

ETH ZÜRICH

Analysis I: One Variable

Lecture Notes 2025

ALESSIO FIGALLI

February 2, 2026

Preface

These notes originate from the Analysis I course at ETH Zürich. Over the years, various internal lecture notes have circulated; building on that tradition, I wrote a complete set of notes during the academic year 2023/2024 and have since undertaken a thorough revision to improve clarity, coherence, and self-containment. The present text focuses on analysis in one real variable and is intended to reflect the first-semester course as it is taught to students in Mathematics, Physics, and Interdisciplinary Natural Sciences.

The guiding philosophy is simple: definitions and theorems are introduced only when they are truly needed, proofs are written with an emphasis on structure rather than length, and examples are used to illuminate ideas rather than to replace them. While many topics will look familiar from high school (limits, derivatives, integrals), the viewpoint is different: we develop a rigorous theory that explains why the methods of calculus work and how they fit together. The course requires little beyond basic algebra and an intuitive understanding of functions; from there, we build the logical framework needed to support later studies, starting from multivariable calculus in Analysis II.

The organization mirrors the flow of a first encounter with rigorous analysis. An introductory chapter places classical problems (such as the quadrature of the parabola) in a historical and conceptual perspective, and it is followed by practical advice on studying effectively. We then introduce the real numbers through their algebraic and order properties and isolate completeness as the key axiom; consequences such as the Archimedean principle and decimal expansions are discussed. Complex numbers are introduced early to streamline later discussions of series and power series.

With this foundation, we turn to sequences, limits, and the first qualitative properties of functions (boundedness, monotonicity, continuity). Compactness on closed intervals leads to fundamental results such as the intermediate value property, existence of extrema, and uniform continuity, and prepares the ground for a precise treatment of exponential and logarithmic functions. The material on limits of functions and Landau notation formalizes the asymptotic language used throughout analysis and applications. A short chapter on sequences of functions (pointwise versus uniform convergence) anticipates the role of uniformity in interchange of limits, differentiation, and integration.

Series are handled in parallel with sequences: we discuss non-negative and conditionally convergent series, absolute convergence and its criteria, reordering, and products. Power series are studied both in the real and complex settings, with radius of convergence, termwise operations, and the construction of the elementary functions (exponential, trigonometric, logarithmic) from their series expansions, including polar coordinates and the complex logarithm.

Differential calculus begins with the derivative as a limit and its geometric meaning, proceeds through the calculus rules, and culminates in the mean value theorem and its consequences (monotonicity, convexity, l'Hôpital's rule). Integration is developed via step functions and Riemann sums, followed by integrability criteria and basic properties. The Fundamental Theorem of Calculus binds differentiation and integration, after which we present standard

techniques (parts, substitution, rational functions), improper integrals, and a first look at special functions such as the Gamma function. A chapter on Taylor polynomials and analytic functions closes the circle between local approximation and global information.

The final part provides a concise introduction to ordinary differential equations: linear and autonomous first-order equations, linear second-order equations with constant coefficients, and an existence–uniqueness theorem for first-order problems. The emphasis is on illustrating how the tools developed earlier—continuity, differentiability, integration, and series—combine to yield both qualitative and quantitative information about solutions.

Throughout, proofs are included for the core results used later in the text; when an argument is instructive but not essential, it will be marked as “Extra material”. Numerous examples and exercises are interspersed to consolidate understanding; serious engagement with them is indispensable, as mastery in mathematics is achieved primarily through problem solving.

These notes will continue to evolve. I have aimed for a balance between brevity and completeness, avoiding unnecessary generality while keeping the pathway to further topics as transparent as possible. I hope the text will serve both as a reliable companion during the semester and as a reference to which students can return in later studies.

Contents

1	Introduction	2
1.1	QUADRATURE OF THE PARABOLA	2
1.2	NAIVE SET THEORY	6
1.3	TIPS ON STUDYING	9
2	The Real Numbers: Maximum, Supremum, and Sequences	10
2.1	THE AXIOMS OF THE REAL NUMBERS	10
2.1.1	Ordered Fields	10
2.1.2	Completeness Axiom	26
2.1.3	Intervals	31
2.2	COMPLEX NUMBERS	35
2.2.1	Definition of Complex Numbers	35
2.2.2	The Absolute Value on the Complex Numbers	38
2.3	MAXIMUM AND SUPREMUM	42
2.3.1	Existence of the Supremum	42
2.3.2	Two-point Compactification	45
2.4	CONSEQUENCES OF COMPLETENESS	47
2.4.1	The Archimedean Principle	47
2.4.2	Decimal Fraction Expansion and Uncountability	50
2.5	SEQUENCES OF REAL NUMBERS	58
2.5.1	Convergence of Sequences	58
2.5.2	Convergent Subsequences and Accumulation Points	59
2.5.3	Addition, Multiplication, and Inequalities	61
2.5.4	Bounded and Monotone Sequences	64
2.5.5	Superior and Inferior Limits	67
2.5.6	Cauchy Sequences	71
2.5.7	Improper Limits	72
2.6	SEQUENCES OF COMPLEX NUMBERS	74
3	Functions of one Real Variable	76
3.1	REAL-VALUED FUNCTIONS	76
3.1.1	Boundedness and Monotonicity	76
3.1.2	Continuity	78

3.1.3	Sequential Continuity	83
3.2	CONTINUOUS FUNCTIONS	85
3.2.1	The Intermediate Value Theorem	85
3.2.2	Inverse Function Theorem	87
3.3	CONTINUOUS FUNCTIONS ON COMPACT INTERVALS	90
3.3.1	Boundedness and Extrema	90
3.3.2	Uniform Continuity	91
3.4	EXAMPLE: EXPONENTIAL AND LOGARITHMIC FUNCTIONS	94
3.4.1	Definition of the Exponential Function	94
3.4.2	Properties of the Exponential Function	97
3.4.3	The Natural Logarithm	99
3.5	LIMITS OF FUNCTIONS	103
3.5.1	Limit in the Vicinity of a Point	103
3.5.2	One-sided Limits	105
3.5.3	Landau Notation	108
3.6	SEQUENCES OF FUNCTIONS	112
3.6.1	Pointwise Convergence	112
3.6.2	Uniform Convergence	113
4	Series and Power Series	117
4.1	SERIES OF REAL NUMBERS	117
4.1.1	Series with Nonnegative Elements	119
4.1.2	Conditional Convergence	122
4.1.3	Leibniz Criterion for Alternating Series	123
4.2	ABSOLUTE CONVERGENCE	125
4.2.1	Criteria for Absolute Convergence	125
4.2.2	Reordering Series	128
4.2.3	Products of Series	129
4.3	SERIES OF COMPLEX NUMBERS	133
4.4	POWER SERIES	134
4.4.1	Radius of Convergence	134
4.4.2	Complex Power Series	137
4.5	EXAMPLE: EXPONENTIAL AND TRIGONOMETRIC FUNCTIONS	140
4.5.1	The Exponential Map as a Power Series	140
4.5.2	Sine and Cosine	143
4.5.3	The Circle Number	145
4.5.4	Polar Coordinates and Multiplication of Complex Numbers	148
4.5.5	The Complex Logarithm	150
4.5.6	Other Trigonometric and Hyperbolic Functions	150
5	Differential Calculus	153
5.1	THE DERIVATIVE	153

5.1.1	Definition and Geometrical Interpretation	153
5.1.2	Differentiation Rules	157
5.2	MAIN THEOREMS OF DIFFERENTIAL CALCULUS	164
5.2.1	Local Extrema	164
5.2.2	The Mean Value Theorem	165
5.2.3	L'Hôpital's Rule	169
5.2.4	Monotonicity and Convexity via Differential Calculus	173
5.3	EXAMPLE: DIFFERENTIATION OF TRIGONOMETRIC FUNCTIONS	177
5.3.1	Sine and Arc Sine	177
5.3.2	Cosine and Arc Cosine	178
5.3.3	Tangent and Arc Tangent	178
5.3.4	Hyperbolic Functions	180
6	The Riemann Integral	182
6.1	STEP FUNCTIONS AND THEIR INTEGRAL	182
6.1.1	Decompositions and Step Functions	182
6.1.2	The Integral of a Step Function	184
6.2	DEFINITION AND FIRST PROPERTIES OF THE RIEMANN INTEGRAL	188
6.2.1	Integrability of Real-valued Functions	188
6.2.2	Linearity and Monotonicity of the Riemann Integral	190
6.3	INTEGRABILITY THEOREMS	195
6.3.1	Integrability of Monotone Functions	195
6.3.2	Integrability of Continuous Functions	196
6.3.3	Integration and Sequences of Functions	198
7	The Derivative and the Riemann Integral	200
7.1	THE FUNDAMENTAL THEOREM OF CALCULUS	200
7.1.1	The Fundamental Theorem	200
7.1.2	Integration by Parts and by Substitution	203
7.1.3	Improper Integrals	205
7.2	INTEGRATION AND DIFFERENTIATION OF POWER SERIES	209
7.3	INTEGRATION METHODS	214
7.3.1	Integration by Parts and by Substitution in Leibniz Notation	214
7.3.2	Integration by Parts: Examples	215
7.3.3	Integration by Substitution: Examples	216
7.3.4	Integration of Rational Functions	219
7.3.5	Definite Integrals	222
7.3.6	The Gamma Function	223
7.4	TAYLOR SERIES	224
7.4.1	Taylor Approximation	224
7.4.2	Analytic Functions	231
8	Ordinary Differential Equations	234

