upper bound of B, then there exists $M' \in \mathbb{Q}$ with $\sqrt{2} < M' < M$, so M is not a least upper bound.)

1.1.2. Cauchy sequences. We assume familiarity with the convergence of real sequences, but we recall the definition of Cauchy sequences and their relation with the completeness of $\mathbb{R}$.

Definition 1.9. A sequence (x_n) of real numbers is a Cauchy sequence if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

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

Every convergent sequence is Cauchy. Conversely, it follows from Theorem 1.7 that every Cauchy sequence of real numbers has a limit.

Theorem 1.10. A sequence of real numbers converges if and only if it is a Cauchy sequence.

The fact that real Cauchy sequences have a limit is an equivalent way to formulate the completeness of $\mathbb{R}$. By contrast, the rational numbers $\mathbb{Q}$ are not complete.

Example 1.11. Let (x_n) be a sequence of rational numbers such that $x_n \to \sqrt{2}$ as $n \to \infty$. Then (x_n) is Cauchy in $\mathbb{Q}$ but (x_n) does not have a limit in $\mathbb{Q}$.

1.2. 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 compactness or continuity) that can be defined entirely in terms of open sets is called a topological property.

Definition 1.12. A set $G \subset \mathbb{R}$ is open in $\mathbb{R}$ if for every $x \in G$ there exists a $\delta > 0$ such that $G \supset (x - \delta, x + \delta)$.

Another way to state this definition is in terms of interior points.

Definition 1.13. Let $A \subset \mathbb{R}$ be a subset of $\mathbb{R}$. A point $x \in A$ is an interior point of A a if there is a $\delta > 0$ such that $A \supset (x - \delta, x + \delta)$. A point $x \in \mathbb{R}$ is a boundary point of A if every interval $(x - \delta, x + \delta)$ contains points in A and points not in A.

Thus, a set is open if and only if every point in the set is an interior point.

Example 1.14. The open interval $I = (0, 1)$ is open. If $x \in I$ then I contains an open interval about x,

$$I \supset \left(\frac{x}{2}, \frac{1+x}{2}\right), \qquad x \in \left(\frac{x}{2}, \frac{1+x}{2}\right),$$

and, for example, $I \supset (x - \delta, x + \delta)$ if

$$\delta = \min\left(\frac{x}{2}, \frac{1-x}{2}\right) > 0.$$

Similarly, every finite or infinite open interval (a, b), $(-\infty, b)$, (a, ∞) is open.

An arbitrary union of open sets is open; one can prove that every open set in $\mathbb{R}$ is a countable union of disjoint open intervals. A *finite* intersection of open sets is open, but an intersection of infinitely many open sets needn't be open.

Example 1.15. 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.

Instead of using intervals to define open sets, we can use neighborhoods, and it is frequently simpler to refer to neighborhoods instead of open intervals of radius $\delta > 0$.

Definition 1.16. 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. Sometimes a neighborhood is also required to be an open set, but we don't do this and will specify that a neighborhood is open when it is needed.

Example 1.17. 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 Definition 1.12 in terms of neighborhoods as follows.

Definition 1.18. A set $G \subset \mathbb{R}$ is open if every $x \in G$ has a neighborhood U such that $G \supset U$.

We define relatively open sets by restricting open sets in $\mathbb{R}$ to a subset.

Definition 1.19. If $A \subset \mathbb{R}$ then $B \subset A$ is relatively open in A, or open in A, if $B = A \cap U$ where U is open in $\mathbb{R}$.

Example 1.20. Let $A = [0, 1]$. Then the half-open intervals $(a, 1]$ and $[0, b)$ are open in A for every $0 \leq a < 1$ and $0 < b \leq 1$, 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.

Definition 1.21. If $A \subset \mathbb{R}$ then a relative neighborhood of $x \in A$ is a set $C = A \cap V$ where V is a neighborhood of x in $\mathbb{R}$.

As for open sets in $\mathbb{R}$, a set is relatively open if and only if it contains a relative neighborhood of every point. Since we use this fact at one point later on, we give a proof.

Proposition 1.22. A set $B \subset A$ is relatively open in A if and only if every $x \in B$ has a relative neighborhood C such that $B \supset C$.

Proof. Assume that $B = A \cap U$ is open in A, where U is open in $\mathbb{R}$. If $x \in B$, then $x \in U$. Since U is open, there is a neighborhood V of x in $\mathbb{R}$ such that $U \supset V$. Then $C = A \cap V$ is a relative neighborhood of x with $B \supset C$. (Alternatively, we could observe that B itself is a relative neighborhood of every $x \in B$.)

Conversely, assume that every point $x \in B$ has a relative neighborhood $C_x = A \cap V_x$ such that $C_x \subset B$. Then, since V_x is a neighborhood of x in $\mathbb{R}$, there is an open neighborhood $U_x \subset V_x$ of x, for example a δ-neighborhood. We claim that that $B = A \cap U$ where

$$U = \bigcup_{x \in B} U_x.$$

To prove this claim, we show that $B \subset A \cap U$ and $B \supset A \cap U$. First, $B \subset A \cap U$ since $x \in A \cap U_x \subset A \cap U$ for every $x \in B$. Second, $A \cap U_x \subset A \cap V_x \subset B$ for every $x \in B$. Taking the union over $x \in B$, we get that $A \cap U \subset B$. Finally, U is open since it's a union of open sets, so $B = A \cap U$ is relatively open in A. □

1.3. Closed sets

Closed sets are complements of open sets.

Definition 1.23. A set $F \subset \mathbb{R}$ is closed if $F^c = \{x \in \mathbb{R} : x \notin F\}$ is open.

Closed sets can also be characterized in terms of sequences.

Definition 1.24. A set $F \subset \mathbb{R}$ is sequentially closed if the limit of every convergent sequence in F belongs to F.

A subset of $\mathbb{R}$ is closed if and only if it is sequentially closed, so we can use either definition, and we don't distinguish between closed and sequentially closed sets.

Example 1.25. The closed interval $[0, 1]$ is closed. To verify this from Definition 1.23, note that

$$[0, 1]^c = (-\infty, 0) \cup (1, \infty)$$

is open. To verify this from Definition 1.24, note that if (x_n) is a convergent sequence in $[0, 1]$, then $0 \leq x_n \leq 1$ for all $n \in \mathbb{N}$. Since limits preserve (non-strict) inequalities, we have

$$0 \leq \lim_{n \to \infty} x_n \leq 1,$$

meaning that the limit belongs to $[0, 1]$. Similarly, every finite or infinite closed interval $[a, b]$, $(-\infty, b]$, $[a, \infty)$ is closed.

An arbitrary intersection of closed sets is closed and a *finite* union of closed sets is closed. A union of infinitely many closed sets needn't be closed.

Example 1.26. 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

$$\bigcup_{n=1}^{\infty} I_n = (0, 1).$$

The only sets that are both open and closed are the real numbers $\mathbb{R}$ and the empty set $\varnothing$. In general, sets are neither open nor closed.

Example 1.27. The half-open interval $I = (0, 1]$ is neither open nor closed. It's not open since I doesn't contain any neighborhood of the point $1 \in I$. It's not closed since $(1/n)$ is a convergent sequence in I whose limit 0 doesn't belong to I.

1.4. Accumulation points and isolated points

An accumulation point of a set A is a point in $\mathbb{R}$ that has points in A arbitrarily close to it.

Definition 1.28. A point $x \in \mathbb{R}$ is an accumulation point of $A \subset \mathbb{R}$ if for every $\delta > 0$ the interval $(x - \delta, x + \delta)$ contains a point in A that is different from x.

Accumulation points are also called limit points or cluster points. By taking smaller and smaller intervals about x, we see that if x is an accumulation point of A then every neighborhood of x contains infinitely many points in A. This leads to an equivalent sequential definition.

Definition 1.29. A point $x \in \mathbb{R}$ is an accumulation point of $A \subset \mathbb{R}$ 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 $n \to \infty$.

An accumulation point of a set may or may not belong to the set (a set is closed if and only if all its accumulation points belong to the set), and a point that belongs to the set may or may not be an accumulation point.

Example 1.30. The set $\mathbb{N}$ of natural numbers has no accumulation points.

Example 1.31. If

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

then 0 is an accumulation point of A since every open interval about 0 contains $1/n$ for sufficiently large n. Alternatively, the sequence $(1/n)$ in A converges to 0 as $n \to \infty$. In this case, the accumulation point 0 does not belong to A. Moreover, 0 is the only accumulation point of A; in particular, none of the points in A are accumulation points of A.

Example 1.32. The set of accumulation points of a bounded, open interval $I = (a, b)$ is the closed interval $[a, b]$. Every point in I is an accumulation point of I. In addition, the endpoints a, b are accumulation points of I that do not belong to I. The set of accumulation points of the closed interval $[a, b]$ is again the closed interval $[a, b]$.

Example 1.33. Let $a < c < b$ and suppose that

$$A = (a, c) \cup (c, b)$$

is an open interval punctured at c. Then the set of accumulation points of A is the closed interval $[a, b]$. The points a, b, c are accumulation points of A that do not belong to A.

An isolated point of a set is a point in the set that does not have other points in the set arbitrarily close to it.

Definition 1.34. Let $A \subset \mathbb{R}$. A point $x \in A$ is an isolated point of A if there exists $\delta > 0$ such that x is the only point belonging to A in the interval $(x - \delta, x + \delta)$.

Unlike accumulation points, isolated points are required to belong to the set. Every point $x \in A$ is either an accumulation point of A (if every neighborhood contains other points in A) or an isolated point of A (if some neighborhood contains no other points in A).

Example 1.35. If

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

then every point $1/n \in A$ is an isolated point of A since the interval $(1/n-\delta, 1/n+\delta)$ does not contain any points $1/m$ with $m \in \mathbb{N}$ and $m \neq n$ when $\delta > 0$ is sufficiently small.

Example 1.36. An interval has no isolated points (excluding the trivial case of closed intervals of zero length that consist of a single point $[a, a] = \{a\}$).

1.5. Compact sets

Compactness is not as obvious a property of sets as being open, but it plays a central role in analysis. One motivation for the property is obtained by turning around the Bolzano-Weierstrass and Heine-Borel theorems and taking their conclusions as a definition.

We will give two equivalent definitions of compactness, one based on sequences (every sequence has a convergent subsequence) and the other based on open covers (every open cover has a finite subcover). A subset of $\mathbb{R}$ is compact if and only if it is closed and bounded, in which case it has both of these properties. For example, every closed, bounded interval $[a, b]$ is compact. There are also other, more exotic, examples of compact sets, such as the Cantor set.

1.5.1. Sequential compactness.

Intuitively, a compact set confines every infinite 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 the accumulation point and leads to the following definition.

Definition 1.37. 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.

Example 1.38. The open interval $I = (0, 1)$ is not sequentially 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 1.39. The set $\mathbb{N}$ is closed, but it is not sequentially compact since the sequence (n) in $\mathbb{N}$ has no convergent subsequence. (Every subsequence diverges to infinity.)

As these examples illustrate, a sequentially compact set must be closed and bounded. Conversely, the Bolzano-Weierstrass theorem implies that that every closed, bounded subset of $\mathbb{R}$ is sequentially compact.

Theorem 1.40. A set $K \subset \mathbb{R}$ is sequentially compact if and only if it is closed and bounded.

Proof. First, assume that K is sequentially compact. Let (x_n) be any 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 $x \in K$, meaning that K is closed.

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 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 is closed and bounded. Let (x_n) be a sequence in K. Then (x_n) is bounded since K is bounded, and the Bolzano-Weierstrass theorem 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. □

For later use, we explicitly state and prove one other property of compact sets.

Proposition 1.41. If $K \subset \mathbb{R}$ is sequentially compact, then K has a maximum and minimum.

Proof. Since K is sequentially compact it is bounded and, by the completeness of $\mathbb{R}$, 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

$$M - \frac{1}{n} < x_n \leq M.$$

It follows (from the 'sandwich' 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. □

1.5.2. Compactness. Next, we give a topological definition of compactness in terms of open sets. If A is a subset of $\mathbb{R}$, an open cover of A is a collection of open sets

$$\{G_i \subset \mathbb{R} : i \in \mathcal{I}\}$$

whose union contains A,

$$\bigcup_{i \in \mathcal{I}} G_i \supset A.$$

A finite subcover of this open cover is a finite collection of sets in the cover

$$\{G_{i_1}, G_{i_2}, \ldots, G_{i_N}\}$$

whose union still contains A,

$$\bigcup_{n=1}^{N} G_{i_n} \supset A.$$

Definition 1.42. A set $K \subset \mathbb{R}$ is compact if every open cover of K has a finite subcover.

We illustrate the definition with several examples.

Example 1.43. The collection of open intervals

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

is an open cover of the natural numbers $\mathbb{N}$ since

$$\bigcup_{n=1}^{\infty} I_n = (0, \infty) \supset \mathbb{N}.$$

However, no finite subcollection $\{I_1, I_2, \ldots, I_N\}$ of intervals covers $\mathbb{N}$ since their union

$$\bigcup_{n=1}^{N} I_n = (0, N+1)$$

does not contain sufficiently large integers with $n \geq N+1$. (A finite subcover that omits some of the intervals I_i for $1 \leq i \leq N$ would have an even smaller union.) Thus, $\mathbb{N}$ is not compact. A similar argument, using the intervals $I_n = (-n, n)$, shows that a compact set must be bounded.

Example 1.44. The collection of open intervals (which get smaller as they get closer to 0)

$$\{I_n : n = 0, 1, 2, 3, \ldots\}, \qquad I_n = \left(\frac{1}{2^n} - \frac{1}{2^{n+1}}, \frac{1}{2^n} + \frac{1}{2^{n+1}}\right)$$

is an open cover of the open interval $(0, 1)$; in fact

$$\bigcup_{n=0}^{\infty} I_n = \left(0, \frac{3}{2}\right) \supset (0, 1).$$

However, no finite subcollection $\{I_0, I_1, I_2, \ldots, I_N\}$ of intervals covers $(0, 1)$ since their union

$$\bigcup_{n=0}^{N} I_n = \left(\frac{1}{2^N} - \frac{1}{2^{N+1}}, \frac{3}{2}\right),$$

does not contain points in $(0, 1)$ that are sufficiently close to 0. Thus, $(0, 1)$ is not compact.

