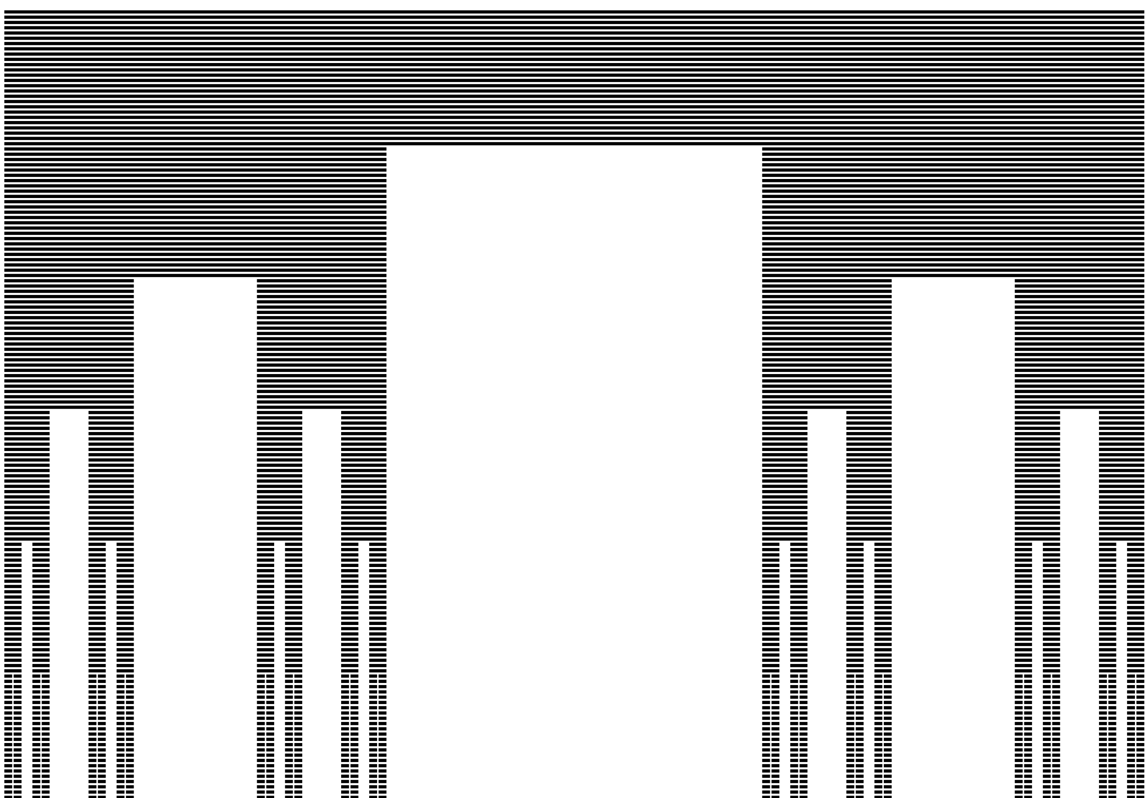
An Introduction to Real Analysis



Cesar O. Aguilar

May 4, 2022

Contents

Preface

The material in these notes constitute my personal notes that are used in the course lectures for MATH 324 and 325 (Real Analysis I, II). You will find that the lectures and these notes are very closely aligned. The notes highlight the important ideas and examples that you should master as a student. You may find these notes useful if:

- you miss a lecture and need to know what was covered,
- you want to know what material you are expected to master,
- you want to know the level of difficulty of questions that you should expect in a test, and
- you want to see more worked out examples in addition to those worked out in the lectures.

If you find any typos or errors in these notes, no matter how small, please email me a short description (with a page number) of the typo/error. Suggestions and comments on how to improve the notes are also welcomed.

Cesar O. Aguilar
SUNY Geneseo

1

Preliminaries

In this short chapter, we will briefly review some basic set notation, proof methods, functions, and countability. The presentation of these topics is intentionally brief for two reasons: (1) the reader is likely familar with these topics, and (2) we include only the necessary material needed to start *doing* real analysis.

1.1 Sets, Numbers, and Proofs

Let S be a set. If x is an **element** of S then we write $x \in S$, otherwise we write that $x \notin S$. A set A is called a **subset** of S if each element of A is also an element of S, that is, if $a \in A$ then also $a \in S$. To denote that A is a subset of S we write $A \subset S$.

Now let A and B be subsets of S. If $A \subset B$ and $B \subset A$ then A and B are said to be **equal** and we write that $A = B$. The **union** of A and B is the set

$$A \cup B = \{x \in S \mid x \in A \text{ or } x \in B\}$$

and the **intersection** of A and B is the set

$$A \cap B = \{x \in S \mid x \in A \text{ and } x \in B\}.$$

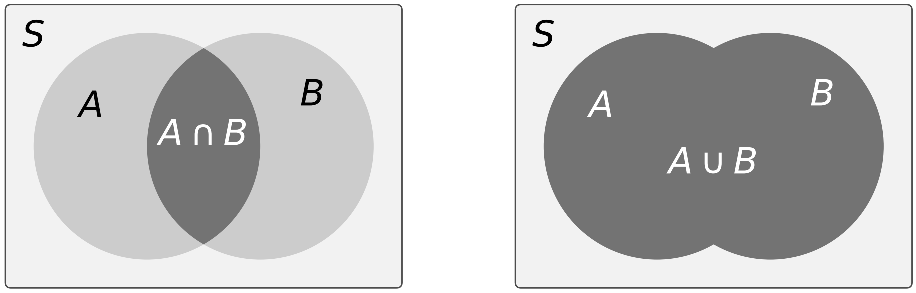


Figure 1.1: Set intersection $A \cap B$ and union $A \cup B$

A graphical representation of set unions and intersections are shown in Figure 1.1.

The **empty set** is the set that does not contain any elements and is denoted by $\emptyset$. We note that $\emptyset \subset S$ for any set S. The sets A and B are **disjoint** if $A \cap B = \emptyset$. The **complement** of A in S is the set

$$S \backslash A = \{x \in S \mid x \notin A\},$$

in other words, $S \backslash A$ consists of the elements in S not contained in A. We sometimes use the shorter notation A^c for $S \backslash A$ when it is clear that it is the complement of A relative to S.

The **Cartesian product** of A and B, denoted by $A \times B$, is the set of ordered pairs (a, b) where $a \in A$ and $b \in B$, in other words,

$$A \times B = \{(a, b) \mid a \in A,\ b \in B\}.$$

A **partition** of a set S is a set Π whose elements are subsets of S such that Π does not contain the empty set, the union of the elements of Π equals S, and any two distinct elements of Π are disjoint.

Lastly, for any set S, the **power set** of S is the set of all subsets of S, and is denoted by $\mathcal{P}(S)$.

Example 1.1.1. Let A and B be subsets of a set S. Show that

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

Solution. We first show that $(A \cup B)^c \subset A^c \cap B^c$. If $x \in (A \cup B)^c$ then by definition $x \notin (A \cup B)$ and therefore $x \notin A$ and $x \notin B$. Thus, $x \in A^c$ and $x \in B^c$, that is, $x \in A^c \cap B^c$.

Now suppose that $x \in A^c \cap B^c$, that is, suppose that $x \in A^c$ and $x \in B^c$. Thus, $x \notin A$ and $x \notin B$ and thus $x \notin (A \cup B)$. By definition, $x \in (A \cup B)^c$ and this proves that $A^c \cap B^c \subset (A \cup B)^c$. □

We use the symbol $\mathbb{N}$ to denote the set of **natural numbers**, that is,

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

The set of **integers** is denoted by $\mathbb{Z}$ so that

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

The set of **rational** numbers is denoted by

$$\mathbb{Q} = \left\{ \frac{p}{q} \mid p, q \in \mathbb{Z},\ q \neq 0 \right\}.$$

Notice that we have the following chain of set inclusions:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q}.$$

We now review the most commonly used methods of proof. To that end, recall that a **logical statement** is a declarative sentence that can be unambiguously decided to be either true or false. A **theorem** is a logical statement that has been proved to be true using a sequence of true statements and deductive reasoning. Many theorems are usually written as a conditional statement of the form "if P then Q" or symbolically "$P \Rightarrow Q$". The statement P is called the **hypothesis** or **assumption** and Q is called the **conclusion**. Below are the main techniques used to prove the statement "$P \Rightarrow Q$":

- **Direct Proof**: To prove the statement "$P \Rightarrow Q$", assume that the statement P is true and show by combining axioms, definitions, and earlier theorems that Q is true. This should be the first method you attempt.

- **Mathematical Induction**: Covered in Section 1.2.

- **By Contraposition**: Proving the statement "$P \Rightarrow Q$" by proving the logically equivalent statement "not $Q \Rightarrow$ not P". Do not confuse this with proof by contradiction.

- **By Contradiction**: To prove the statement "$P \Rightarrow Q$" by contradiction, one assumes that "$P \Rightarrow Q$" is false and then show that **some** contradiction results. Assuming that "$P \Rightarrow Q$" is false is to assume that P is true and Q is false. Using these to latter assumptions, one attempts to derive at a contradiction of the form "R and not R", where R is **some** statement. One disadvantage with proof by contradiction is that the logical contradiction that one is seeking "R and not R" is not known in advance so that the goal of the proof is unclear. Proof by contradiction frequently gets confused with proof by contraposition in the following way (do not do this): *To prove that "$P \Rightarrow Q$", assume that P is true and suppose that Q is not true. After some work using **only** the assumption that Q is not true you show that P is not true and thus you say that there is a contradiction because you assumed that P is true.* What you have really done is proved the contrapositive statement. Thus, if you believe that you are proving a statement by contradiction, take a close look at your proof to see if what you have is a proof by contraposition.

In the following example, we use both proof by contradiction and proof by contraposition.

Example 1.1.2. Prove that if x and y are consecutive integers then $x+y$ is odd.

Solution. Assume that x and y are consecutive integers (i.e. assume P) and assume that $x+y$ is not odd (i.e. assume not Q). Since $x+y$ is not odd then $x+y \neq 2n+1$ for all integers n. However, since x and y are consecutive, and assuming without loss of generality that $x<y$, we have that $x+y=2x+1$. Thus, we have that $x+y \neq 2n+1$ for all integers n and also that $x+y=2x+1$. Since x is an integer we have reached a contradiction. Hence, if x and y are consecutive integers then $x+y$ is odd.

Now we prove the statement by contraposition. Without loss of generality, suppose that $x<y$. Assume that $x+y$ is even. Then there exists an integer n such that $x+y=2n$ and therefore $x=2n-y$. Consequently,

$$y-x=y-(2n-y)=2(y-n).$$

Hence, since $(y-n)$ is an integer, $y-x \neq 1$ and consequently x and y are not consecutive integers. □

In general, if the statement "$P \Rightarrow Q$" is true then the converse conditional statement "$Q \Rightarrow P$" is not necessarily true. For example, the converse conditional statement in Example 1.1.2 is "if $x+y$ is odd then x and y are consecutive integers" is easily shown to be false (e.g., $x=2$ and $y=5$). The conjoined statement "$P \Rightarrow Q$ and $Q \Rightarrow P$", alternatively written as "P if and only if Q" or symbolically

"$P \Leftrightarrow Q$", is called a **biconditional statement**. Thus, to prove that the biconditional statement "$P \Leftrightarrow Q$" is true one must prove that both "$P \Rightarrow Q$" and "$Q \Rightarrow P$" are true.

Example 1.1.3. Let A, B, and C be subsets of a set S. Prove that $(A \cup B) \subset C$ if and only if $A \subset C$ and $B \subset C$.

Exercises

Exercise 1.1.1. Let A and B be subsets of a set S. Show that $A \subset B$ if and only if $B^c \subset A^c$

Exercise 1.1.2. Find the power set of $S = \{x, y, z, w\}$.

Exercise 1.1.3. Let $A = \{\alpha_1, \alpha_2, \alpha_3\}$ and let $B = \{\beta_1, \beta_2\}$. Find $A \times B$.

Exercise 1.1.4. Let $x \in \mathbb{Z}$. Prove that if x^2 is even then x is even. Do not use proof by contradiction.

Exercise 1.1.5. Prove that if x and y are even natural numbers then xy is even. Do not use proof by contradiction.

Exercise 1.1.6. Prove that if x and y are rational numbers then $x + y$ is a rational number. Do not use proof by contradiction.

Exercise 1.1.7. Let x and y be natural numbers. Prove that x and y are odd if and only if xy is odd. Do not use proof by contradiction.

1.2 Mathematical Induction

Mathematical induction is a powerful proof technique that relies on the following property of $\mathbb{N}$.

Axiom 1.2.1: Well-Ordering Principle

Every non-empty subset of $\mathbb{N}$ contains a smallest element.

In other words, if S is a non-empty subset of $\mathbb{N}$ then there exists $a \in S$ such that $a \leq x$ for all $x \in S$. The smallest element of S is denoted by $\min(S)$. Thus, $\min(S) \in S$ and $\min(S) \leq x$ for all $x \in S$. We now state and prove the principle of Mathematical Induction.

Theorem 1.2.2: Mathematical Induction

Suppose that S is a subset of $\mathbb{N}$ with the following properties:

(i) $1 \in S$

(ii) If $k \in S$ then also $k + 1 \in S$.

Then $S = \mathbb{N}$.

Proof. Suppose that S is a subset of $\mathbb{N}$ with properties (i) and (ii) and let $T = \mathbb{N} \backslash S$. Proving that $S = \mathbb{N}$ is equivalent to proving that T is the empty set. Suppose then by contradiction that T is non-empty. By the well-ordering principle of $\mathbb{N}$, T has a smallest element, say it is $a \in T$. Because S satisfies property (i) we know that $1 \notin T$ and therefore $a > 1$. Now since a is the least element of T, then $a - 1 \in S$ (we know that $a - 1 > 0$ because $a > 1$). But since S satisfies property (ii) then $(a-1)+1 \in S$, that is, $a \in S$. This is a contradiction because we cannot have both $a \in T$ and $a \in S$. Thus, T is the empty set, and therefore $S = \mathbb{N}$. $\square$

Mathematical induction is frequently used to prove formulas or inequalities involving the natural numbers. For example, consider the validity of the formula

$$1 + 2 + 3 + \cdots + n = \tfrac{1}{2}n(n+1) \tag{1.1}$$

where $n \in \mathbb{N}$. In words, the identity (1.1) says that the sum of all the integers from 1 to n equals $\frac{1}{2}n(n+1)$. We use induction to show that this formula is true for all $n \in \mathbb{N}$. Let S be the subset of $\mathbb{N}$ consisting of the natural numbers that satisfy (1.1), that is,

$$S = \{n \in \mathbb{N} \mid 1 + 2 + 3 + \cdots + n = \tfrac{1}{2}n(n+1)\}.$$

If $n = 1$ then

$$\tfrac{1}{2}n(n+1) = \tfrac{1}{2}(1+1) = 1.$$

Thus, $\frac{1}{2}n(n+1)$ is equal to the sum of all the integers from 1 to $n = 1$. Hence, (1.1) is true when $n = 1$ and thus $1 \in S$. Now suppose that some $k \in \mathbb{N}$ satisfies (1.1), that is, suppose that $k \in S$. Then we may write that

$$1 + 2 + \cdots + k = \tfrac{1}{2}k(k+1). \tag{1.2}$$

We will prove that the integer $k+1$ also satisfies (1.1). To that end, adding $k+1$ to both sides of (1.2) we obtain

$$1 + 2 + \cdots + k + (k+1) = \tfrac{1}{2}k(k+1) + (k+1).$$

Now notice that we can factor $(k+1)$ from the right-hand side and through some algebraic steps we obtain that

$$\begin{aligned} 1 + 2 + \cdots + k + (k+1) &= \tfrac{1}{2}k(k+1) + (k+1) \\ &= (k+1)[\tfrac{1}{2}k + 1] \\ &= \tfrac{1}{2}(k+1)(k+2). \end{aligned}$$

Hence, (1.1) also holds for $n = k + 1$ and thus $k + 1 \in S$. We have therefore proved that S satisfies properties (i) and (ii), and therefore by mathematical induction $S = \mathbb{N}$, or equivalently that (1.1) holds for all $n \in \mathbb{N}$.

Example 1.2.3. Use mathematical induction to show that $2^n \leq (n+1)!$ holds for all $n \in \mathbb{N}$.

Example 1.2.4. Let $r \neq 1$ be a constant. Use mathematical induction to show that

$$1 + r + r^2 + \cdots + r^n = \frac{1 - r^{n+1}}{1 - r}$$

holds for all $n \in \mathbb{N}$.

Example 1.2.5 (Bernoulli's inequality)**.** Prove that if $x > -1$ then $(1 + x)^n \geq 1 + nx$ for all $n \in \mathbb{N}$.

Proof. The statement is trivial for $n = 1$. Assume that for some $k \in \mathbb{N}$ it holds that $(1 + x)^k \geq 1 + kx$. Since $x > -1$ then $x + 1 > 0$ and therefore

$$\begin{aligned}(1 + x)^k(1 + x) &\geq (1 + kx)(1 + x) \\ &= 1 + (k + 1)x + kx^2 \\ &\geq 1 + (k + 1)x.\end{aligned}$$

Therefore, $(1 + x)^{k+1} \geq 1 + (k + 1)x$, and the proof is complete by induction. ☐

There is another version of mathematical induction called the **Principle of Strong Induction** which we now state.

Theorem 1.2.6: Strong Induction

Suppose that S is a subset of $\mathbb{N}$ with the following properties:

(i) $1 \in S$

(ii) If $\{1, 2, \ldots, k\} \subset S$ then also $k + 1 \in S$.

Then $S = \mathbb{N}$.

Do you notice the difference between induction and strong induction? It turns out that the two statements are equivalent, in other words, if S satisies either one of properties (i)-(ii) of induction or strong induction then we may conclude that $S = \mathbb{N}$. The upshot with strong induction is that one is able to use the stronger condition that $\{1, 2, \ldots, k\} \subset S$ to prove that $k + 1 \in S$.

Exercises

Exercise 1.2.1. Prove that $n < 2^n$ for all $n \in \mathbb{N}$.

Exercise 1.2.2. Prove that $2^n < n!$ for all $n \geq 4$, $n \in \mathbb{N}$.

Exercise 1.2.3. Use induction to prove that if S has n elements then $\mathcal{P}(S)$ has 2^n elements. **Hint**: If S is a set with $n+1$ elements, for instance $S = \{x_1, x_2, \ldots, x_n, x_{n+1}\}$, then argue that $\mathcal{P}(S) = \mathcal{P}(\tilde{S}) \cup \mathcal{T}$ where $\tilde{S} = S \backslash \{x_{n+1}\}$ and $\mathcal{T}$ consists of subsets of S that contain x_{n+1}. How many sets are in $\mathcal{P}(\tilde{S})$ and how many are in $\mathcal{T}$? And what is $\mathcal{P}(\tilde{S}) \cap \mathcal{T}$? Explain carefully.

1.3 Functions

Let A and B be sets. A **function** from A to B is a rule that assigns to each element $x \in A$ one element $y \in B$. The set A is called the **domain** of f and B is called the **co-domain** of f. We usually denote a function with the notation $f : A \to B$, and the assignment of x to y is written as $y = f(x)$. We also say that f is a **mapping** from A to B, or that f **maps** A into B. The element y assigned to x is called the **image** of x under f. The **range** of f, denoted by $f(A)$, is the set

$$f(A) = \{y \in B \mid \exists\, x \in A,\ y = f(x)\}.$$

In the above definition of $f(A)$, we use the symbol $\exists$ as a short-hand for "there exsits". By definition, $f(A) \subset B$ but in general we do not have that $f(A) = B$, in other words, the range of a function is generally a strict subset of the function's co-domain.

Example 1.3.1. Consider the mapping $f : \mathbb{Q} \to \mathbb{Z}$ defined by

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

The image of $x = 1/2$ under f is $f(1/2) = 1$. The range of f is $f(\mathbb{Q}) = \{1, -1\}$.

Example 1.3.2. Consider the function $f : \mathbb{N} \to \mathcal{P}(\mathbb{N})$ defined by

$$f(n) = \{1, 2, \ldots, n\}.$$

The set $S = \{2, 4, 6, 8, \ldots,\}$ of even numbers is an element of the co-domain $\mathcal{P}(\mathbb{N})$ but is not in the range of f. As another example, the set $\mathbb{N} \in \mathcal{P}(\mathbb{N})$ itself is not in the range of f.

Function's whose range is equal to it's co-domain are given a special name.

Definition 1.3.3: Surjection

A function $f : A \to B$ is said to be a **surjection** if for any $y \in B$ there exists $x \in A$ such that $f(x) = y$.

In other words, $f : A \to B$ is a surjection if $f(A) = B$.

Example 1.3.4. The function $f : \mathbb{Q} \to \mathbb{Q}$ defined by $f(x) = x^2$ is not a surjection. For example, $y = -1$ is clearly not in the range of f since $f(x) = x^2 \neq -1$ for all $x \in \mathbb{Q}$. On the other hand, $y = \frac{121}{64}$ is in the range of f since $f(11/8) = \frac{121}{64}$. Is $y = 2$ in the range of f?

Example 1.3.5. Consider the function $f : \mathcal{P}(\mathbb{N}) \to \mathbb{N}$ defined by

$$f(S) = \min(S).$$

Prove that f is a surjection.

Solution. To prove that f is a surjection, we must show that for any element $y \in \mathbb{N}$ (the co-domain), there exists $S \in \mathcal{P}(\mathbb{N})$ (the domain) such that $f(S) = y$. Consider then an arbitrary $y \in \mathbb{N}$. Let $S = \{y\}$ and thus clearly $S \in \mathcal{P}(\mathbb{N})$. Moreover, it is clear that $\min(S) = y$ and thus $f(S) = \min(S) = y$. This proves that f is a surjection. □

Notice that in Example 1.3.5, given any $y \in \mathbb{N}$ there are many sets $S \in \mathcal{P}(\mathbb{N})$ such that $f(S) = y$. This leads us to the following definition.

Definition 1.3.6: Injection

A function $f : A \to B$ is said to be an **injection** if no two distinct elements of A are mapped to the same element in B, in other words, for any $x_1, x_2 \in A$, if $x_1 \neq x_2$ then $f(x_1) \neq f(x_2)$.

In other words, f is an injection if whenever $f(x_1) = f(x_2)$ then necessarily $x_1 = x_2$.

Example 1.3.7. The function $f : \mathbb{Q} \to \mathbb{Q}$ defined by $f(x) = x^2$ is not an injection. For example, $f(-2) = f(2) = 4$.

Example 1.3.8. Consider again the function $f : \mathbb{N} \to \mathcal{P}(\mathbb{N})$ defined by

$$f(n) = \{1, 2, \ldots, n\}.$$

This function is an injection. Indeed, if $f(n) = f(m)$ then $\{1, 2, \ldots, n\} = \{1, 2, \ldots, m\}$ and therefore $n = m$. Hence, whenever $f(n) = f(m)$ then necessarily $n = m$ and this proves that f is an injection.

Example 1.3.9. Consider the function $f : \mathcal{P}(\mathbb{N}) \to \mathbb{N}$ defined by

$$f(S) = \min(S)$$

Is f an injection?

Example 1.3.10. Consider the function $f : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ defined by

$$f(n) = (2n^2, n + 1)$$

Show that f is an injection.

Definition 1.3.11: Bijection

A function $f : A \to B$ is said to be a **bijection** if it is a surjection and an injection.

Example 1.3.12. Suppose that $f : P \to Q$ is an injection. Prove that the function $\tilde{f} : P \to f(P)$ defined by $\tilde{f}(x) = f(x)$ for $x \in P$ is a bijection.

Solution. By construction, $\tilde{f}$ is a surjection. If $\tilde{f}(x) = \tilde{f}(y)$ then $f(x) = f(y)$ and then $x = y$ since f is an injection. Thus, $\tilde{f}$ is an injection and this proves that $\tilde{f}$ is a bijection. ☐

Example 1.3.13. Prove that $f : \mathbb{Q}\backslash\{0\} \to \mathbb{Q}\backslash\{0\}$ defined by $f(\frac{p}{q}) = \frac{q}{p}$ is a bijection, where $\gcd(p, q) = 1$.

Suppose that $f : A \to B$ is a bijection and define the function $g : B \to A$ as follows: for $b \in B$ let $g(b)$ be the (necessarily unique) element in A such that $f(g(b)) = b$. Notice that by definition, $g(f(a)) = a$. The function g is called the **inverse** of f and we write instead $g = f^{-1}$. It is not hard to show that g is a bijection and that $g^{-1} = f$.

Given functions $f : A \to B$ and $g : B \to C$, the **composition** of g and f is the function $(g \circ f) : A \to C$ defined as $(g \circ f)(a) = g(f(a))$.

Theorem 1.3.14

If $f : A \to B$ and $g : B \to C$ are injections (surjections) then the composition $(g \circ f) : A \to C$ is an injection (surjection).

Proof. Assume that $f : A \to B$ and $g : B \to C$ are injections. To prove that $(g \circ f)$ is an injection, suppose that $(g \circ f)(x_1) = (g \circ f)(x_2)$. Then by definition of $(g \circ f)$, it follows that $g(f(x_1)) = g(f(x_2))$. Now since g is an injection then necessarily $f(x_1) = f(x_2)$ and since f is an injection then necessarily $x_1 = x_2$. Thus if $(g \circ f)(x_1) = (g \circ f)(x_2)$ then $x_1 = x_2$ and this proves that $(g \circ f)$ is an injection.

Now suppose that f and g are surjections. To prove that $(g \circ f) : A \to C$ is a surjection, let $z \in C$ be arbitrary. Since g is a surjection, there exists $y \in B$ such that $g(y) = z$. Since f is a surjection, there exists $x \in A$ such that $y = f(x)$. Thus, for $x \in A$ we have that $(g \circ f)(x) = g(f(x)) = g(y) = z$. Hence, for arbitrary $z \in C$ there exists $x \in A$ such that $(g \circ f)(x) = z$ and this proves that $(g \circ f)$ is a surjection. $\square$

The following result is then an immediate application of Theorem 1.3.14 and the definition of a bijection.

Corollary 1.3.15

The composition of two bijections is a bijection.

Exercises

Exercise 1.3.1. Consider the function $f : \mathbb{N} \to \mathbb{Q}$ defined as $f(n) = \frac{1}{n}$. Is f an injection? Is f a surjection?

Exercise 1.3.2. Consider the function $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ defined as $f(n, m) = nm$. Is f an injection? Is f a surjection?

Exercise 1.3.3. Consider the function $f : \mathbb{Q} \to \mathbb{Q}$ defined as $f(x) = (x - 2)(x - 6)$. Is f an injection? If f a surjection?

Exercise 1.3.4. Let $f : \mathbb{N} \to \mathbb{Z}$ be the function defined as

$$f(n) = \begin{cases} \frac{n}{2}, & \text{if } n \text{ is even,} \\ -\frac{(n-1)}{2}, & \text{if } n \text{ is odd.} \end{cases}$$

Prove that f is a bijection.

Exercise 1.3.5. The **sign** of a rational number $x \in \mathbb{Q}$ is defined as $\text{sgn}(x) = x/|x|$ if $x \neq 0$, where $|x|$ is the absolute value of x, and $\text{sgn}(0) = 1$. For example, $\text{sgn}(-3) = -1$ and $\text{sgn}(2) = 1$. Prove that the function $f : \mathbb{Z} \to \{-1, 1\} \times \mathbb{N}$ defined as

$$f(x) = (\text{sgn}(x), |x| + 1)$$

is a bijection.

1.4 Countability

A non-empty set S is said to be **finite** if there is a bijection from $\{1, 2, \ldots, n\}$ onto S for some $n \in \mathbb{N}$. In this case, we say that S contains n elements and we write $|S| = n$. If S is not finite then we say that S is **infinite**.

Example 1.4.1. Let $f : P \to Q$ be an injection. If P is an infinite set then $f(P)$ is an infinite set.

Solution. The proof is by contraposition. Suppose that $f(P)$ is a finite set containing n elements. Then there exists a bijection $g : \{1, 2, \ldots, n\} \to f(P)$. The function $\tilde{f} : P \to f(P)$ defined as $\tilde{f}(x) = f(x)$ for $x \in P$ is a bijection (Example 1.3.12) and therefore $(\tilde{f}^{-1} \circ g) : \{1, 2, \ldots, n\} \to P$ is a bijection. Thus P is a finite set and completes the proof. □

We now introduce the notion of a **countable** set.

Definition 1.4.2: Countability

Let S be a set.

(i) The set S is **countably infinite** if there is a bijection from $\mathbb{N}$ onto S.

(ii) The set S is **countable** if S is either finite or countably infinite.

(iii) The set S is **uncountable** if S is not countable.

Roughly speaking, a set S is countable if the elements of S can be **listed** or **enumerated** in a systematic manner. To see how, suppose that S is countably infinite and let $f : \mathbb{N} \to S$ be a bijection. Then the elements of S can be listed as

$$S = \{f(1), f(2), f(3), f(4), \ldots\}.$$

Hence, although sets have no predetermined order, the elements of a countable set *can* be ordered.

Example 1.4.3. The set S of odd natural numbers is countable. Recall that $n \in \mathbb{N}$ is an odd positive integer if $n = 2k - 1$ for some $k \in \mathbb{N}$. A bijection $f : \mathbb{N} \to S$ from $\mathbb{N}$ to S is $f(k) = 2k - 1$. The function f can be interpreted as a listing of the odd natural numbers in the natural way:

$$S = \{f(1), f(2), f(3), \ldots\} = \{1, 3, 5, \ldots\}.$$

Example 1.4.4. The natural numbers $S = \mathbb{N}$ are countable. A bijection $f : \mathbb{N} \to S$ is $f(n) = n$, i.e., the identity mapping.

Example 1.4.5. The set of integers $\mathbb{Z}$ is countable. A bijection f from $\mathbb{N}$ to $\mathbb{Z}$ can be defined by listing the elements of $\mathbb{Z}$ as follows:

$$\begin{array}{rccccccccc}
\mathbb{N}: & 1 & 2 & 3 & 4 & 5 & 6 & 7 & \ldots \\
 & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \ldots \\
\mathbb{Z}: & 0 & 1 & -1 & 2 & -2 & 3 & -3 & \ldots
\end{array}$$

To be more precise, f is the function

$$f(n) = \begin{cases} \frac{n}{2}, & \text{if } n \text{ is even,} \\ -\frac{(n-1)}{2}, & \text{if } n \text{ is odd.} \end{cases}$$

It is left as an exercise to show that f is indeed a bijection.

Example 1.4.6. The set $\mathbb{N}\times\mathbb{N}$ is countable. There are many bijections from $\mathbb{N}$ to $\mathbb{N}\times\mathbb{N}$ but a particularly interesting one is the function defined as follows. Consider the family of lines

$$L_k = \{(x, y) \in \mathbb{N} \times \mathbb{N} \,|\, y = -x + k + 1\}$$

for $k \in \mathbb{N}$. There are k points on the line L_k, namely $(j, k+1-j)$ for $1 \leq j \leq k$, and we say that $(j, k+1-j)$ is the jth point on the line L_k. The point $(x, y) \in \mathbb{N} \times \mathbb{N}$ is contained in the line L_k where $k = x+y-1$. Thus, one way to enumerate the points in $\mathbb{N} \times \mathbb{N}$ is to assign to each $(x, y) \in L_k$ the number $\rho(x, y)$ obtained by adding all points on the lines $L_1, \ldots, L_{k-1}$ and adding the position of (x, y) on line L_k, namely x. Thus,

$$\begin{aligned}\rho(x, y) &= 1 + 2 + \ldots + (k-1) + x \\ &= \frac{1}{2}(k-1)k + x \\ &= \frac{1}{2}(x+y-2)(x+y-1) + x.\end{aligned}$$

The function ρ is called the **Cantor pairing** function. Alternatively, we use a modified version of ρ which we call $\tau : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ and defined as

$$\tau(x, y) = \begin{cases} \rho(x, y), & \text{if } (x+y-1) \text{ is odd,} \\ \rho(y, x), & \text{if } (x+y-1) \text{ is even.} \end{cases}$$

We now find a formula for the inverse of τ which we call the **Cantor snake**. To write down the formula for τ^{-1}, we first let for each $n \in \mathbb{N}$

$$m = \text{floor}\left(\frac{-1+\sqrt{1+8(n-1)}}{2}\right)$$

and we note that $m \geq 0$ is the smallest integer such that $\frac{1}{2}m(m+1) < n$. We then set $p(n) = n - \frac{1}{2}m(m+1)$ a then

$$\tau^{-1}(n) = \begin{cases} (p(n), -p(n)+m+2), & \text{if } m \text{ is even} \\ (-p(n)+m+2, p(n)), & \text{if } m \text{ is odd.} \end{cases}$$

We note that the point $(x, y) = \tau^{-1}(n)$ is on the line L_k with $k = m+1$. The range of τ^{-1} is

$$\tau^{-1}(\mathbb{N}) = \{(1,1), (2,1), (1,2), (1,3), (2,2), (3,1), \ldots\}.$$

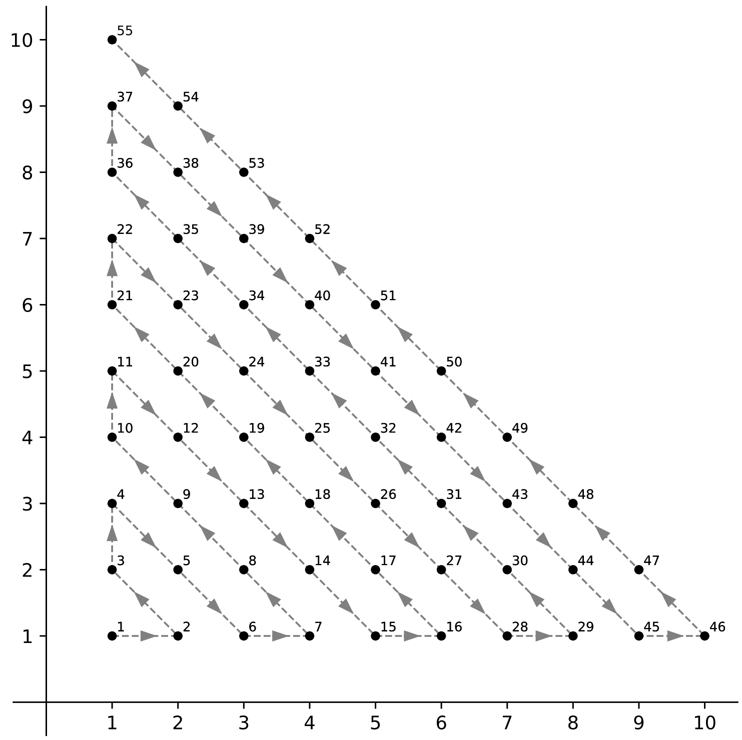


Figure 1.2: The image of the Cantor snake

The sequence of pairs $(x, y) \in \mathbb{N} \times \mathbb{N}$ generated by τ^{-1} for $1 \leq n \leq 55$ are shown in Figure 1.2.

Example 1.4.7. Suppose that $f : T \to S$ is a bijection. Prove that T is countable if and only if S is countable.

Solution. Suppose that T is countable. Then by definition there exists a bijection $g : \mathbb{N} \to T$. Since g and f are bijections, the composite function $(f \circ g) : \mathbb{N} \to S$ is a bijection. Hence, S is countable.

Now suppose that S is countable. Then by definition there exists a bijection $h : \mathbb{N} \to S$. Since f is a bijection then the inverse function $f^{-1} : S \to T$ is also a bijection. Therefore, the composite function $(f^{-1} \circ h) : \mathbb{N} \to T$ is a bijection. Thus, T is countable. □

As the following theorem states, countability, or lack thereof, can be inherited via set inclusion.

Theorem 1.4.8: Inheriting Countability

Let S and T be sets and suppose that $T \subset S$.

(i) If S is countable then T is also countable.

(ii) If T is uncountable then S is also uncountable.

Proof. To prove (i), let S be a countable set and let $f : S \to \mathbb{N}$ be a bijection. Define the mapping $\tilde{f} : T \to f(T)$ by $\tilde{f}(x) = f(x)$. By Example 1.3.12, $\tilde{f}$ is a bijection. Therefore, since $f(T)$ is a subset of $\mathbb{N}$, and thus countable, then T is countable. The proof of (ii) is left as an exercise. □

Example 1.4.9. Let S be the set of odd natural numbers. In Example 1.4.3, we proved that the odd natural numbers are countable by explicitly constructing a bijection from $\mathbb{N}$ to S. Alternatively, since $\mathbb{N}$ is countable and $S \subset \mathbb{N}$ then by Theorem 1.4.8 S is countable. More generally, any subset of $\mathbb{N}$ is countable.

If S is known to be a finite set then by Definition 1.4.2 S is countable, while if S is infinite then S may or may not be countable (we have yet to encounter an uncountable set but soon we will). To prove that a given infinite set S is countable we could use Theorem 1.4.8 if it is applicable but otherwise we must use Definition 1.4.2, that is, we must show that there is a bijection from S to $\mathbb{N}$, or equivalently from $\mathbb{N}$ to S. However, suppose that we can only prove the existence of a surjection f from $\mathbb{N}$ to S. The problem might be that f is not an injection and thus not a bijection. However, the fact that f is a surjection from $\mathbb{N}$ to S somehow says that S is no "larger" than $\mathbb{N}$ and gives evidence that perhaps S is countable. Could we use a surjection $f : \mathbb{N} \to S$ to construct a bijection $g : \mathbb{N} \to S$? Or, what if instead we had an injection $g : S \to \mathbb{N}$; could we use g to construct a bijection $f : S \to \mathbb{N}$? The following theorem says that it is indeed possible to do both.

Theorem 1.4.10: Countability Relaxations

Let S be a set.

(i) If there exists an injection $g : S \to \mathbb{N}$ then S is countable.

(ii) If there exists a surjection $f : \mathbb{N} \to S$ then S is countable.

Proof. (i) Let $g : S \to \mathbb{N}$ be an injection. Then the function $\tilde{g} : S \to g(S)$ defined by $\tilde{g}(s) = g(s)$ for $s \in S$ is a bijection. Since $g(S) \subset \mathbb{N}$ then $g(S)$ is countable. Therefore, S is countable also.

(ii) Now let $f : \mathbb{N} \to S$ be a surjection. For $s \in S$ let $f^{-1}(s) = \{n \in \mathbb{N} \mid f(n) = s\}$. Since f is a surjection, $f^{-1}(s)$ is non-empty for each $s \in S$. Consider the function $h : S \to \mathbb{N}$ defined by $h(s) = \min f^{-1}(s)$. Then $f(h(s)) = s$ for each $s \in S$. We claim that h is an injection. Indeed, if $h(s) = h(t)$ then $f(h(s)) = f(h(t))$ and thus $s = t$, and the claim is proved. Thus, h is an injection and then by (i) we conclude that S is countable. $\square$

We must be careful when using Theorem 1.4.10; if $f : \mathbb{N} \to S$ is known to be an injection then we cannot conclude that S is countable and similarly if $f : S \to \mathbb{N}$ is known to be a surjection then we cannot conclude that S is countable.

Example 1.4.11. In this example we will prove that the union of countable sets is countable. Hence, suppose that A and B are countable. By definition, there exist bijections $f : \mathbb{N} \to A$ and $g : \mathbb{N} \to B$.

Consider the function $h : \mathbb{N} \to A \cup B$ defined as follows:

$$h(n) = \begin{cases} f((n+1)/2), & \text{if } n \text{ is odd,} \\ g(n/2), & \text{if } n \text{ is even.} \end{cases}$$

We claim that h is a surjection (Loosely speaking, if $A = \{a_1, a_2, \ldots, \}$ and $B = \{b_1, b_2, \ldots, \}$, then the function h lists the elements of $A \cup B$ as $A \cup B = \{a_1, b_1, a_2, b_2, a_3, b_3, \ldots, \}$.). To see this, let $x \in A \cup B$. If $x \in A$ then $x = f(k)$ for some $k \in \mathbb{N}$. Then $h(2k-1) = f(k) = x$. If on the other hand $x \in B$ then $x = g(k)$ for some $k \in \mathbb{N}$. Then $h(2k) = g(k) = x$. In either case, there exists $n \in \mathbb{N}$ such that $h(n) = x$, and thus h is a surjection. By Theorem 1.4.10, the set $A \cup B$ is countable.

This example can be generalized as follows. Let $A_1, A_2, A_3, \ldots,$ be countable sets and let $S = \bigcup_{k=1}^{\infty} A_k$. Then S is countable. To prove this, we first enumerate the elements of each A_k as follows:

$$\begin{aligned} A_1 &= \{a_{1,1}, a_{1,2}, a_{1,3}, \ldots\} \\ A_2 &= \{a_{2,1}, a_{2,2}, a_{2,3}, \ldots\} \\ A_3 &= \{a_{3,1}, a_{3,2}, a_{3,3}, \ldots\} \\ \cdots & \quad \cdots \quad \cdots \end{aligned}$$

Formally, we have surjections $f_k : \mathbb{N} \to A_k$ for each $k \in \mathbb{N}$. Consider the mapping $\varphi : \mathbb{N} \times \mathbb{N} \to S$ defined by

$$\varphi(m, n) = a_{m,n} = f_m(n).$$

It is clear that φ is a surjection. Since $\mathbb{N} \times \mathbb{N}$ is countable, there is a surjection $\phi : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$, and therefore the composition $(\varphi \circ \phi) : \mathbb{N} \to S$ is a surjection. Therefore, S is countable.

The following theorem is perhaps surprising.

Theorem 1.4.12

The set of rational numbers $\mathbb{Q}$ is countable.

Proof. Let $\mathbb{Q}_{>0}$ be the set of all positive rational numbers and let $\mathbb{Q}_{<0}$ be the set of all negative rational numbers. Clearly, $\mathbb{Q} = \mathbb{Q}_{<0} \cup \{0\} \cup \mathbb{Q}_{>0}$, and thus it is enough to show that $\mathbb{Q}_{<0}$ and $\mathbb{Q}_{>0}$ are countable. In fact, we have the bijection $h : \mathbb{Q}_{>0} \to \mathbb{Q}_{<0}$ given by $h(x) = -x$, and thus if we can show that $\mathbb{Q}_{>0}$ is countable then this implies that $\mathbb{Q}_{<0}$ is also countable. In summary, to show that $\mathbb{Q}$ is countable it is enough to show that $\mathbb{Q}_{>0}$ is countable. To show that $\mathbb{Q}_{>0}$ is countable, consider the function $f : \mathbb{N} \times \mathbb{N} \to \mathbb{Q}_{>0}$ defined as

$$f(p, q) = \frac{p}{q}.$$

By definition, any rational number $x \in \mathbb{Q}_{>0}$ can be written as $x = \frac{p}{q}$ for some $p, q \in \mathbb{N}$. Hence, $x = f(p, q)$ and thus x is in the range of f. This shows that f is a surjection. Now, because $\mathbb{N} \times \mathbb{N}$ is countable, there is a surjection $g : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ and thus $(f \circ g) : \mathbb{N} \to \mathbb{Q}_{>0}$ is a surjection. By Theorem 1.4.10, this proves that $\mathbb{Q}_{>0}$ is countable and therefore $\mathbb{Q}$ is countable. $\square$

We end this section with Cantor's theorem named after mathematician Georg Cantor (1845-1918). Cantor is considered as the creator of set theory. Originally interested in analytical problems having as their root problems in physics, and in particular in characterizing solutions to equations describing heat conduction, Cantor discovered that infinite sets come in many possible sizes. One of Cantor's fascinating discoveries, which initially were very controversial at the time, led to the following theorem [1].

Theorem 1.4.13: Cantor's Theorem (1891)

For any set S, there is no surjection of S onto the power set $\mathcal{P}(S)$.

Proof. Suppose by contradiction that $f : S \to \mathcal{P}(S)$ is a surjection. By definition, for each $a \in S$, $f(a)$ is a subset of S. Consider the set

$$\mathcal{C} = \{a \in S \mid a \notin f(a)\}.$$

Since $\mathcal{C}$ is a subset of S then $\mathcal{C} \in \mathcal{P}(S)$. Since f is a surjection, there exists $x \in S$ such that $\mathcal{C} = f(x)$. One of the following must be true: either $x \in \mathcal{C}$ or $x \notin \mathcal{C}$. If $x \in \mathcal{C}$ then $x \notin f(x)$ by definition of $\mathcal{C}$. But $\mathcal{C} = f(x)$ and thus we reach contradiction. If $x \notin \mathcal{C}$ then by definition of $\mathcal{C}$ we have $x \in f(x)$. But $\mathcal{C} = f(x)$ and thus we reach a contradiction. Hence, neither of the possibilities are true, and thus we have reached an absurdity. Hence, we conclude that there is no such surjection f. ☐

Cantor's theorem implies that $\mathcal{P}(\mathbb{N})$ is uncountable. Indeed, if we take $S = \mathbb{N}$ in Cantor's Theorem then there is no surjection from $\mathbb{N}$ to $\mathcal{P}(\mathbb{N})$, and thus certainly no bijection from $\mathbb{N}$ to $\mathcal{P}(\mathbb{N})$. In summary:

Corollary 1.4.14

The set $\mathcal{P}(\mathbb{N})$ is uncountable.

Exercises

Exercise 1.4.1. Let P and Q be infinite sets. Prove that if $f : Q \to P$ is a bijection then Q is uncountable if and only if P is uncountable. Do not use proof by contradiction.

Exercise 1.4.2. In this exercise you will provide another proof that $\mathbb{N} \times \mathbb{N}$ is countable.

(a) Prove that 3^n is odd for each $n \in \mathbb{N}$.

(b) Consider the function $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ defined as $f(p, q) = 2^p 3^q$. Prove that f is an injection. Hint: Use part (a) in some way.

(c) Explain how part (b) proves that $\mathbb{N} \times \mathbb{N}$ is countable.

Exercise 1.4.3. Prove that if A and B are countably infinite then $A \times B$ is countable.

Exercise 1.4.4. Recall that a sequence $a = \{a_k\}_{k=1}^{\infty}$ of numbers is an infinite list

$$a = (a_1, a_2, a_3, a_4, \ldots)$$

where each element a_k is a number (We will cover sequences in detail but you are already familiar with them from calculus.). Let Q be the set of sequences whose elements are either 0 or 1, in other words,

$$Q = \left\{ \{a_k\}_{k=1}^{\infty} \mid a_k = 0 \text{ or } a_k = 1 \right\}.$$

For example, the following sequences are elements of the set Q:

$$a = (0, 0, 0, 0, 0, 0, \ldots) \in Q$$
$$b = (1, 0, 1, 0, 1, 0, \ldots) \in Q$$
$$c = (0, 0, 1, 1, 0, 0, \ldots) \in Q$$
$$d = (1, 1, 1, 1, 1, 1, \ldots) \in Q$$

(a) Give two non-trivial examples of functions from $\mathbb{N}$ to Q.

(b) Consider the function $f : Q \to \mathcal{P}(\mathbb{N})$ defined as

$$f(a) = \{k \in \mathbb{N} \mid a_k = 1\}$$

where $\mathcal{P}(\mathbb{N})$ is the power set of $\mathbb{N}$. Hence, f takes a sequence $a \in Q$ and outputs the indices where a is equal to 1. For example:

$$f(0, 0, 0, 0, 0, 0, 0, 0, \ldots) = \emptyset$$

$$f(1, 0, 1, 0, 1, 0, 1, 0, \ldots) = \{1, 3, 5, 7, \ldots\}$$

$$f(0, 0, 1, 1, 0, 0, 0, 0, \ldots) = \{3, 4\}$$

$$f(1, 1, 1, 1, 1, 1, 1, 1, \ldots) = \{1, 2, 3, 4, 5, 6, \ldots\} = \mathbb{N}$$

Prove that $f : Q \to \mathcal{P}(\mathbb{N})$ is a bijection.

(c) Combine part (b), Exercise 1.4.1, and Cantor's theorem to thoroughly explain whether Q is countable or uncountable.

2

The Real Numbers

2.1 Introduction

Recall that a *rational number* is a number x that can be written in the form $x = \frac{p}{q}$ where p, q are integers with $q \neq 0$. The rational number system is all you need to accomplish most everyday tasks. For instance, to measure distances when building a house it suffices to use a tape measure with an accuracy of about $\frac{1}{16}$ of an inch. However, to do mathematical analysis the rational numbers have some very serious shortcomings; here is a an example.

Theorem 2.1.1

If $x^2 = 2$ then x is not a rational number.

Proof. Suppose by contradiction that there exists $x \in \mathbb{Q}$ such that $x^2 = 2$. We may write $x = \frac{p}{q}$ for some integers p, q, and we can assume that p and q have no common factor other than 1 (that is, p and q are *relatively prime*). Now, since $x^2 = 2$ then $p^2 = 2q^2$ and thus p^2 is an even number. This implies that p is also even. Since p is even, we may write $p = 2k$ for some $k \in \mathbb{N}$ and therefore $(2k)^2 = 2q^2$, from which it follows that $2k^2 = q^2$. Hence, q^2 is even and thus q is also even.

Thus, both p and q are even, which contradicts the fact that p and q are relatively prime. □

The previous theorem highlights that the set of rational numbers are in some sense incomplete, or that there are gaps in $\mathbb{Q}$, and that a larger number system is needed to enlarge the set of math problems that can be analyzed and solved. Although mathematicians in the 1700s were using the real number system and resorting to limiting processes to analyze problems in physics, it was not until the late 1800s that mathematicians gave a rigorous construction of the real number system. Motivated by Theorem 2.1.1, we might be tempted to define the real numbers as the set of solutions of all polynomial equations with integer coefficients. As it turns out, however, this definition of the reals would actually leave out almost *all* the real numbers, including some of our favorites like π and e. In fact, the set of all numbers that are solutions to polynomial equations with rational coefficients is countable!

There are two standard ways to construct the set of real numbers. One standard method to construct $\mathbb{R}$ uses the notion of **Cauchy sequences** of rational numbers and is attributed to Georg Cantor [2]. We will cover Cauchy sequences in Section 3.6 and therefore postpone describing some of the details of the construction until then. The second standard method to construct the reals relies on the notion of a Dedekind cut and is attributed to Richard Dedekind (1831-1916). A **Dedekind cut** is a partition $\{A, B\}$ of $\mathbb{Q}$ such that both A and B are non-empty and

(i) if $b \in A$ and $a < b$ then $a \in A$, and

(ii) for any $a \in A$ there exists $b \in A$ such that $a < b$.

The set of real numbers $\mathbb{R}$ is then defined to be the set of all Dedekind cuts. For example, let $A = \{x \in \mathbb{Q} \mid x^2 < 2 \text{ or } x < 0\}$ and thus

$B = \mathbb{Q} \backslash A$. Then one can show that $\{A, B\}$ is a Dedekind cut of $\mathbb{Q}$ and the idea is that $x = \{A, B\}$ represents the real number x such that $x^2 = 2$, that is, the irrational number $\sqrt{2}$. Having defined $\mathbb{R}$ as the set of Dedekind cuts we then proceed to define all the usual operations of arithmetic and arrive at the familiar model of $\mathbb{R}$ [2]. Additionally, if $x = \{A, B\}$ and $y = \{C, D\}$ then we write that $x \leq y$ if $A \subset C$ and write $x < y$ if $A \subset C$ and $A \neq C$. Refer to [2] for further details.

In this book, we instead adopt the familiar viewpoint that the real numbers $\mathbb{R}$ are in a one-to-one correspondence with the points on an infinite line:

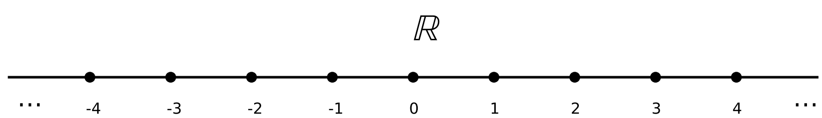


Figure 2.1: The real numbers are in a one-to-one correspondence with points on an infinite line

The essential feature that we want to capture with this view of $\mathbb{R}$ is that there are no "holes" in the real number system. This view of $\mathbb{R}$ allows us to quickly start learning the properties of $\mathbb{R}$ instead of focusing on the details of constructing a model for $\mathbb{R}$. Naturally, the rational numbers $\mathbb{Q}$ are a subset of $\mathbb{R}$ and we say that a real number $x \in \mathbb{R}$ is **irrational** if it is not rational. As we saw in Thereom 2.1.1, the positive number $x \in \mathbb{R}$ such that $x^2 = 2$ is irrational.

Exercises

Exercise 2.1.1. Let $x \in \mathbb{Q}$ be fixed. Prove the following statements without using proof by contradiction.

(a) Prove that if $y \in \mathbb{R}\backslash\mathbb{Q}$ then $x + y \in \mathbb{R}\backslash\mathbb{Q}$.

(b) Suppose in addition that $x > 0$. Prove that if $y \in \mathbb{R}\backslash\mathbb{Q}$ then $xy \in \mathbb{R}\backslash\mathbb{Q}$.

Exercise 2.1.2. Prove that if $0 < x < 1$ then $x^n < x$ for all natural numbers $n \geq 2$. Do not assume that x is rational.

2.2 Algebraic and Order Properties

We will soon see the main difference between $\mathbb{Q}$ and $\mathbb{R}$ from an analysis point of view but in this section we will discuss one important thing that $\mathbb{Q}$ and $\mathbb{R}$ have in common, namely, both are *ordered fields*. We begin then with the definition of a field.

Definition 2.2.1

A **field** is a set $\mathbb{F}$ with two binary operations $+ : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ and $\times : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$, the former called *addition* and the latter *multiplication*, satisfying the following properties:

(i) $a + b = b + a$ for all $a, b \in \mathbb{F}$

(ii) $(a + b) + c = a + (b + c)$ for all $a, b, c \in \mathbb{F}$

(iii) $a \times (b + c) = a \times b + a \times c$

(iv) There exists an element $1 \in \mathbb{F}$ such that $a \times 1 = 1 \times a = a$ for all $a \in \mathbb{F}$

(v) There exists an element $0 \in \mathbb{F}$ such that $a + 0 = 0 + a = a$ for all $a \in \mathbb{F}$

(vi) For each $a \in \mathbb{F}$, there exists an element $-a \in \mathbb{F}$ such that $a + (-a) = (-a) + a = 0$.

(vii) For each $a \in \mathbb{F}$, there exists an element $a^{-1} \in \mathbb{F}$ such that $a \times a^{-1} = a^{-1} \times a = 1$.

Example 2.2.2. It is not hard to see that $\mathbb{N}$ and $\mathbb{Z}$ are not fields. In each case, what property of a field fails to hold?

Example 2.2.3. Both $\mathbb{Q}$ and $\mathbb{R}$ are fields.

Besides being fields, both $\mathbb{Q}$ and $\mathbb{R}$ are **totally ordered** sets. By totally ordered we mean that for any $a, b \in \mathbb{R}$ either $a = b$, $a < b$, or $b < a$. This property of $\mathbb{R}$ is referred to as the Law of Trichotomy. For $a, b \in \mathbb{R}$, the relation $a \leq b$ means that either $a < b$ or $a = b$. Similarly, $a \geq b$ means that $b \leq a$. From our number line viewpoint of $\mathbb{R}$, if $a < b$ then a is on the left of b.

We now present some very important rules of inequalities that we will use frequently in this book.

Theorem 2.2.4

Let $a, b, c \in \mathbb{R}$.

(i) If $a < b$ and $b < c$ then $a < c$. (transitivity)

(ii) If $a < b$ then $a + c < b + c$.

(iii) If $a < b$ and $c > 0$ then $ac < bc$.

(iv) If $a < b$ and $c < 0$ then $ac > bc$.

(v) If $ab > 0$ then either $a, b > 0$ or $a, b < 0$.

(vi) If $a \neq 0$ then $a^2 > 0$.

The two inequalities $a \leq b$ and $b \leq c$ as sometimes combined as

$$a \leq b \leq c.$$

Example 2.2.5. Suppose that $a > 0$ and $b > 0$, or written more compactly as $a, b > 0$. Prove that if $a < b$ then $\frac{1}{b} < \frac{1}{a}$.

Example 2.2.6. Suppose that $a \leq x$ and $b \leq y$. Prove that $a + b \leq x + y$. Deduce that if $a \leq x \leq \xi$ and $b \leq y \leq \zeta$ then

$$a + b \leq x + y \leq \zeta + \xi.$$

We will encounter situations where we will need to prove that if two numbers $a, b \in \mathbb{R}$ satisfy a certain property then $a = b$. Proving that $a = b$ is equivalent to proving that $x = a - b = 0$. In such situations, the following theorem will be very useful.

Theorem 2.2.7

Let $x \in \mathbb{R}$ be non-negative, that is, $x \geq 0$. If for every $\varepsilon > 0$ it holds that $x < \varepsilon$ then $x = 0$.

Proof. We prove the contrapositive, that is, we prove that if $x > 0$ then there exists $\varepsilon > 0$ such that $\varepsilon < x$. Assume then that $x > 0$ and let $\varepsilon = \frac{x}{2}$. Then $\varepsilon > 0$ and clearly $\varepsilon < x$. □

The next few examples will give us practice with working with inequalities.

Example 2.2.8. Let $\varepsilon = 0.0001$. Find a natural number $n \in \mathbb{N}$ such that

$$\frac{1}{n+1} < \varepsilon$$

Example 2.2.9. Let $\varepsilon = 0.0001$. Find analytically a natural number $n \in \mathbb{N}$ such that

$$\frac{n+2}{n^2+3} < \varepsilon$$

Example 2.2.10. Let $\varepsilon = 0.001$. Find analytically a natural number $n \in \mathbb{N}$ such that

$$\frac{5n-4}{2n^3-n} < \varepsilon$$

Exercises

Exercise 2.2.1. Let $\varepsilon = 0.0001$ and find analytically a natural number $n \in \mathbb{N}$ such that

$$\frac{3n-2}{n^3+2n} < \varepsilon.$$

Exercise 2.2.2. Let $\varepsilon = 0.0001$ and find analytically a natural number $n \in \mathbb{N}$ such that

$$\frac{3n+2}{4n^3-n} < \varepsilon.$$

Exercise 2.2.3. Let $\varepsilon = 0.0001$ and find analytically a natural number $n \in \mathbb{N}$ such that

$$\frac{\cos^2(3n)+1}{\arctan(n)+n} < \varepsilon.$$

2.3 The Absolute Value

To solve problems in calculus you need to master differentiation and integration. To solve problems in analysis you need to master inequalities. The content of this section, mostly on inequalities, is fundamental to everything else that follows in this book.

Given any $a \in \mathbb{R}$, we define the **absolute value** of a as the number

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

Clearly, $|a| = 0$ if and only if $a = 0$, and $0 \leq |a|$ for all $a \in \mathbb{R}$. Below we record some important properties of the absolute value function.

Theorem 2.3.1

Let $a, b \in \mathbb{R}$ and $c \geq 0$.

(i) $|ab| = |a| \cdot |b|$

(ii) $|a|^2 = a^2$

(iii) $|a| \leq c$ if and only if $-c \leq a \leq c$

(iv) $-|a| \leq a \leq |a|$

Proof. Statements (i) and (ii) are trivial. To prove (iii), first suppose that $|a| \leq c$. If $a > 0$ then $a \leq c$. Hence, $-a \geq -c$ and since $a > 0$ then $a > -a \geq -c$. Hence, $-c \leq a \leq c$. If $a < 0$ then $-a \leq c$, and thus $a \geq -c$. Since $a < 0$ then $a < -a \leq c$. Thus, $-c \leq a \leq c$. Now suppose that $-c \leq a \leq c$. If $0 < a \leq c$ then $|a| = a \leq c$. If $a < 0$ then from multiplying the inequality by (-1) we have $c \geq -a \geq -c$ and thus $|a| = -a \leq c$.

To prove part (iv), notice that $|a| \leq |a|$ and thus applying (iii) we get $-|a| \leq a \leq |a|$. □

Example 2.3.2. If $b \neq 0$ prove that $\left|\frac{a}{b}\right| = \frac{|a|}{|b|}$.

Example 2.3.3. From Theorem 2.3.1 part (i), we have that $|a^2| = |a \cdot a| = |a| \cdot |a| = |a|^2$. Therefore, $|a^2| = |a|^2$. Similarly, one can show that $|a^3| = |a|^3$. By induction, for each $n \in \mathbb{N}$ it holds that $|a^n| = |a|^n$.

Below is the most important inequality in this book.

Theorem 2.3.4: Triangle Inequality

For any $x, y \in \mathbb{R}$ it holds that

$$|x + y| \leq |x| + |y|.$$

Proof. We have that

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

from which it follows that

$$-(|x| + |y|) \leq x + y \leq |x| + |y|$$

and thus

$$|x + y| \leq |x| + |y|.$$

□

By induction, one can prove the following corollary to the Triangle inequality.

Corollary 2.3.5

For any $x_1, x_2, \ldots, x_n \in \mathbb{R}$ it holds that

$$|x_1 + x_2 + \cdots + x_n| \leq |x_1| + |x_2| + \cdots + |x_n|.$$

A compact way to write the triangle inequality using summation notation is

$$\left|\sum_{i=1}^{n} x_i\right| \leq \sum_{i=1}^{n} |x_i|.$$

Here is another consequence of the Triangle inequality.

Corollary 2.3.6

For $x, y \in \mathbb{R}$ it holds that

(i) $|x - y| \leq |x| + |y|$

(ii) $||x| - |y|| \leq |x - y|$

Proof. For part (i), we have

$$\begin{aligned} |x - y| &= |x + (-y)| \\ &\leq |x| + |-y| \\ &= |x| + |y|. \end{aligned}$$

For part (ii), consider

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

and therefore $|x| - |y| \leq |x - y|$. Switching the role of x and y we obtain $|y| - |x| \leq |y - x| = |x - y|$, and therefore multiplying this last inequality by -1 yields $-|x - y| \leq |x| - |y|$. Therefore,

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

which is the stated inequality. □

Example 2.3.7. For $a, b \in \mathbb{R}$ prove that $|a+b| \geq |a| - |b|$.

Example 2.3.8. Let $f(x) = 2x^2 - 3x + 7$ for $x \in [-2, 2]$. Find a number $M > 0$ such that $|f(x)| \leq M$ for all $-2 \leq x \leq 2$.

Solution. Clearly, if $-2 \leq x \leq 2$ then $|x| \leq 2$. Apply the triangle inequality and the properties of the absolute value:

$$\begin{aligned}
|f(x)| &= |2x^2 - 3x + 7| \\
&\leq |2x^2| + |3x| + |7| \\
&= 2|x|^2 + 3|x| + 7 \\
&\leq 2(2)^2 + 3(2) + 7 \\
&= 21.
\end{aligned}$$

Therefore, if $M = 21$ then $|f(x)| \leq M$ for all $x \in [-2, 2]$. □

Example 2.3.9. Let $f(x) = \frac{2x^2+3x+1}{1-2x}$. Find a number $M > 0$ such that $|f(x)| \leq M$ for all $2 \leq x \leq 3$.

Solution. It is clear that if $2 \leq x \leq 3$ then $|x| \leq 3$. Using the properties of the absolute value and the triangle inequality repeatedly on the numerator:

$$
\begin{aligned}
|f(x)| &= \left|\frac{2x^2+3x+1}{1-2x}\right| \\
&= \frac{|2x^2+3x+1|}{|1-2x|} \\
&\leq \frac{|2x^2|+|3x|+|1|}{|1-2x|} \\
&= \frac{2|x|^2+3|x|+1|}{|1-2x|} \\
&\leq \frac{2\cdot 3^2+3\cdot 3+1}{|1-2x|} \\
&= \frac{28}{|2x-1|}.
\end{aligned}
$$

Now, for $2 \leq x \leq 3$ we have that $-5 \leq 1-2x \leq -3$ and therefore $3 \leq |1-2x| \leq 5$ and then $\frac{1}{|2x-1|} \leq \frac{1}{3}$. Therefore,

$$|f(x)| \leq \frac{28}{|1-2x|} \leq \frac{28}{3}.$$

Hence, we can take $M = \frac{28}{3}$. □

Example 2.3.10. Let $f(x) = \frac{\sin(2x)-3}{x^2+1}$. Find a number $M > 0$ such that $|f(x)| \leq M$ for all $-3 \leq x \leq 2$.

In analysis, the absolute value is used to measure distance between points in $\mathbb{R}$. For any $a \in \mathbb{R}$, the absolute value $|a|$ is the distance from a to 0. This interpretation of the absolute value can be used to measure the difference (in magnitude) between two points. That is,

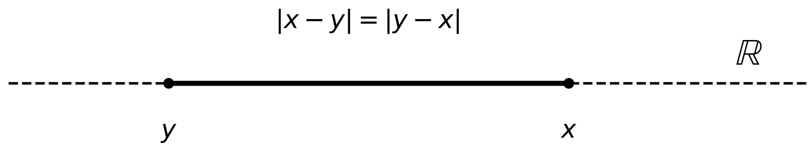


Figure 2.2: The number $|x - y|$ is the distance between x and y.

given $x, y \in \mathbb{R}$, the distance between x and y is $|x - y|$. From the properties of the absolute value, this distance is also $|y - x|$.

We will often be concerned with how close a given number x is to a fixed number $a \in \mathbb{R}$. To do this, we introduce the notion of neighborhoods based at a.

Definition 2.3.11: Neighborhoods

Let $a \in \mathbb{R}$ and let $\varepsilon > 0$. The **ε-neighborhood of a of radius** is the set

$$B_\varepsilon(a) = \{x \in \mathbb{R} \mid |x - a| < \varepsilon\} = (a - \varepsilon, a + \varepsilon)$$

Notice that if $\varepsilon_1 < \varepsilon_2$ then $B_{\varepsilon_1}(a) \subset B_{\varepsilon_2}(a)$.

Example 2.3.12. Let $f(n) = \frac{3n+1}{2n+3}$ and let $\varepsilon = 0.0001$. From calculus, we know that $\lim_{n\to\infty} f(n) = \frac{3}{2}$. Find a natural number N such that $|f(n) - \frac{3}{2}| < \varepsilon$ for every $n \geq N$.

Solution. The inequality $|f(n) - \frac{3}{2}| < \varepsilon$ means that $f(n)$ is within ε of $\frac{3}{2}$. That is,

$$\frac{3}{2} - \varepsilon < f(n) < \frac{3}{2} + \varepsilon.$$

This inequality does not hold for all n, but it will eventually hold for some $N \in \mathbb{N}$ and for all $n \geq N$. For example, $f(1) = \frac{4}{5} = 0.8$ and $|f(1) - \frac{3}{2}| = 0.7 > \varepsilon$, and similarly for $f(2) = \frac{7}{7} = 1$ and $|f(2) - \frac{3}{2}| = 0.5$.

In fact, $|f(20) - \frac{3}{2}| = 0.081 > \varepsilon$. However, because we know that $\lim_{n\to\infty} f(n) = \frac{3}{2}$, eventually $|f(n) - \frac{3}{2}| < \varepsilon$ for large enough n. To find out how large, let's analyze the magnitude $|f(n) - \frac{3}{2}|$:

$$\begin{aligned}
|f(n) - \tfrac{3}{2}| &= \left|\frac{3n+1}{2n+3} - \frac{3}{2}\right| \\
&= \left|\frac{6n+2-6n-9}{2(2n+3)}\right| \\
&= \frac{7}{2(2n+3)}
\end{aligned}$$

Hence, $|f(n) - \frac{3}{2}| < \varepsilon$ if and only if

$$\frac{7}{2(2n+3)} < \varepsilon$$

which after re-arranging can be written as

$$n > \frac{7}{4\varepsilon} - \frac{3}{2}.$$

With $\varepsilon = 0.0001$ we obtain

$$n > 17,498.5.$$

Hence, if $N = 17,499$ then if $n \geq N$ then $|f(n) - \frac{3}{2}| < \varepsilon$. ☐

Example 2.3.13. Let $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$, and let $a \in \mathbb{R}$. Show that $B_{\varepsilon_1}(a) \cap B_{\varepsilon_2}(a)$ and $B_{\varepsilon_1}(a) \cup B_{\varepsilon_2}(a)$ are ε-neighborhoods of a for some appropriate value of ε.

Exercises

Exercise 2.3.1. Prove that if $a < x < b$ and $a < y < b$ then $|x - y| < b - a$. Draw a number line with points a, b, x, y satisfying the inequalities and graphically interpret the inequality $|x - y| < b - a$.

Exercise 2.3.2. Let $a_0, a_1, a_2, \ldots, a_n$ be **positive** real numbers and consider the polynomial

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n.$$

Prove that

$$|f(x)| \leq f(|x|)$$

for all $x \in \mathbb{R}$. HINT: For example, if say $f(x) = 2 + 3x^2 + 7x^3$ then you are asked to prove that

$$\underbrace{|2 + 3x^2 + 7x^3|}_{|f(x)|} \leq \underbrace{2 + 3|x|^2 + 7|x|^3}_{f(|x|)}.$$

However, prove the claim for a general polynomial $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ **with** $a_i > 0$.

Exercise 2.3.3. Let $f(x) = 3x^2 - 7x + 11$ for $x \in [-4, 2]$. Find analytically a number $M > 0$ such that $|f(x)| \leq M$ for all $x \in [-4, 2]$. Do not use calculus to find M.

Exercise 2.3.4. Let $f(x) = \dfrac{x - 1}{x^2 + 7}$ for $x \in [0, 10]$. Find analytically a number $M > 0$ such that $|f(x)| \leq M$ for all $x \in [0, 10]$. Do not use calculus to find M.

Exercise 2.3.5. Let $f(x) = \dfrac{3\cos(\pi x)}{x^2 - 2x + 3}$ for $x \in [0, 2]$. Find analytically a number $M > 0$ such that $|f(x)| \leq M$ for all $x \in [0, 2]$. Do not use calculus to find M. (HINT: Complete the square.)

Exercise 2.3.6. Let $a, b \in \mathbb{R}$ be distinct points. Show that there exists neighborhoods $B_\varepsilon(a)$ and $B_\delta(b)$ such that $B_\epsilon(a) \cap B_\delta(b) \neq \emptyset$.