8.1	ORDINARY DIFFERENTIAL EQUATIONS (ODEs)	234
8.1.1	Linear First Order ODEs	237
8.1.2	Autonomous First Order ODEs	242
8.1.3	Homogeneous Linear Second Order ODEs with Constant Coefficients	246
8.1.4	Non-Homogeneous Linear Second Order ODEs with Constant Coefficients	253
8.2	EXISTENCE AND UNIQUENESS FOR ODEs	259
8.2.1	Existence and Uniqueness for First Order ODEs	259
8.2.2	Extra material: Proof of Theorem 8.26	263
8.2.3	Higher Order ODEs	267

Chapter 1

Introduction

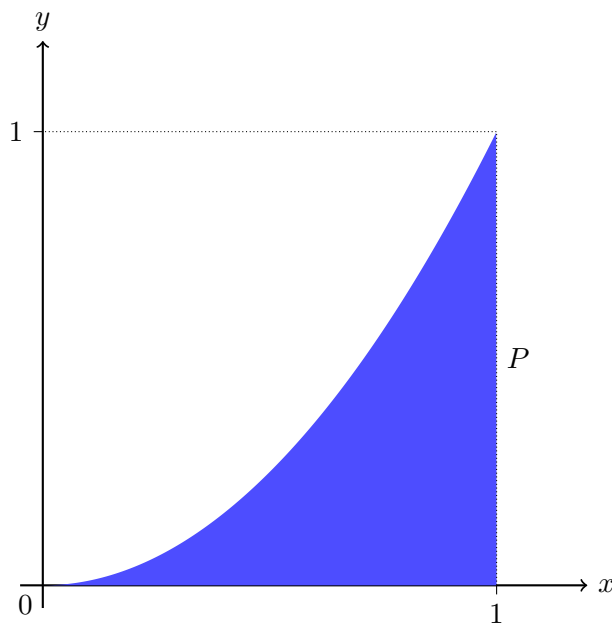
1.1 Quadrature of the Parabola

Before embarking on our journey into the world of mathematical analysis, let us examine a simple example that illustrates the way of thinking we aim to develop. At the same time, it will serve as a first encounter with the ideas behind integral calculus.

Consider the set

$$P = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq x \leq 1, 0 \leq y \leq x^2\}. \quad (1.1)$$

Our goal is to determine its area.



This region, bounded above by a parabola, was already studied by Archimedes (ca. 287–212 BC), who computed it as the first example of a curvilinear area in the 3rd century BC.

For now, let us assume that we understand the symbols in (1.1) and that P corresponds to the colored region in the figure. In particular, we take for granted the existence of the real numbers \mathbb{R} .

Of course, once we know integral calculus, the computation is straightforward. But at this stage we do not want to rely on integrals. Instead, we must face the more basic question:

What is an area?

If we cannot answer this question precisely, then the task of “computing the area of P ” is not well defined. To proceed, we reformulate our aim as follows.

PROPOSITION 1.1. — *Suppose there is a notion of area in \mathbb{R}^2 satisfying:*

1. *The area of the rectangle $[a, b] \times [c, d] = \{(x, y) \in \mathbb{R}^2 \mid a \leq x \leq b, c \leq y \leq d\}$ is equal to $(b - a)(d - c)$, where a, b, c, d are real numbers with $a \leq b, c \leq d$.*
2. *If F, G are domains in \mathbb{R}^2 with F contained in G , then the area of F is less than or equal to the area of G .*
3. *For sets F, G in \mathbb{R}^2 without common points, the area of the union $F \cup G$ is the sum of the areas of F and G .*

Then the area of the set P defined in (1.1) (if it exists) must be equal to $\frac{1}{3}$.

In other words, without deciding in general how area is defined, we can already prove that $\frac{1}{3}$ is the only consistent value for the area of P .

For the proof of Proposition 1.1 we need the following lemma:

LEMMA 1.2. — *For every natural number $n \geq 1$,*

$$1^2 + 2^2 + \cdots + n^2 = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6}. \quad (1.2)$$

The proof of this result uses a standard technique that will appear many times throughout your studies: the *principle of mathematical induction*. The idea of induction is the following:

- (i) First, you verify that the statement holds for the initial natural number where it is intended to be true (usually $n = 1$). This is called the *base case*.
- (ii) Then, you assume that the statement is true for some number n , and under this assumption, you prove that it is also true for $n + 1$. This is called the *induction step*.
- (iii) If both steps (i) and (ii) succeed, one concludes that the statement is valid for all natural numbers because starting from $n = 1$, the induction step allows one to go to $n = 2$, then to $n = 3$, then to $n = 4$, and so on.

Proof of Lemma 1.2. As mentioned above, we apply induction.

(i) *Base case.* For $n = 1$, the left-hand side of (1.2) is 1, while the right-hand side equals

$$\frac{1}{3} + \frac{1}{2} + \frac{1}{6} = 1.$$

Hence (1.2) holds for $n = 1$.

(ii) *Induction step.* Assume that for some $n \geq 1$,

$$1^2 + 2^2 + \cdots + n^2 = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6}. \quad (1.3)$$

We want to show the validity of the same formula for $n + 1$, that is

$$1^2 + 2^2 + \cdots + n^2 + (n + 1)^2 = \frac{(n + 1)^3}{3} + \frac{(n + 1)^2}{2} + \frac{n + 1}{6}. \quad (1.4)$$

To this aim, we can rely on the assumption that the formula already holds for n . So, using (1.3) and expanding $(n + 1)^2$ we get

$$\begin{aligned} 1^2 + \cdots + n^2 + (n + 1)^2 &= \left(\frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6} \right) + (n^2 + 2n + 1) \\ &= \frac{n^3}{3} + \frac{3n^2}{2} + \frac{13n}{6} + 1. \end{aligned}$$

On the other hand, expanding the terms $(n + 1)^3$ and $(n + 1)^2$, we get

$$\begin{aligned} \frac{(n + 1)^3}{3} + \frac{(n + 1)^2}{2} + \frac{n + 1}{6} &= \frac{n^3 + 3n^2 + 3n + 1}{3} + \frac{n^2 + 2n + 1}{2} + \frac{n + 1}{6} \\ &= \frac{n^3}{3} + \frac{3n^2}{2} + \frac{13n}{6} + 1, \end{aligned}$$

which matches the previous expression. Hence (1.4) holds.

(iii) *Conclusion.* Since the identity is true for $n = 1$, and the induction step shows it remains true from n to $n + 1$, the principle of induction implies that (1.2) holds for all $n \geq 1$. \square

Proof of Proposition 1.1. Assume that there is a notion of area satisfying the three properties of the proposition, and that it is defined for P . Let \mathcal{A} denote the area of P .

Fix $n \geq 1$. We partition the interval $[0, 1]$ into n subintervals of equal length $\frac{1}{n}$. Above each subinterval we place a rectangle of width $\frac{1}{n}$. If we choose the *right endpoint* of each subinterval to determine the height, we obtain the rectangles shown in Figure 1.1 (left images).