Example 1.45. The collection of open intervals $\{I_n\}$ in Example 1.44 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 $\{I_n, J\}$ of $[0, 1]$ by adding to the I_n an open interval $J = (-\delta, \delta)$ about zero, where $\delta > 0$ can be 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 $\{I_0, I_1, I_2, \ldots, I_N, J\}$ is a finite subcover of $[0, 1]$ since

$$\bigcup_{n=0}^{N} I_n \cup J = \left(-\delta, \frac{3}{2}\right) \supset [0, 1].$$

Points sufficiently close to 0 belong to J, while points further away belong to I_i for some $0 \leq i \leq N$. As this example illustrates, $[0, 1]$ is compact and every open cover of $[0, 1]$ has a finite subcover.

Theorem 1.46. A subset of $\mathbb{R}$ is compact if and only if it is closed and bounded.

This result follows from the Heine-Borel theorem, that every open cover of a closed, bounded interval has a finite subcover, but we omit a detailed proof.

It follows that a subset of $\mathbb{R}$ is sequentially compact if and only if it is compact, since the subset is closed and bounded in either case. We therefore refer to any such set simply as a compact set. We will use the sequential definition of compactness in our proofs.

Chapter 2

Limits of Functions

In this chapter, we define limits of functions and describe some of their properties.

2.1. Limits

We begin with the ϵ-δ definition of the limit of a function.

Definition 2.1. 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 we exclude $x = c$, so the function need not be defined at c for the limit as $x \to c$ to exist. Also note that it follows directly from the definition that

$$\lim_{x \to c} f(x) = L \quad \text{if and only if} \quad \lim_{x \to c} |f(x) - L| = 0.$$

Example 2.2. 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. For $x \in A$, we have from the difference of two squares that $f(x) = \sqrt{x} + 3$, and

$$|f(x) - 6| = \left|\sqrt{x} - 3\right| = \left|\frac{x - 9}{\sqrt{x} + 3}\right| \leq \frac{1}{3}|x - 9|.$$

Thus, if $\delta = 3\epsilon$, then $|x - 9| < \delta$ and $x \in A$ implies that $|f(x) - 6| < \epsilon$.

We can rephrase the ϵ-δ definition of limits in terms of neighborhoods. Recall from Definition 1.16 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. A set U is a punctured (or deleted) neighborhood of c if $U \supset (c-\delta, c) \cup (c, c+\delta)$ for some $\delta > 0$, and $(c-\delta, c) \cup (c, c+\delta)$ 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 2.3. 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 2.1 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 that maps into V, which verifies Definition 2.3. Conversely, if Definition 2.3 holds and $\epsilon > 0$, let $V = (L-\epsilon, L+\epsilon)$ be an ϵ-neighborhood of L. Then there is a punctured neighborhood U of c that maps into V and U contains a punctured δ-neighborhood of c, which verifies Definition 2.1.

The next theorem gives an equivalent sequential characterization of the limit.

Theorem 2.4. 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. 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 2.1, 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. 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. □

This theorem gives a way to show that a limit of a function does not exist.

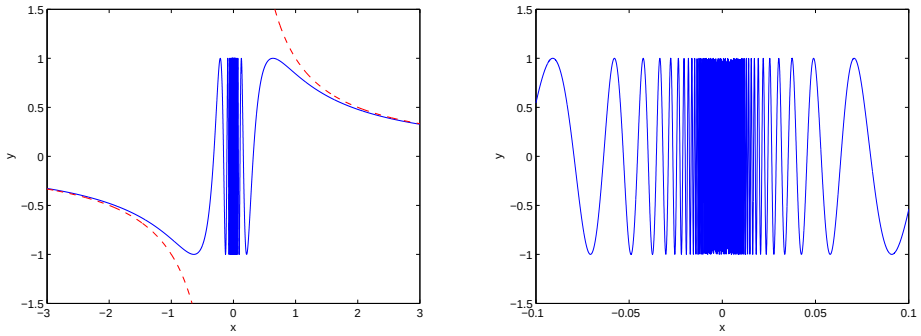

Figure 1. A plot of the function $y = \sin(1/x)$, with the hyperbola $y = 1/x$ shown in red, and a detail near the origin.

Corollary 2.5. 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 $(f(x_n))$ does not converge.

Example 2.6. 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 2.5 implies that the limit does not exist.

Example 2.7. 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, consider the non-zero sequence (x_n) given by $x_n = 1/n$. Then $1/n \to 0$ but the sequence of values (n) doesn't converge.

Example 2.8. 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.

2.2. 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 2.9 (Right and left limits)**.** 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 (right limit)

$$\lim_{x\to c^+} f(x) = L$$

if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$c < x < c + \delta \text{ and } x \in A \text{ implies that } |f(x) - L| < \epsilon,$$

and (left limit)

$$\lim_{x\to c^-} f(x) = L$$

if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$c - \delta < x < c \text{ and } x \in A \text{ implies that } |f(x) - L| < \epsilon.$$

Example 2.10. For the sign function in Example 2.6, we have

$$\lim_{x\to 0^+} \operatorname{sgn} x = 1, \qquad \lim_{x\to 0^-} \operatorname{sgn} x = -1.$$

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 2.11 (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 this definition 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 2.12. 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 2.13 (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 2.14. We have

$$\lim_{x\to 0} \frac{1}{x^2} = \infty, \qquad \lim_{x\to\infty} \frac{1}{x^2} = 0.$$

Example 2.15. 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 2.16. 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 2.17. 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?

2.3. Properties of limits

The properties of limits of functions follow immediately from the corresponding properties of sequences and the sequential characterization of the limit in Theorem 2.4. We can also prove them directly from the ϵ-δ definition of the limit, and we shall do so in a few cases below.

2.3.1. Uniqueness and boundedness. The following result might be taken for-granted, but it requires proof.

Proposition 2.18. 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 $L_1, L_2 \in \mathbb{R}$. Then 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

$$|L_1 - L_2| \le |L_1 - f(x)| + |f(x) - L_2| < \epsilon.$$

Since this holds for arbitrary $\epsilon > 0$, we must have $L_1 = L_2$. □

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.

Definition 2.19. A function $f : A \to \mathbb{R}$ is bounded on $B \subset A$ if there exists $M \ge 0$ such that

$$|f(x)| \le M \text{ for every } x \in B.$$

A function is bounded if it is bounded on its domain.

Equivalently, f is bounded on B if $f(B)$ is a bounded subset of $\mathbb{R}$.

Example 2.20. The function $f : (0, 1] \to \mathbb{R}$ defined by $f(x) = 1/x$ is unbounded, but it is bounded on any interval $[\delta, 1]$ with $0 < \delta < 1$. The function $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = x^2$ is unbounded, but is it bounded on any finite interval $[a, b]$.

If a function has a limit as $x \to c$, it must be locally bounded at c, as stated in the next proposition.

Proposition 2.21. Suppose that $f : A \to \mathbb{R}$ and c is an accumulation point of A. If $\lim_{x\to c} f(x)$ exists, then there is a punctured neighborhood U of c such that f is bounded on $A \cap U$.

Proof. Suppose that $f(x) \to L$ as $x \to c$. 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 $U = (c - \delta, c) \cup (c, c + \delta)$, which is a punctured neighborhood of c. Then for $x \in A \cap U$, we have

$$|f(x)| \leq |f(x) - L| + |L| < 1 + |L|,$$

so f is bounded on $A \cap U$. □

2.3.2. Algebraic properties. Limits of functions respect algebraic operations.

Theorem 2.22. 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

$$\begin{aligned}
&\lim_{x\to c} kf(x) = kL && \text{for every } k \in \mathbb{R},\\
&\lim_{x\to c} [f(x) + g(x)] = L + M,\\
&\lim_{x\to c} [f(x)g(x)] = LM,\\
&\lim_{x\to c} \frac{f(x)}{g(x)} = \frac{L}{M} && \text{if } M \neq 0.
\end{aligned}$$

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

$$\begin{aligned}
&0 < |x - c| < \delta_1 \text{ and } x \in A \text{ implies that } |f(x) - L| < \epsilon/2,\\
&0 < |x - c| < \delta_2 \text{ and } x \in A \text{ implies that } |g(x) - M| < \epsilon/2,
\end{aligned}$$

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 2.21) 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

$$\begin{aligned}
&0 < |x - c| < \delta_1 \text{ and } x \in A \text{ implies that } |f(x) - L| < \epsilon/(2K),\\
&0 < |x - c| < \delta_2 \text{ and } x \in A \text{ implies that } |g(x) - M| < \epsilon/(2|L| + 1).
\end{aligned}$$

Let $\delta = \min(\delta_0, \delta_1, \delta_2) > 0$. Then for $0 < |x - c| < \delta$ and $x \in A$,

$$\begin{aligned} |f(x)g(x) - LM| &= |(f(x) - L)\, g(x) + L\,(g(x) - M)| \\ &\le |f(x) - L|\ |g(x)| + |L|\,|g(x) - M| \\ &< \frac{\epsilon}{2K} \cdot K + |L| \cdot \frac{\epsilon}{2|L| + 1} \\ &< \epsilon, \end{aligned}$$

which proves that $\lim(fg) = \lim f \lim g$. □

2.3.3. Order properties. As for limits of sequences, limits of functions preserve (non-strict) inequalities.

Theorem 2.23. Suppose that $f, g : A \to \mathbb{R}$ and c is an accumulation point of A. If

$$f(x) \le g(x) \qquad \text{for all } x \in A,$$

and $\lim_{x\to c} f(x)$, $\lim_{x\to c} g(x)$ exist, then

$$\lim_{x\to c} f(x) \le \lim_{x\to c} g(x).$$

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

$$\begin{aligned} |f(x) - L| < \epsilon \qquad &\text{if } x \in A \text{ and } 0 < |x - c| < \delta_1, \\ |g(x) - M| < \epsilon \qquad &\text{if } x \in A \text{ and } 0 < |x - c| < \delta_2. \end{aligned}$$

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

$$\begin{aligned} f(x) - g(x) &= [f(x) - L] + L - M + [M - g(x)] \\ &> L - M - 2\epsilon \\ &> 0, \end{aligned}$$

which contradicts the assumption that $f(x) \le g(x)$. □

Finally, we state a useful "sandwich" or "squeeze" criterion for the existence of a limit.

Theorem 2.24. Suppose that $f, g, h : A \to \mathbb{R}$ and c is an accumulation point of A. If

$$f(x) \le g(x) \le h(x) \qquad \text{for all } x \in A$$

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

$$0 \le f(x) \le g(x) \quad \text{or} \quad |f(x)| \le g(x)$$

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 2.24 to have the same limit.

Example 2.25. We have

$$-1 \le \sin\left(\frac{1}{x}\right) \le 1 \qquad \text{for all } x \neq 0$$

and

$$\lim_{x\to 0}(-1) = -1, \qquad \lim_{x\to 0} 1 = 1,$$

but

$$\lim_{x\to 0} \sin\left(\frac{1}{x}\right) \qquad \text{does not exist.}$$

Chapter 3

Continuous Functions

In this chapter, we define continuous functions and study their properties.

3.1. Continuity

According to the definition introduced by Cauchy, and developed by Weierstrass, continuous functions are functions that take nearby values at nearby points.

Definition 3.1. 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 on a set $B \subset A$ if it is continuous at every point in B, and continuous if it is continuous at every point of its domain A.

The definition of continuity at a point may be stated in terms of neighborhoods as follows.

Definition 3.2. 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. We leave it as an exercise to prove that these definitions are equivalent.

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 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 3.3. 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 3.4. 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 3.5. Suppose that

$$A = \left\{ 0, 1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots \right\}$$

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 2.4.

Theorem 3.6. 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 3.7. 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 $0 \leq x < \infty$,

$$|f(x) - f(c)| = \left|\sqrt{x} - \sqrt{c}\right| = \left| \frac{x - c}{\sqrt{x} + \sqrt{c}} \right| \leq \frac{1}{\sqrt{c}} |x - c|,$$

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 3.8. 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 $|\sin x| \le |x|$:

$$\begin{aligned} |\sin x - \sin c| &= \left| 2\cos\left(\frac{x+c}{2}\right) \sin\left(\frac{x-c}{2}\right) \right| \\ &\le 2\left|\sin\left(\frac{x-c}{2}\right)\right| \\ &\le |x - c|. \end{aligned}$$

It follows that we can take $\delta = \epsilon$ in the definition of continuity for every $c \in \mathbb{R}$.

Example 3.9. The sign function $\operatorname{sgn} : \mathbb{R} \to \mathbb{R}$, defined by

$$\operatorname{sgn} x = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -1 & \text{if } x < 0, \end{cases}$$

is not continuous at 0 since $\lim_{x\to 0} \operatorname{sgn} x$ does not exist (see Example 2.6). 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 3.10. 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 2.7). Neither the left or right limits of f at 0 exist either, and we say that f has an essential discontinuity at 0.

Example 3.11. 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 3.20 below) but discontinuous at 0 because $\lim_{x\to 0} f(x)$ does not exist (see Example 2.8).

Example 3.12. 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 3.21 below. To prove continuity at 0, note that for $x \neq 0$,

$$|f(x) - f(0)| = |x\sin(1/x)| \le |x|,$$

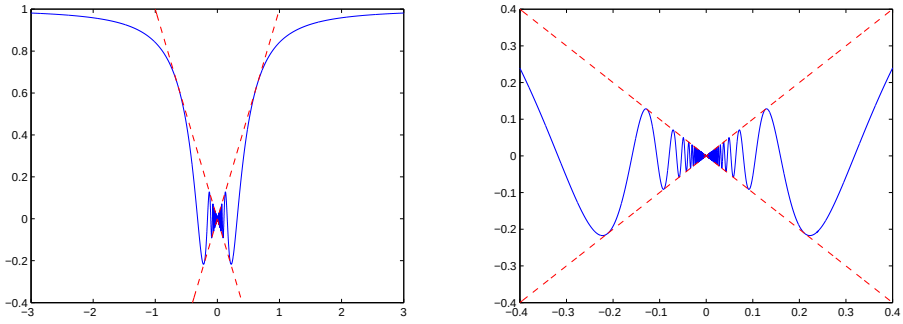

Figure 1. A plot of the function $y = x\sin(1/x)$ and a detail near the origin with the lines $y = \pm x$ shown in red.

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.

Example 3.13. 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 $f(c) = 1$. In fact, taking a rational sequence (x_n) and an irrational sequence $(\tilde{x}_n)$ that converge to c, we see that $\lim_{x\to c} f(x)$ does not exist for any $c \in \mathbb{R}$.

Example 3.14. The Thomae function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1/q & \text{if } x = p/q \text{ where } p \text{ and } q > 0 \text{ are relatively prime,} \\ 0 & \text{if } x \notin \mathbb{Q} \text{ or } x = 0 \end{cases}$$

is continuous at 0 and every irrational number and discontinuous at every nonzero rational number. See Figure 2 for a plot.

We can give a rough classification of a discontinuity 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 $L \neq f(c)$, in which case we can make f continuous at c by redefining $f(c) = L$ (see Example 3.12).