2.4 The Completeness Axiom

In this section, we introduce the **Completeness Axiom** of $\mathbb{R}$. Recall that an axiom is a statement or proposition that is accepted as true without justification. In mathematics, axioms are the first principles that are accepted as truths and are used to build mathematical theories; in this case real analysis. Roughly speaking, the Completeness Axiom is a way to say that the real numbers have "no gaps" or "no holes", contrary to the case of the rational numbers. As you will see below, the Completeness Axiom is centered around the notions of bounded sets and least upper bounds; let us begin then with some definitions.

Definition 2.4.1: Boundedness

Let $S \subset \mathbb{R}$ be a non-empty set.

(i) We say that S is **bounded above** if there exists $u \in \mathbb{R}$ such that $x \leq u$ for all $x \in S$. We then say that u is an **upper bound** of S.

(ii) We say that S is **bounded below** if there exists $w \in \mathbb{R}$ such that $w \leq x$ for all $x \in S$. We then say that w is a **lower bound** of S.

(iii) We say that S is **bounded** if it is both bounded above and bounded below.

(iv) We say that S is **unbounded** if it is not bounded.

Example 2.4.2. For each case, determine if S is bounded above, bounded below, bounded, or unbounded. If the set is bounded below, determine the set of lower bounds, and similarly if it is bounded above.

(i) $S = [0, 1]$

(ii) $S = (-\infty, 3)$

(iii) $\mathbb{N} = \{1, 2, 3, \ldots, \}$

(iv) $S = \{\frac{1}{x^2+1} \mid -\infty < x < \infty\}$

(v) $S = \{x \in \mathbb{R} \mid x^2 < 2\}$

Example 2.4.3. Let A and B be sets and suppose that $A \subset B$.

(a) Prove that if B is bounded above (below) then A is bounded above (below).

(b) Give an example of sets A and B such that A is bounded below but B is not bounded below.

We now come to an important notion that will be at the root of what we do from now.

Definition 2.4.4: Supremum and Infimum

Let $S \subset \mathbb{R}$ be non-empty.

(i) Let S be bounded above. An upper bound u of S is said to be a **least upper bound** of S if $u \leq u'$ for any upper bound u' of S. In this case we also say that u is a **supremum** of S and write $u = \sup(S)$.

(ii) Let S be bounded below. A lower bound w of S is said to be a **greatest lower bound** of S if $w' \leq w$ for any lower bound w' of S. In this case we also say that w is an **infimum** of S and write $w = \inf(S)$.

It is straightforward to show that a set that is bounded above (bounded below) can have at most one supremum (infimum). At the risk of being repetitive, when it exists, $\sup(S)$ is a number that is an upper bound of S and is the smallest possible upper bound of S. Therefore, $x \leq \sup(S)$ for all $x \in S$, and any number less than $\sup(S)$ is not an upper bound of S (because $\sup(S)$ is the **least** upper bound!). Similarly, when it exists, $\inf(S)$ is a number that is a lower bound of S and is the largest possible lower bound of S. Therefore, $\inf(S) \leq x$ for all $x \in S$ and any number greater than $\inf(S)$ is not a lower bound of S (because $\inf(S)$ is the **greatest** lower bound of S!).

Remark 2.4.5. In some analysis texts, $\sup(S)$ is written as $\text{lub}(S)$ and $\inf(S)$ is written as $\text{glb}(S)$. In other words, $\sup(S) = \text{lub}(S)$ and $\inf(S) = \text{glb}(S)$.

Does every non-empty bounded set S have a supremum/infimum? You might say "Yes, of course!!" and add that "It is a self-evident principle and needs no justification!". Is not that what an axiom is?

Axiom 2.4.6: Completeness Axiom

Every non-empty subset of $\mathbb{R}$ that is bounded above has a least upper bound (a supremum) in $\mathbb{R}$. Similarly, every non-empty subset of $\mathbb{R}$ that is bounded below has a greatest lower bound (an infimum) in $\mathbb{R}$.

As you will see in the pages that follow, The Completeness Axiom is the key notion upon which the theory of real analysis depends on.

Example 2.4.7. Determine $\sup(S)$ and $\inf(S)$, if they exist.

(a) $S = \{-5, -9, 2, -1, 11, 0, 4\}$

(b) $S = [0, \infty)$

(c) $S = (-\infty, 3)$

The Completeness Axiom is sometimes called the **supremum property** of $\mathbb{R}$ or the **least upper bound property** of $\mathbb{R}$. The Completeness property makes $\mathbb{R}$ into a **complete ordered field**. The following example shows that $\mathbb{Q}$ does not have the completeness property.

Example 2.4.8. The set of rational numbers is an ordered field but it is not complete. Consider the set $S = \{x \in \mathbb{Q} \mid x^2 < 2\}$. By definition, $S \subset \mathbb{Q}$. Clearly, S is bounded above, for example $u = 10$ is an upper bound of S, but the least upper bound of S is $u = \sqrt{2}$ which is not a rational number. Therefore $S \subset \mathbb{Q}$ does not have a supremum **in** $\mathbb{Q}$ and therefore $\mathbb{Q}$ does not have the Completeness property. From the point of view of analysis, this is the main distinction between $\mathbb{Q}$ and $\mathbb{R}$ (note that both are ordered fields).

In some cases, it is obvious what sup and inf are, however, to do analysis rigorously, we need systematic ways to determine $\sup(S)$ and $\inf(S)$. To start with, we first need to be a bit more rigorous about what it means to be the least upper bound, or at least have a more concrete description that we can work with, i.e., using inequalities. The following lemma does that and, as you will observe, it is simply a direct consequence of the definition of the supremum.

Lemma 2.4.9

Let $S \subset \mathbb{R}$ be non-empty and suppose that $u \in \mathbb{R}$ is an upper bound of S. Then u is the least upper bound of S if and only if for any $\varepsilon > 0$ there exists $x \in S$ such that

$$u - \varepsilon < x \leq u.$$

Proof. Suppose that u is the supremum of S, that is, u is the least upper bound of S. Since $u - \varepsilon < u$ then $u - \varepsilon$ is not an upper bound of S. Thus, there exists $x \in S$ such that $u - \varepsilon < x$.

Now suppose that for any $\varepsilon > 0$ there exists $x \in S$ such that $u - \varepsilon < x \leq u$. Now let $v \in \mathbb{R}$ be such that $v < u$. Then there exists $\varepsilon > 0$ such that $v = u - \varepsilon$, and thus by assumption, there exists $x \in S$ such that $v < x$. Hence, v is not an upper bound of S and this shows that u is the least upper bound of S, that is, $u = \sup(S)$. □

Example 2.4.10. If $A \subset B \subset \mathbb{R}$, and B is bounded above, prove that A is bounded above and that $\sup(A) \leq \sup(B)$.

Solution. Since B is bounded above, $\sup(B)$ exists by the Completeness property of $\mathbb{R}$. Let $x \in A$. Then $x \in B$ and therefore $x \leq \sup(B)$. This proves that A is bounded above by $\sup(B)$ and therefore $\sup(A)$ exists. Since $\sup(A)$ is the least upper bound of A we must have $\sup(A) \leq \sup(B)$. For example, if say $B = [1, 3]$ and $A = [1, 2]$ then $\sup(A) < \sup(B)$, while if $A = [2, 3]$ then $\sup(A) = \sup(B)$. □

Example 2.4.11. Let $A \subset \mathbb{R}$ be non-empty and bounded above. Let $c \in \mathbb{R}$ and define the set

$$cA = \{y \in \mathbb{R} \mid \exists\, x \in A \text{ s.t. } y = cx\}.$$

Prove that if $c > 0$ then cA is bounded above and $\sup(cA) = c\sup(A)$. Show by example that if $c < 0$ then cA need not be bounded above even if A is bounded above.

Proof. Let $y \in cA$ be arbitrary. Then there exists $x \in A$ such that $y = cx$. By the Completeness property, $\sup(A)$ exists and $x \leq \sup(A)$. If $c > 0$ then $cx \leq c\sup(A)$ which is equivalent to $y \leq c\sup(A)$. Since y is arbitrary, this shows that $c\sup(A)$ is an upper bound of the set

cA and thus cA is bounded above and thus $\sup(cA)$ exists. Because $\sup(cA)$ is the least upper bound of cA then

$$\sup(cA) \le c\sup(A). \tag{2.1}$$

Now, by definition, $y \le \sup(cA)$ for all $y \in cA$. Thus, $cx \le \sup(cA)$ for all $x \in A$ and therefore $x \le \frac{1}{c}\sup(cA)$ for all $x \in A$. Therefore, $\frac{1}{c}\sup(cA)$ is an upper bound of A and consequently $\sup(A) \le \frac{1}{c}\sup(cA)$ because $\sup(A)$ is the least upper bound of A. We have therefore proved that

$$c\sup(A) \le \sup(cA). \tag{2.2}$$

Combining (2.1) and (2.2) we conclude that

$$\sup(cA) = c\sup(A).$$

□

Example 2.4.12. Suppose that A and B are non-empty and bounded above. Prove that $A \cup B$ is bounded above and that

$$\sup(A \cup B) = \sup\{\sup(A), \sup(B)\}.$$

Proof. Let $u = \sup\{\sup(A), \sup(B)\}$. Then clearly $\sup(A) \le u$ and $\sup(B) \le u$. We first show that $A \cup B$ is bounded above by showing that u is an upper bound of $A \cup B$. Let $x \in A \cup B$. Then either $z \in A$ or $x \in B$ (or both). If $z \in A$ then $z \le \sup(A) \le u$ and if $z \in B$ then $z \le \sup(B) \le u$. Hence, $A \cup B$ is bounded above and u is an upper bound of $A \cup B$. Consequently, $\sup(A \cup B)$ exists by the Completeness axiom and moreover $\sup(A \cup B) \le u$, that is,

$$\sup(A \cup B) \le \sup\{\sup(A), \sup(B)\}. \tag{2.3}$$

Now, by definition of the supremum, $z \leq \sup(A \cup B)$ for all $z \in A \cup B$. Now since $A \subset A \cup B$ this implies that $x \leq \sup(A \cup B)$ for all $x \in A$ and $y \leq \sup(A \cup B)$ for all $y \in B$. In other words, $\sup(A \cup B)$ is an upper bound of both A and B and thus $\sup(A) \leq \sup(A \cup B)$ and $\sup(B) \leq \sup(A \cup B)$. Then clearly

$$\sup\{\sup(A), \sup(B)\} \leq \sup(A \cup B) \tag{2.4}$$

and combining (2.3) and (2.4) we have proved that $\sup(A \cup B) = \sup\{\sup(A), \sup(B)\}$. □

Example 2.4.13. Suppose that A and B are non-empty and bounded below, and suppose that $A \cap B$ is non-empty. Prove that $A \cap B$ is bounded below and that

$$\sup\{\inf(A), \inf(B)\} \leq \inf(A \cap B).$$

Proof. If $x \in A \cap B$ then $x \in A$ and therefore $\inf(A) \leq x$, and also $x \in B$ and thus $\inf(B) \leq x$. Therefore, both $\inf(A)$ and $\inf(B)$ are lower bounds of $A \cap B$, and by definition of $\inf(A \cap B)$ we have that $\inf(A) \leq \inf(A \cap B)$ and $\inf(B) \leq \inf(A \cap B)$, and consequently $\sup\{\inf(A), \inf(B)\} \leq \inf(A \cap B)$. □

Example 2.4.14. For any set A define the set

$$-A = \{y \in \mathbb{R} \mid \exists\, x \in A \text{ s.t. } y = -x\}.$$

Prove that if $A \subset \mathbb{R}$ is non-empty and bounded then $\sup(-A) = -\inf(A)$.

Proof. It holds that $\inf(A) \leq x$ for all $x \in A$ and therefore, $-\inf(A) \geq -x$ for all $x \in A$, which is equivalent to $-\inf(A) \geq y$ for all $y \in -A$. Therefore, $-\inf(A)$ is an upper bound of the set $-A$ and therefore

$\sup(-A) \leq -\inf(A)$. Now, $y \leq \sup(-A)$ for all $y \in -A$, or equivalently $-x \leq \sup(-A)$ for all $x \in A$. Hence, $x \geq -\sup(-A)$ for all $x \in A$. This proves that $-\sup(-A)$ is a lower bound of A and therefore $-\sup(-A) \leq \inf(A)$, or $\sup(-A) \geq -\inf(A)$. This proves that $\sup(-A) = -\inf(A)$. □

Example 2.4.15. Let $A \subset \mathbb{R}$ be non-empty and bounded below. Let $c \in \mathbb{R}$ and define the set

$$c + A = \{y \in \mathbb{R} \mid \exists\, x \in A \text{ s.t. } y = c + x\}.$$

Prove that $c + A$ is bounded below and that $\inf(c + A) = c + \inf(A)$.

Proof. For all $x \in A$ it holds that $\inf(A) \leq x$ and therefore $c+\inf(A) \leq c + x$. This proves that $c + \inf(A)$ is a lower bound of the set $c + A$, and therefore $c + \inf(A) \leq \inf(c + A)$. Now, $\inf(c + A) \leq y$ for all $y \in c + A$ and thus $\inf(c + A) \leq c + x$ for all $x \in A$, which is the same as $\inf(c + A) - c \leq x$ for all $x \in A$. Hence, $\inf(c + A) - c \leq \inf(A)$ or equivalently $\inf(c + A) \leq c + \inf(A)$. This proves the claim. □

Example 2.4.16. Let A and B be non-empty subsets of $\mathbb{R}_{>0} = \{x \in \mathbb{R} : x > 0\}$, and suppose that A and B are bounded below. Define the set $AB = \{xy : x \in A,\, y \in B\}$.

(a) Prove that AB is bounded below.

(b) Prove that $\inf(AB) = \inf(A) \cdot \inf(B)$. Hint: Consider two cases, when say $\inf(A)\inf(B) = 0$ and when $\inf(A)\inf(B) \neq 0$.

(c) How do things change if we do not assume A and B are subsets of $\mathbb{R}_{>0}$.

Example 2.4.17. Give an example of a non-empty set $A \subset \mathbb{R}_{>0}$ such that $\inf(A) = 0$.

Example 2.4.18. For any two non-empty sets P and Q of $\mathbb{R}$ let us write that $P \leq Q$ if, for each $x \in P$, there exists $y \in Q$ such that $x \leq y$.

(a) Prove that if $P \leq Q$ then $\sup(P) \leq \sup(Q)$.

(b) Show via an example that if $P \leq Q$ then it is not necessarily true that $\inf(P) \leq \inf(Q)$.

Example 2.4.19. Let A and B be non-empty bounded sets of positive real numbers. Define the set

$$\frac{A}{B} = \left\{ z \in \mathbb{R} \mid \exists\, x \in A, \exists\, y \in B \text{ s.t. } z = \tfrac{x}{y} \right\}.$$

Assume that $\inf(B) > 0$. Prove that

$$\sup\left(\tfrac{A}{B}\right) = \frac{\sup(A)}{\inf(B)}.$$

Proof. Since A is bounded above, $\sup(A)$ exists and $x \leq \sup(A)$ for all $x \in A$. Since B is bounded below, $\inf(B)$ exists and $\inf(B) \leq y$ for all $y \in B$. Let $z = x/y$ be an arbitrary point in $\frac{A}{B}$. Then since $\inf(B) \leq y$ and $y > 0$ and $\inf(B) > 0$ we obtain that

$$\frac{1}{y} \leq \frac{1}{\inf(B)}.$$

Since $x \leq \sup(A)$ and $x > 0$ (and then clearly $\sup(A) > 0$) we obtain

$$\frac{x}{y} \leq \frac{x}{\inf(B)} \leq \frac{\sup(A)}{\inf(B)}.$$

This proves that $z \leq \frac{\sup(A)}{\inf(B)}$ and since $z \in \frac{A}{B}$ was arbitrary we have proved that $\frac{\sup(A)}{\inf(B)}$ is an upper bound of the set $\frac{A}{B}$. This proves that

$\sup(\frac{A}{B})$ exists. Moreover, by the definition of the supremum, we also have that

$$\sup\left(\tfrac{A}{B}\right) \leq \frac{\sup(A)}{\inf(B)}. \tag{2.5}$$

Now, $z \leq \sup(\frac{A}{B})$ for all $z \in \frac{A}{B}$ and thus $\frac{x}{y} \leq \sup(\frac{A}{B})$ for all $x \in A$ and all $y \in B$. If y is held fixed then $x \leq y \cdot \sup(\frac{A}{B})$ for all $x \in A$ and thus $\sup(A) \leq y \cdot \sup(\frac{A}{B})$. Therefore, $\sup(A)/\sup(\frac{A}{B}) \leq y$, which holds for all $y \in B$. Therefore, $\sup(A)/\sup(\frac{A}{B}) \leq \inf(B)$ and consequently

$$\frac{\sup(A)}{\inf(B)} \leq \sup\left(\tfrac{A}{B}\right). \tag{2.6}$$

Combining (2.5) and (2.6) completes the proof. □

Exercises

Exercise 2.4.1. Let $S \subset \mathbb{R}$ be a non-empty set and suppose that u is an upper bound of S. Prove that if $u \in S$ then necessarily $u = \sup S$.

Exercise 2.4.2. Let A and B be non-empty subsets of $\mathbb{R}$, and suppose that $A \subset B$. Prove that if B is bounded below then $\inf B \leq \inf A$.

Exercise 2.4.3. If P and Q are non-empty subsets of $\mathbb{R}$ such that $\sup P = \sup Q$ and $\inf P = \inf Q$ does it follow that $P = Q$? Support your answer with either a proof or a counterexample.

Exercise 2.4.4. Let $A \subset \mathbb{R}$ be a bounded set, let $x \in \mathbb{R}$ be fixed, and define the set

$$x + A = \{y \in \mathbb{R} \mid \exists\, a \in A \text{ s.t. } y = x + a\}.$$

Prove that $\sup(x + A) = x + \sup A$.

Exercise 2.4.5. Let $A, B \subset \mathbb{R}$ be bounded above. Let

$$A + B = \{z \in \mathbb{R} \mid \exists\, x \in A,\ \exists\, y \in B \text{ s.t. } z = x + y\}.$$

Prove that $A+B$ is bounded above and that $\sup(A+B) = \sup A + \sup B$.

Exercise 2.4.6. Let $\mathbb{R}_{>0}$ denote the set of all positive real numbers and let $A, B \subset \mathbb{R}_{>0}$ be bounded. Assume that $\inf(B) > 0$. Define the set

$$\frac{A}{B} = \left\{z \in \mathbb{R} \mid \exists\, x \in A,\ \exists\, y \in B \text{ s.t. } z = \tfrac{x}{y}\right\}.$$

Prove that

$$\sup\left(\tfrac{A}{B}\right) = \frac{\sup(A)}{\inf(B)}.$$

2.5 Applications of the Supremum

Having defined the notion of boundedness for a set, we can define a notion of boundedness for a function.

Definition 2.5.1: Bounded Functions

Let $D \subset \mathbb{R}$ be a non-empty set and let $f : D \to \mathbb{R}$ be a function.

(i) f is **bounded below** if the range $f(D) = \{f(x) \mid x \in D\}$ is bounded below.

(ii) f is **bounded above** if the range $f(D) = \{f(x) \mid x \in D\}$ is bounded above.

(iii) f is **bounded** if the range $f(D) = \{f(x) \mid x \in D\}$ is bounded.

Hence, boundedness of a function f is boundedness of its range.

Example 2.5.2. Suppose $f, g : D \to \mathbb{R}$ are bounded functions and $f(x) \leq g(x)$ for all $x \in D$. Show that $\sup(f(D)) \leq \sup(g(D))$.

Solution. Since g is bounded, $\sup(g(D))$ exists and is by definition an upper bound for the set $g(D)$, that is, $g(x) \leq \sup(g(D))$ for all $x \in D$. Now, by assumption, for all $x \in D$ it holds that $f(x) \leq g(x)$ and therefore $f(x) \leq \sup(g(D))$. This shows that $\sup(g(D))$ is an upper bound of the set $f(D)$, and therefore by definition of the supremum, $\sup(f(D)) \leq \sup(g(D))$. □

Example 2.5.3. Let $f, g : D \to \mathbb{R}$ be bounded functions and suppose that $f(x) \leq g(y)$ for all $x, y \in D$. Show that $\sup(f(D)) \leq \inf(g(D))$.

Solution. Fix $x^* \in D$. Then, $f(x^*) \leq g(y)$ for all $y \in D$. Therefore, $f(x^*)$ is a lower bound of $g(D)$, and thus $f(x^*) \leq \inf(g(D))$ by definition of the infimum. Since $x^* \in D$ was arbitrary, we have proved that $f(x) \leq \inf(g(D))$ for all $x \in D$. Hence, $\inf(g(D))$ is an upper bound of $f(D)$ and thus by definition of the supremum we have that $\sup(f(D)) \leq \inf(g(D))$. □

The following sequence of results will be used to prove an important property of the rational numbers $\mathbb{Q}$ as seen from within $\mathbb{R}$.

Theorem 2.5.4: Archimedean Property

If $x \in \mathbb{R}$ then there exists $n \in \mathbb{N}$ such that $x \leq n$.

Proof. Suppose not. Hence, $n \leq x$ for all $n \in \mathbb{N}$, and thus x is an upper bound for $\mathbb{N}$, and therefore $\mathbb{N}$ is bounded. Let $u = \sup(\mathbb{N})$. By definition of u, $u - 1$ is not an upper bound of $\mathbb{N}$ and therefore there exists $m \in \mathbb{N}$ such that $u - 1 < m$. But then $u < m + 1$ and this contradicts the definition of u. □

Corollary 2.5.5

If $S = \{\frac{1}{n} \mid n \in \mathbb{N}\}$ then $\inf(S) = 0$.

Proof. Since $0 < \frac{1}{n}$ for all n then 0 is a lower bound of S. Now suppose that $0 < y$. By the Archimedean Property, there exists $n \in \mathbb{N}$ such that $\frac{1}{y} < n$ and thus $\frac{1}{n} < y$. Hence, y is not a lower bound of S. Therefore 0 is the greatest lower bound of S, that is, $\inf(S) = 0$. □

Corollary 2.5.6

For any $y > 0$ there exists $n \in \mathbb{N}$ such that $\frac{1}{n} < y$.

Corollary 2.5.7

Given $y > 0$ there exists $n \in \mathbb{N}$ such that $n - 1 \leq y < n$.

Proof. Let $E = \{k \in \mathbb{R} \mid y < k\}$. By the Archimedean property, E is non-empty. By the Well-Ordering Principle of $\mathbb{N}$, E has a least element, say that it is m. Hence, $m - 1 \notin E$ and thus $m - 1 \leq y < m$. □

We now come to an important result that we will use frequently.

Theorem 2.5.8: Density of the Rationals

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

Proof. We first prove the claim for the case that $0 < x < y$. Suppose that $y - x > 1$ and thus $x + 1 < y$. There exists $m \in \mathbb{N}$ such that $m - 1 \leq x < m$ and thus $m \leq x + 1$. Therefore,

$$x < m \leq x + 1 < y$$

and thus $x < m < y$ and we may take $r = m$. In general, if $y - x > 0$ then there exists $n \in \mathbb{N}$ such that $\frac{1}{n} < y - x$ and thus $1 + nx < ny$. Since $ny - nx > 1$ there exists $m \in \mathbb{N}$ such that $1 + nx < m < ny$ and thus

$$nx < 1 + nx < m < ny$$

and dividing by n yields $x < \frac{m}{n} < y$ and thus we may take $r = \frac{m}{n}$. This proves the claim when both x and y are positive. If $x < 0 < y$

then take $r = 0$ and if $x < y < 0$ then apply the previous arguments to $0 < -y < -x$. □

Hence, between any two distinct real numbers there is a rational number. This implies that any irrational number can be approximated by a rational number to within any degree of accuracy.

Example 2.5.9. Let $\zeta \in \mathbb{R}\backslash\mathbb{Q}$ be an irrational number and let $\varepsilon > 0$ be arbitrary. Prove that there exists a rational number $x \in \mathbb{Q}$ such that $\zeta - \varepsilon < x < \zeta + \varepsilon$, that is, x is in the ε-neighborhood of ζ.

Corollary 2.5.10: Density of the Irrationals

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

Proof. We have that $\sqrt{2}x < \sqrt{2}y$. By the Density Theorem, $\sqrt{2}x < r < \sqrt{2}y$ for some $r \in \mathbb{Q}$. Then $\xi = r/\sqrt{2}$. □

Exercises

Exercise 2.5.1. Let $D \subset \mathbb{R}$ be non-empty and let $f, g : D \to \mathbb{R}$ be functions. Let $f + g$ denote the function defined by

$$(f + g)(x) = f(x) + g(x)$$

for any $x \in D$. If $f(D)$ and $g(D)$ are bounded above prove that $(f + g)(D)$ is also bounded above and that

$$\sup(f + g)(D) \leq \sup f(D) + \sup g(D)$$

Exercise 2.5.2. If $y > 0$ prove that there exists $n \in \mathbb{N}$ such that $\frac{1}{2^n} < y$. (**Note:** If you begin with $\frac{1}{2^n} < y$ and solve for n then you are assuming that such an n exists. You are not asked to *find* an n, you are asked to prove that such an n *exists*.)

2.6 Nested Interval Theorem

If $a, b \in \mathbb{R}$ and $a < b$ define

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

The set (a, b) is called an **open interval** from a to b. Define also

$$[a, b] := \{x \in \mathbb{R} \mid a \leq x \leq b\}$$

which we call the **closed interval** from a to b. The following are called **half-open** (or **half-closed**) intervals:

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

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

If $a = b$ then $(a, a) = \emptyset$ and $[a, a] = \{a\}$. Infinite intervals are

$$\begin{aligned}
(a, \infty) &= \{x \in \mathbb{R} \mid x > a\} \\
(-\infty, b) &= \{x \in \mathbb{R} \mid x < b\} \\
[a, \infty) &= \{x \in \mathbb{R} \mid x \geq a\} \\
(-\infty, b] &= \{x \in \mathbb{R} \mid x \leq b\}.
\end{aligned}$$

Below is a characterization of intervals, we will omit the proof.

Theorem 2.6.1

Let $S \subset \mathbb{R}$ contain at least two points. Suppose that if $x, y \in S$ and $x < y$ then $[x, y] \subset S$. Then S is an interval.

A sequence $I_1, I_2, I_3, I_4, \ldots$ of intervals is **nested** if

$$I_1 \supseteq I_2 \supseteq I_3 \supseteq I_4 \supseteq \cdots$$

As an example, consider $I_n = [0, \frac{1}{n}]$ where $n \in \mathbb{N}$. Then $I_1, I_2, I_3, \ldots$ is nested:

$$[0,1] \supseteq [0, \tfrac{1}{2}] \supseteq [0, \tfrac{1}{3}] \supseteq [0, \tfrac{1}{4}] \cdots$$

Notice that since $0 \in I_n$ for each $n \in \mathbb{N}$ then

$$0 \in \bigcap_{n=1}^{\infty} I_n.$$

Is there another point in $\bigcap_{n=1}^{\infty} I_n$? Suppose that $x \neq 0$ and $x \in \bigcap_{n=1}^{\infty} I_n$. Then necessarily $x > 0$. Then there exists $m \in \mathbb{N}$ such that $\frac{1}{m} < x$. Thus, $x \notin [0, \frac{1}{m}]$ and therefore $x \notin \bigcap_{n=1}^{\infty} I_n$. Therefore,

$$\bigcap_{n=1}^{\infty} I_n = \{0\}.$$

In general, we have the following.

Theorem 2.6.2: Nested Interval Property

Let $I_1, I_2, I_3, \ldots$ be a sequence of nested closed bounded intervals. Then there exists $\xi \in \mathbb{R}$ such that $\xi \in I_n$ for all $n \in \mathbb{N}$, that is, $\xi \in \bigcap_{n=1}^{\infty} I_n$. In particular, if $I_n = [a_n, b_n]$ for $n \in \mathbb{N}$, and $a = \sup\{a_1, a_2, a_3, \ldots\}$ and $b = \inf\{b_1, b_2, b_3, \ldots\}$ then

$$\{a, b\} \subset \bigcap_{n=1}^{\infty} I_n.$$

Proof. Since I_n is a closed interval, we can write $I_n = [a_n, b_n]$ for some $a_n, b_n \in \mathbb{R}$ and $a_n \leq b_n$ for all $n \in \mathbb{N}$. The nested property can be written as

$$[a_1, b_1] \supseteq [a_2, b_2] \supseteq [a_3, b_3] \supseteq [a_4, b_4] \supseteq \cdots$$

Since $[a_n, b_n] \subseteq [a_1, b_1]$ for all $n \in \mathbb{N}$ then $a_n \leq b_1$ for all $n \in \mathbb{N}$. Therefore, the set $S = \{a_n \mid n \in \mathbb{N}\}$ is bounded above. Let $\xi = \sup(S)$

and thus $a_n \leq \xi$ for all $n \in \mathbb{N}$. We will show that $\xi \leq b_n$ for all $n \in \mathbb{N}$ also. Let $n \in \mathbb{N}$ be arbitrary. If $k \leq n$ then $[a_n, b_n] \subseteq [a_k, b_k]$ and therefore $a_k \leq a_n \leq b_n$. On the other hand, if $n < k$ then $[a_k, b_k] \subset [a_n, b_n]$ and therefore $a_n \leq a_k \leq b_n$. In any case, $a_k \leq b_n$ for all $k \in \mathbb{N}$. Hence, b_n is an upper bound of S, and thus $\xi \leq b_n$. Since $n \in \mathbb{N}$ was arbitrary, we have that $\xi \leq b_n$ for all $n \in \mathbb{N}$. Therefore, $a_n \leq \xi \leq b_n$ for all $n \in \mathbb{N}$, that is $\xi \in \bigcap_{n=1}^{\infty}[a_n, b_n]$. The proof that $\inf\{b_1, b_2, b_3, \ldots\} \in \bigcap_{n=1}^{\infty} I_n$ is similar. □

The following theorem gives a condition when $\bigcap_{n=1}^{\infty} I_n$ contains a single point.

Theorem 2.6.3

Let $I_n = [a_n, b_n]$ be a sequence of nested closed bounded intervals. If

$$\inf\{b_n - a_n \mid n \in \mathbb{N}\} = 0$$

then $\bigcap_{n=1}^{n} I_n$ is a singleton set.

Example 2.6.4. Let $I_n = \left[1 - \frac{1}{n}, 1 + \frac{1}{n}\right]$ for $n \in \mathbb{N}$.

(a) Prove that $I_1, I_2, I_3, \ldots$ is a sequence of nested intervals.

(b) Find $\bigcap_{n=1}^{\infty} I_n$.

Using the Nested Interval property of $\mathbb{R}$, we give a proof that $\mathbb{R}$ is uncountable

Theorem 2.6.5: Reals Uncountable

The real numbers $\mathbb{R}$ are uncountable.

Proof. We will prove that the interval $[0, 1]$ is uncountable. Suppose by contradiction that $I = [0, 1]$ is countable, and let $I = \{x_1, x_2, x_3, \ldots, \}$

be an enumeration of I (formally this means we have a bijection $f : \mathbb{N} \to [0,1]$ and $f(n) = x_n$). Since $x_1 \in I = [0,1]$, there exists a closed and bounded interval $I_1 \subset [0,1]$ such that $x_1 \notin I_1$. Next, consider x_2. There exists a closed and bounded interval $I_2 \subset I_1$ such that $x_2 \notin I_2$. Next, consider x_3. There exists a closed and bounded interval $I_3 \subset I_2$ such that $x_3 \notin I_3$. By induction, there exists a sequence $I_1, I_2, I_3, \ldots$ of closed and bounded intervals such that $x_n \notin I_n$ for all $n \in \mathbb{N}$. Moreover, by construction the sequence I_n is nested and therefore $\bigcap_{n=1}^{\infty} I_n$ is non-empty, say it contains ξ. Clearly, since $\xi \in I_n \subset [0,1]$ for all $n \in \mathbb{N}$ then $\xi \in [0,1]$. Now, since $x_n \notin I_n$ for each $n \in \mathbb{N}$ then $x_n \notin \bigcap_{n=1}^{\infty} I_n$. Therefore, $\xi \neq x_n$ for all $n \in \mathbb{N}$ and thus $\xi \notin I = \{x_1, x_2, \ldots\} = [0,1]$, which is a contradiction since $\xi \in [0,1]$. Therefore, $[0,1]$ is uncountable and this implies that $\mathbb{R}$ is also uncountable. □

We now give an alternative proof that $\mathbb{R}$ is uncountable. To that end, consider the following subset $S \subset \mathbb{R}$:

$$S = \{0.a_1a_2a_3a_4 \cdots \in \mathbb{R} \mid a_k = 0 \text{ or } a_k = 1,\ k \in \mathbb{N}\}.$$

In other words, S consists of numbers $x \in [0,1)$ whose decimal expansion consists of only 0's and 1's. For example, some elements of S are

$$x = 0.000000\ldots$$
$$x = 0.101010\ldots$$
$$x = 0.100000\ldots$$
$$x = 0.010100\ldots$$

If we can construct a bijection $f : S \to \mathcal{P}(\mathbb{N})$, then since $\mathcal{P}(\mathbb{N})$ is uncountable then by Example 1.4.7 this would show that S is uncountable. Since $S \subset \mathbb{R}$ then this would show that $\mathbb{R}$ is uncountable (by

Theorem 1.4.8). To construct f, given $x = 0.a_1a_2a_3\ldots$ in S define $f(x) \in \mathcal{P}(\mathbb{N})$ as

$$f(0.a_1a_2a_3\cdots) = \{k \in \mathbb{N} \mid a_k = 1\}.$$

In other words, $f(x)$ consists of the decimal places in the decimal expansion of x that have a value of 1. For example,

$$\begin{aligned} f(0.000000\ldots) &= \emptyset \\ f(0.101010\ldots) &= \{1, 3, 5, 7, \ldots\} \\ f(0.100100\ldots) &= \{1, 4\} \\ f(0.011000\ldots) &= \{2, 3\} \end{aligned}$$

It is left as an exercise to show that f is a bijection (see Exercise 1.4.4).

Exercises

Exercise 2.6.1. Let $I_n = \left[0, \frac{1}{n}\right]$ for $n \in \mathbb{N}$. Prove that $\bigcap_{n=1}^{\infty} I_n = \{0\}$.

3

Sequences

In the tool box used to build analysis, if the Completeness property of the real numbers is the hammer then sequences are the nails. Almost everything that can be said in analysis can be, and is, done using sequences. For this reason, the study of sequences will occupy us for the next foreseeable future.

3.1 Limits of Sequences

A **sequence** of real numbers is a function $X : \mathbb{N} \to \mathbb{R}$. Informally, the sequence X can be written as an infinite list of real numbers as $X = (x_1, x_2, x_3, \ldots)$, where $x_n = X(n)$. Other notations for sequences are (x_n) or $\{x_n\}_{n=1}^{\infty}$; we will use (x_n).

Some sequences can be written explicitly with a formula such as $x_n = \frac{1}{n}$, $x_n = \frac{1}{2^n}$, or

$$x_n = (-1)^n \cos(n^2 + 1),$$

or we could be given the first few terms of the sequence, such as

$$X = (3, 3.1, 3.14, 3.141, 3.1415, \ldots).$$

Some sequences may be given **recursively**. For example,

$$x_1 = 1, \qquad x_{n+1} = \frac{x_n}{n+1}, \quad n \geq 1.$$

Using the definition of x_{n+1} and the initial value x_1 we can in principle find all the terms:

$$x_2 = \frac{1}{2}, \quad x_3 = \frac{1/2}{3}, \quad x_4 = \frac{1/6}{4}, \ \ldots$$

A famous sequence given recursively is the **Fibonacci** sequence which is defined as $x_1 = 1$, $x_2 = 1$, and

$$x_{n+1} = x_{n-1} + x_n, \quad n \geq 2.$$

Then

$$x_3 = 2, \quad x_4 = 3, \quad x_5 = 5, \ldots$$

The **range** of a sequence (x_n) is the set

$$\{x_n \mid n \in \mathbb{N}\},$$

that is, the usual range of a function. However, the range of a sequence is not the actual sequence (the range is a set and a sequence is a function). For example, if $X = (1, 2, 3, 1, 2, 3, \ldots)$ then the range of X is $\{1, 2, 3\}$. If $x_n = \sin(\frac{n\pi}{2})$ then the range of (x_n) is $\{1, 0, -1\}$.

Many concepts in analysis can be described using the long-term or limiting behavior of sequences. In calculus, you undoubtedly developed techniques to compute the limit of basic sequences (and hence show convergence) but you might have omitted the rigorous definition of the convergence of a sequence. Perhaps you were told that a given sequence (x_n) converges to L if as $n \to \infty$ the values x_n get closer to L. Although this is intuitively sound, we need a more precise way to describe the meaning of the convergence of a sequence. Before we give the precise definition, we will consider an example.

Example 3.1.1. Consider the sequence (x_n) whose nth term is given by $x_n = \frac{3n+2}{n+1}$. The values of (x_n) for several values of n are displayed in Table 3.1.

n	x_n
1	2.50000000
2	2.66666667
3	2.75000000
4	2.80000000
5	2.83333333
50	2.98039216
101	2.99019608
10,000	2.99990001
1,000,000	2.99999900
2,000,000	2.99999950

Table 3.1: Values of the sequence $x_n = \frac{3n+2}{n+1}$

The above data suggests that the values of the sequence (x_n) become closer and closer to the number $L = 3$. For example, suppose that $\varepsilon = 0.005$ and consider the ε-neighborhood of $L = 3$, that is, the interval $(3-\varepsilon, 3+\varepsilon) = (2.995, 3.005)$. Not all the terms of the sequence (x_n) are in the ε-neighborhood, however, it seems that all the terms of the sequence from x_{101} and **onward** are inside the ε-neighborhood. In other words, **IF** $n \geq 101$ then $3 - \varepsilon < x_n < 3 + \varepsilon$, or equivalently $|x_n - 3| < \varepsilon$. Suppose now that $\varepsilon = 0.00001$ and thus the new ε-neighborhood is $(3-\varepsilon, 3+\varepsilon) = (2.99999, 3.000001)$. Then it is no longer true that $|x_n - 3| < \varepsilon$ for all $n \geq 101$. However, it seems that all the terms of the sequence from $x_{1000000}$ and onward are inside the smaller ε-neighborhood, in other words, $|x_n - 3| < \varepsilon$ for all $n \geq 1,000,000$. We can extrapolate these findings and make the following hypothesis: For any given $\varepsilon > 0$ there exists a natural number $K \in \mathbb{N}$ such that if $n \geq K$ then $|x_n - L| < \varepsilon$.

The above example and our analysis motivates the following definition.

Definition 3.1.2: Convergence of Sequences

The sequence (x_n) is said to **converge** if there exists a number $L \in \mathbb{R}$ such that for any given $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that $|x_n - L| < \varepsilon$ for all $n \geq K$. In this case, we say that (x_n) has limit L and we write

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

If (x_n) is not convergent then we say that it is **divergent**.

Hence, x_n converges to L if for any given $\varepsilon > 0$ (no matter how small), there exists a point in the sequence x_K such that $|x_K - L| < \varepsilon$, $|x_{K+1} - L| < \varepsilon$, $|x_{K+2} - L| < \varepsilon, \ldots$, that is, $|x_n - L| < \varepsilon$ for all $n \geq K$. We will sometimes write $\lim\limits_{n\to\infty} x_n = L$ simply as $\lim x_n = L$ or $(x_n) \to L$.

Example 3.1.3. Using the definition of the limit of a sequence, prove that $\lim_{n\to\infty} \frac{1}{n} = 0$.

Proof. Let $\varepsilon > 0$ be arbitrary but fixed. By the Archimedean property of $\mathbb{R}$, there exists $K \in \mathbb{N}$ such that $\frac{1}{K} < \varepsilon$. Then, if $n \geq K$ then $\frac{1}{n} \leq \frac{1}{K} < \varepsilon$. Therefore, if $n \geq K$ then

$$\begin{aligned} |x_n - 0| &= \left|\tfrac{1}{n} - 0\right| \\ &= \frac{1}{n} \\ &\leq \frac{1}{K} \\ &< \varepsilon. \end{aligned}$$

This proves, by definition, that $\lim_{n\to\infty} \frac{1}{n} = 0$. □

Example 3.1.4. Using the definition of the limit of a sequence, prove that $\lim_{n\to\infty} \frac{3n+2}{n+1} = 3$.

Proof. Given an arbitrary $\varepsilon > 0$, we want to prove that there exists $K \in \mathbb{N}$ such that

$$\left|\frac{3n+2}{n+1} - 3\right| < \varepsilon, \quad \forall\, n \geq K.$$

Start by analyzing $|x_n - L|$:

$$\begin{aligned} |x_n - L| &= \left|\frac{3n+2}{n+1} - 3\right| \\ &= \left|\frac{-1}{n+1}\right| \\ &= \frac{1}{n+1}. \end{aligned}$$

Now, the condition that

$$\left|\frac{3n+2}{n+1} - 3\right| = \frac{1}{n+1} < \varepsilon$$

it is equivalent to

$$\frac{1}{\varepsilon} - 1 < n.$$

Now let's write the formal proof.

Let $\varepsilon > 0$ be arbitrary and let $K \in \mathbb{N}$ be such that $\frac{1}{\varepsilon} - 1 < K$. Then, $\frac{1}{K+1} < \varepsilon$. Now if $n \geq K$ then $\frac{1}{n+1} \leq \frac{1}{K+1}$ and thus if $n \geq K$ then

$$\begin{aligned} \left|\frac{3n+2}{n+1} - 3\right| &= \left|\frac{-1}{n+1}\right| \\ &= \frac{1}{n+1} \\ &\leq \frac{1}{K+1} \\ &< \varepsilon. \end{aligned}$$

By definition, this proves that $(x_n) \to 3$. ☐

Example 3.1.5. Using the definition of the limit of a sequence, prove that $\lim_{n\to\infty} \frac{4n^3+3n}{n^3+6} = 4$.

Proof. Let $x_n = \frac{4n^3+3n}{n^3+6}$. We want to show that for any given $\varepsilon > 0$, there exists $K \in \mathbb{N}$ such that if $n \geq K$ then

$$|x_n - 4| = \left|\frac{4n^3+3n}{n^3+6} - 4\right| < \varepsilon.$$

Start by analyzing $|x_n - 4|$:

$$\begin{aligned} |x_n - 4| &= \left|\frac{4n^3+3n}{n^3+6} - 4\right| \\ &= \left|\frac{3n-24}{n^3+6}\right| \end{aligned}$$

In this case, it is difficult to explicitly isolate for n in terms of ε. Instead we take a different approach; we find an upper bound for $|x_n - 4| = \left|\frac{3n-24}{n^3+6}\right|$:

$$\begin{aligned} \left|\frac{3n-24}{n^3+6}\right| &\leq \frac{3n+24}{n^3+6} \\ &\leq \frac{27n}{n^3+6} \\ &< \frac{27n}{n^3} \\ &= \frac{27}{n^2}. \end{aligned}$$

Hence, if $\frac{27}{n^2} < \varepsilon$ then also $|x_n - 4| < \varepsilon$ by the transitivity property of inequalities. The inequality $\frac{27}{n^2} < \varepsilon$ holds true if and only if $\sqrt{\frac{27}{\varepsilon}} < n$. Now that we have done a detailed preliminary analysis, we can proceed with the proof.

Suppose that $\varepsilon > 0$ is given and let $K \in \mathbb{N}$ be such that $\sqrt{\frac{27}{\varepsilon}} < K$. Then $\frac{27}{\varepsilon} < K^2$, and thus $\frac{27}{K^2} < \varepsilon$. Then, if $n \geq K$ then $\frac{27}{n^2} \leq \frac{27}{K^2}$ and therefore

$$\begin{aligned}
|x_n - 4| &= \left|\frac{3n - 24}{n^3 + 6}\right| \\
&\leq \frac{3n + 24}{n^3 + 6} \\
&\leq \frac{27n}{n^3 + 6} \\
&< \frac{27n}{n^3} \\
&= \frac{27}{n^2} \\
&\leq \frac{27}{K^2} \\
&< \varepsilon.
\end{aligned}$$

This proves that $\lim_{n\to\infty} x_n = 4$. □

Example 3.1.6 (Important)**.** Prove that for any irrational number ζ there exists a sequence of rational numbers (x_n) converging to ζ.

Proof. Let (δ_n) be any sequence of positive numbers converging to zero, for example, $\delta_n = \frac{1}{n}$. Now since $\zeta - \delta_n < \zeta + \delta_n$ for each $n \in \mathbb{N}$, then by the Density theorem there exists a rational number x_n such that $\zeta - \delta_n < x_n < \zeta + \delta_n$. In other words, $|x_n - \zeta| < \delta_n$. Now let $\varepsilon > 0$ be arbitrary. Since (δ_n) converges to zero, there exists $K \in \mathbb{N}$ such that $|\delta_n - 0| < \varepsilon$ for all $n \geq K$, or since $\delta_n > 0$, then $\delta_n < \varepsilon$ for all $n \geq K$. Therefore, if $n \geq K$ then $|x_n - \zeta| < \delta_n < \varepsilon$. Thus, for arbitrary $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that if $n \geq K$ then $|x_n - \zeta| < \varepsilon$. This proves that (x_n) converges to ζ. □

Example 3.1.7. Let $x_n = \frac{\cos(n)}{n^2-1}$ where $n \geq 2$. Prove that $\lim_{n\to\infty} x_n = 0$.

Proof. We want to prove given any $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that

$$\left|\frac{\cos(n)}{n^2-1}\right| < \varepsilon, \quad n \geq K.$$

Now, since $|\cos(x)| \leq 1$ for all $x \in \mathbb{R}$ we have that

$$\begin{aligned}\left|\frac{\cos(n)}{n^2-1}\right| &= \frac{|\cos(n)|}{n^2-1}\\ &\leq \frac{1}{n^2-1}\\ &\leq \frac{1}{n^2-\frac{1}{2}n^2}\\ &= \frac{2}{n^2}.\end{aligned}$$

Thus, if $\frac{2}{n^2} < \varepsilon$ then $\left|\frac{\cos(n)}{n^2-1}\right|$.

Let $\varepsilon > 0$ be arbitrary. Let $K \in \mathbb{N}$ be such that $\sqrt{\frac{2}{\varepsilon}} < K$. Then $\frac{2}{K^2} < \varepsilon$. Therefore, if $n \geq K$ then $\frac{2}{n^2} \leq \frac{2}{K^2}$ and therefore

$$\begin{aligned}\left|\frac{\cos(n)}{n^2-1}\right| &= \frac{|\cos(n)|}{n^2-1}\\ &\leq \frac{1}{n^2-1}\\ &\leq \frac{1}{n^2-\frac{1}{2}n^2}\\ &= \frac{2}{n^2} \leq \frac{2}{K^2}\\ &< \varepsilon\end{aligned}$$

This proves that $\lim_{n\to\infty} x_n = 0$. ☐

Example 3.1.8. Does the sequence (x_n) defined by $x_n = \frac{(-1)^n n}{n+1}$ converge?

A useful tool for proving convergence is the following.

Theorem 3.1.9

Let (x_n) be a sequence and let $L \in \mathbb{R}$. Let (a_n) be a sequence of positive numbers such that $\lim\limits_{n\to\infty} a_n = 0$. Suppose that there exists $M \in \mathbb{N}$ such that

$$|x_n - L| \leq a_n, \qquad \forall\, n \geq M.$$

Then $\lim\limits_{n\to\infty} x_n = L$.

Proof. Let $\varepsilon > 0$ be arbitrary. Since $a_n \to 0$, there exists $K_1 \in \mathbb{N}$ such that $a_n < \varepsilon$ for all $n \geq K_1$. Let $K = \max\{M, K_1\}$. Then, if $n \geq K$ then $a_n < \varepsilon$ and $|x_n - L| \leq a_n$. Thus, if $n \geq K$ then

$$|x_n - L| \leq a_n < \varepsilon.$$

□

Example 3.1.10. Suppose that $0 < r < 1$. Prove that $\lim_{n\to\infty} r^n = 0$.

Proof. We first note that

$$r = \frac{1}{\frac{1}{r}} = \frac{1}{1+x}$$

where $x = \frac{1}{r} - 1$ and since $r < 1$ then $x > 0$. Now, by Bernoulli's inequality (Example 1.2.5) it holds that $(1+x)^n \geq 1 + xn$ for all $n \in \mathbb{N}$ and therefore

$$\begin{aligned} r^n &= \frac{1}{(1+x)^n} \\ &\leq \frac{1}{1+nx} \\ &< \frac{1}{nx}. \end{aligned}$$

Now since $\lim_{n\to\infty} \frac{1}{nx} = 0$ then it follows by Theorem 3.1.9 that

$$\lim_{n\to\infty} r^n = 0.$$

□

Example 3.1.11. Consider the sequence $x_n = \frac{n^2-1}{2n^2+3}$. Prove that $\lim_{n\to\infty} x_n = \frac{1}{2}$.

Proof. We have that

$$\begin{aligned} \left|\frac{n^2-1}{2n^2+3} - \frac{1}{2}\right| &= \frac{5}{2}\frac{1}{(2n^2+3)} \\ &< \frac{5/2}{2n^2}. \end{aligned}$$

Using the definition of the limit of a sequence, one can show that $\lim_{n\to\infty} \frac{5}{4n^2} = 0$ and therefore $\lim_{n\to\infty} x_n = \frac{1}{2}$. □

Notice that in the definition of the limit of a sequence, we wrote "there exists a number L". Could there be more than one number L satisfying the definition of convergence of a sequence? Before we go any further, we prove that if a sequence converges then it has a unique limit.

Theorem 3.1.12: Uniqueness of Limits

A convergent sequence can have at most one limit.

Proof. Suppose that $(x_n) \to L_1$ and that $(x_n) \to L_2$. Let $\varepsilon > 0$ be arbitrary. Then there exists K_1 such that $|x_n - L_1| < \varepsilon/2$ for all $n \geq K_1$ and there exists K_2 such that $|x_n - L_2| < \varepsilon/2$ for all $n \geq K_2$. Let $K = \max\{K_1, K_2\}$. Then for $n \geq K$ it holds that $|x_n - L_1| < \varepsilon/2$

and also $|x_n - L_2| < \varepsilon/2$ and therefore

$$\begin{aligned} |L_1 - L_2| &= |L_1 - x_n + x_n - L_2| \\ &< |x_n - L_1| + |x_n - L_2| \\ &< \varepsilon/2 + \varepsilon/2 \\ &= \varepsilon. \end{aligned}$$

Hence, $|L_1 - L_2| < \varepsilon$ for all $\varepsilon > 0$, and therefore by Theoreom 2.2.7 we conclude that $|L_1 - L_2| = 0$, that is, $L_1 - L_2 = 0$. □

The ultimate long-time behavior of a sequence will not change if we discard a finite number of terms of the sequence. To be precise, suppose that $X = (x_1, x_2, x_3, \ldots)$ is a sequence and let $Y = (x_{m+1}, x_{m+2}, x_{m+3}, \ldots)$, that is, Y is the sequence obtained from X by discarding the first m terms of X. In this case, we will call Y the **m-tail** of X. The next theorem states, not surprisingly, that the convergence properties of X and Y are the same.

Theorem 3.1.13: Tails of Sequences

Let $X : \mathbb{N} \to \mathbb{R}$ be a sequence and let $Y : \mathbb{N} \to \mathbb{R}$ be the sequence obtained from X by discarding the first $m \in \mathbb{N}$ terms of X, in other words, $Y(n) = X(m + n)$. Then X converges to L if and only if Y converges to L.

Exercises

Exercise 3.1.1. Write the first three terms of the recursively defined sequence $x_1 = 1$, $x_{n+1} = \frac{1}{2}(x_n + \frac{2}{x_n})$ for $n \geq 1$.

Exercise 3.1.2. Use the definition of the limit of a sequence to establish the following limits:

(a) $\displaystyle\lim_{n\to\infty} \frac{n+1}{3n} = \frac{1}{3}$

(b) $\displaystyle\lim_{n\to\infty} \frac{3n^2+2}{4n^2+1} = \frac{3}{4}$

(c) $\displaystyle\lim_{n\to\infty} \frac{(-1)^n n}{n^2+1} = 0$

Exercise 3.1.3.

(a) Prove that $\lim_{n\to\infty} |x_n| = 0$ if and only if $\lim_{n\to\infty} x_n = 0$.

(b) Combining the previous result and Example 3.1.10, prove that if $1 < r < 0$ then $\lim_{n\to\infty} r^n = 0$.

(c) Conclude that for any real number $r \in \mathbb{R}$, if $|r| < 1$ then $\lim_{n\to\infty} r^n = 0$.

Exercise 3.1.4. Let $m \in \mathbb{N}$ and assume that $m \geq 2$.

(a) Prove that $\frac{1}{m^n} < \frac{1}{n}$ for all $n \in \mathbb{N}$.

(b) Use Theorem 3.1.9 to show that $\lim_{n\to\infty} \frac{1}{m^n} = 0$.

Note: Do not use Example 3.1.10 to show that $\lim_{n\to\infty} \frac{1}{m^n} = 0$.

Exercise 3.1.5. Suppose that $S \subset \mathbb{R}$ is non-empty and bounded above and let $u = \sup S$. Show that there exists a sequence (x_n) such that $x_n \in S$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} x_n = u$. HINT: If $\varepsilon > 0$ then clearly $u - \varepsilon < u$. Since $u = \sup(S)$ there exists $x \in S$ such that $u - \varepsilon < x < u$. Example 3.1.6 is similar.

Exercise 3.1.6. Let (x_n) be the sequence defined as

$$x_n = \begin{cases} 2n^2 + 1, & n < 50 \\ \frac{\sin(2n)}{n^2+1}, & n \geq 50. \end{cases}$$

Using the definition of the limit of a sequence, find $\lim_{n\to\infty} x_n$.

3.2 Limit Theorems

Proving that a particular number $L \in \mathbb{R}$ is the limit of a given sequence is usually not easy because there is no systematic way to determine a candidate limit L for a given arbitrary sequence. Instead, we are frequently interested in just knowing if a given sequence converges or not, and not so much on finding the actual limit. The theorems in this section help us do just that. We begin with a definition.

Definition 3.2.1: Boundedness

A sequence (x_n) is said to be **bounded** if there exists $R \geq 0$ such that $|x_n| \leq R$ for all $n \in \mathbb{N}$.

Example 3.2.2. Prove that (x_n) is bounded if and only if there exists numbers R_1 and R_2 such that $R_1 \leq x_n \leq R_2$ for all $n \in \mathbb{N}$.

Theorem 3.2.3: Convergence implies Boundedness

A convergent sequence is bounded.

Proof. Suppose that (x_n) converges to L. Then there exists $K \in \mathbb{N}$ such that $|x_n - L| < 1$ for all $n \geq K$. Let

$$R = 1 + \max\{|x_1 - L|, |x_2 - L|, \ldots, |x_{K-1} - L|\},$$

and we note that $R \geq 1$. Then for all $n \geq 1$ it holds that $|x_n - L| \leq R$. Indeed, if $n \geq K$ then $|x_n - L| < 1 \leq R$ and if $1 \leq n \leq K - 1$ then

$$|x_n - L| \leq \max\{|x_1 - L|, \ldots, |x_{K-1} - L|\} \leq R.$$

Thus, for all $n \geq 1$ it holds that $L - R \leq x_n \leq R + L$ and this proves that (x_n) is bounded. □

Theorem 3.2.4: Convergence under Absolute Value

If $(x_n) \to L$ then $(|x_n|) \to |L|$.

Proof. Follows by the inequality $||x_n| - |L|| \leq |x_n - L|$ (see Corollary 2.3.6). Indeed, for any given $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that $|x_n - L| < \varepsilon$ for all $n \geq K$ and therefore $||x_n| - |L|| \leq |x_n - L| < \varepsilon$ for all $n \geq K$. □

The following theorem describes how the basic operations of arithmetic preserve convergence.

Theorem 3.2.5: Limit Laws

Suppose that $(x_n) \to L$ and $(y_n) \to M$.

(a) Then $(x_n + y_n) \to L + M$ and $(x_n - y_n) \to L - M$.

(b) Then $(x_n y_n) \to LM$.

(c) If $y_n \neq 0$ and $M \neq 0$ then $\left(\frac{x_n}{y_n}\right) \to \frac{L}{M}$.

Proof. (i) By the triangle inequality

$$\begin{aligned} |x_n + y_n - (L + M)| &= |x_n - L + y_n - M| \\ &< |x_n - L| + |y_n - M|. \end{aligned}$$

Let $\varepsilon > 0$. There exists K_1 such that $|x_n - L| < \varepsilon/2$ for $n \geq K_1$ and there exists K_2 such that $|y_n - M| < \varepsilon/2$ for $n \geq K_2$. Let $K = \max\{K_1, K_2\}$. Then for $n \geq K$

$$\begin{aligned} |x_n + y_n - (L + M)| &\leq |x_n - L| + |y_n - M| \\ &< \varepsilon/2 + \varepsilon/2 \\ &= \varepsilon. \end{aligned}$$

The proof for $(x_n - y_n) \to L - M$ is similar.
(ii) We have that

$$\begin{aligned} |x_n y_n - LM| &= |x_n y_n - y_n L + y_n L - LM| \\ &\le |x_n y_n - y_n L| + |y_n L - LM| \\ &= |y_n||x_n - L| + |L||y_n - M|. \end{aligned}$$

Now, (y_n) is bounded because it is convergent, and therefore $|y_n| \le R$ for all $n \in \mathbb{N}$ for some $R > 0$. By convergence of (y_n) and (x_n), there exists $K \in \mathbb{N}$ such that $|x_n - L| < \frac{\varepsilon}{2R}$ and $|y_n - M| < \frac{\varepsilon}{2(|L|+1)}$ for all $n \ge K$. Therefore, if $n \ge K$ then

$$\begin{aligned} |x_n y_n - LM| &< |y_n||x_n - L| + |L||y_n - M| \\ &< R|x_n - L| + (|L| + 1)|y_n - M| \\ &< R\frac{\varepsilon}{2R} + (|L| + 1)\frac{\varepsilon}{2(|L| + 1)} \\ &= \varepsilon. \end{aligned}$$

(iii) It is enough to prove that $\left(\frac{1}{y_n}\right) \to \frac{1}{M}$ and then use (ii). Now, since $M \ne 0$ and $y_n \ne 0$ then $|y_n|$ is bounded below by some positive number, say $R > 0$. Indeed, $(|y_n|) \to |M|$ and $|y_n| > 0$. Thus, $\frac{1}{|y_n|} < \frac{1}{R}$ for all $n \in \mathbb{N}$. Now,

$$\begin{aligned} \left|\frac{1}{y_n} - \frac{1}{M}\right| &= \frac{1}{|y_n||M|}|y_n - M| \\ &< \frac{1}{R|M|}|y_n - M|. \end{aligned}$$

For $\varepsilon > 0$, there exists $K \in \mathbb{N}$ such that $|y_n - M| < R|M|\varepsilon$ for all

$n \geq K$. Therefore, for $n \geq K$ we have that

$$\begin{aligned}\left|\frac{1}{y_n} - \frac{1}{M}\right| &< \frac{1}{R|M|}|y_n - M| \\ &< \frac{1}{R|M|}R|M|\varepsilon \\ &= \varepsilon.\end{aligned}$$

□

Corollary 3.2.6

Suppose that $(x_n) \to L$. Then $(x_n^k) \to L^k$ for any $k \in \mathbb{N}$.

The next theorem states that the limit of a convergent sequence of non-negative terms is non-negative.

Theorem 3.2.7

Suppose that $(x_n) \to L$. If $x_n \geq 0$ for all $n \in \mathbb{N}$ then $L \geq 0$.

Proof. We prove the contrapositive, that is, we prove that if $L < 0$ then there exists $K \in \mathbb{N}$ such that $x_K < 0$. Suppose then that $L < 0$. Let $\varepsilon > 0$ be such that $L + \varepsilon < 0$. Since $(x_n) \to L$, there exists $K \in \mathbb{N}$ such that $x_K < L + \varepsilon$, and thus by transitivity we have $x_K < 0$. □

Corollary 3.2.8: Comparison

Suppose that (x_n) and (y_n) are convergent and suppose that there exists $M \in \mathbb{N}$ such that $x_n \leq y_n$ for all $n \geq M$. Then $\lim\limits_{n\to\infty} x_n \leq \lim\limits_{n\to\infty} y_n$.

Proof. Suppose for now that $M = 1$, that is, $x_n \leq y_n$ for all $n \in \mathbb{N}$. Consider the sequence $z_n = y_n - x_n$. Then $z_n \geq 0$ for all $n \in \mathbb{N}$ and (z_n) is convergent since it is the difference of convergent sequences. By Theorem 3.2.7, we conclude that $\lim_{n\to\infty} z_n \geq 0$. But

$$\begin{aligned}\lim_{n\to\infty} z_n &= \lim_{n\to\infty} (y_n - x_n)\\ &= \lim_{n\to\infty} y_n - \lim_{n\to\infty} x_n\end{aligned}$$

and therefore $\lim_{n\to\infty} y_n - \lim_{n\to\infty} x_n \geq 0$, which is the same as $\lim_{n\to\infty} y_n \geq \lim_{n\to\infty} x_n$. If $M > 1$, then we can apply the theorem to the M-tail of the sequences of (x_n) and (y_n) and the result follows. □

Corollary 3.2.9

Suppose that $a \leq x_n \leq b$ and $\lim\limits_{n\to\infty} x_n = L$. Then $a \leq L \leq b$.

Proof. We have that $0 \leq x_n - a \leq b - a$. The sequence $y_n = b - a$ is constant and converges to $b - a$. The sequence $z_n = x_n - a$ converges to $L - a$. Therefore, by the previous theorem, $0 \leq L - a \leq b - a$, or $a \leq L \leq b$. □

Theorem 3.2.10: Squeeze Theorem

Suppose that $y_n \leq x_n \leq z_n$ for all $n \in \mathbb{N}$. Assume that $(y_n) \to L$ and also $(z_n) \to L$. Then (x_n) is convergent and $(x_n) \to L$.

Proof. Let $\varepsilon > 0$ be arbitrary. There exists $K_1 \in \mathbb{N}$ such that $L - \varepsilon < y_n < L + \varepsilon$ for all $n \geq K_1$ and there exists $K_2 \in \mathbb{N}$ such that $L - \varepsilon < z_n < L + \varepsilon$ for all $n \geq K_2$. Let $K = \max\{K_1, K_2\}$. Then if $n \geq K$ then

$$L - \varepsilon < y_n \leq x_n \leq z_n < L + \varepsilon.$$

Therefore, for $n \geq K$ we have that

$$L - \varepsilon < x_n < L + \varepsilon$$

and thus $\lim_{n\to\infty} x_n = L$. □

Remark 3.2.11. Some people call the Squeeze Theorem the Sandwich Theorem; we are not those people.

Example 3.2.12. Let $0 < a < b$ and let $x_n = (a^n + b^n)^{1/n}$. Prove that $\lim\limits_{n\to\infty} x_n = b$.

Theorem 3.2.13: Ratio Test

Let (x_n) be a sequence such that $x_n > 0$ for all $n \in \mathbb{N}$ and suppose that $L = \lim\limits_{n\to\infty} \dfrac{x_{n+1}}{x_n}$ exists. If $L < 1$ then (x_n) converges and $\lim\limits_{n\to\infty} x_n = 0$.

Proof. Let $r \in \mathbb{R}$ be such that $L < r < 1$ and set $\varepsilon = r - L$. There exists $K \in \mathbb{N}$ such that

$$\frac{x_{n+1}}{x_n} < L + \varepsilon = r$$

for all $n \geq K$. Therefore, for all $n \geq K$ we have that

$$0 < x_{n+1} < r x_n.$$

Thus, $x_{K+1} < r x_K$, and therefore $x_{K+2} < r x_{K+1} < r^2 x_K$, and inductively for $m \geq 1$ it holds that

$$x_{K+m} < r^m x_K.$$

Hence, the tail of the sequence (x_n) given by $(y_m) = (x_{K+1}, x_{K+2}, \ldots,)$ satisfies

$$0 < y_m < r^m x_K.$$

Since $0 < r < 1$ it follows that $\lim_{m\to\infty} r^m = 0$ and therefore $\lim_{m\to\infty} y_m = 0$ by the Squeeze theorem. This implies that (x_n) converges to 0 also. □

Exercises

Exercise 3.2.1. Use the Limit Theorems to prove that if (x_n) converges and $(x_n + y_n)$ converges then (y_n) converges. Give an example of two sequences (x_n) and (y_n) such that **both** (x_n) and (y_n) diverge but $(x_n + y_n)$ converges.

Exercise 3.2.2. Is the sequence $y_n = (-1)^n n^4$ convergent? Explain.

Exercise 3.2.3. Let (x_n) and (y_n) be sequences in $\mathbb{R}$. Suppose that $\lim_{n\to\infty} x_n = 0$ and that (y_n) is bounded. Prove that $\lim_{n\to\infty} x_n y_n = 0$.

Exercise 3.2.4. Show that if (x_n) and (y_n) are sequences such that (x_n) and $(x_n + y_n)$ are convergent, then (y_n) is convergent.

Exercise 3.2.5. Give examples of the following:

(a) Divergent sequences (x_n) and (y_n) such that $z_n = x_n y_n$ converges.

(b) Divergent sequences (x_n) and (y_n) such that $z_n = x_n y_n$ diverges.

(c) A divergent sequence (x_n) and a convergent sequence (y_n) such that $z_n = x_n y_n$ converges.

(d) A divergent sequence (x_n) and a convergent sequence (y_n) such that $z_n = x_n y_n$ diverges.

Exercise 3.2.6. Let (x_n) and (y_n) be sequences and suppose that (x_n) converges to L. Assume that for every $\varepsilon > 0$ there exists $M \in \mathbb{N}$ such that $|x_n - y_n| < \varepsilon$ for all $n \geq M$. Prove that (y_n) also converges to L.

Exercise 3.2.7. Let (x_n) be a sequence and define a sequence (y_n) as

$$y_n = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

for $n \in \mathbb{N}$. Show that if $\lim_{n\to\infty} x_n = 0$ then $\lim_{n\to\infty} y_n = 0$.

Exercise 3.2.8. Let (x_n) be a convergent sequence with limit L. Let $f(x) = a_0 + a_1 x + \cdots + a_k x^k$ be a polynomial. Use the Limit Theorems to prove that the sequence (y_n) defined by $y_n = f(x_n)$ is convergent and find the limit of (y_n).

Exercise 3.2.9. Apply the Limit Theorems to find the limits of the following sequences:

(a) $x_n = \sqrt{\dfrac{2n^2+3}{n^2+1}}$

(b) $x_n = (2 + 1/n)^2$

(c) $x_n = \dfrac{n+1}{n\sqrt{n}}$

(d) $x_n = 2^n/n!$

Exercise 3.2.10. Let (x_n) be a sequence such that $x_n \neq 0$ for all $n \in \mathbb{N}$. Suppose that $\lim_{n\to\infty} x_n = L$ and $L > 0$. Let $w = \inf\{|x_n| \ : \ n \in \mathbb{N}\}$. Prove that $w > 0$.

Exercise 3.2.11. Let (x_n) be a sequence of positive numbers such that $\displaystyle\lim_{n\to\infty} \frac{x_{n+1}}{x_n} = L > 1$. Show that (x_n) is not bounded and hence is not convergent.

3.3 Monotone Sequences

As we have seen, a convergent sequence is necessarily bounded, and it is straightforward to construct examples of sequences that are bounded but not convergent, for example, $(x_n) = (1, 0, 1, 0, 1, 0, \ldots)$. In this section, we prove the Monotone Convergence Theorem which says that a bounded sequence whose terms increase (or decrease) must necessarily converge.

Definition 3.3.1: Monotone Sequences

Let (x_n) be a sequence.

(i) We say that (x_n) is **increasing** if $x_n \leq x_{n+1}$ for all $n \in \mathbb{N}$.

(ii) We say that (x_n) is **decreasing** if $x_{n+1} \leq x_n$ for all $n \in \mathbb{N}$.

(iii) We say that (x_n) is **monotone** if (x_n) is either increasing or decreasing.

Example 3.3.2. Prove that if (x_n) is increasing then (x_n) is bounded below. Similarly, prove that if (x_n) is decreasing then (x_n) is bounded above.

Theorem 3.3.3: Monotone Convergence Theorem

If (x_n) is bounded and monotone then (x_n) is convergent. In particular:

(i) if (x_n) is bounded above and increasing then

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

(ii) if (x_n) is bounded below and decreasing then

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

Proof. Suppose that (x_n) is bounded above and increasing. Let $u = \sup\{x_n \mid n \in \mathbb{N}\}$ and let $\varepsilon > 0$ be arbitrary. Then by the properties of the supremum, there exists x_K such that $u - \varepsilon < x_K \leq u$. Since (x_n) is increasing, and u is an upper bound for the range of the sequence, it follows that $x_K \leq x_n \leq u$ for all $n \geq K$. Therefore, $u - \varepsilon < x_n \leq u$ for all $n \geq K$. Clearly, this implies that $u - \varepsilon < x_n < u + \varepsilon$ for all $n \geq K$. Since $\varepsilon > 0$ was arbitrary, this proves that (x_n) converges to u.

Suppose now that (x_n) is bounded below and decreasing. Let $w = \inf\{x_n \mid n \in \mathbb{N}\}$ and let $\varepsilon > 0$ be arbitrary. Then by the properties of the infimum, there exists x_K such that $w \leq x_K < w + \varepsilon$. Since (x_n) is decreasing, and w is a lower bound for the range of the sequence, it follows that $w \leq x_n \leq x_K$ for all $n \geq K$. Therefore, $w \leq x_n < w + \varepsilon$ for all $n \geq K$. Hence, $w - \varepsilon < x_n < w + \varepsilon$ for all $n \geq K$. Since $\varepsilon > 0$ was arbitrary, this proves that (x_n) converges to w. □

The Monotone Convergence Theorem (MCT) is an important tool in real analysis and we will use it frequently; notice that it is more-or-less a direct consequence of the Completeness Axiom. In fact, we could have taken as our starting axiom the MCT and then proved the Completeness property of $\mathbb{R}$.