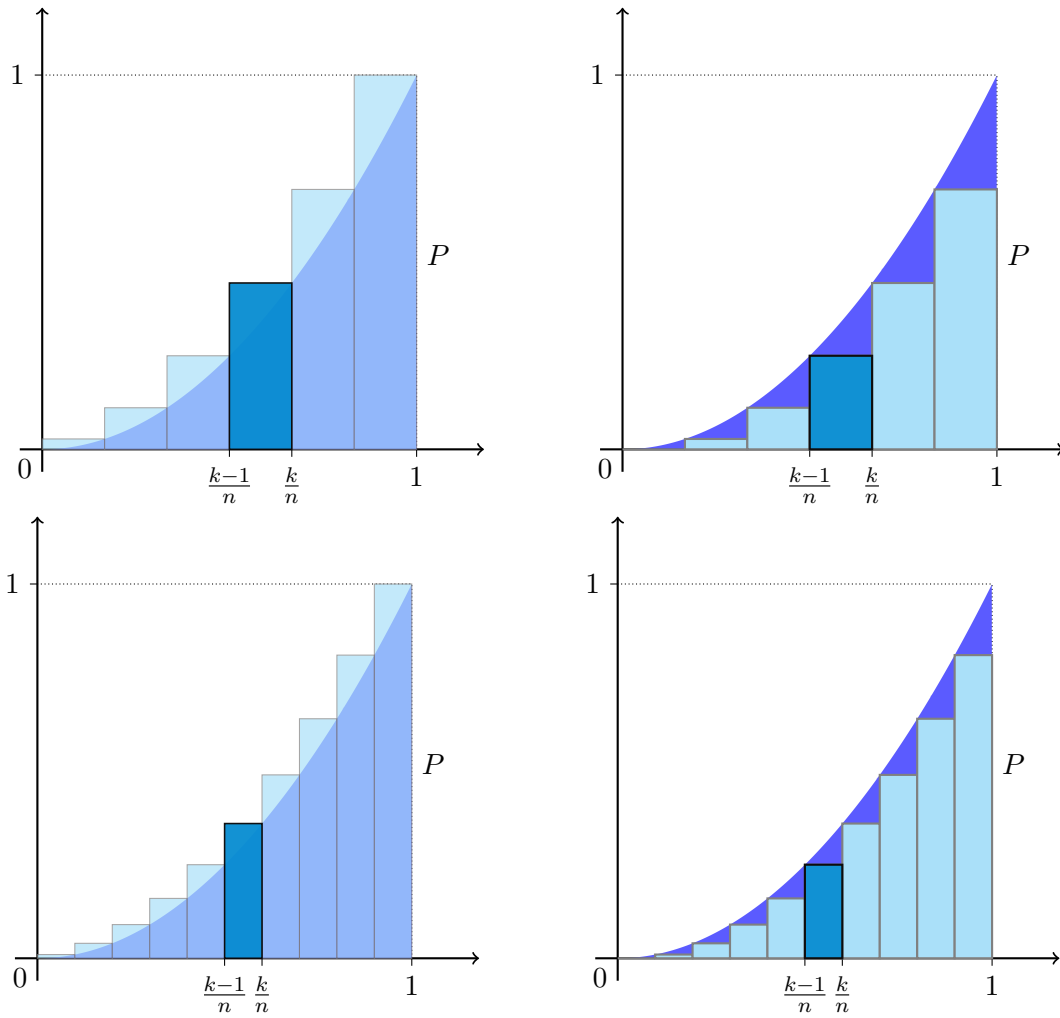


Figure 1.1: The cases $n = 6$ and $n = 10$. In the left pictures the k -th rectangle has height $\left(\frac{k}{n}\right)^2$, while in the right pictures it has height $\left(\frac{k-1}{n}\right)^2$.

In this construction, the k -th rectangle has height $\left(\frac{k}{n}\right)^2$. Hence, by additivity of the area and Lemma 1.2, their total area gives an *upper sum*:

$$\begin{aligned}
 \mathcal{A} &\leq \frac{1}{n} \left(\frac{1^2}{n^2} + \frac{2^2}{n^2} + \cdots + \frac{n^2}{n^2} \right) \\
 &= \frac{1}{n^3} (1^2 + 2^2 + \cdots + n^2) \\
 &= \frac{1}{n^3} \left(\frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6} \right) \\
 &= \frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2}.
 \end{aligned}$$

(The line segments where the rectangles meet have zero area and can be ignored.)

Similarly, if instead we choose the *left endpoint* of each subinterval, we obtain the rectangles shown in Figure 1.1 (right images). In this case, the k -th rectangle has height $\left(\frac{k-1}{n}\right)^2$, and

their total area gives a *lower sum*:

$$\begin{aligned}
 \mathcal{A} &\geq \frac{1}{n} \left(\frac{0}{n^2} + \frac{1^2}{n^2} + \cdots + \frac{(n-1)^2}{n^2} \right) \\
 &= \frac{1}{n^3} (1^2 + 2^2 + \cdots + (n-1)^2) \\
 &= \frac{1}{n^3} \left(\frac{(n-1)^3}{3} + \frac{(n-1)^2}{2} + \frac{n-1}{6} \right) \\
 &= \frac{1}{n^3} \left(\frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6} - n^2 \right) \\
 &= \frac{1}{3} - \frac{1}{2n} + \frac{1}{6n^2}.
 \end{aligned}$$

In summary, for every $n \geq 1$,

$$-\frac{1}{2n} + \frac{1}{6n^2} \leq \mathcal{A} - \frac{1}{3} \leq \frac{1}{2n} + \frac{1}{6n^2}. \quad (1.5)$$

The only real number $\mathcal{A} - \frac{1}{3}$ that satisfies these inequalities for all n is 0. Hence, $\mathcal{A} = \frac{1}{3}$. \square

1

REMARK 1.3. — In the language of Riemann integration (that we will discuss in Chapter 6), we have shown that the lower sums converge to $\frac{1}{3}$ from below, while the upper sums converge to $\frac{1}{3}$ from above; therefore, $\mathcal{A} = \frac{1}{3}$.

To make the last step rigorous, one must prove that the only real number satisfying (1.5) for all $n \geq 1$ is 0. Intuitively, this is clear: as n grows, both bounds $\frac{1}{2n} + \frac{1}{6n^2}$ and $-\frac{1}{2n} + \frac{1}{6n^2}$ shrink toward 0, leaving no other possible value. At this stage, however, we cannot yet give a fully rigorous proof, because we have not defined the real numbers precisely.

Moreover, we have not yet settled the question of which subsets of \mathbb{R}^2 admit an area, or what exactly we mean by “area”. Here we have worked under the implicit assumption that such a notion exists for sets like P . These foundational questions will be addressed in later classes, with the introduction of the Riemann and Lebesgue integrals and the concept of measurable sets.

1.2 Naive Set Theory

Note that in the previous examples we informally used the notion of a “set.” For completeness, we now give a more precise (though still informal, or *naive*) description.

NAIVE SET THEORY

The central assumptions of naive set theory are the following postulates:

- (1) A **set** is a collection of distinct objects, called its **elements**.
- (2) A set is completely determined by its **elements** (axiom of extensionality).
- (3) If $A(x)$ is any property of elements of a set X , then

$$\{x \in X \mid A(x)\}$$

denotes the set of all elements x in X for which $A(x)$ holds (set-builder notation).

The **empty set**, written \emptyset (or sometimes $\{\}$), is the set containing no elements.

In addition, Zermelo–Fraenkel set theory assumes the *axiom of regularity*, which implies that no set is an element of itself.

1

We write $x \in X$ if x is an **element** of the set X , and $x \notin X$ otherwise.

A set can sometimes be described by listing its elements explicitly, for example

$$X = \{x_1, x_2, \dots, x_n\}.$$

More often, we describe sets using a property that characterizes their elements. According to postulate (3), this is done using the so-called **set-builder notation**. For instance, both

$$\{n \in \mathbb{Z} \mid n \text{ is even}\} \quad \text{and} \quad \{n \in \mathbb{Z} \mid \exists m \in \mathbb{Z} : n = 2m\}$$

denote the set of even integers.

Here, the symbol “ \exists ” means “there exists”, while “ \forall ” means “for all”. The symbols “ \mid ” and “ $:$ ” both mean “such that”.

For convenience, we summarize below some of the logical symbols that will be used throughout these notes. With time and practice, this mathematical language will become familiar.

LOGICAL SYMBOLS

- \forall : “for all” or “for every”.
- \exists : “there exists”.
- \Rightarrow : “implies” or “if ... then ...”.
- \Leftarrow : “is implied by”.
- \Leftrightarrow : “if and only if” (logical equivalence).
- \nRightarrow : “does not imply”.

For example,

$$\begin{aligned}x > 2 &\Rightarrow x^2 > 4, & \text{or equivalently} & & x^2 > 4 &\Leftarrow x > 2; \\x^2 > 4 &\nRightarrow x > 2, & x \in \mathbb{Q} &\Rightarrow x \in \mathbb{R}, & x = 3 &\Leftrightarrow 3x = 9.\end{aligned}$$

1.3 Tips on Studying

All beginnings are difficult. Starting university mathematics is not easy. The subject has its own language and way of thinking, and you will need some time to become fluent in it. The earlier you begin engaging actively with this new way of working, the more rewarding the lectures will become.

Mathematics is learned by doing. You cannot learn mathematics by watching others, just as you cannot learn to ski or play tennis by watching competitions on television. You have to practice. Treat mathematics like a language: use it, speak it, and write it.

Exercises are the best teachers. The most effective way to learn is through solving problems. Work on as many exercises as you can, try to explain the solutions to yourself and to others, and attempt variations until you feel comfortable. Discussing the material with classmates is also extremely helpful: explaining a proof or a solution often reveals whether you have really understood it.

Collaboration vs. independence. Working in small groups is highly recommended. Discussion makes abstract ideas more concrete and easier to grasp. But remember: in the end you must be able to solve problems on your own. Always revisit exercises without help and check that you can reproduce the reasoning independently.

Asking questions is a strength. Never hesitate to ask questions. Many of your classmates will have the same doubts, and raising them helps everyone. It also provides valuable feedback to lecturers and assistants about which points need further explanation. Learning to formulate precise questions is itself an important skill, and the first year is the perfect time to practice it.