(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 3.9).

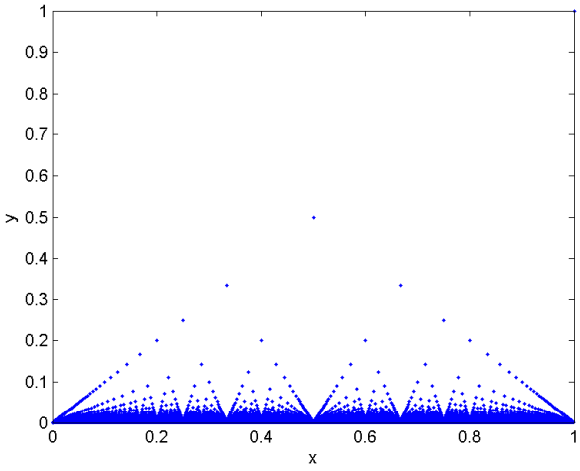

Figure 2. A plot of the Thomae function on $[0, 1]$

(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 3.10, 3.11, 3.13).

3.2. Properties of continuous functions

The basic properties of continuous functions follow from those of limits.

Theorem 3.15. 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 $g(c) \neq 0$ then f/g is continuous at c.

Proof. This result follows immediately Theorem 2.22. □

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 3.16. 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 3.15 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 3.17. 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 if $f : A \to \mathbb{R}$ and $g : B \to \mathbb{R}$ where $f(A) \subset B$, meaning that the domain of g contains the range of f, then we define the composition $g \circ f : A \to \mathbb{R}$ by

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

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 3.18. 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

$$|y - f(c)| < \eta \text{ and } y \in B \text{ implies that } |g(y) - g\left(f(c)\right)| < \epsilon.$$

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

$$|x - c| < \delta \text{ and } x \in A \text{ implies that } |g\left(f(x)\right) - g\left(f(c)\right)| < \epsilon,$$

which proves that $g \circ f$ is continuous at c. □

Corollary 3.19. 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 3.20. 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 3.21. 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 3.12, f is also continuous at 0, so f is continuous on $\mathbb{R}$.

3.3. Uniform continuity

Uniform continuity is a subtle but powerful strengthening of continuity.

Definition 3.22. 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, as we explain next.

If a function is continuous on A, then given $\epsilon > 0$ there exists $\delta(c) > 0$ for every $c \in A$ such that

$$|x - c| < \delta(c) \text{ and } x \in A \text{ implies that } |f(x) - f(c)| < \epsilon.$$

In general, $\delta(c)$ depends on both ϵ and c, but we don't show the ϵ-dependence explicitly since we're thinking of ϵ as fixed. If

$$\inf_{c \in A} \delta(c) = 0$$

however we choose $\delta(c)$, then no $\delta_0 > 0$ depending only on ϵ works simultaneously for all $c \in A$. In that case, the function is continuous on A but not uniformly continuous.

Before giving examples, we state a sequential condition for uniform continuity to fail.

Proposition 3.23. 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 3.22 cannot hold for any $\delta > 0$ if $\epsilon = \epsilon_0$, so f is not uniformly continuous. □

Example 3.24. Example 3.8 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 3.25. 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 3.26. 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 3.23 that f is not uniformly continuous on $\mathbb{R}$. The problem here is that, for given $\epsilon > 0$, we need to make $\delta(c)$ smaller as c gets larger to prove the continuity of f at c, and $\delta(c) \to 0$ as $c \to \infty$.

Example 3.27. The function $f : (0, 1] \to \mathbb{R}$ defined by

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

is continuous but not uniformly continuous on $(0, 1]$. We have already proved that f is continuous on $(0, 1]$ (it's a rational function whose denominator x is nonzero in $(0, 1]$). To prove that f is not uniformly continuous, 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 $x_n \to 0$, $y_n \to 0$, and $|x_n - y_n| \to 0$ 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 3.23 that f is not uniformly continuous on $(0, 1]$. The problem here is that, for given $\epsilon > 0$, we need to make $\delta(c)$ smaller as c gets closer to 0 to prove the continuity of f at c, and $\delta(c) \to 0$ 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 3.28. 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 $x_n \to 0$, $y_n \to 0$, and $|x_n - y_n| \to 0$ 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 a domain that is either unbounded or not closed. We will prove in Section 3.5 that a continuous function on a closed, bounded set is uniformly continuous.

3.4. Continuous functions and open sets

Let $f : A \to \mathbb{R}$ be a function. Recall that if $B \subset A$, the set

$$f(B) = \{y \in \mathbb{R} : y = f(x) \text{ for some } x \in B\}$$

is called the image of B under f, and if $C \subset \mathbb{R}$, the set

$$f^{-1}(C) = \{x \in A : f(x) \in C\}$$

is called the inverse image or preimage of C under f. Note that $f^{-1}(C)$ is a well-defined set even if the function f does not have an inverse.

Example 3.29. Suppose $f : \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = x^2$. If $I = (1, 4)$, then

$$f(I) = (1, 16), \qquad f^{-1}(I) = (-2, -1) \cup (1, 2).$$

Note that we get two intervals in the preimage because f is two-to-one on $f^{-1}(I)$. If $J = (-1, 1)$, then

$$f(J) = [0, 1), \qquad f^{-1}(J) = (-1, 1).$$

In the previous example, the preimages of the open sets I, J under the continuous function f are open, but the image of J under f isn't open. 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 implies that the function is continuous at every point of its domain.

Recall from Section 1.2 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 1.22, 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 3.30. 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

$$V_\epsilon(f(c)) = (f(c) - \epsilon, f(c) + \epsilon)$$

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

$$f(A \cap U_\delta(c)) \subset V_\epsilon(f(c)).$$

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. □

3.5. 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 is compact.

Theorem 3.31. If $f : K \to \mathbb{R}$ is continuous and $K \subset \mathbb{R}$ is compact, then $f(K)$ is compact.

Proof. 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.

Let us also give an alternative proof based on the Heine-Borel property. Suppose that $\{V_i : i \in I\}$ is an open cover of $f(K)$. Since f is continuous, Theorem 3.30 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. □

Note that compactness is essential here; it is not true, in general, that a continuous function maps closed sets to closed sets.

Example 3.32. Define $f : [0, \infty) \to \mathbb{R}$ by

$$f(x) = \frac{1}{1 + x^2}.$$

Then $[0, \infty)$ is closed but $f([0, \infty)) = (0, 1]$ is not.

The following result is the most important property of continuous functions on compact sets.

Theorem 3.33 (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. Since $f(K)$ is compact, Theorem 1.40 implies that it is bounded, which means that f is bounded on K. Proposition 1.41 implies that the maximum M and minimum m of $f(K)$ 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 3.34. 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 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 3.35. 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 3.36. 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 $f(x) \neq 0$, $f(x) \neq 1$ 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 3.37. 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 3.33 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 $x = 2/\pi$.

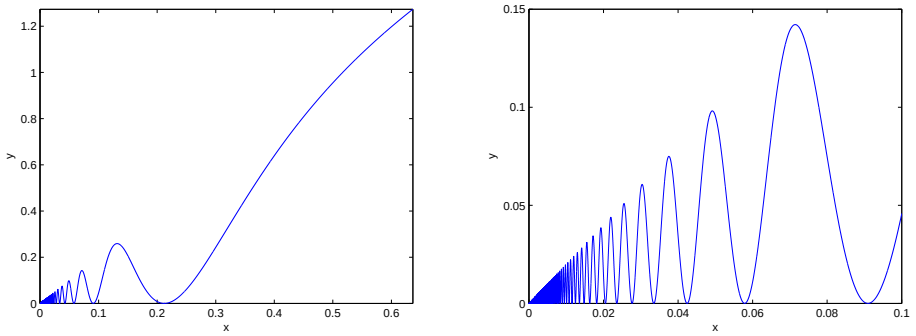

Figure 3. A plot of the function $y = x + x\sin(1/x)$ on $[0, 2/\pi]$ and a detail near the origin.

Finally, we prove that continuous functions on compact sets are uniformly continuous

Theorem 3.38. 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 3.23 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. □

Example 3.39. The function $f : [0, 2/\pi] \to \mathbb{R}$ defined in Example 3.37 is uniformly continuous on $[0, 2/\pi]$ since it is is continuous and $[0, 2/\pi]$ is compact.

3.6. 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 3.40. 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 $a < c < b$. □

We give some examples to show that all of the hypotheses in this theorem are necessary.

Example 3.41. 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 3.40 fails if the domain of f is not a connected interval $[a, b]$.

Example 3.42. Define $f : [-1, 1] \to \mathbb{R}$ by

$$f(x) = \begin{cases} -1 & \text{if } -1 \le x < 0 \\ 1 & \text{if } 0 \le x \le 1 \end{cases}$$

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 3.40 fails for discontinuous functions.

Example 3.43. 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 3.40 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, so 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 3.40. 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 3.40 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 3.44 (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 3.40 implies that $g(c) = 0$ for some $a < c < b$, meaning that $f(c) = d$. □

As one consequence of our previous results, we prove that a continuous function maps compact intervals to compact intervals.

Theorem 3.45. 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 3.33 implies that $m \le f(x) \le M$ for all $x \in [a, b]$, where m and M are the maximum and minimum values of f, 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 $c \le d$ or $J = [d, c]$ if $d < c$. Then $J \subset [a, b]$, and Theorem 3.44 implies that f takes on all values in $[m, M]$ on J. It follows that $f([a, b]) \supset [m, M]$, so $f([a, b]) = [m, M]$. □

First we give an example to illustrate the theorem.

Example 3.46. 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

$$f([-1, 1]) = [-M, M], \qquad M = \frac{2}{3\sqrt{3}}.$$

This example illustrates that $f([a, b]) \neq [f(a), f(b)]$ 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 3.45 to hold.

Example 3.47. Let $\operatorname{sgn} : [-1, 1] \to \mathbb{R}$ be the sign function defined in Example 2.6. 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 3.48. In Example 3.41, 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 3.49. The continuous function $f : [0, \infty) \to \mathbb{R}$ in Example 3.32 maps the unbounded, closed interval $[0, \infty)$ to a 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, it follows from the fact that a continuous function maps compact intervals to compact intervals that it maps intervals to intervals (where the intervals may be open, closed, half-open, bounded, or unbounded). We omit a detailed proof.

3.7. Monotonic functions

Monotonic functions have continuity properties that are not shared by general functions.

Definition 3.50. 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 3.51. 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 $L - \epsilon < y_0 \le L$, 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 $f(x_0) \le f(x) \le L$ since f is increasing and L is an upper bound of E. It follows that

$$L - \epsilon < f(x) \le L \qquad \text{if } c - \delta < x < c,$$

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. □

Similarly, if $I = [a, b]$ is a closed interval and f is monotonic on I, then the left limit $\lim_{x\to b^-} f(x)$ exists at the right endpoint, although it may not equal $f(b)$, and the right limit $\lim_{x\to a^+} f(x)$ exists at the left endpoint, although it may not equal $f(a)$.

Corollary 3.52. 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

$$f(x) \le f(c) \le f(y) \qquad \text{for all } x, y \in I \text{ with } x < c < y,$$

and since limits preserve inequalities

$$\lim_{x\to c^-} f(x) \le f(c) \le \lim_{x\to c^+} f(x).$$

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. □

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 $\mathbb{R}$.

Chapter 4

Differentiable Functions

A differentiable function is a function that can be approximated locally by a linear function.

4.1. The derivative

Definition 4.1. 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 $c \in (a, b)$.

4.1.1. Examples of derivatives. Let us give a number of examples that illustrate differentiable and non-differentiable functions.

Example 4.2. 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 4.3. 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,

$$f'(0) = \lim_{h\to 0}\frac{f(h)}{h}.$$

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, so does the limit

$$\lim_{h\to 0}\left[\frac{f(h) - f(0)}{h}\right] = 0,$$

and f is differentiable at 0 with $f'(0) = 0$.

Next, we consider some examples of non-differentiability at discontinuities, corners, and cusps.

Example 4.4. 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

$$\begin{aligned}\lim_{h\to 0}\left[\frac{f(c+h)-f(c)}{h}\right] &= \lim_{h\to 0}\left[\frac{1/(c+h)-1/c}{h}\right] \\ &= \lim_{h\to 0}\left[\frac{c-(c+h)}{hc(c+h)}\right] \\ &= -\lim_{h\to 0}\frac{1}{c(c+h)} \\ &= -\frac{1}{c^2}.\end{aligned}$$

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 4.5. The sign function $f(x) = \operatorname{sgn} x$, defined in Example 2.6, 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 4.6. 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.

Example 4.7. 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, we use the identity for the difference of cubes,

$$a^3 - b^3 = (a-b)(a^2+ab+b^2),$$

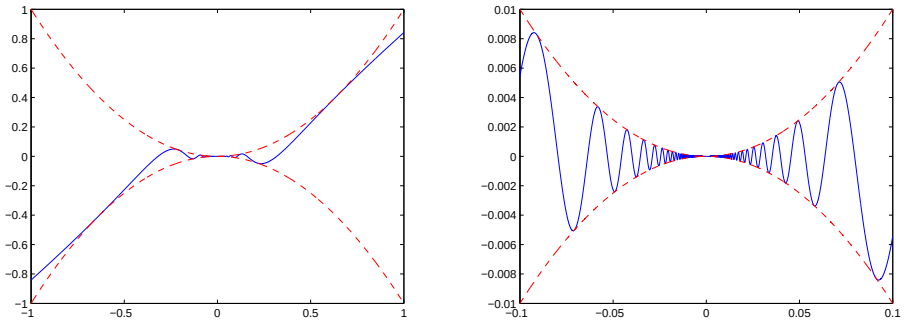

Figure 1. A plot of the function $y = x^2 \sin(1/x)$ and a detail near the origin with the parabolas $y = \pm x^2$ shown in red.

and get for $c \neq 0$ that

$$\begin{aligned}\lim_{h\to 0}\left[\frac{f(c+h)-f(c)}{h}\right] &= \lim_{h\to 0}\frac{(c+h)^{1/3}-c^{1/3}}{h}\\ &= \lim_{h\to 0}\frac{(c+h)-c}{h\left[(c+h)^{2/3}+(c+h)^{1/3}c^{1/3}+c^{2/3}\right]}\\ &= \lim_{h\to 0}\frac{1}{(c+h)^{2/3}+(c+h)^{1/3}c^{1/3}+c^{2/3}}\\ &= \frac{1}{3c^{2/3}}.\end{aligned}$$

However, f is not differentiable at 0, since

$$\lim_{h\to 0}\frac{f(h)-f(0)}{h} = \lim_{h\to 0}\frac{1}{h^{2/3}},$$

which does not exist.