Example 3.3.4. By the MCT, a bounded sequence that is also monotone is convergent. However, it is easy to construct a convergent sequence that is not monotone. Provide such an example.

The MCT can be used to show convergence of recursively defined sequences. To see how, suppose that (x_n) is defined recursively as $x_1 = a$ and

$$x_{n+1} = f(x_n)$$

where f is some given function. For example, say $x_1 = 2$ and

$$x_{n+1} = 2 + \frac{1}{x_n}.$$

Hence, in this case $f(x) = 2 + \frac{1}{x}$. If (x_n) is bounded and increasing then by the MCT (x_n) converges, but we do not know what the limit is. However, for example, if f is a polynomial/rational function of x then we can conclude that $L = \lim_{n\to\infty} x_n$ must satisfy the equation

$$L = f(L).$$

Indeed, if f is a polynomial/rational function then by the Limit Laws we have

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

But $x_{n+1} = f(x_n)$ and therefore $\lim_{n\to\infty} x_{n+1} = f(L)$, which is equivalent to $\lim_{n\to\infty} x_n = f(L)$ since (x_{n+1}) is just the 1-tail of the sequence (x_n). Therefore, $L = f(L)$ as claimed. From the equation $L = f(L)$ we can solve for L if possible.

Example 3.3.5. Consider the sequence (x_n) defined recursively as $x_1 = 1$ and

$$x_{n+1} = \tfrac{1}{2}x_n + \frac{1}{4}, \qquad \forall\ n \geq 1.$$

Prove that (x_n) converges and find the limit.

Proof. We prove by induction that $\frac{1}{2} \leq x_n$ for all $n \in \mathbb{N}$, that is, (x_n) is bounded below by $\frac{1}{2}$. First of all, it is clear that $\frac{1}{2} \leq x_1$. Now assume that $\frac{1}{2} \leq x_n$ for some $n \in \mathbb{N}$. Then

$$\begin{aligned} x_{n+1} &= \tfrac{1}{2}x_n + \frac{1}{4} \\ &\geq \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{4} \\ &= \frac{1}{2}. \end{aligned}$$

Hence, (x_n) is bounded below by $\frac{1}{2}$. We now prove that (x_n) is decreasing. We compute that $x_2 = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$, and thus $x_2 < x_1$. Assume now that $x_n < x_{n-1}$ for some $n \in \mathbb{N}$. Then

$$x_{n+1} < \frac{1}{2}x_{n-1} + \frac{1}{4} = x_n.$$

Hence, by induction we have shown that (x_n) is decreasing. By the MCT, (x_n) is convergent. Suppose that $(x_n) \to L$. Then also $(x_{n+1}) \to L$ and the sequence $y_n = \frac{1}{2}x_n + \frac{1}{4}$ converges to $\frac{1}{2}L + \frac{1}{4}$. Therefore, $L = \frac{1}{2}L + \frac{1}{4}$ and thus $L = 1/2$. □

Before we embark on the next example, we recall that

$$1 + r + r^2 + \cdots + r^{n-1} = \frac{1 - r^n}{1 - r}$$

and if $0 < r < 1$ then $0 < r^n < r < 1$ and therefore

$$\frac{1 - r^n}{1 - r} < \frac{1}{1 - r}.$$

Example 3.3.6. Consider the sequence

$$x_n = 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!}.$$

Note that this can be defined recursively as $x_1 = 1$ and $x_{n+1} = x_n + \frac{1}{(n+1)!}$. Prove that (x_n) converges.

Proof. We will prove by the MCT that (x_n) converges. By induction, one can show that $2^{n-1} < n!$ for all $n \geq 3$. Therefore,

$$\begin{aligned} x_n &< 1 + \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{2^{n-1}} \\ &= 1 + \frac{1 - (1/2)^n}{1 - (1/2)} \\ &< 1 + \frac{1}{1 - (1/2)} \\ &= 3. \end{aligned}$$

Hence, $x_n < 3$ and therefore (x_n) is bounded. Now, since $x_{n+1} = x_n + \frac{1}{(n+1)!}$ then clearly $x_n < x_{n+1}$. Thus (x_n) is increasing. By the MCT, (x_n) converges. You might recognize that $\lim_{n\to\infty} x_n = e = 2.71828\ldots$. □

Example 3.3.7. Let $x_1 = 0$ and let $x_{n+1} = \sqrt{2 + x_n}$ for $n \geq 1$. Prove that (x_n) converges and find its limit.

Proof. Clearly, $x_1 < x_2 = \sqrt{2}$. Assume by induction that $x_k > x_{k-1}$ for some $k \in \mathbb{N}$. Then

$$\begin{aligned} x_{k+1} &= \sqrt{2 + x_k} \\ &> \sqrt{2 + x_{k-1}} \\ &= x_k. \end{aligned}$$

Hence, (x_n) is an increasing sequence. We now prove that (x_n) is bounded above. Clearly, $x_1 < 2$. Assume that $x_k < 2$ for some $k \in \mathbb{N}$.

Then $x_{k+1} = \sqrt{2 + x_k} < \sqrt{2+2} = 2$. This proves that (x_n) is bounded above. By the MCT, (x_n) converges, say to L. Moreover, since $x_n \geq 0$ (as can be proved by induction), then $L \geq 0$. Therefore, $L = \sqrt{2+L}$ and then $L^2 - L - 2 = 0$. Hence, $L = \frac{1\pm\sqrt{1+8}}{2} = \frac{1\pm 3}{2}$. Since $L \geq 0$ then $L = 2$. ☐

Example 3.3.8. Consider the sequence $x_n = \left(1 + \frac{1}{n}\right)^n$. We will show that (x_n) is bounded and increasing, and therefore by the MCT (x_n) convergent. The limit of this sequence is the number e. From the binomial theorem

$$\begin{aligned}
x_n &= \sum_{k=0}^{n} \binom{n}{k} \frac{1}{n^k} \\
&= 1 + \frac{n}{n} + \frac{1}{2!}\frac{n(n-1)}{n^2} + \frac{1}{3!}\frac{n(n-1)(n-2)}{n^3} + \cdots \\
&\quad + \frac{1}{n!}\frac{n(n-1)(n-2)\cdots(n-(n-1))}{n^k} \\
&= 1 + 1 + \frac{1}{2!}(1 - \tfrac{1}{n}) + \frac{1}{3!}(1-\tfrac{1}{n})(1-\tfrac{1}{n}) + \cdots \\
&\quad + \frac{1}{n!}(1-\tfrac{1}{n})(1-\tfrac{1}{n})\cdots(1-(n-1)/n) \\
&< 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} \\
&< 1 + 1 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^{n-1}} \\
&= 1 + \frac{1-(1/2)^n}{1-1/2} \\
&< 3
\end{aligned}$$

where we used that $2^{n-1} < n!$ for all $n \geq 3$. This shows that (x_n) is bounded. Now, for each $1 \leq k \leq n$, we have that

$$\begin{aligned}\binom{n}{k}\frac{1}{n^k} &= \frac{n(n-1)(n-2)\cdots(n-(k-1))}{n^k}\\ &= (1-\tfrac{1}{n})(1-\tfrac{2}{n})\cdots(1-\tfrac{k-1}{n}).\end{aligned}$$

And similarly,

$$\binom{n+1}{k}\frac{1}{(n+1)^k} = (1-\tfrac{1}{n+1})(1-\tfrac{2}{n+1})\cdots(1-\tfrac{k-1}{n+1}).$$

It is clear that $(1-\frac{j}{n}) < (1-\frac{j}{n+1})$ for all $1 \leq j \leq n$. Hence, $\binom{n}{k}\frac{1}{n^k} < \binom{n+1}{k}\frac{1}{(n+1)^k}$. Therefore, $x_n < x_{n+1}$, that is, (x_n) is increasing. By the MCT, (x_n) converges to $\sup\{x_n : n \in \mathbb{N}\}$.

Exercises

Exercise 3.3.1. Let (x_n) be an increasing sequence, let (y_n) be a decreasing sequence, and assume that $x_n \leq y_n$ for all $n \in \mathbb{N}$. Prove that $\lim_{n\to\infty} x_n$ and $\lim_{n\to\infty} y_n$ exist, and that $\lim_{n\to\infty} x_n \leq \lim_{n\to\infty} y_n$. NOTE: Recall that a sequence (x_n) is bounded if there exist constants $R_1, R_2 > 0$ (**independent of** n) such that $R_1 \leq x_n \leq R_2$ for all $n \in \mathbb{N}$.

Exercise 3.3.2. Let $x_1 = 8$ and let $x_{n+1} = \frac{1}{2}x_n + 2$ for $n \geq 1$. Prove that (x_n) is bounded and monotone. Find the limit of (x_n).

Exercise 3.3.3. Let $x_1 = 1$ and let $x_{n+1} = \dfrac{n}{n+1}x_n^2$ for $n \geq 1$. Prove that (x_n) is bounded and monotone. Find the limit of (x_n). HINT: Using induction to prove that (x_n) is monotone will not work with this sequence. Instead, work with x_{n+1} directly to prove that (x_n) is monotone.

Exercise 3.3.4. True or false, a convergent sequence is necessarily monotone? If it is true, prove it. If it is false, give an example.

3.4 Bolzano-Weierstrass Theorem

We can gather information about a sequence by studying its **subsequences**. Loosely speaking, a subsequence of (x_n) is a new sequence (y_k) such that each term y_k is from the original sequence (x_n) and the term y_{k+1} appears to the "right" of the term y_k in the original sequence (x_n). Let us be precise about what we mean to the "right".

Definition 3.4.1: Subsequences

Let (x_n) be a sequence. A **subsequence** of (x_n) is a sequence of the form $(x_{n_1}, x_{n_2}, x_{n_3}, \ldots)$ where $n_1 < n_2 < n_3 < \cdots$ is a sequence of strictly increasing natural numbers. A subsequence of (x_n) will be denoted by (x_{n_k}).

The notation (x_{n_k}) of a subsequence indicates that the indexing variable is $k \in \mathbb{N}$. The selection of the elements of (x_n) to form a subsequence (x_{n_k}) does not need to follow any particular well-defined pattern but only that $n_1 < n_2 < n_3 < \cdots$. Notice that for any increasing sequence $n_1 < n_2 < n_3 < \cdots$ of natural numbers, we have

$$k \leq n_k$$

for all $k \geq 1$.

Example 3.4.2. An example of a subsequence of $x_n = \frac{1}{n}$ is the sequence $(y_k) = (1, \frac{1}{3}, \frac{1}{5}, \frac{1}{7}, \ldots)$. Here we have chosen the odd terms of the sequence (x_n) to create (y_k). In other words, if we write that $y_k = x_{n_k}$ then $n_k = 2k - 1$. Another example of a subsequence of (x_n) is obtained by taking the even terms to get the subsequence $(\frac{1}{2}, \frac{1}{4}, \frac{1}{6}, \ldots)$, so that here $n_k = 2k$. In general, we can take any increasing selection

$n_1 < n_2 < n_3 < \cdots$, such as

$$(\tfrac{1}{11}, \tfrac{1}{303}, \tfrac{1}{2000}, \ldots)$$

to form a subsequence of (x_n).

Example 3.4.3. Two subsequences of

$$(x_n) = (1, -1, \tfrac{1}{2}, -1, \tfrac{1}{3}, -1, \tfrac{1}{4}, -1, \ldots)$$

$(-1, -1, -1, \ldots,)$ and $(1, \frac{1}{2}, \frac{1}{3}, \ldots,)$. Both of these subsequences converge to distinct limits.

Example 3.4.4 (Important)**.** We proved that $\mathbb{Q}$ is countable and hence there is a bijection $f : \mathbb{N} \to \mathbb{Q}$. The bijection f defines a sequence

$$(x_n) = (x_1, x_2, x_3, x_4, \ldots)$$

where $x_n = f(n)$. Let $L \in \mathbb{R}$ be arbitrary. By the density of $\mathbb{Q}$ in $\mathbb{R}$, there exists $x_{n_1} \in \mathbb{Q}$ such that $x_{n_1} \in (L-1, L+1)$. Now consider the interval $(L - \frac{1}{2}, L + \frac{1}{2})$. It has infinitely many distinct rational numbers (by the Density Theorem). Therefore, there exists $n_2 > n_1$ such that $x_{n_2} \in (L - \frac{1}{2}, L + \frac{1}{2})$. Consider now the interval $(L - \frac{1}{3}, L + \frac{1}{3})$. It has infinitely many rational numbers, and therefore there exists $n_3 > n_2$ such that $x_{n_3} \in (L - \frac{1}{3}, L + \frac{1}{3})$. By induction, there exists a subsequence (x_{n_k}) of (x_n) such that $|x_{n_k} - L| < \frac{1}{k}$ for all $k \geq 1$. Therefore, $\lim_{k\to\infty} x_{n_k} = L$. We proved the following: For any real number L there exists a sequence of rational numbers that converges to L.

The following theorem is a necessary condition for convergence.

Theorem 3.4.5

If $(x_n) \to L$ then every subsequence of (x_n) converges to L.

Proof. Let $\varepsilon > 0$. Then there exists $K \in \mathbb{N}$ such that $|x_n - L| < \varepsilon$ for all $n \geq K$. Since $n_K \geq K$ and $n_K < n_{K+1} < \ldots$, then $|x_{n_k} - L| < \varepsilon$ for all $k \geq K$. ☐

The contrapositive of the previous theorem is worth stating.

Theorem 3.4.6

Let (x_n) be a sequence.

(i) If (x_n) has two subsequences converging to distinct limits then (x_n) is divergent.

(ii) If (x_n) has a subsequence that diverges then (x_n) diverges.

The following is a very neat result that will supply us with a very short proof of the main result of this section, namely, the Bolzano-Weierstrass Theorem.

Theorem 3.4.7

Every sequence has a monotone subsequence.

Proof. Let (x_n) be an arbitrary sequence. We will say that the term x_m is a **peak** if $x_m \geq x_n$ for all $n \geq m$. In other words, x_m is a peak if it is an upper bound of all the terms coming after it. There are two possible cases for (x_n), either it has an infinite number of peaks or it has a finite number of peaks. Suppose that it has an infinite number of peaks, say $x_{m_1}, x_{m_2}, \ldots$, and we may assume that $m_1 < m_2 < m_3 < \cdots$. Then,

$x_{m_1} \geq x_{m_2} \geq x_{m_3} \geq \cdots$, and therefore (x_{m_k}) is a decreasing sequence. Now suppose that there are only a finite number of peaks and that x_m is the last peak. Then $x_{n_1} = x_{m+1}$ is not a peak and therefore there exists $n_2 > n_1$ such that $x_{n_2} \geq x_{n_1}$. Similarly, x_{n_2} is not a peak and therefore there exists $n_3 > n_2$ such that $x_{n_3} \geq x_{n_2}$. Hence, by induction, there exists a subsequence (x_{n_k}) that is increasing. □

Theorem 3.4.8: Bolzano-Weierstrass

Every bounded sequence contains a convergent subsequence.

Proof. Let (x_n) be an arbitrary bounded sequence. By Theorem 3.4.7, (x_n) has a monotone subsequence (x_{n_k}). Since (x_n) is bounded then so is (x_{n_k}). By the MCT applied to (x_{n_k}) we conclude that (x_{n_k}) is convergent. □

We will give a second proof of the Bolzano-Weierstrass Theorem that is more "hands-on".

Another proof of Bolzano-Weierstrass. If (x_n) is a bounded sequence, then there exists $a_1, b_1 \in \mathbb{R}$ such that $a_1 \leq x_n \leq b_1$ for all $n \in \mathbb{N}$. We will apply a recursive bisection algorithm to hunt down a converging subsequence of (x_n). Let $m_1 = \frac{(a_1+b_1)}{2}$ be the mid-point of the interval $[a_1, b_1]$. Then at least one of the subsets $I_1 = \{n \in \mathbb{N} : a_1 \leq x_n \leq m_1\}$ or $J_1 = \{n \in \mathbb{N} : m_1 \leq x_n \leq b_1\}$ is infinite; if it is I_1 then choose some $x_{n_1} \in [a_1, m_1]$ and let $a_2 = a_1$ and $b_2 = m_1$; otherwise choose some $x_{n_1} \in [m_1, b_1]$ and let $a_2 = m_1$, $b_2 = b_1$. In any case, it is clear that $(b_2 - a_2) = \frac{(b_1-a_1)}{2}$, that $a_1 \leq a_2$ and that $b_1 \geq b_2$. Now let $m_2 = \frac{(a_2+b_2)}{2}$ be the mid-point of the interval $[a_2, b_2]$ and let $I_2 = \{n \in \mathbb{N} : a_2 \leq x_n \leq m_2\}$ and let $J_2 = \{n \in \mathbb{N} : m_2 \leq x_n \leq b_2\}$. If I_2 is infinite then choose some $x_{n_2} \in [a_2, m_2]$ and let $a_3 = a_2$, $b_3 = m_2$; otherwise choose some $x_{n_2} \in [m_2, b_2]$ and let $a_3 = m_2$ and $b_3 = b_2$. In any case, it is clear that $(b_3 - a_3) = \frac{(b_1-a_1)}{2^2}$, that $a_2 \leq a_3$ and $b_3 \leq b_2$. By induction, there exists sequences (a_k), (b_k), and (x_{n_k}) such that $a_k \leq x_{n_k} \leq b_k$ and $(b_k - a_k) = \frac{(b_1-a_1)}{2^{k-1}}$, (a_k) is increasing and (b_k) is decreasing. It is clear that $a_k \leq b$ and $a \leq b_k$ for all $k \in \mathbb{N}$. Hence, by the MCT, (a_k) and (b_k) are convergent. Moreover, since $(b_k - a_k) = \frac{(b_1-a_1)}{2^{k-1}}$ then $\lim(b_k - a_k) = 0$ and consequently $\lim a_k = \lim b_k = L$. By the Squeeze theorem we conclude that $\lim_{k\to\infty} x_{n_k} = L$. □

Notice that the proofs of the Bolzano-Weierstrass Theorem rely on the Monotone Convergence Theorem and the latter relies on the Completeness Axiom. We therefore have the following chain of implications:

$$\text{Completeness} \Longrightarrow \text{MCT} \Longrightarrow \text{Bol-Wei}$$

It turns out that if we had taken as our starting axiom the Bolzano-Weierstrass theorem then we could prove the Completeness property and then of course the MCT. In other words, all three statements are equivalent:

$$\text{Completeness} \Longleftrightarrow \text{MCT} \Longleftrightarrow \text{Bol-Wei}$$

Exercises

Exercise 3.4.1. Prove that the following sequences are divergent.

(a) $x_n = 1 - (-1)^n + 1/n$

(b) $x_n = \sin(n\pi/4)$

(Hint: Theorem 3.4.6)

Exercise 3.4.2. Suppose that $x_n \geq 0$ for all $n \in \mathbb{N}$ and suppose that $\lim_{n\to\infty} (-1)^n x_n = L$ exists. Prove that $L = 0$ and that also $\lim_{n\to\infty} x_n = L$. (Hint: Consider subsequences of $(-1)^n x_n$.)

Exercise 3.4.3. Let (x_n) be a sequence.

(a) Suppose that (x_n) is increasing. Prove that if (x_n) has a subsequence (x_{n_k}) that is bounded above then (x_n) is also bounded above.

(b) Suppose that (x_n) is decreasing. Prove that if (x_n) has a subsequence (x_{n_k}) that is bounded below then (x_n) is also bounded below.

Exercise 3.4.4. True or false: If (x_n) is bounded and diverges then (x_n) has two subsequences that converge to distinct limits. Explain.

Exercise 3.4.5. Give an example of a sequence (x_n) with the following property: For each number $L \in \left\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\right\}$ there exists a subsequence (x_{n_k}) such that $x_{n_k} \to L$. **Hint:** If you are spending a lot of time on this question then you have not been reading this textbook carefully.

Exercise 3.4.6. Suppose that x_1 and y_1 satisfy $0 < x_1 < y_1$ and define

$$x_{n+1} = \sqrt{x_n y_n}, \qquad y_{n+1} = \frac{x_n + y_n}{2}$$

for $n \geq 1$. Prove that (x_n) and (y_n) are convergent and that $\lim x_n = \lim y_n$. (Hint: First show that $\sqrt{xy} \leq (x+y)/2$ for any $x, y > 0$.)

Exercise 3.4.7. Let (x_n) be a sequence and define the sequence (y_n) as

$$y_n = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Prove that if $(x_n) \to L$ then $(y_n) \to L$.

3.5 limsup and liminf

The behavior of a convergent sequence is easy to understand. Indeed, if $(x_n) \to L$ then eventually the terms of (x_n) will be arbitrarily close to L for n sufficiently large. What else is there to say? In this section, we focus on bounded sequences that do not necessarily converge. The idea is that we would like to develop a limit concept for these sequences, and in particular, a "limiting upper bound".

Let (x_n) be an arbitrary sequence and introduce the set S defined as the set of all the limits of convergent subsequences of (x_n), that is,

$$S = \{L \in \mathbb{R} \mid (x_{n_k}) \to L\}.$$

We will call S the **subsequences limit set** of (x_n).

Example 3.5.1. If (x_n) is bounded and S is the subsequences limit set of (x_n) explain why S is non-empty.

Example 3.5.2. Here are six examples of sequences and the corresponding subsequences limit set. Notice that in the cases where (x_n)

(x_n)	S
$(1, \frac{1}{2}, \frac{1}{3}, \ldots)$	$\{0\}$
$(1, -1, 1, -1, \ldots)$	$\{1, -1\}$
$(1, 2, 3, 4, \ldots)$	$\emptyset$
$(1, \frac{3}{2}, \frac{1}{3}, \frac{5}{4}, \frac{1}{5}, \frac{7}{6}, \ldots)$	$\{0, 1\}$
$(0, 1, \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \ldots, \frac{7}{8}, \ldots)$	$[0, 1]$
(r_n) enumeration of $\mathbb{Q}$	$\mathbb{R}$

Table 3.2: Limits of subsequences

is bounded, the set S is also bounded, which is as expected since if $a \leq x_n \leq b$ then for any convergent subsequence (x_{n_k}) of (x_n) we necessarily have $a \leq x_{n_k} \leq b$ and therefore $a \leq \lim_{k\to\infty} x_{n_k} \leq b$.

In general, we have seen that for a general set S, $\sup(S)$ and $\inf(S)$ are not necessarily in S. This, however, is not the case for the subsequences limit set.

Lemma 3.5.3

Let (x_n) be a bounded sequence and let S be its subsequences limit set. Then $\sup(S) \in S$ and $\inf(S) \in S$. In other words, there exists a subsequence (x_{n_k}) of (x_n) such that $\lim_{k\to\infty} x_{n_k} = \sup(S)$ and similarly there exists a subsequence (y_{n_k}) of (x_n) such that $\lim_{k\to\infty} y_{n_k} = \inf(S)$.

Proof. Let $u = \sup(S)$. If $\varepsilon > 0$ then there exists $s \in S$ such that $u - \varepsilon < s \leq u$. Since $s \in S$, there exists a subsequence (x_{n_k}) of (x_n) that converges to s. Therefore, there exists $K \in \mathbb{N}$ such that $u - \varepsilon < x_{n_k} < u + \varepsilon$ for all $k \geq K$. Hence, for each ε, the inequality $u - \varepsilon < x_n < u + \varepsilon$ holds for infinitely many n. Consider $\varepsilon_1 = 1$. Then there exists x_{n_1} such that $u - \varepsilon_1 < x_{n_1} < u + \varepsilon_1$. Now take $\varepsilon_2 = \frac{1}{2}$. Since $u - \varepsilon_2 < x_n < u + \varepsilon$ holds for infinitely many n, there exists x_{n_2} such that $u - \varepsilon_2 < x_{n_2} < u + \varepsilon_2$ and $n_2 > n_1$. By induction, for $\varepsilon_k = \frac{1}{k}$, there exists x_{n_k} such that $u - \varepsilon_k < x_{n_k} < u + \varepsilon_k$ and $n_k > n_{k-1}$. Hence, the subsequence (x_{n_k}) satisfies $|x_{n_k} - u| < \frac{1}{k}$ and therefore $(x_{n_k}) \to u$. Therefore, $u = \sup(S) \in S$. □

By Lemma 3.5.3, if (x_n) is a bounded sequence then there exists a convergent subsequence of (x_n) whose limit is larger than any other limit of a convergent subsequence of (x_n). This leads to the following definition.

Definition 3.5.4

Let (x_n) be a bounded sequence and let S be its subsequences limit set. We define the **limit superior** of (x_n) as

$$\limsup x_n = \sup S$$

and the **limit inferior** of (x_n) as

$$\liminf x_n = \inf S.$$

By Lemma 3.5.3, $\limsup x_n$ is simply the largest limit of all convergent subsequences of (x_n) while $\liminf x_n$ is the smallest limit of all convergent subsequences of (x_n). Notice that by definition it is clear that $\liminf x_n \leq \limsup x_n$. The next theorem gives an alternative characterization of $\limsup x_n$ and $\liminf x_n$. The idea is that $\limsup x_n$ is a sort of limiting supremum and $\liminf x_n$ is a sort of limiting infimum of a bounded sequence (x_n).

Theorem 3.5.5

Let (x_n) be a bounded sequence and let $L^* \in \mathbb{R}$. The following are equivalent:

(i) $L^* = \limsup x_n$

(ii) If $\varepsilon > 0$ then there are at most finitely many x_n such that $L^* + \varepsilon < x_n$ and infinitely many x_n such that $L^* - \varepsilon < x_n$.

(iii) Let $u_m = \sup\{x_n : n \geq m\}$. Then $L^* = \lim_{m\to\infty} u_m = \inf\{u_m : m \in \mathbb{N}\}$.

Proof. (i)→(ii) Let S denote the subsequences limit set of (x_n). By definition, $L^* = \limsup x_n = \sup(S)$ and by Lemma 3.5.3 we have that $L^* \in S$. Hence, there exists a subsequence of (x_n) converging to L^* and thus $L^* - \varepsilon < x_n < L^* + \varepsilon$ holds for infinitely many n. In particular $L^* - \varepsilon < x_n$ holds for infinitely many n. Suppose that $L^* + \varepsilon < x_n$ holds infinitely often. Now $x_n \leq M$ for all n and some $M > 0$. Since the inequality $L^* + \varepsilon < x_n$ holds infinitely often, there exists a sequence $n_1 < n_2 < \cdots$ such that $L^* + \varepsilon < x_{n_k} \leq M$ for all $k \in \mathbb{N}$. We can assume that (x_{n_k}) is convergent (because it is bounded and we can pass to a subsequence by the MCT) and thus $L^* + \varepsilon \leq \lim\limits_{n\to\infty} x_{n_k} \leq M$. Hence we have proved that the subsequence (x_{n_k}) converges to a number greater than L^* which contradicts the definition of $L^* = \sup(S)$.

(ii)→(iii) Let $\varepsilon > 0$. Since $L^* + \varepsilon/2 < x_m$ holds for finitely many m, there exists M such that $x_m \leq L^* + \varepsilon/2$ for all $m \geq M$. Hence, $L^* + \varepsilon/2$ is an upper bound of $\{x_n \mid n \geq m\}$ and thus $u_m < L^* + \varepsilon$. Since (u_m) is decreasing, we have that $u_m < L^* + \varepsilon$ for all $m \geq M$. Now, $L^* - \varepsilon/2 < x_n$ holds infinitely often and thus $L^* - \varepsilon < u_m$ for all $m \in \mathbb{N}$. Hence, $L^* - \varepsilon < u_m < L^* + \varepsilon$ for all $m \geq M$. This proves the claim.

(iii)→(i) Let (x_{n_k}) be a convergent subsequence. Since $n_k \geq k$, by definition of u_k, we have that $x_{n_k} \leq u_k$. Therefore, $\lim x_{n_k} \leq \lim u_k = L^*$. Hence, L^* is an upper bound of S. By definition of u_k, there exists x_{n_1} such that $u_1 - 1 < x_{n_1} \leq u_1$. By induction, there exists a subsequence (x_{n_k}) such that $u_k - \frac{1}{k} < x_{n_k} \leq u_k$. Hence, by the Squeeze Theorem, $L^* = \lim x_{n_k}$. Hence, $L^* \in S$ and thus $L^* = \sup S = \liminf x_n$. □

Example 3.5.6. Let p_n denote the nth prime number, that is $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, and so on. The numbers p_n and p_{n+1} are called **twin**

primes if $p_{n+1} - p_n = 2$. The **Twin Prime Conjecture** is that

$$\liminf(p_{n+1} - p_n) = 2$$

In other words, the Twin Prime Conjecture is that there are infinitely many pairs of twin primes.

We end this section with the following interesting theorem that says that if the subsequences limit set of a bounded sequence (x_n) consists of a single number L then the sequence (x_n) also converges to L.

Theorem 3.5.7

Let (x_n) be a bounded sequence and let $L \in \mathbb{R}$. If every convergent subsequence of (x_n) converges to L then (x_n) converges to L.

Proof. Suppose that (x_n) does not converge to L. Then, there exists $\varepsilon > 0$ such that for every $K \in \mathbb{N}$ there exists $n \geq K$ such that $|x_n - L| \geq \varepsilon$. Let $K_1 \in \mathbb{N}$. Then there exists $n_1 \geq K_1$ such that $|x_{n_1} - L| \geq \varepsilon$. Then there exists $n_2 > n_1 + 1$ such that $|x_{n_2} - L| \geq \varepsilon$. By induction, there exists a subsequence (x_{n_k}) of (x_n) such that $|x_{n_k} - L| \geq \varepsilon$ for all $k \in \mathbb{N}$. Now (x_{n_k}) is bounded and therefore by Bolzano-Weierstrass has a convergent subsequence, say (z_k), which is also a subsequence of (x_n). By assumption, (z_k) converges to L, which contradicts that $|x_{n_k} - L| \geq \varepsilon$ for all $k \in \mathbb{N}$. □

Another way to say Theorem 3.5.7 is that if (x_n) is bounded and $L = \limsup x_n = \liminf x_n$ then $(x_n) \to L$. The converse, by the way, has already been proved: if $(x_n) \to L$ then every subsequence of (x_n) converges to L and therefore $L = \limsup x_n = \liminf x_n$.

Exercises

Exercise 3.5.1. Determine the $\limsup_{n\to\infty} x_n$ and $\liminf_{n\to\infty} x_n$ for each case:

(a) $x_n = 3 + (-1)^n(1 + 1/n)$

(b) $x_n = 1 + \sin(n\pi/2)$

(c) $x_n = (2 - 1/n)(-1)^n$

Exercise 3.5.2. Let (x_n) and (y_n) be bounded sequences. Let (z_n) be the sequence $z_n = x_n + y_n$. Show that

$$\limsup_{n\to\infty} z_n \leq \limsup_{n\to\infty} x_n + \limsup_{n\to\infty} y_n$$

In other words, prove that

$$\limsup_{n\to\infty}(x_n + y_n) \leq \limsup_{n\to\infty} x_n + \limsup_{n\to\infty} y_n$$

Exercise 3.5.3. Let (x_n) and (y_n) be bounded sequences. Show that if $x_n \leq y_n$ for all n then

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

3.6 Cauchy Sequences

Up until now, the Monotone Convergence theorem is our main tool for determining that a sequence converges without actually knowing what the the limit is. It is a general sufficient condition for convergence. In this section, we prove another groundbreaking general sufficient condition for convergence known as the **Cauchy criterion**. Roughly speaking, the idea is that if the terms of a sequence (x_n) become closer and closer *to one another* as $n \to \infty$ then the sequence ought to converge. A sequence whose terms become closer and closer to one another is called a Cauchy sequence.

Definition 3.6.1: Cauchy Sequences

A sequence (x_n) is said to be a **Cauchy sequence** if for every $\varepsilon > 0$ there exists a natural number K such that if $n, m \geq K$ then $|x_n - x_m| < \varepsilon$.

In other words, (x_n) is a Cauchy sequence if the difference $|x_n - x_m|$ is arbitrarily small provided that both n and m are sufficiently large.

Example 3.6.2. Prove directly using the definition of a Cauchy sequence that if (x_n) and (y_n) are Cauchy sequences then the sequence $z_n = |x_n - y_n|$ is a Cauchy sequence.

Not surprisingly, a convergent sequence is indeed a Cauchy sequence.

Lemma 3.6.3

If (x_n) is convergent then it is a Cauchy sequence.

Proof. Suppose that $(x_n) \to L$. Let $\varepsilon > 0$ and let K be sufficiently large so that $|x_n - L| < \varepsilon/2$ for all $n \geq K$. If $n, m \geq K$ then by the triangle inequality,

$$\begin{aligned} |x_n - x_m| &= |x_n - L + L - x_m| \\ &\leq |x_n - L| + |x_m - L| \\ &< \varepsilon/2 + \varepsilon/2 \\ &= \varepsilon. \end{aligned}$$

This proves that (x_n) is Cauchy. □

A Cauchy sequence is bounded.

Lemma 3.6.4: Cauchy implies Boundedness

If (x_n) is a Cauchy sequence then (x_n) is bounded.

Proof. The proof is similar to the proof that a convergent sequence is bounded. □

Theorem 3.6.5: Cauchy Criterion

The sequence (x_n) is convergent if and only if (x_n) is a Cauchy sequence.

Proof. In Lemma 3.6.3 we already showed that if (x_n) converges then it is a Cauchy sequence. To prove the converse, suppose that (x_n) is a Cauchy sequence. By Lemma 3.6.4, (x_n) is bounded. Therefore, by the Bolzano-Weierstrass theorem there is a subsequence (x_{n_k}) of (x_n) that converges, say it converges to L. We will prove that (x_n) also converges to L. Let $\varepsilon > 0$ be arbitrary. Since (x_n) is Cauchy there exists $K \in \mathbb{N}$ such that if $n, m \geq K$ then $|x_n - x_m| < \varepsilon/2$. On the other hand, since

$(x_{n_k}) \to L$ there exists $n_M \geq K$ such that $|x_{n_M} - L| < \varepsilon/2$. Therefore, if $n \geq K$ then

$$\begin{aligned} |x_n - L| &= |x_n - x_{n_M} + x_{n_M} - L| \\ &\leq |x_n - x_{n_M}| + |x_{n_M} - L| \\ &< \varepsilon/2 + \varepsilon/2 \\ &= \varepsilon. \end{aligned}$$

This proves that (x_n) converges to L. □

Example 3.6.6. Let $0 < r < 1$ and suppose that $|x_n - x_{n+1}| \leq r^n$ for all $n \in \mathbb{N}$. Using the fact that $1 + r + r^2 + \cdots + r^k < \frac{1}{1-r}$ for all $k \in \mathbb{N}$ prove that if $m > n$ then $|x_n - x_m| \leq \frac{r^n}{1-r}$. Deduce that (x_n) is a Cauchy sequence.

When the MCT is not applicable, the Cauchy criterion is another possible tool to show convergence of a sequence.

Example 3.6.7. Consider the sequence (x_n) defined by $x_1 = 1$, $x_2 = 2$, and

$$x_n = \tfrac{1}{2}(x_{n-2} + x_{n-1})$$

for $n \geq 2$. One can show that (x_n) is not monotone and therefore the MCT is not applicable.

(a) Prove that $1 \leq x_n \leq 2$ for all $n \in \mathbb{N}$.

(b) Prove that $|x_n - x_{n+1}| = \frac{1}{2^{n-1}}$ for all $n \in \mathbb{N}$.

(c) Prove that if $m > n$ then

$$|x_n - x_m| < \frac{1}{2^{n-2}}$$

Hint: Use part (b) and the Triangle inequality.

(d) Deduce that (x_n) is a Cauchy sequence and thus convergent.

(e) Show by induction that $x_{2n+1} = 1 + \frac{2}{3}\left(1 - \frac{1}{4^n}\right)$ and deduce that $\lim x_n = \frac{5}{3}$.

Notice that the main result used in the Cauchy Criterion is the Bolzano–Weierstrass (B–W) theorem. We therefore have the following chain of implications:

$$\text{Completeness} \Longrightarrow \text{MCT} \Longrightarrow \text{B-W} \Longrightarrow \text{Cauchy}$$

A close inspection of the Cauchy Criterion reveals that it is really a statement about the real numbers not having any gaps or holes. In fact, the same can be said about the MCT and the Bolzano-Weierstrass theorem. Regarding the Cauchy Criterion, if (x_n) is a Cauchy sequence then the terms of (x_n) are clustering around a number and that number must be in $\mathbb{R}$ if $\mathbb{R}$ has no holes. It is natural to ask then if we could have used the Cauchy Criterion as our starting axiom (instead of the Completeness Axiom) and then prove the Completeness property, and then the MCT and the Bolzano-Weierstrass theorem. Unfortunately, the Cauchy Criterion is not enough and we also need to take as an axiom the Archimedean Property.

Theorem 3.6.8

Suppose that every Cauchy sequence in $\mathbb{R}$ converges to a number in $\mathbb{R}$ and the Archimedean Property holds in $\mathbb{R}$. Then $\mathbb{R}$ satisfies the Completeness property, that is, every non-empty bounded above subset of $\mathbb{R}$ has a least upper bound in $\mathbb{R}$.

Proof. Let $S \subset \mathbb{R}$ be a non-empty set that is bounded above. If u is an upper bound of S and $u \in S$ then u is the least upper bound of S

and since $u \in \mathbb{R}$ there is nothing to prove. Suppose then that no upper bound of S is an element of S. Let $a_1 \in S$ be arbitrary and let $b_1 \in \mathbb{R}$ be an upper bound of S. Then $a_1 < b_1$, and we set $M = b_1 - a_1 > 0$. Consider the mid-point $m_1 = \frac{a_1+b_1}{2}$ of the interval $[a_1, b_1]$. If m_1 is an upper bound of S then set $b_2 = m_1$ and set $a_2 = a_1$, otherwise set $b_2 = b_1$ and $a_2 = m_1$. In any case, we have $|a_2 - a_1| \leq \frac{M}{2}$, $|b_2 - b_1| \leq \frac{M}{2}$, and $|b_2 - a_2| = \frac{M}{2}$. Now consider the mid-point $m_2 = \frac{a_2+b_2}{2}$ of the interval $[a_2, b_2]$. If m_2 is an upper bound of S then set $b_3 = m_2$ and $a_3 = a_2$, otherwise set $b_3 = b_2$ and $a_3 = m_2$. In any case, we have $|a_3 - a_2| \leq \frac{M}{2^2}$, $|b_3 - b_2| \leq \frac{M}{2^2}$, and $|b_3 - a_3| = \frac{M}{2^2}$. By induction, there exists a sequence (a_n) such that a_n is not an upper bound of S and a sequence (b_n) such that b_n is an upper bound of S, and $|a_n - a_{n+1}| \leq \frac{M}{2^n}$, $|b_n - b_{n+1}| \leq \frac{M}{2^n}$, and $|b_n - a_n| = \frac{M}{2^{n-1}}$. By Exercise 3.6.6, and using the fact that $\lim_{n\to\infty} r^n = 0$ if $0 < r < 1$ (this is where the Archimedean property is needed), it follows that (a_n) and (b_n) are Cauchy sequences and therefore by assumption both (a_n) and (b_n) are convergent. Since $|b_n - a_n| = \frac{M}{2^{n-1}}$ it follows that $u = \lim a_n = \lim b_n$. We claim that u is the least upper bound of S. First of all, for fixed $x \in S$ we have that $x < b_n$ for all $n \in \mathbb{N}$ and therefore $x \leq \lim b_n$, that is, u is an upper bound of S. Since a_n is not an upper bound of S, there exists $x_n \in S$ such that $a_n < x_n < b_n$ and therefore by the Squeeze theorem we have $u = \lim x_n$. Given an arbitrary $\varepsilon > 0$ then there exists $K \in \mathbb{N}$ such that $u - \varepsilon < x_K$ and thus $u - \varepsilon$ is not an upper bound of S. This proves that u is the least upper bound of S. □

Exercises

Exercise 3.6.1. Show that if (x_n) is a Cauchy sequence then (x_n) is bounded. (Note: Do not use the fact that a Cauchy sequence converges but show directly that if (x_n) is Cauchy then (x_n) is bounded.)

Exercise 3.6.2. Show that if (x_n) converges then it is a Cauchy sequence.

Exercise 3.6.3. Show **by definition** that $x_n = \frac{n+1}{n}$ is a Cauchy sequence.

Exercise 3.6.4. Show **by definition** that $x_n = \frac{\cos(n^2+1)}{n}$ is a Cauchy sequence.

Exercise 3.6.5. Suppose that (x_n) and (y_n) are sequences such that $|x_m - x_n| \leq |y_m - y_n|$ for all $n, m \in \mathbb{N}$. Show that if the sequence (y_n) is convergent then so is the sequence (x_n).

Exercise 3.6.6. Suppose that $0 < r < 1$. Show that if the sequence (x_n) satisfies $|x_n - x_{n-1}| < r^{n-1}$ for all $n \geq 2$ then (x_n) is a Cauchy sequence and therefore convergent. Hint: If $m > n$ then

$$|x_m - x_n| = |x_m - x_{m-1} + x_{m-1} - x_{m-2} + x_{m-2} - \cdots + x_{n+1} - x_n|$$

Also, if $0 < r < 1$ then $\sum_{j=0}^{k} r^j = \frac{1-r^{k+1}}{1-r} < \frac{1}{1-r}$.

3.7 Infinite Series

Informally speaking, a series is an infinite sum:

$$x_1 + x_2 + x_3 + \cdots$$

Using summation notation:

$$\sum_{n=1}^{\infty} x_n = x_1 + x_2 + x_3 + \cdots$$

The series $\sum_{n=1}^{\infty} x_n$ can be thought of as the sum of the sequence $(x_n) = (x_1, x_2, x_3, \ldots)$. It is of course not possible to actually sum an infinite number of terms and so we need a precise way to talk about what it means for a series to have a finite value. Take for instance

$$\sum_{n=1}^{\infty} (\tfrac{2}{3})^{n-1} = 1 + (\tfrac{2}{3}) + (\tfrac{2}{3})^2 + (\tfrac{2}{3})^3 + \cdots$$

so that the sequence being summed is $x_n = (\frac{2}{3})^{n-1}$. Let's compute the first 10 terms of the sequence of partial sums $\{s_n\}_{n=1}^{\infty} = (s_1, s_2, s_3, s_4, s_5, \ldots)$ defined as follows:

$$\begin{aligned}
s_1 &= 1 \\
s_2 &= 1 + (\tfrac{2}{3}) = 1.6666 \\
s_3 &= 1 + (\tfrac{2}{3}) + (\tfrac{2}{3})^2 = 2.1111 \\
s_4 &= 1 + (\tfrac{2}{3}) + (\tfrac{2}{3})^2 + (\tfrac{2}{3})^3 = 2.4074 \\
s_5 &= 1 + (\tfrac{2}{3}) + (\tfrac{2}{3})^2 + (\tfrac{2}{3})^3 + (\tfrac{2}{3})^4 = 2.6049 \\
s_6 &= 1 + (\tfrac{2}{3}) + (\tfrac{2}{3})^2 + (\tfrac{2}{3})^3 + (\tfrac{2}{3})^4 + (\tfrac{2}{3})^5 = 2.7366 \\
s_7 &= 1 + (\tfrac{2}{3}) + (\tfrac{2}{3})^2 + (\tfrac{2}{3})^3 + (\tfrac{2}{3})^4 + (\tfrac{2}{3})^5 + (\tfrac{2}{3})^6 = 2.8244 \\
s_8 &= 1 + (\tfrac{2}{3}) + (\tfrac{2}{3})^2 + (\tfrac{2}{3})^3 + (\tfrac{2}{3})^4 + (\tfrac{2}{3})^5 + (\tfrac{2}{3})^6 + (\tfrac{2}{3})^7 = 2.8829 \\
s_9 &= 1 + (\tfrac{2}{3}) + (\tfrac{2}{3})^2 + (\tfrac{2}{3})^3 + (\tfrac{2}{3})^4 + (\tfrac{2}{3})^5 + (\tfrac{2}{3})^6 + (\tfrac{2}{3})^7 + (\tfrac{2}{3})^8 = 2.9219 \\
s_{10} &= 1 + (\tfrac{2}{3}) + (\tfrac{2}{3})^2 + (\tfrac{2}{3})^3 + (\tfrac{2}{3})^4 + (\tfrac{2}{3})^5 + (\tfrac{2}{3})^6 + (\tfrac{2}{3})^7 + (\tfrac{2}{3})^8 + (\tfrac{2}{3})^9 = 2.9479
\end{aligned}$$

With the help of a computer, one can compute

$$s_{20} = \sum_{k=1}^{20} (\tfrac{2}{3})^{k-1} = 2.999097813\ldots$$

$$s_{50} = \sum_{k=1}^{50} (\tfrac{2}{3})^{k-1} = 2.999999994\ldots$$

$$s_{100} = \sum_{k=1}^{100} (\tfrac{2}{3})^{k-1} = 2.999999998\ldots.$$

It seems as though the sequence (s_n) is converging to $L = 3$, that is, $\lim_{n\to\infty} s_n = 3$. It is then reasonable to say that the infinite series sums or converges to

$$\sum_{n=1}^{\infty} (\tfrac{2}{3})^{n-1} = 3 = \lim_{n\to\infty} s_n.$$

We now introduce some definitions to formalize our example.

Definition 3.7.1

Let (x_n) be a sequence. The **infinite series generated** by (x_n) is the sequence (s_n) defined by

$$s_n = x_1 + x_2 + \cdots + x_n$$

or recursively,

$$\begin{aligned} s_1 &= x_1 \\ s_{n+1} &= s_n + x_{n+1},\ n \geq 1. \end{aligned}$$

The sequence (s_n) is also called the **sequence of partials sums** generated by (x_n). The nth term of the sequence of partial sums (s_n) can instead be written using summation notation:

$$s_n = x_1 + x_2 + \cdots + x_n = \sum_{k=1}^{n} x_k.$$

Example 3.7.2. Let (x_n) be the sequence $x_n = \frac{3}{n}$. The first few terms of the sequence of partials (s_n) is

$$\begin{aligned} s_1 &= x_1 = 3 \\ s_2 &= x_1 + x_2 = 3 + \tfrac{3}{2} = \tfrac{9}{2} \\ s_3 &= x_1 + x_2 + x_3 = \tfrac{9}{2} + 1 = \tfrac{11}{2} \\ s_4 &= s_3 + x_4 = \tfrac{11}{2} + \tfrac{3}{4} = \tfrac{25}{4} \end{aligned}$$

In both examples above, we make the following important observation: if (x_n) is a sequence of non-negative terms then the sequence of partials sums (s_n) is increasing. Indeed, if $x_n \geq 0$ then

$$s_{n+1} = s_n + x_{n+1} \geq s_n$$

and thus $s_{n+1} \geq s_n$.

Example 3.7.3. Find the sequence of partial sums generated by $x_n = (-1)^n$.

Solution. We compute:

$$\begin{aligned} s_1 &= x_1 = -1 \\ s_2 &= x_1 + x_2 = -1 + 1 = 0 \\ s_3 &= s_2 + x_3 = 0 - 1 = -1 \end{aligned}$$

Hence, $(s_n) = (-1, 0, -1, 0, -1, 0, \ldots)$. □

The limit of the sequence (s_n), if it exists, makes precise what it means for an infinite series

$$\sum_{n=1}^{\infty} x_n = x_1 + x_2 + x_3 + \cdots$$

to converge.

Definition 3.7.4: Convergence of Series

Let (x_n) be a sequence and let (s_n) be the sequence of partial sums generated by (x_n). If $\lim_{n\to\infty} s_n$ exists and equals L then we say that the series generated by (x_n) **converges to** L and we write that

$$\sum_{n=1}^{\infty} x_n = L = \lim_{n\to\infty} s_n = \lim_{n\to\infty} \sum_{k=1}^{n} x_k.$$

The notation $\sum_{n=1}^{\infty} x_n$ is therefore a compact way of writing

$$\lim_{n\to\infty} \sum_{k=1}^{n} x_k,$$

and the question of whether the series $\sum_{n=1}^{\infty} x_n$ converges is really about whether the limit $\lim_{n\to\infty} s_n$ exists. Often we will write a series such as $\sum_{n=1}^{\infty} x_n$ simply as $\sum x_n$ when either the initial value of n is understood or is unimportant. Sometimes, the initial n value may be $n = 0$, $n = 2$, or some other $n = n_0$.

Example 3.7.5 (Geometric Series)**.** The geometric series is perhaps the most important series we will encounter. Let $x_n = r^n$ where $r \in \mathbb{R}$ is a constant. The generated series is

$$\sum_{n=0}^{\infty} x_n = \sum_{n=0}^{\infty} r^n = 1 + r + r^2 + r^3 + \cdots$$

and is called the geometric series. The nth term of the sequence of partial sums is

$$s_n = 1 + r + r^2 + \cdots + r^n.$$

If $r = 1$ then $\lim_{n\to\infty} s_n$ does not exist (why?), so suppose that $r \neq 1$. Using the fact that $(1-r)(1 + r + r^2 + \cdots + r^n) = \frac{1-r^{n+1}}{1-r}$ we can write

$$s_n = \frac{1 - r^{n+1}}{1-r}.$$

Now if $|r| < 1$ then $\lim_{n\to\infty} r^n = 0$, while if $|r| > 1$ then $\lim_{n\to\infty} r^n$ does not exist. Therefore, if $|r| < 1$ then

$$\lim_{n\to\infty} s_n = \lim_{n\to\infty} \frac{1 - r^{n+1}}{1-r} = \frac{1}{1-r}.$$

Therefore,

$$\sum_{n=0}^{\infty} r^n = \lim_{n\to\infty} s_n = \frac{1}{1-r}.$$

In summary: The series $\sum_{n=0}^{\infty} r^n$ is called the **geometric series** and converges if and only if $|r| < 1$ and in this case converges to $\frac{1}{1-r}$.

Example 3.7.6. Consider the series $\sum_{n=0}^{\infty} \frac{(-1)^n 2^n}{3^n}$. The series can be written as $\sum_{n=0}^{\infty} \left(\frac{-2}{3}\right)^n$ and thus it is a geometric series with $r = -\frac{2}{3}$. Since $|r| = |-\frac{2}{3}| < 1$, the series converges and it converges to

$$\sum_{n=0}^{\infty} \frac{(-1)^n 2^n}{3^n} = \frac{1}{1-(-2/3)} = \frac{3}{5}$$

Example 3.7.7. Use the geometric series to show that

$$0.999999\ldots = 1.$$

Solution. We can write

$$\begin{aligned}
0.999999\ldots &= 0.9 + 0.09 + 0.009 + \cdots \\
&= \frac{9}{10} + \frac{9}{100} + \frac{9}{1000} + \cdots \\
&= \frac{9}{10} + \frac{9}{10^2} + \frac{9}{10^3} + \cdots \\
&= \frac{9}{10}\left(1 + \frac{1}{10} + \frac{1}{10^2} + \cdots\right) \\
&= \frac{9}{10} \sum_{n=0}^{\infty} \frac{1}{10^n} \\
&= \frac{9}{10}\left(\frac{1}{1-\frac{1}{10}}\right) \\
&= 1.
\end{aligned}$$

□

Example 3.7.8 (Telescoping Series)**.** Consider $\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \frac{1}{3\cdot 4} + \cdots$. Using partial fraction decomposition

$$\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}.$$

Therefore, the nth term of the sequence of partial sums (s_n) is

$$s_n = \sum_{k=1}^{n} \frac{1}{k(k+1)} = \sum_{k=1}^{n} \left[\frac{1}{k} - \frac{1}{k+1}\right]$$

This is a **telescoping sum** because all terms in the middle cancel and only the first and last remain. For example:

$$\begin{aligned} s_1 &= 1 - \tfrac{1}{2} \\ s_2 &= 1 - \tfrac{1}{2} + \tfrac{1}{2} - \tfrac{1}{3} = 1 - \tfrac{1}{3} \\ s_3 &= 1 - \tfrac{1}{2} + \tfrac{1}{2} - \tfrac{1}{3} + \tfrac{1}{3} - \tfrac{1}{4} = 1 - \tfrac{1}{4}. \end{aligned}$$

By induction one can show that $s_n = 1 - \frac{1}{n+1}$ and therefore $\lim_{n\to\infty} s_n = 1$. Therefore, the given series converges and it converges to $\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \lim_{n\to\infty} s_n = 1$.

Example 3.7.9 (Harmonic Series)**.** Consider the series $\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots$. We are going to analyze a subsequence (s_{n_k}) of the sequence of partial sums (s_n). We will show that (s_{n_k}) is unbounded and thus (s_n) is unbounded and therefore divergent. Consequently, the series $\sum_{n=1}^{\infty} \frac{1}{n}$ is divergent. Consider

$$s_4 = 1 + \tfrac{1}{2} + \left(\tfrac{1}{3} + \tfrac{1}{4}\right) > 1 + \tfrac{1}{2} + \tfrac{1}{2} = 1 + 2\tfrac{1}{2}.$$

Now consider

$$s_8 = s_4 + \tfrac{1}{5} + \tfrac{1}{6} + \tfrac{1}{7} + \tfrac{1}{8} > 1 + 2\tfrac{1}{2} + 4\tfrac{1}{8} = 1 + 3\tfrac{1}{2}.$$

Lastly consider,

$$s_{16} = s_8 + \tfrac{1}{9} + \tfrac{1}{10} + \cdots + \tfrac{1}{16} > 1 + 3\tfrac{1}{2} + 8\tfrac{1}{16} = 1 + 4\tfrac{1}{2}.$$

In general, one an show by induction that for $k \geq 2$ we have

$$s_{2^k} > 1 + \frac{k}{2}.$$

Therefore, the subsequence $(s_4, s_8, s_{16}, \ldots)$ is unbounded and thus the series $\sum_{n=1}^{\infty} \frac{1}{n}$ is divergent.

We now present some basic theorems on the convergence of series; most of them are a direct consequence of results from limit theorems for sequences. The first theorem we present can be used to show that a series diverges.

Theorem 3.7.10: Series Divergence Test

If $\sum x_n$ converges then $\lim\limits_{n\to\infty} x_n = 0$. Equivalently, if $\lim\limits_{n\to\infty} x_n \neq 0$ then $\sum x_n$ diverges.

Proof. By definition, if $\sum x_n$ converges then the sequence of partial sums (s_n) is convergent. Suppose then that $L = \lim_{n\to\infty} s_n = \sum x_n$. Recall that (s_n) has the recursive definition $s_{n+1} = s_n + x_{n+1}$. The sequences (s_{n+1}) and (s_n) both converge to L and thus

$$\begin{aligned}
\lim_{n\to\infty} x_{n+1} &= \lim_{n\to\infty} (s_{n+1} - s_n) \\
&= L - L \\
&= 0.
\end{aligned}$$

□

Example 3.7.11. The series $\sum_{n=1}^{\infty} \frac{3n+1}{2n+5}$ is divergent because $\lim_{n\to\infty} \frac{3n+1}{2n+5} = \frac{3}{2} > 1$.

Example 3.7.12. The Series Divergence Test can only be used to show that a series diverges. For example, consider the harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$. Clearly $\lim_{n\to\infty} \frac{1}{n} = 0$. However, we already know that $\sum_{n=1}^{\infty} \frac{1}{n}$ is a divergent series. Hence, in general, the condition $\lim_{n\to\infty} x_n = 0$ is not sufficient to establish convergence of the series $\sum_{n=1}^{\infty} x_n$.

Example 3.7.13. A certain series $\sum_{k=1}^{\infty} x_k$ has sequence of partial sums (s_n) whose general nth term is $s_n = \frac{2n}{n+1}$.

(a) What is $\sum_{k=1}^{10} x_k$?

(b) Does the series $\sum_{k=1}^{\infty} x_k$ converge? If yes, what does it converge to? Explain.

(c) Does the sequence (x_n) converge? If yes, what does it converge to? Explain.

The next theorem is just an application of the Cauchy criterion for convergence of sequences to the sequence of partial sums (s_n).

Theorem 3.7.14: Cauchy Criterion

The series $\sum x_n$ converges if and only if for every $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that if $m > n \geq K$ then

$$|s_m - s_n| = |x_{n+1} + x_{n+2} + \cdots + x_m| < \varepsilon.$$

The following theorem is very useful and is a direct application of the Monotone convergence theorem.

Theorem 3.7.15

Suppose that (x_n) is a sequence of non-negative terms, that is, $x_n \geq 0$ for all $n \in \mathbb{N}$. Then $\sum x_n$ converges if and only if (s_n) is bounded. In this case,

$$\sum_{n=1}^{\infty} x_n = \sup\{s_n \mid n \geq 1\}.$$

Proof. Clearly, if $\sum x_n = \lim_{n\to\infty} s_n$ exists then (s_n) is bounded. Now suppose that (s_n) is bounded. Since $x_n \geq 0$ for all $n \in \mathbb{N}$ then $s_{n+1} = s_n + x_{n+1} \geq s_n$ and thus $s_{n+1} \geq s_n$ shows that (s_n) is an increasing sequence. By the Monotone convergence theorem, (s_n) converges and $\lim_{n\to\infty} s_n = \sup\{s_n \mid n \geq 1\}$. □

Example 3.7.16. Consider the series $\sum_{n=1}^{\infty} \frac{1}{n^2}$. Since $\frac{1}{n^2} > 0$, to prove that the series converges it is enough to show that the sequence of partial sums (s_n) is bounded. We will consider a subsequence of (s_n), namely, (s_{n_k}) where $n_k = 2^k - 1$ for $k \geq 2$. We have

$$\begin{aligned} s_3 &= 1 + \frac{1}{2^2} + \frac{1}{3^2} \\ &< 1 + \frac{1}{2^2} + \frac{1}{2^2} \\ &= 1 + \frac{1}{2} \end{aligned}$$

and

$$\begin{aligned} s_7 &= s_3 + \frac{1}{4^2} + \frac{1}{5^2} + \frac{1}{6^2} + \frac{1}{7^2} \\ &< 1 + \frac{1}{2} + \frac{4}{4^2} \\ &= 1 + \frac{1}{2} + \frac{1}{2^2}. \end{aligned}$$

By induction, one can show that

$$s_{n_k} < 1 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^{k-1}}$$

and therefore using the geometric series with $r = 1/2$ we have

$$\begin{aligned} s_{n_k} &< 1 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^{k-1}} \\ &< \sum_{n=0}^{\infty} \frac{1}{2^n} = 2. \end{aligned}$$

This shows that the subsequence (s_{n_k}) is bounded. In general, the existence of a bounded subsequence does not imply that the original sequence is bounded but if the original sequence is increasing then it does. In this case, (s_n) is indeed increasing, and thus since (s_{n_k}) is a bounded subsequence then (s_n) is also bounded. Therefore, (s_n) converges, that is, $\sum_{n=1}^{\infty} \frac{1}{n^2}$ is a convergent series.

Theorem 3.7.17

Suppose that $\sum x_n$ and $\sum y_n$ are convergent series.

(i) Then $\sum (x_n + y_n)$ and $\sum (x_n - y_n)$ are also convergent and

$$\sum (x_n \pm y_n) = \sum x_n \pm \sum y_n.$$

(ii) For any constant $c \in \mathbb{R}$, $\sum c x_n$ is also convergent and $\sum c x_n = c \sum x_n$.

Proof. Using the limit laws:

$$\begin{aligned}\lim_{n\to\infty}\sum_{k=1}^{n}(x_k \pm y_k) &= \lim_{n\to\infty}\left(\sum_{k=1}^{n} x_k \pm \sum_{k=1}^{n} y_k\right)\\ &= \lim_{n\to\infty}\sum_{k=1}^{n} x_k \pm \lim_{n\to\infty}\sum_{k=1}^{n} y_k\\ &= \sum_{n=1}^{\infty} x_n \pm \sum_{n=1}^{\infty} y_n.\end{aligned}$$

Hence,

$$\sum_{n=1}^{\infty}(x_n \pm y_n) = \sum_{n=1}^{\infty} x_n \pm \sum_{n=1}^{\infty} y_n.$$

If c is a constant then by the Limit Laws,

$$\lim_{n\to\infty}\sum_{k=1}^{n} cx_k = c\lim_{n\to\infty}\sum_{k=1}^{n} x_k = c\sum_{n=1}^{\infty} x_n.$$

Therefore,

$$\sum_{n=1}^{\infty} cx_n = c\sum_{n=1}^{\infty} x_n.$$

□

Once establishing the convergence/divergence of some sequences, we can use comparison tests to the determine convergence/divergence properties of new sequences.

Theorem 3.7.18: Comparison Test

Let (x_n) and (y_n) be non-negative sequences and suppose that $x_n \leq y_n$ for all $n \geq 1$.

(i) If $\sum y_n$ converges then $\sum x_n$ converges.

(ii) If $\sum x_n$ diverges then $\sum y_n$ diverges.

Proof. Let $t_n = \sum_{k=1}^n x_n$ and $s_n = \sum_{k=1}^n y_k$ be the sequences of partial sums. Since $x_n \leq y_n$ then $t_n \leq s_n$. To prove (a), if $\sum y_n$ converges then (s_n) is bounded. Thus, (t_n) is also bounded. Since (t_n) is increasing and bounded it is convergent by the MCT. To prove (b), if $\sum x_n$ diverges then (t_n) is necessarily unbounded and thus (s_n) is also unbounded and therefore (s_n) is divergent. □

Example 3.7.19. Let $p \geq 2$ be an integer and let (d_n) be a sequence of integers such that $0 \leq d_n \leq p-1$. Use the comparison test to show that the series $\sum_{n=1}^\infty \frac{d_n}{p^n}$ converges and converges to a point in $[0,1]$.

Example 3.7.20. Determine whether the given series converge.

(a) $\sum_{n=1}^\infty \frac{n^2}{3n^2+n}$: First compute $\lim_{n\to\infty} \frac{n^2}{3n^2+n} = 1/3$. Therefore, by the divergence test, the series diverges.

(b) $\sum_{n=1}^\infty \frac{1}{n^p}$ where $p \geq 2$: We know that $\sum_{n=1}^\infty \frac{1}{n^2}$ is convergent. Now, if $p \geq 2$ then $\frac{1}{n^p} \leq \frac{1}{n^2}$. Therefore by the Comparison test, $\sum_{n=1}^\infty \frac{1}{n^p}$ is also convergent. This is called a p-series and actually converges for any $p > 1$.

(c) $\sum_{n=1}^{\infty} \frac{n+7}{n^3+3}$: We will use the comparison test. We have

$$\begin{aligned} \frac{n+7}{n^3+3} &\leq \frac{n+7n}{n^3+3} \\ &< \frac{8n}{n^3} \\ &= \frac{8}{n^2}. \end{aligned}$$

The series $\sum_{n=1}^{\infty} \frac{8}{n^2}$ converges and thus the original series $\sum_{n=1}^{\infty} \frac{n+7}{n^3+3}$ also converges.

Example 3.7.21. Suppose that $0 \leq x_n \leq 1$ for all $n \in \mathbb{N}$. If $\sum x_n$ converges prove that $\sum (x_n)^2$ also converges. Does the claim hold if we only assume that $x_n > 0$?

Solution. Since $0 \leq x_n \leq 1$ then $0 \leq x_n^2 \leq x_n$. Since $\sum x_n$ converges then by the Comparison test then $\sum (x_n)^2$ also converges. More generally, suppose that $x_n > 0$ and $\sum x_n$ converges. Then (x_n) converges to zero and thus there exists $K \in \mathbb{N}$ such that $0 < x_n < 1$ for all $n \geq K$. Then since the series $\sum_{n=1}^{\infty} x_{n+K}$ converges it follows that $\sum_{n=1}^{\infty} x_{n+K}^2$ converges and consequently $\sum_{n=1}^{\infty} x_n^2$ converges. □

The following two tests can be used for series whose terms are not necessarily non-negative.

Theorem 3.7.22: Absolute Convergence

If the series $\sum |x_n|$ converges then $\sum x_n$ converges.

Proof. From $-|x_n| \leq x_n \leq |x_n|$ we obtain that

$$0 \leq x_n + |x_n| \leq 2|x_n|.$$

If $\sum |x_n|$ converges then so does $\sum 2|x_n|$. By the Comparison test 3.7.18, the series $\sum(x_n + |x_n|)$ converges also. Then the following series is a difference of two converging series and therefore converges:

$$\sum(x_n + |x_n|) - \sum |x_n| = \sum x_n$$

and the proof is complete. □

In the case that $\sum |x_n|$ converges we say that $\sum x_n$ **converges absolutely**. We end the section with the Ratio test for series.

Theorem 3.7.23: Ratio Test for Series

Consider the series $\sum x_n$ and let $L = \lim_{n\to\infty} \frac{|x_{n+1}|}{|x_n|}$. If $L < 1$ then the series $\sum x_n$ converges absolutely and if $L > 1$ then the series diverges. If $L = 1$ or the limit does not exist then the test is inconclusive.

Proof. Suppose that $L < 1$. Let $\varepsilon > 0$ be such that $r = L + \varepsilon < 1$. There exists $K \in \mathbb{N}$ such that $\frac{|x_{n+1}|}{|x_n|} < L + \varepsilon = r$ for all $n \geq K$ and thus $|x_{n+1}| < |x_n| r$ for all $n \geq K$. By induction, it follows that $|x_{K+m}| < |x_K| r^m$ for all $m \geq 1$. Since $\sum_{m=1}^{\infty} |x_k| r^m$ is a geometric series with $r < 1$ then, by the comparison test, the series $\sum_{m=1}^{\infty} |x_{K+m}|$ converges. Therefore, the series $\sum |x_n|$ converges and by the Absolute convergence criterion we conclude that $\sum x_n$ converges absolutely. If $L > 1$ then a similar argument shows that $\sum x_n$ diverges. The case $L = 1$ follows from the fact that some series converge and some diverge when $L = 1$. □

Exercises

Exercise 3.7.1. Suppose that $\sum x_n$ is a convergent series. Is it true that **if** $\sum y_n$ is divergent then $\sum(x_n + y_n)$ is divergent? If it is true, prove it, otherwise give an example to show that it is not true.

Exercise 3.7.2. Suppose that $x_n \geq 0$ for all $n \in \mathbb{N}$. Prove that if $\sum_{n=1}^{\infty} x_n$ converges then $\sum_{n=1}^{\infty} \frac{x_n}{n}$ also converges. Is the converse true? That is, if $\sum_{n=1}^{\infty} \frac{x_n}{n}$ converges then does it necessarily follow that $\sum_{n=1}^{\infty} x_n$ converges?

Exercise 3.7.3. Using only the tests derived in this section, determine whether the given series converge or diverge:

(a) $$\sum_{n=1}^{\infty} \frac{2n^2+3}{\sqrt{n^2+3n+2}}$$

(b) $$\sum_{n=1}^{\infty} \frac{\cos(n\pi)3^n}{2^n}$$

(c) $$\sum_{n=1}^{\infty} \frac{n-3}{n^3+1}$$

Exercise 3.7.4. Suppose that (x_n) and (y_n) are non-negative sequences. Prove that if $\sum x_n$ and $\sum y_n$ are convergent then $\sum x_n y_n$ is convergent. (Hint: Recall that if $z_n \geq 0$ then $\sum z_n$ converges iff (s_n) is bounded, where $s_n = \sum_{k=1}^{n} z_k$ is the sequence of partial sums. Alternatively, use the identity $(x+y)^2 = x^2 + 2xy + y^2$.)

Exercise 3.7.5. Suppose that $x_n > 0$ for all $n \in \mathbb{N}$ and $\sum_{n=1}^{\infty} x_n$ converges. You will prove that $\sum_{n=1}^{\infty} \sqrt{x_n x_{n+1}}$ converges.

(a) Let $y_n = x_n + x_{n+1}$. Prove that $\sum_{n=1}^{\infty} y_n$ converges.

(b) Using the fact that $(a+b)^2 = a^2+2ab+b^2$, prove that $\sqrt{ab} \leq a+b$ if $a, b > 0$. Deduce that

$$\sqrt{x_n x_{n+1}} \leq x_n + x_{n+1} = y_n.$$

(c) Deduce that $\sum_{n=1}^{\infty} \sqrt{x_n x_{n+1}}$ converges.

Exercise 3.7.6. Any number of the form $x = 0.a_1a_2a_3a_4\ldots$ can be written as

$$x = \frac{a_1}{10} + \frac{a_2}{10^2} + \frac{a_3}{10^3} + \frac{a_4}{10^4} + \cdots$$

Using this fact, and a geometric series, prove that

$$0.2555555555555555\ldots\ldots\ldots = \frac{23}{90}.$$

Exercise 3.7.7. Show that the following series converge and find their sum.

(a) $\displaystyle\sum_{n=0}^{\infty} \frac{(-1)^n}{e^{2n}}$

(b) $\displaystyle\sum_{j=2}^{\infty} \frac{3}{2^j}$

(c) $\displaystyle\sum_{k=-2}^{\infty} \frac{1}{3^k}$

Exercise 3.7.8. Using the fact that $2^{k-1} < k!$ for all $k \geq 3$, prove that the series $\sum_{n=0}^{\infty} \frac{1}{n!}$ converges.

Exercise 3.7.9. Let $X : \mathbb{N} \to \mathbb{R}$ be a decreasing sequence of non-negative terms. Let (s_n) be the sequence of partial sums of the series $\sum_{n=1}^{\infty} X(n)$, let $n_k = 2^k - 1$ for $k \in \mathbb{N}$, and consider the subsequence (s_{n_k}).

(a) Show by induction that

$$s_{n_k} < \sum_{n=0}^{k-1} 2^n X(2^n)$$

for all $k \in \mathbb{N}$.

(b) Conclude that if $\sum_{n=0}^{\infty} 2^n X(2^n)$ converges then $\sum_{n=1}^{\infty} X(n)$.

(Note: This is a generalization of Example 3.7.16.)

4

Limits of Functions

In this chapter, we study another notion of convergence that is surely familiar to the reader, namely, the limit of a function at a given point. After introducing the precise definition of the limit of a function, and working through some examples, we will relate limits of functions with limits of sequences resulting in the Sequential Criterion for Limits (Theorem 4.1.11). In this chapter, when not explicitly stated, the letter A will denote a subset of $\mathbb{R}$.

4.1 Limits of Functions

Before we can give the definition of the limit of a function, we need the notion of a *cluster* point of a set.