Chapter 2

The Real Numbers: Maximum, Supremum, and Sequences

2.1 The Axioms of the Real Numbers

2.1.1 Ordered Fields

We all have an intuitive feeling for what real numbers are, but our goal is to introduce them carefully and precisely. To get there, we first need a few building blocks.

2

Groups, Rings, and Fields

We begin with the notion of *group*. Loosely speaking, a group is a set equipped with an “operation” that satisfies a list of properties. For our purposes, it is enough to know that a **operation** is something that *takes two elements of a set and gives back a third element of the set*. We now specify the properties that the operation must satisfy so that we can speak of a group.

DEFINITION 2.1: GROUPS

A **group** is a non-empty set G together with a rule (called an *operation* and usually written as \star) that combines any two elements of G into another element of G . This operation must satisfy three conditions:

- **Associativity:** No matter how you place parentheses, the result is the same: for all $a, b, c \in G$,

$$(a \star b) \star c = a \star (b \star c).$$

- **Neutral element:** There is a special element $e \in G$ such that combining it with any $a \in G$ leaves a unchanged: for any $a \in G$,

$$a \star e = e \star a = a.$$

- **Inverse element:** Every $a \in G$ has a “partner” $a^{-1} \in G$ that “cancels it out,” giving the neutral element: for any $a \in G$ there exists $a^{-1} \in G$ such that

$$a \star a^{-1} = a^{-1} \star a = e.$$

Note that, in general, one does not require that $a \star b = b \star a$. If the order of the operation does not matter, i.e. $a \star b = b \star a$ for all $a, b \in G$, the group is called **commutative** or **abelian**.

2

EXAMPLE 2.2. — Let us check some familiar sets and operations.

1. *Natural numbers with addition.* Take $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ with the addition $+$.

- Addition is associative: $(k + l) + m = k + (l + m)$ for all $k, l, m \in \mathbb{N}$.
- The number 0 is neutral: $0 + n = n + 0 = n$.
- However, no natural number except 0 has an inverse in \mathbb{N} . For example, the inverse of 3 is -3 , which is not a natural number.

So $(\mathbb{N}, +)$ is not a group.

2. *Integers with addition.* Now take $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ with addition $+$.

- Addition is associative.
- 0 is the neutral element.
- Every integer has an inverse: the inverse of n is $-n$.

Thus $(\mathbb{Z}, +)$ is a group. It is also commutative since $n + m = m + n$.

3. *Nonzero rational numbers with multiplication.* Consider

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

with the multiplication \cdot . One can easily check that the multiplication is associative and commutative, the neutral element is 1, and the inverse of $\frac{p}{q}$ is $\frac{q}{p}$. Also, $\frac{p}{q} \cdot \frac{r}{s} = \frac{r}{s} \cdot \frac{p}{q}$. Hence (\mathbb{Q}^*, \cdot) is a commutative group.

4. *Permutations of three objects.* Imagine we have three objects $\{1, 2, 3\}$. A *permutation* is just a way of rearranging them. There are 6 possible rearrangements and together they form the set S_3 , whose elements are the following:

id	(do nothing)
(1 2)	(swap 1 and 2, leave 3)
(1 3)	(swap 1 and 3, leave 2)
(2 3)	(swap 2 and 3, leave 1)
(1 2 3)	(send $1 \mapsto 2$, $2 \mapsto 3$, $3 \mapsto 1$)
(1 3 2)	(send $1 \mapsto 3$, $3 \mapsto 2$, $2 \mapsto 1$).

2

The operation in this group is *composition*: do one rearrangement after another. More explicitly, if σ and τ are two permutations, then $\sigma \circ \tau$ means “first apply τ , then apply σ ”. With this operation, one can check all the properties:

- Associativity holds because if you do three rearrangements in a row, it does not matter how you group them.
- The identity id is the neutral element.
- Every permutation can be undone, so each element has an inverse. For example, the inverse of (1 2 3) is (1 3 2), since doing one after the other brings everything back to the start.

Therefore (S_3, \circ) is a group, but in this case it is not commutative. For instance, let $\sigma = (1\ 2)$ and $\tau = (2\ 3)$.

- If we compute $\sigma \circ \tau$: τ sends $1 \mapsto 1$, $2 \mapsto 3$, $3 \mapsto 2$. Then σ swaps 1 and 2. So overall we get $1 \mapsto 2$, $2 \mapsto 3$, $3 \mapsto 1$, which is the permutation (1 2 3).

- If we compute $\tau \circ \sigma$: σ swaps 1 and 2. Then τ swaps 2 and 3. So overall we get $1 \mapsto 3$, $3 \mapsto 2$, $2 \mapsto 1$, which is the permutation (1 3 2).

Since $(1\ 2\ 3) \neq (1\ 3\ 2)$, we conclude that $\sigma \circ \tau \neq \tau \circ \sigma$. Hence S_3 is not commutative.

A couple of useful facts follow from the definition of group.

LEMMA 2.3: BASIC PROPERTIES OF GROUPS

Let G be a group. Then:

1. The neutral element is unique.
2. The inverse of an element is unique.
3. The inverse of the inverse of an element is the element itself, namely $(a^{-1})^{-1} = a$ for all $a \in G$.

Proof. (i) Assume that, in addition to $e \in G$, we have a second element e' with the property that $e' \star a = a \star e' = a$ for all elements $a \in G$. Then, we can choose $a = e$ to obtain

$$e \star e' = e.$$

Similarly, since e is a neutral element,

$$e \star e' = e'.$$

Combining the two identities, we get

$$e = e \star e' = e'.$$

This proves that $e = e'$, so we can speak of *the* neutral element of a group.

(ii) Assume that for an element $a \in G$, there exist two elements $b, c \in G$ that are both inverse of a , namely

$$a \star b = b \star a = e, \quad a \star c = c \star a = e.$$

Then, using associativity, we observe that

$$b = b \star e = b \star (a \star c) = (b \star a) \star c = e \star c = c.$$

This proves that the inverse element of a is unique, so we can speak of *the* inverse element, and the notation a^{-1} makes sense.

(iii) Since $a \star a^{-1} = e$, we deduce that a is the inverse of a^{-1} , thus

$$(a^{-1})^{-1} = a. \tag{2.1}$$

□

Groups capture the idea of combining elements with a single operation. But to describe the arithmetic of numbers more faithfully, we also need a second operation (as we do with addition and multiplication). This leads us to the notions of *rings* and *fields*.

DEFINITION 2.4: RINGS AND FIELDS

A **ring** is a non-empty set R in which we can both “add” and “multiply” elements with two operations “+” and “ \cdot ”. Also, these two operations are compatible with each other. More precisely:

- $(R, +)$ is a **commutative group**, with neutral element denoted by 0.
- Multiplication \cdot is **associative**, has a **neutral element** (usually written as 1), and **distributes over addition**:

$$a \cdot (b + c) = a \cdot b + a \cdot c, \quad (b + c) \cdot a = b \cdot a + c \cdot a \quad \text{for all } a, b, c \in R.$$

If multiplication is also commutative, we call R a **commutative ring**.

Note that, unlike addition, we do not require that every element has an inverse for multiplication.

A **field** is a special kind of commutative ring: every nonzero element has an inverse for the multiplication. In other words, if R is a commutative ring, then R is a field if $R \setminus \{0\}$ forms a commutative group under multiplication.

Traditionally, we use the letter F to denote a field. We also write $F^* = F \setminus \{0\}$ for the set of all invertible elements of F .

2

EXAMPLE 2.5. — Let us test these ideas on examples.

1. *The integers \mathbb{Z} .* We already know that $(\mathbb{Z}, +)$ is a commutative group. To see if it is a ring with the usual multiplication, we must check:

- Associativity of the multiplication: For all integers $k, l, m \in \mathbb{Z}$, we have

$$(k \cdot l) \cdot m = k \cdot (l \cdot m).$$

- Neutral element for the multiplication: The neutral element for the multiplication is $1 \in \mathbb{Z}$ as, for all integers $k \in \mathbb{Z}$, we have

$$1 \cdot k = k \cdot 1 = k.$$

- Distributivity: For all $k, l, m \in \mathbb{Z}$ we have

$$k \cdot (l + m) = k \cdot l + k \cdot m$$

and

$$(k + l) \cdot m = k \cdot m + l \cdot m.$$

Thus \mathbb{Z} is a ring. Moreover, since the multiplication is commutative (namely, $k \cdot l = l \cdot k$), it is a commutative ring.

However, \mathbb{Z} is not a field: most integers do not have a multiplicative inverse inside \mathbb{Z} . For example, the inverse of 2 would be $\frac{1}{2}$, which is not an integer.

2. *The rational numbers* \mathbb{Q} . With the usual addition and multiplication, \mathbb{Q} is a commutative ring. Moreover, every nonzero rational number $\frac{p}{q}$ has a multiplicative inverse $\frac{q}{p}$, which is also rational. Hence \mathbb{Q} is a field.

So, in short: rings generalize the integers, while fields generalize the rationals.