Finally, we consider some examples of highly oscillatory functions.

Example 4.8. 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 4.9. Define $f : \mathbb{R} \to \mathbb{R}$ by

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

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.

4.1.2. Derivatives as linear approximations. Another way to view Definition 4.1 is to write

$$f(c + h) = f(c) + f'(c)h + r(h)$$

as the sum of a linear 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 4.10. 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, as in Definition 4.1, and define

$$r(h) = f(c+h) - f(c) - f'(c)h.$$

Then

$$\lim_{h\to 0} \frac{r(h)}{h} = \lim_{h\to 0} \left[\frac{f(c+h) - f(c)}{h} - f'(c) \right] = 0.$$

Conversely, suppose that $f(c+h) = f(c) + Ah + r(h)$ where $r(h)/h \to 0$ as $h \to 0$. Then

$$\lim_{h\to 0} \left[\frac{f(c+h) - f(c)}{h} \right] = \lim_{h\to 0} \left[A + \frac{r(h)}{h} \right] = A,$$

which proves that f is differentiable at c with $f'(c) = A$. □

Example 4.11. In Example 4.2 with $f(x) = x^2$,

$$(c+h)^2 = c^2 + 2ch + h^2,$$

and $r(h) = h^2$, which goes to zero at a quadratic rate as $h \to 0$.

Example 4.12. In Example 4.4 with $f(x) = 1/x$,

$$\frac{1}{c+h} = \frac{1}{c} - \frac{1}{c^2}h + r(h),$$

for $c \neq 0$, where the quadratically small remainder is

$$r(h) = \frac{h^2}{c^2(c+h)}.$$

4.1.3. Left and right derivatives. We can use left and right limits to define one-sided derivatives, for example at the endpoint of an interval, but for the most part we will consider only two-sided derivatives defined at an interior point of the domain of a function.

Definition 4.13. Suppose $f : [a, b] \to \mathbb{R}$. 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 exist at c and are equal.

Example 4.14. 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

$$f(x) = \begin{cases} x^2 & \text{if } 0 \le x \le 1, \\ 0 & \text{if } x > 1, \\ 1/x & \text{if } x < 0. \end{cases}$$

For this extended function we have $f'(1^+) = 0$, which is not equal to $f'(1^-)$, and $f'(0^-)$ does not exist, so it is not differentiable at 0 or 1.

Example 4.15. The absolute value function $f(x) = |x|$ in Example 4.6 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.

4.2. Properties of the derivative

In this section, we prove some basic properties of differentiable functions.

4.2.1. Differentiability and continuity. First we discuss the relation between differentiability and continuity.

Theorem 4.16. 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

$$\begin{aligned} \lim_{h\to 0} f(c+h) - f(c) &= \lim_{h\to 0} \left[\frac{f(c+h) - f(c)}{h} \cdot h \right] \\ &= \lim_{h\to 0} \left[\frac{f(c+h) - f(c)}{h} \right] \cdot \lim_{h\to 0} h \\ &= f'(c) \cdot 0 \\ &= 0, \end{aligned}$$

which implies that f is continuous at c. □

For example, the sign function in Example 4.5 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 4.6, 4.7, 4.8 are continuous but not differentiable at 0. Example 5.24 describes a function that is continuous on $\mathbb{R}$ but not differentiable anywhere.

In Example 4.9, 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' needn't be continuous. This leads to the following definition.

Definition 4.17. 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 any interval (a, b). As Example 4.9 illustrates, functions that are differentiable but not continuously differentiable may still behave in rather pathological ways. On the other hand, continuously differentiable functions, whose tangent lines vary continuously, are relatively well-behaved.

4.2.2. Algebraic properties of the derivative. Next, we state the linearity of the derivative and the product and quotient rules.

Theorem 4.18. 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 2.22. For the product rule, we write

$$\begin{aligned}
(fg)'(c) &= \lim_{h\to 0}\left[\frac{f(c+h)g(c+h) - f(c)g(c)}{h}\right] \\
&= \lim_{h\to 0}\left[\frac{(f(c+h) - f(c))\,g(c+h) + f(c)\,(g(c+h) - g(c))}{h}\right] \\
&= \lim_{h\to 0}\left[\frac{f(c+h) - f(c)}{h}\right] \lim_{h\to 0} g(c+h) + f(c) \lim_{h\to 0}\left[\frac{g(c+h) - g(c)}{h}\right] \\
&= f'(c)g(c) + f(c)g'(c),
\end{aligned}$$

where we have used the properties of limits in Theorem 2.22 and Theorem 4.18, 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 4.21 below.) □

Example 4.19. 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 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.

4.2.3. The chain rule. The chain rule states the differentiability of a composition of functions. 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. We will prove the chain rule by making this observation rigorous.

Theorem 4.20 (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

$$g(f(c)+k) = g(f(c)) + g'(f(c))\, k + s(k), \qquad \lim_{k\to 0} \frac{s(k)}{k} = 0.$$

It follows that

$$\begin{aligned}(g \circ f)(c+h) &= g(f(c) + f'(c)h + r(h)) \\ &= g(f(c)) + g'(f(c))\,(f'(c)h + r(h)) + s(f'(c)h + r(h)) \\ &= g(f(c)) + g'(f(c))\, f'(c)h + t(h)\end{aligned}$$

where

$$t(h) = r(h) + s(\phi(h)), \qquad \phi(h) = f'(c)h + r(h).$$

Then, since $r(h)/h \to 0$ as $h \to 0$,

$$\lim_{h\to 0} \frac{t(h)}{h} = \lim_{h\to 0} \frac{s(\phi(h))}{h}.$$

We claim that this is limit is zero, and then it follows from Proposition 4.10 that $g \circ f$ is differentiable at c with

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

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

$$\frac{s(\phi(h))}{h} \sim \frac{s(f'(c)h)}{h} \to 0 \qquad \text{as } h \to 0.$$

To prove this in detail, let $\epsilon > 0$ be given. We want to show that there exists $\delta > 0$ such that

$$\left| \frac{s(\phi(h))}{h} \right| < \epsilon \qquad \text{if } 0 < |h| < \delta.$$

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 to avoid a division by 0 if $f'(c) = 0$.) Next, 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

$$\begin{aligned} |\phi(h)| &\le |f'(c)|\,|h| + |r(h)| \\ &< |f'(c)|\,|h| + (|f'(c)|+1)|h| \\ &< (2|f'(c)|+1)\,|h|. \end{aligned}$$

Define $\delta_2 > 0$ by

$$\delta_2 = \frac{\eta}{2|f'(c)|+1},$$

and let $\delta = \min(\delta_1, \delta_2) > 0$. If $0 < |h| < \delta$, then $|\phi(h)| < \eta$ and

$$|\phi(h)| < (2|f'(c)|+1)\,|h|.$$

It follows that for $0 < |h| < \delta$

$$|s\,(\phi(h))\,| < \frac{\epsilon|\phi(h)|}{2|f'(c)|+1} < \epsilon|h|.$$

(If $\phi(h) = 0$, then $s(\phi(h)) = 0$, so the inequality holds in that case also.) This proves that

$$\lim_{h\to 0} \frac{s\,(\phi(h))}{h} = 0.$$

□

Example 4.21. 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 4.4). It follows that $1/f = g \circ f$ is differentiable at c with

$$\left(\frac{1}{f}\right)'(c) = -\frac{f'(c)}{f(c)^2}.$$

4.2.4. The derivative of inverse functions. 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 4.22. 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 $B = f\,(A)$. If f is differentiable at an interior point $c \in A$ with $f'(c) \neq 0$, $f(c)$ is an interior point of B, and f^{-1} is differentiable at $f(c)$, then

$$(f^{-1})'\,(f(c)) = \frac{1}{f'(c)}.$$

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 $f'(c) = 0$. □

Alternatively, setting $d = f(c)$, we can write the result as

$$(f^{-1})'(d) = \frac{1}{f'(f^{-1}(d))}.$$

The following example illustrates the necessity of the condition $f'(c) \neq 0$ for the differentiability of the inverse.

Example 4.23. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^3$. Then f is strictly increasing, one-to-one, and onto with inverse $f^{-1} : \mathbb{R} \to \mathbb{R}$ 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 $y = x^3$,

$$(f^{-1})'(y) = \frac{1}{3y^{2/3}},$$

in agreement with Example 4.7.

Proposition 4.22 is not entirely satisfactory because it assumes the differentiability of f^{-1} at $f(c)$. One can show that if $f : I \to \mathbb{R}$ is a continuous and one-to-one function on an interval I, then f is strictly monotonic and f^{-1} is also continuous and strictly monotonic. In that case, f^{-1} is differentiable at $f(c)$ if f is differentiable at c and $f'(c) \neq 0$. We omit the proof of these statements.

Another condition for the existence and differentiability of f^{-1}, which generalizes to functions of several variables, is given by the inverse function theorem: If f is differentiable in a neighborhood of c, $f'(c) \neq 0$, *and* f' is continuous at c, then f has a local inverse f^{-1} defined in a neighborhood of $f(c)$ and the inverse is differentiable at $f(c)$ with derivative given by Proposition 4.22.

4.3. Extreme values

Definition 4.24. 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 3.33 states that a continuous function on a compact set has a global maximum and minimum. The following fundamental result goes back to Fermat.

Theorem 4.25. Suppose that $f : A \to \mathbb{R}$ 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 $f'(c) = 0$. □

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 an inequality condition on the derivative. If $f : [a, b] \to \mathbb{R}$, the right derivative of f exists at a, and f has a local maximum at a, then $f(x) \leq f(a)$ for $a \leq x < a + \delta$, so $f'(a^+) \leq 0$. 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.

Definition 4.26. Suppose that $f : A \to \mathbb{R}$. 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 4.25 limits the search for points where f has a maximum or minimum value on A to:

(1) Boundary points of A;

(2) Interior points where f is not differentiable;

(3) Stationary points of f.

4.4. The mean value theorem

We begin by proving a special case.

Theorem 4.27 (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 3.33, 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 4.25 implies that $f'(c) = 0$. □

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 4.28 (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. □

Graphically, this result says that there is point $a < c < b$ at which the slope of the graph $y = f(x)$ is equal to the slope of the chord between the endpoints $(a, f(a))$ and $(b, f(b))$.

Analytically, the mean value theorem is a key result that connects the local behavior of a function, described by the derivative $f'(c)$, to its global behavior, described by the difference $f(b) - f(a)$. As a first application we prove a converse to the obvious fact that the derivative of a constant functions is zero.

Theorem 4.29. 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). □

Corollary 4.30. 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 $(f - g)' = 0$. □

We can also use the mean value theorem to relate the monotonicity of a differentiable function with the sign of its derivative.

Theorem 4.31. 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, 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

$$\frac{f(y) - f(x)}{y - x} = f'(c) \geq 0$$

for some $x < c < y$, 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 $-f$. □

Note that if f is strictly increasing, it does *not* follow that $f'(x) > 0$ for every $x \in (a, b)$.

Example 4.32. 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 4.31 implies that f is strictly increasing near c. This conclusion may fail if f is not continuously differentiable at c.

Example 4.33. The function

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

is differentiable, but not continuously differentiable, at 0 and $f'(0) = 1/2 > 0$. However, f is not increasing in any neighborhood of 0 since

$$f'(x) = \frac{1}{2} - \cos\left(\frac{1}{x}\right) + 2x \sin\left(\frac{1}{x}\right)$$

is continuous for $x \neq 0$ and takes negative values in any neighborhood of 0, so f is strictly decreasing near those points.

4.5. 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 4.34. 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 4.35. If $P(x)$ is a polynomial of degree n, then $P_n(x) = P(x)$.

Example 4.36. 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 4.37. 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 $P_{2n+1} = P_{2n}$.

Example 4.38. 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 4.39. 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 4.40. 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 $n = 0$.

The proof is a bit tricky, but the essential idea 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 4.41 (Taylor)**.** Suppose $f : (a, b) \to \mathbb{R}$ 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. □

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 an (unknown) intermediate point ξ between c and x, instead of at c.

Example 4.42. 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 Taylor's theorem not only proves the limit, but it also gives an explicit upper bound for the difference between $(1-\cos x)/x^2$ and its limit $1/2$.

Chapter 5

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.

5.1. Pointwise convergence

Pointwise convergence defines the convergence of functions in terms of the convergence of their values at each point of their domain.

Definition 5.1. 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 natural 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 5.2. 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 their pointwise limit f is not. Thus, pointwise convergence does not, in general, preserve boundedness.

Example 5.3. Suppose that $f_n : [0,1] \to \mathbb{R}$ is defined by $f_n(x) = x^n$. If $0 \le x < 1$, 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

$$f(x) = \begin{cases} 0 & \text{if } 0 \le x < 1, \\ 1 & \text{if } x = 1. \end{cases}$$

Although each f_n is continuous on $[0,1]$, their pointwise limit f is not (it is discontinuous at 1). Thus, pointwise convergence does not, in general, preserve continuity.

Example 5.4. Define $f_n : [0,1] \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} 2n^2 x & \text{if } 0 \le x \le 1/(2n) \\ 2n^2(1/n - x) & \text{if } 1/(2n) < x < 1/n, \\ 0 & 1/n \le x \le 1. \end{cases}$$

If $0 < x \le 1$, then $f_n(x) = 0$ for all $n \ge 1/x$, 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 of functions need not be bounded, even if it converges to zero.

Example 5.5. 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 is because the derivative of a small, rapidly oscillating function may be large.

Example 5.6. 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 $\mathbb{R}$. The limit $|x|$ not differentiable at 0 even though all of the f_n are differentiable on $\mathbb{R}$. (The f_n "round off" the corner in the absolute value function.)

Example 5.7. 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, which we do not prove here, $f_n \to e^x$ pointwise on $\mathbb{R}$.

5.2. 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 5.8. 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 for some $\epsilon > 0$ one needs to choose arbitrarily large N for different $x \in A$, meaning that there are sequences of values which converge arbitrarily slowly on A, then a pointwise convergent sequence of functions is not uniformly convergent.

Example 5.9. The sequence $f_n(x) = x^n$ in Example 5.3 converges pointwise on $[0, 1]$ but not uniformly on $[0, 1]$. For $0 \le x < 1$ and $0 < \epsilon < 1$, we have

$$|f_n(x) - f(x)| = |x^n| < \epsilon$$

if and only if $0 \le x < \epsilon^{1/n}$. Since $\epsilon^{1/n} < 1$ for all $n \in \mathbb{N}$, no N works for all x sufficiently close to 1 (although there is no difficulty at $x = 1$). The sequence does, however, converge uniformly on $[0, b]$ for every $0 \le b < 1$; for $0 < \epsilon < 1$, we can take $N = \log \epsilon / \log b$.