Definition 4.1.1: Cluster Point

A number $c \in \mathbb{R}$ is called a **cluster point** of A if for any given $\delta > 0$ there exists at least one point $x \in A$, with $x \neq c$, such that $|x - c| < \delta$.

Hence, c is a cluster point of A if there are points in A that are arbitrarily close to c. In general, a cluster point of A is not necessarily an element of A. Naturally, cluster points can be characterized using limits of sequences.

Lemma 4.1.2

A point c is a cluster point of A if and only if there exists a sequence (x_n) in A such that $x_n \neq c$ and $\lim_{n\to\infty} x_n = c$.

Proof. Let c be a cluster point of A and let $\delta_n = \frac{1}{n}$ for $n \in \mathbb{N}$. Then by definition of a cluster point, there exists $x_n \in A$, $x_n \neq c$, such that $|x_n - c| < \delta_n$. Since $\delta_n \to 0$ then $x_n \to c$.

To prove the converse, suppose that $x_n \to c$, $x_n \neq c$ and $x_n \in A$ for $n \in \mathbb{N}$. Then by convergence of (x_n) to c, for any $\delta > 0$ there exists $K \in \mathbb{N}$ such that $|x_K - c| < \delta$. Since $x = x_K \in A$, this proves that c is a cluster point of A. $\square$

Example 4.1.3. Below are some examples of cluster points for a given set:

- Consider the set $A = [0, 1]$. Every point $c \in A$ is a cluster point of A.

- On the other hand, for $A = (0, 1]$, the point $c = 0$ is a cluster point of A but does not belong to A.

- For $A = \left\{\frac{1}{n} \mid n \in \mathbb{N}\right\}$, the only cluster point of A is $c = 0$.

- A finite set does not have any cluster points.

- The set $A = \mathbb{N}$ has no cluster points.

- Consider the set $A = \mathbb{Q} \cap [0, 1]$. By the Density theorem, every point $c \in [0, 1]$ is a cluster point of A.

We now give the definition of the limit of a function $f : A \to \mathbb{R}$ at a cluster point c of A.

Definition 4.1.4: Limit of a Function

Consider a function $f : A \to \mathbb{R}$ and let c be a cluster point of A. We say that f **has a limit at** c, or **converges at** c, if there exists a number $L \in \mathbb{R}$ such that for any given $\varepsilon > 0$ there exists $\delta > 0$ such that if $x \in A$ and $0 < |x - c| < \delta$ then $|f(x) - L| < \varepsilon$. In this case, we write that

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

and we say that f **converges to** L **at c**, or that f **has limit** L **at** c. If f does not converge at c then we say that f **diverges** at c.

Another short-hand notation to denote that f converges to L at c is $f(x) \to L$ as $x \to c$.

By definition, if $\lim_{x \to c} f(x) = L$, then for any $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $x \in (c - \delta, c + \delta) \cap A$ not equal to c it holds that $f(x) \in (L - \varepsilon, L + \varepsilon)$.

Theorem 4.1.5: Uniqueness of Limits

A function $f : A \to \mathbb{R}$ can have at most one limit at c.

Proof. Suppose that $f(x) \to L$ and $f(x) \to L'$ as $x \to c$, and let $\varepsilon > 0$. Then there exists $\delta > 0$ such that $|f(x) - L| < \varepsilon/2$ and $|f(x) - L'| < \varepsilon/2$, for all $x \in A$ satisfying $0 < |x - c| < \delta$. Then if $0 < |x - c| < \delta$ then

$$\begin{aligned} |L - L'| &\le |f(x) - L| + |f(x) - L'| \\ &< \varepsilon/2 + \varepsilon/2 \\ &= \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, Theorem 2.2.7 implies that $L = L'$. ☐

Example 4.1.6. Consider the function $f(x) = 5x + 3$ with domain $A = \mathbb{R}$. Prove that

$$\lim_{x \to 2} f(x) = 13.$$

Proof. We begin by analyzing the quantity $|f(x) - 13|$:

$$\begin{aligned}|f(x) - 13| &= |5x + 3 - 13| \\ &= |5x - 10| \\ &= 5|x - 2|.\end{aligned}$$

Hence, if $0 < |x - 2| < \varepsilon/5$ then

$$\begin{aligned}|f(x) - 13| &= |5x + 3 - 13| \\ &= 5|x - 2| \\ &< 5(\varepsilon/5) \\ &= \varepsilon.\end{aligned}$$

Thus, given $\varepsilon > 0$ we let $\delta = \varepsilon/5$ and thus if $0 < |x - c| < \delta$ then $|f(x) - 13| < \varepsilon$. Thus, by definition, $\lim_{x \to 2} f(x) = 13$. ☐

Example 4.1.7. Consider the function $f(x) = \frac{x+1}{x^2+3}$ with domain $A = \mathbb{R}$. Prove that

$$\lim_{x \to 1} f(x) = \frac{1}{2}.$$

Proof. We have that

$$\begin{aligned}
|f(x) - \tfrac{1}{2}| &= \left|\frac{x+1}{x^2+3} - \frac{1}{2}\right| \\
&= \left|\frac{x^2-2x+1}{2(x^2+3)}\right| \\
&= \frac{|x-1|^2}{2(x^2+3)} \\
&< |x-1|^2.
\end{aligned}$$

Let $\varepsilon > 0$ be arbitrary and let $\delta = \sqrt{\varepsilon}$. Then if $0 < |x-1| < \delta$ then $|x-1|^2 < \delta^2 = \varepsilon$. Hence, if $0 < |x-1| < \delta$ then

$$\begin{aligned}
|f(x) - \tfrac{1}{2}| &= \left|\frac{x+1}{x^2+3} - \frac{1}{2}\right| \\
&< |x-1|^2 \\
&< \varepsilon.
\end{aligned}$$

Thus, by definition, $\lim_{x\to 1} f(x) = \frac{1}{2}$. □

Example 4.1.8. Consider the function $f(x) = x^2$ with domain $A = \mathbb{R}$. Prove that for any $c \in \mathbb{R}$,

$$\lim_{x\to c} f(x) = c^2.$$

Proof. We first note that

$$|f(x) - c^2| = |x^2 - c^2| = |x+c||x-c|.$$

By the triangle inequality, $|x+c| \leq |x| + |c|$ and therefore

$$\begin{aligned}
|f(x) - c^2| &= |x+c||x-c| \\
&\leq (|x| + |c|)|x-c|.
\end{aligned}$$

We now need to analyze how large $|x|$ can become when x is say within $\overline{\delta} > 0$ of c. To be concrete, suppose that $\overline{\delta} = 1/2$. Hence, if $0 < |x-c| < \overline{\delta}$ then

$$\begin{aligned} |x| &= |x - c + c| \\ &\leq |x - c| + |c| \\ &< \overline{\delta} + |c| \\ &< 1 + |c|. \end{aligned}$$

Therefore, if $0 < |x - c| < \overline{\delta}$ it holds that

$$\begin{aligned} |f(x) - c^2| &\leq (|x| + c)|x - c| \\ &< (1 + c)|x - c|. \end{aligned}$$

Now suppose that $\varepsilon > 0$ is arbitrary and let $\delta = \min\{\overline{\delta}, \frac{\varepsilon}{1+|c|}\}$. Then if $0 < |x - c| < \delta$ then $|x| < 1 + |c|$ and therefore

$$\begin{aligned} |f(x) - c^2| &= |x^2 - c^2| \\ &= |x + c||x - c| \\ &\leq (|x| + |c|) \cdot \delta \\ &< (1 + |c|) \cdot \frac{\varepsilon}{1 + |c|} \\ &= \varepsilon. \end{aligned}$$

This proves, by definition, that $\lim_{x\to c} x^2 = c^2$ for any $c \in \mathbb{R}$. □

Example 4.1.9. Consider the function $f(x) = \frac{x^2-3x}{x+3}$ with domain $A = \mathbb{R}\backslash\{-3\}$. Prove that

$$\lim_{x\to 6} f(x) = 2.$$

Proof. We first note that $c = -3$ is indeed a cluster point of $A = \mathbb{R}\backslash\{-3\}$. Now,

$$\begin{aligned} |f(x) - 2| &= \left|\frac{x^2 - 3x}{x+3} - 2\right| \\ &= \left|\frac{x^2 - 5x - 6}{x+3}\right| \\ &= \left|\frac{(x+1)(x-6)}{(x+3)}\right| \\ &= \frac{|x+1|}{|x+3|}|x-6|. \end{aligned}$$

We now obtain a bound for $\frac{|x+1|}{|x+3|}$ when x is close to 6. Suppose then that $|x - 6| < 1$. Then $5 < x < 7$ and therefore, $6 < x + 1 < 8$, which implies that $|x + 1| < 8$. Similarly, if $|x - 6| < 1$ then $8 < x + 3 < 10$ and therefore $8 < |x + 3|$, which implies that $\frac{1}{|x+3|} < \frac{1}{8}$. Therefore, if $|x - 6| < 1$ then

$$\frac{|x+1|}{|x+3|} < 8 \cdot \frac{1}{8} = 1.$$

Suppose now that $\varepsilon > 0$ is arbitrary and let $\delta = \min\{1, \varepsilon\}$. If $0 < |x - 6| < \delta$ then from our analysis above it follows that $\frac{|x+1|}{|x+3|} < 1$. Therefore, if $0 < |x - 6| < \delta$ then

$$\begin{aligned} |f(x) - 2| &= \left|\frac{x^2 - 3x}{x+3} - 2\right| \\ &= \frac{|x+1|}{|x+3|}|x-6| \\ &< 1 \cdot \delta \\ &\le \varepsilon. \end{aligned}$$

This proves that $\lim_{x\to 6} f(x) = 2$. $\square$

Example 4.1.10. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined as

$$f(x) = \begin{cases} (x-1)\arctan(x), & x \in \mathbb{Q} \\ \frac{3(x-1)}{1+x^2}, & x \notin \mathbb{Q}. \end{cases}$$

Prove that $\lim_{x\to 1} f(x) = 0$.

Proof. If $x \in \mathbb{Q}$ then

$$\begin{aligned} |f(x)| &= |(x-1)\arctan(x)| \\ &= |x-1||\arctan(x)| \\ &= |x-1| \cdot \tfrac{\pi}{2} \end{aligned}$$

and if $x \notin \mathbb{Q}$ then

$$\begin{aligned} |f(x)| &= \left|\frac{3(x-1)}{1+x^2}\right| \\ &= \frac{3|x-1|}{1+x^2} \\ &\leq 3|x-1|. \end{aligned}$$

Therefore, for all $x \in \mathbb{R}$ it holds that $|f(x)| \leq 3|x-1|$ since $\pi/2 < 3$. Thus, given $\varepsilon > 0$ let $\delta = \varepsilon/3$ and thus if $0 < |x-1| < \delta$ then

$$\begin{aligned} |f(x)| &\leq 3|x-1| \\ &< 3 \cdot \delta \\ &= 3 \cdot \varepsilon/3 \\ &= \varepsilon. \end{aligned}$$

This proves that $\lim_{x\to 1} f(x) = 0$. $\square$

The following important result states that limits of functions can be studied using limits of sequences.

Theorem 4.1.11: Sequential Criterion for Limits

Let $f : A \to \mathbb{R}$ be a function and let c be a cluster point of A. Then $\lim_{x\to c} f(x) = L$ if and only if for every sequence (x_n) in A converging to c (with $x_n \neq c$ for all $n \in \mathbb{N}$) the sequence $(f(x_n))$ converges to L.

Proof. Suppose that $\lim_{x\to c} f(x) = L$. Let (x_n) be a sequence in A converging to c, with $x_n \neq c$ for all $n \in \mathbb{N}$. We must prove that the sequence $(f(x_n))$ converges to L. To that end, let $\varepsilon > 0$ be arbitrary. Then, by convergence of f to L at c, there exists $\delta > 0$ such that if $0 < |x - c| < \delta$ then $|f(x) - L| < \varepsilon$. Now, since $(x_n) \to c$, there exists $K \in \mathbb{N}$ such that $|x_n - c| < \delta$ for all $n \geq K$. Therefore, for $n \geq K$ we have that $|f(x_n) - L| < \varepsilon$. This proves that $\lim_{n\to\infty} f(x_n) = L$.

To prove the converse, we prove the contrapositive. Hence, we must show that if f does not converge to L then there exists a sequence (x_n) in A (with $x_n \neq c$) converging to c but the sequence $(f(x_n))$ does not converge to L. Assume then that f does not converge to L. Then, negating the definition of the limit of a function, there exists $\varepsilon > 0$ such for all $\delta > 0$ there exists $x \in A$ such that $0 < |x - c| < \delta$ and $|f(x) - L| \geq \varepsilon$. Then, let $\delta_n = \frac{1}{n}$ for $n \in \mathbb{N}$. Then there exists $x_n \neq c$ such that $0 < |x_n - c| < \delta_n$ and $|f(x_n) - L| \geq \varepsilon$. Since $\delta_n \to 0$ then $(x_n) \to c$ but clearly $f(x_n)$ does not converge to L. This ends the proof. □

The following theorem follows immediately from Theorem 4.1.11.

Corollary 4.1.12

Let $f : A \to \mathbb{R}$ be a function and let c be a cluster point of A and let $L \in \mathbb{R}$. Then f does not converge to L at c if and only if there exists a sequence (x_n) in A converging c, with $x_n \neq c$ for all $n \in \mathbb{N}$, and such that $(f(x_n))$ does not converge to L.

Note that in Corollary 4.1.12, if the sequence $(f(x_n))$ diverges then by definition it does not converge to any $L \in \mathbb{R}$ and then f does not have a limit at c. When applicable, the following corollary is a useful tool to prove that a limit of a function does not exist.

Corollary 4.1.13

Let $f : A \to \mathbb{R}$ be a function and let c be a cluster point of A. Suppose that (x_n) and (y_n) are sequences in A converging to c, with $x_n \neq c$ and $y_n \neq c$ for all $n \in \mathbb{N}$. If $f(x_n)$ and $f(y_n)$ converge but

$$\lim_{n\to\infty} f(x_n) \neq \lim_{n\to\infty} f(y_n)$$

then f does not have a limit at c.

Example 4.1.14. Prove that $\lim_{x\to 0} \frac{1}{x}$ does not exist.

Proof. Consider $x_n = \frac{1}{n}$, which clearly converges to $c = 0$ and $x_n \neq 0$ for all $n \in \mathbb{N}$. Then $f(x_n) = n$ which is unbounded and thus does not converge. Thus, by Corollary 4.1.12, $\lim_{x\to 0} \frac{1}{x}$ does not exist. □

Example 4.1.15. Prove that $\lim_{x\to 0} \sin\left(\frac{1}{x}\right)$ does not exist.

Proof. Let $f(x) = \sin\left(\frac{1}{x}\right)$ with domain $A = \mathbb{R}\backslash\{0\}$. Consider the sequence $x_n = \frac{1}{\pi/2+n\pi}$. It is clear that $(x_n) \to 0$ and $x_n \neq 0$ for all $n \in \mathbb{N}$. Now $(f(x_n)) = (-1, 1, -1, -1, \ldots)$ and therefore $(f(x_n))$ does not converge. Therefore, f has no limit at $c = 0$. In fact, for each $\alpha \in [0, 2\pi)$, consider the sequence $x_n = \frac{1}{\alpha+2n\pi}$. Clearly $(x_n) \to 0$ and $x_n \neq 0$ for all $n \in \mathbb{N}$. Now, $f(x_n) = \sin(\alpha + 2n\pi) = \sin(\alpha)$. Hence, $(f(x_n))$ converges to $\sin(\alpha)$. This shows that f oscillates within the interval $[-1, -1]$ as x approaches $c = 0$. □

Example 4.1.16. The *sign* function, denoted by $\text{sgn} : \mathbb{R} \to \mathbb{R}$, is defined as

$$\text{sgn}(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases}$$

Prove that $\lim_{x \to 0} \text{sgn}(x)$ does not exist.

Proof. Consider the sequence $x_n = \frac{(-1)^n}{n}$. Then $(x_n) \to 0$ and $x_n \neq 0$ for all $n \in \mathbb{N}$. Now, $y_n = \text{sgn}(x_n) = (-1)^n$ and thus (y_n) does not converge. Therefore, by Corollary 4.1.12, the function sgn has no limit at $c = 0$. □

Exercises

Exercise 4.1.1. Use the definition of the limit of a function to prove that the following limits do indeed hold.

(a) $\displaystyle\lim_{x\to 3}\frac{2x+3}{4x-9}=3$

(b) $\displaystyle\lim_{x\to 6}\frac{x^2-3x}{x+3}=2$

(c) $\displaystyle\lim_{x\to 4}|x-3|=1$

Exercise 4.1.2. Let $A\subset\mathbb{R}$, let $f:A\to\mathbb{R}$, and suppose that c is a cluster point of A. Suppose that there exists a constant $K>0$ such that $|f(x)-L|\le K|x-c|$ for all $x\in A$. Prove that $\displaystyle\lim_{x\to c}f(x)=L$.

Exercise 4.1.3. Consider the function

$$f(x)=\begin{cases} x^2\sin(1/x), & x\in\mathbb{Q}\backslash\{0\}\\ \frac{x^2}{1+x^2}, & x\notin\mathbb{Q}.\end{cases}$$

Prove that $\lim_{x\to 0}f(x)=0$.

Exercise 4.1.4. Let $f:\mathbb{R}\to\mathbb{R}$ be defined as follows:

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

(a) Prove that f has a limit at $c=0$.

(b) Now suppose that $c\neq 0$. Prove that f has no limit at c.

(c) Define $g:\mathbb{R}\to\mathbb{R}$ by $g(x)=(f(x))^2$. Prove that g has a limit at any $c\in\mathbb{R}$.

Hint: The Density Theorem will be helpful for (b). In particular, the Density Theorem implies that for any point $c \in \mathbb{R}$, there exists a sequence (x_n) of rational numbers such that $(x_n) \to c$, and that there exists a sequence (y_n) of irrational numbers such that $(y_n) \to c$.

Exercise 4.1.5. Use any applicable theorem to explain why the following limits do not exist.

(a) $\lim_{x\to 0} \frac{1}{x^2}$

(b) $\lim_{x\to 0}(x + \operatorname{sgn}(x))$

(c) $\lim_{x\to 0} \sin(1/x^2)$

Recall that the function $\operatorname{sgn} : \mathbb{R} \to \mathbb{R}$ is defined as follows:

$$\operatorname{sgn}(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases}$$

4.2 Limit Theorems

In this section, we establish basic limit theorems for limits of functions. The reader should compare the results of this section with Section 3.2 where we established limit theorems for sequences. In fact, thanks to the sequential criterion for limits of functions (Theorem 4.1.11), all of the theorems in this section can be proved using limits of sequences.

To begin, we first show that if f has a limit at c then f satisfies a local boundedness property at c. Let us first define then what it means for a function to be *locally bounded* at a given point.

Definition 4.2.1: Local Boundedness

Consider a function $f : A \to \mathbb{R}$ and let c be a cluster point of A. We say that f is **bounded locally at** c if there exists $\delta > 0$ and $M > 0$ such that if $x \in (c - \delta, c + \delta) \cap A$ then $|f(x)| \leq M$.

Theorem 4.2.2

Consider a function $f : A \to \mathbb{R}$ and let c be a cluster point of A. If $\lim\limits_{x \to c} f(x)$ exists then f is bounded locally at c.

Proof. Let $L = \lim_{x\to c} f(x)$ and let $\varepsilon > 0$ be arbitrary. Then there exists $\delta > 0$ such that $|f(x) - L| < \varepsilon$ for all $x \in A$ such that $0 < |x - c| < \delta$. Therefore, for all $x \in A$ and $0 < |x - c| < \delta$ we have that

$$\begin{aligned} |f(x)| &= |f(x) - L + L| \\ &\leq |f(x) - L| + |L| \\ &< \varepsilon + |L|. \end{aligned}$$

If $c \in A$ then let $M = \max\{|f(c)|, \varepsilon + |L|\}$ and if $c \notin A$ then let

$M = \varepsilon + |L|$. Then $|f(x)| \leq M$ for all $x \in A$ such that $0 < |x - c| < \delta$, that is, f is bounded locally at c. □

Example 4.2.3. Consider the function $f(x) = \frac{1}{x}$ defined on the set $A = (0, \infty)$. Clearly, $c = 0$ is a cluster point of A. For any $\delta > 0$ and any $M > 0$ let $x \in A$ be such that $0 < x < \min\{\delta, \frac{1}{M}\}$. Then $0 < x < \frac{1}{M}$, that is, $M < \frac{1}{x} = f(x)$. Since M was arbitrary, this proves that f is unbounded at $c = 0$ and consequently f does not have a limit at $c = 0$.

We now state and prove some limit laws for functions. Let $f, g : A \to \mathbb{R}$ be functions and define the functions $(f + g)$, $(f - g)$, fg, and f/g on A as follows:

$$(f \pm g)(x) = f(x) \pm g(x)$$

$$(fg)(x) = f(x)g(x)$$

$$\left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)}$$

where for f/g we require that $g(x) \neq 0$ for all $x \in A$.

Theorem 4.2.4: Limit Laws

Let $f, g : A \to \mathbb{R}$ be functions and let c be a cluster point of A. Suppose that $\lim_{x\to c} f(x) = L$ and $\lim_{x\to c} g(x) = M$. Then

(i) $\lim\limits_{x\to c}(f \pm g)(x) = L \pm M$

(ii) $\lim\limits_{x\to c}(fg)(x) = LM$

(iii) $\lim\limits_{x\to c}\left(\frac{f}{g}\right)(x) = \frac{L}{M}$, if $M \neq 0$

The proofs are left as an exercises. (To prove the results, use the sequential criterion for limits and the limits laws for sequences).

Corollary 4.2.5

Let $f_1, \ldots, f_k : A \to \mathbb{R}$ be functions and let c be a cluster point of A. If $\lim_{x\to c} f_i(x)$ exists for each $i = 1, 2, \ldots, k$ then

(i) $\displaystyle \lim_{x\to c} \sum_{i=1}^{k} f_i(x) = \sum_{i=1}^{k} \lim_{x\to c} f_i(x)$

(ii) $\displaystyle \lim_{x\to c} \prod_{i=1}^{k} f_i(x) = \prod_{i=1}^{k} \lim_{x\to c} f_i(x)$

Example 4.2.6. If $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ is a polynomial function then $\lim_{x\to c} f(x) = f(c)$ for every $c \in \mathbb{R}$. If $g(x) = b_0 + b_1 x + b_2 x + \cdots + b_m x^m$ is another polynomial function and $g(x) \neq 0$ in a neighborhood of $x = c$ and $\lim_{x\to c} g(x) = g(c) \neq 0$ then

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

Example 4.2.7. Prove that $\displaystyle \lim_{x\to 2} \frac{x^2 - 4}{x - 2} = 4.$

Proof. We cannot use the Limit Laws directly since $\lim_{x\to 2}(x-2) = 0$. Instead, notice that if $x \neq 2$ then $\frac{x^2-4}{x-2} = x + 2$. Hence, the functions $f(x) = \frac{x^2-4}{x-2}$ and $g(x) = x + 2$ are equal at every point in $\mathbb{R}\backslash\{0\}$. It is clear that $\lim_{x\to 2} g(x) = 4$ and therefore it follows that also $\lim_{x\to 2} f(x) = 4$. □

Theorem 4.2.8

Let $f : A \to \mathbb{R}$ be a function and let c be a cluster point of A. Suppose that f has limit L at c. If $f(x) \geq 0$ for all $x \in A$ then $L \geq 0$.

Proof. We prove the contrapositive. Suppose then that $L < 0$. Let $\varepsilon > 0$ be such that $L + \varepsilon < 0$. Then since $\lim_{x\to c} f(x) = 0$, there exists $\delta > 0$ such that if $0 < |x - c| < \delta$ then $f(x) < L + \varepsilon < 0$. Hence, $f(x) < 0$ for some $x \in A$.

We give another proof using the sequential criterion for limits. To that end, if f converges to L at c then for any sequence (x_n) converging to c, $x_n \neq 0$, we have that $f(x_n) \to L$. Now $f(x_n) \geq 0$ and therefore $L \geq 0$ from our results on limits of sequences (Theorem 3.2.7). □

Theorem 4.2.9

Let $f : A \to \mathbb{R}$ be a function and let c be a cluster point of A. Suppose that $M_1 \leq f(x) \leq M_2$ for all $x \in A$ and suppose that $\lim_{x\to c} f(x) = L$. Then $M_1 \leq L \leq M_2$.

Proof. We have that $0 \leq f(x) - M_1$ and therefore by Theorem 4.2.8 we have that $0 \leq L - M_1$. Similarly, from $0 \leq M_2 - f(x)$ we deduce that $0 \leq M_2 - L$. From this we conclude that $M_1 \leq L \leq M_2$. An alternative proof: Since $f \to L$ at c, for any sequence $(x_n) \to c$ with $x_n \neq 0$, we have that $f(x_n) \to L$. Clearly, $M_1 \leq f(x_n) \leq M_2$ and therefore $M_1 \leq L \leq M_2$ from our results on limits of sequences (Theorem 3.2.7). □

The following is the Squeeze Theorem for functions.

Theorem 4.2.10: Squeeze Theorem

Let $f, g, h : A \to \mathbb{R}$ be functions and let c be a cluster point of A. Suppose that $\lim\limits_{x\to c} g(x) = L$ and $\lim\limits_{x\to c} h(x) = L$. If $g(x) \leq f(x) \leq h(x)$ for all $x \in A$, $x \neq c$, then $\lim\limits_{x\to c} f(x) = L$.

Proof. Let (x_n) be a sequence in A converging to c with $x_n \neq c$ for all $n \in \mathbb{N}$. Then, by the sequential criterion,

$$L = \lim_{n\to\infty} g(x_n) = \lim_{n\to\infty} h(x_n).$$

By assumption, it holds that $g(x_n) \leq f(x_n) \leq h(x_n)$ for all $n \in \mathbb{N}$, and therefore by the Squeeze Theorem for sequences, we have that $\lim_{n\to\infty} f(x_n) = L$. This holds for every such sequence and therefore $\lim_{x\to c} f(x) = L$. ☐

Example 4.2.11. Let

$$f(x) = \begin{cases} x^2\sin(1/x), & x \in \mathbb{Q}\backslash\{0\} \\ x^2\cos(1/x), & x \notin \mathbb{Q} \\ 0, & x = 0. \end{cases}$$

Show that $\lim_{x\to 0} f(x) = 0$.

We end this section with the following theorem.

Theorem 4.2.12

Let $f : A \to \mathbb{R}$ be a function and let c be a cluster point of A. Suppose that $\lim_{x\to c} f(x) = L$. If $L > 0$ then there exists $\delta > 0$ such that $f(x) > 0$ for all $x \in (c - \delta, c + \delta)$, $x \neq c$.

Proof. Choose $\varepsilon > 0$ so that $L - \varepsilon > 0$, take for example $\varepsilon = L/2$. Then there exists $\delta > 0$ such that $L - \varepsilon < f(x) < L + \varepsilon$ for all

$x \in (c-\delta, c+\delta)$, $x \neq c$, and thus by transitivity it follows that $0 < f(x)$ for all $x \in (c-\delta, c+\delta)$, $x \neq c$. □

Exercises

Exercise 4.2.1. Let $f, g : A \to \mathbb{R}$ and suppose that $c \in \mathbb{R}$ is a cluster point of A. Suppose that at c, f converges to L and g converges to M. Prove that fg converges to LM at c in two ways: (1) using the definition of the limit of a function, and (2) using the sequential criterion for limits.

Exercise 4.2.2. Give an example of a set $A \subset \mathbb{R}$, a cluster point c of A, and two functions $f, g : A \to \mathbb{R}$ such that $\lim_{x\to c} f(x)g(x)$ exists but $\lim_{x\to c} f(x)$ does not exist.

Exercise 4.2.3. Give an example of a function $f : \mathbb{R} \to \mathbb{R}$ that is bounded locally at $c = 0$ but does not have a limit at $c = 0$. Your answer should not be in the form of a graph.

Exercise 4.2.4. Let $f : \mathbb{R} \to \mathbb{R}$ be a function that is bounded locally at c and suppose that $g : \mathbb{R} \to \mathbb{R}$ converges to $L = 0$ at c. Prove that $\lim_{x\to c} f(x)g(x) = 0$.

5

Continuity

Throughout this chapter, A is a non-empty subset of $\mathbb{R}$ and $f : A \to \mathbb{R}$ is a function.

5.1 Continuous Functions

Definition 5.1.1: Continuity

The function f is **continuous at** $c \in A$ if for any given $\varepsilon > 0$ there exists $\delta > 0$ such that if $x \in A$ and $|x - c| < \delta$ then $|f(x) - f(c)| < \varepsilon$. If f is not continuous at c then we say that f is **discontinuous at** c. The function f is **continuous on** A if f is continuous at every point in A.

Suppose that $c \in A$ is a cluster point of A and f is continuous at c. Then from the definition of continuity, $\lim_{x \to c} f(x)$ exists and equal to $f(c)$. If c is not a cluster point of A then there exists $\delta > 0$ such that $(c - \delta, c + \delta) \cap A = \{c\}$ and continuity of f at c is immediate. In either case, we see that f is continuous at c if and only if

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

The following is then immediate.

Theorem 5.1.2: Sequential Criterion for Continuity

The function f is continuous at $c \in A$ if and only if for every sequence (x_n) in A converging to c, $f(x_n)$ converges to $f(c)$:

$$\lim_{n\to\infty} f(x_n) = f(\lim_{x\to\infty} x_n) = f(c).$$

Notice that here (x_n) is allowed to take on the value c. The following is immediate.

Theorem 5.1.3

The function f is discontinuous at c if and only if there exists a sequence (x_n) in A converging to c but $f(x_n)$ does not converge to $f(c)$.

Example 5.1.4. If $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ is a polynomial function then $\lim_{x\to c} f(x) = f(c)$ for every $c \in \mathbb{R}$. Thus, f is continuous everywhere. If $g(x) = b_0 + b_1 x + b_2 x + \cdots + b_m x^m$ is another polynomial function and $g(c) \neq 0$ then

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

Hence, $h(x) = f(x)/g(x)$ is continuous at every c where g is non-zero.

Example 5.1.5. Determine the points of continuity of

$$f(x) = \begin{cases} \frac{1}{x}, & x \neq 0 \\ 0, & x = 0. \end{cases}$$

Solution. Suppose that $c \neq 0$. Then $\lim_{x\to c} f(x) = \frac{1}{c} = f(c)$. Hence, f is continuous at $c \in \mathbb{R}\backslash\{0\}$. Consider now $c = 0$. The sequence

$x_n = \frac{1}{n}$ converges to $c = 0$ but $f(x_n) = n$ does not converge. Hence, $\lim_{x\to 0} f(x)$ does not exist. Thus, even though $f(0) = 0$ is well-defined, f is discontinuous at $c = 0$. □

Example 5.1.6 (Dirichlet Function)**.** The following function was considered by Peter Dirichlet in 1829:

$$f(x) = \begin{cases} 1, & x \in \mathbb{Q} \\ 0, & x \in \mathbb{R}\backslash\mathbb{Q} \end{cases} \tag{5.1}$$

Prove that f is discontinuous everywhere.

Proof. Let c be irrational and let $\varepsilon = 1/2$. Then for all $\delta > 0$, there exists $x \in \mathbb{Q} \cap (c - \delta, c + \delta)$ (by the Density theorem) and therefore $|f(x) - f(c)| = 1 > \varepsilon$. Hence, f is discontinuous at c. A similar argument shows that f is discontinuous $c \in \mathbb{Q}$. Alternatively, if $c \in \mathbb{Q}$ then there exists a sequence of irrational numbers (x_n) converging to c. Now $f(x_n) = 0$ and $f(c) = 1$, and this proves that f is discontinuous at c. A similar arguments holds for c irrational. □

Example 5.1.7 (Thomae Function)**.** Let $A = \{x \in \mathbb{R} \ : \ x > 0\}$ and define $f : A \to \mathbb{R}$ as

$$f(x) = \begin{cases} 0, & x \in \mathbb{R}\backslash\mathbb{Q}, \\ \frac{1}{n}, & x = \frac{m}{n} \in \mathbb{Q}, \ \gcd(m, n) = 1, \ n \in \mathbb{N} \end{cases}$$

The graph of f is shown in Figure 5.1. Prove that f is continuous at every irrational number in A and is discontinuous at every rational number in A.

Proof. Let $c = \frac{m}{n} \in \mathbb{Q}$ with $\gcd(m, n) = 1$. There exists a sequence (x_n) of irrational numbers in A converging to c. Hence, $f(x_n) = 0$ while $f(c) = \frac{1}{n}$. This shows that f is discontinuous at c. Now let c be

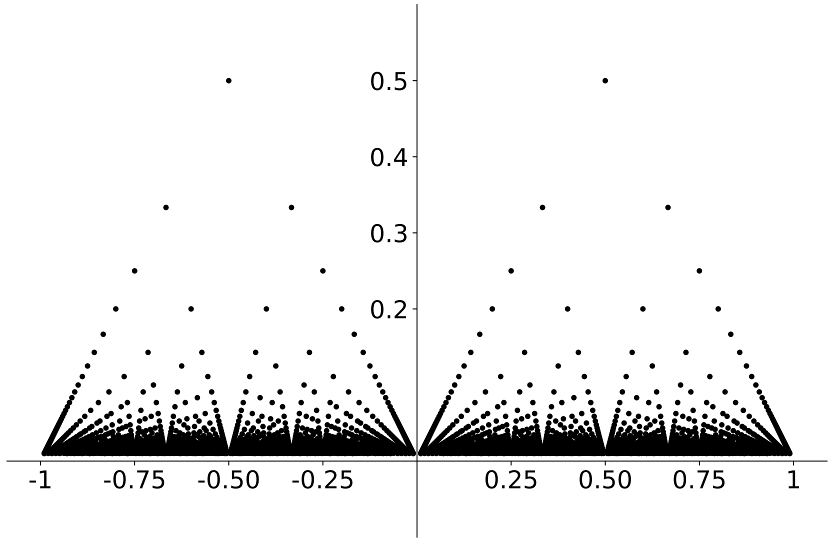


Figure 5.1: Thomae's function is continuous at each irrational and discontinuous at each rational

irrational and let $\varepsilon > 0$ be arbitrary. Let $N \in \mathbb{N}$ be such that $\frac{1}{N} < \varepsilon$. In the interval $(c-1, c+1)$, there are only a finite number of rationals $\frac{m}{n}$ with $n < N$, otherwise we can create a sequence $\frac{m_k}{n_k}$ with $n_k < N$, all the rationals $\frac{m_k}{n_k}$ distinct and thus necessarily $\frac{m_k}{n_k}$ is unbounded. Hence, there exists $\delta > 0$ such that the interval $(c-\delta, c+\delta)$ contains only rational numbers $x = \frac{m}{n}$ with $n > N$. Hence, if $x = \frac{m}{n} \in (c-\delta, c+\delta)$ then $f(x) = \frac{1}{n} < \frac{1}{N}$ and therefore $|f(x) - f(c)| = \frac{1}{n} < \frac{1}{N} < \varepsilon$. On the other hand, if $x \in (c-\delta, c+\delta)$ is irrational then $|f(x) - f(c)| = |0 - 0| < \varepsilon$. This proves that f is continuous at c. □

Suppose that f has a limit L at c but f is not defined at c. We can extend the definition of f by defining

$$F(x) = \begin{cases} f(x), & x \neq c \\ L, & x = c. \end{cases}$$

Now, $\lim_{x\to c} F(x) = \lim_{x\to c} f(x) = L = F(c)$, and thus F is continuous at c. Hence, functions that are not defined at a particular point c but have a limit at c can be extended to a function that is continuous at c. Points of discontinuity of this type are called **removal singularities**. On the other hand, the function $f(x) = \sin(1/x)$ is not defined at $c = 0$ and has not limit at $c = 0$, and therefore cannot be extended at $c = 0$ to a continuous function.

Exercises

Exercise 5.1.1. Let

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

Prove that f is continuous at $c = 0$.

Exercise 5.1.2. Let

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

Prove that f is discontinuous at $c = 0$.

Exercise 5.1.3. This is an interesting exercise.

(a) Suppose that $h : \mathbb{R} \to \mathbb{R}$ is continuous on $\mathbb{R}$ and that $h(r) = 0$ for every rational number $r \in \mathbb{Q}$. Prove that in fact $h(x) = 0$ for all $x \in \mathbb{R}$.

(b) Let $f, g : \mathbb{R} \to \mathbb{R}$ be continuous functions on $\mathbb{R}$ such that $f(r) = g(r)$ for every rational number $r \in \mathbb{Q}$. Prove that in fact $f(x) = g(x)$ for all $x \in \mathbb{R}$. Hint: Part (a) will be useful here.

Exercise 5.1.4. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a continuous function such that $f(p + q) = f(p) + f(q)$ for every $p, q \in \mathbb{Q}$. Prove that in fact $f(x + y) = f(x) + f(y)$ for every $x, y \in \mathbb{R}$.

5.2 Combinations of Continuous Functions

Not surprisingly, the set of continuous functions is closed under the basic operation of arithmetic.

Theorem 5.2.1

Let $f, g : A \to \mathbb{R}$ be continuous functions at $c \in A$ and let $b \in \mathbb{R}$. Then

(i) $f + g$, $f - g$, fg, and bf are continuous at c.

(ii) If $h : A \to \mathbb{R}$ is continuous at $c \in A$ and $h(x) \neq 0$ for all $x \in A$ then $\frac{f}{h}$ is continuous at c.

Proof. Let $\varepsilon > 0$ be arbitrary. By continuity of f and g at c, there exists $\delta_1 > 0$ such that $|f(x) - f(c)| < \varepsilon/2$ for all $x \in A$ such that $0 < |x - c| < \delta_1$, and there exists $\delta_2 > 0$ such that $|g(x) - g(c)| < \varepsilon/2$ for all $x \in A$ such that $0 < |x - c| < \delta_2$. Let $\delta = \min\{\delta_1, \delta_2\}$. Then for $x \in A$ such that $0 < |x - c| < \delta$ we have that

$$\begin{aligned} |f(x) + g(x) - (f(c) + g(c))| &\leq |f(x) - f(c)| + |g(x) - g(c)| \\ &< \varepsilon/2 + \varepsilon/2 \\ &= \varepsilon. \end{aligned}$$

This proves that $f + g$ is continuous at c. A similar proof holds for $f - g$.

Consider now the function bf. If $b = 0$ then $bf(x) = 0$ for all $x \in A$ and continuity is trivial. So assume that $b \neq 0$. Let $\varepsilon > 0$ be arbitrary. Then there exists $\delta > 0$ such that if $x \in A \cap (c - \delta, c + \delta)$, $x \neq c$, then $|f(x) - f(c)| < \varepsilon/(|b|)$. Therefore, for $x \in A \cap (c - \delta, c + \delta)$, $x \neq c$, we have that

$$\begin{aligned} |bf(x) - bf(c)| &= |b||f(x) - f(c)| \\ &< |b|\varepsilon/(|b|) \\ &= \varepsilon. \end{aligned}$$

We now prove continuity of fg. Let (x_n) be any sequence in A converging to c. Then $y_n = f(x_n)$ converges to $f(c)$ by continuity of f at c, and $z_n = g(x_n)$ converges to $g(c)$ by continuity of g at c. Hence the sequence $w_n = y_n z_n$ converges to $f(c)g(c)$. Hence, for every sequence (x_n) converging to c, $f(x_n)g(x_n)$ converges to $f(c)g(c)$. This shows that fg is continuous at c. □

Corollary 5.2.2

Let $f, g : A \to \mathbb{R}$ be continuous functions on A and let $b \in \mathbb{R}$. Then

(i) $f + g$, $f - g$, fg, and bf are continuous on A.

(ii) If $h : A \to \mathbb{R}$ is continuous on A and $h(x) \neq 0$ for all $x \in A$ then $\frac{f}{h}$ is continuous on A.

Example 5.2.3. Prove that $f(x) = x$ is continuous on $\mathbb{R}$.

Proof. Let $\varepsilon > 0$ be arbitrary. Let $\delta = \varepsilon$. If $0 < |x - c| < \delta$ then $|f(x) - f(c)| = |x - c| < \delta = \varepsilon$. □

Example 5.2.4. All polynomials $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ are continuous everywhere.

Example 5.2.5. Rational functions $f(x) = p(x)/q(x)$, with $q(x) \neq 0$ on $A \subset \mathbb{R}$, are continuous on A.

Lemma 5.2.6: Continuity Under Shifting

If $f : \mathbb{R} \to \mathbb{R}$ is continuous then $g(x) = f(x+\alpha)$ is continuous, where $\alpha \in \mathbb{R}$ is arbitrary.

Proof. Let $\varepsilon > 0$ be arbitrary. Then there exists $\delta > 0$ such that $|f(y) - f(d)| < \varepsilon$ for all $0 < |y - d| < \delta$. Therefore, if $0 < |x - c| = |(x+\alpha) - (c+\alpha)| < \delta$ then $|f(x+\alpha) - f(c+\alpha)| < \varepsilon$ and therefore

$$|g(x) - g(c)| = |f(x+\alpha) - f(c+\alpha)| < \varepsilon.$$

□

To prove continuity of $\sin(x)$ and $\cos(x)$ we use the following facts. For all $x \in \mathbb{R}$, $|\sin(x)| \leq |x|$ and $|\cos(x)| \leq 1$, and for all $x, y, \in \mathbb{R}$

$$\sin(x) - \sin(y) = 2\sin\left(\tfrac{1}{2}(x-y)\right)\cos\left(\tfrac{1}{2}(x+y)\right).$$

Example 5.2.7. Prove that $\sin(x)$ and $\cos(x)$ are continuous everywhere.

Proof. We have that

$$\begin{aligned} |\sin(x) - \sin(c)| &\leq 2|\sin(\tfrac{1}{2}(x-c))||\cos(\tfrac{1}{2}(x-c))| \\ &\leq 2\frac{1}{2}|x-c| \\ &= |x-c|. \end{aligned}$$

Hence given $\varepsilon > 0$ we choose $\delta = \varepsilon$. The proof that $\cos(x)$ is continuous follows from the fact that $\cos(x) = \sin(x + \pi/2)$ and Lemma 5.2.6. □

Example 5.2.8. The functions $\tan(x) = \frac{\sin(x)}{\cos(x)}$, $\cot(x) = \frac{\cos(x)}{\sin(x)}$, $\sec(x) = \frac{1}{\cos(x)}$, and $\csc(x) = \frac{1}{\sin(x)}$ are continuous on their domain.

Example 5.2.9. Prove that $f(x) = \sqrt{x}$ is continuous on $A = \{x \in \mathbb{R} \mid x \geq 0\}$.

Proof. For $c = 0$, we must consider $|\sqrt{x} - \sqrt{0}| = \sqrt{x}$. Given $\varepsilon > 0$ let $\delta = \varepsilon^2$. Then if $x \in A$ and $x < \delta = \varepsilon^2$ then $\sqrt{x} < \varepsilon$. This shows that f is continuous at $c = 0$. Now suppose that $c \neq 0$. Then

$$\begin{aligned} |\sqrt{x} - \sqrt{c}| &= |\sqrt{x} - \sqrt{c}| \cdot \frac{\sqrt{x} + \sqrt{c}}{\sqrt{x} + \sqrt{c}} \\ &= \frac{|x - c|}{\sqrt{x} + \sqrt{c}} \\ &\leq \frac{1}{\sqrt{c}} |x - c|. \end{aligned}$$

Hence, given $\varepsilon > 0$, suppose that $0 < |x - c| < \sqrt{c}\varepsilon$. Then $|\sqrt{x} - \sqrt{c}| < \varepsilon$. □

Example 5.2.10. Prove that $f(x) = |x|$ is continuous everywhere.

Proof. Follows from the inequality $||x| - |c|| \leq |x - c|$. □

The last theorem of this section is concerned with the composition of continuous functions.

Theorem 5.2.11: Continuity of Composite Functions

Let $f : A \to \mathbb{R}$ and let $g : B \to \mathbb{R}$ be continuous functions and suppose that $f(A) \subset B$. Then the composite function $(g \circ f) : A \to \mathbb{R}$ is continuous.

Proof. Let $\varepsilon > 0$ be given. Let $c \in A$ and let $d = f(c) \in B$. Then there exists $\delta_1 > 0$ such that if $0 < |y - d| < \delta_1$ then $|g(y) - g(d)| < \varepsilon$. Now since f is continuous at c, there exists $\delta_2 > 0$ such that if $0 < |x - c| < \delta_2$

then $|f(x)-f(c)| < \delta_1$. Therefore, if $0 < |x-c| < \delta_2$ then $|f(x)-d| < \delta_1$ and therefore $|g(f(x))-g(d)| < \varepsilon$. This proves that $(g\circ f)$ is continuous at $c \in A$. Since c is arbitrary, $(g \circ f)$ is continuous on A. $\square$

Corollary 5.2.12

If $f : A \to \mathbb{R}$ is continuous then $g(x) = |f(x)|$ is continuous. If $f(x) \geq 0$ for all $x \in A$ then $h(x) = \sqrt{f(x)}$ is continuous.

5.3 Continuity on Closed Intervals

In this section we develop properties of continuous functions on closed intervals.

Definition 5.3.1

We say that $f : A \to \mathbb{R}$ is **bounded on** A if there exists $M > 0$ such that $|f(x)| \leq M$ for all $x \in A$.

If f is not bounded on A then for any given $M > 0$ there exists $x \in A$ such that $|f(x)| > M$.

Example 5.3.2. Consider the function $f(x) = \frac{1}{x}$ defined on the interval $A = (0, \infty)$. Let $M > 0$ be arbitrary. Then if $0 < x < \frac{1}{M}$ then $f(x) = \frac{1}{x} > M$. For instance, take $x = \frac{1}{M+1}$. However, on the interval $[2, 3]$, f is bounded by $M = \frac{1}{2}$.

Theorem 5.3.3

Let $f : A \to \mathbb{R}$ be a continuous function. If $A = [a, b]$ is a closed and bounded interval then f is bounded on A.

Proof. Suppose that f is unbounded. Then for each $n \in \mathbb{N}$ there exists $x_n \in [a, b]$ such that $|f(x_n)| > n$. Now the sequence (x_n) is bounded since $a \leq x_n \leq b$. By the Bolzano-Weierstrass theorem, (x_n) has a convergent subsequence, say (x_{n_k}), whose limit $u = \lim_{k\to\infty} x_{n_k}$ satisfies $a \leq u \leq b$. Since f is continuous at u then $\lim_{k\to\infty} f(x_{n_k})$ exists and equal to $f(u)$. This is a contradiction since $|f(x_{n_k})| > n_k \geq k$ implies that $f(x_{n_k})$ is unbounded. □

Definition 5.3.4: Extrema of Functions

Let $f : A \to \mathbb{R}$ be a function.

(i) The function f has an **absolute maximum** on A if there exists $x^* \in A$ such that $f(x) \leq f(x^*)$ for all $x \in A$. We call x^* a **maximum point** and $f(x^*)$ the **maximum value** of f on A.

(ii) The function f has an **absolute minimum** on A if there exists $x_* \in A$ such that $f(x_*) \leq f(x)$ for all $x \in A$. We call x_* a **minimum point** and $f(x_*)$ the **minimum value** of f on A.

Suppose that $f : [a, b] \to \mathbb{R}$ is continuous. By Theorem 5.3.3, the range of f, that is $S = \{f(x) \mid x \in [a, b]\}$, is bounded and therefore $\inf(S)$ and $\sup(S)$ exist. In this case, we want to answer the question as to whether $\inf(S)$ and $\sup(S)$ are elements of S. In other words, as to whether f achieves its maximum and/or minimum value on $[a, b]$. That is, if there exists $x_*, x^* \in [a, b]$ such that $f(x_*) \leq f(x) \leq f(x^*)$ for all $x \in [a, b]$.

Example 5.3.5. The function $f(x) = \frac{1}{x}$ is continuous on $A = (0, 1]$.

However, f is unbounded on A and never achieves a maximum value on A.

Example 5.3.6. The function $f(x) = x^2$ is continuous on $[0, 2)$, is bounded on $[0, 2)$ but never reaches its maximum value on $[0, 2)$, that is, if $S = \{x^2 \,:\, x \in [0, 2)\}$ then $\sup(S) = 4 \notin S$.

Theorem 5.3.7: Extreme Value Theorem

Let $f : A \to \mathbb{R}$ be a continuous function. If $A = [a, b]$ is a closed and bounded interval then f has a maximum and minimum point on $[a, b]$.

Proof. Let $S = \{f(x) \mid x \in [a, b]\}$ be the range of f. By Theorem 5.3.3, $\sup(S)$ exists; set $M = \sup(S)$. By the definition of the supremum, for each $\varepsilon > 0$ there exists $x \in [a, b]$ such that $M - \varepsilon < f(x) \leq M$. In particular, for $\varepsilon_n = 1/n$, there exists $x_n \in [a, b]$ such that $M - \varepsilon_n < f(x_n) \leq M$. Then $\lim_{n\to\infty} f(x_n) = M$. The sequence (x_n) is bounded and thus has a convergent subsequence, say (x_{n_k}). Let $x^* = \lim_{k\to\infty} x_{n_k}$. Clearly, $a \leq x^* \leq b$. Since f is continuous at x^*, we have that $M = \lim_{k\to\infty} f(x_{n_k}) = f(x^*)$. Hence, x^* is a maximum point. A similar proof establishes that f has a minimum point on $[a, b]$. $\square$

By Theorem 5.3.7, we can replace $\sup\{f(x) \mid x \in [a, b]\}$ with $\max\{f(x) \mid x \in [a, b]\}$, and $\inf\{f(x) \mid x \in [a, b]\}$ with $\min\{f(x) \mid x \in [a, b]\}$. When the interval $[a, b]$ is clear from the context, we will simply write $\max(f)$ and $\min(f)$. The following example shows the importance of continuity in achieving a maximum/minimum.

Example 5.3.8. The function $f : [-1, 1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 3 - x^2, & 0 < x \leq 1 \\ x^2, & -1 \leq x \leq 0 \end{cases}$$

does not achieve a maximum value on the closed interval $[-1, 1]$.

The next theorem, called the Intermediate Value Theorem, is the main result of this section, and one of the most important results in this course.

Theorem 5.3.9: Intermediate Value Theorem

Consider the function $f : [a, b] \to \mathbb{R}$ and suppose that $f(a) < f(b)$. If f is continuous then for any $\xi \in \mathbb{R}$ such that $f(a) < \xi < f(b)$ there exists $c \in (a, b)$ such that $f(c) = \xi$.

Proof. Let $S = \{x \in [a, b] \ : \ f(x) < \xi\}$. The set S is non-empty because $a \in S$. Moreover, S is clearly bounded above. Let $c = \sup(S)$. By the definition of the supremum, there exists a sequence (x_n) in S such that $\lim_{n\to\infty} x_n = c$. Since $a \leq x_n \leq b$ it follows that $a \leq c \leq b$. By definition of x_n, $f(x_n) < \xi$ and since f is continuous at c we have that $f(c) = \lim_{n\to\infty} f(x_n) \leq \xi$, and thus $f(c) \leq \xi$. Now let $\delta_n > 0$ be such that $\delta_n \to 0$ and $c + \delta_n < b$. Then $z_n = c + \delta_n$ converges to c and $\xi \leq f(z_n)$ because $z_n \notin S$. Therefore, since $\lim_{n\to\infty} f(z_n) = f(c)$ we have that $\xi \leq f(c)$. Therefore, $f(c) \leq \xi \leq f(c)$ and this proves that $\xi = f(c)$. This shows also that $a < c < b$. □

The Intermediate Value Theorem has applications in finding points where a function is zero.

Corollary 5.3.10

Let $f : [a, b] \to \mathbb{R}$ be a function and suppose that $f(a)f(b) < 0$. If f is continuous then there exists $c \in (a, b)$ such that $f(c) = 0$.

Example 5.3.11. A hiker begins his climb at 7:00 am on a marked trail and arrives at the summit at 7:00 pm. The next day, the hiker begins his trek down the mountain at 7:00 am, takes the same trail down as he did going up, and arrives at the base at 7:00 pm. Use the Intermediate Value Theorem to show that there is a point on the path that the hiker crossed at exactly the same time of day on both days.

Proof. Let $f(t)$ be the distance traveled along the trail on the way up the mountain after t units of time, and let $g(t)$ be the distance remaining to travel along the trail on the way down the mountain after t units of time. Both f and g are defined on the same time interval, say $[0, 12]$ if t is measured in hours. If M is the length of the trail, then $f(0) = 0$, $f(12) = M$, $g(0) = M$ and $g(12) = 0$. Let $h(t) = g(t) - f(t)$. Then $h(0) = M$ and $h(12) = -M$. Hence, there exists $t^* \in (0, 12)$ such that $h(t^*) = 0$. In other words, $f(t^*) = g(t^*)$, and therefore t^* is the time of day when the hiker is at exactly the same point on the trail. ☐

Example 5.3.12. Prove by the Intermediate Value Theorem that $f(x) = xe^x - 2$ has a root in the interval $[0, 1]$.

Proof. The function f is continuous on $[0, 1]$. We have that $f(0) = -2 < 0$ and $f(1) = e - 2 > 0$. Therefore, there exists $x^* \in (0, 1)$ such that $f(x^*) = 0$, i.e., f has a zero in the interval $(0, 1)$. ☐

The next results says, roughly, that continuous functions preserve closed and bounded intervals. In the following theorem, we use the short-hand notation $f([a, b]) = \{f(x) \,|\, x \in [a, b]\}$ for the range of f under $[a, b]$.

Theorem 5.3.13

If $f : [a, b] \to \mathbb{R}$ is continuous then $f([a, b]) = [\min(f), \max(f)]$.

Proof. Since f achieves its maximum and minimum value on $[a, b]$, there exists $x^*, x_* \in [a, b]$ such that $f(x_*) \leq f(x) \leq f(x^*)$ for all $x \in [a, b]$.

Hence, $f([a,b]) \subset [f(x_*), f(x^*)]$. Assume for simplicity that $x_* < x^*$. Then $[x_*, x^*] \subset [a,b]$. Let $\xi \in \mathbb{R}$ be such that $f(x_*) < \xi < f(x^*)$. Then by the Intermediate Value Theorem, there exists $c \in (x_*, x^*)$ such that $\xi = f(c)$. Hence, $\xi \in f([a,b])$, and this shows that $[f(x_*), f(x^*)] \subset f([a,b])$. Therefore, $f([a,b]) = [f(x_*), f(x^*)] = [\min(f), \max(f)]$. □

It is worth noting that the previous theorem does not say that $f([a,b]) = [f(a), f(b)]$.

Exercises

Exercise 5.3.1. Let $f : A \to \mathbb{R}$ be any function. Show that if $-f$ achieves its maximum at $x_0 \in A$ then f achieves its minimum at x_0.

Exercise 5.3.2. Let f and g be continuous functions on $[a, b]$. Suppose that $f(a) \geq g(a)$ and $f(b) \leq g(b)$. Prove that $f(x_0) = g(x_0)$ for at least one x_0 in $[a, b]$.

Exercise 5.3.3. Let $f : [0, 1] \to \mathbb{R}$ be a continuous function and suppose that $f(x) \in [0, 1]$ for all $x \in [0, 1]$. Show that there exists $x_0 \in [0, 1]$ such that $f(x_0) = x_0$. Hint: Consider the function $g(x) = f(x) - x$ on the interval $[0, 1]$.

Exercise 5.3.4. Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Prove that if $f(x) \in \mathbb{Q}$ for all $x \in [a, b]$ then f is a constant function. **Hint:** You will need the Density Theorem and the Intermediate Value Theorem.

5.4 Uniform Continuity

In the definition of continuity of $f : A \to \mathbb{R}$ at $c \in A$, the δ will in general not only depend on ε but also on c. In other words, given two points c_1, c_2 and fixed $\varepsilon > 0$, the minimum δ_1 and δ_2 needed for c_1 and c_2 in the definition of continuity are generally going to be different. To see this, consider the continuous function $f(x) = x^2$. Then it is straightforward to verify that if $c^2 - \varepsilon > 0$ then $|x^2 - c^2| < \varepsilon$ if and only if

$$\sqrt{c^2 - \varepsilon} - c < x - c < \sqrt{c^2 + \varepsilon} - c.$$

Let $c_1 = 1$ and $c_2 = 3$, and let $\varepsilon = 1/2$. Then $|x - 1| < \delta_1$ implies that $|f(x) - f(1)| < \varepsilon$ if and only $\delta_1 \leq \sqrt{1 + \varepsilon} - 1 \approx 0.2247$. On the other hand, $|x - c| < \delta_2$ implies that $|f(x) - f(3)| < \varepsilon$ if and only if $\delta_2 \leq \sqrt{9 + \varepsilon} - 9 \approx 0.0822$. The reason that a smaller delta is needed at $c = 3$ is that the slope of f at $c = 3$ is larger than that at $c = 1$. On the other hand, consider the function $f(x) = \sin(2x)$. For any c it holds that

$$\begin{aligned} |f(x) - f(c)| &= |\sin(2x) - \sin(2c)| \\ &\leq 2|x - c|. \end{aligned}$$

Hence, given any $\varepsilon > 0$ we can set $\delta = \varepsilon/2$ and then $|x - c| < \delta$ implies that $|f(x) - f(c)| < \varepsilon$. The punchline is that $\delta = \varepsilon/2$ will work for *any* c. These motivating examples lead to the following definition.

Definition 5.4.1: Uniform Continuity

The function $f : A \to \mathbb{R}$ is said to be **uniformly continuous** on A if for each $\varepsilon > 0$ there exists $\delta > 0$ such that for all $x, u \in A$ satisfying $|x - u| < \delta$ it holds that $|f(x) - f(u)| < \varepsilon$.

Example 5.4.2. Let $k \neq 0$ be any non-zero constant. Show that $f(x) = kx$ is uniformly continuous on $\mathbb{R}$.

Proof. We have that $|f(x)-f(c)| = |kx-kc| = |k||x-c|$. Hence, for any $\varepsilon > 0$ we let $\delta = \varepsilon/|k|$, and thus if $|x-c| < \delta$ then $|f(x)-f(c)| < \varepsilon$. ☐

Example 5.4.3. Prove that $f(x) = \sin(x)$ is uniformly continuous.

Proof. We have that

$$\begin{aligned} |\sin(x) - \sin(c)| &\leq 2|\sin(\tfrac{1}{2}(x-c))| \\ &\leq |x-c|. \end{aligned}$$

Hence, for $\varepsilon > 0$ let $\delta = \varepsilon$ and if $|x-c| < \delta$ then $|\sin(x)-\sin(c)| < \varepsilon$. ☐

Example 5.4.4. Show that $f(x) = \dfrac{1}{1+x^2}$ is uniformly continuous on $\mathbb{R}$.

Proof. We have that

$$\begin{aligned} |f(x) - f(c)| &= \left|\frac{1}{1+x^2} - \frac{1}{1+c^2}\right| \\ &= \left|\frac{1+c^2-1-x^2}{(1+x^2)(1+c^2)}\right| \\ &= \left|\frac{x+c}{(1+x^2)(1+c^2)}\right| |x-c| \\ &= \left|\frac{x}{(1+x^2)(1+c^2)} + \frac{c}{(1+x^2)(1+c^2)}\right| |x-c|. \end{aligned}$$

Now, $|x| \leq 1 + x^2$ implies that $\frac{|x|}{1+x^2} \leq 1$ and therefore

$$\frac{|x|}{(1+x^2)(1+c^2)} \leq \frac{1}{1+c^2} \leq 1.$$

It follows that $|f(x) - f(c)| \leq 2|x - c|$. Hence, given $\varepsilon > 0$ we let $\delta = \varepsilon/2$, and if $|x - c| < \delta$ then $|f(x) - f(c)| < \varepsilon$. □

The following is a simple consequence of the definition of uniform continuity.

Theorem 5.4.5

Let $f : A \to \mathbb{R}$ be a function. The following are equivalent:

(i) The function f is not uniformly continuous on A.

(ii) There exists $\varepsilon_0 > 0$ such that for every $\delta > 0$ there exists $x, u \in A$ such that $|x - u| < \delta$ but $|f(x) - f(c)| \geq \varepsilon_0$.

(iii) There exists $\varepsilon_0 > 0$ and two sequences (x_n) and (u_n) such that $\lim_{n\to\infty}(x_n - u_n) = 0$ and $|f(x_n) - f(u_n)| \geq \varepsilon_0$.

Example 5.4.6. Let $f(x) = \frac{1}{x}$ and let $A = (0, \infty)$. Let $x_n = \frac{1}{n}$ and let $u_n = \frac{1}{n+1}$. Then $\lim(x_n - u_n) = 0$. Now $|f(x_n) - f(u_n)| = |-1| = 1$. Hence, if $\varepsilon = 1/2$ then $|f(x_n) - f(u_n)| > \varepsilon$. This proves that f is not uniformly continuous on $A = (0, \infty)$.

Theorem 5.4.7: Uniform Continuity on Intervals

Let $f : A \to \mathbb{R}$ be a continuous function with domain $A = [a, b]$. If f is continuous on A then f is uniformly continuous on A.

Proof. We prove the contrapositive, that is, we will prove that if f is not uniformly continuous on $[a, b]$ then f is not continuous on $[a, b]$. Suppose then that f is not uniformly continuous on $[a, b]$. Then there exists $\varepsilon > 0$ such that for $\delta_n = 1/n$, there exists $x_n, u_n \in [a, b]$ such that $|x_n - u_n| < \delta_n$ but $|f(x_n) - f(u_n)| \geq \varepsilon$. Clearly, $\lim(x_n - u_n) = 0$. Now, since $a \leq x_n \leq b$, by the Bolzano-Weierstrass theorem there is a subsequence (x_{n_k}) of (x_n) that converges to a point $z \in [a, b]$. Now,

$$|u_{n_k} - z| \leq |u_{n_k} - x_{n_k}| + |x_{n_k} - z|$$

and therefore also $\lim_{k\to\infty} u_{n_k} = z$. Hence, both (x_{n_k}) and (u_{n_k}) are sequences in $[a, b]$ converging to z but $|f(x_{n_k}) - f(u_{n_k})| \geq \varepsilon$. Hence $f(x_{n_k})$ and $f(u_{n_k})$ do not converge to the same limit and thus f is not continuous at z. This completes the proof. □

The following example shows that boundedness of a function does not imply uniform continuity.

Example 5.4.8. Show that $f(x) = \sin(x^2)$ is not uniformly continuous on $\mathbb{R}$.

Proof. Consider $x_n = \sqrt{\pi n}$ and $u_n = \sqrt{\pi n} + \frac{\sqrt{\pi}}{4\sqrt{n}}$. Clearly $\lim(x_n - u_n) = 0$. On the other hand

$$\begin{aligned}
|f(x_n) - f(u_n)| &= \left|\sin(\pi n) - \sin(\sqrt{\pi n} + \tfrac{\sqrt{\pi}}{4\sqrt{n}})^2\right| \\
&= \left|\sin\left(\pi n + \tfrac{\pi}{2} + \tfrac{\pi}{16n}\right)\right| \\
&= \left|\cos\left(\pi n + \tfrac{\pi}{16n}\right)\right| \\
&= \left|\cos(n\pi)\cos(\tfrac{\pi}{16n}) - \sin(n\pi)\sin(\tfrac{\pi}{16n})\right| \\
&= |(-1)^n \cos(\tfrac{\pi}{16n})| \\
&= \cos(\tfrac{\pi}{16n}) \\
&\geq \cos(\tfrac{\pi}{16})
\end{aligned}$$

□

The reason that $f(x) = \sin(x^2)$ is not uniformly continuous is that f is increasing rapidly on arbitrarily small intervals. Explicitly, it does not satisfy the following property.

Definition 5.4.9

A function $f : A \to \mathbb{R}$ is called a **Lipschitz function** on A if there exists a constant $K > 0$ such that $|f(x) - f(u)| \leq K|x - u|$ for all $x, u \in A$.

Suppose that f is Lipschitz with Lipschitz constant $K > 0$. Then for all $x, u \in A$ we have that

$$\left| \frac{f(x) - f(u)}{x - u} \right| \leq K.$$

Hence, the secant line through the points $(x, f(x))$ and $(u, f(u))$ has slope no larger than K in magnitude. Hence, a Lipschitz function has a constraint on how quickly it can change (measured by $|f(x) - f(u)|$) relative to the change in its inputs (measured by $|x - u|$).

Example 5.4.10. Since $|\sin(x) - \sin(u)| \leq |x - u|$, $f(x) = \sin(x)$ is a Lipschitz function on $\mathbb{R}$ with constant $K = 1$.

Example 5.4.11. If $a \neq 0$, the function $f(x) = ax + b$ is Lipschitz on $\mathbb{R}$ with constant $K = |a|$. When $a = 0$, f is clearly Lipschitz with arbitrary constant $K > 0$.

Theorem 5.4.12: Lipschitz and Uniform Continuity

If $f : A \to \mathbb{R}$ is a Lipschitz function on A then f is uniformly continuous on A.

Proof. By assumption, $|f(x) - f(u)| \leq K|x - u|$ for all $x, u \in A$ for some constant $K > 0$. Let $\varepsilon > 0$ be arbitrary and let $\delta = \varepsilon/K$. Then if $|x-u| < \delta$ then $|f(x)-f(u)| \leq K|x-u| < \varepsilon$. Hence, f is uniformly continuous on A. □

The following example shows that a uniformly continuous function is not necessarily Lipschitzian.

Example 5.4.13. Consider the function $f(x) = \sqrt{x}$ defined on $A = [0, 2]$. Since f is continuous, f is uniformly continuous on $[0, 2]$. To show that f is not Lipschitzian on A, let $u = 0$ and consider the inequality $|f(x)| = \sqrt{x} \leq K|x|$ for some $K > 0$. If $x \in [0, 2]$ then $\sqrt{x} \leq K|x|$ if and only if $x \leq K^2x^2$ if and only if $x(K^2x - 1) \geq 0$. If $x \in (0, 1/K^2) \cap A$, it holds that $K^2x - 1 < 0$, and thus no such K can exist. Thus, f is not Lipschitzian on $[0, 2]$.

Exercises

Exercise 5.4.1. Let $f : A \to \mathbb{R}$ and let $g : A \to \mathbb{R}$ be uniformly continuous functions on A. Prove that $f + g$ is uniformly continuous on A

Exercise 5.4.2. Let $f : A \to \mathbb{R}$ and let $g : A \to \mathbb{R}$ be Lipschitz functions on A. Prove that $f + g$ is a Lipschitz function on A.

Exercise 5.4.3. Give an example of a function that is uniformly continuous on $\mathbb{R}$ but is not bounded on $\mathbb{R}$.

Exercise 5.4.4. Prove that $f(x) = x^2$ is not uniformly continuous on $\mathbb{R}$.

Exercise 5.4.5. Give an example of distinct functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ that are uniformly continuous on $\mathbb{R}$ but fg is not uniformly continuous on $\mathbb{R}$. Prove that your resulting function fg is indeed not uniformly continuous.

Exercise 5.4.6. A function $f : \mathbb{R} \to \mathbb{R}$ is said to be T-**periodic** on $\mathbb{R}$ if there exists a number $T > 0$ such that $f(x + T) = f(x)$ for all $x \in \mathbb{R}$. Prove that a T-periodic continuous function on $\mathbb{R}$ is bounded and uniformly continuous on $\mathbb{R}$. Hint: First consider f on the interval $[0, T]$.

6

Differentiation

6.1 The Derivative

We begin with the definition of the derivative of a function.

Definition 6.1.1: The Derivative

Let $I \subset \mathbb{R}$ be an interval and let $c \in I$. We say that $f : I \to \mathbb{R}$ **is differentiable at** c or **has a derivative at** c if

$$\lim_{x \to c} \frac{f(x) - f(c)}{x - c}$$

exists. We say that f is **differentiable** on I if f is differentiable at every point in I.

By definition, f has a derivative at c if there exists a number $L \in \mathbb{R}$ such that for every $\varepsilon > 0$ there exists $\delta > 0$ such that if $|x - c| < \delta$ then

$$\left| \frac{f(x) - f(c)}{x - c} - L \right| < \varepsilon.$$

If f is differentiable at c, we will denote $\lim_{x\to c} \dfrac{f(x)-f(c)}{x-c}$ by $f'(c)$, that is,

$$f'(c) = \lim_{x\to c} \frac{f(x)-f(c)}{x-c}.$$

The rule that sends c to the number $f'(c)$ defines a function on a possibly smaller subset $J \subset I$. The function $f' : J \to \mathbb{R}$ is called the **derivative** of f.

Example 6.1.2. Let $f(x) = 1/x$ for $x \in (0, \infty)$. Prove that $f'(x) = -\frac{1}{x^2}$.

Example 6.1.3. Let $f(x) = \sin(x)$ for $x \in \mathbb{R}$. Prove that $f'(x) = \cos(x)$.

Proof. Recall that

$$\sin(x) - \sin(c) = 2\sin\left(\tfrac{x-c}{2}\right)\cos\left(\tfrac{x+c}{2}\right)$$

and that $\lim_{x\to 0} \frac{\sin(x)}{x} = 1$. Therefore,

$$\begin{aligned}
\lim_{x\to c} \frac{\sin(x)-\sin(c)}{x-c} &= \lim_{x\to c} \frac{2\sin\left(\frac{x-c}{2}\right)\cos\left(\frac{x+c}{2}\right)}{x-c} \\
&= \lim_{x\to c} \left(\frac{\sin\left(\frac{x-c}{2}\right)}{\frac{x-c}{2}}\right)\cos\left(\tfrac{x+c}{2}\right) \\
&= 1\cdot\cos(c) = \cos(c).
\end{aligned}$$

Hence $f'(c) = \cos(c)$ for all c and thus $f'(x) = \cos(x)$. ☐

Example 6.1.4. Prove by definition that $f(x) = \frac{x}{1+x^2}$ is differentiable on $\mathbb{R}$.