EXAMPLE 2.6. — *A non-commutative ring.* We consider the set $M_2(\mathbb{Z})$ of all 2×2 tables of integers (called “matrices”) usually written as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Addition and multiplication are defined as follows:

- Addition is done entry by entry:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} a + a' & b + b' \\ c + c' & d + d' \end{pmatrix}.$$

- Multiplication is slightly more involved:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} aa' + bc' & ab' + bd' \\ ca' + dc' & cb' + dd' \end{pmatrix}.$$

This structure satisfies the axioms of a ring with neutral elements $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ for addition and $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ for multiplication, but multiplication is *not* commutative. Let us check explicitly with

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

First compute AB :

$$AB = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 \cdot 0 + 1 \cdot 1 & 0 \cdot 0 + 1 \cdot 0 \\ 0 \cdot 0 + 0 \cdot 1 & 0 \cdot 0 + 0 \cdot 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Now compute BA :

$$BA = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 \cdot 0 + 0 \cdot 0 & 0 \cdot 1 + 0 \cdot 0 \\ 1 \cdot 0 + 0 \cdot 0 & 1 \cdot 1 + 0 \cdot 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

We see that

$$AB = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = BA.$$

Thus $M_2(\mathbb{Z})$ is a ring in which multiplication is not commutative.

Just as with groups, some useful facts follow immediately:

- The additive inverse of a is written $-a$, and the multiplicative inverse of $a \neq 0$ is written a^{-1} .
- In the current context, (2.1) applied to both addition and multiplication implies that

$$-(-a) = a, \quad (a^{-1})^{-1} = a \quad \text{whenever } a \neq 0. \quad (2.2)$$

Also, we have the following

LEMMA 2.7: BASIC PROPERTIES OF FIELDS

Let F be a field and let $a, b \in F$. Then:

1. $0 \cdot a = a \cdot 0 = 0$.
2. $a \cdot (-b) = -(a \cdot b) = (-a) \cdot b$. In particular, $(-1) \cdot a = -a$.
3. $(-a) \cdot (-b) = a \cdot b$. In particular, $(-a)^{-1} = -(a^{-1})$ whenever $a \neq 0$.

Proof. (i) Since 0 is the neutral element for the addition, we have $0 = 0 + 0$. Hence, using distributivity, we get

$$0 \cdot a = (0 + 0) \cdot a = (0 \cdot a) + (0 \cdot a).$$

Adding $-0 \cdot a$ (i.e., the inverse of $0 \cdot a$ for the addition), we deduce that $0 \cdot a = 0$. The case of $a \cdot 0$ is analogous.

(ii) By the distributive law,

$$a \cdot b + a \cdot (-b) = a \cdot (b + (-b)) = a \cdot 0 = 0.$$

So $a \cdot (-b)$ is the additive inverse of $a \cdot b$, i.e., $-(a \cdot b) = a \cdot (-b)$. Taking $b = 1$ gives $-a = (-1) \cdot a$.

The validity of $(-a) \cdot b = -(a \cdot b)$ follows exchanging a and b in the argument above.

(iii) By (ii) we know that $-(a \cdot b) = a \cdot (-b)$. Hence, recalling (2.2),

$$a \cdot b = -(a \cdot (-b)).$$

On the other hand, applying (ii) with $(-b)$ instead of b , we also have

$$-(a \cdot (-b)) = (-a) \cdot (-b).$$

Combining the two identities above, we conclude that $(-a) \cdot (-b) = a \cdot b$. Finally, taking $b = a^{-1}$ yields $(-a) \cdot (-(a^{-1})) = a \cdot a^{-1} = 1$, which gives the second assertion. \square

REMARK 2.8. — A natural question one may ask is the following: Can 0 (the additive neutral element) and 1 (the multiplicative neutral element) be the same? If $0 = 1$, then for any $a \in F$, we would have

$$a = a \cdot 1 = a \cdot 0 = 0.$$

So the field would collapse to just one element, namely $\{0\}$. From now on, we will always assume that a field has at least two distinct elements, therefore $0 \neq 1$.

Order relation

Next, we introduce the second ingredient of an ordered field: the *order relation*. To get there, we first need a few basic notions.

2

CARTESIAN PRODUCT

Let X and Y be two sets. The **cartesian product** $X \times Y$ is the set of ordered pairs of elements of X and Y :

$$X \times Y = \{(x, y) \mid x \in X, y \in Y\}.$$

EXAMPLE 2.9. — The cartesian product $X \times Y$ of $X = \{A, B, C, D, E, F, G, H\}$ and $Y = \{1, 2, 3, 4, 5, 6, 7, 8\}$ is what we use to describe positions on a chessboard. Each pair corresponds to a unique square. For instance, the black king starts the game on the square $(E, 8) \in X \times Y$.

SUBSETS

Let P and Q be sets.

- P is a **subset** of Q , written $P \subset Q$ (or $P \subseteq Q$), if every element of P also belongs to Q .
- P is a **proper subset** of Q , written $P \subsetneq Q$, if P is a subset of Q but $P \neq Q$.
- We write $P \not\subset Q$ (or $P \not\subseteq Q$) if P is not a subset of Q .

Equivalent ways of saying “ P is a subset of Q ” are “ P is contained in Q ” or “ Q is a superset of P ,” written $Q \supset P$. Likewise, “ Q is a proper superset of P ” is written $Q \supsetneq P$.

Because of the axioms of naive set theory, two sets P and Q are equal exactly when both $P \subset Q$ and $Q \subset P$ hold. For example, $\{x, y\} = \{z\}$ holds if $x = y = z$. Notice that multiplicities do not matter: $\{x, x, x\} = \{x\}$.

DEFINITION 2.10: RELATIONS

Let X be a set. A **relation** on X is a subset $\mathcal{R} \subset X \times X$, that is, a collection of ordered pairs of elements of X . If $(x, y) \in \mathcal{R}$ we write $x\mathcal{R}y$. Common symbols for relations include $<, \leq, \sim, \equiv, \cong$.

If \sim is a relation on X , we write $x \not\sim y$ if $x \sim y$ does not hold. A relation \sim may have the following properties:

1. **Reflexive:** $x \sim x$ for all $x \in X$.
2. **Transitive:** if $x \sim y$ and $y \sim z$, then $x \sim z$.
3. **Symmetric:** if $x \sim y$, then $y \sim x$.
4. **Antisymmetric:** if $x \sim y$ and $y \sim x$, then $x = y$.

A relation is an **equivalence relation** if it is reflexive, transitive, and symmetric. It is an **order relation** if it is reflexive, transitive, and antisymmetric.

2

EXAMPLE 2.11. — On the set of integers \mathbb{Z} , we consider three different relations.

- The usual relation \leq (“less than or equal to”). This relation satisfies the following properties:

1. Reflexive: $n \leq n$ for all n .
2. Transitive: if $n \leq m$ and $m \leq p$, then $n \leq p$.
3. Not symmetric: e.g. $7 \leq 8$ but $8 \not\leq 7$.
4. Antisymmetric: if $n \leq m$ and $m \leq n$, then $n = m$.

So \leq is an order relation.

- The relation $<$ (“strictly smaller than”).
 1. Not reflexive: no n satisfies $n < n$.
 2. Transitive: if $n < m$ and $m < p$, then $n < p$.
 3. Not symmetric: e.g. $3 < 5$ but $5 \not< 3$.
 4. Antisymmetric (this is subtle): there are no integers $n, m \in \mathbb{Z}$ that satisfy both $n < m$ and $m < n$. Hence, the condition of antisymmetry holds because there is nothing to check.

So $<$ is neither an equivalence relation nor an order relation.

- On \mathbb{Z} , define $m \equiv n \pmod{3}$ if $m - n$ is divisible by 3.
 1. Reflexive: for any m , $m - m = 0$ is divisible by 3, so $m \equiv m \pmod{3}$.
 2. Symmetric: if $m \equiv n \pmod{3}$, then $m - n$ is divisible by 3, hence also $n - m = -(m - n)$ is divisible by 3, so $n \equiv m \pmod{3}$.
 3. Transitive: if $m \equiv n \pmod{3}$ and $n \equiv p \pmod{3}$, then both $m - n$ and $n - p$ are divisible by 3, hence so is $(m - n) + (n - p) = m - p$, so $m \equiv p \pmod{3}$.

Therefore \equiv is an equivalence relation (this relation is called “congruence modulo 3” as is usually written as “ $a \equiv b \pmod{3}$ ”).

Ordered Fields

We now introduce the concept of an ordered field. From now on, we use the arrow “ \implies ” to denote implication: $A \implies B$ means “ A implies B ”.