Example 5.10. The pointwise convergent sequence in Example 5.4 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 $n \ge \epsilon$ if $x = 1/(2n)$.

Example 5.11. The functions in Example 5.5 converge uniformly to 0 on $\mathbb{R}$, since

$$|f_n(x)| = \frac{|\sin nx|}{n} \le \frac{1}{n},$$

so $|f_n(x) - 0| < \epsilon$ for all $x \in \mathbb{R}$ if $n > 1/\epsilon$.

5.3. Cauchy condition for uniform convergence

The Cauchy condition in Definition 1.9 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 5.12. 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 5.13. 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 Cauchy condition to get for all $x \in A$ and $n > N$ that

$$|f(x) - f_n(x)| = \lim_{m\to\infty} |f_m(x) - f_n(x)| \leq \frac{\epsilon}{2} < \epsilon.$$

□

5.4. 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.

5.4.1. Boundedness. First, we consider the uniform convergence of bounded functions.

Theorem 5.14. 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 $M_n \geq 0$ such that

$$|f_n(x)| \leq M_n \qquad \text{for all } x \in A.$$

It follows that

$$|f(x)| \leq |f(x) - f_n(x)| + |f_n(x)| < 1 + M_n \qquad \text{for all } x \in A,$$

meaning that f is bounded on A (by $1 + M_n$). □

We do not assume here that all the functions in the sequence are bounded by the same constant. (If they were, the pointwise limit would also be bounded by that constant.) In particular, it follows that if a sequence of bounded functions converges pointwise to an unbounded function, then the convergence is not uniform.

Example 5.15. The sequence of functions $f_n : (0, 1) \to \mathbb{R}$ in Example 5.2, 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 their 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 $a \leq x < 1$ that

$$|f_n(x) - f(x)| = \left|\frac{n}{nx + 1} - \frac{1}{x}\right| = \frac{1}{x(nx + 1)} < \frac{1}{nx^2} \leq \frac{1}{na^2}.$$

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

$$|f(x)| \leq \frac{1}{a} \qquad \text{for all } x \in [a, 1)$$

so the uniform limit f is bounded on $[a, 1)$, as Theorem 5.14 requires.

5.4.2. Continuity. One of the most important property 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 5.16. 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 $\epsilon > 0$ is given. Then, for every $n \in \mathbb{N}$,

$$|f(x) - f(c)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(c)| + |f_n(c) - f(c)| .$$

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. □

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 5.16 that if a sequence of continuous functions converges pointwise to a discontinuous function, as in Example 5.3, 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 5.4.

5.4.3. 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 5.5). 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 5.17. 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

$$\frac{t}{1 + t^2} \leq \frac{1}{2} \qquad \text{for all } t \in \mathbb{R},$$

since $(1 - t)^2 \geq 0$, which implies that $2t \leq 1 + t^2$. Using this inequality, we get

$$|f_n(x)| \leq \frac{1}{2\sqrt{n}} \qquad \text{for all } x \in \mathbb{R}.$$

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 5.18. 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:

$$\left|\frac{f(x)-f(c)}{x-c} - g(c)\right| \le \left|\frac{f(x)-f(c)}{x-c} - \frac{f_n(x)-f_n(c)}{x-c}\right| + \left|\frac{f_n(x)-f_n(c)}{x-c} - f_n'(c)\right| + |f_n'(c) - g(c)|$$

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 5.13. 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

$$\left|\frac{f(x)-f(c)}{x-c} - \frac{f_n(x)-f_n(c)}{x-c}\right| \le \frac{\epsilon}{3} \qquad \text{for } n > N_1.$$

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 $f'(c) = g(c)$. □

Like Theorem 5.16, Theorem 5.18 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 5.18 the derivatives f_n' are not assumed to be continuous. If they are continuous, one can use Riemann integration and the fundamental theorem of calculus (FTC) to give a simpler proof of the theorem, as follows. Fix some $x_0 \in (a, b)$. The uniform convergence $f_n' \to g$ implies that

$$\int_{x_0}^{x} f_n' dx \to \int_{x_0}^{x} g\, dx.$$

(This is the main point: although we cannot differentiate a uniformly convergent sequence, we can integrate it.) It then follows from one direction of the FTC that

$$f_n(x) - f_n(x_0) \to \int_{x_0}^{x} g\, dx,$$

and the pointwise convergence $f_n \to f$ implies that

$$f(x) = f(x_0) + \int_{x_0}^{x} g\, dx.$$

The other direction of the FTC then implies that f is differentiable and $f' = g$.

5.5. 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 5.19. Suppose that (f_n) is a sequence of functions $f_n : A \to \mathbb{R}$, and define a sequence (S_n) of partial sums $S_n : A \to \mathbb{R}$ 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 5.20. 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 5.14 implies that the convergence cannot be uniform.

The series does, however, converges uniformly on $[-\rho,\rho]$ for every $0 \le \rho < 1$. To prove this, we estimate for $|x| \le \rho$ that

$$\left| S_n(x) - \frac{1}{1-x} \right| = \frac{|x|^{n+1}}{1-x} \le \frac{\rho^{n+1}}{1-\rho}.$$

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

$$0 \le \frac{\rho^{n+1}}{1-\rho} < \epsilon \qquad \text{for all } n > N.$$

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 5.21. 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 5.13 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. □

This condition says that the sum of any number of consecutive terms in the series gets arbitrarily small sufficiently far down the series.

5.6. The Weierstrass M-test

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 5.22 (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

$$\begin{aligned} \left| \sum_{k=m+1}^{n} f_k(x) \right| &\leq \sum_{k=m+1}^{n} |f_k(x)| \\ &\leq \sum_{k=m+1}^{n} M_k \\ &< \epsilon. \end{aligned}$$

Thus, $\sum f_n$ satisfies the uniform Cauchy condition in Theorem 5.21, so it converges uniformly. □

This proof illustrates the value of the Cauchy condition: we can prove the convergence of the series without having to know what its sum is.

Example 5.23. Returning to Example 5.20, 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 $[-\rho, \rho]$.

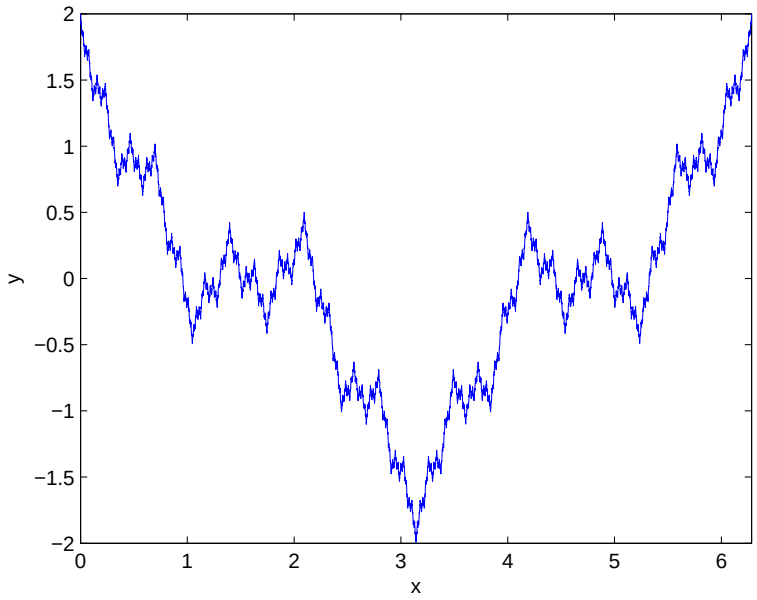

Figure 1. Graph of the Weierstrass continuous, nowhere differentiable function $y = \sum_{n=0}^{\infty} 2^{-n} \cos(3^n x)$ on one period $[0, 2\pi]$.

Example 5.24. The series

$$f(x) = \sum_{n=1}^{\infty} \frac{1}{2^n} \cos\left(3^n x\right)$$

converges uniformly on $\mathbb{R}$ by the M-test since

$$\left| \frac{1}{2^n} \cos\left(3^n x\right) \right| \leq \frac{1}{2^n}, \qquad \sum_{n=1}^{\infty} \frac{1}{2^n} = 1.$$

It then follows from Theorem 5.16 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 does not appear to be smooth at all length-scales. 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.

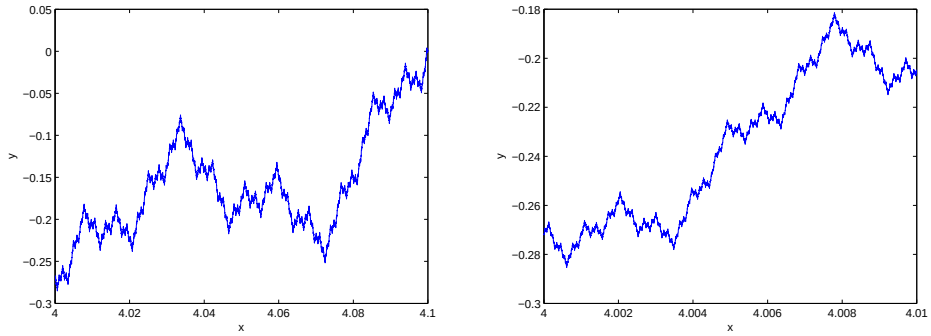

Figure 2. Details of the Weierstrass function showing its self-similar, fractal behavior under rescalings.

If the Weierstrass M-test applies to a series of functions to prove uniform convergence, 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 should not 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 5.25. 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 is not absolute at any $x \in \mathbb{R}$ since the harmonic series

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

diverges to infinity.

5.7. The sup-norm

An equivalent, and often clearer, way to describe uniform convergence is in terms of the uniform, or sup, norm.

Definition 5.26. Suppose that $f : A \to \mathbb{R}$. The uniform, or sup, norm $\|f\|$ of f on A is

$$\|f\| = \sup_{x \in A} |f(x)|.$$

A function is bounded on A if and only if $\|f\| < \infty$.

Example 5.27. Let $A = (0, 1)$ and define $f, g, h : (0, 1) \to \mathbb{R}$ by

$$f(x) = x^2, \qquad g(x) = x^2 - x, \qquad h(x) = \frac{1}{x}.$$

Then

$$\|f\| = 1, \quad \|g\| = \frac{1}{4}, \qquad \|h\| = \infty.$$

We have the following characterization of uniform convergence.

Definition 5.28. A sequence (f_n) of functions $f_n : A \to \mathbb{R}$ converges uniformly on A to a function $f : A \to \mathbb{R}$ if

$$\lim_{n \to \infty} \|f_n - f\| = 0.$$

Similarly, we can define a uniformly Cauchy sequence in terms of the sup-norm.

Definition 5.29. 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

$$m, n > N \text{ implies that } \|f_m - f_n\| < \epsilon.$$

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 $|\cdot|$ value replaced by the sup-norm $\|\cdot\|$.

5.8. Spaces of continuous functions

Our previous theorems about continuous functions on compact sets can be restated in a more geometrical way using the sup-norm.

Definition 5.30. Let $K \subset \mathbb{R}$ be a compact set. The space $C(K)$ consists of all continuous functions $f : K \to \mathbb{R}$.

Thus, we think of a function f as a point in a function space $C(K)$, just as we think of a real number x as a point in $\mathbb{R}$.

Theorem 5.31. The space $C(K)$ is a vector space with respect to the usual pointwise definitions of scalar multiplication and addition of functions: If $f, g \in C(K)$ and $k \in \mathbb{R}$, then

$$(kf)(x) = kf(x), \qquad (f + g)(x) = f(x) + g(x).$$

This follows from Theorem 3.15, which states that scalar multiples and sums of continuous functions are continuous and therefore belong to $C(K)$. The algebraic vector-space properties of $C(K)$ follow immediately from those of the real numbers.

Definition 5.32. A normed vector space $(X, \|\cdot\|)$ is a vector space X (which we assume to be real) together with a function $\|\cdot\| : X \to \mathbb{R}$, called a norm on X, such that for all $f, g \in X$ and $k \in \mathbb{R}$:

(1) $0 \leq \|f\| < \infty$ and $\|f\| = 0$ if and only if $f = 0$;
(2) $\|kf\| = |k|\|f\|$;
(3) $\|f + g\| \leq \|f\| + \|g\|$.

We think of $\|f\|$ as defining a "length" of the vector $f \in X$ and $\|f - g\|$ as the corresponding "distance" between $f, g \in X$. (There are typically many ways to define a norm on a vector space satisfying Definition 5.32, each leading to a different notion of the distance between vectors.)

The properties in Definition 5.32 are natural one to require of a length: The length of f is 0 if and only if f is the 0-vector; multiplying a vector by k multiplies its length by $|k|$; and the length of the "hypoteneuse" $f + g$ is less than or equal to the sum of the lengths of the "sides" f, g. Because of this last interpretation, property (3) is referred to as the triangle inequality.

It is straightforward to verify that the sup-norm on $C(K)$ has these properties.

Theorem 5.33. The space $C(K)$ with the sup-norm $\|\cdot\| : C(K) \to \mathbb{R}$ given in Definition 5.26 is a normed vector space.

Proof. From Theorem 3.33, the sup-norm of a continuous function $f : K \to \mathbb{R}$ on a compact set K is finite, and it is clearly nonnegative, so $0 \leq \|f\| < \infty$. If $\|f\| = 0$, then $\sup_{x \in K} |f(x)| = 0$, which implies that $f(x) = 0$ for every $x \in K$, meaning that $f = 0$ is the zero function.

We also have

$$\|kf\| = \sup_{x \in K} |k(f(x)| = |k| \sup_{x \in K} |f(x)| = k\|f\|,$$

and

$$\begin{aligned} \|f + g\| &= \sup_{x \in K} |(f(x) + g(x)| \\ &\leq \sup_{x \in K} \{|f(x)| + |g(x)|\} \\ &\leq \sup_{x \in K} |f(x)| + \sup_{x \in K} |g(x)| \\ &\leq \|f\| + \|g\|, \end{aligned}$$

which verifies the properties of a norm. □

Definition 5.34. A sequence (f_n) in a normed vector space $(X, \|\cdot\|)$ converges to $f \in X$ if $\|f_n - f\| \to 0$ as $n \to \infty$. That is, if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$n > N \text{ implies that } \|f_n - f\| < \epsilon.$$

The sequence is a Cauchy sequence for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$m, n > N \text{ implies that } \|f_m - f_n\| < \epsilon.$$

Definition 5.35. A normed vector space is complete if every Cauchy sequence converges. A complete normed linear space is called a Banach space.

Theorem 5.36. The space $C(K)$ with the sup-norm is a Banach space.