Proof. We have that

$$\begin{aligned}\frac{f(x)-f(c)}{x-c} &= \frac{\frac{x}{1+x^2}-\frac{c}{1+c^2}}{x-c}\\ &= \frac{x(1+c^2)-c(1+x^2)}{(1+x^2)(1+c^2)(x-c)}\\ &= \frac{1-cx}{(1+c^2)(1+x^2)}.\end{aligned}$$

Now

$$\lim_{x\to c}\frac{f(x)-f(c)}{x-c} = \frac{1-c^2}{(1+c^2)^2}.$$

Hence, $f'(c)$ exists for all $c\in\mathbb{R}$ and the derivative function of f is

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

□

Example 6.1.5. Prove that $f'(x)=\alpha$ if $f(x)=\alpha x+b$.

Proof. We have that $f(x)-f(c)=\alpha x-\alpha c=\alpha(x-c)$. Therefore, $\lim_{x\to c}\frac{f(x)-f(c)}{x-c}=\alpha$. This proves that $f'(x)=\alpha$ for all $x\in\mathbb{R}$. □

Example 6.1.6. Compute the derivative function of $f(x)=|x|$ for $x\in\mathbb{R}$.

Solution. If $x>0$ then $f(x)=x$ and thus $f'(x)=1$ for $x>0$. If $x<0$ then $f(x)=-x$ and therefore $f'(x)=-1$ for $x<0$. Now consider $c=0$. We have that

$$\frac{f(x)-f(c)}{x-c} = \frac{|x|}{x}.$$

We claim that the limit $\lim_{x\to 0}\frac{|x|}{x}$ does not exist and thus $f'(0)$ does not exist. To see this, consider $x_n=1/n$. Then $(x_n)\to 0$ and $f(x_n)=1$

for all n. On the other hand, consider $y_n = -1/n$. Then $(y_n) \to 0$ and $f(y_n) = -1$. Hence, $\lim_{n\to\infty} f(x_n) \neq \lim_{n\to\infty} f(y_n)$, and thus the claim holds by the Sequential criterion for limits. The derivative function f' of f is therefore defined on $A = \mathbb{R}\backslash\{0\}$ and is given by

$$f'(x) = \begin{cases} 1, & x > 0 \\ -1, & x > 0. \end{cases}$$

Hence, even though f is continuous at every point in its domain $\mathbb{R}$, it is not differentiable at every point in its domain. In other words, continuity is not a sufficient condition for differentiability. □

Theorem 6.1.7: Differentiability Implies Continuity

Suppose that $f : I \to \mathbb{R}$ is differentiable at c. Then f is continuous at c.

Proof. To prove that f is continuous at c we must show that $\lim_{x\to c} f(x) = f(c)$. By assumption $\lim_{x\to c} \frac{f(x)-f(c)}{x-c} = f'(c)$ exists, and clearly $\lim_{x\to c}(x-c) = 0$. Hence we can apply the Limits laws and compute

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

and the proof is complete. □

Theorem 6.1.8: Combinations of Differentiable Functions

Let $f : I \to \mathbb{R}$ and $g : I \to \mathbb{R}$ be differentiable at $c \in I$. The following hold:

(i) If $\alpha \in \mathbb{R}$ then (αf) is differentiable and $(\alpha f)'(c) = \alpha f'(c)$.

(ii) $(f + g)$ is differentiable at c and $(f + g)'(c) = f'(c) + g'(c)$.

(iii) fg is differentiable at c and $(fg)'(c) = f'(c)g(c) + f(c)g'(c)$.

(iv) If $g(c) \neq 0$ then (f/g) is differentiable at c and

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

Proof. Parts (i) and (ii) are straightforward. We will prove only (iii) and (iv). For (iii), we have that

$$\begin{aligned}\frac{f(x)g(x) - f(c)g(c)}{x - c} &= \frac{f(x)g(x) - f(c)g(x) + f(c)g(x) - f(c)g(c)}{x - c} \\ &= \frac{f(x) - f(c)}{x - c}g(x) + f(c)\frac{g(x) - g(c)}{x - c}.\end{aligned}$$

Now $\lim_{x\to c} g(x) = g(c)$ because g is differentiable at c. Therefore,

$$\begin{aligned}\lim_{x\to c}\frac{f(x)g(x) - f(c)g(c)}{x - c} &= \lim_{x\to c}\frac{f(x) - f(c)}{x - c}g(x) + \lim_{x\to c} f(c)\frac{g(x) - g(c)}{x - c} \\ &= f'(c)g(c) + f(c)g'(c).\end{aligned}$$

To prove part (iv), since $g(c) \neq 0$, then there exist a δ-neighborhood

$J = (c - \delta, c + \delta)$ such that $g(x) \neq 0$ for all $x \in J$. If $x \in J$ then

$$\begin{aligned}
\frac{\frac{f(x)}{g(x)} - \frac{f(c)}{g(x)}}{x - c} &= \frac{f(x)g(c) - g(x)f(c)}{g(x)g(c)(x - c)} \\
&= \frac{f(x)g(c) - f(c)g(c) + f(c)g(c) - g(x)f(c)}{g(x)g(c)(x - c)} \\
&= \frac{\frac{f(x)g(c) - f(c)g(c)}{x - c} - \frac{f(c)g(x) - f(c)g(c)}{x - c}}{g(x)g(c)}
\end{aligned}$$

Since $g(c) \neq 0$, it follows that

$$\lim_{x \to c} \frac{\frac{f(x)}{g(x)} - \frac{f(c)}{g(x)}}{x - c} = \frac{f'(c)g(c) - f(c)g'(c)}{g(c)^2}$$

and the proof is complete. □

We now prove the Chain Rule.

Theorem 6.1.9: Chain Rule

Let $f : I \to \mathbb{R}$ and $g : J \to \mathbb{R}$ be functions such that $f(I) \subset J$ and let $c \in I$. If $f'(c)$ exists and $g'(f(c))$ exists then $(g \circ f)'(c)$ exists and $(g \circ f)'(c) = g'(f(c))f'(c)$.

Proof. Suppose that there exists a neighborhood of c where $f(x) \neq f(c)$. Otherwise, the composite function $(g \circ f)(x)$ is constant in a neighborhood of c, and then clearly differentiable at c. Consider the function $h : J \to \mathbb{R}$ defined by

$$h(y) = \begin{cases} \frac{g(y) - g(f(c))}{y - f(c)}, & y \neq f(c) \\ g'(f(c)), & y = f(c). \end{cases}$$

Now

$$\lim_{y\to f(c)} h(y) = \lim_{y\to f(c)} \frac{g(y)-g(f(c))}{y-c}$$

$$= g'(f(c))'$$

$$= h(f(c)).$$

Hence, h is differentiable at $f(c)$ and therefore h is at $f(c)$. Now,

$$\frac{g(f(x))-g(f(c))}{x-c} = h(f(x))\frac{f(x)-f(c)}{x-c}$$

and therefore

$$\begin{aligned}\lim_{x\to c} \frac{g(f(x))-g(f(c))}{x-c} &= \lim_{x\to c} h(f(x))\frac{f(x)-f(c)}{x-c}\\ &= h(f(c))f'(c)\\ &= g'(f(c))f'(c).\end{aligned}$$

Therefore, $(g\circ f)'(c) = g'(f(c))f'(c)$ as claimed. □

Example 6.1.10. Compute $f'(x)$ if

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

Where is $f'(x)$ continuous?

Solution. When $x \neq 0$, $f(x)$ is the composition and product of differentiable functions at x, and therefore f is differentiable at $x \neq 0$. For instance, on $A = \mathbb{R}\backslash\{0\}$, the functions $1/x$, $\sin(x)$ and x^2 are differentiable at every $x \in A$. Hence, if $x \neq 0$ we have that

$$f'(x) = 2x\sin(\tfrac{1}{x}) - \cos(\tfrac{1}{x}).$$

Consider now $c = 0$. If $f'(0)$ exists it is equal to

$$\lim_{x\to 0} \frac{f(x) - f(c)}{x - c} = \lim_{x\to 0} \frac{x^2 \sin(\frac{1}{x})}{x}$$
$$= \lim_{x\to 0} x \sin(\tfrac{1}{x}).$$

Using the Squeeze Theorem, we deduce that $f'(0) = 0$. Therefore,

$$f'(x) = \begin{cases} 2x\sin(\frac{1}{x}) - \cos(\frac{1}{x}), & x \neq 0 \\ 0, & x = 0. \end{cases}$$

From the above formula obtained for $f'(x)$, we observe that when $x \neq 0$ f' is continuous since it is the product/difference/composition of continuous functions. To determine continuity of f' at $x = 0$ consider $\lim_{x\to 0} f'(x)$. Consider the sequence $x_n = \frac{1}{n\pi}$, which clearly converges to $c = 0$. Now, $f'(x_n) = \frac{2}{n\pi}\sin(n\pi) - \cos(n\pi)$. Now, $\sin(n\pi) = 0$ for all n and therefore $f'(x_n) = -\cos(n\pi) = (-1)^{n+1}$. The sequence $f'(x_n)$ does not converge and therefore $\lim_{x\to 0} f'(x)$ does not exist. Thus, f' is not continuous at $x = 0$. □

Example 6.1.11. Compute $f'(x)$ if

$$f(x) = \begin{cases} x^3 \sin(\frac{1}{x}), & x \neq 0 \\ 0, & x = 0 \end{cases}$$

Where is $f'(x)$ continuous?

Solution. When $x \neq 0$, $f(x)$ is the composition and product of differentiable functions, and therefore f is differentiable at $x \neq 0$. For instance, on $A = \mathbb{R}\backslash\{0\}$, the functions $1/x$, $\sin(x)$ and x^3 are differentiable on A. Hence, if $x \neq 0$ we have that

$$f'(x) = 3x^2\sin(\tfrac{1}{x}) - x\cos(\tfrac{1}{x}).$$

Consider now $c = 0$. If $f'(0)$ exists it is equal to

$$\begin{aligned}\lim_{x\to 0}\frac{f(x)-f(c)}{x-c} &= \lim_{x\to 0}\frac{x^3\sin(\frac{1}{x})}{x}\\ &= \lim_{x\to 0} x^2\sin(\tfrac{1}{x})\end{aligned}$$

and using the Squeeze Theorem we deduce that $f'(0) = 0$. Therefore,

$$f'(x) = \begin{cases} 3x^2\sin(\frac{1}{x}) - x\cos(\frac{1}{x}), & x \neq 0\\ 0, & x = 0.\end{cases}$$

When $x \neq 0$, f' is continuous since it is the product/difference/composition of continuous functions. To determine continuity of f' at $c = 0$ we consider the limit $\lim_{x\to 0} f'(x)$. Now $\lim_{x\to 0} 3x^2\sin(\frac{1}{x}) = 0$ using the Squeeze Theorem, and similarly $\lim_{x\to 0} x\cos(\frac{1}{x}) = 0$ using the Squeeze Theorem. Therefore, $\lim_{x\to 0} f'(x)$ exists and is equal to 0, which is equal to $f'(0)$. Hence, f' is continuous at $x = 0$, and thus continuous everywhere. □

Example 6.1.12. Consider the function

$$f(x) = \begin{cases} x^2\sin(\frac{1}{x}), & x \in \mathbb{Q}\backslash\{0\}\\ x^2\cos(\frac{1}{x}), & x \notin \mathbb{Q}\\ 0, & x = 0.\end{cases}$$

Show that $f'(0) = 0$.

Exercises

Exercise 6.1.1. Use the definition of the derivative of a function to find $f'(x)$ if $f(x) = \dfrac{3x+4}{2x-1}$. Clearly state the domain of $f'(x)$.

Exercise 6.1.2. Use the definition of the derivative of a function to find $f'(x)$ if $f(x) = x|x|$. Clearly state the domain of $f'(x)$.

Exercise 6.1.3. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} x^2, & x \in \mathbb{Q} \\ 0, & x \in \mathbb{R}\backslash\mathbb{Q} \end{cases}$$

1. Show that f is differentiable at $c = 0$ and find $f'(0)$.

2. Prove that if $c \neq 0$ then f is not differentiable at c.

Exercise 6.1.4. Let $g(x) = |x^3|$ for $x \in \mathbb{R}$. Determine whether $g'(0)$, $g^{(2)}(0), g^{(3)}(0)$ exist and if yes find them. Hint: Consider writing g as a piecewise function and use the definition of the derivative.

Exercise 6.1.5. If $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $c \in \mathbb{R}$, explain why

$$f'(c) = \lim_{n\to\infty} [n(f(c+1/n) - f(c))]$$

Give an example of a function f and a number c such that

$$\lim_{n\to\infty} [n(f(c+1/n) - f(c))]$$

exists but $f'(c)$ does not exist.

6.2 The Mean Value Theorem

Definition 6.2.1: Relative Extrema

Let $f : I \to \mathbb{R}$ be a function and let $c \in I$.

(i) We say that f has a **relative maximum** at c if there exists $\delta > 0$ such that $f(x) \leq f(c)$ for all $x \in (c - \delta, c + \delta)$.

(ii) We say that f has a **relative minimum** at c if there exists δ such that $f(c) \leq f(x)$ for all $x \in (c - \delta, c + \delta)$.

A point $c \in I$ is called a **critical point** of $f : I \to \mathbb{R}$ if $f'(c) = 0$. The next theorem says that a relative maximum/minimum of a differentiable function can only occur at a critical point.

Theorem 6.2.2: Critical Point at Extrema

Let $f : I \to \mathbb{R}$ be a function and let c be an interior point of I. Suppose that f has a relative maximum (or minimum) at c. If f is differentiable at c then c is a critical point of f, that is, $f'(c) = 0$.

Proof. Suppose that f has a relative maximum at c; the relative minimum case is similar. Then for $x \neq c$, it holds that $f(x) - f(c) \leq 0$ for $x \in (c - \delta, c + \delta)$ and some $\delta > 0$. Consider the function $h : (c-\delta, c+\delta) \to \mathbb{R}$ defined by $h(x) = \frac{f(x)-f(c)}{x-c}$ for $x \neq c$ and $h(c) = f'(c)$. Then the function h is continuous at $c = 0$ because $\lim_{x\to c} h(x) = h(c)$. Now for $x \in A = (c, c + \delta)$ it holds that $h(x) \leq 0$ and therefore $f'(c) = \lim_{x\to c} h(x) \leq 0$. Similarly, for $x \in B = (c - \delta, c)$ it holds that $h(x) \geq 0$ and therefore $0 \leq f'(c)$. Thus $f'(c) = 0$. $\square$

Corollary 6.2.3

If $f : I \to \mathbb{R}$ has a relative maximum (or minimum) at c then either $f'(c) = 0$ or $f'(c)$ does not exist.

Example 6.2.4. The function $f(x) = |x|$ has a relative minimum at $x = 0$, however, f is not differentiable at $x = 0$.

Theorem 6.2.5: Rolle

Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). If $f(a) = f(b)$ then there exists $c \in (a, b)$ such that $f'(c) = 0$.

Proof. Since f is continuous on $[a, b]$ it achieves its maximum and minimum at some point x^* and x_*, respectively, that is $f(x_*) \leq f(x) \leq f(x^*)$ for all $x \in [a, b]$. If f is constant then $f'(x) = 0$ for all $x \in (a, b)$. If f is not constant then $f(x_*) < f(x^*)$. Since $f(a) = f(b)$ it follows that at least one of x_* and x^* is not contained in $\{a, b\}$, and hence by Theorem 6.2.2 there exists $c \in \{x_*, x^*\}$ such that $f'(c) = 0$. □

We now state and prove the main result of this section.

Theorem 6.2.6: Mean Value

Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). Then there exists $c \in (a, b)$ such that $f'(c) = \frac{f(b)-f(a)}{b-a}$.

Proof. If $f(a) = f(b)$ then the result follows from Rolle's Theorem ($f'(c) = 0$ for some $c \in (a, b)$). Let $h : [a, b] \to \mathbb{R}$ be the line from $(a, f(a))$ to $(b, f(b)$, that is,

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

and define the function

$$g(x) = f(x) - h(x)$$

for $x \in [a, b]$. Then $g(a) = f(a) - f(a) = 0$ and $g(b) = f(b) - f(b) = 0$, and thus $g(a) = g(b)$. Clearly, g is continuous on $[a, b]$ and differentiable on (a, b), and it is straightforward to verify that $g'(x) = f'(x) - \frac{f(b)-f(a)}{b-a}$. By Rolle's Theorem, there exists $c \in (a, b)$ such that $g'(c) = 0$, and therefore $f'(c) = \frac{f(b)-f(a)}{b-a}$. □

Theorem 6.2.7

Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). If $f'(x) = 0$ for all $x \in (a, b)$ then f is constant on $[a, b]$.

Proof. Let $y \in (a, b]$. Now f restricted to $[a, y]$ satisfies all the assumptions needed in the Mean Value Theorem. Therefore, there exists $c \in (a, y)$ such that $f'(c) = \frac{f(y)-f(a)}{y-a}$. But $f'(c) = 0$ and thus $f(y) = f(a)$. This holds for all $y \in (a, b]$ and thus f is constant on $[a, b]$. □

Example 6.2.8. Show by example that Theorem 6.2.7 is not true for a function $f : A \to \mathbb{R}$ if A is not a closed and bounded interval.

Corollary 6.2.9

If $f, g : [a, b] \to \mathbb{R}$ are continuous and differentiable on (a, b) and $f'(x) = g'(x)$ for all $x \in (a, b)$ then $f(x) = g(x) + C$ for some constant C.

Example 6.2.10. Use the Mean Value theorem to show that $-x \leq \sin(x) \leq x$ for all $x \in \mathbb{R}$.

Proof. Suppose that $x > 0$ and let $g(x) = \sin(x)$ so that $g'(x) = \cos(x)$. By the MVT, there exists $c \in (0, x)$ such that $\cos(c) = \frac{\sin(x)}{x}$, that is $\sin(x) = x\cos(c)$. Now $|\cos(c)| \leq 1$ and therefore $|\sin(x)| \leq |x| = x$. Therefore, $-x \leq \sin(x) \leq x$. The case $x < 0$ can be treated similarly. ☐

Definition 6.2.11: Monotone Functions

The function $f : I \to \mathbb{R}$ is **increasing** if $f(x_1) \leq f(x_2)$ whenever $x_1 < x_2$. Similarly, f is **decreasing** if $f(x_2) \leq f(x_1)$ whenever $x_1 < x_2$. In either case, we say that f is **monotone**.

The sign of the derivative f' determines where f is increasing/decreasing.

Theorem 6.2.12

Suppose that $f : I \to \mathbb{R}$ is differentiable.

(i) Then f is increasing if and only if $f'(x) \geq 0$ for all $x \in I$.

(ii) Then f is decreasing if and only if $f'(x) \leq 0$ for all $x \in I$.

Proof. Suppose that f is increasing. Then for all $x, c \in I$ with $x \neq c$ it holds that $\frac{f(x)-f(c)}{x-c} \geq 0$ and therefore $f'(c) = \lim_{x\to c} \frac{f(x)-f(c)}{x-c} \geq 0$. Hence, this proves that $f'(x) \geq 0$ for all $x \in I$.

Now suppose that $f'(x) \geq 0$ for all $x \in I$. Suppose that $x < y$. Then by the Mean Value Theorem, there exists $c \in (x, y)$ such that $f'(c) = \frac{f(y)-f(x)}{y-x}$. Therefore, since $f'(c) \geq 0$ it follows that $f(y) - f(x) \geq 0$.

Part (ii) is proved similarly. ☐

Exercises

Exercise 6.2.1. Use the Mean Value Theorem to show that

$$|\cos(x) - \cos(y)| \leq |x - y|.$$

In general, suppose that $f : [a, b] \to \mathbb{R}$ is such that f' exists on $[a, b]$ and f' is continuous on $[a, b]$. Prove that f is Lipschitz on $[a, b]$.

Exercise 6.2.2. Give an example of a uniformly continuous function on $[0, 1]$ that is differentiable on $(0, 1)$ but whose derivative is not bounded on $(0, 1)$. Justify your answer.

Exercise 6.2.3. Let I be an interval and let $f : I \to \mathbb{R}$ be differentiable on I. Prove that if $f'(x) > 0$ for $x \in I$ then f is strictly increasing on I.

Exercise 6.2.4. Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). Show that if $\lim_{x \to a} f'(x) = A$ then $f'(a)$ exists and equals A. **Hint**: Use the definition of $f'(a)$, the Mean Value Theorem, and the Sequential Criterion for limits.

Exercise 6.2.5. Let $f : [a, b] \to \mathbb{R}$ be continuous and suppose that f' exists on (a, b). Prove that if $f'(x) > 0$ for $x \in (a, b)$ then f is strictly increasing on $[a, b]$.

Exercise 6.2.6. Suppose that $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b). We proved that if $f'(x) = 0$ for all $x \in (a, b)$ then f is constant on $[a, b]$. Give an example of a function $f : A \to \mathbb{R}$ such that $f'(x) = 0$ for all $x \in A$ but f is not constant on A.

Exercise 6.2.7. Let $f : [a, b] \to \mathbb{R}$ be differentiable. Prove that if $f'(x) \neq 0$ on $[a, b]$ then f is injective.

6.3 Taylor's Theorem

Taylor's theorem is a higher-order version of the Mean Value Theorem and it has abundant applications in numerical analysis. Taylor's theorem involves Taylor polynomials which you are familiar with from calculus.

Definition 6.3.1: Taylor Polynomials

Let $x_0 \in [a,b]$ and suppose that $f : [a,b] \to \mathbb{R}$ is such that the derivatives $f'(x_0)$, $f^{(2)}(x_0)$, $f^{(3)}(x_0)$,...,$f^{(n)}(x_0)$ exist for some positive integer n. Then the polynomial

$$P_n(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2!}f^{(2)}(x_0)(x - x_0)^2 + \cdots + \frac{1}{n!}f^{(n)}(x_0)(x - x_0)^n$$

is called the **nth order Taylor polynomial of f based at** x_0. Using summation convention, $P_n(x)$ can be written as

$$P_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k.$$

By construction, the derivatives of f and P_n up to order n are identical at x_0 (verify this!):

$$\begin{aligned} P_n(x_0) &= f(x_0) \\ P_n^{(1)}(x_0) &= f^{(1)}(x_0) \\ \vdots &= \vdots \\ P^{(n)}(x_0) &= f^{(n)}(x_0). \end{aligned}$$

It is reasonable then to suspect that $P_n(x)$ is a good approximation to $f(x)$ for points x near x_0. If $x \in [a,b]$ then the difference between $f(x)$

and $P_n(x)$ is

$$R_n(x) = f(x) - P_n(x)$$

and we call $R_n(x)$ the n**th order remainder based at** x_0. Hence, for each $x^* \in [a,b]$, the remainder $R_n(x^*)$ is the **error** in approximating $f(x^*)$ with $P_n(x^*)$. You may be asking yourself why we would need to approximate $f(x)$ if the function f is known and given. For example, if say $f(x) = \sin(x)$ then why would we need to approximate say $f(1) = \sin(1)$ since any basic calculator could easily compute $\sin(1)$? Well, what your calculator is actually computing is an approximation to $\sin(1)$ using a (rational) number such as $P_n(1)$ and using a large value of n for accuracy (although modern numerical algorithms for computing trigonometric functions have superseded Taylor approximations but Taylor approximations are a good start). Taylor's theorem provides an expression for the remainder term $R_n(x)$ using the derivative $f^{(n+1)}$.

Theorem 6.3.2: Taylor's Theorem

Let $f : [a,b] \to \mathbb{R}$ be a function such that for some $n \in \mathbb{N}$ the functions $f, f^{(1)}, f^{(2)}, \ldots, f^{(n)}$ are continuous on $[a,b]$ and $f^{(n+1)}$ exists on (a,b). Fix $x_0 \in [a,b]$. Then for any $x \in [a,b]$ there exists c between x_0 and x such that

$$f(x) = P_n(x) + R_n(x)$$

where

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x - x_0)^{n+1}.$$

Proof. If $x = x_0$ then $P_n(x_0) = f(x_0)$ and then c can be chosen arbitrarily. Thus, suppose that $x \neq x_0$, let $m = \frac{f(x)-P_n(x)}{(x-x_0)^{n+1}}$, and define the function $g : [a, b] \to \mathbb{R}$ by

$$g(t) = f(t) - P_n(t) - m(t - x_0)^{n+1}.$$

Since $f^{(n+1)}$ exists on (a, b) then $g^{(n+1)}$ exists on (a, b). Moreover, since $P^{(k)}(x_0) = f^{(k)}(x_0)$ for $k = 0, 1, \ldots, n$ then $g^{(k)}(x_0) = 0$ for $k = 0, 1, \ldots, n$. Now $g(x) = 0$ and therefore since $g(x_0) = 0$ by Rolle's theorem there exists c_1 in between x and x_0 such that $g'(c_1) = 0$. Now we can apply Rolle's theorem to g' since $g'(c_1) = 0$ and $g'(x_0) = 0$, and therefore there exists c_2 in between c_1 and x_0 such that $g''(c_2) = 0$. By applying this same argument repeatedly, there exists c in between x_0 and c_{n-1} such that $g^{(n+1)}(c) = 0$. Now,

$$g^{(n+1)}(t) = f^{(n+1)}(t) - m(n + 1)!$$

and since $g^{(n+1)}(c) = 0$ then

$$0 = f^{(n+1)}(c) - m(n + 1)!$$

from which we conclude that

$$f(x) - P(x) = \frac{f^{(n+1)}(c)}{(n + 1)!}(x - x_0)^{n+1}$$

and the proof is complete. □

Example 6.3.3. Consider the function $f : [0, 2] \to \mathbb{R}$ given by $f(x) = \ln(1 + x)$. Use P_4 based at $x_0 = 0$ to estimate $\ln(2)$ and give a bound on the error with your estimation.

Solution. Note that $f(1) = \ln(2)$ and so the estimate of $\ln(2)$ using P_4 is $\ln(2) \approx P_4(1)$. To determine P_4 we need $f(0), f^{(1)}(0), \ldots, f^{(4)}(0)$. We compute

$$\begin{aligned} f^{(1)}(x) &= \frac{1}{1+x} & f^{(1)}(0) &= 1 \\ f^{(2)}(x) &= \frac{-1}{(1+x)^2} & f^{(2)}(0) &= -1 \\ f^{(3)}(x) &= \frac{2}{(1+x)^3} & f^{(3)}(0) &= 2 \\ f^{(4)}(x) &= \frac{-6}{(1+x)^4} & f^{(4)}(0) &= -6. \end{aligned}$$

Therefore,

$$P_4(x) = x - \tfrac{1}{2}x^2 + \tfrac{1}{3}x^3 - \tfrac{1}{4}x^4.$$

Now $P_4(1) = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} = \frac{7}{12}$ and therefore

$$\ln(2) \approx P_4(1) = \frac{7}{12}.$$

The error is $R_4(1) = f(1) - P_4(1)$ which is unknown but we can approximate it using Taylor's theorem. To that end, by Taylor's theorem, for any $x \in [0, 2]$ there exists c in between $x_0 = 0$ and x such that

$$\begin{aligned} R_4(x) &= \frac{f^{(5)}(c)}{5!}x^5 \\ &= \frac{1}{5!}\frac{24}{(1+c)^5}x^4 \\ &= \frac{1}{5(1+c)^5}. \end{aligned}$$

Therefore, for $x = 1$, there exists $0 < c < 1$ such that

$$R_4(1) = \frac{1}{5(1+c)^5}.$$

Therefore, a bound for the error is

$$|R_4(1)| = \left|\frac{1}{5(1+c)^5}\right| \leq \frac{1}{5}$$

since $1 + c > 1$. □

Example 6.3.4. Let $f : \mathbb{R} \to \mathbb{R}$ be the sine function, that is, $f(x) = \sin(x)$.

(a) Approximate $f(3) = \sin(3)$ using P_8 centered at $x_0 = 0$ and give a bound on the error.

(b) Restrict f to a closed and bounded interval of the form $[-R, R]$. Show that for any $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that if $n \geq K$ then $|f(x) - P_n(x)| < \varepsilon$ **for all** $x \in [-a, a]$.

Solution. (a) It is straightforward to compute that

$$P_8(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7$$

and $f^{(9)}(x) = \sin(x)$. Thus, by Taylor's theorem for any x there exists c in between $x = 0$ and x such that

$$f(x) - P_8(x) = R_8(x) = \frac{\sin(c)}{9!}x^9.$$

The estimate for $f(3) = \sin(3)$ is

$$\begin{aligned}\sin(3) &\approx P_8(3)\\ &= 3 - \frac{1}{3!}3^3 + \frac{1}{5!}3^5 - \frac{1}{7!}3^7\\ &= \frac{51}{560}\\ &= 0.0910714286\end{aligned}$$

By Taylor's theorem, there exists c such that $0 < c < 3$ and

$$\sin(3) - P_8(3) = R_8(3) = \frac{\sin(c)}{9!}3^9$$

Now since $|\sin(c)| \leq 1$ for all $c \in \mathbb{R}$, we have

$$\begin{aligned} |R_8(3)| &= \frac{|\sin(c)|}{9!}3^9 \\ &= \frac{3^9}{9!} \\ &= 0.054241\ldots \end{aligned}$$

(b) Since $f(x) = \sin(x)$ has derivatives of all orders, for any $n \in \mathbb{N}$ we have by Taylor's theorem that

$$\begin{aligned} |f(x) - P_n(x)| &= |R_n(x)| \\ &= \left| \frac{f^{(n+1)}(c)}{(n+1)!} x^{n+1} \right| \\ &= \frac{|f^{(n+1)}(c)|}{(n+1)!} |x|^{n+1} \end{aligned}$$

where c is in between $x_0 = 0$ and x. Now, the derivative of $f(x) = \sin(x)$ of any order is one of $\pm\cos(x)$ or $\pm\sin(x)$, and therefore $|f^{(n+1}(c)| \leq 1$. Since $x \in [-a, a]$ then $|x| \leq a$ and therefore $|x|^{n+1} \leq a^{n+1}$. Therefore, for all $x \in [-a, a]$ we have

$$\begin{aligned} |R_n(x)| &\leq \frac{1}{(n+1)!} a^{n+1} \\ &= \frac{a^{n+1}}{(n+1)!}. \end{aligned}$$

Consider the sequence $x_n = \frac{a^n}{n!}$. Applying the Ratio test we obtain

$$\begin{aligned} \lim_{n\to\infty} \frac{x_{n+1}}{x_n} &= \lim_{n\to\infty} \frac{a^{n+1} n!}{(n+1)! a^n} \\ &= \lim_{n\to\infty} \frac{a}{n+1} = 0. \end{aligned}$$

Therefore, by the Ratio test $\lim_{n\to\infty} x_n = 0$. Hence, for any $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that $|x_n - 0| = x_n < \varepsilon$ for all $n \geq K$. Therefore, for all $n \geq K$ we have that

$$|R_n(x)| \leq \frac{a^{n+1}}{(n+1)!} < \varepsilon$$

for all $x \in [-a, a]$. □

Taylor's theorem can be used to derive useful inequalities.

Example 6.3.5. Prove that for all $x \in \mathbb{R}$ it holds that

$$1 - \frac{1}{2}x^2 \leq \cos(x).$$

Solution. Let $f(x) = \cos(x)$. Applying Taylor's theorem to f at $x_0 = 0$ we obtain

$$\cos(x) = 1 - \frac{1}{2}x^2 + R_2(x)$$

where

$$R_2(x) = \frac{f^{(3)}(c)}{3!}x^3 = \frac{\sin(c)}{6}x^3$$

and c is in between $x_0 = 0$ and x. Now, if $0 \leq x \leq \pi$ then $0 < c < \pi$ and then $\sin(c) > 0$, from which it follows that $R_2(x) \geq 0$. If on the other hand $-\pi \leq x \leq 0$ then $-\pi < c < 0$ and then $\sin(c) < 0$, from which it follows that $R_2(x) \geq 0$. Hence, the inequality holds for $|x| \leq \pi$. Now if $|x| \geq \pi > 3$ then

$$1 - \frac{1}{2}x^2 < -3 < \cos(x).$$

Hence the inequality holds for all $x \in \mathbb{R}$. □

Exercises

Exercise 6.3.1. Use Taylor's theorem to prove that if $x > 0$ then

$$1 + \frac{1}{2}x - \frac{1}{8}x^2 \leq \sqrt{1+x} \leq 1 + \frac{1}{2}x$$

Then use these inequalities to approximate $\sqrt{1.2}$ and $\sqrt{2}$, and for each case determine a bound on the error of your approximation.

Exercise 6.3.2. Let $f : \mathbb{R} \to \mathbb{R}$ be such that $f^{(k)}(x)$ exists for all $x \in \mathbb{R}$ and for all $k \in \mathbb{N}$ (such a function is called **infinitely differentiable on** $\mathbb{R}$). Suppose further that there exists $M > 0$ such that $|f^{(k)}(x)| \leq M$ for all $x \in \mathbb{R}$ and all $k \in \mathbb{N}$. Let $P_n(x)$ be the nth order Taylor polynomial of f centered at $x_0 = 0$. Let $I = [-R, R]$, where $R > 0$. Prove that for any fixed $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that for $n \geq K$ it holds that

$$|f(x) - P_n(x)| < \varepsilon$$

for all $x \in [-R, R]$. **Hint:** $f^{(n)}$ is continuous on $[-R, R]$ for every $n \in \mathbb{N}$.

Exercise 6.3.3. Euler's number is approximately $e \approx 2.718281828\ldots$. Use Taylor's theorem at $x_0 = 0$ on $f(x) = e^x$ and the estimate $e < 3$ to show that, for all $n \in \mathbb{N}$,

$$0 < e - \left(1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!}\right) < \frac{3}{(n+1)!}$$

Exercise 6.3.4. Let $f : \mathbb{R} \to \mathbb{R}$ be the cosine function $f(x) = \cos(x)$. Approximate $f(2) = \cos(2)$ using P_8 centered at $x_0 = 0$ and give a bound on the error of your estimation.

7

Riemann Integration

7.1 The Riemann Integral

We begin with the definition of a partition.

Definition 7.1.1: Partitions

Let $a, b \in \mathbb{R}$ and suppose $a < b$. By a **partition** of the interval $[a, b]$ we mean a collection of intervals

$$\mathcal{P} = \{[x_0, x_1], [x_1, x_2], \ldots, [x_{n-1}, x_n]\}$$

such that $a = x_0 < x_1 < x_2 < \cdots < x_n = b$ and where $n \in \mathbb{N}$.

Hence, a partition $\mathcal{P}$ defines a finite collection of non-overlapping intervals $I_k = [x_{k-1}, x_k]$, where $k = 1, \ldots, n$. The **norm** of a partition $\mathcal{P}$ is defined as

$$\|\mathcal{P}\| = \max\{x_1 - x_0, x_2 - x_1, \ldots, x_n - x_{n-1}\}.$$

In other words, $\|\mathcal{P}\|$ is the maximum length of the intervals in $\mathcal{P}$. To ease our notation, we will denote a partition as $\mathcal{P} = \{[x_{k-1}, x_k]\}_{k=1}^n$.

Let $\mathcal{P} = \{[x_{k-1}, x_k]\}_{k=1}^n$ be a partition of $[a, b]$. If $t_k \in I_k = [x_{k-1}, x_k]$

then we say that t_k is a **sample** of I_k and the set of ordered pairs

$$\dot{\mathcal{P}} = \{([x_{k-1}, x_k], t_k)\}_{k=1}^n$$

will be called a **sampled partition**.

Example 7.1.2. Examples of sampled partitions are mid-points, right-end points, and left-end points partitions.

Now consider a function $f : [a, b] \to \mathbb{R}$ and let $\dot{\mathcal{P}} = \{([x_{k-1}, x_k], t_k)\}_{k=1}^n$ be a sampled partition of the interval $[a, b]$. The **Riemann sum** of f corresponding to $\dot{\mathcal{P}}$ is the number

$$S(f; \dot{\mathcal{P}}) = \sum_{k=1}^{n} f(t_k)(x_k - x_{k-1}).$$

When $f(x) > 0$ on the interval $[a, b]$, the Riemann sum $S(f; \dot{\mathcal{P}})$ is the sum of the areas of the rectangles with height $f(t_k)$ and width $(x_k - x_{k-1})$.

We now define the notion of Riemann integrability.

Definition 7.1.3: Riemann Integrability

The function $f : [a, b] \to \mathbb{R}$ is said to be **Riemann integrable** if there exists a number $L \in \mathbb{R}$ such that for every $\varepsilon > 0$ there exists $\delta > 0$ such that for any sampled partition $\dot{\mathcal{P}}$ that satisfies $\|\dot{\mathcal{P}}\| < \delta$ it holds that $|S(f; \dot{\mathcal{P}}) - L| < \varepsilon$.

The set of all Riemann integrable functions on the interval $[a, b]$ will be denoted by $\mathcal{R}[a, b]$.

Theorem 7.1.4

If $f \in \mathcal{R}[a, b]$ then the number L in the definition of Riemann integrability is unique.

Proof. Let L_1 and L_2 be two real numbers satisfying the definition of Riemann integrability and let $\varepsilon > 0$ be arbitrary. Then there exists $\delta > 0$ such that $|S(f;\dot{\mathcal{P}}) - L_1| < \varepsilon/2$ and $|S(f;\dot{\mathcal{P}}) - L_2| < \varepsilon/2$, for all sampled partitions $\dot{\mathcal{P}}$ with $\|\dot{\mathcal{P}}\| < \delta$. Then, if $\|\dot{\mathcal{P}}\| < \delta$ it holds that

$$\begin{aligned} |L_1 - L_2| &\leq |S(f;\dot{\mathcal{P}}) - L_1| + |S(f;\dot{\mathcal{P}}) - L_2| \\ &< \varepsilon. \end{aligned}$$

By Theorem 2.2.7 this proves that $L_1 = L_2$. □

If $f \in \mathcal{R}[a,b]$, we call the number L the **integral of** f **over** $[a,b]$ and we denote it by

$$L = \int_a^b f$$

Example 7.1.5. Show that a constant function on $[a,b]$ is Riemann integrable.

Proof. Let $f : [a,b] :\to \mathbb{R}$ be such that $f(x) = C$ for all $x \in [a,b]$ and let $\dot{\mathcal{P}} = \{([x_{k-1}, x_k], t_k)\}$ be a sampled partition of $[a,b]$. Then

$$\begin{aligned} S(f;\dot{\mathcal{P}}) &= \sum_{k=1}^{n} f(t_k)(x_k - x_{k-1}) \\ &= C\sum_{k=1}^{n}(x_k - x_{k-1}) \\ &= C(x_n - x_0) \\ &= C(b-a). \end{aligned}$$

Hence, with $L = C(b-a)$, we obtain that $|S(f;\dot{\mathcal{P}}) - L| = 0 < \varepsilon$ for any $\varepsilon > 0$ and therefore $\int_a^b f = C(b-a)$. This proves that f is Riemann integrable. □

Example 7.1.6. Prove that $f(x) = x$ is Riemann integrable on $[a,b]$.

Proof. We consider the special case that $[a, b] = [0, 1]$, the general case is similar. Let $\dot{\mathcal{Q}} = \{([x_{k-1}, x_k], q_k)\}$ be a sampled partition of $[0, 1]$ chosen so that $q_k = \frac{1}{2}(x_k + x_{k-1})$, i.e., q_k is the midpoint of the interval $[x_{k-1}, x_k]$. Then

$$\begin{aligned} S(f; \dot{\mathcal{Q}}) &= \sum_{k=1}^{n} f(q_k)(x_k - x_{k-1}) \\ &= \sum_{k=1}^{n} \tfrac{1}{2}(x_k + x_{k-1})(x_k - x_{k-1}) \\ &= \frac{1}{2} \sum_{k=1}^{n} (x_k^2 - x_{k-1}^2) \\ &= \frac{1}{2}(x_n^2 - x_0^2) \\ &= \frac{1}{2}(1^2 - 0^2) \\ &= \frac{1}{2}. \end{aligned}$$

Now let $\dot{\mathcal{P}} = \{([x_{k-1}, x_k]), t_k\}_{k=1}^n$ be an arbitrary sampled partition of $[0, 1]$ and suppose that $\|\dot{\mathcal{P}}\| < \delta$, so that $(x_k - x_{k-1}) \leq \delta$ for all $k = 1, 2, \ldots, n$. If $\dot{\mathcal{Q}} = \{([x_{k-1}, x_k], q_k)\}_{k=1}^n$ is the corresponding midpoint sampled partition then $|t_k - q_k| < \delta$. Therefore,

$$
\begin{aligned}
|S(f;\dot{\mathcal{P}}) - S(f;\dot{\mathcal{Q}})| &= \left|\sum_{k=1}^{n} t_k(x_k - x_{k-1}) - q_k(x_k - x_{k-1})\right| \\
&\leq \sum_{k=1}^{n} |t_k - q_k|(x_k - x_{k-1}) \\
&< \delta(1-0) \\
&= \delta.
\end{aligned}
$$

Hence, we have proved that for arbitrary $\dot{\mathcal{P}}$ that satisfies $\|\dot{\mathcal{P}}\| < \delta$ it holds that $|S(f;\dot{\mathcal{P}}) - 1/2| < \delta$. Hence, given $\varepsilon > 0$ we let $\delta = \varepsilon$ and then if $\|\dot{\mathcal{P}}\| < \delta$ then $|S(f;\dot{\mathcal{P}}) - 1/2| < \varepsilon$. Therefore, $\int_0^1 f = \frac{1}{2}$. □

The next result shows that if $f \in \mathcal{R}[a,b]$ then changing f at a finite number of points in $[a,b]$ does not affect the value of $\int_a^b f$.

Theorem 7.1.7

Let $f \in \mathcal{R}[a,b]$ and let $g : [a,b] \to \mathbb{R}$ be a function such that $g(x) = f(x)$ for all $x \in [a,b]$ except possibly at a finite number of points in $[a,b]$. Then $g \in \mathcal{R}[a,b]$ and in fact $\int_a^b g = \int_a^b f$.

Proof. Let $L = \int_a^b f$. Suppose that $g(x) = f(x)$ except at one point $x = c$. Let $\dot{\mathcal{P}} = \{([x_{k-1}, x_k], t_k)\}$ be a sampled partition. We consider mutually exclusive cases. First, if $c \neq t_k$ and $c \neq x_k$ for all k then $S(f;\dot{\mathcal{P}}) = S(g;\dot{\mathcal{P}})$. If $c = t_k \notin \{x_0, x_1, \ldots, x_n\}$ for some k then

$$S(g;\dot{\mathcal{P}}) - S(f;\dot{\mathcal{P}}) = (f(c) - g(c))(x_k - x_{k-1}).$$

If $c = t_k = t_{k-1}$ for some k then necessarily $c = x_{k-1}$ and then

$$S(g;\dot{\mathcal{P}}) - S(f;\dot{\mathcal{P}}) = (f(c) - g(c))(x_k - x_{k-1}) + (f(c) - g(c))(x_{k-1} - x_{k-2}).$$

Hence, in any case, by the triangle inequality

$$\begin{aligned}|S(g;\dot{\mathcal{P}}) - S(f;\dot{\mathcal{P}})| &\leq 2(|f(c)| + |g(c)|)\|\dot{\mathcal{P}}\| \\ &= M\|\dot{\mathcal{P}}\|.\end{aligned}$$

where $M = 2(|f(c)| + |g(c)|)$. Let $\varepsilon > 0$ be arbitrary. Then there exists $\delta_1 > 0$ such that $|S(f;\dot{\mathcal{P}}) - L| < \varepsilon/2$ for all partitions $\dot{\mathcal{P}}$ such that $\|\dot{\mathcal{P}}\| < \delta_1$. Let $\delta = \min\{\delta_1, \varepsilon/(2M)\}$. Then if $\|\dot{\mathcal{P}}\| < \delta$ then

$$\begin{aligned}\|S(g;f) - L\| &\leq |S(g;\dot{\mathcal{P}}) - S(f;\dot{\mathcal{P}})| + |S(f;\dot{\mathcal{P}}) - L| \\ &< M\|\dot{\mathcal{P}}| + \varepsilon/2 \\ &< M\varepsilon/(2M) + \varepsilon/2 \\ &= \varepsilon.\end{aligned}$$

This proves that $g \in \mathcal{R}[a,b]$ and $\int_a^b g = L = \int_a^b f$. Now suppose by induction that if $g(x) = f(x)$ for all $x \in [a,b]$ except at a $j \geq 1$ number of points in $[a,b]$ then $g \in \mathcal{R}[a,b]$ and $\int_a^b g = \int_a^b f$. Now suppose that $h : [a,b] \to \mathbb{R}$ is such that $h(x) = f(x)$ for all $x \in [a,b]$ except at the points $c_1, c_2, \ldots, c_j, c_{j+1}$. Define the function g by $g(x) = h(x)$ for all $x \in [a,b]$ except at $x = c_{j+1}$ and define $g(c_{j+1}) = f(c_{j+1})$. Then g and f differ at the points $c_1, \ldots, c_j$. Then by the induction hypothesis, $g \in \mathcal{R}[a,b]$ and $\int_a^b g = \int_a^b f$. Now g and h differ at the point c_{j+1} and therefore $h \in \mathcal{R}[a,b]$ and $\int_a^b h = \int_a^b g = \int_a^b f$. This ends the proof. $\square$

We now state some properties of the Riemann integral.

Theorem 7.1.8: Properties of the Riemann Integral

Suppose that $f, g \in \mathcal{R}[a,b]$. The following hold.

(i) If $k \in \mathbb{R}$ then $(kf) \in \mathcal{R}[a,b]$ and $\int_a^b kf = k\int_a^b f$.

(ii) $(f+g) \in \mathcal{R}[a,b]$ and $\int_a^b (f+g) = \int_a^b f + \int_a^b g$.

(iii) If $f(x) \leq g(x)$ for all $x \in [a,b]$ then $\int_a^b f \leq \int_a^b g$.

Proof. If $k = 0$ then $(kf)(x) = 0$ for all x and then clearly $\int_a^b kf = 0$, so assume that $k \neq 0$. Let $\varepsilon > 0$ be given. Then there exists $\delta > 0$ such that if $\|\dot{\mathcal{P}}\| < \delta$ then $|S(f;\dot{\mathcal{P}}) - \int_a^b f\| < \varepsilon/|k|$. Now for any partition $\dot{\mathcal{P}}$, it holds that $S(kf;\dot{\mathcal{P}}) = kS(f;\dot{\mathcal{P}})$. Therefore, if $\|\dot{\mathcal{P}}\| < \delta$ then

$$\begin{aligned}\left|S(kf;\dot{\mathcal{P}}) - k\int_a^b f\right| &= |k|\left|S(f;\dot{\mathcal{P}}) - \int_a^b f\right| \\ &< |k|(\varepsilon/|k|) \\ &= \varepsilon.\end{aligned}$$

To prove (b), it is easy to see that $S(f+g;\dot{\mathcal{P}}) = S(f;\dot{\mathcal{P}}) + S(g;\dot{\mathcal{P}})$. Given $\varepsilon > 0$ there exists $\delta > 0$ such that $|S(f;\dot{\mathcal{P}}) - \int_a^b f| < \varepsilon/2$ and $|S(g;\dot{\mathcal{P}}) - \int_a^b g| < \varepsilon/2$, whenever $\|\dot{\mathcal{P}}\| < \delta$. Therefore, if $\|\dot{\mathcal{P}}\| < \delta$ we have that

$$\begin{aligned}\left|S(f+g;\dot{\mathcal{P}}) - \left(\int_a^b f + \int_a^b g\right)\right| &\leq \left|S(f;\dot{\mathcal{P}}) - \int_a^b f\right| + \left|S(g;\dot{\mathcal{P}}) - \int_a^b g\right| \\ &< \varepsilon.\end{aligned}$$

To prove (c), let $\varepsilon > 0$ be arbitrary and let $\delta > 0$ be such that if

$\|\dot{\mathcal{P}}\| < \delta$ then

$$\int_a^b f - \varepsilon/2 < S(f;\dot{\mathcal{P}}) < \int_a^b f + \varepsilon/2$$

$$\int_a^b g - \varepsilon/2 < S(g;\dot{\mathcal{P}}) < \int_a^b g + \varepsilon/2$$

Now, by assumption, $S(f;\dot{\mathcal{P}}) \leq S(g;\dot{\mathcal{P}})$ and therefore

$$\int_a^b f - \varepsilon/2 < S(f;\dot{\mathcal{P}}) \leq S(g;\dot{\mathcal{P}}) < \int_a^b g + \varepsilon/2.$$

Therefore,

$$\int_a^b f < \int_a^b g + \varepsilon.$$

Since ε is arbitrary, we can choose $\varepsilon_n = 1/n$ and then passing to the limit we deduce that $\int_a^b f \leq \int_a^b g$. □

Properties (i), (ii), and (iii) in Theorem 7.1.8 are known as homogeneity, additivity, and monotonicity, respectively.

We now give a necessary condition for Riemann integrability.

Theorem 7.1.9: Integrable Functions are Bounded

If $f \in \mathcal{R}[a,b]$ then f is bounded on $[a,b]$.

Proof. Let $f \in \mathcal{R}[a,b]$ and put $L = \int_a^b f$. There exists $\delta > 0$ such that if $\|\dot{\mathcal{P}}\| < \delta$ then $|S(f;\dot{\mathcal{P}}) - L| < 1$ and therefore $|S(f;\dot{\mathcal{P}})| < |L| + 1$. Suppose by contradiction that f is unbounded on $[a,b]$. Let $\mathcal{P}$ be a partition of $[a,b]$, with sets $I_1, \ldots, I_n$, and with $\|\mathcal{P}\| < \delta$. Then f is unbounded on some I_j, i.e., for any $M > 0$ there exists $x \in I_j = [x_{j-1}, x_j]$ such that $f(x) > M$. Choose samples in $\mathcal{P}$ by asking that

$t_k = x_k$ for $k \neq j$ and t_j is such that

$$|f(t_j)(x_j - x_{j-1})| > |L| + 1 + \left| \sum_{k \neq j} f(t_k)(x_k - x_{k-1}) \right|.$$

Therefore, (using $|a| = |a + b - b| \leq |a + b| + |b|$ implies that $|a + b| \geq |a| - |b|$)

$$\begin{aligned} |S(f; \dot{\mathcal{P}})| &= \left| f(t_j)(x_j - x_{j-1}) + \sum_{k \neq j} f(t_k)(x_k - x_{k-1}) \right| \\ &\geq |f(t_j)(x_j - x_{j-1})| - \left| \sum_{k \neq j} f(t_k)(x_k - x_{k-1}) \right| \\ &> |L| + 1. \end{aligned}$$

This is a contradiction and thus f is bounded on $[a, b]$. □

Example 7.1.10 (Thomae)**.** Consider Thomae's function $h : [0, 1] \to \mathbb{R}$ defined as $h(x) = 0$ if x is irrational and $h(m/n) = 1/n$ for every rational $m/n \in [0, 1]$, where $\gcd(m, n) = 1$. In Example 5.1.7, we proved that h is continuous at every irrational but discontinuous at every rational. Prove that h is Riemann integrable.

Proof. Let $\varepsilon > 0$ be arbitrary and let $E = \{x \in [0, 1] \;:\; h(x) \geq \varepsilon/2\}$. By definition of h, the set E is finite, say consisting of n elements. Let $\delta = \varepsilon/(4n)$ and let $\dot{\mathcal{P}}$ be a sampled partition of $[0, 1]$ with $\|\dot{\mathcal{P}}\| < \delta$. We can separate the partition $\dot{\mathcal{P}}$ into two sampled partitions $\dot{\mathcal{P}}_1$ and $\dot{\mathcal{P}}_2$ where $\dot{\mathcal{P}}_1$ has samples in the the set E and $\dot{\mathcal{P}}_2$ has no samples in E. Then the number of intervals in $\dot{\mathcal{P}}_1$ can be at most $2n$, which occurs when all the elements of E are samples and they are at the endpoints of the subintervals of $\dot{\mathcal{P}}_1$. Therefore, the total length of the intervals in $\dot{\mathcal{P}}_1$ can be at most $2n\delta = \varepsilon/2$. Now $0 < h(t_k) \leq 1$ for every sample t_k in $\dot{\mathcal{P}}_1$ and therefore $S(f;\dot{\mathcal{P}}_1) \leq 2n\delta = \varepsilon/2$. For samples t_k in $\dot{\mathcal{P}}_2$ we have that $h(t_k) < \varepsilon/2$. Therefore, since the sum of the lengths of the subintervals of $\dot{\mathcal{P}}_2$ is ≤ 1, it follows that $S(f;\dot{\mathcal{P}}_2) \leq \varepsilon/2$. Hence $0 \leq S(f;\dot{\mathcal{P}}) = S(f;\dot{\mathcal{P}}_1) + S(f;\dot{\mathcal{P}}_2) < \varepsilon$. Thus $\int_a^b h = 0$. □

Exercises

Exercise 7.1.1. Suppose that $f, g \in \mathcal{R}[a,b]$ and let $\alpha, \beta \in \mathbb{R}$. Prove by definition that $(\alpha f + \beta g) \in \mathcal{R}[a,b]$.

Exercise 7.1.2. If f is Riemann integrable on $[a,b]$ and $|f(x)| \leq M$ for all $x \in [a,b]$, prove that $|\int_a^b f| \leq M(b-a)$. **Hint**: The inequality $|f(x)| \leq M$ is equivalent to $-M \leq f(x) \leq M$. Then use the fact that constants functions are Riemann integrable whose integrals are easily computed. Finally, apply a theorem from this section.

Exercise 7.1.3. If f is Riemann integrable on $[a,b]$ and $(\dot{\mathcal{P}}_n)$ is a sequence of tagged partitions of $[a,b]$ such that $\|\dot{\mathcal{P}}_n\| \to 0$ prove that

$$\int_a^b f = \lim_{n\to\infty} S(f; \dot{\mathcal{P}}_n)$$

Hint: For each $n \in \mathbb{N}$ we have the real number $s_n = S(f; \dot{\mathcal{P}}_n)$, and we therefore have a sequence (s_n). Let $L = \int_a^b f$. We therefore want to prove that $\lim\limits_{n\to\infty} s_n = L$.

Exercise 7.1.4. Give an example of a function $f : [0,1] \to \mathbb{R}$ that is Riemann integrable on $[c,1]$ for every $c \in (0,1)$ but which is not Riemann integrable on $[0,1]$. **Hint**: What is a necessary condition for Riemann integrability?

7.2 Riemann Integrable Functions

To ease our notation, if I is a bounded interval with end-points $a < b$ we denote by $\mu(I)$ the length of I, that is $\mu(I) = b - a$. Hence, if $I = [a, b]$, $I = [a, b)$, $I = (a, b]$, or $I = (a, b)$ then $\mu(I) = b - a$.

Thus far, to establish the Riemann integrability of f, we computed a candidate integral L and showed that in fact $L = \int_a^b f$. The following theorem is useful when a candidate integral L is unknown. The proof is omitted.

Theorem 7.2.1: Cauchy Criterion

A function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if for every $\varepsilon > 0$ there exists $\delta > 0$ such that if $\dot{\mathcal{P}}$ and $\dot{\mathcal{Q}}$ are sampled partitions of $[a, b]$ with norm less than δ then

$$|S(f; \dot{\mathcal{P}}) - S(f; \dot{\mathcal{Q}})| < \varepsilon.$$

Using the Cauchy Criterion, we show next that the Dirichlet function is not Riemann integrable.

Example 7.2.2 (Non-Riemann integrable function). Let $f : [0, 1] \to \mathbb{R}$ be defined as $f(x) = 1$ if x is rational and $f(x) = 0$ if x is irrational. Show that f is not Riemann integrable.

Proof. To show that f is not in $\mathcal{R}[0, 1]$, we must show that there exists $\varepsilon_0 > 0$ such that for all $\delta > 0$ there exists sampled partitions $\dot{\mathcal{P}}$ and $\dot{\mathcal{Q}}$ with norm less than δ but $|S(f; \dot{\mathcal{P}}) - S(f; \dot{\mathcal{Q}})| \geq \varepsilon_0$. To that end, let $\varepsilon_0 = 1/2$, and let $\delta > 0$ be arbitrary. Let n be sufficiently large so that $1/n < \delta$. Let $\dot{\mathcal{P}}$ be a sampled partition of $[0, 1]$ with intervals all of equal length $1/n < \delta$ and let the samples of $\dot{\mathcal{P}}$ be rational numbers. Similarly, let $\dot{\mathcal{Q}}$ be a partition of $[0, 1]$ with intervals all of equal length $1/n$ and with samples irrational numbers. Then $S(f;\dot{\mathcal{P}}) = 1$ and $S(f;\dot{\mathcal{Q}}) = 0$, and therefore $|S(f;\dot{\mathcal{Q}}) - S(f;\dot{\mathcal{Q}})| \geq \varepsilon_0$. ☐

We now state a sort of squeeze theorem for integration.

Theorem 7.2.3: Squeeze Theorem

Let f be a function on $[a,b]$. Then $f \in \mathcal{R}[a,b]$ if and only if for every $\varepsilon > 0$ there exist functions α and β in $\mathcal{R}[a,b]$ with $\alpha(x) \leq f(x) \leq \beta(x)$ for all $x \in [a,b]$ and $\int_a^b (\beta - \alpha) < \varepsilon$.

Proof. If $f \in \mathcal{R}[a,b]$ then let $\alpha(x) = \beta(x) = f(x)$. Then clearly $\int_a^b (\beta - \alpha) = 0 < \varepsilon$ for all $\varepsilon > 0$. Now suppose the converse and let $\varepsilon > 0$ be arbitrary. Let α and β satisfy the conditions of the theorem, with $\int_a^b (\beta - \alpha) < \frac{\varepsilon}{3}$. Now, there exists $\delta > 0$ such that if $\|\dot{\mathcal{P}}\| < \delta$ then

$$\int_a^b \alpha - \frac{\varepsilon}{3} < S(\alpha;\dot{\mathcal{P}}) < \int_a^b \alpha + \frac{\varepsilon}{3}$$

and

$$\int_a^b \beta - \frac{\varepsilon}{3} < S(\beta;\dot{\mathcal{P}}) < \int_a^b \beta + \frac{\varepsilon}{3}.$$

For any sampled partition $\dot{\mathcal{P}}$ it holds that $S(\alpha;\dot{\mathcal{P}}) \leq S(f;\dot{\mathcal{P}}) \leq S(\beta;\dot{\mathcal{P}})$, and therefore

$$\int_a^b \alpha - \frac{\varepsilon}{3} < S(f;\dot{\mathcal{P}}) < \int_a^b \beta + \frac{\varepsilon}{3}. \tag{7.1}$$

If $\dot{\mathcal{Q}}$ is another sampled partition with $\|\dot{\mathcal{Q}}\| < \delta$ then also

$$\int_a^b \alpha - \frac{\varepsilon}{3} < S(f;\dot{\mathcal{Q}}) < \int_a^b \beta + \frac{\varepsilon}{3}. \tag{7.2}$$

Subtracting the two inequalities (7.1)-(7.2), we deduce that

$$-\int_a^b (\beta - \alpha) - 2\frac{\varepsilon}{3} < S(f;\dot{\mathcal{P}}) - S(f;\dot{\mathcal{Q}}) < \int_a^b (\beta - \alpha) + 2\frac{\varepsilon}{3}.$$

Therefore, since $\int_a^b (\beta - \alpha) < \frac{\varepsilon}{3}$ it follows that

$$-\varepsilon < S(f;\dot{\mathcal{P}}) - S(f;\dot{\mathcal{Q}}) < \varepsilon.$$

By the Cauchy criterion, this proves that $f \in \mathcal{R}[a,b]$. ☐

Step-functions, defined below, play an important role in integration theory.

Definition 7.2.4

A function $s : [a,b] \to \mathbb{R}$ is called a **step-function** on $[a,b]$ if there is a finite number of disjoint intervals $I_1, I_2, \ldots, I_n$ contained in $[a,b]$ such that $[a,b] = \bigcup_{k=1}^n I_k$ and such that s is constant on each interval.

In the definition of a step-function, the intervals I_k may be of any form, i.e., half-closed, open, or closed.

Lemma 7.2.5

Let J be a subinterval of $[a,b]$ and define φ_J on $[a,b]$ as $\varphi_J(x) = 1$ if $x \in J$ and $\varphi_J(x) = 0$ otherwise. Then $\varphi_J \in \mathcal{R}[a,b]$ and $\int_a^b \varphi_J = \mu(J)$.

Theorem 7.2.6

If $\varphi : [a,b] \to \mathbb{R}$ is a step function then $\varphi \in \mathcal{R}[a,b]$.

Proof. Let $I_1, \ldots, I_n$ be the intervals where φ is constant, and let $c_1, \ldots, c_n$ be the constant values taken by φ on the intervals $I_1, \ldots, I_n$, respectively. Then it is not hard to see that $\varphi = \sum_{k=1}^n c_k \varphi_{I_k}$. Then φ is the sum of Riemann integrable functions and therefore is also Riemann integrable. Moreover, $\int_a^b \varphi = \sum_{k=1}^n c_k \mu(I_k)$. □

We will now show that any continuous function on $[a, b]$ is Riemann integrable. To do that we will need the following.

Lemma 7.2.7: Continuity and Step-Functions

Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then for every $\varepsilon > 0$ there exists a step-function $s : [a, b] \to \mathbb{R}$ such that $|f(x) - s(x)| < \varepsilon$ for all $x \in [a, b]$.

Proof. Let $\varepsilon > 0$ be arbitrary. Since f is uniformly continuous on $[a, b]$ there exists $\delta > 0$ such that if $|x - u| < \delta$ then $|f(x) - f(u)| < \varepsilon$. Let $n \in \mathbb{N}$ be sufficiently large so that $(b - a)/n < \delta$. Partition $[a, b]$ into n subintervals of equal length $(b-a)/n$, and denote them by $I_1, I_2, \ldots, I_n$, where $I_1 = [x_0, x_1]$ and $I_k = (x_{k-1}, x_k]$ for $1 < k \leq n$. Then for $x, u \in I_k$ it holds that $|f(x) - f(u)| < \varepsilon$. For $x \in I_k$ define $s(x) = f(x_k)$. Therefore, for any $x \in I_k$ it holds that $|f(x)-s(x)| = |f(x)-f(x_k)| < \varepsilon$. Since $\bigcup_{k=1}^n I_k = [a, b]$, it holds that $|f(x)-s(x)| < \varepsilon$ for all $x \in [a, b]$. □

We now prove that continuous functions are integrable.

Theorem 7.2.8: Continuous Functions are Integrable

A continuous function on $[a, b]$ is Riemann integrable on $[a, b]$.

Proof. Suppose that $f : [a, b] \to \mathbb{R}$ is continuous. Let $\varepsilon > 0$ be arbitrary and let $\tilde{\varepsilon} = (\varepsilon/4)/(b - a)$. Then there exists a step-function $s : [a, b] \to \mathbb{R}$ such that $|f(x) - s(x)| < \tilde{\varepsilon}$ for all $x \in [a, b]$. In other words, for all $x \in [a, b]$ it holds that

$$s(x) - \tilde{\varepsilon} < f(x) < s(x) + \tilde{\varepsilon}.$$

The functions $\alpha(x) := s(x) - \tilde{\varepsilon}$ and $\beta(x) := s(x) + \tilde{\varepsilon}$ are Riemann integrable integrable on $[a, b]$, and $\int_a^b (\beta - \alpha) = 2\tilde{\varepsilon}(b - a) = \varepsilon/2 < \varepsilon$. Hence, by the Cauchy criterion, f is Riemann integrable. ☐

Recall that a function is called monotone if it is decreasing or increasing.

Theorem 7.2.9: Monotone Functions are Integrable

A monotone function on $[a, b]$ is Riemann integrable on $[a, b]$.

Proof. Assume that $f : [a, b] \to \mathbb{R}$ is increasing and that $M = f(b) - f(a) > 0$ (if $M = 0$ then f is the zero function which is clearly integrable). Let $\varepsilon > 0$ be arbitrary. Let $n \in \mathbb{N}$ be such that $\frac{M(b-a)}{n} < \varepsilon$. Partition $[a, b]$ into subintervals of equal length $\Delta x = \frac{(b-a)}{n}$, and as usual let $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$ denote the resulting points of the partition. On each subinterval $[x_{k-1}, x_k]$, it holds that $f(x_{k-1}) \leq f(x) \leq f(x_k)$ for all $x \in [x_{k-1}, x_k]$ since f is increasing. Let $\alpha : [a, b] \to \mathbb{R}$ be the step-function whose constant value on the interval $[x_{k-1}, x_k)$ is $f(x_{k-1})$ and similarly let $\beta : [a, b] \to \mathbb{R}$ be the step-function whose constant value on the interval $[x_{k-1}, x_k)$ is $f(x_k)$, for all $k = 1, \ldots, n$. Then $\alpha(x) \leq f(x) \leq \beta(x)$ for all $x \in [a, b]$. Both α and β are Riemann integrable and

$$\begin{aligned}
\int_a^b (\beta - \alpha) &= \sum_{k=1}^{n} [f(x_k) - f(x_{k-1})] \Delta x \\
&= (f(x_n) - f(x_{n-1}) \Delta x \\
&= M \frac{(b-a)}{n} \\
&< \varepsilon.
\end{aligned}$$

Hence by the Squeeze theorem for integrals (Theorem 7.2.3), $f \in \mathcal{R}[a, b]$. ☐

Our las theorem is the additivity property of the integral, the proof is omitted.

Theorem 7.2.10: Additivity Property

Let $f : [a, b] \to \mathbb{R}$ be a function and let $c \in (a, b)$. Then $f \in \mathcal{R}[a, b]$ if and only if its restrictions to $[a, c]$ and $[c, b]$ are both Riemann integrable. In this case,

$$\int_a^b f = \int_a^c f + \int_c^b f$$

Exercises

Exercise 7.2.1. Suppose that $f : [a, b] \to \mathbb{R}$ is continuous and assume that $f(x) > 0$ for all $x \in [a, b]$. Prove that $\int_a^b f > 0$. **Hint**: A continuous function on a closed and bounded interval achieves its minimum value.

Exercise 7.2.2. Suppose that f is continuous on $[a, b]$ and that $f(x) \geq 0$ for all $x \in [a, b]$.

(a) Prove that if $\int_a^b f = 0$ then necessarily $f(x) = 0$ for all $x \in [a, b]$.

(b) Show by example that if we drop the assumption that f is continuous on $[a, b]$ then it may not longer hold that $f(x) = 0$ for all $x \in [a, b]$.

Exercise 7.2.3. Show that if $f : [a, b] \to \mathbb{R}$ is Riemann integrable then $|f| : [a, b] \to \mathbb{R}$ is also Riemann integrable.

7.3 The Fundamental Theorem of Calculus

Theorem 7.3.1: FTC Part I

Let $f : [a, b] \to \mathbb{R}$ be a function. Suppose that there exists a finite set $E \subset [a, b]$ and a function $F : [a, b] \to \mathbb{R}$ such that F is continuous on $[a, b]$ and $F'(x) = f(x)$ for all $x \in [a, b]\backslash E$. If f is Riemann integrable then $\int_a^b f = F(b) - F(a)$.

Proof. Assume for simplicity that $E := \{a, b\}$. Let $\varepsilon > 0$ be arbitrary. Then there exists $\delta > 0$ such that if $\|\dot{\mathcal{P}}\| < \varepsilon$ then $|S(f; \dot{\mathcal{P}}) - \int_a^b f| < \varepsilon$. For any $\dot{\mathcal{P}}$, with intervals $I_k = [x_{k-1}, x_k]$ for $k = 1, 2, \ldots, n$, there exists, by the Mean Value Theorem applied to F on I_k, a point $u_k \in (x_{k-1}, x_k)$ such that $F(x_k) - F(x_{k-1}) = F'(u_k)(x_k - x_{k-1})$. Therefore,

$$\begin{aligned} F(b) - F(a) &= \sum_{k=1}^{n} F(x_k) - F(x_{k-1}) \\ &= \sum_{k=1}^{n} f(u_k)(x_k - x_{k-1}) \\ &= S(f; \dot{\mathcal{P}}_u) \end{aligned}$$

where $\dot{\mathcal{P}}_u$ has the same intervals as $\dot{\mathcal{P}}$ but with samples u_k. Therefore, if $\|\dot{\mathcal{P}}\| < \delta$ then

$$\begin{aligned} \left| F(b) - F(a) - \int_a^b f \right| &= \left| S(f; \dot{\mathcal{P}}_u) - \int_a^b f \right| \\ &< \varepsilon. \end{aligned}$$

Hence, for any ε we have that $\left| F(b) - F(a) - \int_a^b f \right| < \varepsilon$ and this shows that $\int_a^b f = F(b) - F(a)$. □

Definition 7.3.2: Indefinite Integral

Let $f \in \mathcal{R}[a,b]$. The **indefinite integral** of f with basepoint a is the function $F : [a,b] \to \mathbb{R}$ defined by

$$F(x) := \int_a^x f$$

for $x \in [a,b]$.

Theorem 7.3.3

Let $f \in \mathcal{R}[a,b]$. The indefinite integral $F : [a,b] \to \mathbb{R}$ of $f : [a,b] \to \mathbb{R}$ is a Lipschitz function on $[a,b]$, and thus continuous on $[a,b]$.

Proof. For any $w, z \in [a,b]$ such that $w \leq z$ it holds that

$$\begin{aligned} F(z) &= \int_a^z f \\ &= \int_a^w f + \int_w^z f \\ &= F(w) + \int_w^z f \end{aligned}$$

and therefore $F(z) - F(w) = \int_w^z f$. Since f is Riemann integrable on $[a,b]$ it is bounded and therefore $|f(x)| \leq K$ for all $x \in [a,b]$. In particular, $-K \leq f(x) \leq K$ for all $x \in [w,z]$ and thus $-K(z-w) \leq \int_w^z f \leq K(z-w)$, and thus

$$\begin{aligned} |F(z) - F(w)| &= \left| \int_w^z f \right| \\ &\leq K|z-w| \end{aligned}$$

□

Under the additional hypothesis that $f \in \mathcal{R}[a,b]$ is continuous, the indefinite integral of f is differentiable.

Theorem 7.3.4: FTC Part 2

Let $f \in \mathcal{R}[a,b]$ and let f be continuous at a point $c \in [a,b]$. Then the indefinite integral F of f is differentiable at c and $F'(c) = f(c)$.