DEFINITION 2.12: ORDERED FIELD

Let F be a field, and let \leq be an order relation on F . We call (F, \leq) , or simply F , an **ordered field** if the following hold:

1. **Linearity of order:** for all $x, y \in F$, at least one of $x \leq y$ or $y \leq x$ holds.
2. **Compatibility with addition:** for all $x, y, z \in F$,

$$x \leq y \implies x + z \leq y + z.$$

3. **Compatibility with multiplication:** for all $x, y \in F$,

$$0 \leq x \text{ and } 0 \leq y \implies 0 \leq x \cdot y.$$

The following terminology is standard and will be used throughout:

- $x \leq y$ is read as “ x is less than or equal to y .”
- $y \geq x$ means $x \leq y$ (“ y is greater than or equal to x ”).
- $x < y$ means $x \leq y$ and $x \neq y$ (“ x is strictly smaller than y ”).
- $x > y$ means $y < x$ (“ x is strictly greater than y ”).
- An element $x \in F$ is *non-negative* if $x \geq 0$, and *non-positive* if $x \leq 0$.
- An element $x \in F$ is *positive* if $x > 0$, and *negative* if $x < 0$.

We often chain inequalities, e.g.

$$x \leq y < z = a$$

stands for “ $x \leq y$, and $y < z$, and $z = a$.”

EXAMPLE 2.13. — A standard example of an ordered field is the field of rational numbers \mathbb{Q} with the usual order. For two fractions $\frac{p}{q}$ and $\frac{p'}{q'}$ (with $p, p' \in \mathbb{Z}$ and $q, q' \in \mathbb{N}$), we define

$$\frac{p}{q} \leq \frac{p'}{q'} \quad \text{if} \quad pq' \leq p'q,$$

where the inequality on the right is the usual order on the integers. It is easy to check that this order is compatible with addition and multiplication, so (\mathbb{Q}, \leq) is indeed an ordered field.

Given (F, \leq) an ordered field, we want to prove a series of properties that follow from the definitions. To simplify the notation, it is customary to write \cdot for multiplication only if it would otherwise be confusing. This is why, in proofs, \cdot may disappear. For example, we may write xy instead of $x \cdot y$.

LEMMA 2.14: ORDERED FIELD: BASIC CONSEQUENCES

Let (F, \leq) be an ordered field, and let $x, y, z, w \in F$. Then:

- (a) (Trichotomy) Either $x < y$, or $x = y$, or $x > y$.
- (b) If $x < y$ and $y \leq z$, then $x < z$. (Analogously, $x \leq y$ and $y < z$ imply $x < z$.)
- (c) (Addition of inequalities) If $x \leq y$ and $z \leq w$, then $x + z \leq y + w$. (Analogously, $x < y$ and $z \leq w$ imply $x + z < y + w$.)
- (d) $x \leq y$ if and only if $0 \leq y - x$.
- (e) $x \leq 0$ if and only if $0 \leq -x$.
- (f) $x^2 \geq 0$, and $x^2 > 0$ if $x \neq 0$.
- (g) $0 < 1$.
- (h) If $0 \leq x$ and $y \leq z$, then $xy \leq xz$.
- (i) If $x \leq 0$ and $y \leq z$, then $xy \geq xz$.
- (j) If $0 < x \leq y$, then $0 < y^{-1} \leq x^{-1}$.
- (k) If $0 \leq x \leq y$ and $0 \leq z \leq w$, then $0 \leq xz \leq yw$.
- (l) If $x + y \leq x + z$, then $y \leq z$.
- (m) If $xy \leq xz$ and $x > 0$, then $y \leq z$.

Proof. (a) By linearity of the order, for any x, y at least one of $x \leq y$ or $y \leq x$ holds. If $x \leq y$ and $y \leq x$, antisymmetry gives $x = y$. Otherwise, exactly one strict inequality holds, giving $x < y$ or $y < x$.

(b) From $x < y$ we have in particular $x \leq y$. Since $y \leq z$, transitivity yields $x \leq z$. So, to prove $x < z$, we need to exclude that $x = z$.

To see that, assume by contradiction that $x = z$. Then $y \leq z$ yields $y \leq x$, which contradicts $x < y$. Hence $x < z$.

The variant in parentheses is analogous.

(c) From $x \leq y$, compatibility with addition (Definition 2.12) gives $x + z \leq y + z$. From $z \leq w$ we get $y + z \leq y + w$. Transitivity yields $x + z \leq y + w$.

For the strict variant, use (b).

(d) If $x \leq y$, add $-x$ to both sides to obtain $0 \leq y - x$. Conversely, if $0 \leq y - x$, add x to get $x \leq y$.

(e) Apply (d) with $y = 0$.

(f) If $x \geq 0$, then $x^2 \geq 0$ by multiplicative compatibility in Definition 2.12. If $x \leq 0$, then $-x \geq 0$ by (e), hence $x^2 = (-x)^2 \geq 0$; here $x^2 = (-x)^2$ follows from Lemma 2.7 (iii).

For strict positivity: if $x \neq 0$ and $x^2 = 0$, multiply by x^{-1} and use Lemma 2.7(i) to get $0 = 0 \cdot x^{-1} = x^2 x^{-1} = x \cdot x \cdot x^{-1} = x$, a contradiction.

(g) By (f), $1 = 1^2 \geq 0$, and by Remark 2.8, $1 \neq 0$. Hence $0 < 1$.

(h) From $y \leq z$ and (d), we have $0 \leq z - y$. If also $0 \leq x$, multiplicative compatibility gives $0 \leq x(z - y) = xz - xy$. Apply again (d) to conclude that $xy \leq xz$.

(i) From $x \leq 0$ and (e) we get $-x \geq 0$, and from $y \leq z$ and (d) we get $0 \leq z - y$. Then $0 \leq (-x)(z - y) = xy - xz$, so by (d) we have $xy \geq xz$.

(j) We first assert that $x^{-1} > 0$: if by contradiction $x^{-1} \leq 0$, then (h) with $x \geq 0$ and $x^{-1} \leq 0$ would give $1 = xx^{-1} \leq 0$, contradicting (g).

Similarly, $y^{-1} > 0$.

3

Now, by (h), $x^{-1} \geq 0$ and $y^{-1} > 0$ implies that $x^{-1}y^{-1} \geq 0$. Combining this with $x \leq y$ and (h), we get

$$y^{-1} = x x^{-1} y^{-1} \leq y x^{-1} y^{-1} = x^{-1}$$

as desired.

(k)–(m) These follow by combining (h)–(i) with (c) and (d). Detailed arguments are assigned to Exercise 2.15. \square

EXERCISE 2.15. — Prove inferences (k)–(m) in Lemma 2.14.

What happens in (m) if the assumption $x > 0$ is replaced with $x < 0$? For each of these inferences, also try to formulate and prove analogous statements for the strict relation “ $<$ ”.

Now that we have a notion of an ordered field, we can construct the integers and rational numbers.

LEMMA 2.16: INTEGERS AND RATIONALS INSIDE AN ORDERED FIELD

Let (F, \leq) be an ordered field, and denote by 0 and 1 the neutral elements for addition and multiplication, respectively. Then:

(i) The elements $\dots, -2, -1, 0, 1, 2, \dots$ defined by

$$2 = 1 + 1, \quad 3 = 2 + 1, \quad \dots, \quad -n = (-1) \cdot n$$

are all distinct and satisfy

$$\dots < -2 < -1 < 0 < 1 < 2 < 3 < \dots$$

We denote this set of elements by \mathbb{Z} , and we call them “integers”.

(ii) Every fraction pq^{-1} with $p, q \in \mathbb{Z}$, $q \neq 0$, lies in F and the set of all such elements is denoted by \mathbb{Q} . Also,

$$\mathbb{Z} \subsetneq \mathbb{Q} \subseteq F.$$

Proof. (i) By Lemma 2.14(g), we have $0 < 1$. Then Lemma 2.14(c) yields $0 < 1 < 2 < 3 < \dots$, and taking negatives gives $\dots < -2 < -1 < 0$. Hence all these elements are distinct.

(ii) For $q \neq 0$, q is invertible in F ; define $\frac{p}{q} = pq^{-1}$. The set of such fractions is a field contained in F , which we denote by \mathbb{Q} .

To show that \mathbb{Q} strictly contains \mathbb{Z} , consider $\frac{1}{2}$ (the inverse of 2). Since $2 > 1$, it follows from Lemma 2.14(j) that $0 < \frac{1}{2} < 1$, so $\frac{1}{2} \notin \mathbb{Z}$. \square

EXERCISE 2.17. — Show that

$$\left\{x \in \mathbb{R} \setminus \{0\} \mid x + \frac{3}{x} + 4 \geq 0\right\} = \{x \in \mathbb{R} \setminus \{0\} \mid -3 \leq x \leq -1 \text{ or } x > 0\}.$$

Hint: note that $x + \frac{3}{x} + 4 = \frac{(x+3)(x+1)}{x}$.

FUNCTIONS

A **function** f from a set X to a set Y is an assignment of an element of Y to each element of X . The element $y \in Y$ to which $x \in X$ is assigned is denoted $f(x)$. We write $f : X \rightarrow Y$ for a function from X to Y and sometimes also speak of a **map**, **mapping**, or **transformation**.

The set X is the **domain** and Y the **codomain**.

We refer to the set X as **domain**, and the set Y as **domain of values** or **codomain**.

The set

$$\{(x, f(x)) \mid x \in X\} \subset X \times Y$$

is called the **graph** of f .