Proof. The space $C(K)$ with the sup-norm is a normed space from Theorem 5.33. Theorem 5.13 implies that it is complete. □

Chapter 6

Power Series

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 (and many other less familiar functions).

6.1. 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 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 $|x| > R$, and the convergence is uniform on every interval $|x| < \rho$ where $0 \leq \rho < R$. If $R > 0$, 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, but we restrict our attention here to real-valued power series.

Definition 6.1. 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

$$\sum_{n=0}^{\infty} a_n (x-c)^n.$$

Here are some power series centered at 0:

$$\sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + x^4 + \dots,$$
$$\sum_{n=0}^{\infty} \frac{1}{n!} x^n = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{24}x^4 + \dots,$$
$$\sum_{n=0}^{\infty} (n!) x^n = 1 + x + 2x^2 + 6x^3 + 24x^4 + \dots,$$
$$\sum_{n=0}^{\infty} (-1)^n x^{2^n} = x - x^2 + x^4 - x^8 + \dots;$$

and here is a power series centered at 1:

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} (x-1)^n = (x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3 - \frac{1}{4}(x-1)^4 + \dots.$$

The power series in Definition 6.1 is a formal expression, since we have not said anything about its convergence. By changing variables $x \mapsto (x-c)$, we can assume without loss of generality that a power series is centered at 0, and we will do so when it's convenient.

6.2. Radius of convergence

First, we prove that every power series has a radius of convergence.

Theorem 6.2. Let

$$\sum_{n=0}^{\infty} a_n (x-c)^n$$

be a power series. There is an $0 \le R \le \infty$ 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. Assume without loss of generality that $c = 0$ (otherwise, replace x by $x-c$). 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 $M \ge 0$ such that

$$|a_n x_0^n| \le M \qquad \text{for } n = 0, 1, 2, \dots.$$

If $|x| < |x_0|$, then

$$|a_n x^n| = |a_n x_0^n| \left| \frac{x}{x_0} \right|^n \le M r^n, \qquad r = \left| \frac{x}{x_0} \right| < 1.$$

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 $|x| < R$, because 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 5.22) implies that the series converges uniformly on $|x| \leq \rho$, and then it follows from Theorem 5.16 that the sum is continuous on $|x| \leq \rho$. Since this holds for every $0 \leq \rho < R$, the sum is continuous in $|x| < R$. □

The following definition therefore makes sense for every power series.

Definition 6.3. 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 6.2 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 will not discuss any general theorems about the convergence of power series at the endpoints (e.g. the Abel theorem).

Theorem 6.2 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 6.4. 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 $0 \le r < 1$, or $|x-c| < R$, and diverges if $1 < r \le \infty$, 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 6.5 (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 lim sup diverges to ∞, and $R = \infty$ if the lim sup 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 $0 \le r < 1$, or $|x-c| < R$, and diverges if $1 < r \le \infty$, or $|x-c| > R$, which proves the result. □

This theorem provides an alternate proof of Theorem 6.2 from the root test; in fact, our proof of Theorem 6.2 is more-or-less a proof of the root test.

6.3. Examples of power series

We consider a number of examples of power series and their radii of convergence.

Example 6.6. The geometric series

$$\sum_{n=0}^{\infty} x^n = 1 + x + x^2 + \dots$$

has radius of convergence

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

so it converges for $|x| < 1$, to $1/(1-x)$, 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 5.20. Note that although the function $1/(1-x)$ is well-defined for all $x \ne 1$, the power series only converges to it when $|x| < 1$.

Example 6.7. 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 6.8. 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 $x \in \mathbb{R}$. Its sum provides a definition of the exponential function $\exp : \mathbb{R} \to \mathbb{R}$. (See Section 6.5.)

Example 6.9. 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 provides a definition of the cosine function $\cos : \mathbb{R} \to \mathbb{R}$.

Example 6.10. The 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 provides a definition of the sine function $\sin : \mathbb{R} \to \mathbb{R}$.

Example 6.11. The power series

$$\sum_{n=0}^{\infty} (n!) x^n = 1 + x + (2!)x + (3!)x^3 + (4!)x^4 + \dots$$

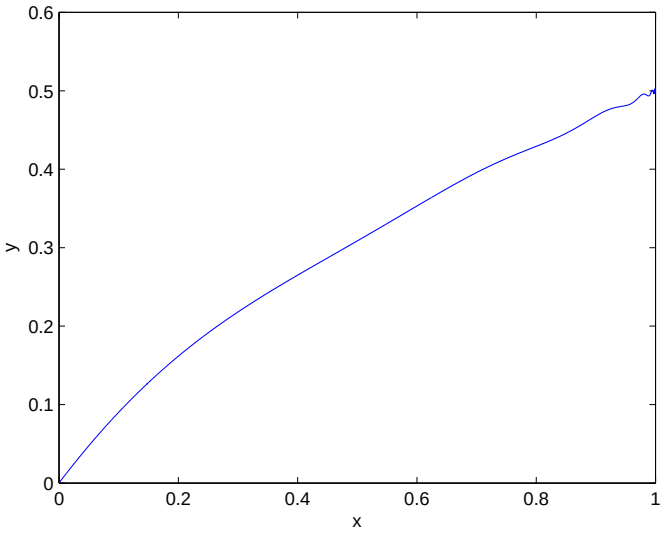

Figure 1. Graph of the lacunary power series $y = \sum_{n=0}^{\infty}(-1)^n x^{2^n}$ on $[0, 1)$. It appears relatively well-behaved; however, the small oscillations visible near $x = 1$ are not a numerical artifact.

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 6.12. 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 $(0, 2]$.

Example 6.13. 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

$$a_n = \begin{cases} 1 & \text{if } n = 2^k, \\ 0 & \text{if } n \neq 2^k, \end{cases}$$

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

$$|a_n x^n| = \begin{cases} |x|^n & \text{if } n = 2^k, \\ 0 \leq |x|^n & \text{if } n \neq 2^k. \end{cases}$$

If $|x| > 1$, 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 root test (Theorem 6.5) gives $R = 1$. The series does not converge at either endpoint $x = \pm 1$, so its interval of convergence is $(-1, 1)$.

There are successively longer gaps (or "lacuna") between the powers with non-zero coefficients. Such series are called lacunary power series, and 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

6.4. Differentiation of power series

We saw in Section 5.4.3 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

$$f'(x) = a_1 + 2a_2 x + 3a_3 x^2 + 4a_4 x^3 + \dots$$

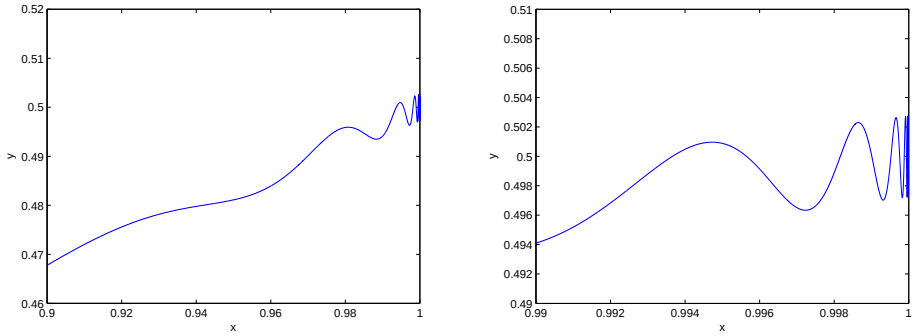

Figure 2. Details of the lacunary power series $\sum_{n=0}^{\infty}(-1)^n x^{2^n}$ near $x = 1$, showing its oscillatory behavior and the nonexistence of a limit as $x \to 1^-$.

is given by term-by-term differentiation. To prove this, 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.

Theorem 6.14. 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 $(n r^{n-1})$ is bounded, by M say. It follows that

$$\left|n a_n x^{n-1}\right| \le \frac{M}{\rho}|a_n \rho^n| \qquad \text{for all } n \in \mathbb{N}.$$

The series $\sum |a_n \rho^n|$ converges, since $\rho < R$, so the comparison test implies that $\sum n a_n x^{n-1}$ converges absolutely.

Conversely, suppose $|x| > R$. Then $\sum |a_n x^n|$ diverges (since $\sum a_n x^n$ diverges) and

$$\left|n a_n x^{n-1}\right| \geq \frac{1}{|x|} |a_n x^n|$$

for $n \geq 1$, so the comparison test implies that $\sum n a_n x^{n-1}$ diverges. Thus the series have the same radius of convergence. □

Theorem 6.15. 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

$$f'(x) = \sum_{n=1}^{\infty} n a_n (x-c)^{n-1} \qquad \text{for } |x-c| < R.$$

Proof. The term-by-term differentiated power series converges in $|x-c| < R$ by Theorem 6.14. We denote its sum by

$$g(x) = \sum_{n=1}^{\infty} n a_n (x-c)^{n-1}.$$

Let $0 < \rho < R$. Then, by Theorem 6.2, the power series for f and g both converge uniformly in $|x-c| < \rho$. Applying Theorem 5.18 to their partial sums, we conclude that f is differentiable in $|x-c| < \rho$ and $f' = g$. Since this holds for every $0 \leq \rho < R$, it follows that f is differentiable in $|x-c| < R$ and $f' = g$, which proves the result. □

Repeated application Theorem 6.15 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 6.16. 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 6.16 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

$$\begin{aligned}
f'(x) &= a_1 + 2a_2x + 3a_3x^2 + \cdots + na_nx^{n-1} + \ldots,\\
f''(x) &= 2a_2 + (3\cdot 2)a_3x + \cdots + n(n-1)a_nx^{n-2} + \ldots,\\
f'''(x) &= (3\cdot 2\cdot 1)a_3 + \cdots + n(n-1)(n-2)a_nx^{n-3} + \ldots,\\
&\vdots\\
f^{(k)}(x) &= (k!)a_k + \cdots + \frac{n!}{(n-k)!}x^{n-k} + \ldots,
\end{aligned}$$

where all of these power series have radius of convergence R. Setting $x = 0$ in these series, we get

$$a_0 = f(0), \quad a_1 = f'(0), \quad \ldots \quad a_k = \frac{f^{(k)}(0)}{k!},$$

which proves the result (after replacing 0 by c). □

One consequence of this result is that convergent power series with different coefficients cannot converge to the same sum.

Corollary 6.17. 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 on 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 □

6.5. The exponential function

We showed in Example 6.8 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 6.15, 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, and they 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.

First, we show that $e^x e^y = e^{x+y}$. We continue to write the function as $E(x)$ to emphasise that we use nothing beyond its power series definition.

Proposition 6.18. 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, which is justified by the absolute converge of the power series, we get

$$\begin{aligned} E(x)E(y) &= \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \frac{x^j y^k}{j!\,k!} \\ &= \sum_{n=0}^{\infty}\sum_{k=0}^{n} \frac{x^{n-k} y^k}{(n-k)!\,k!}. \end{aligned}$$

From the binomial theorem,

$$\sum_{k=0}^{n} \frac{x^{n-k} y^k}{(n-k)!\,k!} = \frac{1}{n!}\sum_{k=0}^{n} \frac{n!}{(n-k)!\,k!} x^{n-k} y^k = \frac{1}{n!}\,(x+y)^n\,.$$

Hence,

$$E(x)E(y) = \sum_{n=0}^{\infty} \frac{(x+y)^n}{n!} = E(x+y),$$

which proves the result. □

In particular, it follows that

$$E(-x) = \frac{1}{E(x)}.$$

Note that $E(x) > 0$ for all $x > 0$ since all the terms in its power series are positive, so $E(x) > 0$ for every $x \in \mathbb{R}$.

The following proposition, which we use below in Section 6.6.2, states that e^x grows faster than any power of x as $x \to \infty$.

Proposition 6.19. 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 $k \in \mathbb{N}$

$$E(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} > \frac{x^k}{k!} \qquad \text{for all } x > 0.$$

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. □

Finally, we prove that the exponential is characterized by the properties $E' = E$ and $E(0) = 1$. This is a uniqueness result for an initial value problem for a simple linear ordinary differential equation.

Proposition 6.20. Suppose $f : \mathbb{R} \to \mathbb{R}$ is a differentiable function such that

$$f' = f, \qquad f(0) = 1.$$

Then $f = E$.

Proof. Suppose that $f' = f$. Then 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 4.29 that f/E is constant on $\mathbb{R}$. Since $f(0) = E(0) = 1$, we have $f/E = 1$, which implies that $f = E$. □

The logarithm can be defined as the inverse of the exponential. 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 do this 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.

6.6. Taylor's theorem and power series

Theorem 6.16 looks similar to Taylor's theorem, Theorem 4.41. There is, however, a fundamental difference. Taylor's theorem gives an expression for the error between a function and its Taylor polynomial of degree n. No questions of convergence are involved here. On the other hand, Theorem 6.16 asserts the convergence of an infinite power series to a function f, and gives an expression for the coefficients of the power series in terms of f. The coefficients of the Taylor polynomials and the power series are the same in both cases, but the Theorems are different.

Roughly speaking, Taylor's theorem describes the behavior of the Taylor polynomials $P_n(x)$ of f as $x \to c$ with n fixed, while the power series theorem describes the behavior of $P_n(x)$ as $n \to \infty$ with x fixed.

6.6.1. Smooth functions and analytic functions. 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 6.21. 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 6.22. A function $f : (a, b) \to \mathbb{R}$ is analytic on (a, b) if for every $c \in (a, b)$ 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, we abbreviate "real analytic" to "analytic."

Theorem 6.16 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 6.16 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, in which case it diverges for every $x \neq c$, or the power series may converge, but not to f.

6.6.2. 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 6.19 that if

$$p(x) = \sum_{k=0}^{n} a_k x^k$$

is any polynomial function, then

$$\lim_{x \to \infty} \frac{p(x)}{e^x} = \sum_{k=0}^{n} a_k \lim_{x \to \infty} \frac{x^k}{e^x} = 0.$$

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.)

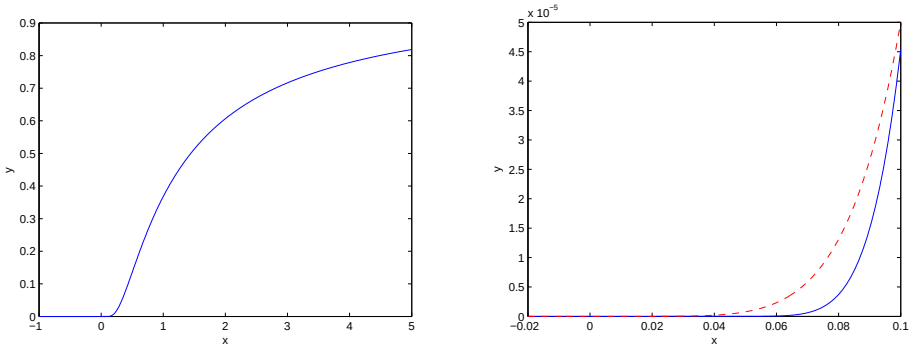

Figure 3. Left: Plot $y = \phi(x)$ of the smooth, non-analytic function defined in Proposition 6.23. Right: A detail of the function near $x = 0$. The dotted line is the power-function $y = x^6/50$. The graph of ϕ near 0 is "flatter' than the graph of the power-function, illustrating that $\phi(x)$ goes to zero faster than any power of x as $x \to 0$.