7.4 Riemann-Lebesgue Theorem

In this section we present a complete characterization of Riemann integrability for a bounded function. Roughly speaking, a bounded function is Riemann integrable if the set of points were it is discontinuous is not too large. We first begin with a definition of "not too large".

Definition 7.4.1

A set $E \subset \mathbb{R}$ is said to be of **measure zero** if for every $\varepsilon > 0$ there exists a countable collection of open intervals I_k such that

$$E \subset \bigcup_{k=1}^{\infty} I_k \quad \text{and} \quad \sum_{k=1}^{\infty} \mu(I_k) < \varepsilon.$$

Example 7.4.2. Show that a subset of a set of measure zero also has measure zero. Show that the union of two sets of measure zero is a set of measure zero.

Example 7.4.3. Let $S \subset \mathbb{R}$ be a countable set. Show that S has measure zero.

Solution. Let $S = \{s_1, s_2, s_3, \ldots\}$. Consider the interval

$$I_k = \left(s_k - \frac{\varepsilon}{2^{k+1}}, s_k + \frac{\varepsilon}{2^{k+1}}\right).$$

Clearly, $s_k \in I_k$ and thus $S \subset \bigcup I_k$. Moreover,

$$\sum_{k=1}^{\infty} \mu(I_k) = \sum_{k=1}^{\infty} \frac{\varepsilon}{2^k} = \varepsilon.$$

As a corollary, $\mathbb{Q}$ has measure zero. □

However, there exists uncountable sets of measure zero.

Example 7.4.4. The Cantor set is defined as follows. Start with $I_0 = [0, 1]$ and remove the middle third $J_1 = (\frac{1}{3}, \frac{2}{3})$ yielding the set $I_1 = I_0 \backslash J_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. Notice that $\mu(J_1) = \frac{1}{3}$. Now remove from each subinterval of I_1 the middle third resulting in the set

$$I_2 = I_1 \backslash \left((\tfrac{1}{9}, \tfrac{2}{9}) \cup (\tfrac{7}{9}, \tfrac{8}{9})\right) = [0, \tfrac{1}{9}] \cup [\tfrac{2}{9}, \tfrac{3}{9}] \cup [\tfrac{6}{9}, \tfrac{7}{9}] \cup [\tfrac{8}{9}, 1]$$

The two middle thirds $J_2 = (\frac{1}{9}, \frac{2}{9}) \cup (\frac{7}{9}, \frac{8}{9})$ removed have total length $\mu(J_2) = 2\frac{1}{9}$. By induction, having constructed I_n which consists of the union of 2^n closed subintervals of $[0, 1]$, we remove from each subinterval of I_n the middle third resulting in the set $I_{n+1} = I_n \backslash J_{n+1}$, where J_{n+1} is the union of the 2^n middle third open intervals and I_{n+1} now consists of 2^{n+1} disjoint closed-subintervals. By induction, the total length of J_{n+1} is $\mu(J_{n+1}) = \frac{2^n}{3^{n+1}}$. The Cantor set is defined as

$$C = \bigcap_{n=1}^{\infty} I_n.$$

We now state the Riemann-Lebesgue theorem.

Theorem 7.4.5: Riemann-Lebesgue

Let $f : [a, b] \to \mathbb{R}$ be a bounded function. Then f is Riemann integrable if and only if the points of discontinuity of f forms a set of measure zero.

8

Sequences of Functions

In the previous sections, we have considered real-number sequences, that is, sequences (x_n) such that $x_n \in \mathbb{R}$ for each $n \in \mathbb{N}$. In this section, we consider sequences whose terms are *functions*. Sequences of functions arise naturally in many applications in physics and engineering. A typical way that sequences of functions arise is in the problem of solving an equation in which the unknown is a function f. In many of these types of problems, one is able to generate a sequence of functions $(f_n) = (f_1, f_2, f_3, \ldots)$ through some algorithmic process with the intention that the sequence of functions (f_n) converges to the solution f. Moreover, it would be desirable that the limiting function f inherit as many properties possessed by each function f_n such as, for example, continuity, differentiability, or integrability. We will see that this latter issue is rather delicate. In this section, we develop a notion of the limit of a sequence of functions and then investigate if the fundamental properties of boundedness, continuity, integrability, and differentiability are preserved under the limit operation.

8.1 Pointwise Convergence

Let $A \subset \mathbb{R}$ be a non-empty subset and suppose that for each $n \in \mathbb{N}$ we have a function $f_n : A \to \mathbb{R}$. We then say that $(f_n) = (f_1, f_2, f_3, \ldots,)$ is a **sequence of functions on** A.

Example 8.1.1. Let $A = [0, 1]$ and let $f_n(x) = x^n$ for $n \in \mathbb{N}$ and $x \in A$. Then $(f_n) = (f_1, f_2, f_3, \ldots)$ is a sequence of functions on A. As another example, for $n \in \mathbb{N}$ and $x \in A$ let $g_n(x) = nx(1 - x^2)^n$. Then $(g_n) = (g_1, g_2, g_3, \ldots)$ is a sequence of functions on A. Or how about

$$f_n(x) = a_n \cos(nx) + b_n \sin(nx)$$

where $a_n, b_n \in \mathbb{R}$ and $x \in [-\pi, \pi]$.

Let (f_n) be a sequence of functions on A. For each fixed $x \in A$ we obtain a sequence of real numbers (x_n) by simply evaluating each f_n at x, that is, $x_n = f_n(x)$. For example, if $f_n(x) = x^n$ and we fix $x = \frac{3}{4}$ then we obtain the sequence $x_n = f_n(\frac{3}{4}) = \left(\frac{3}{4}\right)^n$. If $x \in A$ is fixed we can then easily talk about the convergence of the sequence of numbers $(f_n(x))$ in the usual way. This leads to our first definition of convergence of function sequences.

Definition 8.1.2: Pointwise Convergence

Let (f_n) be a sequence of functions on $A \subseteq \mathbb{R}$. We say that (f_n) **converges pointwise on** A **to the function** $f : A \to \mathbb{R}$ if for each $x \in A$ the sequence $(f_n(x))$ converges to the number $f(x)$, that is,

$$\lim_{n\to\infty} f_n(x) = f(x).$$

In this case, we call the function f the **pointwise limit** of the sequence (f_n).

By uniqueness of limits of sequences of real numbers (Theorem 3.1.12), the pointwise limit of a sequence (f_n) is unique. Also, when the domain A is understood, we will simply say that (f_n) converges pointwise to f.

Example 8.1.3. Consider the sequence (f_n) defined on $\mathbb{R}$ by $f_n(x) = (2xn + (-1)^n x^2)/n$. For fixed $x \in \mathbb{R}$ we have

$$\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \frac{2xn + (-1)^n x^2}{n} = 2x.$$

Hence, (f_n) converges pointwise to $f(x) = 2x$ on $\mathbb{R}$. In Figure 8.1, we graph $f_n(x)$ for the values $n = 1, 2, 3, 4$ and the function $f(x) = 2x$. Notice that $f_n'(x) = (2n + 2(-1)^n x)/n$ and therefore $\lim_{n\to\infty} f_n'(x) = 2$, and for the limit function $f(x) = 2x$ we have $f'(x) = 2$. Hence, the sequence of derivatives (f_n') converges pointwise to f'. Also, after some basic computations,

$$\begin{aligned}\int_{-1}^{1} f_n(x)\,dx &= \int_{-1}^{1} \frac{2xn + (-1)^n x^2}{n}\,dx \\ &= \frac{2(-1)^n}{3n}\end{aligned}$$

and therefore

$$\begin{aligned}\lim_{n\to\infty} \int_{-1}^{1} f_n(x)\,dx &= \lim_{n\to\infty} \frac{2(-1)^n}{3n} \\ &= 0.\end{aligned}$$

On the other hand it is clear that $\int_{-1}^{1} f(x)\,dx = 0$.

Before considering more examples, we state the following result which is a direct consequence of the definition of the limit of a sequence of numbers and the definition of pointwise convergence.

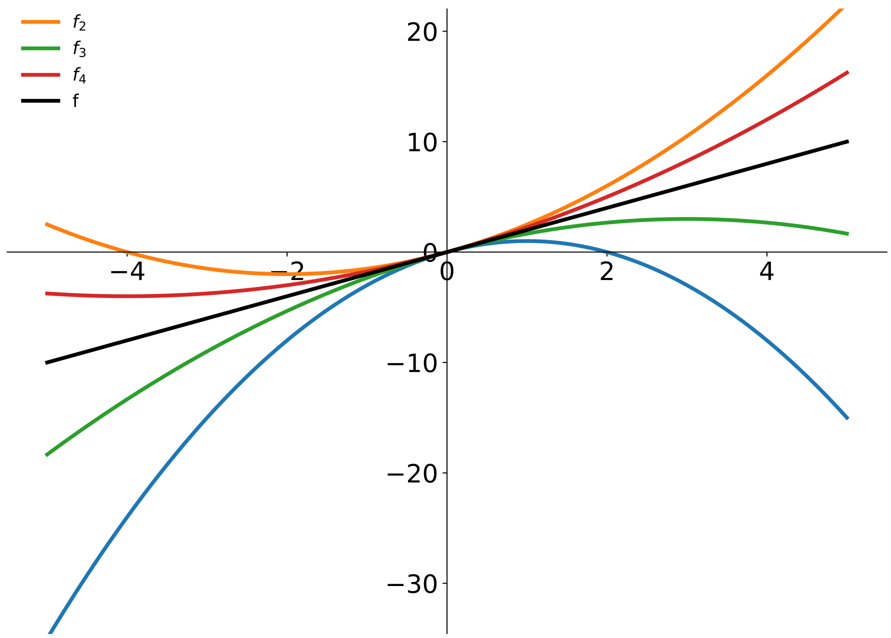


Figure 8.1: Graph of $f_n(x) = \frac{2xn+(-1)^n x^2}{n}$ for $n = 1, 2, 3, 4$ and $f(x) = 2x$

Lemma 8.1.4

Let (f_n) be a sequence of functions on A. Then (f_n) converges pointwise to $f : A \to \mathbb{R}$ if and only if for each $x \in A$ and each $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that $|f_n(x) - f(x)| < \varepsilon$ for all $n \geq K$.

As the following example shows, it is important to note that the K in Lemma 8.1.4 depends not only on $\varepsilon > 0$ but in general will also depend on $x \in A$.

Example 8.1.5. Consider the sequence (f_n) defined on $A = [0, 1]$ by $f_n(x) = x^n$. For all $n \in \mathbb{N}$ we have $f_n(1) = 1^n = 1$ and therefore $\lim_{n\to\infty} f_n(1) = 1$. On the other hand if $x \in [0, 1)$ then

$$\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} x^n = 0.$$

Therefore, (f_n) converges pointwise on A to the function

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

In Figure 8.2, we graph $f_n(x) = x^n$ for various values of n. Consider a fixed $x \in (0,1)$. Since $\lim_{n\to\infty} x^n = 0$ it follows that for $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that $|x^n - 0| < \varepsilon$ for all $n \geq K$. For $\varepsilon < 1$, in order for $|x^K| = x^K < \varepsilon$ we can choose $K > \ln(\varepsilon)/\ln(x)$. Notice that K clearly depends on both ε and x, and in particular, as x get closer to 1 then a larger K is needed. We note that each f_n is continuous while f is not.

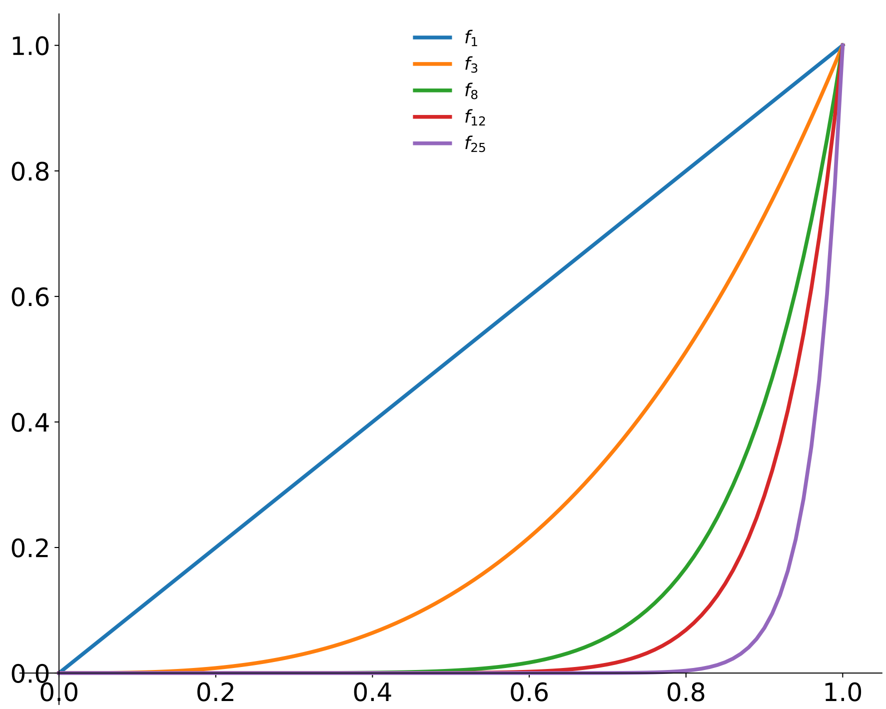


Figure 8.2: Graph of $f_n(x) = x^n$ for $n = 1, 3, 8, 12, 25$

Example 8.1.5 also illustrates a weakness of pointwise convergence, namely, that if (f_n) is a sequence of continuous functions on A and (f_n)

converges pointwise to f on A then f is not necessarily continuous on A.

Example 8.1.6. Consider the sequence (f_n) defined on $A = [-1, 1]$ by $f_n(x) = \sqrt{\frac{nx^2+1}{n}}$. For fixed $x \in A$ we have

$$\begin{aligned}
\lim_{n\to\infty} f_n(x) &= \lim_{n\to\infty} \sqrt{\frac{nx^2+1}{n}} \\
&= \lim_{n\to\infty} \sqrt{x^2 + \frac{1}{n}} \\
&= \sqrt{x^2} \\
&= |x|.
\end{aligned}$$

Hence, (f_n) converges pointwise on A to the function $f(x) = |x|$. Notice that each function f_n is continuous on A and the pointwise limit f is also continuous. After some basic calculations we find that

$$f_n'(x) = \frac{x}{\sqrt{\frac{nx^2+1}{n}}}$$

and $f_n'(x)$ exists for each $x \in [-1, 1]$, in other words, f_n is differentiable on A. However, $f(x) = |x|$ is not differentiable on A since f does not have a derivative at $x = 0$. In Figure 8.3, we graph f_n for various values of n.

Example 8.1.6 illustrates another weakness of pointwise convergence, namely, that if (f_n) is a sequence of differentiable functions on A and (f_n) converges pointwise to f on A then f is not necessarily differentiable on A.

Example 8.1.7. Consider the sequence (f_n) on $A = [0, 1]$ defined by $f_n(x) = 2nxe^{-nx^2}$. For fixed $x \in [0, 1]$ we find (using l'Hôpital's rule)

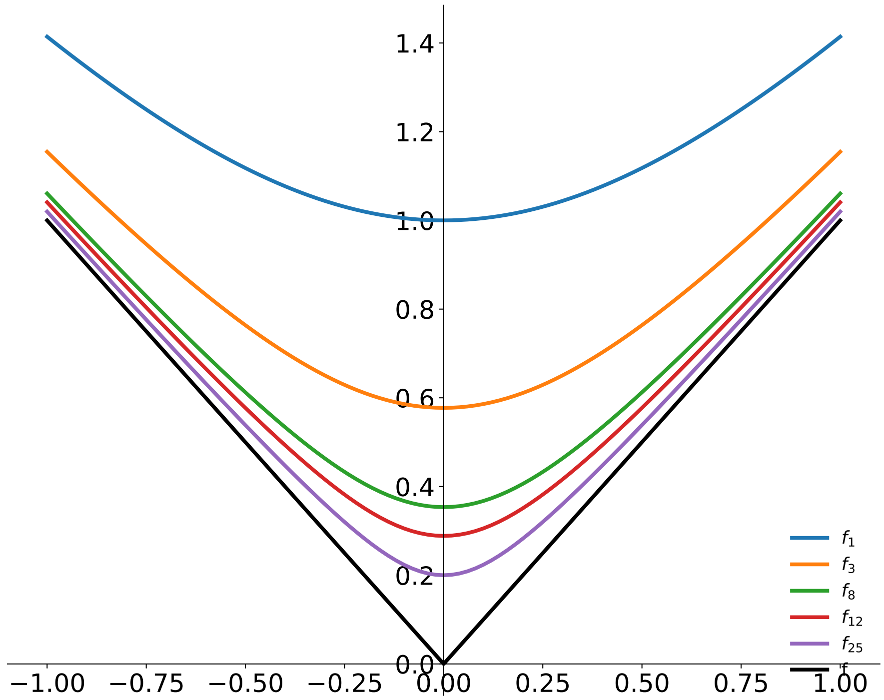


Figure 8.3: Graph of $f_n(x) = \sqrt{\frac{nx^2+1}{n}}$ for $n = 1, 3, 8, 12, 25$ and $f(x) = |x|$

that

$$\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \frac{2nx}{e^{nx^2}} = 0.$$

Hence, (f_n) converges pointwise to $f(x) = 0$ on A. Consider

$$\begin{aligned}\int_0^1 f_n(x)\,dx &= \int_0^1 2nxe^{-nx^2}\,dx \\ &= -e^{-nx^2}\Big|_0^1 \\ &= 1 - e^{-n}\end{aligned}$$

and therefore

$$\lim_{n\to\infty} \int_0^1 f_n(x) = \lim_{n\to\infty} (1 - e^{-n}) = 1.$$

On the other hand, $\int_0^1 f(x)\,dx = 0$. Therefore,

$$\int_0^1 f(x)\,dx \neq \lim_{n\to\infty} \int_0^1 f_n(x)\,dx$$

or another way to write this is

$$\lim_{n\to\infty} \int_0^1 f_n(x)\,dx \neq \int_0^1 \lim_{n\to\infty} f_n(x)\,dx.$$

Examples 8.1.5-8.1.6 illustrate that the pointwise limit f of a sequence of functions (f_n) does not always inherit the properties of continuity and/or differentiability, and Example 8.1.7 illustrates that unexpected (or surprising) results can be obtained when combining the operations of integration and limits, and in particular, one cannot in general interchange the limit operation with integration.

Exercises

Exercise 8.1.1. Suppose that $f_n : [a,b] \to \mathbb{R}$ is a sequence of functions such that f_n is increasing for each $n \in \mathbb{N}$. Suppose that $f(x) = \lim_{n\to\infty} f_n(x)$ exists for each $x \in [a,b]$. Is f an increasing function?

Exercise 8.1.2. Let (a_n) be a sequence of positive numbers and define $f_n : [0,1] \to \mathbb{R}$ as

$$f_n(x) = \begin{cases} 2na_n x, & 0 \le x \le 1/(2n), \\ 2a_n - 2na_n x, & 1/(2n) \le x \le 1/n, \\ 0, & 1/n \le x \le 1. \end{cases}$$

(a) Find the pointwise limit $f : [0,1] \to \mathbb{R}$ of the sequence (f_n).
(b) Find $\int_0^1 f(x)\,dx$.
(c) If $a_n = 4n$, find

$$\lim_{n\to\infty} \int_0^1 f_n(x)\,dx.$$

Exercise 8.1.3. Recall that $\mathbb{Q}$ is countable and thus there exists a bijection $r : \mathbb{N} \to \mathbb{Q}$. Define the sequence (r_n) by letting $r_n = r(n)$. Now define $f_n : \mathbb{R} \to \mathbb{R}$ as

$$f_n(x) = \begin{cases} 1, & x \in \{r_1, r_2, \ldots, r_n\} \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the pointwise limit $f : \mathbb{R} \to \mathbb{R}$ of the sequence (f_n).
(b) Is f_n Riemann integrable? Explain.
(c) Is f Riemann integrable? Explain.

8.2 Uniform Convergence

In the previous section we saw that pointwise convergence is a rather weak form of convergence since the limiting function will not in general inherit any of the properties possessed by the terms of the sequence. Examining the concept of pointwise convergence one observes that it is a very localized definition of convergence of a sequence of functions; all that is asked for is that $(f_n(x))$ converge for each $x \in A$. This allows the possibility that the "speed" of convergence of $(f_n(x))$ may differ wildly as x varies in A. For example, for the sequence of functions $f_n(x) = x^n$ and $x \in (0, 1)$, convergence of $(f_n(x))$ to zero is much faster for values of x near 0 than for values of x near 1. What is worse, as $x \to 1$ convergence of $(f_n(x))$ to zero is arbitrarily slow. Specifically, recall in Example 8.1.5 that $|x^K - 0| < \varepsilon$ if and only if $K > \ln(\varepsilon)/\ln(x)$. Thus, for a fixed $\varepsilon > 0$, as $x \to 1$ we have $K \to \infty$. Hence, there is no single K that will work for all values of $x \in (0, 1)$, that is, the convergence is not uniform.

Definition 8.2.1: Uniform Convergence

Let (f_n) be a sequence of functions on $A \subseteq \mathbb{R}$. We say that (f_n) **converges uniformly on A to the function** $f : A \to \mathbb{R}$ if for any $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that if $n \geq K$ then $|f_n(x) - f(x)| < \varepsilon$ for all $x \in A$.

Notice that in Definition 8.2.1, the K only depends on the given (but fixed) $\varepsilon > 0$ and the inequality $|f_n(x) - f(x)| < \varepsilon$ holds **for all** $x \in A$ provided $n \geq K$. The inequality $|f_n(x) - f(x)| < \varepsilon$ for all $x \in A$ is equivalent to

$$f(x) - \varepsilon < f_n(x) < f(x) + \varepsilon$$

for all $x \in A$ and can therefore be interpreted as saying that the graph of f_n lies in the tube of radius $\varepsilon > 0$ and centered along the graph of f, see Figure 8.4.

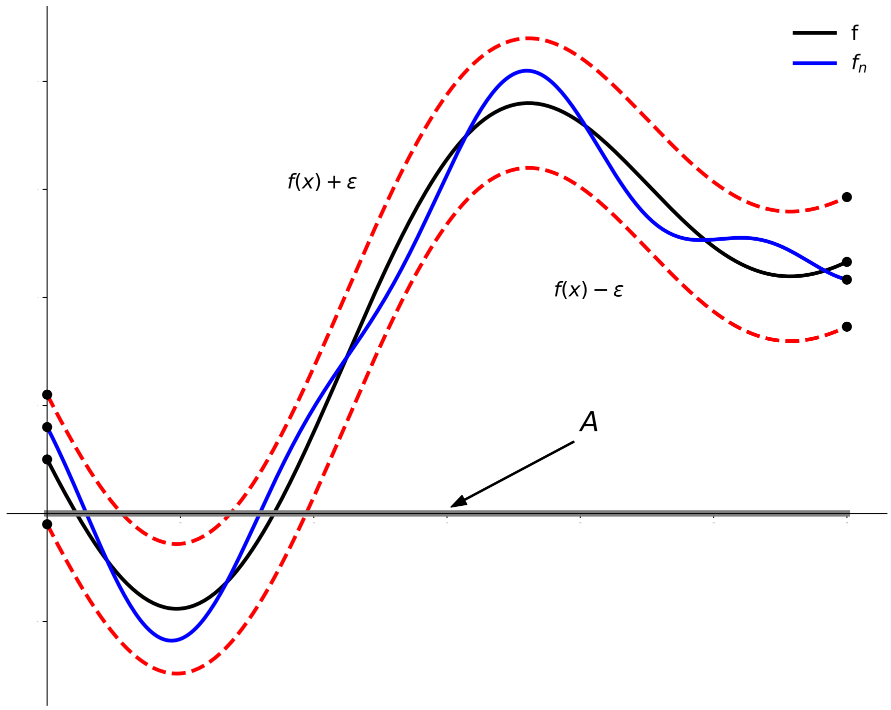


Figure 8.4: ε-tubular neighborhood along the graph of f; if $|f_n(x) - f(x)| < \varepsilon$ for all $x \in A$ then the graph of f_n is within the ε-tubular neighborhood of f

The following result is a direct consequence of the definitions but it is worth stating anyhow.

Proposition 8.2.2

If (f_n) converges uniformly to f then (f_n) converges pointwise to f.

Example 8.2.3. Let $A = [-5, 5]$ and let (f_n) be the sequence of functions on A defined by $f_n(x) = (2xn + (-1)^n x^2)/n$. Prove that (f_n) converges uniformly to $f(x) = 2x$.

Solution. We compute that

$$\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \frac{2xn + (-1)^n x^2}{n} = 2x$$

and thus (f_n) converges pointwise to $f(x) = 2x$ on A. To prove that the convergence is uniform, consider

$$\begin{aligned} |f_n(x) - f(x)| &= \left|\frac{2xn + (-1)^n x^2}{n} - 2x\right| \\ &= \left|\frac{(-1)^n x^2}{n}\right| \\ &= \frac{|x|^2}{n} \\ &\leq \frac{5^2}{n}. \end{aligned}$$

For any given $\varepsilon > 0$ if $K \in \mathbb{N}$ is such that $\frac{5^2}{K} < \varepsilon$ then if $n \geq K$ then for any $x \in A$ we have

$$\begin{aligned} |f_n(x) - f(x)| &\leq \frac{5^2}{n} \\ &\leq \frac{5^2}{K} \\ &< \varepsilon. \end{aligned}$$

This proves that (f_n) converges uniformly to $f(x) = 2x$ on $A = [-5, 5]$. Note that a similar argument will not hold if we take $A = \mathbb{R}$. □

Example 8.2.4. Show that the sequence of functions $f_n(x) = \sin(nx)/\sqrt{n}$ converges uniformly to $f(x) = 0$ on $\mathbb{R}$.

Solution. We compute

$$\begin{aligned}|f_n(x) - f(x)| &= \left|\frac{\sin(nx)}{\sqrt{n}}\right| \\ &= \frac{|\sin(nx)|}{\sqrt{n}} \\ &\leq \frac{1}{\sqrt{n}}\end{aligned}$$

and therefore if $K \in \mathbb{N}$ is such that $\frac{1}{\sqrt{K}} < \varepsilon$ then if $n \geq K$ then $|f_n(x) - 0| < \varepsilon$ for all $x \in \mathbb{R}$. Hence, (f_n) converges uniformly to $f = 0$ on $\mathbb{R}$. ☐

On a close examination of the previous examples on uniform convergence, one observes that in proving that (f_n) converges uniformly to f on A, we used an inequality of the form:

$$|f_n(x) - f(x)| \leq M_n, \ \forall x \in A$$

for some sequence (M_n) of non-negative numbers such that $\lim_{n\to\infty} M_n = 0$. It follows that

$$\sup_{x\in A} |f_n(x) - f(x)| \leq M_n.$$

This observation is worth formalizing.

Theorem 8.2.5

Let $f_n : A \to \mathbb{R}$ be a sequence of functions. Then (f_n) converges uniformly to f on A if and only if there exists a sequence (M_n) of non-negative numbers converging to zero such that $\sup_{x\in A} |f_n(x) - f(x)| \leq M_n$ for n sufficiently large.

Proof. Suppose that (f_n) converges uniformly on A to f. There exists $N \in \mathbb{N}$ such that $|f_n(x) - f(x)| < 1$ for all $n \geq N$ and $x \in A$. Hence,

$M_n = \sup_{x\in A} |f_n(x) - f(x)| \geq 0$ is well-defined for all $n \geq N$. Define $M_n \geq 0$ arbitrarily for $1 \leq n \leq N-1$. Given an arbitrary $\varepsilon > 0$, there exists $K \in \mathbb{N}$ such that if $n \geq K$ then $|f_n(x) - f(x)| < \varepsilon$ for all $x \in A$. We can assume that $K \geq N$. Therefore, if $n \geq K$ then $M_n = \sup_{x\in A} |f_n(x) - f(x)| \leq \varepsilon$. This prove that $\lim_{n\to\infty} M_n = 0$.

Conversely, suppose that there exists $M_n \geq 0$ such that $\lim_{n\to\infty} M_n = 0$ and $\sup_{x\in A} |f_n(x) - f(x)| \leq M_n$ for all $n \geq N$. Let $\varepsilon > 0$ be arbitrary. Then there exists $K \in \mathbb{N}$ such that if $n \geq K$ then $M_n < \varepsilon$. Hence, if $n \geq K \geq N$ then $\sup_{x\in A} |f_n(x) - f(x)| \leq M_n < \varepsilon$. This implies that if $n \geq K$ then $|f_n(x) - f(x)| < \varepsilon$ for all $x \in A$, and thus (f_n) converges uniformly to f on A. ☐

Example 8.2.6. Let f be a continuous function on $[a, b]$. Prove that there exists a sequence of step functions (s_n) on $[a, b]$ that converges uniformly to f on $[a, b]$.

We end this section by stating and proving a Cauchy criterion for uniform convergence.

Theorem 8.2.7: Cauchy Criterion for Uniform Convergence

The sequence (f_n) converges uniformly on A if and only if for every $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that if $n, m \geq K$ then $|f_m(x) - f_n(x)| < \varepsilon$ for all $x \in A$.

Proof. Suppose that $(f_n) \to f$ uniformly on A and let $\varepsilon > 0$. There exists $K \in \mathbb{N}$ such that if $n \geq K$ then $|f_n(x) - f(x)| < \varepsilon/2$ for all

$x \in A$. Therefore, if $n, m \geq K$ then for all $x \in A$ we have

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

To prove the converse, suppose that for every $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that if $n, m \geq K$ then $|f_m(x) - f_n(x)| < \varepsilon$ for all $x \in A$. Therefore, for each $x \in A$ the sequence $(f_n(x))$ is a Cauchy sequence and therefore converges. Let $f : A \to \mathbb{R}$ be defined by $f(x) = \lim_{n\to\infty} f_n(x)$. If $\varepsilon > 0$ let $K \in \mathbb{N}$ be such that $|f_m(x) - f_n(x)| < \varepsilon$ for all $x \in A$ and $n, m \geq K$. Fix $m \geq K$ and consider the sequence $z_n = |f_m(x) - f_n(x)|$ and thus $z_n < \varepsilon$. Now since $\lim_{n\to\infty} f_n(x) = f(x)$ then $\lim z_n$ exists and $\lim z_n \leq \varepsilon$, that is,

$$\begin{aligned}
\lim_{n\to\infty} z_n &= \lim_{n\to\infty} |f_m(x) - f_n(x)| \\
&= |f_m(x) - f(x)| \\
&\leq \varepsilon.
\end{aligned}$$

Therefore, if $m \geq K$ then $|f_m(x) - f(x)| \leq \varepsilon$ for all $x \in A$. $\square$

Exercises

Exercise 8.2.1. Let $f_n : A \to \mathbb{R}$ be a sequence of functions converging uniformly to $f : A \to \mathbb{R}$. Let $g : A \to \mathbb{R}$ be a function and let $g_n = gf_n$ for each $n \in \mathbb{N}$. Under what condition on g does the sequence (g_n) converge uniformly? Prove it. What is the uniform limit of (g_n)?

Exercise 8.2.2. Prove that if (f_n) converges uniformly to f on A and (g_n) converges uniformly to g on A then $(f_n + g_n)$ converges uniformly to $f + g$ on A.

Exercise 8.2.3. Let $f_n : [0, 1] \to$ be the sequence defined in Exercise 8.1.2. Show that if $\lim_{n\to\infty} a_n = 0$ then (f_n) converges uniformly.

Exercise 8.2.4. Let $f_n(x) = \sin(nx)/\sqrt{n}$ for $x \in \mathbb{R}$. Prove that (f_n) converges uniformly on $\mathbb{R}$.

8.3 Properties of Uniform Convergence

A sequence (f_n) on A is said to be **uniformly bounded** on A if there exists a constant $M > 0$ such that $|f_n(x)| < M$ for all $x \in A$ and for all $n \in \mathbb{N}$.

Theorem 8.3.1: Uniform Boundedness

Suppose that $(f_n) \to f$ uniformly on A. If each f_n is bounded on A then the sequence (f_n) is uniformly bounded on A and f is bounded on A.

Proof. By definition, there exists $K \in \mathbb{N}$ such that

$$\begin{aligned} |f(x)| &\leq |f_n(x) - f(x)| + |f_n(x)| \\ &< 1 + |f_n(x)| \end{aligned}$$

for all $x \in A$ and all $n \geq K$. Since f_K is bounded, then $|f(x)| \leq 1 + \max_{x\in A} |f_K(x)|$ for all $x \in A$ and thus f is bounded on A with upper bound $M' = 1 + \max_{x\in A} |f_K(x)|$. Therefore, $|f_n(x)| \leq |f_n - f(x)| + |f(x)| < 1 + M'$ for all $n \geq K$ and all $x \in A$. Let M_n be an upper bounded for f_n on A for each $n \in \mathbb{N}$. Then if $M = \max\{M_1, \ldots, M_{K-1}, 1 + M'\}$ then $|f_n(x)| < M$ for all $x \in A$ and all $n \in \mathbb{N}$. ☐

Example 8.3.2. Give an example of a set A and a sequence of functions (f_n) on A such that f_n is bounded for each $n \in \mathbb{N}$, (f_n) converges pointwise to f but (f_n) is not uniformly bounded on A.

Unlike the case with pointwise convergence, a sequence of continuous functions converging uniformly does so to a continuous function.

Theorem 8.3.3: Uniform Convergence and Continuity

Let (f_n) be a sequence of functions on A converging uniformly to f on A. If each f_n is continuous on A then f is continuous on A.

Proof. To prove that f is continuous on A we must show that f is continuous at each $c \in A$. Let $\varepsilon > 0$ be arbitrary. Recall that to prove that f is continuous at c we must show there exists $\delta > 0$ such that if $|x - c| < \delta$ then $|f(x) - f(c)| < \varepsilon$. Consider the following:

$$\begin{aligned} |f(x) - f(c)| &= |f(x) - f_n(x) + f_n(x) - f_n(c) + f_n(c) - f(c)| \\ &\leq |f(x) - f_n(x)| + |f_n(x) - f_n(c)| + |f_n(c) - f(c)|. \end{aligned}$$

Since $(f_n) \to f$ uniformly on A, there exists $K \in \mathbb{N}$ such that $|f(x) - f_K(x)| < \varepsilon/3$ for all $x \in A$. Moreover, since f_K is continuous there exists $\delta > 0$ such that if $|x - c| < \delta$ then $|f_K(x) - f_K(c)| < \varepsilon/3$. Therefore, if $|x - c| < \delta$ then

$$\begin{aligned} |f(x) - f(c)| &\leq |f(x) - f_K(x)| + |f_K(x) - f_K(c)| + |f_K(c) - f(c)| \\ &< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 \\ &= \varepsilon. \end{aligned}$$

This proves that f is continuous at $c \in A$. □

A direct consequence of Theorem 8.3.3 is that if $(f_n) \to f$ pointwise and each f_n is continuous then if f is discontinuous then the convergence cannot be uniform.

Example 8.3.4. Let $g_n(x) = \dfrac{x}{\sqrt{x^2 + \frac{1}{n}}}$ for $x \in [-1, 1]$ and $n \in \mathbb{N}$. Each function g_n is clearly continuous. Now $g_n(0) = 0$ and thus

$\lim_{n\to\infty} g_n(0) = 0$. If $x \neq 0$ then

$$\begin{aligned}
\lim_{n\to\infty} g_n(x) &= \lim_{n\to\infty} \frac{x}{\sqrt{x^2+\frac{1}{n}}} \\
&= \frac{x}{\sqrt{x^2}} \\
&= \frac{x}{|x|} \\
&= \begin{cases} 1, & x > 0 \\ -1, & x < 0, \end{cases}
\end{aligned}$$

Therefore, (g_n) converges pointwise to the function

$$g(x) = \begin{cases} -1, & -1 \leq x < 0 \\ 0, & x = 0 \\ 1, & 0 < x \leq 1. \end{cases}$$

The function g is discontinuous and therefore, by Theorem 8.3.3, (g_n) does not converge uniformly to g.

The next property that we can deduce from uniform convergence is that the limit and integration operations can be interchanged. Recall from Example 8.1.7 that if $(f_n) \to f$ pointwise then it is not necessarily true that

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

Since $\lim_{n\to\infty} f_n(x) = f(x)$, then it in general we can say that

$$\lim_{n\to\infty} \int_A f_n \neq \int_A \lim_{n\to\infty} f_n$$

However, when the convergence is uniform we can indeed interchange the limit and integration operations.

Theorem 8.3.5: Uniform Convergence and Integration

Let (f_n) be a sequence of Riemann integrable functions on $[a, b]$. If (f_n) converges uniformly to f on $[a, b]$ then $f \in \mathcal{R}[a, b]$ and

$$\lim_{n\to\infty} \int_a^b f_n = \int_a^b f.$$

Proof. Let $\varepsilon > 0$ be arbitrary. By uniform convergence, there exists $K \in \mathbb{N}$ such that if $n \geq K$ then for all $x \in [a, b]$ we have

$$|f_n(x) - f(x)| < \frac{\varepsilon}{4(b-a)}$$

or

$$f_n(x) - \frac{\varepsilon}{4(b-a)} < f(x) < f_n(x) + \frac{\varepsilon}{4(b-a)}.$$

By assumption, $f_n \pm \frac{\varepsilon}{4(b-a)}$ is Riemann integrable and thus if $n \geq N$ then

$$\int_a^b [(f_n + \varepsilon/4(b-a)) - (f_n - \varepsilon/4(b-a))] = \frac{\varepsilon}{2} < \varepsilon.$$

By the Squeeze Theorem of Riemann integration (Theorem 7.2.3), f is Riemann integrable. Moreover, if $n \geq N$ then

$$-\frac{\varepsilon}{4(b-a)} < f_n(x) - f(x) < \frac{\varepsilon}{4(b-a)}$$

implies (by monotonicity of integration)

$$-\frac{\varepsilon}{4} < \int_a^b f_n - \int_a^b f < \frac{\varepsilon}{4}$$

and thus

$$\left| \int_a^b f_n - \int_a^b f \right| < \frac{\varepsilon}{4}.$$

This proves that the sequence $\int_a^b f_n$ converges to $\int_a^b f$. □

The following corollary to Theorem 8.3.5 is worth noting.

Corollary 8.3.6: Uniform Convergence and Integration

Let (f_n) be a sequence of continuous functions on the interval $[a, b]$. If (f_n) converges uniformly to f then $f \in \mathcal{R}[a, b]$ and

$$\lim_{n\to\infty} \int_a^b f_n = \int_a^b f.$$

Proof. If each f_n is continuous then $f_n \in \mathcal{R}[a, b]$ and Theorem 8.3.5 applies. □

Example 8.3.7. Consider the sequence of functions (f_n) defined on $[0, 1]$ given by

$$f_n(x) = \begin{cases} (n+1)^2 x, & 0 \leq x \leq \frac{1}{n+1} \\ -(n+1)^2\left(x - \frac{2}{n+1}\right), & \frac{1}{n+1} \leq x \leq \frac{2}{n+1} \\ 0, & \frac{2}{n+1} < x \leq 1. \end{cases}$$

(a) Draw a typical function f_n.

(b) Prove that (f_n) converges pointwise.

(c) Use Theorem 8.3.5 to show that the convergence is not uniform.

We now consider how the operation of differentiation behaves under uniform convergence. One would hope, based on the results of Theorem 8.3.3, that if $(f_n) \to f$ uniformly and each f_n is differentiable then f' is also differentiable and maybe even that $(f_n') \to f'$ at least pointwise and maybe even uniformly. Unfortunately, the property of differentiability is not generally inherited under uniform convergence. An example of this occurred in Example 8.1.6 where $f_n(x) = \sqrt{(nx^2+1)/n}$ and $(f_n) \to f$ where $f(x) = |x|$ for $x \in [-1, 1]$. The convergence in this case is uniform on $[-1, 1]$ but although each f_n is differentiable the limit function $f(x) = |x|$ is not. It turns out that the main assumption needed for all to be well is that the sequence (f'_n) converge uniformly.

Theorem 8.3.8: Uniform Convergence and Differentiation

Let (f_n) be a sequence of differentiable functions on $[a, b]$. Assume that f'_n is Riemann integrable on $[a, b]$ for each $n \in \mathbb{N}$ and suppose that (f'_n) converges uniformly to g on $[a, b]$. Suppose there exists $x_0 \in [a, b]$ such that $(f_n(x_0))$ converges. Then the sequence (f_n) converges uniformly on $[a, b]$ to a differentiable function f and $f' = g$.

Proof. Let $x \in [a, b]$ be arbitrary but with $x \neq x_0$. By the Mean Value theorem applied to the differentiable function $f_m - f_n$, there exists y in between x and x_0 such that

$$\frac{(f_m(x) - f_n(x)) - (f_m(x_0) - f_n(x_0))}{x - x_0} = f'_m(y) - f'_n(y)$$

or equivalently

$$f_m(x) - f_n(x) = f_m(x_0) - f_n(x_0) + (x - x_0)(f'_m(y) - f'_n(y))$$

Therefore,

$$|f_m(x) - f_n(x)| \leq |f_m(x_0) - f_n(x_0)| + (b - a)|f'_m(y) - f'_n(y)|.$$

Since $(f_n(x_0))$ converges and (f'_n) is uniformly convergent, by the Cauchy criterion, for any $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that if $n, m \geq K$ then $|f_m(x_0) - f_n(x_0)| < \varepsilon/2$ and $|f'_m(y) - f'_n(y)| < (\varepsilon/2)/(b - a)$ for all $y \in [a, b]$. Therefore, if $m, n \geq K$ then

$$\begin{aligned}|f_m(x) - f_n(x)| &\leq |f_m(x_0) - f_n(x_0)| + (b - a)|f'_m(y) - f'_n(y)| \\ &< \varepsilon\end{aligned}$$

and this holds for all $x \in [a,b]$. By the Cauchy criterion for uniform convergence, (f_n) converges uniformly. Let f be the uniform limit of (f_n). We now prove that f is differentiable and $f' = g$. By the Fundamental theorem of Calculus (FTC), we have that

$$f_n(x) = f_n(a) + \int_a^x f_n'(t)\,dt$$

for each $x \in [a,b]$. Since (f_n) converges to f and (f_n') converges uniformly to g we have

$$\begin{aligned} f(x) &= \lim_{n\to\infty} f_n(x) \\ &= \lim_{n\to\infty} \left(f_n(a) + \int_a^x f_n'(t)\,dt \right) \\ &= \lim_{n\to\infty} f_n(a) + \lim_{n\to\infty} \int_a^x f_n'(t)\,dt \\ &= f(a) + \int_a^x g(t)\,dt. \end{aligned}$$

Thus $f(x) = f(a) + \int_a^x g(t)\,dt$ and by the FTC we obtain $f'(x) = g(x)$. □

Notice that in the statement of Theorem 8.3.8, all that is required is that $(f_n(x_0))$ converge for one $x_0 \in [a,b]$. The assumption that (f_n') converges uniformly then guarantees that in fact (f_n) converges uniformly.

Example 8.3.9. Consider the sequence (f_n) defined on $[-1,1]$ by $f_n(x) = (2xn + (-1)^n x^2)/n$. We compute that $f_n'(x) = (2n + 2(-1)^n x)/n$ and clearly f_n' is continuous on $[-1,1]$ for each $n \in \mathbb{N}$. Now $\lim_{n\to\infty} f_n'(x) = 2$ for all x and thus (f_n') converges pointwise to $g(x) = 2$. To prove that the convergence is uniform we note that

$$\begin{aligned} |f_n'(x) - g(x)| &= |2 + (-1)^n \frac{x}{n} - 2| \\ &= \frac{|x|}{n} \\ &\leq \frac{1}{n}. \end{aligned}$$

Therefore, (f_n') converges uniformly to g on $[-1, 1]$. Now $f_n(0) = 0$ and thus $(f_n(0))$ converges to 0. By Theorem 8.3.8, (f_n) converges uniformly to say f with $f(0) = 0$ and $f' = g$. Now by the FTC, $f(x) = \int g(x)\, dx + C = 2x + C$ and since $f(0) = 0$ then $f(x) = 2x$.

Exercises

Exercise 8.3.1. Give an example of a set A and a sequence of functions (f_n) on A such that f_n is bounded for each $n \in \mathbb{N}$, (f_n) converges pointwise to f but (f_n) is not uniformly bounded on A.

Exercise 8.3.2. Suppose that $(f_n) \to f$ uniformly on A and $(g_n) \to g$ uniformly on A. Prove that if (f_n) and (g_n) are uniformly bounded on A then $(f_n g_n)$ converges uniformly to fg on A. Then give an example to show that if one of (f_n) or (g_n) is not uniformly bounded then the result is false.

Exercise 8.3.3. Let

$$f_n(x) = \frac{nx^2}{1 + nx^2}$$

for $x \in \mathbb{R}$ and $n \in \mathbb{N}$.

(a) Show that (f_n) converges pointwise on $\mathbb{R}$.

(b) Show that (f_n) does not converge uniformly on any closed interval containing 0.

(c) Show that (f_n) converges uniformly on any closed interval not containing 0. For instance, take $[a, b]$ with $0 < a < b$.

Exercise 8.3.4. Suppose that $f : \mathbb{R} \to \mathbb{R}$ has that property that $|f(x) - f(y)| \leq K|x - y|$ for all $x, y, \in \mathbb{R}$ and some $K > 0$. Prove that if (g_n) converges uniformly on $\mathbb{R}$ to g then the sequence $(f \circ g_n)$ converges uniformly to $f \circ g$ on $\mathbb{R}$. Note: $f \circ g_n$ and $f \circ g$ are compositions of functions and not function multiplication.

Exercise 8.3.5. Let $f_n(x) = nx/(nx + 1)$ for $n \in \mathbb{N}$ and $x \in [a, 1]$ where $0 < a < 1$.

(a) Prove directly that the sequence (f_n) is uniformly Cauchy.

(b) If f is the uniform limit of (f_n), find $\int_a^1 f$ without computing f.

Exercise 8.3.6. Consider the sequence of functions (f_n) on $A = [0, \infty)$ defined as follows:

$$f_n(x) = \begin{cases} 1/n, & 0 \leq x \leq n^2, \\ 0, & x > n^2. \end{cases}$$

(a) Prove that (f_n) converges uniformly to $f = 0$ on A.

(b) For each fixed $n \in \mathbb{N}$, find the improper integral

$$\int_0^\infty f_n$$

and show that

$$\lim_{n\to\infty} \int_0^\infty f_n = \infty.$$

(c) The results above seem to contradict Theorem 8.3.5. Explain why there is no contradiction.

8.4 Infinite Series of Functions

In this section, we consider series whose terms are functions. You have already encountered such objects when studying power series in Calculus. An example of an infinite series of functions (more specifically a power series) is

$$\sum_{n=0}^{\infty} \frac{(-1)^n x^n}{(2n)!}.$$

In this case, if we set $f_n(x) = \frac{(-1)^n x^n}{(2n)!}$ then the above infinite series is $\sum_{n=0}^{\infty} f_n(x)$. Let us give the general definition.

Definition 8.4.1: Infinite Series of Functions

Let A be a non-empty subset of $\mathbb{R}$. An **infinite series of functions** on A is a series of the form $\sum_{n=1}^{\infty} f_n(x)$ for each $x \in A$ where (f_n) is a sequence of functions on A. The **sequence of partial sums** generated by the series $\sum f_n$ is the sequence of functions (s_n) on A defined as $s_n(x) = f_1(x) + \cdots + f_n(x)$ for each $x \in A$.

Recall that a series of numbers $\sum x_n$ converges if the sequence of partial sums (t_n), defined as $t_n = x_1 + x_2 + \cdots + x_n$, converges. Hence, convergence of an infinite series of functions $\sum f_n$ is treated by considering the convergence of the sequence of partial sums (s_n) (which are functions). For example, to say that the series $\sum f_n$ converges uniformly to a function f we mean that the sequence of partial sums (s_n) converges uniformly to f, etc. It is now clear that our previous work in Sections 8.1-8.3 translate essentially directly to infinite series of functions. As an example:

Theorem 8.4.2

Let (f_n) be a sequence of functions on A and suppose that $\sum f_n$ converges uniformly to f. If each f_n is continuous on A then f is continuous on A.

Proof. By assumption, the sequence of functions $s_n(x) = \sum_{k=1}^n f_k(x)$ for $x \in A$ converges uniformly to f. Since each function f_n is continuous, and the sum of continuous functions is continuous, it follows that s_n is continuous. The result now follows by Theorem 8.3.3. □

The following translate of Theorem 8.3.5 is worth explicitly writing out.

Theorem 8.4.3: Term-by-Term Integration

Let (f_n) be a sequence of functions on $[a, b]$ and suppose that $\sum f_n$ converges uniformly to f. If each f_n is Riemann integrable on $[a, b]$ then $f \in \mathcal{R}[a, b]$ and

$$\int_a^b \left(\sum_{n=1}^\infty f_n(x) \right) dx = \sum_{n=1}^\infty \int_a^b f_n(x)\, dx.$$

Proof. By assumption, the sequence (s_n) defined as $s_n(x) = f_1(x)+\cdots+f_n(x)$ converges uniformly to f. Since each f_n is Riemann integrable then s_n is Riemann integrable and therefore $f = \lim s_n = \sum f_n$ is Riemann integrable by Theorem 8.3.5. Also by Theorem 8.3.5, we have

$$\int_a^b f = \lim_{n\to\infty} \int_a^b s_n$$

or written another way is

$$\int_a^b \sum_{n=1}^\infty f_n = \lim_{n\to\infty} \int_a^b \sum_{k=1}^n f_k$$

or

$$\int_a^b \sum_{n=1}^{\infty} f_n = \lim_{n\to\infty} \sum_{k=1}^{n} \int_a^b f_k$$

or

$$\int_a^b \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty} \int_a^b f_n$$

□

We now state the derivative theorem (similar to Theorem 8.3.8) for infinite series of functions.

Theorem 8.4.4: Term-by-Term Differentiation

Let (f_n) be a sequence of differentiable functions on $[a, b]$ and suppose that $\sum f_n$ converges at some point $x_0 \in [a, b]$. Assume further that $\sum f_n'$ converges uniformly on $[a, b]$ and each f_n' is continuous. Then $\sum f_n$ converges uniformly to some differentiable function f on $[a, b]$ and $f' = \sum f_n'$.

We now state a useful theorem for uniform convergence of infinite series of functions.

Theorem 8.4.5: Weierstrass M-Test

Let (f_n) be a sequence of functions on A and suppose that there exists a sequence of non-negative numbers (M_n) such that $|f_n(x)| \leq M_n$ for all $x \in A$, and all $n \in \mathbb{N}$. If $\sum M_n$ converges then $\sum f_n$ converges uniformly on A.

Proof. Let $\varepsilon > 0$ be arbitrary. Let $t_n = \sum_{k=1}^{n} M_k$ be the sequence of partial sums of the series $\sum M_n$. By assumption, (t_n) converges and thus (t_n) is a Cauchy sequence. Hence, there exists $K \in \mathbb{N}$ such that

$|t_m - t_n| < \varepsilon$ for all $m > n \geq K$. Let (s_n) be the sequence of partial sums of $\sum f_n$. Then if $m > n \geq K$ then for all $x \in A$ we have

$$\begin{aligned}
|s_m(x) - s_n(x)| &= |f_m(x) + f_{m-1}(x) + \cdots + f_{n+1}(x)| \\
&\leq |f_m(x)| + |f_{m-1}(x)| + \cdots + |f_{n+1}(x)| \\
&\leq M_m + M_{m-1} + \cdots + M_{n+1} \\
&= |t_m - t_n| \\
&< \varepsilon.
\end{aligned}$$

Hence, the sequence (s_n) satisfies the Cauchy Criterion for uniform convergence (Theorem 8.2.7) and the proof is complete. □

Example 8.4.6. Prove that

$$\int_0^{\pi} \left(\sum_{n=1}^{\infty} \frac{n \sin(nx)}{e^n} \right) = \frac{2e}{e^2 - 1}$$

Proof. For any $x \in \mathbb{R}$ it holds that

$$\left| \frac{n \sin(nx)}{e^n} \right| \leq \frac{n}{e^n}.$$

A straightforward application of the Ratio test shows that $\sum_{n=1}^{\infty} \frac{n}{e^n}$ is a convergent series. Hence, by the M-Test, the given series converges uniformly on $A = \mathbb{R}$, and in particular on $[0, \pi]$. By Theorem 8.4.3,

$$\begin{aligned}
\int_0^\pi \sum_{n=1}^\infty \frac{n\sin(nx)}{e^n}\,dx &= \sum_{n=1}^\infty \int_0^\pi \frac{n\sin(nx)}{e^n}\,dx \\
&= \sum_{n=1}^\infty -\frac{\cos(nx)}{e^n}\Big|_0^\pi \\
&= \sum_{n=1}^\infty \left[\left(\tfrac{1}{e}\right)^n - \left(\tfrac{-1}{e}\right)^n\right] \\
&= \left(\tfrac{1}{1-1/e} - 1\right) - \left(\tfrac{1}{1+1/e} - 1\right) \\
&= \frac{2e}{e^2-1}
\end{aligned}$$

□

Example 8.4.7 (Riemann (1853)). Consider the function $r(x)$ whose graph is given in Figure 8.5; one can write down an explicit expression for $r(x)$ but the details are unimportant. Consider the series

$$\sum_{n=1}^\infty \frac{r(nx)}{n^2}.$$

Since

$$\left|\frac{r(nx)}{n^2}\right| \le \frac{1/2}{n^2}$$

and $\sum_{n=1}^\infty \frac{1}{2n^2}$ converges, then by the M-test the above series converges uniformly on any interval $[a, b]$. Let f be the function defined by the series on $[a, b]$. Now, on $[a, b]$, the function $f_n(x) = \frac{r(nx)}{n^2}$ has only a finite number of discontinuities and thus f_n is Riemann integrable. Therefore, by Theorem 8.3.5, the function f is Riemann integrable. The graph of f is shown in Figure 8.6. One can show that f has discontinuities at the rational points $x = \frac{p}{2q}$ where $\gcd(p, q) = 1$.

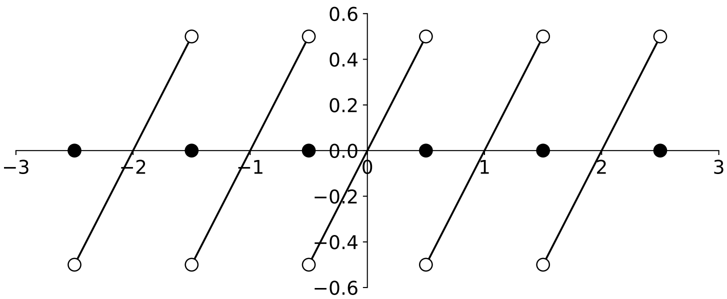


Figure 8.5: The function $r(x)$

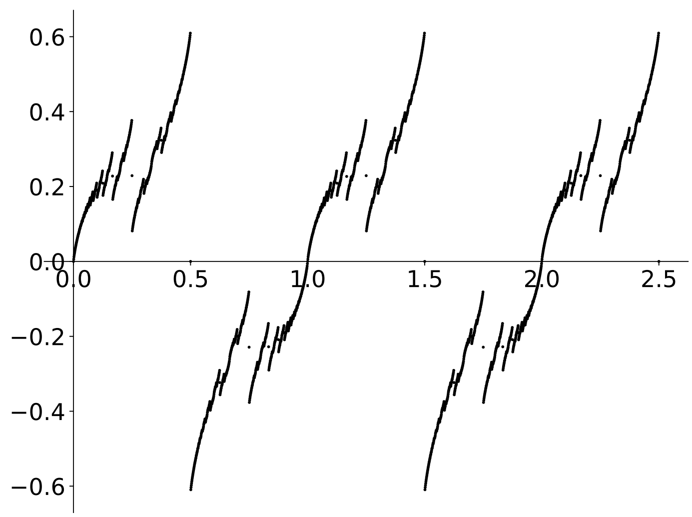


Figure 8.6: The function $f(x) = \sum_{n=1}^{\infty} \frac{r(nx)}{n^2}$

Example 8.4.8 (Power Series)**.** Recall that a **power series** is a series of the form

$$\sum_{n=0}^{\infty} c_n(x-a)^n$$

where $c_n \in \mathbb{R}$ and $a \in \mathbb{R}$. Hence, in this case if we write the series as $\sum f_n(x)$ then $f_n(x) = c_n(x-a)^n$ for each $n \in \mathbb{N}$ and $f_0(x) = c_0$. In calculus courses, the main problem you were asked to solve is to find the interval of convergence of the given power series. The main tool is to apply the Ratio test (Theorem 3.7.23):

$$\lim_{n\to\infty} \frac{|c_{n+1}||x-a|^{n+1}}{|c_n||x-a|^n} = |x-a| \lim_{n\to\infty} \frac{|c_{n+1}|}{|c_n|}.$$

Suppose that $\lim_{n\to\infty} \frac{|c_{n+1}|}{|c_n|}$ exists and is non-zero and $\lim_{n\to\infty} \frac{|c_{n+1}|}{|c_n|} = \frac{1}{R}$ (a similar argument can be done when the limit is zero). Then by the Ratio test, the power series converges if $|x-a|\frac{1}{R} < 1$, that is, if $|x-a| < R$. The number $R > 0$ is called the **radius of convergence** and the interval $(a-R, a+R)$ is the **interval of convergence** (if $\lim_{n\to\infty} \frac{|c_{n+1}|}{|c_n|}$ is zero then $R > 0$ can be chosen arbitrarily and the argument that follows is applicable). Let $\rho < r < R$ and consider the closed interval $[a-\rho, a+\rho] \subset (a-R, a+R)$. Then if $x \in [a-\rho, a+\rho]$ then

$$\begin{aligned} |f_n(x)| &= |c_n||x-a|^n \\ &= |c_n| r^n \frac{|x-a|^n}{r^n} \\ &\leq |c_n| r^n \left(\frac{\rho}{r}\right)^n. \end{aligned}$$

Now if $x = a + r \in (a-R, a+R)$ then by assumption the series $\sum c_n(x-a)^n = \sum c_n r^n$ converges, and in particular the sequence $|c_n|r^n$ is bounded, say by M. Therefore,

$$|f_n(x)| \leq M \left(\frac{\rho}{r}\right)^n.$$

Since $\rho/r < 1$, the geometric series $\sum \left(\frac{\rho}{r}\right)^n$ converges. Therefore, by the M-test, the series $\sum c_n(x-a)^n$ converges uniformly on the interval $[a-\rho, a+\rho]$. Let $f(x) = \sum f_n(x)$ for $x \in [a-\rho, a+\rho]$. Now consider the series of the derivatives

$$\sum_{n=1}^{\infty} f_n'(x) = \sum_{n=1}^{\infty} c_n n (x-a)^{n-1}.$$

Applying the Ratio test again we conclude that the series of the derivatives converges for each $x \in (a - R, a + R)$ and a similar argument as before shows that the series of derivatives converges uniformly on any interval $[a - \rho, a + \rho]$ where $\rho < R$. It follows from the Term-by-Term Differentiation theorem that f is differentiable and $f'(x) = \sum c_n n(x - a)^{n-1}$. By the Term-by-Term Integration theorem, we can also integrate the series and

$$\int_I \left(\sum f_n(x)\right) dx = \sum \int_I f_n(x)\, dx$$

where $I \subset (a - R, a + R)$ is any closed and bounded interval.

Example 8.4.9. Consider the power series

$$\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!} \quad \text{and} \quad \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}.$$

(a) Prove that the series converges at every $x \in \mathbb{R}$.

(b) Let f denote the function defined by the series on the left and let g denote the function defined by the series on the right. Justifying each step, show that f' exists and that $f' = g$.

(c) Similarly, show that g' exists and $g' = -f$.

Example 8.4.10. A **Fourier series** is a series of the form

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos(nx) + b_n \sin(nx))$$

where $a_n, b_n \in \mathbb{R}$.

(a) Suppose that for a given (a_n) and (b_n), the associated Fourier series converges pointwise on $[-\pi, \pi]$ and let f be the pointwise limit. Prove that in fact the Fourier series converges on $\mathbb{R}$. Hint: For any $y \in \mathbb{R}$ there exists $x \in [-\pi, \pi]$ such that $y = x + 2\pi$.

(b) Prove that if $\sum |a_n|$ and $\sum |b_n|$ are convergent series then the associated Fourier series converges uniformly on $\mathbb{R}$.

(c) Suppose that for a given (a_n) and (b_n), the associated Fourier series converges uniformly on $[-\pi, \pi]$ and let f be the uniform limit. Prove the following:

$$a_0 = \int_{-\pi}^{\pi} f(x)dx$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)dx$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)dx$$

You will need the following identities:

$$\int_{-\pi}^{\pi} \sin(nx)\cos(mx)dx = 0, \quad \forall\, n, m \in N$$

$$\int_{-\pi}^{\pi} \sin(nx)\sin(mx)dx = \int_{-\pi}^{\pi} \cos(nx)\cos(mx)dx = \begin{cases} \pi, & m = n \\ 0, & m \neq n \end{cases}$$

Example 8.4.11 (Dini). (a) By an **open cover** of an interval $[a, b]$, we mean a collection of open intervals $\{I_\mu\}_{\mu\in X}$ such that $I_\mu \cap [a, b] \neq \emptyset$ for each $\mu \in X$ and

$$[a, b] \subset \bigcup_{\mu\in X} I_\mu.$$

Here X is some set, possibly uncountable. Prove that if $\{I_\mu\}_{\mu\in X}$ is any open cover of $[a, b]$ then there there exists finitely many $\mu_1, \mu_2, \ldots, \mu_N \in X$ such that

$$[a, b] \subset \bigcup_{k=1}^{N} I_{\mu_k}.$$

(b) Let f_n be continuous and suppose that $f_{n+1}(x) \leq f_n(x)$ for all $x \in [a, b]$ and all $n \in \mathbb{N}$. Suppose that (f_n) converges pointwise to a continuous function f. Prove that the convergence is actually uniform. Give an example to show that if f is not continuous then we only have pointwise convergence.

Proof. We first prove (a). For convenience, write $I_\mu = (a_\mu, b_\mu)$ for each $\mu \in X$ and assume without loss of generality that $\{b_\mu \mid \mu \in X\}$ is bounded above. Let $b_0 = a$ and let I_{μ_1} be such that $b_0 \in I_{\mu_1} = (a_1, b_1)$ and

$$b_1 = \sup\{b_\mu \mid b_0 \in I_\mu\}.$$

If $b_1 > b$ then we are done because then $[a, b] \subset I_{\mu_1}$. By induction, having defined $b_{k-1} \in [a, b]$, let $I_{\mu_k} = (a_k, b_k)$ be such that $b_{k-1} \in I_{\mu_k}$ and $b_k = \sup\{b_\mu \mid b_{k-1} \in I_\mu\}$. We claim that $b_N > b$ for some $N \in \mathbb{N}$ and thus $[a, b] \subset \cup_{k=1}^N I_{\mu_k}$. To prove the claim, suppose that $b_k \leq b$ for all $k \in \mathbb{N}$. Then the increasing sequence (b_k) converges by the Monotone Convergence theorem, say to $L = \sup\{b_1, b_2, \ldots\}$. Since $L \in [a, b]$ then $L \in I_\mu$ for some $\mu \in X$ and thus by convergence there exists $k \in \mathbb{N}$ such that $b_k \in I_\mu = (a_\mu, b_\mu)$. However, by definition of b_{k+1} we must have that $L < b_\mu \leq b_{k+1}$ which is a contradiction to the definition of L. This completes the proof.

Now we prove (b). First of all since $f_m(x) \leq f_n(x)$ for all $m \geq n$ it holds that $f(x) \leq f_n(x)$ for all $n \in \mathbb{N}$ and all $x \in [a, b]$. Fix $\tilde{x} \in [a, b]$ and let $\varepsilon > 0$. By pointwise convergence, there exists $N \in \mathbb{N}$ such that $|f_n(\tilde{x}) - f(\tilde{x})| < \varepsilon/3$ for all $n \geq N$. By continuity of f_N and f, there exists $\delta_{\tilde{x}} > 0$ such that $|f_N(x) - f_N(\tilde{x})| < \varepsilon/3$ and $|f(x) - f(\tilde{x})| < \varepsilon/3$

for all $x \in I_{\tilde{x}} = (\tilde{x} - \delta_{\tilde{x}}, \tilde{x} + \delta_{\tilde{x}})$. Therefore, if $n \geq N$ then

$$\begin{aligned} |f_n(x) - f(x)| &= f_n(x) - f(x) \\ &\leq f_N(x) - f(x) \\ &\leq |f_N(x) - f_N(\tilde{x})| + |f_N(\tilde{x} - f(\tilde{x})| + |f(\tilde{x}) - f(x)| \\ &< \varepsilon \end{aligned}$$

for all $x \in I_{\tilde{x}}$. Hence, (f_n) converges uniformly to f on the interval $I_{\tilde{x}}$. It is clear that $\{I_{\tilde{x}}\}_{\tilde{x} \in [a,b]}$ is an open cover of $[a, b]$. Therefore, by part (a), there exists $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_k$ such that $[a, b] \subset I_{\tilde{x}_1} \cup \cdots \cup I_{\tilde{x}_k}$. Hence, for arbitrary $\varepsilon > 0$ there exists $N_j \in \mathbb{N}$ such that if $n \geq N_j$ then $|f_n(x) - f(x)| < \varepsilon$ for all $x \in I_{\tilde{x}_j}$. If $N = \max\{N_1, N_2, \ldots, N_k\}$ then $|f_n(x) - f(x)| < \varepsilon$ for all $n \geq N$ and all $x \in [a, b]$. This completes the proof. The sequence $f_n(x) = x^n$ on $[0, 1]$ satisfies $f_{n+1}(x) \leq f_n(x)$ for all n and all $x \in [0, 1]$, and (f_n) converges to $f(x) = 0$ if $x \in [0, 1)$ and $f(1) = 1$. Since f is not continuous on $[0, 1]$, the convergence is not uniform. □

Example 8.4.12. Let g_n be continuous and suppose that $g_n \geq 0$ for all $n \in \mathbb{N}$. Prove that if $\sum_{n=1}^{\infty} g_n$ converges pointwise to a continuous function f on $[a, b]$ then in fact the convergence is uniform.

Exercises

Exercise 8.4.1. Show that $\sum_{n=0}^{\infty} x^n$ converges uniformly on $[-a, a]$ for every a such that $0 < a < 1$. Then show that the given series does not converge uniformly on $(-1, 1)$. Hint: This is an important series and you should know what function the series converges uniformly to.

Exercise 8.4.2. If $\sum_{n=1}^{\infty} |a_n| < \infty$ prove that $\sum_{n=1}^{\infty} a_n \sin(nx)$ converges uniformly on $\mathbb{R}$.

Exercise 8.4.3. Prove, justifying each step, that

$$\int_1^2 \left(\sum_{n=1}^{\infty} ne^{-nx} \right) dx = \frac{e}{e^2 - 1}$$

Exercise 8.4.4. For any number $q \in \mathbb{R}$ let $\chi_q : \mathbb{R} \to \mathbb{R}$ be the function defined as $\chi_q(x) = 1$ if $x = q$ and $\chi_q(x) = 0$ if $x \neq q$. Let $\{q_1, q_2, q_3, \ldots\} = \mathbb{Q}$ be an enumeration of the rational numbers. Define $f_n : \mathbb{R} \to \mathbb{R}$ as

$$f_n = \chi_{q_1} + \chi_{q_2} + \cdots + \chi_{q_n}.$$

Find the pointwise limit f of the sequence (f_n). Is the convergence uniform? Explain.

9

Metric Spaces

9.1 Metric Spaces

The main concepts of real analysis on $\mathbb{R}$ can be carried over to a general set M once a notion of "distance" $d(x, y)$ has been defined for points $x, y \in M$. When $M = \mathbb{R}$, the distance we have been using all along is $d(x, y) = |x - y|$. The set $\mathbb{R}$ along with the distance function $d(x, y) = |x - y|$ is an example of a metric space.

Definition 9.1.1: Metric Space

Let M be a non-empty set. A **metric** on M is a function $d : M \times M \to [0, \infty)$ satisfying the following properties:

(i) $d(x, y) = 0$ if and only if $x = y$

(ii) $d(x, y) = d(y, x)$ for all $x, y \in M$ (symmetry)

(iii) $d(x, y) \leq d(x, z) + d(z, y)$ for all $x, y, z \in M$ (triangle inequality)

A **metric space** is a pair (M, d) where d is a metric on M.

If the metric d is understood, then we simply refer to M as a metric space instead of formally referring to the pair (M, d).

Example 9.1.2. The set $M = \mathbb{R}$ and function $d(x, y) = |x - y|$ is a metric space. To see this, first of all $|x - y| = 0$ iff $x - y = 0$ iff $x = y$. Second of all, $|x - y| = |-(y - x)| = |y - x|$, and finally by the usual triangle inequality on $\mathbb{R}$ we have

$$d(x, y) = |x - y| = |x - z + y - z| \leq |x - z| + |z - y| = d(x, z) + d(z, y)$$

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

Example 9.1.3. Let $\mathcal{B}([a, b])$ denote the set of bounded functions on the interval $[a, b]$, that is, $f \in \mathcal{B}([a, b])$ if there exists $M > 0$ such that $|f(x)| \leq M$ for all $x \in [a, b]$. For $f, g \in \mathcal{B}([a, b])$ let

$$d(f, g) = \sup_{x \in [a,b]} |f(x) - g(x)|.$$

We claim that $(\mathcal{B}([a, b]), d)$ is a metric space. First of all, if $f, g \in \mathcal{B}([a, b])$ then using the triangle inequality it follows that $(f - g) \in \mathcal{B}([a, b])$. Therefore, $d(f, g)$ is well-defined for all $f, g \in \mathcal{B}([a, b])$. Next, by definition, we have that $0 \leq d(f, g)$ and it is clear that $d(f, g) = d(g, f)$. Lastly, for $f, g, h \in \mathcal{B}([a, b])$ since

$$|f(x) - g(x)| \leq |f(x) - h(x)| + |h(x) - g(x)|$$

then

$$\begin{aligned} d(f,g) &= \sup_{x\in[a,b]} |f(x)-g(x)| \\ &\le \sup_{x\in[a,b]} (|f(x)-h(x)| + |h(x)-g(x)|) \\ &\le \sup_{x\in[a,b]} (|f(x)-h(x)|) + \sup_{x\in[a,b]} (|h(x)-g(x)|) \\ &= d(f,h) + d(h,g). \end{aligned}$$

This proves that $(\mathcal{B}[a,b], d)$ is a metric space. It is convention to denote the metric $d(f,g)$ as $d_\infty(f,g)$, and we will follow this convention.