In the context of a function $f : X \rightarrow Y$, an element x of the domain of definition is also called **argument**, and an element $y = f(x) \in Y$ assumed by the function is also called **value** of the function.

If $f : X \rightarrow Y$ is a function, one also writes

$$\begin{aligned} f : X &\rightarrow Y \\ x &\mapsto f(x), \end{aligned}$$

where $f(x)$ could be a concrete formula. We pronounce “ \mapsto ” as “is mapped to”.

Two functions $f_1 : X_1 \rightarrow Y_1$ and $f_2 : X_2 \rightarrow Y_2$ are said to be equal if $X_1 = X_2$, $Y_1 = Y_2$, and $f_1(x) = f_2(x)$ for all $x \in X_1$.

DEFINITION 2.18: ABSOLUTE VALUE AND SIGN

Let (F, \leq) be an ordered field.

- The **absolute value** (or **modulus**) is the function $|\cdot| : F \rightarrow F$ defined by

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

- The **sign** is the function $\text{sgn} : F \rightarrow \{-1, 0, 1\}$ defined by

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

LEMMA 2.19: ABSOLUTE VALUE AND SIGN: BASIC PROPERTIES

Let (F, \leq) be an ordered field and let $x, y \in F$. Then:

- (a) $x = \operatorname{sgn}(x)|x|$, $|-x| = |x|$, $\operatorname{sgn}(-x) = -\operatorname{sgn}(x)$.
- (b) $|x| \geq 0$, and $|x| = 0$ if and only if $x = 0$ (by trichotomy, Lemma 2.14 a).
- (c) (Multiplicativity) $\operatorname{sgn}(xy) = \operatorname{sgn}(x)\operatorname{sgn}(y)$ and $|xy| = |x||y|$.
- (d) If $x \neq 0$, then $|x^{-1}| = |x|^{-1}$.
- (e) $|x| \leq y$ iff $-y \leq x \leq y$.
- (f) $|x| < y$ iff $-y < x < y$.
- (g) (Triangle inequality) $|x + y| \leq |x| + |y|$.
- (h) (Inverse triangle inequality) $||x| - |y|| \leq |x - y|$.

Proof. (a) This follows directly from the definition.

(b) This follows from the definition and Lemma 2.14(a).

(c) Check the four sign cases (depending on whether x, y are positive or negative).

(d) This follows from (c), because $|x^{-1}||x| = |1| = 1$.

(e) We consider the case $x \geq 0$, so that $|x| = x$ (the case $x \leq 0$ is similar).

If $|x| \leq y$, since $x = |x| \geq 0$ we get

$$-y \leq 0 \leq x \leq y.$$

Conversely, if $-y \leq x \leq y$, since $|x| = x$ we get $|x| \leq y$.

(f) This is proved similarly to (e), with strict inequalities.

(g) Thanks to (e) we have $-|x| \leq x \leq |x|$ and $-|y| \leq y \leq |y|$. Adding these two inequalities, we get

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

Applying (e) again, we conclude.

(h) From (g) we have $|x| \leq |x - y| + |y|$, therefore

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

Exchanging the roles of x and y we also have $|y| - |x| \leq |y - x| = |x - y|$. Combining these two inequalities yields

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

and the result follows by applying (e). □

EXERCISE 2.20. — For which $x, y \in \mathbb{R}$ does equality hold in the triangle inequality? And in the inverse triangle inequality?

2.1.2 Completeness Axiom

An arbitrary ordered field is not enough for calculus: it may have “gaps”. The basic example is the ordered field of rationals \mathbb{Q} . Consider

$$X = \{q \in \mathbb{Q} : \text{either } q < 0, \text{ or } q \geq 0 \text{ and } q^2 < 2\}, \quad Y = \{q \in \mathbb{Q} : q \geq 0 \text{ and } q^2 > 2\}.$$

Then X and Y are non-empty, and every $x \leq y$ for every $x \in X$ and $y \in Y$. However, there is no rational number c with $x \leq c \leq y$ for all such x, y (since there is no $c \in \mathbb{Q}$ with $c^2 = 2$). This “gap” shows that \mathbb{Q} fails the completeness property below. To rule out such gaps we add the *completeness axiom*. The need for such an axiom was already felt by the ancient Greeks (Pythagoras, Euclid, Archimedes), but a precise formalisation came only in the 19th century through the work of Weierstrass, Heine, Cantor, Dedekind, and others.

DEFINITION 2.21: COMPLETENESS AXIOM

Let (K, \leq) be an ordered field. We say that (K, \leq) is **complete** (or a **completely ordered field**) if the following statement holds:

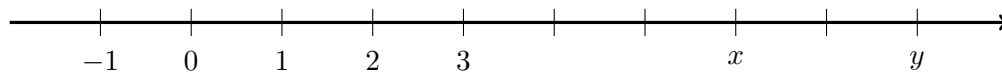
Let X, Y be non-empty subsets of K such that $x \leq y$ for all $x \in X$ and $y \in Y$. Then there exists $c \in K$ lying between X and Y , in the sense that $x \leq c \leq y$ for all $x \in X$ and $y \in Y$.

The statement above is called the **completeness axiom**.

DEFINITION 2.22: REAL NUMBERS

We call **the field of real numbers** any completely ordered field and denote it by \mathbb{R} .

We will often visualise the real numbers as points on a straight line, also called the **number line**.



Given $x, y \in \mathbb{R}$, we interpret the relation $x < y$ as “on the straight line, the point y lies to the right of the point x ”. With this representation of the real numbers, the following figure represents the completeness axiom.

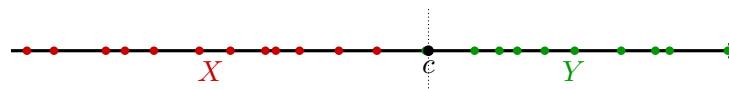


Figure 2.1: Given $X, Y \subset \mathbb{R}$ with $x \leq y$ for all $x \in X$ and $y \in Y$, there exists a number c in between.

We note that while the number line is a helpful aid to build intuition, it should not replace rigorous proofs.

INJECTIVE, SURJECTIVE AND BIJECTIVE FUNCTIONS

Let $f : X \rightarrow Y$ be a function. We call f :

1. **injective** (or an **injection**) if $x_1 \neq x_2 \implies f(x_1) \neq f(x_2)$ for all $x_1, x_2 \in X$;
2. **surjective** (or a **surjection**) if for every $y \in Y$ there exists $x \in X$ with $f(x) = y$;
3. **bijective** (or a **bijection**) if it is both injective and surjective.

Thus, a function $f : X \rightarrow Y$ is *not* injective if there exist distinct $x_1 \neq x_2 \in X$ with $f(x_1) = f(x_2)$, and *not* surjective if there exists $y \in Y$ such that $f(x) \neq y$ for all $x \in X$.

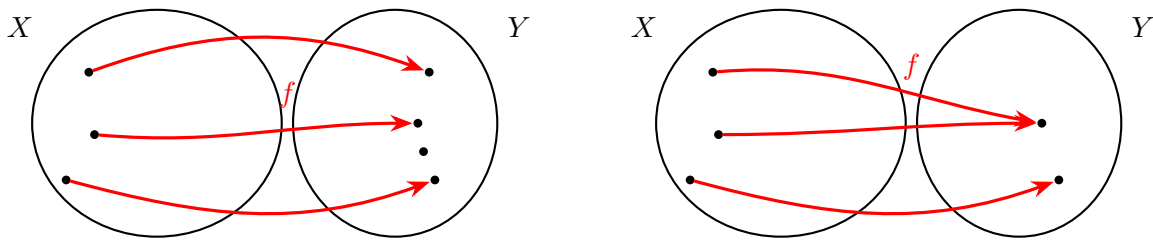


Figure 2.2: Left: injective but not surjective. Right: surjective but not injective.

IMAGE AND PREIMAGE OF A FUNCTION

For $f : X \rightarrow Y$ and $A \subset X$, define the **image** of A under the function f as

$$f(A) = \{ y \in Y \mid \exists x \in A \text{ with } f(x) = y \}.$$

For $B \subset Y$ define the **preimage** of B under the function f as

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

REMARK 2.23. — Saying that $f : X \rightarrow Y$ is surjective is equivalent to $f(X) = Y$. Equivalently, f is surjective if $f^{-1}(\{y\}) \neq \emptyset$ for every $y \in Y$.

EXAMPLE 2.24. — Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be the constant function $x \mapsto 0$. Then $f^{-1}(\{0\}) = \mathbb{R}$, while $f^{-1}(\{y\}) = \emptyset$ for every $y \neq 0$.

EXAMPLE 2.25. — Let X, Y be finite sets with the same number of elements. Then for any $f : X \rightarrow Y$, injectivity and surjectivity are equivalent.

Proof. Suppose X and Y both have n elements, and list $X = \{x_1, \dots, x_n\}$. If f is injective, the n values $f(x_i)$ are all distinct, so $f(X) = \{f(x_1), \dots, f(x_n)\}$ has n elements. Since $f(X) \subset Y$ and $|Y| = n$, we must have $f(X) = Y$, i.e. f is surjective.