Proposition 6.23. 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 in $1/x$. (This follows, for example, by induction.) 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 clearly 0 since $\phi(0) = 0$ and $\phi(h) = 0$ for all $h < 0$. For the right derivative, writing $1/h = x$ and using Proposition 6.19, we get

$$\begin{aligned} \phi'(0^+) &= \lim_{h \to 0^+} \left[\frac{\phi(h) - \phi(0)}{h} \right] \\ &= \lim_{h \to 0^+} \frac{\exp(-1/h)}{h} \\ &= \lim_{x \to \infty} \frac{x}{e^x} \\ &= 0. \end{aligned}$$

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 6.19, we get that

$$\begin{aligned}\phi^{(n+1)}(0^+) &= \lim_{h\to 0^+} \left[\frac{\phi^{(n)}(h) - \phi^{(n)}(0)}{h}\right] \\ &= \lim_{h\to 0^+} \frac{p_n(1/h)\exp(-1/h)}{h} \\ &= \lim_{x\to\infty} \frac{xp_n(x)}{e^x} \\ &= 0,\end{aligned}$$

which proves the result. □

Corollary 6.24. The function $\phi : \mathbb{R} \to \mathbb{R}$ defined by

$$\phi(x) = \begin{cases} \exp(-1/x) & \text{if } x > 0, \\ 0 & \text{if } x \le 0, \end{cases}$$

is smooth but not analytic on $\mathbb{R}$.

Proof. From Proposition 6.23, 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 states that for $x > 0$ there exists $0 < \xi < x$ such that

$$\phi(x) = \frac{p_{n+1}(1/\xi)}{(n+1)!} e^{-1/\xi} x^{n+1}.$$

Since the derivatives of ϕ are bounded, this shows that for every $n \in \mathbb{N}$ there exists a constant C_{n+1} such that

$$0 \le \phi(x) \le C_{n+1} x^{n+1} \qquad \text{for all } 0 \le x < \infty,$$

but this does not imply that $\phi(x) = 0$.

We can construct other smooth, non-analytic functions from ϕ.

Example 6.25. The function

$$\psi(x) = \begin{cases} \exp(-1/x^2) & \text{if } x \ne 0, \\ 0 & \text{if } x = 0, \end{cases}$$

is infinitely differentiable on $\mathbb{R}$, since $\psi(x) = \phi(x^2)$ is a composition of smooth functions.

The following example is useful in many parts of analysis.

Definition 6.26. A function $f : \mathbb{R} \to \mathbb{R}$ has compact support if there exists $R \ge 0$ such that $f(x) = 0$ for all $x \in \mathbb{R}$ with $|x| \ge R$.

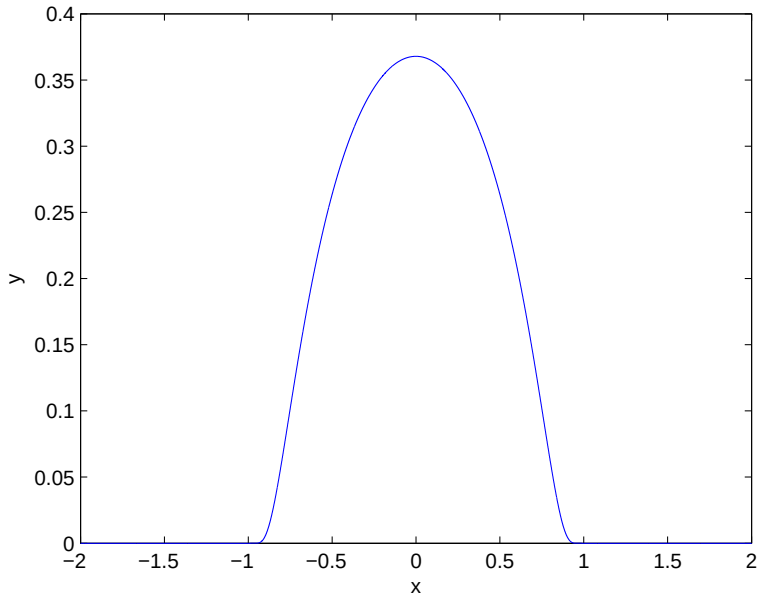

Figure 4. Plot of the smooth, compactly supported "bump" function defined in Example 6.27.

It isn't hard to construct continuous functions with compact support; one example that vanishes for $|x| \geq 1$ is

$$f(x) = \begin{cases} 1 - |x| & \text{if } |x| < 1, \\ 0 & \text{if } |x| \geq 1. \end{cases}$$

By matching left and right derivatives of a piecewise-polynomial function, 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 6.27. 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.

The function ϕ defined in Proposition 6.23 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 other points. By contrast, an analytic function on an interval has the remarkable property that the value of the function and all of its derivatives at one point of the interval determine its values at all other points

of the interval, since we can extend the function from point to point by summing its power series. (This claim requires a proof, which we omit.)

For example, it is impossible to construct an analytic function with compact support, since if an analytic function on $\mathbb{R}$ vanishes in any interval $(a, b) \subset \mathbb{R}$, then it must be identically zero on $\mathbb{R}$. Thus, the non-analyticity of the "bump"-function η in Example 6.27 is essential.

6.7. Appendix: Review of series

We summarize the results and convergence tests that we use to study power series. Power series are closely related to geometric series, so most of the tests involve comparisons with a geometric series.

Definition 6.28. 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. The series converges absolutely if

$$\sum_{n=1}^{\infty} |a_n|$$

converges.

The following Cauchy condition for series is an immediate consequence of the Cauchy condition for the sequence of partial sums.

Theorem 6.29 (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. □

Since

$$\left|\sum_{k=m+1}^{n} a_k\right| \leq \sum_{k=m+1}^{n} |a_k|$$

the series $\sum a_n$ is Cauchy if $\sum |a_n|$ is Cauchy, so an absolutely convergent series converges. We have the following necessary, but not sufficient, condition for convergence of a series.

Theorem 6.30. 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 6.29, 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 $n \to \infty$. □

Next, we derive the comparison, ratio, and root tests, which provide explicit sufficient conditions for the convergence of a series.

Theorem 6.31 (Comparison test)**.** Suppose that $|b_n| \leq a_n$ and $\sum a_n$ converges. Then $\sum b_n$ converges absolutely.

Proof. Since $\sum a_n$ converges it satisfies the Cauchy condition, and since

$$\sum_{k=m+1}^{n} |b_k| \leq \sum_{k=m+1}^{n} a_k$$

the series $\sum |b_n|$ also satisfies the Cauchy condition. Therefore $\sum b_n$ converges absolutely. □

Theorem 6.32 (Ratio test)**.** Suppose that (a_n) is a sequence of real numbers such that a_n is nonzero for all sufficiently large $n \in \mathbb{N}$ and 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 $0 \leq r < 1$ and diverges if $1 < r \leq \infty$.

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

$$|a_n| \leq Ms^n \qquad \text{for all } n > N$$

where M is a suitable constant. Therefore $\sum a_n$ converges absolutely by comparison with the convergent geometric series $\sum Ms^n$.

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 $|a_n| \geq Ms^n$ 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. □

Before stating the root test, we define the lim sup.

Definition 6.33. If (a_n) is a sequence of real numbers, then

$$\limsup_{n\to\infty} a_n = \lim_{n\to\infty} b_n, \qquad b_n = \sup_{k\geq n} a_k.$$

If (a_n) is a bounded sequence, then $\limsup a_n \in \mathbb{R}$ always exists since (b_n) is a monotone decreasing sequence of real numbers that is bounded from below. If (a_n) isn't bounded from above, then $b_n = \infty$ for every $n \in \mathbb{N}$ (meaning that $\{a_k : k \geq n\}$ isn't bounded from above) and we write $\limsup a_n = \infty$. If (a_n) is bounded from above but (b_n) diverges to $-\infty$, then (a_n) diverges to $-\infty$ and we write $\limsup a_n = -\infty$. With these conventions, every sequence of real numbers has a lim sup, even if it doesn't have a limit or diverge to $\pm\infty$.

We have the following equivalent characterization of the lim sup, which is what we often use in practice. If the lim sup is finite, it states that every number bigger than the lim sup eventually bounds all the terms in a tail of the sequence from above, while infinitely many terms in the sequence are greater than every number less than the lim sup.

Proposition 6.34. Let (a_n) be a real sequence with

$$L = \limsup_{n\to\infty} a_n.$$

(1) If $L \in \mathbb{R}$ is finite, then for every $M > L$ there exists $N \in \mathbb{N}$ such that $a_n < M$ for all $n > N$, and for every $m < L$ there exist infinitely many $n \in \mathbb{N}$ such that $a_n > m$.

(2) If $L = -\infty$, then for every $M \in \mathbb{R}$ there exists $N \in \mathbb{N}$ such that $a_n < M$ for all $n > N$.

(3) If $L = \infty$, then for every $m \in \mathbb{R}$, there exist infinitely many $n \in \mathbb{N}$ such that $a_n > m$.

Theorem 6.35 (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 $t > \limsup |a_n|^{1/n}$, Proposition 6.34 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 $1 < r \leq \infty$. 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 $t < \limsup |a_n|^{1/n}$, Proposition 6.34 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. □

Chapter 7

Metric Spaces

A metric space is a set X that has a notion of the distance $d(x, y)$ between every pair of points $x, y \in X$. The purpose of this chapter is to introduce metric spaces and give some definitions and examples. We do not develop their theory in detail, and we leave the verifications and proofs as an exercise. In most cases, the proofs are essentially the same as the ones for real functions or they simply involve chasing definitions.

7.1. Metrics

A metric on a set is a function that satisfies the minimal properties we might expect of a distance.

Definition 7.1. A metric d on a set X is a function $d : X \times X \to \mathbb{R}$ such that for all $x, y \in X$:

(1) $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$;

(2) $d(x, y) = d(y, x)$ (symmetry);

(3) $d(x, y) \leq d(x, z) + d(z, x)$ (triangle inequality).

A metric space (X, d) is a set X with a metric d defined on X.

We can define many different metrics on the same set, but if the metric on X is clear from the context, we refer to X as a metric space and omit explicit mention of the metric d.

Example 7.2. A rather trivial example of a metric on any set X is the discrete metric

$$d(x, y) = \begin{cases} 0 & \text{if } x = y, \\ 1 & \text{if } x \neq y. \end{cases}$$

Example 7.3. Define $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ by

$$d(x, y) = |x - y|.$$

Then d is a metric on $\mathbb{R}$. Nearly all the concepts we discuss for metric spaces are natural generalizations of the corresponding concepts for $\mathbb{R}$ with this absolute-value metric.

Example 7.4. 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 requires an analytical proof (see Section 7.6).

Example 7.5. The Euclidean metric $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ on $\mathbb{R}^n$ is defined by

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots (x_n - y_n)^2}$$

where

$$x = (x_1, x_2, \ldots, x_n), \qquad y = (y_1, y_2, \ldots, y_n).$$

For $n = 1$ this metric reduces to the absolute-value metric on $\mathbb{R}$, and for $n = 2$ it is the previous example. We will mostly consider the case $n = 2$ for simplicity. The triangle inequality for this metric follows from the Minkowski inequality, which is proved in Section 7.6.

Example 7.6. 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. It is also referred to informally as the "taxicab" metric, since it's the distance one would travel by taxi on a rectangular grid of streets.

Example 7.7. 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 7.8. 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 Britain, d is sometimes called the "British Rail" metric, because all the train lines radiate from London (located at the origin). 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 7.9. Let $C(K)$ denote the set of continuous functions $f : K \to \mathbb{R}$, where $K \subset \mathbb{R}$ is compact; for example, we could take $K = [a, b]$ to be a closed, bounded interval. For $f, g \in C(K)$ define

$$d(f, g) = \sup_{x \in K} |f(x) - g(x)| .$$

The function $d : C(K) \times C(K) \to \mathbb{R}$ is well-defined, since a continuous function on a compact set is bounded; in fact, such a function attains it maximum value, so we could also write

$$d(f, g) = \max_{x \in K} |f(x) - g(x)| .$$

Then d is a metric on $C(K)$. Two functions are close with respect to this metric if their values are close at every point of K.

Subspaces of a metric space (X, d) are subsets $A \subset X$ with the metric d_A obtained by restricting the metric d on X to A.

Definition 7.10. Let (X, d) be a metric space. A 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 properties of subsets $A \subset X$ of a metric space (X, d) in terms of properties of the corresponding metric subspace (A, d_A).

7.2. Norms

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

Definition 7.11. A normed vector space $(X, \|\cdot\|)$ is a vector space X (which we assume to be real) 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 7.11 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 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 referred to as the triangle inequality.

Proposition 7.12. If $(X, \|\cdot\|)$ is a normed vector space X, 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 immediately from properties (1) and (3) of a norm in Definition 7.11. □

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 do not even make sense in a general metric space since we cannot add points or multiply them by scalars. If X is a normed vector space, we always use the metric associated with its norm, unless stated specifically otherwise.

Example 7.13. The set of real numbers $\mathbb{R}$ with the absolute-value norm $|\cdot|$ is a one-dimensional normed vector space.

Example 7.14. The set $\mathbb{R}^2$ with any of the norms defined for $x = (x_1, x_2)$ by

$$\|x\|_1 = |x_1| + |x_2|, \qquad \|x\|_2 = \sqrt{x_1^2 + x_2^2}, \qquad \|x\|_\infty = \max(|x_1|, |x_2|)$$

is a two-dimensional normed vector space. The corresponding metrics are the "taxicab" metric, the Euclidean metric, and the maximum metric, respectively.

These norms are special cases of the following example.