Example 9.1.4. Let M be a non-empty set. Define $d(x,y) = 1$ if $x \neq y$ and $d(x,y) = 0$ if $x = y$. It is straightforward to show that (M,d) is a metric space. The metric d is called the **discrete metric** and (M,d) is a **discrete space**.

Example 9.1.5. Let (M,d) be a metric space and let $M' \subset M$ be a non-empty subset. Let d' be the restriction of d onto M', that is, $d' : M' \times M' \to [0,\infty)$ is defined as $d'(x,y) = d(x,y)$ for $x,y \in M'$. Then (M',d') is a metric space. We may therefore say that M' is a **metric subspace** of M.

Example 9.1.6. Let $C([a,b])$ denote the set of continuous functions on the inteval $[a,b]$. Then $C([a,b]) \subset \mathcal{B}([a,b])$ and thus $(C([a,b]), d_\infty)$ is a metric subspace of $(\mathcal{B}([a,b]), d_\infty)$.

Example 9.1.7. For $f,g \in C([a,b])$ let $d(f,g) = \int_a^b |f(t)-g(t)|\, dt$. Prove that d defines a metric on $C([a,b])$.

Example 9.1.8. Let $\mathbb{R}^{n\times n}$ denote the set of $n \times n$ matrices with real entries. For $A, B \in \mathbb{R}^{n\times n}$ define

$$d(A,B) = \max_{1\le i,j\le n} |a_{i,j} - b_{i,j}|.$$

It is clear that $d(A,B) = d(B,A)$ and $d(A,B) = 0$ if and only if $A = B$. For $A, B, C \in \mathbb{R}^{n\times n}$ we have that

$$\begin{aligned}
d(A,B) &= \max_{1\le i,j\le n} |a_{i,j} - b_{i,j}| \\
&= \max_{1\le i,j\le n} |a_{i,j} - c_{i,j} + c_{i,j} - b_{i,j}| \\
&\le \max_{1\le i,j\le n} (|a_{i,j} - c_{i,j}| + |c_{i,j} - b_{i,j}|) \\
&\le \max_{1\le i,j\le n} |a_{i,j} - c_{i,j}| + \max_{1\le i,j\le n} |c_{i,j} - b_{i,j}| \\
&= d(A,C) + d(C,B).
\end{aligned}$$

Hence, $(\mathbb{R}^{n\times n}, d)$ is a metric space.

An important class of metric spaces are **normed vector spaces**.

Definition 9.1.9: Norms

Let V be a vector space over $\mathbb{R}$ (or $\mathbb{C}$). A **norm** on V is a function $\psi : V \to [0, \infty)$ satisfying the following properties:

(i) $\psi(x) = 0$ if and only if $x = 0$,

(ii) $\psi(\alpha x) = |\alpha|\psi(x)$ for any scalar $\alpha \in \mathbb{R}$ and any $x \in V$, and

(iii) $\psi(x + y) \leq \psi(x) + \psi(y)$ for all $x, y \in V$.

The number $\psi(x)$ is called the norm of $x \in V$. A vector space V together with a norm ψ is called a **normed vector space**.

Instead of using the generic letter ψ to denote a norm, it is convention to use instead $\|\cdot\|$. Hence, using $\|x\|$ to denote the norm $\psi(x)$, properties (i)-(iii) are:

(i) $\|x\| = 0$ if and only if $x = 0$,

(ii) $\|\alpha x\| = |\alpha| \, \|x\|$ for any scalar $\alpha \in \mathbb{R}$ and any $x \in V$, and

(iii) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in V$.

Let $(V, \|\cdot\|)$ be a normed vector space and define $d : V \times V \to [0, \infty)$ by

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

It is a straightforward exercise (which you should do) to show that (V, d) is a metric space. Hence, every normed vector space induces a metric space.

Example 9.1.10. The real numbers $V = \mathbb{R}$ form a vector space over $\mathbb{R}$ under the usual operations of addition and multiplication. The absolute value function $x \mapsto |x|$ is a norm on $\mathbb{R}$. The induced metric is then $(x, y) \mapsto |x - y|$.

Example 9.1.11. The **Euclidean norm** on $\mathbb{R}^n$ is defined as

$$\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

for $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$. It can be verified that $\|\cdot\|_2$ is indeed a norm on $\mathbb{R}^n$. Hence, we define the distance between $x, y \in \mathbb{R}^n$ as

$$\|x - y\|_2 = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}.$$

Notice that when $n = 1$, $\|\cdot\|_2$ is the absolute value function since $\|x\|_2 = \sqrt{x^2} = |x|$ for $x \in \mathbb{R}$. When not specified otherwise, whenever we refer to $\mathbb{R}^n$ as a normed vector space we implicitly assume that the norm is $\|\cdot\|_2$ and simply use the notation $\|\cdot\|$.

Example 9.1.12 (Important)**.** The set $\mathcal{B}([a, b])$ of bounded functions forms a vector space over $\mathbb{R}$ with addition defined as $(f + g)(x) = f(x) + g(x)$ for $f, g \in \mathcal{B}([a, b])$ and scalar multiplication defined as $(\alpha f)(x) = \alpha f(x)$ for $\alpha \in \mathbb{R}$ and $f \in \mathcal{B}([a, b])$. For $f \in \mathcal{B}([a, b])$ let

$$\|f\|_\infty = \sup_{a \leq x \leq b} |f(x)|.$$

It is left as an (important) exercise to show that $\|\cdot\|_\infty$ is indeed a norm on $\mathcal{B}([a, b])$. The induced metric is

$$d_\infty(f, g) = \|f - g\|_\infty = \sup_{a \leq x \leq b} |f(x) - g(x)|.$$

The norm $\|f\|_\infty$ is called the **sup-norm** of f. Notice that the metric in Example 9.1.3 is induced by the norm $\|\cdot\|_\infty$.

Example 9.1.13. Two examples of norms on $C([a, b])$ are

$$\|f\|_1 = \int_a^b |f(x)|\, dx$$

and

$$\|f\|_2 = \left(\int_a^b |f(x)|^2\, dx \right)^{1/2}.$$

These norms are important in the analysis of Fourier series.

Example 9.1.14. For $A \in \mathbb{R}^{n\times n}$ let

$$\|A\|_\infty = \max_{1\leq i,j\leq n} |a_{i,j}|.$$

It is left as an exercise to show that $\|\cdot\|_\infty$ defined above is a norm on $\mathbb{R}^{n\times n}$.

Let (M, d) be a metric space. For $x \in M$ and $r > 0$, the **open ball** centered at x of radius r is by definition the set

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

Example 9.1.15. Interpret geometrically the open balls in the normed spaces $(\mathbb{R}^n, \|\cdot\|)$ for $n \in \{1, 2, 3\}$.

Example 9.1.16. Give a graphical/geometric description of the open balls in the normed space $(C([a, b]), \|\cdot\|_\infty)$.

A subset S of a metric space M is called **bounded** if $S \subset B_r(x)$ for some $x \in M$ and $r > 0$.

Example 9.1.17. Let (M, d) be a metric space. Prove that if S is bounded then there exists $y \in S$ and $r > 0$ such that $S \subset B_r(y)$.

Exercises

Exercise 9.1.1. Let H be the set of all real sequences $x = (x_1, x_2, x_3, \ldots)$ such that $|x_n| \leq 1$ for all $n \in \mathbb{N}$. For $x, y \in H$ let

$$d(x, y) = \sum_{n=1}^{\infty} 2^{-n}|x_n - y_n|.$$

Prove that d is a metric on H. **Note:** Part of what you have to show is that $d(x, y)$ is well-defined which means to show that $\sum_{n=1}^{\infty} 2^{-n}|x_n - y_n|$ converges if $x, y \in H$.

9.2 Sequences and Limits

Let M be a metric space. A **sequence** in M is a function $z : \mathbb{N} \to M$. As with sequences of real numbers, we identify a sequence $z : \mathbb{N} \to M$ with the infinite list $(z_n) = (z_1, z_2, z_3, \ldots)$ where $z_n = z(n)$ for $n \in \mathbb{N}$.

Definition 9.2.1: Convergence of Sequences

Let (M, d) be a metric space. A sequence (z_n) in M is said to **converge** if there exists $p \in M$ such that for any given $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that $d(z_n, p) < \varepsilon$ for all $n \geq K$. In this case, we write

$$\lim_{n\to\infty} z_n = p$$

or $(z_n) \to p$, and we call p the **limit** of (z_n). If (z_n) does not converge then we say it is **divergent**.

One can indeed show, just as in Theorem 3.1.12 for sequences of real numbers, that the point p in Definition 9.2.1 is indeed unique.

Remark 9.2.2. Suppose that (z_n) converges to p and let $x_n = d(z_n, p) \geq 0$. Hence, (x_n) is a sequence of real numbers. If $(z_n) \to p$ then for any $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that $x_n < \varepsilon$. Thus, $x_n = d(z_n, p) \to 0$. Conversely, if $d(z_n, p) \to 0$ then clearly $(z_n) \to p$.

Example 9.2.3 (Important)**.** Prove that a sequence (f_n) converges to f in the normed vector space $(\mathcal{B}([a, b], \|\cdot\|_\infty)$ if and only if (f_n) converges uniformly to f on $[a, b]$.

Several of the results for sequences of real numbers carry over to sequences on a general metric space. For example, a sequence (z_n) in M is said to be **bounded** if the set $\{z_n \mid n \in \mathbb{N}\}$ is bounded in M. Then:

Lemma 9.2.4

In a metric space, a convergent sequence is bounded.

Proof. Suppose that p is the limit of (z_n). There exists $K \in \mathbb{N}$ such that $d(z_n, p) < 1$ for all $n \geq K$. Let $r = 1 + \max\{d(z_1, p), \ldots, d(z_{K-1}, p)\}$ and we note that $r \geq 1$. Then $\{z_n \mid n \in \mathbb{N}\} \subset B_r(p)$. To see this, if $n \geq K$ then $d(z_n, p) < 1 \leq r$ and thus $z_n \in B_r(p)$. On the other hand, for $z_j \in \{z_1, \ldots, z_{K-1}\}$ we have that $d(z_j, p) \leq \max\{d(z_1, p), \ldots, d(z_{K-1}, p)\} < r$, and thus $z_j \in B_r(p)$. This proves that (z_n) is bounded. $\square$

A **subsequence** of a sequence (z_n) is a sequence of the form $(y_k) = (z_{n_k})$ where $n_1 < n_2 < n_3 < \cdots$. Then (compare with Theorem 3.4.5):

Lemma 9.2.5

Let M be a metric space and let (z_n) be a sequence in M. If $(z_n) \to p$ then $(z_{n_k}) \to p$ for any subsequence (z_{n_k}) of (z_n).

A sequence (z_n) in M is called a **Cauchy sequence** if for any given $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that $d(z_n, z_m) < \varepsilon$ for all $n, m \geq K$. Then (compare with Lemma 3.6.3-3.6.4):

Lemma 9.2.6

Let M be a metric space and let (z_n) be a sequence in M. The following hold:

(i) If (z_n) is convergent then (z_n) is a Cauchy sequence.

(ii) If (z_n) is a Cauchy sequence then (z_n) is bounded.

(iii) If (z_n) is a Cauchy sequence and if (z_n) has a convergent subsequence then (z_n) converges.

Proof. Proofs for (i) and (ii) are left as exercises (see Lemma 3.6.3-3.6.4). To prove (iii), let (z_{n_k}) be a convergent subsequence of (z_n), say converging to p. Let $\varepsilon > 0$ be arbitrary. There exists $K \in \mathbb{N}$ such that $d(z_n, z_m) < \varepsilon/2$ for all $n, m \geq K$. By convergence of (z_{n_k}) to p, by increasing K if necessary we also have that $d(z_{n_k}, p) < \varepsilon/2$ for all $k \geq K$. Therefore, if $n \geq K$, then since $n_K \geq K$ then

$$\begin{aligned} d(z_n, p) &\leq d(z_n, z_{n_K}) + d(z_{n_K}, p) \\ &< \varepsilon/2 + \varepsilon/2 \\ &= \varepsilon. \end{aligned}$$

Hence, $(z_n) \to p$. □

The previous lemmas (that were applicable on a general metric space) show that some properties of sequences in $\mathbb{R}$ are due entirely to the metric space structure of $\mathbb{R}$. There are, however, important results on $\mathbb{R}$, most notably the Bolzano-Weierstrass theorem and the Cauchy criterion for convergence, that do not generally carry over to a general metric space. The Bolzano-Weierstrass theorem and the Cauchy criterion rely on the completeness property of $\mathbb{R}$ and there is no reason to believe that a general metric space comes equipped with a similar completeness property. Besides, the completeness axiom of $\mathbb{R}$ (Axiom 2.4.6) relies on the order property of $\mathbb{R}$ (i.e., $\leq$) and there is no reason to believe that a general metric space comes equipped with an order. We will have more to say about this in Section 9.4. For now, however, we will consider an important metric space where almost all the results for sequences in $\mathbb{R}$ carry over (an example of a result not carrying over is the Monotone convergence theorem), namely, the normed vector space $(\mathbb{R}^n, \|\cdot\|)$.

Denoting a sequence in $\mathbb{R}^n$ is notationally cumbersome. Formally, a sequence in $\mathbb{R}^n$ is a function $z : \mathbb{N} \to \mathbb{R}^n$. How then should we denote $z(k)$ as a vector in $\mathbb{R}^n$? One way is to simply write $z(k) = (z_1(k), z_2(k), \ldots, z_n(k))$ for each $k \in \mathbb{N}$ and this is the notation we will adopt. It is clear that a sequence $(z(k))$ in $\mathbb{R}^n$ induces n sequences in $\mathbb{R}$, namely, $(z_i(k))$ for each $i \in \{1, 2, \ldots, n\}$ (i.e., the component sequences). The following theorem explains why $\mathbb{R}^n$ inherits almost all the results for sequences in $\mathbb{R}$.

Theorem 9.2.7: Convergence in Euclidean Spaces

Let $(z(k)) = (z_1(k), z_2(k), \ldots, z_n(k))$ be a sequence in the normed vector space $(\mathbb{R}^n, \|\cdot\|)$. Then $(z(k))$ converges if and only if for each $i \in \{1, 2, \ldots, n\}$ the component sequence $(z_i(k))$ converges. Moreover, if $(z(k))$ converges then

$$\lim_{k\to\infty} z(k) = (\lim_{k\to\infty} z_1(k), \lim_{k\to\infty} z_2(k), \ldots, \lim_{k\to\infty} z_n(k)).$$

Proof. Suppose first that $(z(k))$ converges, say to $p = (p_1, p_2, \ldots, p_n)$. For any $i \in \{1, 2, \ldots, n\}$ it holds that

$$|z_i(k) - p_i| \leq \sqrt{(z_1(k) - p_1)^2 + (z_2(k) - p_2)^2 + \cdots + (z_n(k) - p_n)^2}$$

in other words, $|z_i(k) - p_i| \leq \|z(k) - p\|$. Since $(z(k)) \to p$ then $\lim_{k\to\infty} \|z(k) - p\| = 0$ and consequently $\lim_{k\to\infty} |z_i(k) - p_i| = 0$, that is, $\lim_{k\to\infty} z_i(k) = p_i$.

Conversely, now suppose that $(z_i(k))$ converges for each $i \in \{1, 2, \ldots, n\}$. Let $p_i = \lim_{k\to\infty} z_i(k)$ for each $i \in \{1, 2, \ldots, n\}$ and let $p = (p_1, p_2, \ldots, p_n)$. By the basic limit laws of sequences in $\mathbb{R}$, the sequence

$$\begin{aligned} x_k &= \|z(k) - p\| \\ &= \sqrt{(z_1(k) - p_1)^2 + (z_2(k) - p_2)^2 + \cdots + (z_n(k) - p_n)^2} \end{aligned}$$

converges to zero since $\lim_{k\to\infty}(z_i(k) - p_i)^2 = 0$ and the square root function $x \mapsto \sqrt{x}$ is continuous. Thus, $\lim_{k\to\infty} z(k) = p$ as desired. ☐

Corollary 9.2.8

Every Cauchy sequence in $\mathbb{R}^n$ is convergent.

Proof. Let $(z(k))$ be a Cauchy sequence in $\mathbb{R}^n$. Hence, for $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that $\|z(k) - z(m)\| < \varepsilon$ for all $k, m \geq K$. Thus, for any $i \in \{1, 2, \ldots, n\}$, if $k, m \geq K$ then

$$|z_i(k) - z_i(m)| \leq \|z(k) - z(m)\| < \varepsilon.$$

Thus, $(z_i(k))$ is a Cauchy sequence in $\mathbb{R}$, and is therefore convergent by the completeness property of $\mathbb{R}$. By Theorem 9.2.7, this proves that $(z(k))$ is convergent. ☐

Corollary 9.2.9: Bolzano-Weierstrass in Euclidean Space

Every bounded sequence in $(\mathbb{R}^n, \|\cdot\|)$ has a convergent subsequence.

Proof. Let $(z(k))$ be a bounded sequence in $\mathbb{R}^n$. There exists $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $r > 0$ such that $z(k) \in B_r(x)$ for all $k \in \mathbb{N}$, that is, $\|z(k) - x\| < r$ for all $k \in \mathbb{N}$. Therefore, for any $i \in \{1, 2 \ldots, n\}$ we have

$$|z_i(k) - x_i| \leq \|z(k) - x\| < r, \quad \forall\, k \in \mathbb{N}.$$

This proves that $(z_i(k))$ is a bounded sequence in $\mathbb{R}$ for each $i \in \{1, 2, \ldots, n\}$. We now proceed by induction. If $n = 1$ then $(z(k))$ is just a (bounded) sequence in $\mathbb{R}$ and therefore, by the Bolzano-Weierstrass theorem on $\mathbb{R}$, $(z(k))$ has a convergent subsequence. Assume by induction that for some $n \geq 1$, every bounded sequence in $\mathbb{R}^n$ has a convergent subsequence. Let $(z(k))$ be a bounded sequence in $\mathbb{R}^{n+1}$. Let $(\tilde{z}(k))$ be the sequence in $\mathbb{R}^n$ such that $\tilde{z}(k) \in \mathbb{R}^n$ is the vector of the first n components of $z(k) \in \mathbb{R}^{n+1}$. Then $(\tilde{z}(k))$ is a bounded sequence in $\mathbb{R}^n$ (why?). By induction, $(\tilde{z}(k))$ has a convergent subsequence, say it is $(\tilde{z}(k_j))$. Now, the real sequence $y_j = z_{n+1}(k_j) \in \mathbb{R}$ is bounded and therefore by the Bolzano-Weierstrass theorem on $\mathbb{R}$, (y_j) has a convergent subsequence which we denote by $(u_\ell) = (y_{j_\ell})$, that is, $u_\ell = z_{n+1}(k_{j_\ell})$. Now, since $w_\ell = \tilde{z}(k_{j_\ell})$ is a subsequence of the convergent sequence $(\tilde{z}(k_j))$, (w_ℓ) converges in $\mathbb{R}^n$. Thus, each component of the sequence $(z(k_{j_\ell}))$ in $\mathbb{R}^{n+1}$ is convergent and since $(z(k_{j_\ell}))$ is a subsequence of the sequence $(z(k)))$ the proof is complete. ☐

Definition 9.2.10

Let M be a metric space.

(a) A subset U of M is said to be **open** if for any $x \in U$ there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subset U$.

(b) A subset E of M is **closed** if $E^c = M \backslash E$ is open.

Example 9.2.11. Prove that an open ball $B_\varepsilon(x) \subset M$ is open. In other words, prove that for each $y \in B_\varepsilon(x)$ there exists $\delta > 0$ such that $B_\delta(y) \subset B_\varepsilon(x)$.

Example 9.2.12. Below are some facts that are easily proved; once (a) and (b) are proved use DeMorgan's Laws to prove (c) and (d).

(a) If $U_1, \ldots, U_n$ is a finite collection of open sets then $\bigcap_{k=1}^n U_k$ is open.

(b) If $\{U_k\}$ is collection of open sets indexed by a set I then $\bigcup_{k\in I} U_k$ is open.

(c) If $E_1, \ldots, E_n$ is a finite collection of closed sets then $\bigcup_{k=1}^n E_k$ is closed.

(d) If $\{E_k\}$ is collection of closed sets indexed by a set I then $\bigcap_{k\in I} E_k$ is closed.

Below is a characterization of closed sets via sequences.

Theorem 9.2.13: Closed Sets via Sequences

Let M be a metric space and let $E \subset M$. Then E is closed if and only if every sequence in E that converges does so to a point in E, that is, if $(x_n) \to x$ and $x_n \in E$ then $x \in E$.

Proof. Suppose that E is closed and let (x_n) be a sequence in E. If $x \in E^c$ then there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subset E^c$. Hence, $x_n \notin B_\varepsilon(x)$ for all n and thus (x_n) does not converge to x. Hence, if (x_n) converges then it converges to a point in E. Conversely, assume that every sequence in E that converges does so to a point in E and let $x \in E^c$ be arbitrary. Then by assumption, x is not the limit point of any converging sequence in E. Hence, there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subset E^c$ otherwise we can construct a sequence in E converging to x (how?). This proves that E^c is open and thus E is closed. $\square$

Example 9.2.14. Show that $C([a, b])$ is a closed subset of $\mathcal{B}([a, b])$.

Example 9.2.15. Let M be an arbitrary non-empty set and let d be the discrete metric, that is, $d(x, y) = 1$ if $x \neq y$ and $d(x, y) = 0$ if $x = y$. Describe the converging sequences in (M, d). Prove that every subset of M is both open and closed.

Example 9.2.16. For $a \leq b$, prove that $[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$ is closed.

Exercises

Exercise 9.2.1. Let M be a metric space and suppose that (z_n) converges in M. Prove that the limit of (z_n) is unique. In other words, prove that if p and q satisfy the convergence definition for (z_n) then $p = q$.

Exercise 9.2.2. Let (M, d) be a metric space.

(a) Let $y \in M$ be fixed. Prove that if (z_n) converges to p then $\lim_{n\to\infty} d(z_n, y) = d(p, y)$.

(b) Prove that if (z_n) converges to p and (y_n) converges to q then $\lim_{n\to\infty} d(z_n, y_n) = d(p, q)$.

Exercise 9.2.3. Let M be a metric space. Prove that if $U_1, \ldots, U_n \subset M$ are open then $\bigcap_{k=1}^{n} U_k$ is open in M.

Exercise 9.2.4. Let (M, d) be a metric space and let $E \subset M$. A point $x \in M$ is called a **limit point** (or **cluster point**) of E if there exists a sequence (x_n) in E, with $x_n \neq x$ for all n, converging to x. The **closure** of E, denoted by $\mathrm{cl}(E)$, is the union of E and the limit points of E. If $\mathrm{cl}(E) = M$ then we say that E is **dense** in M. As an example, $\mathbb{Q}$ is dense in $\mathbb{R}$ since every irrational number is the limit of a sequence of rational numbers.

(a) Prove that E is dense in M if and only if $E \cap U \neq \emptyset$ for every open set U of M.

(b) Let E be the set of step functions on $[a, b]$. Then clearly $E \subset \mathcal{B}([a, b])$. Prove that the set of continuous function $C([a, b])$ is contained in the closure of E. (Hint: See Example 8.2.6)

(c) Perform an internet search and find dense subsets of $(C([a,b]), \|\cdot\|_\infty)$ (you do not need to supply proofs).

Exercise 9.2.5. For $x \in \mathbb{R}^n$ define $\|x\|_\infty = \max_{1\leq i\leq n} |x_i|$ and $\|x\|_1 = \sum_{i=1}^n |x_i|$. It is not hard to verify that these are norms on $\mathbb{R}^n$. Prove that:

(a) $\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1$,

(b) $\|x\|_1 \leq n \|x\|_\infty$, and

(c) $\|x\|_1 \leq \sqrt{n} \|x\|_2$

Two metrics d and ρ on a set M are **equivalent** if they generate the same convergent sequences, in other words, (x_n) converges in (M, d) if and only if (x_n) converges in (M, ρ). Prove that $\|\cdot\|_1, \|\cdot\|_2, \|\cdot\|_\infty$ are equivalent norms on $\mathbb{R}^n$.

Exercise 9.2.6. How do we (Riemann) integrate functions from $[a, b]$ to $\mathbb{R}^n$? Here is how. First, we equip $\mathbb{R}^n$ with the standard Euclidean norm $\|\cdot\|_2$. For any function $F : [a, b] \to \mathbb{R}^n$ and any tagged partition $\dot{\mathcal{P}} = \{([t_{k-1}, t_k], c_k)\}_{k=1}^n$ of $[a, b]$, define the Riemann sum

$$S(F; \dot{\mathcal{P}}) = \sum_{k=1}^n F(c_k)(x_k - x_{k-1}).$$

We then say that F is Riemann integrable if there exists $v \in \mathbb{R}^n$ such that for any $\varepsilon > 0$ there exists $\delta > 0$ such that for any tagged partition $\dot{\mathcal{P}}$ of $[a, b]$ with $\|\dot{\mathcal{P}}\| < \delta$ it holds that

$$\|S(F; \dot{\mathcal{P}}) - v\|_2 < \varepsilon.$$

We then also write that $\int_a^b F = v$. If F has component functions $F = (f_1, f_2, \ldots, f_n)$, prove that F is Riemann integrable on $[a, b]$ if and only if the component functions $f_1, f_2, \ldots, f_n$ are Riemann integrable on $[a, b]$, and in this case,

$$\int_a^b F = \Big(\int_a^b f_1, \int_a^b f_2, \ldots, \int_a^b f_n \Big).$$

Hint: Recall that for any $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ it holds that $|x_i| \leq \|x\|_2$.

Exercise 9.2.7. Let $\mathbb{R}^\infty$ denote the set of infinite sequences in $\mathbb{R}$. It is not hard to see that $\mathbb{R}^\infty$ is a $\mathbb{R}$-vector space with addition and scalar multiplication defined in the obvious way. Let $\ell_1 \subset \mathbb{R}^\infty$ denote the subset of sequences $x = (x_1, x_2, x_3, \ldots)$ such that $\sum_{n=1}^\infty |x_n|$ converges, that is, ℓ_1 denotes the set of absolutely convergent series.

(a) Prove that ℓ_1 is a subspace of $\mathbb{R}^\infty$, i.e., prove that ℓ_1 is closed under addition and scalar multiplication.

(b) For $x \in \ell_1$ let $\|x\|_1 = \sum_{n=1}^\infty |x_n|$. Prove that $\|\cdot\|_1$ defines a norm on ℓ_1.

9.3 Continuity

Using the definition of continuity for a function $f : \mathbb{R} \to \mathbb{R}$ as a guide, it is a straightforward task to formulate a definition of continuity for a function $f : M_1 \to M_2$ where (M_1, d_1) and (M_2, d_2) are metric spaces.

Definition 9.3.1: Continuity

Let (M_1, d_1) and (M_2, d_2) be metric spaces. A function $f : M_1 \to M_2$ is **continuous at** $x \in M_1$ if given any $\varepsilon > 0$ there exists $\delta > 0$ such that if $d_1(y, x) < \delta$ then $d_2(f(y), f(x)) < \varepsilon$. We say that f is **continuous** if it is continuous at each point of M_1.

Using open balls, $f : M_1 \to M_2$ is continuous at $x \in M_1$ if for any given $\varepsilon > 0$ there exists $\delta > 0$ such that $f(y) \in B_\varepsilon(f(x))$ whenever $y \in B_\delta(x)$. We note that $B_\varepsilon(f(x))$ is an open ball in M_2 while $B_\delta(x)$ is an open ball in M_1.

Below we characterize continuity using sequences (compare with Theorem 5.1.2).

Theorem 9.3.2: Sequential Criterion for Continuity

Let (M_1, d_1) and (M_2, d_2) be metric spaces. A function $f : M_1 \to M_2$ is continuous at $x \in M_1$ if and only if for every sequence (x_n) in M_1 converging to x the sequence $(f(x_n))$ in M_2 converges to $f(x)$.

Proof. Assume that f is continuous at $x \in M_1$ and let (x_n) be a sequence in M_1 converging to x. Let $\varepsilon > 0$ be arbitrary. Then there exists $\delta > 0$ such that $f(y) \in B_\varepsilon(f(x))$ for all $y \in B_\delta(x)$. Since $(x_n) \to x$, there exists $K \in \mathbb{N}$ such that $x_n \in B_\delta(x)$ for all $n \geq K$. Therefore, for

$n \geq K$ we have that $f(x_n) \in B_\varepsilon(f(x))$. Since $\varepsilon > 0$ is arbitrary, this proves that $(f(x_n))$ converges to $f(x)$.

Suppose that f is not continuous at x. Then there exists $\varepsilon^* > 0$ such that for every $\delta > 0$ there exists $y \in B_\delta(x)$ with $f(y) \notin B_{\varepsilon^*}(f(x))$. Hence, if $\delta_n = \frac{1}{n}$ then there exists $x_n \in B_{\delta_n}(x)$ such that $f(x_n) \notin B_{\varepsilon^*}(f(x))$. Since $d(x_n, x) < \delta_n$ then $(x_n) \to x$. On the other hand, it is clear that $(f(x_n))$ does not converge to $f(x)$. Hence, if f is not continuous at x then there exists a sequence (x_n) converging to x such that $(f(x_n))$ does not converge to $f(x)$. This proves that if every sequence (x_n) in M_1 converging to x it holds that $(f(x_n))$ converges to $f(x)$ then f is continuous at x. ☐

Example 9.3.3. A **level set** of a function $f : M \to \mathbb{R}$ is a set of the form $E = \{x \in M \mid f(x) = k\}$ for some $k \in \mathbb{R}$. Prove that if f is continuous then the level sets of f are closed sets.

As a consequence of Theorem 9.3.2, if f is continuous at p and $\lim_{n\to\infty} x_n = p$ then $\lim_{n\to\infty} f(x_n) = f(p)$ can be written as

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

The sequential criteria for continuity can be conveniently used to show that the composition of continuous function is a continuous function.

Lemma 9.3.4

Let $f : M_1 \to M_2$ and let $g : M_2 \to M_3$, where M_1, M_2, and M_3 are metric spaces. If f is continuous at $x \in M_1$ and g is continuous at $f(x)$ then the composite mapping $(g \circ f) : M_1 \to M_3$ is continuous at $x \in M_1$.

Proof. If $\lim_{n\to\infty} x_n = p$ then by Theorem 9.3.2, and using the fact that f is continuous at p, and g is continuous at $g(p)$:

$$\begin{aligned}(g \circ f)(p) &= g(f(p)) \\ &= g(f(\lim_{n\to\infty} x_n)) \\ &= g(\lim_{n\to\infty} f(x_n)) \\ &= \lim_{n\to\infty} g(f(x_n)) \\ &= \lim_{n\to\infty} (g \circ f)(x_n)\end{aligned}$$

□

In general, given functions $f, g : M_1 \to M_2$ on metric spaces M_1 and M_2, there is no general way to define the functions $f \pm g$ or fg since M_2 does not come equipped with a vector space structure nor is it equipped with a product operation. However, when $M_2 = \mathbb{R}$ then $f(x)$ and $g(x)$ are real numbers which can therefore be added/subtracted/multiplied.

Proposition 9.3.5

Let (M, d) be a metric space and let $f, g : M \to \mathbb{R}$ be continuous functions, where $\mathbb{R}$ is equipped with the usual metric. If f and g are continuous at $x \in M$ then $f + g$, $f - g$, and fg are continuous at $x \in M$.

Proof. In all cases, the most economical proof is to use the sequential criterion. The details are left as an exercise. □

Recall that for any function $f : A \to B$ and $S \subset B$ the set $f^{-1}(S)$

is defined as

$$f^{-1}(S) = \{x \in A \mid f(x) \in S\}.$$

Example 9.3.6. For any function $f : A \to B$ prove that $(f^{-1}(S))^c = f^{-1}(S^c)$ for any $S \subset B$.

Proposition 9.3.7: Continuity via Open and Closed Sets

For a given function $f : (M_1, d_1) \to (M_2, d_2)$ the following are equivalent:

(i) f is continuous on M_1.

(ii) $f^{-1}(U)$ is open in M_1 for every open subset $U \subset M_2$.

(iii) $f^{-1}(E)$ is closed in M_1 for every closed subset $E \subset M_2$.

Proof. (i) $\Longrightarrow$ (ii): Assume that f is continuous on M_1 and let $U \subset M_2$ be open. Let $x \in f^{-1}(U)$ and thus $f(x) \in U$. Since U is open, there exists $\varepsilon > 0$ such that $B_\varepsilon(f(x)) \subset U$. By continuity of f, there exists $\delta > 0$ such that if $y \in B_\delta(x)$ then $f(y) \in B_\varepsilon(f(x))$. Therefore, $B_\delta(x) \subset f^{-1}(U)$ and this proves that $f^{-1}(U)$ is open.

(ii) $\Longrightarrow$ (i): Let $x \in M_1$ and let $\varepsilon > 0$ be arbitrary. Since $B_\varepsilon(f(x))$ is open, by assumption $f^{-1}(B_\varepsilon(f(x)))$ is open. Clearly $x \in f^{-1}(B_\varepsilon(f(x)))$ and thus there exists $\delta > 0$ such that $B_\delta(x) \subset f^{-1}(B_\varepsilon(f(x)))$, in other words, if $y \in B_\delta(x)$ then $f(y) \in B_\varepsilon(f(x))$. This proves that f is continuous at x.

(ii) $\Longleftrightarrow$ (iii): This follows from the fact that $(f^{-1}(U))^c = f^{-1}(U^c)$ for any set U. Thus, for instance, if $f^{-1}(U)$ is open for every open set U then if E is closed then $f^{-1}(E^c)$ is open, that is, $(f^{-1}(E))^c$ is open, i.e., $f^{-1}(E)$ is closed. $\square$

Example 9.3.8. Use Proposition 9.3.7 to prove that the level sets of a function $f : M \to \mathbb{R}$ on a metric space M are closed sets.

Example 9.3.9. A function $f : (M_1, d_1) \to (M_2, d_2)$ is called **Lipschitz** on M_1 if there exists $K > 0$ such that $d_2(f(x), f(y)) \leq K d_1(x, y)$ for all $x, y \in M_1$. Prove that a Lipschitz function is continuous.

Example 9.3.10. For $A \in \mathbb{R}^{n\times n}$ recall that $\|A\|_\infty = \sup_{1\leq i,j\leq n} |a_{i,j}|$. Let $\mathrm{tr} : \mathbb{R}^{n\times n} \to \mathbb{R}$ be the trace function on $\mathbb{R}^{n\times n}$, that is, $\mathrm{tr}(A) = \sum_{i=1}^n a_{i,i}$. Show that tr is Lipschitz and therefore continuous.

Let ℓ_∞ denote the set of all real sequences (x_n) that are bounded, that is, $\{|x_n| \; : \; n \in \mathbb{N}\}$ is a bounded set. If $x = (x_n) \in \ell_\infty$, it is straightforward to verify that $\|x\|_\infty = \sup_{n\in\mathbb{N}} |x_n|$ defines a norm on ℓ_∞ with addition and scalar multiplication defined in the obvious way. Let ℓ_1 be the set of absolutely summable sequences (x_n), that is, $(x_n) \in \ell_1$ if and only if $\sum_{n=1}^\infty |x_n|$ converges. It is not too hard to verify that ℓ_1 is a normed vector space with norm defined as $\|x\|_1 = \sum_{n=1}^\infty |x_n|$. If $\sum_{n=1}^\infty |x_n|$ converges then $(|x_n|)$ converges to zero and thus $(x_n) \in \ell_\infty$, thus $\ell_1 \subset \ell_\infty$.

Example 9.3.11. Fix $y = (y_n)_{n=1}^\infty \in \ell_\infty$ and let $h : \ell_1 \to \ell_1$ be defined as $h(x) = (x_n y_n)_{n=1}^\infty$ for $x = (x_n)_{n=1}^\infty$. Verify that h is well-defined and prove that h is continuous.

Example 9.3.12. Let $\det : \mathbb{R}^{n\times n} \to \mathbb{R}$ denote the determinant function. Prove that det is continuous; you may use the formula

$$\det(A) = \sum_{\sigma\in S_n} \left(\mathrm{sgn}(\sigma) \prod_{i=1}^n a_{i,\sigma(i)} \right)$$

where S_n is the set of permutations on $\{1, 2, \ldots, n\}$ and $\mathrm{sgn}(\sigma) = \pm 1$ is the sign of the permutation $\sigma \in S_n$.

Exercises

Exercise 9.3.1. Let (M, d) be a metric space. Fix $y \in M$ and define the function $f : M \to \mathbb{R}$ by $f(x) = d(x, y)$. Prove that f is continuous.

Exercise 9.3.2. Let $(V, \|\cdot\|)$ be a normed vector space. Prove that $f : V \to \mathbb{R}$ defined by $f(x) = \|x\|$ is continuous. **Hint**: $\|a\| \leq \|a - b\| + \|b\|$ for all $a, b, c \in V$.

Exercise 9.3.3. Consider $C([a, b])$ with norm $\|\cdot\|_\infty$. Define the function $\Psi : C([a, b]) \to \mathbb{R}$ by $\Psi(f) = \int_a^b f(x)\, dx$. Prove that Ψ is continuous in two ways, using the definition and the sequential criterion for continuity.

Exercise 9.3.4. Let M be a metric space and let $f : M \to \mathbb{R}$ be continuous. Prove that $E = \{x \in M \mid f(x) = 0\}$ is closed.

Exercise 9.3.5. Consider $\mathbb{R}^{n\times n}$ as a normed vector space with norm $\|\mathbf{A}\|_\infty = \sup_{1\leq i,j\leq n} |a_{i,j}|$ for $\mathbf{A} \in \mathbb{R}^{n\times n}$.

(a) Let $(\mathbf{A}(k))_{k=1}^\infty$ be a sequence in $\mathbb{R}^{n\times n}$ and denote the (i, j) entry of the matrix $\mathbf{A}(k)$ as $a_{i,j}(k)$. Prove that $(\mathbf{A}(k))_{k=1}^\infty$ converges to $\mathbf{B} \in \mathbb{R}^{n\times n}$ if and only if for all $i, j \in \{1, 2, \ldots, n\}$ the real sequence $(a_{i,j}(k))_{k=1}^\infty$ converges to $b_{i,j} \in \mathbb{R}$.

(b) Given matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n\times n}$, recall that the entries of the product matrix $\mathbf{XY}$ are $(\mathbf{XY})_{i,j} = \sum_{\ell=1}^n x_{i,\ell}y_{\ell,j}$. Let $(\mathbf{A}(k))_{k=1}^\infty$ be a sequence in $\mathbb{R}^{n\times n}$ converging to $\mathbf{B}$ and let $(\mathbf{C}(k))_{k=1}^\infty$ be the sequence whose kth term is $\mathbf{C}(k) = \mathbf{A}(k)\mathbf{A}(k) = [\mathbf{A}(k)]^2$. Prove that $(\mathbf{C}(k))_{k=1}^\infty$ converges to $\mathbf{B}^2$. **Hint**: By part (a), it is enough to prove that the (i, j) component of $\mathbf{C}(k)$ converges to the (i, j) component of $\mathbf{B}^2$.

(c) Deduce that if $\mathbf{C}(k) = [\mathbf{A}(k)]^m$ where $m \in \mathbb{N}$ then the sequence $(\mathbf{C}(k))_{k=1}^\infty$ converges to the matrix $\mathbf{B}^m$.

(d) A **polynomial matrix function** is a function $f : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ of the form

$$f(\mathbf{A}) = c_m\mathbf{A}^m + c_{m-1}\mathbf{A}^{m-1} + \cdots + c_1\mathbf{A} + c_0\mathbf{I}$$

where $c_m, \ldots, c_0$ are constants and $\mathbf{I}$ denotes the $n \times n$ identity matrix. Prove that a polynomial matrix function is continuous.

Exercise 9.3.6. According to the sequential criterion for continuity, if (z_n) and (w_n) are sequences in M converging to the same point $p \in M$ and $f : M \to \mathbb{R}$ is a function such that sequences $(f(z_n))$ and $(f(w_n))$ do not have the same limit $f(p)$ (or worse one of them is divergent!) then f is discontinuous at p. Consider $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x, y) = \begin{cases} \frac{2xy}{x^2+y^2}, & (x, y) \neq (0, 0) \\ 0, & (x, y) = (0, 0). \end{cases}$$

Show that f is discontinuous at $p = (0, 0)$.

9.4 Completeness

Consider the space $\mathcal{P}[a,b]$ of polynomial functions on the interval $[a,b]$. Clearly, $\mathcal{P}[a,b] \subset C([a,b])$, and thus $(\mathcal{P}[a,b], \|\cdot\|_\infty)$ is a metric space. The sequence of functions $f_n(x) = \sum_{k=0}^{n} \frac{1}{k!} x^k$ is a sequence in $\mathcal{P}[a,b]$ and it can be easily verified that (f_n) converges in the metric $\|\cdot\|_\infty$, that is, (f_n) converges uniformly in $[a,b]$ (see Example 8.4.8). However, the limiting function f is not an element of $\mathcal{P}[a,b]$ because it can be verified that $f'(x) = f(x)$ and the only polynomial equal to its derivative is the zero polynomial, however, it is clear that $f(0) = \lim_{n\to\infty} f_n(0) = 1$, i.e., f is not the zero function (you may recognize, of course, that $f(x) = e^x$). We do know, however, that f is in $C([a,b])$ because the uniform limit of a sequence of continuous functions is continuous. The set $\mathcal{P}[a,b]$ then suffers from the same "weakness" as do the rationals $\mathbb{Q}$ relative to $\mathbb{R}$, namely, there are sequences in $\mathcal{P}[a,b]$ that converge to elements not in $\mathcal{P}[a,b]$. On the other hand, because (f_n) converges it is a Cauchy sequence in $C([a,b])$ and thus also in $\mathcal{P}[a,b]$ (the Cauchy condition only depends on the metric) and thus (f_n) is a Cauchy sequence in $\mathcal{P}[a,b]$ that does not converge to an element of $\mathcal{P}[a,b]$. The following discussion motivates the following definition.

Definition 9.4.1: Complete Metric Space

A metric space M is called **complete** if every Cauchy sequence in M converges in M.

This seems like a reasonable starting definition of completeness since in $\mathbb{R}$ it can be proved that the Cauchy criterion (plus the Archimedean property) *implies* the Completeness property of $\mathbb{R}$ (Theorem 3.6.8).

Based on our characterization of closed sets via sequences, we have the following first theorem regarding completeness.

Theorem 9.4.2

Let (M, d) be a complete metric space and let $P \subset M$. Then (P, d) is a complete metric space if and only if P is closed.

Proof. If (z_n) is a Cauchy sequence in (P, d) then it is also a Cauchy sequence in (M, d). Since (M, d) is complete then (z_n) converges. If P is closed then by Theorem 9.2.13 the limit of (z_n) is in P. Hence, (P, d) is a complete metric space.

Now suppose that (P, d) is a complete metric space and let (z_n) be a sequence in P that converges to $z \in M$. Then (z_n) is a Cauchy sequence in M and thus also Cauchy in P. Since P is complete then $z \in P$. Hence, by Theorem 9.2.13 P is closed. □

We now consider how to formulate the Bolzano-Weierstrass (BW) property in a general metric space. The proof in Theorem 3.6.8 can be easily modifield to prove that the BW property, namely that every bounded sequence in $\mathbb{R}$ has a convergent subsequence, implies the completeness property of $\mathbb{R}$. We therefore want to develop a BW-type condition in a general metric space M that implies that M is complete. Our first order of business is to develop the correct notion of boundedness. We have already defined what it means for a subset $E \subset M$ to be bounded, namely, that there exists $r > 0$ such that $E \subset B_r(x)$ for some $x \in M$. However, this notion is not enough as the next example illustrates.

Example 9.4.3. Consider $\mathcal{P}[0, 1]$ with induced metric $\|\cdot\|_\infty$ and let $E = \{f \in \mathcal{P}[0, 1] : \|f\|_\infty < 3\}$, in other words, E is the open ball of radius $r = 3$ centered at the zero function. Clearly, E is bounded and thus any sequence in E is bounded. The sequence $f_n(x) = \sum_{k=0}^{n} \frac{1}{k!} x^k$ is in E, that is, $\|f_n\|_\infty < 3$ for all n (see Example 3.3.6). However, as already discussed, (f_n) converges in $C[0,1]$ but not to a point in $\mathcal{P}[0,1]$. On the other hand, (f_n) is a Cauchy sequence in $\mathcal{P}[0,1]$ and thus (f_n) cannot have a converging subsequence in $\mathcal{P}[0,1]$ by part (iii) of Lemma 9.2.6. Thus, (f_n) is a bounded sequence in $P[0,1]$ with no converging subsequence in $P[0,1]$.

The correct notion of boundedness that is needed is the following.

Definition 9.4.4

Let (M,d) be a metric space. A subset $E \subset M$ is called **totally bounded** if for any given $\varepsilon > 0$ there exists $z_1, \ldots, z_N \in E$ such that $E \subset \bigcup_{i=1}^{N} B_\varepsilon(z_i)$.

Example 9.4.5. Prove that a subset of a totally bounded set is also totally bounded.

Example 9.4.6. A totally bounded subset E of a metric space M is bounded. If $E \subset B_\varepsilon(x_1) \cup \cdots \cup B_\varepsilon(x_k)$ then if $r = \max_{2 \leq j \leq k} d(x_1, x_j) + \varepsilon$ then if $x \in E \cap B_\varepsilon(x_j)$ then $d(x_1, x) \leq d(x_1, x_j) + d(x_j, x) < r$. Hence, $E \subset B_r(x_1)$.

The following shows that the converse in the previous example does not hold.

Example 9.4.7. Consider ℓ_1 and let $E = \{e_1, e_2, e_3, \ldots\}$ where e_k is the infinite sequence with entry k equal to 1 and all other entries zero. Then $\|e_k\|_1 = 1$ for all $k \in \mathbb{N}$ and therefore E is bounded, in particular $E \subset B_r(0)$ for any $r > 1$. Now, $\|e_k - e_j\|_1 = 2$ for all $k \neq j$ and thus if $\varepsilon \leq 2$ then no finite collection of open balls $B_\varepsilon(e_{k_1}), B_\varepsilon(e_{k_2}), \ldots, B_\varepsilon(e_{k_N})$ can cover E. Hence, E is not totally bounded.

Example 9.4.8. Prove that a bounded subset E of $\mathbb{R}$ is totally bounded.

Theorem 9.4.9: Bolzano-Weierstrass

Let M be a metric space. Then M is complete if and only if every infinite totally bounded subset of M has a limit point in M.

Proof. Suppose that (M, d) is a complete metric space. Let E be an infinite totally bounded subset of M. Let $\varepsilon_n = \frac{1}{2^n}$ for $n \in \mathbb{N}$. For ε_1 there exists $z_1, \ldots, z_{m_1} \in E$ such that $E \subset \bigcup_{j=1}^{m_1} B_{\varepsilon_1}(z_j)$. Since E is infinite, we can assume without loss of generality that $E_1 = E \cap B_{\varepsilon_1}(z_1)$ contains infinitely many points of E. Let then $x_1 = z_1$. Now, E_1 is totally bounded and thus there exists $w_1, \ldots, w_{m_2} \in E_1$ such that $E_1 \subset \bigcup_{j=1}^{m_2} B_{\varepsilon_2}(w_j)$. Since E_1 is infinite, we can assume without loss of generality that $E_2 = E_1 \cap B_{\varepsilon_2}(w_1)$ contains infinitely many points of E_1. Let $x_2 = w_1$ and therefore $d(x_2, x_1) < \varepsilon_1$. Since E_2 is totally bounded there exists $u_1, \ldots, u_{m_3} \in E_2$ such that $E_2 \subset \bigcup_{j=1}^{m_3} B_{\varepsilon_3}(u_j)$. We can assume without loss of generality that $E_3 = E_2 \cap B_{\varepsilon_3}(u_1)$ contains infinitely many elements of E_2. Let $x_3 = u_1$ and thus $d(x_3, x_2) < \varepsilon_2$. By induction, there exists a sequence (x_n) in E such that $d(x_{n+1}, x_n) < \frac{1}{2^n}$. Therefore, if $m > n$ then by the triangle inequality (and the geometric series) we have

$$\begin{aligned} d(x_m, x_n) &\leq d(x_m, x_{m-1}) + \cdots + d(x_{n+1}, x_n) \\ &< \frac{1}{2^{m-1}} + \cdots + \frac{1}{2^n} \\ &< \frac{1}{2^{n-1}}. \end{aligned}$$

Therefore, (x_n) is a Cauchy sequence and since M is complete (x_n) converges in M. Thus E has a limit point in M.

Conversely, assume that every infinite totally bounded subset of M has a limit point in M. Let (x_n) be a Cauchy sequence in M and let $E = \{x_n \mid n \in \mathbb{N}\}$. For any given $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that $|x_n - x_K| < \varepsilon$ for all $n \geq K$. Therefore, $x_n \in B_\varepsilon(x_K)$ for all $n \geq K$ and clearly $x_j \in B_\varepsilon(x_j)$ for all $j = 1, 2, \ldots, K_1$. Thus, E is totally bounded. By assumption, E has a limit point, that is, there exists a subsequence of (x_n) that converges in M. By part (iii) of Lemma 9.2.6, (x_n) converges in M. Thus, M is a complete metric space. □

The proof in Theorem 9.4.9 that completeness implies that every infinite totally bounded subset has a limit point is reminiscent of the bisection method proof that a bounded sequence in $\mathbb{R}$ contains a convergent subsequence. Also, the proof showed the following.

Lemma 9.4.10

If E is an infinite totally bounded subset of (M, d) then E contains a Cauchy sequence (x_n) such that $x_n \neq x_m$ for all $n \neq m$.

A complete normed vector space is usually referred to as a **Banach space** in honor of Polish mathematician Stefan Banach (1892-1945) who, in his 1920 doctorate dissertation, laid the foundations of these spaces and their applications in integral equations. An important example of a Banach space is the following. Let X be a non-empty set and let $\mathcal{B}(X)$ be the set of bounded functions from X to $\mathbb{R}$ with sup-norm $\|f\|_\infty = \sup_{x \in X} |f(x)|$. Then convergence in $(\mathcal{B}(X), \|\cdot\|_\infty)$ is uniform convergence (Example 9.2.3). We have all the tools necessary to prove that $\mathcal{B}(X)$ is a Banach space.

Theorem 9.4.11

Let X be a non-empty set. The normed space $(\mathcal{B}(X), \|\cdot\|_\infty)$ is a Banach space.

Proof. First of all, it is clear that $\mathcal{B}(X)$ is a real vector space and thus we need only show it is complete. The proof is essentially contained in the Cauchy criterion for uniform convergence for functions on $\mathbb{R}$ (Theorem 8.2.7). Let $f_n : X \to \mathbb{R}$ be a Cauchy sequence of bounded functions. Then for any given $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that if $n, m \geq K$ then $\|f_n - f_m\|_\infty < \varepsilon$. In particular, for any fixed $x \in X$ it holds that

$$|f_n(x) - f_m(x)| \leq \|f_n - f_m\|_\infty < \varepsilon.$$

Therefore, the sequence of real numbers $(f_n(x))$ is a Cauchy sequence and thus $f(x) = \lim_{n\to\infty} f_n(x)$ exists for each $x \in X$. Now, since (f_n) is a Cauchy sequence in $\mathcal{B}(X)$ then (f_n) is bounded in $\mathcal{B}(X)$. Thus, there exists $M > 0$ such that $\|f_n\|_\infty \leq M$ for all $n \geq 1$. Thus, for all $x \in X$ and $n \geq 1$ it holds that

$$|f_n(x)| \leq \|f_n\|_\infty \leq M$$

and by continuity of the absolute value function it holds that

$$|f(x)| = \lim_{n\to\infty} |f_n(x)| \leq M.$$

Thus, f is a bounded function, that is, $f \in \mathcal{B}(X)$. Now, for any fixed $\varepsilon > 0$, let $K \in \mathbb{N}$ be such that $|f_n(x) - f_m(x)| < \varepsilon/2$ for all $x \in X$ and $n, m \geq K$. Therefore, for any $x \in X$ we have that

$$\begin{aligned}\lim_{m\to\infty} |f_n(x) - f_m(x)| &= |f_n(x) - f(x)| \\ &\leq \varepsilon/2.\end{aligned}$$

Therefore, $\|f_n - f\|_\infty < \varepsilon$ for all $n \geq K$. This proves that (f_n) converges to f in $(\mathcal{B}(X), \|\cdot\|_\infty)$. ☐

Corollary 9.4.12

The space of continuous functions $C([a,b])$ on the interval $[a,b]$ with sup-norm is a Banach space.

Proof. A continuous function on the interval $[a,b]$ is bounded and thus $C([a,b]) \subset \mathcal{B}([a,b])$. Convergence in $\mathcal{B}([a,b])$ with sup-norm is uniform convergence. A sequence of continuous functions that converges uniformly on $[a,b]$ does so to a continuous function. Hence, Theorem 9.2.13 implies that $C([a,b])$ is a closed subset of the complete metric space $\mathcal{B}([a,b])$ and then Theorem 9.4.2 finishes the proof. ☐

Example 9.4.13. Prove that ℓ_∞ and ℓ_1 are complete and hence Banach spaces.

In a Banach space, convergence of series can be decided entirely from the convergence of real series.

Theorem 9.4.14: Absolute Convergence Test

Let $(V, \|\cdot\|)$ be a Banach space and let (z_n) be a sequence in V. If the real series $\sum_{n=1}^\infty \|z_n\|$ converges then the series $\sum_{n=1}^\infty z_n$ converges in V.

Proof. Suppose that $\sum_{n=1}^\infty \|z_n\|$ converges, that is, suppose that the sequence of partial sums $t_n = \sum_{k=1}^n \|z_n\|$ converges (note that (t_n) is increasing). Then (t_n) is a Cauchy sequence. Consider the sequence of partial sums $s_n = \sum_{k=1}^{n} z_k$. For $n > m$ we have

$$\begin{aligned} \|s_n - s_m\| &= \left\| \sum_{k=m+1}^{n} z_n \right\| \\ &\le \sum_{k=m+1}^{n} \|z_n\| \\ &= t_n - t_m \\ &= |t_n - t_m| \end{aligned}$$

and since (t_n) is Cauchy then $|t_n - t_m|$ can be made arbitrarily small provided n, m are sufficiently large. This proves that (s_n) is a Cauchy sequence in V and therefore converges since V is complete. $\square$

Remark 9.4.15. We make two remarks. The converse of Theorem 9.4.14 is also true, that is, if every series $\sum z_n$ in $(V, \|\cdot\|)$ converges whenever $\sum \|z_n\|$ converges in $\mathbb{R}$ then V is a Banach space. Notice that the proof of Theorem 9.4.14 is essentially the same as the proof of the Weierstrass M-test.

Example 9.4.16. Consider the set of matrices $\mathbb{R}^{n\times n}$ equipped with the 2-norm

$$\|A\|_2 = \left(\sum_{i,j=1}^{n} a_{i,j}^2 \right)^{1/2}$$

The norm $\|A\|_2$ is called the **Frobenius** norm or the **Hilbert-Schmidt** norm.

(a) Prove that $(\mathbb{R}^{n\times n}, \|\cdot\|_2)$ is complete.

(b) Use the Cauchy-Schwarz inequality

$$\left(\sum_{\ell=1}^{N} x_\ell y_\ell \right)^2 \le \left(\sum_{\ell=1}^{N} x_\ell^2 \right) \left(\sum_{\ell=1}^{N} y_\ell^2 \right)$$

to prove that $\|AB\|_2 \leq \|A\|_2 \|B\|_2$. Conclude that $\left\|A^k\right\|_2 \leq (\|A\|_2)^k$ for all $k \in \mathbb{N}$.

(c) Let $f(x) = \sum_{k=1}^{\infty} c_k x^k$ be a power series converging on $\mathbb{R}$. Define the function $f : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ as

$$f(A) = \sum_{k=1}^{\infty} c_k A^k.$$

Prove that f is well-defined and that if $c_k \geq 0$ then $\|f(A)\|_2 \leq f(\|A\|_2)$, that is, that

$$\left\| \sum_{k=1}^{\infty} c_k A^k \right\|_2 \leq \sum_{k=1}^{\infty} c_k \|A\|_2^k$$

Proof. (a) The norm $\|A\|_2$ is simply the standard Euclidean norm on $(\mathbb{R}^N, \|\cdot\|_2)$ with $N = n^2$ and identifying matrices as elements of $\mathbb{R}^N$. Hence, $(\mathbb{R}^{n\times n}, \|\cdot\|_2)$ is complete.

(b) From the Cauchy-Schwarz inequality we have

$$\begin{aligned} (AB)_{i,j}^2 &= \left(\sum_{\ell=1}^{n} a_{i,\ell} b_{\ell,j} \right)^2 \\ &\leq \left(\sum_{\ell=1}^{n} a_{i,\ell}^2 \right) \left(\sum_{\ell=1}^{n} b_{\ell,j}^2 \right) \end{aligned}$$

and therefore

$$\begin{aligned}
\|AB\|_2 &= \left(\sum_{1\le i,j\le n} (AB)_{i,j}^2 \right)^{1/2} \\
&\le \left(\sum_{1\le i,j\le n} \left(\sum_{\ell=1}^n a_{i,\ell}^2 \right) \left(\sum_{\ell=1}^n b_{\ell,j}^2 \right) \right)^{1/2} \\
&= \left(\sum_{i,\ell=1}^n a_{i,\ell}^2 \right)^{1/2} \left(\sum_{j,\ell=1}^n b_{\ell,j}^2 \right)^{1/2} \\
&= \|A\|_2 \, \|B\|_2
\end{aligned}$$

(c) We first note that for any power series $\sum_{k=1}^{\infty} a_k x^k$ that converges in $(-R, R)$, the power series $\sum_{k=1}^{\infty} |a_k| x^k$ also converges in $(-R, R)$. The normed space $(\mathbb{R}^{n\times n}, \|\cdot\|_2)$ is complete and thus $\sum_{k=1}^{\infty} c_k A^k$ converges whenever $\sum_{k=1}^{\infty} \left\|c_k A^k\right\|_2$ converges. Now by part (b), $\left\|c_k A^k\right\|_2 = |c_k| \left\|A^k\right\|_2 \le |c_k| \, \|A\|_2^k$ and since $\sum_{k=1}^{\infty} |c_k| \, \|A\|_2^k$ converges then by the comparison test for series in $\mathbb{R}$, the series $\sum_{k=1}^{\infty} \left\|c_k A^k\right\|_2$ converges. Therefore, $f(A)$ is well-defined by Theorem 9.4.14. To prove the last inequality, we note that the norm function on a vector space is continuous and thus if $c_k \ge 0$ then

$$\begin{aligned}
\left\| \sum_{k=1}^m c_k A^k \right\|_2 &\le \sum_{k=1}^m |c_k| \, \|A\|_2^k \\
&\le \sum_{k=1}^m c_k \, \|A\|_2^k
\end{aligned}$$

and therefore

$$\begin{aligned}\left\|\sum_{k=1}^{\infty} c_k A^k\right\|_2 &= \lim_{m\to\infty}\left\|\sum_{k=1}^{m} c_k A^k\right\|_2\\ &\le \lim_{m\to\infty}\sum_{k=1}^{m} c_k \|A\|_2^k\\ &= \sum_{k=1}^{\infty} c_k \|A\|_2^k,\end{aligned}$$

in other words, $\|f(A)\|_2 \le f(\|A\|_2)$.

□

Example 9.4.17. In view of the previous example, we can define for $A \in \mathbb{R}^{n\times n}$ the following:

$$\begin{aligned}e^A &= \sum_{k=0}^{\infty}\frac{1}{k!}A^k\\ \sin(A) &= \sum_{k=0}^{\infty}\frac{(-1)^n}{(2n+1)!}A^{2n+1}\\ \cos(A) &= \sum_{k=0}^{\infty}\frac{(-1)^n}{(2n)!}A^{2n}\\ \arctan(A) &= \sum_{k=0}^{\infty}\frac{(-1)^n}{(2n+1)}A^{2n+1}\end{aligned}$$

and for instance $\left\|e^A\right\|_2 \le e^{\|A\|_2}$, etc.

Exercises

Exercise 9.4.1. Let (M_1, d_1) and (M_2, d_2) be metric spaces. There are several ways to define a metric on the Cartesian product $M_1 \times M_2$. One way is to imitate what was done in $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. We can define $d : M_1 \times M_2 \to [0, \infty)$ as

$$d((x,u),(y,v)) = \sqrt{d_1(x,y)^2 + d_2(u,v)^2}$$

(a) Prove that d is a metric on $M_1 \times M_2$.

(b) Prove that $((x_n, u_n))_{n=1}^{\infty}$ converges in $M_1 \times M_2$ if and only if $(x_n)_{n=1}^{\infty}$ and $(y_n)_{n=1}^{\infty}$ converge in M_1 and M_2, respectively. (Hint: Theorem 9.2.7)

(c) Prove that $M_1 \times M_2$ is complete if and only if M_1 and M_2 are complete. (Hint: Corollary 9.2.8)

Exercise 9.4.2. Let M be a complete metric space and let (z_n) be a sequence in M such that $d(z_n, z_{n+1}) < r^n$ for all $n \in \mathbb{N}$ for some fixed $0 < r < 1$. Prove that (z_n) converges. (See Exercise 3.6.6.)

Exercise 9.4.3. Let $\sum_{n=1}^{\infty} x_n$ be a convergent series in a normed vector space $(V, \|\cdot\|)$ and suppose that $\sum_{n=1}^{\infty} \|x_n\|$ converges. Show that

$$\left\| \sum_{n=1}^{\infty} x_n \right\| \leq \sum_{n=1}^{\infty} \|x_n\|.$$

Note: The $\triangle$-inequality can only be used on a **finite** sum. (See Exercise 9.3.2.)

Exercise 9.4.4. Consider the normed space $(\mathcal{B}(X), \|\cdot\|_{\infty})$ where X is a non-empty set. Let $\mathcal{K}(X)$ be the set of constant functions on X. Prove that $(\mathcal{K}(X), \|\cdot\|_{\infty})$ is a Banach space. (Hint: Theorem 9.4.2)

9.5 Compactness

Important results about continuous functions, such as the Extreme Value Theorem (Theorem 5.3.7) and uniform continuity (Theorem 5.4.7), depended heavily on the domain being a closed and bounded interval. On a bounded interval, any sequence (x_n) contains a Cauchy subsequence (x_{n_k}) (use the bisection algorithm), and if the interval is also closed then we are guaranteed that the limit of (x_{n_k}) is contained in the interval. We have already seen that a totally bounded subset E of a metric space M contains a Cauchy sequence (Lemma 9.4.10) and thus if E is complete then Cauchy sequences converge in E. This motivates the following definition.

Definition 9.5.1: Compactness

A metric space M is called **compact** if M is totally bounded and complete.

A closed and bounded subset E of $\mathbb{R}$ is compact. Indeed, E is complete because it is closed (Theorem 9.4.2) and it is easy to see how to cover E with a finite number of open intervals of any given radius $\varepsilon > 0$. Conversely, if $E \subset \mathbb{R}$ is compact then E is bounded and E is closed by Theorem 9.4.2. A similar argument shows that $E \subset \mathbb{R}^n$ is compact if and only if E is closed and bounded. This is called the Heine-Borel theorem.

Theorem 9.5.2: Heine-Borel

A subset $E \subset \mathbb{R}^n$ is compact if and only if E is closed and bounded.

Example 9.5.3. The unit n-sphere $\mathbb{S}^n$ in $\mathbb{R}^{n+1}$ is the set

$$\mathbb{S}^n = \{x = (x_1, x_2, \ldots, x_n, x_{n+1}) \in \mathbb{R}^n \mid x_1^2 + x_2^2 + \cdots + x_{n+1}^2 = 1\}.$$

Explain why $\mathbb{S}^n$ is compact subset of $\mathbb{R}^{n+1}$.

Example 9.5.4. Consider $\mathbb{R}^{n\times n}$ with norm $\|A\|_2 = \left(\sum_{i,j} a_{i,j}^2\right)^{1/2}$. Then $\mathbb{R}^{n\times n}$ is complete. A matrix $Q \in \mathbb{R}^{n\times n}$ is called *orthogonal* if $Q^T Q = I$, where Q^T denotes the transpose of Q and I is the identity matrix. Prove that the set of orthogonal matrices, which we denote by $O(n)$, is compact.

A useful characterization of compactness is stated in the language of sequences.

Theorem 9.5.5: Sequential Criterion for Compactness

A metric space M is compact if and only if every sequence in M has a convergent subsequence.

Proof. Assume that M is compact. If (x_n) is a sequence in M then $\{x_n \mid n \in \mathbb{N}\}$ is totally bounded and thus has a Cauchy subsequence which converges by completeness of M.

Conversely, assume that every sequence in M has a convergent subsequence. If (x_n) is a Cauchy sequence then by assumption it has a convergent subsequence and thus (x_n) converges. This proves M is complete. Suppose that M is not totally bounded. Then there exists $\varepsilon > 0$ such that M cannot be covered by a finite number of open balls of radius $\varepsilon > 0$. Hence, there exists $x_1, x_2 \in M$ such that $d(x_1, x_2) \geq \varepsilon$. By induction, suppose $x_1, \ldots, x_k$ are such that $d(x_i, x_j) \geq \varepsilon$ for $i \neq j$. Then there exists x_{k+1} such that $d(x_i, x_{k+1}) \geq \varepsilon$ for all $i = 1, \ldots, k$. By induction then, there exists a sequence (x_n) such that $d(x_i, x_j) \geq \varepsilon$

if $i \neq j$. Clearly, (x_n) is not a Cauchy sequence and therefore cannot have a convergent subsequence. □

Example 9.5.6. Let M be a metric space.

(a) Prove that if $E \subset M$ is finite then E is compact.

(b) Is the same true if E is countable?

(c) What if M is compact?

We now describe how compact sets behave under continuous functions.

Theorem 9.5.7

Let $f : M_1 \to M_2$ be a continuous mapping. If $E \subset M_1$ is compact then $f(E) \subset M_2$ is compact.

Proof. We use the sequential criterion for compactness. Let $y_n = f(x_n)$ be a sequence in $f(E)$. Since E is compact, by Theorem 9.5.5, there is a convergent subsequence (x_{n_k}) of (x_n). By continuity of f, the subsequence $y_{n_k} = f(x_{n_k})$ of (y_n) is convergent. Hence, $f(E)$ is compact. □

We now prove a generalization of the Extreme value theorem 5.3.7.

Theorem 9.5.8: Extreme Value Theorem

Let (M, d) be a compact metric space. If $f : M \to \mathbb{R}$ is continuous then f achieves a maximum and a minimum on M, that is, there exists $x^*, x_* \in M$ such that $f(x_*) \leq f(x) \leq f(x^*)$ for all $x \in M$. In particular, f is bounded.

Proof. By Theorem 9.5.7, $f(M)$ is a compact subset of $\mathbb{R}$ and thus $f(M)$ is closed and bounded. Let $y_* = \inf f(M)$ and let $y^* = \sup f(M)$. By the properties of the supremum, there exists a sequence (y_n) in $f(M)$ converging to y^*. Since $f(M)$ is closed, then $y^* \in f(M)$ and thus $y^* = f(x^*)$ for some $x^* \in M$. A similar argument shows that $y_* = f(x_*)$ for some $x_* \in M$. Hence, $\inf f(M) = f(x_*) \leq f(x) \leq f(x^*) = \sup f(M)$ for all $x \in M$. □

Let M be a metric space and let $E \subset M$. A **cover** of E is a collection $\{U_i\}_{i\in I}$ of subsets of M whose union contains E. The index set I may be countable or uncountable. The cover $\{U_i\}_{i\in I}$ is called an **open cover** if each set U_i is open. A **subcover** of a cover $\{U_i\}$ of E is a cover $\{U_j\}_{j\in J}$ of E such that $J \subset I$.

Theorem 9.5.9: Compactness via Open Covers

A metric space M is compact if and only if every open cover of M has a finite subcover.

Proof. Assume that M is compact. Then by Theorem 9.5.5, every sequence in M has a convergent subsequence. Let $\{U_i\}$ be an open cover of M. We claim there exists $\varepsilon > 0$ such that for each $x \in M$ it holds that $B_\varepsilon(x) \subset U_i$ for some U_i. If not, then then for each $n \in \mathbb{N}$ there exists $x_n \in M$ such that $B_{1/n}(x_n)$ is not properly contained in a single set U_i. By assumption, the sequence (x_n) has a converging subsequence, say it is (x_{n_k}) and $y = \lim x_{n_k}$. Hence, for each $k \in \mathbb{N}$, $B_{1/n_k}(x_{n_k})$ is not properly contained in a single U_i. Now, $y \in U_j$ for some j, and thus since U_j is open there exists $\delta > 0$ such that $B_\delta(y) \subset U_j$. Since $(x_{n_k}) \to y$, there exists K sufficiently large such that $d(x_{n_K}, y) < \delta/2$ and $\frac{1}{n_K} < \delta/2$. Then $B_{1/n_K}(x_{n_K}) \subset U_j$ which is a contradiction. This proves that such an $\varepsilon > 0$ exists. Now since M is totally bounded, there exists $z_1, z_2, \ldots, z_p \in M$ such that $B_\varepsilon(z_1) \cup \cdots \cup B_\varepsilon(z_p) = M$, and since $B_\varepsilon(z_j) \subset U_{i_j}$ for some U_{i_j} it follows that $\{U_{i_1}, U_{i_2}, \ldots, U_{i_p}\}$ is a finite subcover of $\{U_i\}$.

For the converse, we prove the contrapositive. Suppose then that M is not compact. Then by Theorem 9.5.5, there is a sequence (x_n) in M with no convergent subsequence. In particular, there is a subsequence (y_k) of (x_n) such that all y_k's are distinct and (y_k) has no convergent subsequence. Then there exists $\varepsilon_i > 0$ such that $B_{\varepsilon_i}(y_i)$ contains only the point y_i from the sequence (y_k), otherwise we can construct a subsequence of (y_k) that converges. Hence, $\{B_{\varepsilon_k}(y_k)\}_{k\in\mathbb{N}}$ is an open cover of the set $E = \{y_1, y_2, y_3, \ldots, \}$ that has no finite subcover. The set E is clearly closed since it consists entirely of isolated points of M. Hence, $\{B_{\varepsilon_k}(y_k)\}_{k\in\mathbb{N}} \cup M\backslash E$ is an open cover of M with no finite subcover. □

Definition 9.5.10: Uniform Continuity

A function $f : (M_1, d_1) \to (M_2, d_2)$ is **uniformly continuous** if for any $\varepsilon > 0$ there exists $\delta > 0$ such that if $d_1(x, y) < \delta$ then $d_2(f(x), f(y)) < \varepsilon$.

Example 9.5.11. A function $f : (M_1, d_1) \to (M_2, d_2)$ is **Lipschitz** if there is a constant $K > 0$ such that $d_2(f(x), f(y)) \leq Kd(x, y)$. Show that a Lipschitz map is uniformly continuous.

Example 9.5.12. If $f : M_1 \to M_2$ is uniformly continuous and (x_n) is a Cauchy sequence in M_1, prove that $(f(x_n))$ is a Cauchy sequence in M_2.

Theorem 9.5.13

If $f : (M_1, d_1) \to (M_2, d_2)$ is continuous and M_1 is compact then f is uniformly continuous.

Proof. Let $\varepsilon > 0$. For each $x \in M_1$, there exists $r_x > 0$ such that if $y \in B_{r_x}(x)$ then $f(y) \in B_{\varepsilon/2}(f(x))$. Now $\{B_{r_x/2}(x)\}_{x\in M_1}$ is an open cover of M_1 and since M_1 is compact there exists finite $x_1, x_2, \ldots, x_N$ such that $\{B_{\delta_i}(x_i)\}_{i=1}^N$ is an open cover of M_1, where we have set $\delta_i = r_{x_i}/2$. Let $\delta = \min\{\delta_1, \ldots, \delta_N\}$. If $d_1(x, y) < \delta$, and say $x \in B_{\delta_i}(x_i)$, then $d_1(y, x_i) \leq d_1(y, x) + d_1(x, x_i) < \delta + \delta_i < r_{x_i}$ and thus

$$\begin{aligned} d_2(f(x), f(y)) &\leq d_2(f(x), f(x_i)) + d_2(f(x_i), f(y)) \\ &< \varepsilon/2 + \varepsilon/2 \\ &= \varepsilon. \end{aligned}$$

This proves that f is uniformly continuous. ☐

Exercises

Exercise 9.5.1. Prove that if $E \subset \mathbb{R}$ is compact then $\sup(E)$ and $\inf(E)$ are elements of E.

Exercise 9.5.2. Recall that ℓ_∞ is the set of sequences in $\mathbb{R}$ that are bounded and equipped with the norm $\|(x_n)\|_\infty = \sup_{n\in\mathbb{N}} |x_n|$. Show that the unit ball $B = \{(x_n) : \|(x_n)\|_\infty \leq 1\}$ (which is clearly bounded) is not compact in ℓ_∞. (see Example 9.4.7)

Exercise 9.5.3. Let (x_n) be a sequence in a metric space M and suppose that (x_n) converges to p. Prove that $S = \{p\} \cup \{x_n \,|\, n \in \mathbb{N}\}$ is a compact subset of M.

Exercise 9.5.4. Prove that if M is compact then there exists a countable subset $E \subset M$ that is dense in M.

Exercise 9.5.5. Let E be a compact subset of M and fix $p \in M$. Prove that there exists $z \in E$ such that $d(z, p) \leq d(x, p)$ for all $x \in E$.

9.6 Fourier Series

Motivated by problems involving the conduction of heat in solids and the motion of waves, a major problem that spurred the development of modern analysis (and mathematics in general) was whether an arbitrary function f can be represented by a series of the form

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos(nx) + b_n \sin(nx))$$

for appropriately chosen coefficients a_n and b_n. A central character in this development was mathematician and physicist Joseph Fourier and for this reason such a series is now known as a **Fourier series**.

Fourier made the bold claim (*Theorie analytique de la Chaleur, 1822*) that "there is no function $f(x)$ or part of a function, which cannot be expressed by a trigonometric series". Fourier's claim led B. Riemann (1854) to develop what we now call the Riemann integral. After Riemann, Cantor's (1872) interest in trigonometric series led him to the investigation of the *derived set* of a set S (which nowadays we call the *limit points* of S) and he subsequently developed set theory. The general problem of convergence of a Fourier series led to the realization that by allowing "arbitrary" functions into the picture the theory of integration developed by Riemann would have to be extended to widened the class of "integrable functions". This extension of the Riemann integral was done by Henri Lebesgue (1902) and spurred the development of the theory of measure and integration. The Lebesgue integral is widely accepted as the "official" integral of modern analysis.

In this section, our aim is to present a brief account of Fourier series with the tools that we have already developed. To begin, suppose that $f : [-\pi, \pi] \to \mathbb{R}$ is Riemann integrable and can be represented by a Fourier series, that is,

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos(nx) + b_n \sin(nx)) \tag{9.1}$$

for $x \in [-\pi, \pi]$. In other words, the series on the right of (9.1) converges pointwise to f on $[-\pi, \pi]$. The first question we need to answer is what are the coefficients a_n and b_n in terms of f? To that end, we use the following facts. Let $n, m \in \mathbb{N}$:

(i) For all n:

$$\int_{-\pi}^{\pi} \sin(nx)dx = \int_{-\pi}^{\pi} \cos(nx)dx = 0.$$

(ii) If $n \neq m$ then

$$\int_{-\pi}^{\pi} \sin(nx)\sin(mx)dx = \int_{-\pi}^{\pi} \cos(nx)\cos(mx)dx = 0.$$

(iii) For all n and m:

$$\int_{-\pi}^{\pi} \sin(nx)\cos(mx)dx = 0.$$

(iv) For all n and m:

$$\int_{-\pi}^{\pi} \sin^2(nx)dx = \int_{-\pi}^{\pi} \cos^2(nx)dx = \pi.$$

Then, using these facts, and momentarily ignoring the interchange of the integral and infinite sum,

$$\begin{aligned}
\int_{-\pi}^{\pi} f(x)\cos(nx)dx &= \int_{-\pi}^{\pi} \Big[\frac{a_0}{2}\cos(nx) \\
&\qquad + \sum_{k=1}^{\infty}(a_k\cos(kx)\cos(nx) + b_k\sin(kx)\cos(nx))\Big]dx \\
&= \frac{a_0}{2}\int_{-\pi}^{\pi}\cos(nx)dx + \sum_{k=1}^{\infty} a_k \int_{-\pi}^{\pi}\cos(kx)\cos(nx)dx \\
&\qquad + \sum_{k=1}^{\infty} b_k \int_{-\pi}^{\pi}\sin(kx)\cos(nx)dx \\
&= a_n\pi.
\end{aligned}$$

Therefore,

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)dx.$$

A similar calculation shows that

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)dx.$$

Finally,

$$\begin{aligned}\int_{-\pi}^{\pi} f(x)dx &= \frac{a_0}{2}\int_{-\pi}^{\pi} dx + \sum_{k=1}^{\infty} a_k \int_{-\pi}^{\pi} \cos(kx)dx + \sum_{k=1}^{\infty} b_k \int_{-\pi}^{\pi} \sin(kx)dx \\ &= \frac{a_0}{2}2\pi \\ &= a_0\pi\end{aligned}$$

and therefore

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)dx.$$

Of course, the above calculations are valid provided that the Fourier series converges uniformly to f on $[-\pi, \pi]$ since if all we have is pointwise convergence then in general we cannot interchange the integral sign and the infinite sum. Since the functions $f_n(x) = a_n \cos(nx) + b_n \sin(nx)$ in the Fourier series are clearly continuous, and if we insist that the convergence is uniform, then we have restricted our investigation of Fourier series to continuous functions! Relaxing this restriction led to the development of what we now call the Lebesgue integral; Lebesgue was interested in extending the notion of integration beyond Riemann's development so that a wider class of functions could be integrated and, more importantly, this new integral would be more robust when it came to exchanging limits with integration, i.e., uniform convergence would not be needed. A full development of Lebesgue's theory of integration is beyond the scope of this book, however, we can still say some interesting things about Fourier series.

Motivated by our calculations above, suppose that $f \in C[-\pi, \pi]$

and define

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)dx$$
$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)dx$$
$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)dx.$$

Assume that the Fourier series of f converges uniformly on $C[-\pi,\pi]$ and let

$$g(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n\cos(nx) + b_n\sin(nx)).$$

Then g is continuous on $[-\pi,\pi]$. Does $f = g$? To answer this question, our computations above show that

$$\frac{1}{\pi}\int_{-\pi}^{\pi} g(x)\cos(nx)dx = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)dx$$

and therefore

$$\int_{-\pi}^{\pi}[f(x) - g(x)]\cos(nx)dx = 0$$

for all $n \in \mathbb{N}\cup\{0\}$. Similarly, for all $n \in \mathbb{N}\cup\{0\}$ we have

$$\int_{-\pi}^{\pi}[f(x) - g(x)]\sin(nx)dx = 0.$$

Let $s_n(x) = \frac{a_0}{2} + \sum_{k=1}^{n} a_k\cos(kx) + b_k\sin(kx)$ and recall that (s_n) converges uniformly to g. Consider for the moment

$$\int_{-\pi}^{\pi}[f(x) - s_n(x)]^2dx = \int_{-\pi}^{\pi} f(x)^2dx - 2\int_{-\pi}^{\pi} f(x)s_n(x)dx + \int_{-\pi}^{\pi} s_n^2(x)dx.$$

A straightforward computation shows that

$$\int_{-\pi}^{\pi} f(x)s_n(x)dx = \pi\left[\frac{a_0^2}{2} + \sum_{k=1}^{n}(a_k^2 + b_k^2)\right]$$

and

$$\int_{-\pi}^{\pi} s_n^2(x)dx = \pi\left[\frac{a_0^2}{2} + \sum_{k=1}^{n}(a_k^2 + b_k^2)\right].$$

Therefore,

$$\frac{1}{\pi}\int_{-\pi}^{\pi}[f(x) - s_n(x)]^2dx = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)^2dx - \frac{1}{\pi}\int_{-\pi}^{\pi} s_n^2(x)dx$$

Now since $\int_{-\pi}^{\pi}[f(x) - s_n(x)]^2dx \geq 0$ it follows that

$$\frac{1}{\pi}\int_{-\pi}^{\pi} s_n^2(x)dx \leq \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)^2dx,$$

or equivalently that

$$\frac{a_0^2}{2} + \sum_{k=1}^{n}(a_k^2 + b_k^2) \leq \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)^2dx.$$

This proves that the series $\frac{a_0^2}{2} + \sum_{k=1}^{\infty}(a_k^2 + b_k^2)$ converges and

$$\frac{a_0^2}{2} + \sum_{k=1}^{\infty}(a_k^2 + b_k^2) \leq \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)^2dx.$$

In particular, $\lim_{k\to\infty}(a_k^2+b_k^2) = 0$ and thus $\lim_{k\to\infty} a_k = \lim_{k\to\infty} b_k = 0$.

10

Multivariable Differential Calculus

In this chapter, we consider the differential calculus of mappings from one Euclidean space to another, that is, mappings $F : \mathbb{R}^n \to \mathbb{R}^m$. In first-year calculus, you considered the case $n = 2$ or $n = 3$ and $m = 1$. Examples of functions that you might have encountered were of the type $F(x_1, x_2) = x_1^2 - x_2^2$, $F(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$, or maybe even $F(x_1, x_2) = \sin(x_1)\sin(x_2)$, etc. If now $F : \mathbb{R}^n \to \mathbb{R}^m$ with $m \geq 2$ then F has m **component** functions since $F(x) \in \mathbb{R}^m$. We can therefore write

$$F(x) = (f_1(x), f_2(x), \ldots, f_m(x))$$

and $f_j : \mathbb{R}^n \to \mathbb{R}$ is called the jth component of F.

In this chapter, unless stated otherwise, we equip $\mathbb{R}^n$ with the Euclidean 2-norm $\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$. For this reason, we will omit the subscript in $\|x\|_2$ and simply write $\|x\|$.

10.1 Differentiation

Let $U \subset \mathbb{R}^n$ and let $F : U \to \mathbb{R}^n$ be a function. How should we define differentiability of F at some point $a \in U$? Recall that for a function

$f : I \to \mathbb{R}$, where $I \subset \mathbb{R}$, we say that f is differentiable at $a \in I$ if

$$\lim_{x\to a} \frac{f(x) - f(a)}{x - a}$$

exists. In this case, we denote $f'(a) = \lim_{x\to a} \frac{f(x)-f(a)}{x-a}$ and we call $f'(a)$ the derivative of f at a. As it is written, the above definition does not make sense for F since division of vectors is not well-defined (or at least we have not defined it). An equivalent definition of differentiability of f at a is that there exists a number $m \in \mathbb{R}$ such that

$$\lim_{x\to a} \frac{f(x) - f(a) - m(x - a)}{x - a} = 0$$

which is equivalent to asking that

$$\lim_{x\to a} \frac{|f(x) - f(a) - m(x - a)|}{|x - a|} = 0.$$

The number m is then denoted by $m = f'(a)$ as before. Another way to think about the derivative m is that the affine function $g(x) = f(a)+mx$ is a good approximation to $f(x)$ for points x near a. The linear part of the affine function g is $\ell(x) = mx$. Thought of in this way, the derivative of f at a is a linear function.

Definition 10.1.1: The Derivative

Let U be a subset of $\mathbb{R}^n$. A mapping $F : U \to \mathbb{R}^m$ is said to be **differentiable** at $a \in U$ if there exists a linear mapping $L : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{x\to a} \frac{\|F(x) - F(a) - L(x - a)\|}{\|x - a\|} = 0.$$

In the definition of differentiability, the expression $L(x-a)$ denotes the linear mapping L applied to the vector $(x - a) \in \mathbb{R}^n$. An equivalent definition of differentiability is that

$$\lim_{h\to 0} \frac{\|F(x+h) - F(a) - L(h)\|}{\|h\|} = 0$$

where again $L(h)$ denotes $h \in \mathbb{R}^n$ evaluated at L. It is not hard to show that the linear mapping L in the above definition is unique when $U \subset \mathbb{R}^m$ is an open set. For this reason, we will deal almost exclusively with the case that U is open without further mention. We therefore call L the **derivative of F at a** and denote it instead by $L = DF(a)$. Hence, by definition, the derivative of F at a is the unique linear mapping $DF(a) : \mathbb{R}^n \to \mathbb{R}^m$ satisfying

$$\lim_{x\to a} \frac{\|F(x) - F(a) - DF(a)(x-a)\|}{\|x-a\|} = 0.$$

Applying the definition of the limit, given arbitrary $\varepsilon > 0$ there exists $\delta > 0$ such that if $\|x - a\| < \delta$ then

$$\frac{\|F(x) - F(a) - DF(a)(x-a)\|}{\|x-a\|} < \varepsilon$$

or equivalently

$$\|F(x) - F(a) - DF(a)(x-a)\| < \varepsilon \|x-a\|.$$

If $F : U \to \mathbb{R}^m$ is differentiable at each $x \in U$ then $x \mapsto DF(x)$ is a mapping from U to the space of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$. In other words, if we denote by $\mathcal{L}(\mathbb{R}^n; \mathbb{R}^m)$ the space of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ then we have a well-defined mapping $DF : U \to \mathcal{L}(\mathbb{R}^n; \mathbb{R}^m)$ called the **derivative** of F on U which assigns the derivative of F at each $x \in U$.

We now relate the derivative of F with the derivatives of its component functions. To that end, we need to recall some basic facts from linear algebra and the definition of the partial derivative. For the latter, recall that a function $f : U \subset \mathbb{R}^n \to \mathbb{R}$, has partial derivative at $a \in U$ with respect to x_i, if the following limit exists

$$\lim_{t \to 0} \frac{f(a_1, \ldots, a_{i-1}, a_i + t, a_{i+1}, \ldots, a_n) - f(a)}{t}$$

or equivalently, if there exists a number $m_i \in \mathbb{R}$ such that

$$0 = \lim_{t \to 0} \frac{f(a + e_i t) - f(a) - m_i t}{t}$$

where $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ denotes the ith standard basis vector in $\mathbb{R}^n$. We then denote $m_i = \frac{\partial f}{\partial x_i}(a)$. Now, given any linear map $L : \mathbb{R}^n \to \mathbb{R}^m$, the action of L on vectors in $\mathbb{R}^n$ can be represented as matrix-vector multiplication once we choose a basis for $\mathbb{R}^n$ and $\mathbb{R}^m$. Specifically, if we choose the most convenient bases in $\mathbb{R}^n$ and $\mathbb{R}^m$, namely the standard bases, then

$$L(x) = Ax$$

where $A \in \mathbb{R}^{m \times n}$ and the the (j, i) entry of the matrix A is the jth component of the vector $Ae_i \in \mathbb{R}^m$. We can now prove the following.

Theorem 10.1.2: Jacobian Matrix

Let $U \subset \mathbb{R}^n$ be open and suppose that $F : U \to \mathbb{R}^m$ is differentiable at $a \in U$, and write $F = (f_1, f_2, \ldots, f_m)$. Then the partial derivatives $\frac{\partial f_j}{\partial x_i}(a)$ exist, and the matrix representation of $DF(a)$ in the standard bases in $\mathbb{R}^n$ and $\mathbb{R}^m$ is

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

where all partial derivatives are evaluated at a. The matrix above is called the **Jacobian matrix of F at a.**

Proof. Let $m_{j,i}$ denote the (j, i) entry of the matrix representation of $DF(a)$ in the standard bases in $\mathbb{R}^n$ and $\mathbb{R}^m$, that is, $m_{j,i}$ is the jth component of $DF(a)e_i$. By definition of differentiability, it holds that

$$0 = \lim_{x \to a} \frac{\|F(x) - F(a) - DF(a)(x - a)\|}{\|x - a\|}.$$

Let $x = a + te_i$ where $e_i \in \mathbb{R}^n$ is the ith standard basis vector. Since U is open, $x \in U$ provided t is sufficiently small. Then since $\|x - a\| = \|te_i\| = |t| \to 0$ iff $\|x - a\| \to 0$ we have

$$\begin{aligned} 0 &= \lim_{t \to 0} \frac{\|F(a + te_i) - F(a) - DF(a)e_i t\|}{|t|} \\ &= \lim_{t \to 0} \left\| \frac{1}{t} \big[F(a + te_i) - F(a) - DF(a)e_i t \big] \right\|. \end{aligned}$$

It follows that each component of the vector $\frac{1}{t}\big[F(a + te_i) - F(a) - DF(a)e_i t\big]$ tends to 0 as $t \to 0$. Hence, for each $j \in \{1, 2, \ldots, m\}$ we have

$$0 = \lim_{t \to 0} \frac{1}{t}(f_j(a + te_i) - f_j(a) - m_{j,i}t).$$

Hence, $\frac{\partial f_j}{\partial x_i}(a)$ exists and $m_{j,i} = \frac{\partial f_j}{\partial x_i}(a)$ as claimed. □

It is customary to write

$$DF(a) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

since for any $x \in \mathbb{R}^n$ the vector $DF(a)x$ is the Jacobian matrix of F at a multiplied by x (all partials are evaluated at a). When not explicitly stated, the matrix representation of $DF(a)$ will always mean the Jacobian matrix representation.

We now prove that differentiability implies continuity. To that end, we first recall that if $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$ then

$$\|AB\|_2 \leq \|A\|_2 \|B\|_2 .$$

The proof of this fact is identical to the one in Example 9.4.16. In particular, if $x \in \mathbb{R}^n$ then $\|Ax\|_2 \leq \|A\|_2 \|x\|$.

Theorem 10.1.3: Differentiability implies Continuity

Let $U \subset \mathbb{R}^n$ be an open set. If $F : U \to \mathbb{R}^m$ is differentiable at $a \in U$ then F is continuous at a.

Proof. Let $\varepsilon_1 = 1$. Then there exists $\delta_1 > 0$ such that if $\|x - a\| < \delta_1$ then

$$\|F(x) - F(a) - DF(a)(x - a) + DF(a)(x - a)\| < 1 \cdot \|x - a\| .$$

Then if $\|x-a\| < \delta_1$ then

$$\begin{aligned}\|F(x)-F(a)\| &= \|F(x)-F(a)-DF(a)(x-a)+DF(a)(x-a)\| \\ &\leq \|F(x)-F(a)-DF(a)(x-a)\| + \|DF(a)(x-a)\| \\ &\leq \|x-a\| + \|DF(a)\|_2\,\|x-a\|\end{aligned}$$

and thus $\|F(x)-F(a)\| < \varepsilon$ provided

$$\|x-a\| < \min\{\delta_1, \varepsilon/(1+\|DF(a)\|_2)\}.$$

Hence, F is continuous at a. □

Notice that Theorem 10.1.2 says that if $DF(a)$ exists then all the relevant partials exist. However, it does not generally hold that if all the relevant partials exist then $DF(a)$ exists. The reason is that partial derivatives are derivatives along the coordinate axes whereas, as seen from the definition, the limit used to define $DF(a)$ is along any direction that $x \to a$.

Example 10.1.4. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined as

$$f(x,y) = \begin{cases} \frac{2xy}{x^2+y^2}, & (x,y) \neq (0,0) \\ 0, & (x,y) = (0,0)\end{cases}$$

We determine whether $\frac{\partial f}{\partial x}(0,0)$ and $\frac{\partial f}{\partial y}(0,0)$ exist. To that end, we compute

$$\lim_{t\to 0} \frac{f(x+t,0)-f(0,0)}{t} = \lim_{t\to 0}\frac{0}{t} = 0$$

$$\lim_{t\to 0} \frac{f(0,y+t)-f(0,0)}{t} = \lim_{t\to 0}\frac{0}{t} = 0$$

Therefore, $\frac{\partial f}{\partial x}(0,0)$ and $\frac{\partial f}{\partial y}(0,0)$ exist and are both equal to zero. It is straightforward to show that f is not continuous at $(0,0)$ and therefore not differentiable at $(0,0)$.

The previous examples shows that existence of partial derivatives is a fairly weak assumption with regards to differentiability, in fact, even with regards to continuity. The following theorem gives a sufficient condition for DF to exist in terms of the partial derivatives.

Theorem 10.1.5: Condition for Differentiability

Let $U \subset \mathbb{R}^n$ be an open set and consider $F : U \to \mathbb{R}^m$ with $F = (f_1, f_2, \ldots, f_m)$. If each partial derivative function $\frac{\partial f_j}{\partial x_i}$ exists and is continuous on U then F is differentiable on U.

We will omit the proof of Theorem 10.1.5.

Example 10.1.6. Let $F : \mathbb{R}^2 \to \mathbb{R}^3$ be defined by

$$F(x) = (x_1 \sin(x_2), x_1 x_2^2, \ln(x_1^2 + 1) + 2x_2).$$

Explain why $DF(x)$ exists for each $x \in \mathbb{R}^2$ and find $DF(x)$.

Solution. It is clear that the component functions of F that are given by $f_1(x) = x_1 \sin(x_2)$, $f_2(x) = x_1 x_2^2$, and $f_3(x) = \ln(x_1^2 + 1) + 2x_2$ have partial derivatives that are continuous on all of $\mathbb{R}^2$. Hence, F is differentiable on $\mathbb{R}^2$. Then

$$DF(x) = \begin{bmatrix} \sin(x_2) & x_1 \cos(x_2) \\ x_2^2 & 2x_1 x_2 \\ \frac{2x_1}{x_1^2+1} & 2 \end{bmatrix}$$

□

Example 10.1.7. Prove that the given function is differentiable on $\mathbb{R}^2$.

$$f(x,y) = \begin{cases} \frac{x^2y^2}{\sqrt{x^2+y^2}}, & (x,y) \neq (0,0) \\ 0, & (x,y) = (0,0) \end{cases}$$

Solution. We compute

$$\lim_{t\to 0} \frac{f(0+t,0) - f(0,0)}{t} = \lim_{t\to} \frac{\frac{0}{\sqrt{t^2}}}{t} = 0$$

and thus $\frac{\partial f}{\partial x}(0,0) = 0$. A similar computations shows that $\frac{\partial f}{\partial y}(0,0) = 0$. On the other hand, if $(x,y) \neq (0,0)$ then

$$\frac{\partial f}{\partial x}(x,y) = \frac{xy^2(x^2+2y^2)}{(x^2+y^2)^{3/2}}$$
$$\frac{\partial f}{\partial y}(x,y) = \frac{x^2y(2x^2+y^2)}{(x^2+y^2)^{3/2}}.$$

To prove that $Df(x,y)$ exists for any $(x,y) \in \mathbb{R}^2$, it is enough to show that $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are continuous on $\mathbb{R}^2$ (Theorem 10.1.5). It is clear that $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are continuous on the open set $U = \mathbb{R}^2 \backslash \{(0,0)\}$ and thus Df exists on U. Now consider the continuity of $\frac{\partial f}{\partial x}$ at $(0,0)$. Using polar coordinates $x = r\cos(\theta)$ and $y = r\sin(\theta)$, we can write

$$\begin{aligned} \frac{\partial f}{\partial x}(x,y) &= \frac{xy^2(x^2+2y^2)}{(x^2+y^2)^{3/2}} \\ &= \frac{r^3\cos(\theta)\sin^2(\theta)(r^2\cos^2(\theta)+2r^2\sin^2(\theta)}{r^3} \\ &= r^2\cos(\theta)\sin^2(\theta)(\cos^2(\theta)+2\sin^2(\theta)) \end{aligned}$$

Now $(x,y) \to (0,0)$ if and only if $r \to 0$ and thus

$$\begin{aligned} \lim_{(x,y)\to(0,0)} \frac{\partial f}{\partial x}(x,y) &= \lim_{r\to 0} \left[r^2\cos(\theta)\sin^2(\theta)(\cos^2(\theta)+2\sin^2(\theta))\right] \\ &= 0 \end{aligned}$$

In other words, $\lim_{(x,y)\to(0,0)} \frac{\partial f}{\partial x}(x, y) = \frac{\partial f}{\partial x}(0, 0)$ and thus $\frac{\partial f}{\partial x}$ is continuous at $(0, 0)$. A similar computation shows that $\frac{\partial f}{\partial y}$ is continuous at $(0, 0)$. Hence, by Theorem 10.1.5, Df exists on $\mathbb{R}^2$. □

If $F : U \subset \mathbb{R}^n \to \mathbb{R}^m$ is differentiable on U and $m = 1$, then DF is called the **gradient** of F and we write ∇F instead of DF. Hence, in this case,

$$\nabla F(x) = \begin{bmatrix} \frac{\partial F}{\partial x_1} & \frac{\partial F}{\partial x_2} & \cdots & \frac{\partial F}{\partial x_n} \end{bmatrix}$$

On the other hand, if $n = 1$ and $m \geq 2$ then $F : U \subset \mathbb{R} \to \mathbb{R}^m$ is a **curve** in $\mathbb{R}^m$. In this case, it is customary to use lower-case letters such as c, α, or γ instead of F, and use I for the domain instead of U. In any case, since $c : I \subset \mathbb{R} \to \mathbb{R}^m$ is a function of one variable we use the notation $c(t) = (c_1(t), c_2(t), \ldots, c_m(t))$ and the derivative of c is denoted by

$$\frac{dc}{dt} = c'(t) = (c_1'(t), c_2'(t), \ldots, c_m'(t))$$

where all derivatives are derivatives of single-variable-single-valued functions.

Exercises

Exercise 10.1.1. Let $f, g : U \subset \mathbb{R}^n \to \mathbb{R}^m$ be differentiable functions at $a \in U$. Prove by definition that $h = f + g$ is differentiable at a and that $Dh = Df + Dg$.

Exercise 10.1.2. Recall that a mapping $F : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **linear** if $F(x+y) = F(x)+F(y)$ and $F(\alpha x) = \alpha F(x)$, for all $x, y \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$. Prove that if F is linear then $DF(a) = F$ for all $a \in \mathbb{R}^n$.

Exercise 10.1.3. Let $F : \mathbb{R}^n \to \mathbb{R}^m$ and suppose that there exists $M > 0$ such that $\|F(x)\| \leq M \|x\|^2$ for all $x \in \mathbb{R}^n$. Prove that F is differentiable at $a = 0 \in \mathbb{R}^n$ and that $DF(a) = 0$.

Exercise 10.1.4. Determine if the given function is differentiable at $(x, y) = (0, 0)$.

$$f(x,y) = \begin{cases} \frac{xy}{\sqrt{x^2+y^2}}, & (x,y) \neq (0,0) \\ 0, & (x,y) = (0,0) \end{cases}$$

Exercise 10.1.5. Compute $DF(x, y, z)$ if $F(x, y, z) = (z^{xy}, x^2, \tan(xyz))$.

10.2 Differentiation Rules and the MVT

Theorem 10.2.1: Chain Rule

Let $U \subset \mathbb{R}^n$ and $W \subset \mathbb{R}^m$ be open sets. Suppose that $F : U \to \mathbb{R}^m$ is differentiable at a, $F(U) \subset W$, and $G : W \to \mathbb{R}^p$ is differentiable at $F(a)$. Then $(G \circ F) : U \to \mathbb{R}^p$ is differentiable at a and

$$D(G \circ F)(a) = DG(F(a)) \circ DF(a)$$

Example 10.2.2. Verify the chain rule for the composite function $H = G \circ F$ where $F : \mathbb{R}^3 \to \mathbb{R}^2$ and $G : \mathbb{R}^2 \to \mathbb{R}^2$ are

$$\begin{aligned} F(x_1, x_2, x_3) &= \begin{bmatrix} x_1 - 3x_2 \\ x_1 x_2 x_3 \end{bmatrix} \\ G(y_1, y_2) &= \begin{bmatrix} 2y_1 + y_2 \\ \sin(y_2) \end{bmatrix}. \end{aligned}$$

An important special case of the chain rule is the composition of a curve $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ with a function $f : U \subset \mathbb{R}^n \to \mathbb{R}$. The composite function $f \circ \gamma : I \to \mathbb{R}$ is a single-variable and single-valued function. In this case, if $\gamma'(t)$ is defined for all $t \in I$ and $\nabla f(x)$ exists at each $x \in U$ then

$$\begin{aligned} D(f \circ \gamma)(t) &= \nabla f(\gamma(t)) \cdot \gamma'(t) \\ &= \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix} \begin{bmatrix} \gamma_1'(t) \\ \gamma_2'(t) \\ \vdots \\ \gamma_n'(t) \end{bmatrix} \\ &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\gamma(t)) \gamma_i'(t). \end{aligned}$$

In the case that $\gamma(t) = a + te$ and $e \in \mathbb{R}^n$ is a unit vector, that is, $\|e\| = 1$, then

$$\begin{aligned}\lim_{t\to 0} \frac{f(a+te) - f(a)}{t} &= D(f \circ \gamma)(0) \\ &= \nabla f(\gamma(0)) \cdot \gamma'(0) \\ &= \nabla f(a) \cdot e\end{aligned}$$

is called the **directional derivative of f at a in the direction $e \in \mathbb{R}^n$.**

Example 10.2.3. Let $f : \mathbb{R} \to \mathbb{R}$ and $F : \mathbb{R}^2 \to \mathbb{R}$ be differentiable and suppose that $F(x, f(x)) = 0$. Prove that if $\frac{\partial F}{\partial y} \neq 0$ then $f'(x) = -\frac{\partial F/\partial x}{\partial F/\partial y}$ where $y = f(x)$.

Below is a version of the product rule for multi-variable functions.

Theorem 10.2.4: Product Rule

Let $U \subset \mathbb{R}^n$ be open and suppose that $F : U \to \mathbb{R}^m$ and $g : U \to \mathbb{R}$ are differentiable at $a \in U$. Then the function $G = gF : U \to \mathbb{R}^m$ is differentiable at $a \in U$ and

$$D(gF)(a) = F(a) \cdot \nabla g(a) + g(a)DF(a).$$

Example 10.2.5. Verify the product rule for $G = gF$ if $g : \mathbb{R}^3 \to \mathbb{R}$ and $F : \mathbb{R}^3 \to \mathbb{R}^3$ are

$$\begin{aligned}g(x_1, x_2, x_3) &= x_1^2 x_3 - e^{x_2} \\ F(x_1, x_2, x_3) &= \begin{bmatrix} x_1 x_2 \\ \ln(x_3^2 + 1) \\ 3x_1 - x_2 - x_3 \end{bmatrix}\end{aligned}$$

Example 10.2.6. Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be differentiable functions. Find an expression of $\nabla(fg)$ in terms of $f, g, \nabla f$, and ∇g.

Example 10.2.7. Let $f : U \subset \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. Suppose that $\gamma : [a, b] \to \mathbb{R}^n$ is differentiable. Prove that $f(\gamma(t)) = f(\gamma(a))$ for all $t \in [a, b]$ if and only if $\nabla f(\gamma(t)) \cdot \gamma'(t) = 0$ for all $t \in [a, b]$.

Recall the mean value theorem (MVT) on $\mathbb{R}$. If $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b) then there exists $c \in (a, b)$ such that $f(b) - f(a) = f'(c)(b - a)$. The MVT does not generally hold for a function $F : U \subset \mathbb{R}^n \to \mathbb{R}^m$ without some restrictions on U and, more importantly, on m. For instance, consider $f : [0, 1] \to \mathbb{R}^2$ defined by $f(x) = (x^2, x^3)$. Then $f(1) - f(0) = (1, 1) - (0, 0) = (1, 1)$ while $f'(c)(1 - 0) = (2c, 3c^2)$ and there is no $c \in \mathbb{R}$ such that $(1, 1) = (2c, 3c^2)$. With regards to the domain U, we will be able to generalize the MVT for points $a, b \in U$ provided all points on the line segment joining a and b are contained in U. Specifically, the **line segment joining** $x, y \in U$ is the set of points

$$\{z \in \mathbb{R}^n \mid z = (1 - t)x + ty,\ t \in [0, 1]\}.$$

Hence, the image of the curve $\gamma : [0, 1] \to \mathbb{R}^n$ given by $\gamma(t) = (1-t)x+ty$ is the line segment joining x and y. Even if $U \subset \mathbb{R}^n$ is open, the line segment joining $x, y \in U$ may not be contained in U (see Figure 10.1).

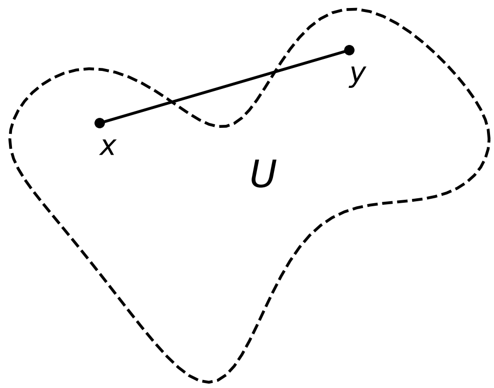


Figure 10.1: Line segment joining x and y not in U

Theorem 10.2.8: Mean Value Theorem

Let $U \subset \mathbb{R}^n$ be open and assume that $f : U \to \mathbb{R}$ is differentiable on U. Let $x, y \in U$ and suppose that the line segment joining $x, y \in U$ is contained entirely in U. Then there exists c on the line segment joining x and y such that $f(y) - f(x) = Df(c)(y - x)$.

Proof. Let $\gamma(t) = (1-t)x + ty$ for $t \in [0,1]$. By assumption, $\gamma(t) \in U$ for all $0 \leq t \leq 1$. Consider the function $h(t) = f(\gamma(t))$ on $[0,1]$. Then h is continuous on $[0,1]$ and by the chain rule is differentiable on $(0,1)$. Hence, applying the MVT on $\mathbb{R}$ to h there exists $t^* \in (0,1)$ such that $h(1) - h(0) = h'(t^*)(1-0)$. Now $h(0) = f(\gamma(0)) = f(x)$ and $h(1) = f(\gamma(1)) = f(y)$, and by the chain rule,

$$\begin{aligned} h'(t^*) &= Df(\gamma(t^*))\gamma'(t^*) \\ &= Df(\gamma(t^*))(y - x). \end{aligned}$$

Hence,

$$f(y) - f(x) = Df(\gamma(t^*))(y - x)$$

and the proof is complete. □

Corollary 10.2.9

Let $U \subset \mathbb{R}^n$ be open and assume that $F = (f_1, f_2, \ldots, f_m) : U \to \mathbb{R}^m$ is differentiable on U. Let $x, y \in U$ and suppose that the line segment joining $x, y \in U$ is contained entirely in U. Then there exists $c_1, c_2, \ldots, c_m \in U$ on the line segment joining x and y such that $f_i(y) - f_i(x) = Df_i(c_i)(y - x)$ for $i = 1, 2, \ldots, m$.

Proof. Apply the MVT to each component function $f_i : U \to \mathbb{R}$ ☐

Example 10.2.10. A set $U \subset \mathbb{R}^n$ is said to be **convex** if for any $x, y \in U$ the line segment joining x and y is contained in U. Let $F : U \to \mathbb{R}^m$ be differentiable. Prove that if U is an open convex set and $DF = 0$ on U then F is constant on U.

Exercises

Exercise 10.2.1. Let $U \subset \mathbb{R}^n$ be an open set satisfying the following property: for any $x, y \in U$ there is a continuous curve $\gamma : [0,1] \to \mathbb{R}^n$ such that γ is differentiable on $(0,1)$ and $\gamma(0) = x$ and $\gamma(1) = y$.

(a) Give an example of a **non-convex** set $U \subset \mathbb{R}^2$ satisfying the above property.

(b) Prove that if U satisfies the above property and $f : U \to \mathbb{R}$ is differentiable on U with $Df = 0$ then f is constant on U.

10.3 The Space of Linear Maps

Let U be an open subset of $\mathbb{R}^n$. Recall that if $F : U \to \mathbb{R}^n$ is differentiable at each $x \in U$ then $DF : U \to \mathcal{L}(\mathbb{R}^n; \mathbb{R}^n)$ denotes the derivative of F on U. The space of linear maps $\mathcal{L}(\mathbb{R}^n; \mathbb{R}^n)$ is a vector space which after

10.4 Solutions to Differential Equations

A differential equation on $\mathbb{R}^n$ is an equation of the form

$$x'(t) = F(x(t)) \tag{10.1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is a given function and $x : \mathbb{R} \to \mathbb{R}^n$ is the unknown in (10.1). A solution to (10.1) is a curve $\gamma : I \to \mathbb{R}^n$ such that

$$\gamma'(t) = F(\gamma(t))$$

where $I \subset \mathbb{R}$ is an interval, possibly infinite. If F is defined

Theorem 10.4.1

Let $U \subset \mathbb{R}^n$ be an open set and let $F : U \to \mathbb{R}^n$ be a differentiable function with a continuous derivative

10.5 High-Order Derivatives

In this section, we consider high-order derivatives of a differentiable mapping $F : U \subset \mathbb{R}^n \to \mathbb{R}^m$. To do this, we will need to make an excursion into the world of **multilinear algebra**. Even though we will discuss high-order derivatives for functions on Euclidean spaces, it will be convenient to first work with general vector spaces.

Definition 10.5.1: Multilinear Maps

Let $V_1, V_2, \ldots, V_k$ and W be vector spaces. A mapping $T : V_1 \times V_2 \times \cdots \times V_k \to W$ is said to be a k**-multilinear map** if T is linear in each variable separately. Specifically, for any $i \in \{1, 2, \ldots, k\}$, and any $v_j \in V_j$ for $j \neq i$, the mapping $T_i : V_i \to W$ defined by

$$T_i(x) = T(v_1, v_2, \ldots, v_{i-1}, x, v_{i+1}, \ldots, v_k)$$

is a linear mapping.

A 1-multilinear mapping is just a linear mapping. A 2-multilinear mapping is called a **bilinear mapping**. Hence, $T : V_1 \times V_2 \to W$ is bilinear if

$$\begin{aligned} T(\alpha u + \beta v, w) &= T(\alpha u, w) + T(\beta v, w) \\ &= \alpha T(u, w) + \beta T(v, w) \end{aligned}$$

and

$$\begin{aligned} T(u, \alpha w + \beta y) &= T(u, \alpha w) + T(u, \beta y) \\ &= \alpha T(u, w) + \beta T(u, y) \end{aligned}$$

for all $u, v \in V_1$, $w \in V_2$, and $\alpha, \beta \in \mathbb{R}$. Roughly speaking, a multilinear mapping is essentially a special type of polynomial multivariable function. We will make this precise after presenting a few examples.

Example 10.5.2. Consider $T : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defined as $T(x, y) = 2xy$. As can be easily verified, T is bilinear. On the other hand, if $T(x, y) = x^2 + y^2$ then T is not bilinear since for example $T(\alpha x, y) = \alpha^2 x^2 + y^2 \neq \alpha T(x, y)$ in general, or $T(a + b, y) = (a + b)^2 + y^2 \neq T(a, y) + T(b, y)$ in general. What about $T(x, y) = 2xy + y^3$?

Example 10.5.3. Let $\{v_1, v_2, \ldots, v_p\}$ be a set of vectors in $\mathbb{R}^n$ and suppose that $x = \sum_{i=1}^p x_i v_i$ and $y = \sum_{i=1}^p y_i v_i$. If $T : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ is bilinear then expand $T(x, y)$ so that it depends only on x_i, y_j and $T(v_i, v_j)$ for $1 \leq i, j \leq p$.

Example 10.5.4. Let M be a $n \times n$ matrix and define $T : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ as $T(u, v) = u^T M v$. Show that T is bilinear. For instance, if say $M = \left[\begin{smallmatrix} 1 & -3 \\ 2 & 1 \end{smallmatrix}\right]$ then

$$\begin{aligned} T(u, v) &= [u_1 \ u_2] \begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \\ &= u_1 v_1 - 3u_1 v_2 + u_2 v_2. \end{aligned}$$

Notice that $T(u, v)$ is a polynomial in the components of u and v.

Example 10.5.5. The function that returns the determinant of a matrix is multilinear in the columns of the matrix. Specifically, if say $A = [a_1 + b_1 \ a_2 \ \cdots \ a_n] \in \mathbb{R}^{n \times n}$ then

$$\det(A) = \det([a_1 \ a_2 \ \cdots \ a_n]) + \det([b_1 \ a_2 \ \cdots \ a_n])$$

and if $A = [\alpha a_1 \ a_2 \ \cdots \ a_n]$ then

$$\det(A) = \alpha \det([a_1 \ a_2 \ \cdots \ a_n]).$$

These facts are proved by expanding the determinant along the first column. The same is true if we perform the same computation with a different column of A. In the case of a 2×2 matrix $A = \left[\begin{smallmatrix} x_1 & y_1 \\ x_2 & y_2 \end{smallmatrix}\right]$ we have

$$\det(A) = x_1 y_2 - y_1 x_2$$

and if A is a 3×3 matrix with columns $x = (x_1, x_2, x_3)$, $y = (y_1, y_2, y_3)$, and $z = (z_1, z_2, z_3)$ then

$$\begin{aligned} \det(A) &= \det([x \ y \ z]) \\ &= x_1 y_2 z_3 - x_1 y_3 z_2 - x_2 y_1 z_3 + x_2 y_3 z_1 + x_3 y_1 z_2 - x_3 y_2 z_1. \end{aligned}$$

We now make precise the statement that a multilinear mapping is a (special type of) multivariable polynomial function. For simplicity, and since this will be the case when we consider high-order derivatives, we consider k-multilinear mappings $T : \mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}^m$. For a positive integer $k \geq 1$ let $(\mathbb{R}^n)^k = \mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n$ where on the right-hand-side $\mathbb{R}^n$ appears k-times. Let $\mathcal{L}^k(\mathbb{R}^n, \mathbb{R}^m)$ denote the space of k-multilinear maps from $(\mathbb{R}^n)^k$ to $\mathbb{R}^m$. It is easy to see that $\mathcal{L}^k(\mathbb{R}^n, \mathbb{R}^m)$ is a vector space under the natural notion of addition and scalar $\mathbb{R}$-multiplication. In what follows we consider the case $k = 3$, the general case is similar but requries more notation. Hence, suppose that $T : (\mathbb{R}^n)^3 \to \mathbb{R}^m$ is a multilinear mapping and let $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n)$, and $z = (z_1, z_2, \ldots, z_n)$. Then $x = \sum_{i=1}^n x_i e_i$ where e_i is the ith standard basis vector of $\mathbb{R}^n$, and similarly for y and z. Therefore, by multilinearity of T we have

$$\begin{aligned} T(x, y, z) &= T\left(\sum_{i=1}^n x_i e_i, \sum_{i=1}^n y_i e_i, \ldots, \sum_{i=1}^n z_i e_i\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n x_i y_j z_k \cdot T(e_i, e_j, e_k). \end{aligned}$$

Thus, to compute $T(x, y, z)$ for any $x, y, z \in \mathbb{R}^n$, we need only know the values $T(e_i, e_j, e_k) \in \mathbb{R}^m$ for all triples (i, j, k) with $1 \leq i, j, k \leq n$. If we set

$$T(e_i, e_j, e_k) = (A^1_{i,j,k}, A^2_{i,j,k}, \ldots, A^m_{i,j,k})$$

where the superscripts are not exponents but indices, then from our computation above

$$T(x, y, z) = \begin{bmatrix} \sum_{i,j,k=1}^{n} A^1_{i,j,k} \cdot x_i y_j z_k \\ \sum_{i,j,k=1}^{n} A^2_{i,j,k} \cdot x_i y_j z_k \\ \vdots \\ \sum_{i,j,k=1}^{n} A^m_{i,j,k} \cdot x_i y_j z_k \end{bmatrix}.$$

Notice that the component functions of T are multilinear, specifically, the mapping

$$(x, y, z) \mapsto T_r(x, y, z) = \sum_{i,j,k=1}^{n} A^r_{i,j,k} \cdot x_i y_j z_k$$

is multilinear for each $r = 1, 2, \ldots, m$. The $n^3 m$ numbers $A^r_{i,j,k} \in \mathbb{R}$ for $1 \leq i, j, k \leq n$ and $1 \leq r \leq m$ completely determine the multilinear mapping T, and we call these the **coefficients** of the multilinear mapping T in the standard bases.

Remark 10.5.6. The general case $k \geq 1$ is just more notation. If $T : (\mathbb{R}^n)^k \to \mathbb{R}^m$ is k-multilinear then there exists $n^k m$ unique coefficients $A^r_{i_1, i_2, \ldots, i_k}$, where $1 \leq i_1, i_2, \ldots, i_k \leq n$ and $1 \leq r \leq m$, such that for any vectors $u_1, u_2, \ldots, u_k \in \mathbb{R}^n$ it holds that

$$T(u_1, u_2, \ldots, u_k) = \sum_{r=1}^{m} \left(\sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_k=1}^{n} A^r_{i_1, i_2, \ldots, i_k} \cdot u_{1,i_1} u_{2,i_2} \cdots u_{k,i_k} \right) e_r$$

where $e_1, e_2, \ldots, e_m$ are the standard basis vectors in $\mathbb{R}^m$.

A multilinear mapping $T \in \mathcal{L}^k(\mathbb{R}^n, \mathbb{R}^m)$ is said to be **symmetric** if the value of T is unchanged after an arbitrary permutation of the inputs to T. In other words, T is symmetric if for any $v_1, v_2, \ldots, v_k \in \mathbb{R}^n$ it holds that

$$T(v_1, v_2, \ldots, v_k) = T(v_{\sigma(1)}, v_{\sigma(2)}, \ldots, v_{\sigma(n)})$$

for any permutation $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$. For instance, if $T : (\mathbb{R}^n)^3 \to \mathbb{R}^m$ is symmetric then for any $u_1, u_2, u_3 \in \mathbb{R}^n$ it holds that

$$\begin{aligned} T(u_1, u_2, u_3) &= T(u_1, u_3, u_2) \\ &= T(u_2, u_1, u_3) \\ &= T(u_2, u_3, u_2) \\ &= T(u_3, u_1, u_2,) \\ &= T(u_3, u_2, u_1). \end{aligned}$$

Example 10.5.7. Consider $T : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ defined by

$$T(x, y) = 2x_1y_1 + 3x_1y_2 + 3y_1x_2 - x_2y_2.$$

Then

$$\begin{aligned} T(y, x) &= 2y_1x_1 + 3y_1x_2 + 3x_1y_2 - y_2x_2 \\ &= T(x, y) \end{aligned}$$

and therefore T is symmetric. Notice that

$$\begin{aligned} T(x, y) &= [x_1 \; x_2] \begin{bmatrix} 2 & 3 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \\ &= x^T M y \end{aligned}$$

and the matrix $M = \left[\begin{smallmatrix} 2 & 3 \\ 3 & -1 \end{smallmatrix}\right]$ is symmetric.

Having introduced the very basics of multilinear mappings, we can proceed with discussing high-order derivatives of vector-valued multi-variable functions. Suppose then that $F : U \to \mathbb{R}$ is differentiable on the open set $U \subset \mathbb{R}^n$ and as usual let $DF : U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ denote the derivative. Now $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is a finite dimensional vector space and can be equipped with a norm (all norms on a given finite dimensional vector space are *equivalent*). Thus, we can speak of differentiability of DF, namely, DF is differentiable at $a \in U$ if there exists a linear mapping $L : \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ such that

$$\lim_{x\to a} \frac{\|DF(x) - DF(a) - L(x-a)\|}{\|x-a\|} = 0.$$

If such an L exists then we denote it by $L = D(DF)(a)$. To simplify the notation, we write instead $D(DF)(a) = D^2F(a)$. Hence, DF is differentiable at $a \in U$ if there exists a linear mapping $D^2F(a) : \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$ such that

$$\lim_{x\to a} \frac{\left\|DF(x) - DF(a) - D^2F(a)(x-a)\right\|}{\|x-a\|} = 0.$$

To say that $D^2F(a)$ is a linear mapping from $\mathbb{R}^n$ to $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is to say that

$$D^2F(a) \in \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)).$$

Let us focus our attention on the space $\mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$. If $L \in \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$ then $L(v) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ for each $v \in \mathbb{R}^n$, and moreover the assignment $v \mapsto L(v)$ is linear, i.e., $L(\alpha v + \beta u) = \alpha L(v) + \beta L(u)$. Now, since $L(v) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, we have that

$$L(v)(\alpha u + \beta w) = \alpha L(v)(u) + \beta L(v)(w).$$

In other words, the mapping

$$(u, v) \mapsto L(u)(v)$$

is bilinear! Hence, L defines (uniquely) a bilinear map $T : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ by

$$T(u, v) = L(u)(v)$$

and the assignment $L \mapsto T$ is linear. Conversely, to any bilinear map $T : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ we associate an element $L \in \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$ defined as

$$L(u)(v) = T(u, v)$$

and the assignment $T \mapsto L$ is linear. We have therefore proved the following.

Lemma 10.5.8

Let V and W be vector spaces. The vector space $\mathcal{L}(V, \mathcal{L}(V, W))$ is isomorphic to the vector space $\mathcal{L}^2(V, W)$ of multilinear maps from $V \times V$ to W.

The punchline is that $D^2F(a) \in \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$ can be viewed in a natural way as a bilinear mapping $D^2F(a) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ and thus from now on we write $D^2F(a)(u, v)$ instead of the more cumbersome $D^2F(a)(u)(v)$.

We now determine a coordinate expression for $D^2F(a)(u, v)$. First of all, if $F = (f_1, f_2, \ldots, f_m)$ then $F(x) = \sum_{j=1}^m f_j(x)e_j$ where $\{e_1, e_2, \ldots, e_m\}$ is the standard basis of $\mathbb{R}^m$. By linearity of the derivative and the product rule of differentiation, we have that $DF = \sum_{j=1}^m Df_j(x)e_j$ and also $D^2F = \sum_{j=1}^m D^2f_j(x)e_j$. Therefore,

$$D^2F(a)(u, v) = \sum_{j=1}^m D^2f_j(a)(u, v)e_j.$$

This shows that we need only consider D^2f for $\mathbb{R}$-valued functions $f : U \subset \mathbb{R}^n \to \mathbb{R}$. Now,

$$Df = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \ldots, \frac{\partial f}{\partial x_n} \end{bmatrix}$$

and thus the Jacobian of $Df : U \to \mathbb{R}^n$ is (Theorem 10.1.2)

$$D^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 x_1} & \frac{\partial^2 f}{\partial x_2 x_1} & \cdots & \frac{\partial^2 f}{\partial x_n x_1} \\ \frac{\partial^2 f}{\partial x_1 x_2} & \frac{\partial^2 f}{\partial x_2 x_2} & \cdots & \frac{\partial^2 f}{\partial x_n x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 x_n} & \frac{\partial^2 f}{\partial x_2 x_n} & \cdots & \frac{\partial^2 f}{\partial x_n x_n} \end{bmatrix}.$$

Therefore,

$$D^2 f(a)(e_i, e_j) = \frac{\partial^2 f}{\partial x_j x_i}(a).$$

Therefore, for any $u = (u_1, u_2, \ldots, u_n)$ and $v = (v_1, v_2, \ldots, v_n)$, by multilinearity we have

$$D^2 f(a)(u, v) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i x_j}(a) u_i v_j.$$

Now, if all second order partials of f are defined and continuous on U we can say more. Let us first introduce some terminology. We say that $f : U \subset \mathbb{R}^n \to \mathbb{R}$ is of **class** C^k if all partial derivatives up to and including order k of f are continuous functions on U.

Theorem 10.5.9: Symmetry of Partial Derivatives

Let $U \subset \mathbb{R}^n$ be an open set and suppose that $f : U \to \mathbb{R}$ is of class C^2. Then

$$\frac{\partial^2 f}{\partial x_i x_j} = \frac{\partial^2 f}{\partial x_j x_i}$$

on U for all $1 \leq i, j \leq n$. Consequently, $D^2 f(a)$ is a symmetric bilinear map on $\mathbb{R}^n \times \mathbb{R}^n$.

If we now go back to a multi-valued function $F : U \to \mathbb{R}^m$ with components $F = (f_1, f_2, \ldots, f_m)$, then if $D^2F(a)$ exists at $a \in U$ then

$$D^2F(a)(u,v) = \begin{bmatrix} \sum_{i,j=1}^n \frac{\partial^2 f_1}{\partial x_i x_j}(a)u_iv_j \\ \sum_{i,j=1}^n \frac{\partial^2 f_2}{\partial x_i x_j}(a)u_iv_j \\ \vdots \\ \sum_{i,j=1}^n \frac{\partial^2 f_m}{\partial x_i x_j}(a)u_iv_j \end{bmatrix}$$

Higher-order derivatives of $F : U \to \mathbb{R}^m$ can be treated similarly. If $D^{k-1}F : U \to \mathcal{L}^{k-1}(\mathbb{R}^n, \mathbb{R}^m)$ is differentiable at $a \in U$ then we denote the derivative at a by $D(D^{k-1})F(a) = D^kF(a)$. Then $D^kF(a) : \mathbb{R}^n \to \mathcal{L}^{k-1}(\mathbb{R}^n, \mathbb{R}^m)$ is a linear map, that is,

$$D^kF(a) \in \mathcal{L}(\mathbb{R}^n, \mathcal{L}^{k-1}(\mathbb{R}^n, \mathbb{R}^m)).$$

The vector space $\mathcal{L}(\mathbb{R}^n, \mathcal{L}^{k-1}(\mathbb{R}^n, \mathbb{R}^m))$ is isomorphic to the space of k-multilinear maps $\mathcal{L}^k(\mathbb{R}^n, \mathbb{R}^m)$. The value of $D^kF(a)$ at $u_1, u_2, \ldots, u_k \in \mathbb{R}^n$ is denoted by $D^kF(a)(u_1, u_2, \ldots, u_k)$. Moreover, $D^kF(a)$ is a symmetric k-multilinear map at each $a \in U$ if F is of class C^k. If $f : U \subset \mathbb{R}^n \to \mathbb{R}$ is of class C^k then for vectors $u_1, u_2, \ldots, u_k \in \mathbb{R}^n$ we have

$$D^kf(a)(u_1, u_2, \ldots, u_k) = \sum_{1 \le i_1, i_2, \ldots, i_k \le n} \frac{\partial^k f}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_k}}(a) u_{1,i_1} u_{2,i_2} \cdots u_{k,i_k}$$

where the summation is over all k-tuples $(i_1, i_2, \ldots, i_k)$ where $i_j \in \{1, 2, \ldots, n\}$. Hence, there are n^k terms in the above summation. In the case that $u_1 = u_2 = \cdots = u_k = x$, the above expression takes the form

$$D^kf(a)(x, x, \ldots, x) = \sum_{1 \le i_1, i_2, \ldots, i_k \le n} \frac{\partial^k f}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_k}}(a) x_{i_1} x_{i_2} \cdots x_{i_k}$$

Example 10.5.10. Compute $D^3f(a)(u, v, w)$ if $f(x, y) = \sin(x - 2y)$, $a = (0, 0)$, and $u, v, w \in \mathbb{R}^2$. Also compute $D^2f(a)(u, u, u)$.

Solution. We compute that $f(0,0)=0$ and

$$\begin{aligned} f_x &= \cos(x-2y) \\ f_y &= -2\cos(x-2y) \end{aligned}$$

and then

$$\begin{aligned} f_{xx} &= -\sin(x-2y) \\ f_{xy} &= f_{yx} = 2\sin(x-2y) \\ f_{yy} &= -4\sin(x-2y) \end{aligned}$$

and then

$$\begin{aligned} f_{xxx} &= -\cos(x-2y) \\ f_{yyy} &= 8\cos(x-2y) \\ f_{xxy} &= f_{xyx} = f_{yxx} = 2\cos(x-2y) \\ f_{xyy} &= f_{yxy} = f_{yyx} = -4\cos(x-2y) \end{aligned}$$

Then,

$$\begin{aligned} D^3f(a)(u,v,w) &= f_{xxx}(a)u_1v_1w_1 + f_{xxy}(a)u_1v_1w_2 + f_{xyx}(a)u_1v_2w_1 \\ &\quad + f_{xyy}(a)u_1v_2w_2 + f_{yxx}(a)u_2v_1w_1 + f_{yxy}(a)u_2v_1w_2 \\ &\quad + f_{yyx}(a)u_2v_2w_1 + f_{yyy}(a)u_2v_2w_2 \\ \\ &= -u_1v_1w_1 + 2u_1v_1w_2 + 2u_1v_2w_1 - 4u_1v_2w_2 \\ &\quad + 2u_2v_1w_1 - 4u_2v_1w_2 - 4u_2v_2w_1 + 8u_2v_2w_2 \\ \\ &= -u_1v_1w_1 + 2(u_1v_1w_2 + u_1v_2w_1 + u_2v_1w_1) \\ &\quad - 4(u_1v_2w_2 + u_2v_1w_2 + u_2v_2w_1) + 8u_2v_2w_2 \end{aligned}$$

If $u=v=w$ then

$$D^3f(a)(u,u,u) = -u_1^3 + 6u_1^2u_2 - 12u_1u_2^2 + 8u_2^3$$

□

10.6 Taylor's Theorem

Taylor's theorem for a function $f : \mathbb{R}^n \to \mathbb{R}$ is as follows.

Theorem 10.6.1: Taylor's Theorem

Let $U \subset \mathbb{R}^n$ be an open set and suppose that $f : U \to \mathbb{R}$ if of class C^{r+1} on U. Let $a \in U$ and suppose that the line segment between a and $x \in U$ lies entirely in U. Then there exists $c \in U$ on the line segment such that

$$f(x) = f(a) + \sum_{k=1}^{r} \frac{1}{k!} D^k f(a)(x-a, x-a, \ldots, x-a) + R_r(x)$$

where

$$R_r(x) = \frac{1}{(r+1)!} D^{r+1} f(c)(x-a, x-a, \ldots, x-a).$$

Furthermore,

$$\lim_{x \to a} \frac{R_r(x)}{\|x-a\|^r} = 0$$

If $x = a + h$ in Taylor's theorem then

$$\begin{aligned} f(a+h) = f(a) + \sum_{k=1}^{r} \frac{1}{k!} D^k f(a)(h, h, \ldots, h) \\ + \frac{1}{(r+1)!} D^{r+1} f(c)(h, h, \ldots, h) \end{aligned}$$

and

$$\lim_{h \to 0} \frac{R_r(h)}{\|h\|^r} = 0.$$

We call

$$T_r(x) = f(a) + \sum_{k=1}^{r} \frac{1}{k!} D^k f(a)(x-a, x-a, \ldots, x-a)$$

the r**th order Taylor polynomial of** f **centered at** a and

$$R_r(x) = \frac{1}{(r+1)!} D^{r+1} f(c)(x-a, x-a, \ldots, x-a)$$

the rth order remainder term. Hence, Taylor's theorem says that

$$f(x) = T_r(x) + R_r(x)$$

Since $\lim_{x \to a} R_r(x) = 0$, for x close to a we get an approximation

$$f(x) \approx T_r(x).$$

Moreover, since $D^{r+1} f$ is continuous, there is a constant $M > 0$ such that if x is sufficiently close to a then the remainder term satisfies the bound

$$|R_r(x)| \leq M \left\| x - a \right\|^{r+1}.$$

From this it follows that

$$\lim_{x \to a} \frac{R_r(x)}{\left\| x - a \right\|^{r}} = 0$$

Example 10.6.2. Compute the third-order Taylor polynomial of $f(x, y) = \sin(x - 2y)$ centered at $a = (0, 0)$.

Solution. Most of the work has been done in Example 10.5.10. Evaluating all derivatives at a we find that

$$Df(a)(u) = f_x(a)u_1 + f_y(a)u_2 = u_1 - 2u_2$$

$$D^2 f(a)(u, u) = 0$$

$$D^3 f(a)(u, u, u) = -u_1^3 + 6u_1^2 u_2 - 12 u_1 u_2^2 + 8u_2^3$$

Therefore,

$$T_r(u) = u_1 - 2u_2 - u_1^3 + 6u_1^2 u_2 - 12 u_1 u_2^2 + 8u_2^3.$$

□

Exercises

Exercise 10.6.1. Find the 2nd order Taylor polynomial of the function $f(x, y, z) = \cos(x + 2y)e^z$ centered at $a = (0, 0, 0)$.

Exercise 10.6.2. A function $L : \mathbb{R}^n \to \mathbb{R}$ is called a **homogeneous function of degree** $k \in \mathbb{N}$ if for all $\alpha \in \mathbb{R}$ and $x \in \mathbb{R}^n$ it holds that $L(\alpha x) = \alpha^k L(x)$. Prove that if $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $a \in \mathbb{R}^n$ then the mapping

$$L(x) = D^k f(a)(x, x, \ldots, x)$$

is a homogeneous function of degree $k \in \mathbb{N}$.

10.7 The Inverse Function Theorem

A square linear system

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n &= y_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n &= y_2 \\ \vdots \qquad\qquad\qquad \vdots \quad &= \vdots \\ a_{n,1}x_1 + a_{n,2}x_2 + \cdots + a_{n,n}x_n &= y_n \end{aligned}$$

or in vector form

$$Ax = y,$$

where the unknown is $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, has a unique solution if and only if A^{-1} exists if and only if $\det(A) \neq 0$. In this case, the solution is $y = A^{-1}x$. Another way to say this is that the mapping $F(x) = Ax$ has a global inverse given by $F^{-1}(x) = A^{-1}x$. Hence, invertibility of $DF = A$ completely determines whether F is invertible. Consider now a system of equations

$$F(x) = y$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is nonlinear. When is it possible to solve for x in terms of y, that is, when does $F^{-1}(x)$ exists? In general, this is a difficult problem and we cannot expect global invertibility even when assuming the most desirable conditions on F. Even in the 1D case, we cannot expect global invertibility. For instance, $f(x) = \cos(x)$ is not globally invertible but is so on any interval where $f'(x) \neq 0$. For instance, on the interval $I = (0, \pi)$, we have that $f'(x) = \sin(x) \neq 0$ and $f^{-1}(x) = \arcsin(x)$. In any neighborhood where $f'(x) = 0$, for instance, at $x = 0$, $f(x) = \cos(x)$ is not invertible. However, having a non-zero derivative is not necessary for invertibility. For instance, the function $f(x) = x^3$ has $f'(0) = 0$ but $f(x)$ has an inverse locally around $x = 0$; in fact it has a global inverse $f^{-1}(x) = x^{1/3}$.

Let's go back to the 1D case and see if we can say something about the invertibility of $f : \mathbb{R} \to \mathbb{R}$ locally about a point a such that $f'(a) \neq 0$. Assume that f' is continuous on $\mathbb{R}$ (or on an open set containing a). Then there is an interval $I = [a - \delta, a + \delta]$ such that $f'(x) \neq 0$ for all $x \in I$. Now if $x, y \in I$ and $x \neq y$, then by the Mean Value Theorem, there exists c in between x and y such that

$$f(y) - f(x) = f'(c)(y - x).$$

Since $f'(c) \neq 0$ and $(y - x) \neq 0$ then $f(y) \neq f(x)$. Hence, if $x \neq y$ then $f(y) \neq f(x)$ and this proves that f is injective on $I = [c - \delta, c + \delta]$. Therefore, the function $f : I \to \mathbb{R}$ has an inverse $f^{-1} : J \to \mathbb{R}$ where $J = f(I)$. Hence, if $f(a) \neq 0$, f has a **local** inverse at a. In fact, we can say even more, namely, one can show that f^{-1} is also differentiable. Then, since $f^{-1}(f(x)) = x$ for $x \in I$, by the chain rule we have

$$(f^{-1})'(f(x)) \cdot f'(x) = 1$$

and therefore since $f'(x) \neq 0$ for all $x \in I$ we have

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

The following theorem is a generalization of this idea.

Theorem 10.7.1: Inverse Function Theorem

Let $V \subset \mathbb{R}^n$ be an open set and let $F : V \to \mathbb{R}^n$ be of class C^1. Suppose that $\det(DF(a)) \neq 0$ for $a \in V$. Then there exists an open set $U \subset \mathbb{R}^n$ containing a such that $W = F(U)$ is open and $F : U \to W$ is invertible. Moreover, the inverse function $F^{-1} : W \to U$ is also C^1 and for $y \in W$ and $x = F^{-1}(y)$ we have

$$DF^{-1}(y) = [DF(x)]^{-1}.$$

Example 10.7.2. Prove that $F(x, y) = (f_1(x, y), f_2(x, y)) = (x^2 - y^2, 2xy)$ is locally invertible at all points $a \neq (0, 0)$.

Proof. Clearly, $DF(x, y)$ exists for all (x, y) since all partials of the components of F are continuous on $\mathbb{R}^2$. A direct computation gives

$$DF(x, y) = \begin{bmatrix} 2x & -2y \\ 2y & 2x \end{bmatrix}$$

and thus $\det(DF(x, y)) = 2x^2 + 2y^2$. Clearly, $\det(DF(x, y)) = 0$ if and only if $(x, y) = (0, 0)$. Therefore, by the Inverse Function theorem, for each non-zero $a \in \mathbb{R}^2$ there exists an open set $U \subset \mathbb{R}^2$ containing a such that $F : U \to F(U)$ is invertible. In this very special case, we can find the local inverse of F about some $a \in \mathbb{R}^2$. Let $(u, v) = F(x, y)$, that is,

$$x^2 - y^2 = u$$
$$2xy = v$$

If $x \neq 0$ then $y = \frac{v}{2x}$ and therefore $x^2 - \frac{v^2}{4x^2} = u$ and therefore $4x^4 - v^2 = 4ux^2$ or

$$4x^4 - 4ux^2 - v^2 = 0.$$

By the quadratic formula,

$$x^2 = \frac{4u \pm \sqrt{16u^2 + 16v^2}}{8}$$

Since $x \in \mathbb{R}$ we must take

$$\begin{aligned} x &= \sqrt{\frac{4u + \sqrt{16u^2 + 16v^2}}{8}} \\ &= \sqrt{\frac{u + \sqrt{u^2 + v^2}}{2}} \end{aligned}$$

and therefore

$$y = \frac{v}{2x} = \frac{\sqrt{2}v}{2\sqrt{u + \sqrt{u^2 + v^2}}}$$

Hence, provided $u \neq 0$ and $v \neq 0$ then

$$F^{-1}(u, v) = \begin{bmatrix} \sqrt{\frac{u+\sqrt{u^2+v^2}}{2}} \\ \frac{\sqrt{2}v}{2\sqrt{u+\sqrt{u^2+v^2}}} \end{bmatrix}.$$

□

Exercises

Exercise 10.7.1. Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$F(x, y) = (f_1(x, y), f_2(x, y)) = (e^x \cos(y), e^x \sin(y))$$

for $(x, y) \in \mathbb{R}^2$.

(a) Prove that the range of F is $\mathbb{R}^2 \backslash \{0\}$. Hint: Think polar coordinates.

(b) Prove that F is not injective.

(c) Prove that F is locally invertible at every $a \in \mathbb{R}^2$.

Exercise 10.7.2. Can the system of equations

$$\begin{aligned} x + xyz &= u \\ y + xy &= v \\ z + 2x + 3z^2 &= w \end{aligned}$$

be solved for x, y, z in terms of u, v, w near $(0, 0, 0)$?

Bibliography

[1] Georg Cantor. *Über eiene elementare Frage der Mannigfaltigskeitslehre*, Jahresber. der DMV, 1: 75-78. 26

[2] Derek Goldrei. *Classical Set Theory: For Guided Independent Study*. Chapman and Hall/CRC, 1996. 32, 33