Example 7.15. The set $\mathbb{R}^n$ with the ℓ^p-norm defined for $x = (x_1, x_2, \ldots, x_n)$ and $1 \le p < \infty$ by

$$\|x\|_p = (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{1/p}$$

and for $p = \infty$ by

$$\|x\|_\infty = \max(|x_1|, |x_2|^p, \ldots, |x_n|^p)$$

is an n-dimensional normed vector space for every $1 \le p \le \infty$. The Euclidean case $p = 2$ is distinguished by the fact that the norm $\|\cdot\|_2$ is derived from an inner product on $\mathbb{R}^n$:

$$\|x\|_2 = \sqrt{\langle x, x \rangle}, \qquad \langle x, y \rangle = \sum_{i=1}^{n} x_i y_i.$$

The triangle inequality for the ℓ^p-norm is called Minkowski's inequality. It is straightforward to verify if $p = 1$ or $p = \infty$, but it is not obvious if $1 < p < \infty$. We give a proof of the simplest case $p = 2$ in Section 7.6.

Example 7.16. Let $K \subset \mathbb{R}$ be compact. Then the space $C(K)$ of continuous functions $f : K \to \mathbb{R}$ with the sup-norm $\|\cdot\| : C(K) \to \mathbb{R}$, defined by

$$\|f\| = \sup_{x \in K} |f(x)|,$$

is a normed vector space. The corresponding metric is the one described in Example 7.9.

Example 7.17. The discrete metric in Example 7.2 and the metric in Example 7.8 are not derived from a norm.

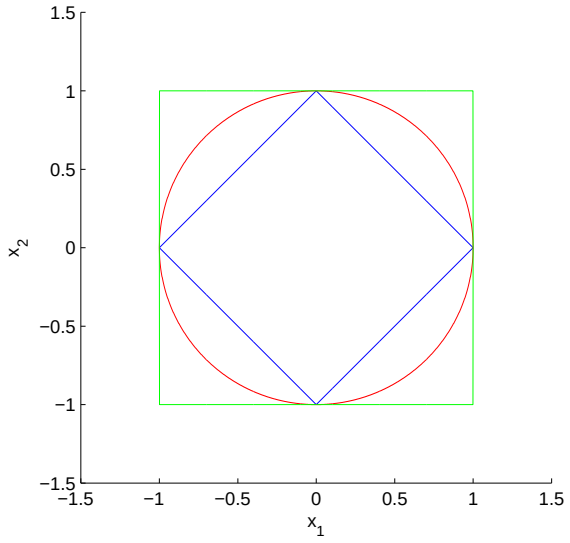

Figure 1. Boundaries of the unit balls $B_1(0)$ in $\mathbb{R}^2$ for the ℓ^1-norm (diamond), the ℓ^2-norm (circle), and the ℓ^∞-norm (square).

7.3. Sets

We first define an open ball in a metric space, which is analogous to a bounded open interval in $\mathbb{R}$.

Definition 7.18. Let (X, d) be a metric space. The open ball of radius $r > 0$ and center $x \in X$ is the set

$$B_r(x) = \{y \in X : d(x, y) < r\}.$$

Example 7.19. Consider $\mathbb{R}$ with its standard absolute-value metric, defined in Example 7.3. Then the open ball

$$B_r(x) = \{y \in \mathbb{R} : |x - y| < r\}$$

is the open interval of radius r centered at x.

Next, we describe the unit balls in $\mathbb{R}^2$ with respect to some different metrics.

Example 7.20. Consider $\mathbb{R}^2$ with the Euclidean metric defined in Example 7.4. Then $B_r(x)$ is a disc of diameter $2r$ centered at x. For the ℓ^1-metric in Example 7.6, the ball $B_r(x)$ is a diamond of diameter $2r$, and for the ℓ^∞-metric in Example 7.7, it is a square of side $2r$ (see Figure 1).

The norms $\|\cdot\|_1$, $\|\cdot\|_2$, $\|\cdot\|_\infty$ on $\mathbb{R}^n$ satisfy

$$\|x\|_\infty \le \|x\|_2 \le \|x\|_1 \le n\|x\|_\infty.$$

These inequalities correspond to the nesting of one ball inside another in Figure 1. Furthermore, the ℓ^∞-ball of radius 1 is included in the ℓ^1-ball of radius 2. As a result, every open ball with respect to one norm contains an open ball with respect

to the other norms, and we say that the norms are equivalent. It follows from the definitions below that, despite the different geometries of their unit balls, these norms define the same collection of open sets and neighborhoods (i.e. the same topologies) and the same convergent sequences, limits, and continuous functions.

Example 7.21. Consider the space $C(K)$ of continuous functions $f : K \to \mathbb{R}$ with the sup-metric defined in Example 7.9. The ball $B_r(f)$ consists of all continuous functions $g : K \to \mathbb{R}$ whose values are strictly within r of the values of f at every point $x \in K$.

One has to be a little careful with the notion of open balls in a general metric space because they do not always behave the way their name suggests.

Example 7.22. Let X be a set with the discrete metric given in Example 7.2. 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 $r \ge 1$.

An another example, what are the open balls for the metric in Example 7.8?

We define open sets in a metric space analogously to open sets in $\mathbb{R}$.

Definition 7.23. 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 $B_r(x) \subset G$.

We can give a more geometrical definition of an open set in terms of neighborhoods.

Definition 7.24. 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 $r > 0$.

Thus, a set is open if and only if every point in the set has a neighborhood that is contained in the set. In particular, an open set is itself a neighborhood of every point in the set.

The following is the topological definition of a closed set.

Definition 7.25. Let X be a metric space. A set $F \subset X$ is closed if

$$F^c = \{x \in X : x \notin F\}$$

is open.

Bounded sets in a metric space are defined in the obvious way.

Definition 7.26. 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

$$d(x, y) \le R \qquad \text{for all } y \in A.$$

Equivalently, this definition says that $A \subset B_R(x)$. The center point $x \in X$ is not important here. The triangle inequality implies that

$$B_R(x) \subset B_S(y), \qquad S = R + d(x, y),$$

so if the definition holds for some $x \in X$, then it holds for every $x \in X$. Alternatively, we define the diameter $0 \le \operatorname{diam} A \le \infty$ of a set $A \subset X$ by

$$\operatorname{diam} A = \sup \{d(x, y) : x, y \in A\}.$$

Then A is bounded if and only if $\operatorname{diam} A < \infty$.

Example 7.27. Let X be a set with the discrete metric given in Example 7.2. Then X is bounded since $X = B_1(x)$ for any $x \in X$.

Example 7.28. Let $C(K)$ be the space of continuous functions $f : K \to \mathbb{R}$ on a compact set $K \subset \mathbb{R}$ equipped with the sup-norm. The set $F \subset C(K)$ of all functions f such that $|f(x)| \leq 1$ for every $x \in K$ is a bounded set since $\|f\| = d(f, 0) \leq 1$ for all $f \in F$.

Compact sets are sets that have the Heine-Borel property

Definition 7.29. A subset $K \subset X$ of a metric space X is compact if every open cover of K has a finite subcover.

A significant property of $\mathbb{R}$ (or $\mathbb{R}^n$) that does *not* generalize to arbitrary metric spaces is that a set is compact if and only if it is closed and bounded. In general, a compact subset of a metric space is closed and bounded; however, a closed and bounded set need not be compact.

Finally, we define some relationships of points to a set that are analogous to the ones for $\mathbb{R}$.

Definition 7.30. 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 points in A and points 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 in the set is an interior point, and a set is closed if and only if every accumulation point of the set belongs to the set.

7.4. Sequences

A sequence (x_n) in a set X is a function $f : \mathbb{N} \to X$, where we write $x_n = f(n)$ for the nth term in the sequence.

Definition 7.31. 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, $x_n \to x$ if $d(x_n, x) \to 0$ as $n \to \infty$. 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 7.32. For $\mathbb{R}$ with its standard absolute value metric, Definition 7.31 is just the definition of the convergence of a real sequence.

Example 7.33. Let $K \subset \mathbb{R}$ be compact. A sequence of continuous functions (f_n) in $C(K)$ converges to $f \in C(K)$ with respect to the sup-norm if and only if $f_n \to f$ as $n \to \infty$ uniformly on K.

We define closed sets in terms of sequences in the same way as for $\mathbb{R}$.

Definition 7.34. A subset $F \subset X$ of a metric space X is sequentially closed if the limit every convergent sequence (x_n) in F belongs to F.

Explicitly, this means that if (x_n) is a sequence of points $x_n \in F$ and $x_n \to x$ as $n \to \infty$ in X, then $x \in F$. A subset of a metric space is sequentially closed if and only if it is closed.

Example 7.35. Let $F \subset C(K)$ be the set of continuous functions $f : K \to \mathbb{R}$ such that $|f(x)| \leq 1$ for all $x \in K$. Then F is a closed subset of $C(K)$.

We can also give a sequential definition of compactness, which generalizes the Bolzano-Weierstrass property.

Definition 7.36. A subset $K \subset X$ of a metric space X is sequentially compact if every sequence in K has a convergent subsequence whose limit belongs to K.

Explicitly, this 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 $k \to \infty$, and $x \in K$.

Theorem 7.37. A subset of a metric space is sequentially compact if and only if it is compact.

We can also define Cauchy sequences in a metric space.

Definition 7.38. 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

$$m, n > N \text{ implies that } d(x_m, x_n) < \epsilon.$$

Completeness of a metric space is defined using the Cauchy condition.

Definition 7.39. A metric space is complete if every Cauchy sequence converges.

For $\mathbb{R}$, completeness is equivalent to the existence of suprema, but general metric spaces are not ordered so this property does not apply to them.

Definition 7.40. A Banach space is a complete normed vector space.

Nearly all metric and normed spaces that arise in analysis are complete.

Example 7.41. The space $(\mathbb{R}, |\cdot|)$ is a Banach space. More generally, $\mathbb{R}^n$ with the ℓ^p-norm defined in Example 7.15 is a Banach space.

Example 7.42. If $K \subset \mathbb{R}$ is compact, the space $C(K)$ with the sup-norm is a Banach space. A sequence of functions (f_n) is Cauchy in $C(K)$ if and only if it is uniformly Cauchy. Thus, Theorem 5.21 states that $C(K)$ is complete.

7.5. Continuous functions

The definitions of limits and continuity of functions between metric spaces parallel the definitions for real functions.

Definition 7.43. Let (X, d_X) and (Y, d_Y) be metric spaces, and suppose that $c \in X$ is an accumulation point of X. If $f : X \setminus \{c\} \to Y$, then $y \in Y$ is the limit of $f(x)$ as $x \to c$, or

$$\lim_{x \to c} f(x) = y,$$

if for every $\epsilon > 0$ there exists $\delta > 0$ such that

$$0 < d_X(x, c) < \delta \text{ implies that } d_Y(f(x), y) < \epsilon.$$

In terms of neighborhoods, the definition says that for every neighborhood V of y in Y there exists a neighborhood U of c in X such that f maps $U \setminus \{c\}$ into V.

Definition 7.44. 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.

In terms of neighborhoods, the definition says that for every neighborhood V of $f(c)$ in Y there exists a neighborhood U of c in X such that f maps U into V.

Example 7.45. 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

$$|f(x_1, x_2) - f(c_1, c_2)| < \epsilon.$$

Example 7.46. 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 where $f_1, f_2 : \mathbb{R} \to \mathbb{R}$, this condition reads: $|x - c| < \delta$ implies that

$$\sqrt{[f_1(x) - f_1(c)]^2 + [f_1(x) - f_1(c)]^2} < \epsilon.$$

The previous examples generalize in a natural way to define the continuity of an m-component vector-valued function of n variables.

Example 7.47. A function $f : \mathbb{R}^n \to \mathbb{R}^m$, where both $\mathbb{R}^n$ and $\mathbb{R}^m$ are equipped with the Euclidean norm, is continuous at c if for every $\epsilon > 0$ there is a $\delta > 0$ such that

$$\|x - c\| < \delta \text{ implies that } \|f(x) - f(c)\| < \epsilon.$$

This definition would look complicated if it was written out explicitly, but it is much clearer when it is expressed in terms or metrics or norms.

We also have a sequential definition of continuity in a metric space.

Definition 7.48. Let X and Y be metric spaces. A function $f : X \to Y$ is sequentially continuous at $c \in X$ if

$$f(x_n) \to f(c) \qquad \text{as } n \to \infty$$

for every sequence (x_n) in X such that

$$x_n \to c \qquad \text{as } n \to \infty$$

As for real functions, this is equivalent to continuity.

Proposition 7.49. A function $f : X \to Y$ is sequentially continuous at $c \in X$ if and only if it is continuous at c.

We define uniform continuity similarly.

Definition 7.50. 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 identical to the proofs we gave for functions $f : \mathbb{R} \to \mathbb{R}$.

First, a function on a metric space is continuous if and only if the inverse images of open sets are open.

Theorem 7.51. 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 7.52. Let K be a compact metric space and Y a metric space. If $f : K \to Y$ is a continuous function, then $f(K)$ is a compact subset of Y.

Third, a continuous functions on a compact set is uniformly continuous.

Theorem 7.53. If $f : K \to Y$ is a continuous function on a compact set K, then f is uniformly continuous.

7.6. Appendix: The Minkowski inequality

Inequalities are essential to analysis. Their proofs, however, are often not obvious and may require considerable ingenuity. Moreover, there may be many different ways to prove the same inequality.

The triangle inequality for the ℓ^p-norm on $\mathbb{R}^n$ defined in Example 7.15 is called the Minkowski inequality, and it is one of the most important inequalities in analysis. In this section, we prove it in the Euclidean case $p = 2$. The general case, with arbitrary $1 < p < \infty$, is more involved, and we will not prove it here.

We first prove the Cauchy-Schwartz inequality, which is itself a fundamental inequality.

Theorem 7.54 (Cauchy-Schwartz)**.** If $(x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$, 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 either $x = 0$ or $y = 0$, so we assume at least one x_i and one y_i is nonzero.

For every $\alpha, \beta \in \mathbb{R}$, we have

$$0 \leq \sum_{i=1}^{n} (\alpha x_i - \beta y_i)^2 .$$

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. □

The Minkowski inequality for $p = 2$ is an immediate consequence of the Cauchy-Schwartz inequality.

Corollary 7.55. If $(x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$, 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 have

$$\begin{aligned}\sum_{i=1}^{n}(x_i+y_i)^2 &= \sum_{i=1}^{n}x_i^2 + 2\sum_{i=1}^{n}x_iy_i + \sum_{i=1}^{n}y_i^2 \\ &\leq \sum_{i=1}^{n}x_i^2 + 2\left(\sum_{i=1}^{n}x_i^2\right)^{1/2}\left(\sum_{i=1}^{n}y_i^2\right)^{1/2} + \sum_{i=1}^{n}y_i^2 \\ &\leq \left[\left(\sum_{i=1}^{n}x_i^2\right)^{1/2} + \left(\sum_{i=1}^{n}y_i^2\right)^{1/2}\right]^2.\end{aligned}$$

Taking the square root of this inequality, we get the result. □