Conversely, to show that surjectivity implies injectivity, we prove that if f is not injective then f is not surjective. So, assume there exist $x_i \neq x_j$ with $f(x_i) = f(x_j)$. Then $f(X)$ has at most $n - 1$ elements, so it cannot coincide with Y and, therefore, f cannot be surjective.

REMARK 2.26. — For infinite sets, injectivity and surjectivity need not be equivalent. Consider $f_1, f_2 : \mathbb{N} \rightarrow \mathbb{N}$ defined by

$$f_1(n) = n + 1, \quad f_2(n) = \begin{cases} 0, & n = 0, \\ n - 1, & n \geq 1. \end{cases}$$

Then f_1 is injective but not surjective, while f_2 is surjective but not injective.

EXERCISE 2.27. — Reformulate the definitions of injectivity, surjectivity, and bijectivity using the notions of image and preimage.

4

We conclude with the **square root function** on $\mathbb{R}_{\geq 0} = \{x \in \mathbb{R} : x \geq 0\}$ as an application of completeness.

EXERCISE 2.28. — Show the existence and uniqueness of a bijective function $\sqrt{\cdot} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ with $(\sqrt{a})^2 = a$ for all $a \in \mathbb{R}_{\geq 0}$.

1. (Existence) For $a \in \mathbb{R}_{\geq 0}$ set

$$X = \{x \in \mathbb{R}_{\geq 0} \mid x^2 \leq a\}, \quad Y = \{y \in \mathbb{R}_{\geq 0} \mid y^2 \geq a\}.$$

By completeness, there exists $c \in \mathbb{R}$ with $x \leq c \leq y$ for all $x \in X, y \in Y$. Show that $c \in X \cap Y$, hence $c^2 = a$.

Hint: If by contradiction $c \notin X$ (i.e. $c^2 > a$), choose small $\varepsilon > 0$ so that $(c - \varepsilon)^2 \geq a$. For instance, one can choose $\varepsilon = \frac{c^2 - a}{2c}$ and note that, with this choice,

$$(c - \varepsilon)^2 = c^2 - 2c\varepsilon + \varepsilon^2 > c^2 - 2c\varepsilon = c^2 - 2c \cdot \frac{c^2 - a}{2c} = a.$$

Hence $c - \varepsilon \in Y$, contradicting $y \geq c$ for all $y \in Y$.

The case $c \notin Y$ is analogous.

2. (Uniqueness) Prove that for every $a \in \mathbb{R}_{\geq 0}$ there is at most one $c \in \mathbb{R}_{\geq 0}$ with $c^2 = a$.
Hint: First, prove that for all $x, y \in \mathbb{R}_{\geq 0}$, the statements $x < y$ and $x^2 < y^2$ are equivalent. Then use this fact to deduce the uniqueness of c .

Define $\sqrt{\cdot} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ by $\sqrt{a} = c$ where c is as above. Show that:

3. $\sqrt{\cdot}$ is increasing: if $0 \leq x < y$ then $\sqrt{x} < \sqrt{y}$;
4. $\sqrt{\cdot}$ is bijective;
5. $\sqrt{xy} = \sqrt{x} \sqrt{y}$ for all $x, y \in \mathbb{R}_{\geq 0}$.

EXERCISE 2.29. — For all $x \in \mathbb{R}$, show that $x^2 = |x|^2$ and $\sqrt{x^2} = |x|$.

4

Working rules. In a field of real numbers as in Definition 2.22, the usual arithmetic rules and equation manipulations hold (as always, division by 0 is undefined). The order relations \leq and $<$ satisfy the familiar laws for inequalities; in particular, multiplying an inequality by a negative number reverses its direction. We will use these laws freely from now on. The new ingredient is the *completeness axiom*, whose power will become clear as we apply it to prove further results.

Do such numbers actually exist? Why do we say *the* real numbers? At this point it is not obvious that a completely ordered field really exists. Moreover, we often speak of *the* real numbers as if there were only one such object. In this course we will *assume* (in line with your school experience) that a field of real numbers exists and is unique. In particular, you may safely use \mathbb{R} with its usual arithmetic, order rules, and completeness, and you can think of \mathbb{R} as the rational line with all gaps filled.

For those interested, the following section provides a brief and concrete explanation of why this assumption is reasonable, with complete proofs reserved for later classes.

Extra material: Why the real numbers exist and are unique

In what follows, we sketch (without all details) the construction of \mathbb{R} via Dedekind cuts and the uniqueness argument for complete ordered fields.

Existence. Start from the rationals \mathbb{Q} . Many real numbers are not rational, but we can still “identify them” by stating which rationals lie to their left. This leads to the idea of *Dedekind cuts*.

A Dedekind cut is a set $C \subset \mathbb{Q}$ with the following properties:

- C is nonempty and is not equal to \mathbb{Q} ;
- C is *downward closed*: if $q \in C$ and $r < q$, then $r \in C$;
- C has no largest element: for every $q \in C$ there exists $r \in C$ with $q < r$.

Intuitively, C collects all rationals “to the left” of a real number. For example, the *cut of $\sqrt{2}$* is

$$\{q \in \mathbb{Q} \mid \text{either } q < 0, \text{ or } q \geq 0 \text{ and } q^2 < 2\}.$$

Define \mathbb{R}_{Ded} to be the set of all Dedekind cuts with the following addition and multiplication:

$$C + D = \{q + r \mid q \in C, r \in D\}, \quad C \cdot D = \{q \cdot r \mid q \in C, r \in D\}.$$

Also, define the order by inclusion: we say that $C \leq D$ if $C \subseteq D$.

With these definitions, one can show that \mathbb{R}_{Ded} is an ordered field and, crucially, that it satisfies the completeness axiom (V). In this construction, the rationals can be seen as a subset of \mathbb{R}_{Ded} via the “natural” cuts

$$\mathbb{Q} \ni r \longmapsto \{q \in \mathbb{Q} \mid q < r\} \in \mathbb{R}_{\text{Ded}}.$$

Thus, at an intuitive level, *real numbers exist*: they can be realized as precise “cuts” of the rational line filling all the “gaps”.

Uniqueness. Suppose K and L are two completely ordered fields, and let $0_K, 1_K$ and $0_L, 1_L$ denote their neutral elements for addition and multiplication. We identify 0_K with 0_L and 1_K with 1_L . In this way, as a consequence of how the rationals are constructed inside any ordered field (see Lemma 2.16), the rationals in K and in L are “the same” (i.e. the rational q in K corresponds to the same q in L). In particular, we may regard a single set of rationals \mathbb{Q} as contained in both K and L .

Now, for $x \in K$, consider the sets

$$X_x = \{q \in \mathbb{Q} \mid q < x \text{ in } K\} \subset \mathbb{Q}, \quad Y_x = \{q \in \mathbb{Q} \mid q > x \text{ in } K\} \subset \mathbb{Q}.$$

These are nonempty, and we have $q \leq r$ for all $q \in X_x$ and $r \in Y_x$. Since $\mathbb{Q} \subset L$, both X_x and Y_x are also contained in L . By completeness of L , there exists $c \in L$ with

$$q \leq c \leq r \quad \text{for all } q \in X_x, r \in Y_x.$$

One then checks that

$$\{q \in \mathbb{Q} \mid q < c \text{ in } L\} = X_x,$$

so c is the unique element of L whose “cut of rationals to its left” is X_x .

This construction allows us to define a map from K to L that assigns to $x \in K$ the element $c \in L$ constructed above:

$$I_{K,L} : K \rightarrow L, \quad I_{K,L}(x) = c.$$

One checks that this map preserves addition, multiplication, and order, and fixes every rational (i.e. $I_{K,L}(q) = q$ for all $q \in \mathbb{Q}$).

Repeating the construction with K and L swapped gives a map $I_{L,K} : L \rightarrow K$, and one can check that

$$I_{L,K}(I_{K,L}(x)) = x \text{ for all } x \in K, \quad I_{K,L}(I_{L,K}(y)) = y \text{ for all } y \in L.$$

Thus $I_{K,L}$ and $I_{L,K}$ are inverses, so $I_{K,L} : K \rightarrow L$ is a bijection that “identifies” K and L .

In short, once the rationals are fixed inside a completely ordered field, every other element is *uniquely* determined by how it “cuts” the rationals. This justifies speaking of *the* real numbers and writing \mathbb{R} .

2.1.3 Intervals

DEFINITION 2.30: INTERVALS

Let $a, b \in \mathbb{R}$. We define:

- the **closed interval**

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

- the **open interval**

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

- the **half-open intervals**

$$[a, b) = \{x \in \mathbb{R} \mid a \leq x < b\} \quad \text{and} \quad (a, b] = \{x \in \mathbb{R} \mid a < x \leq b\};$$

- the **unbounded closed intervals**

$$[a, \infty) = \{x \in \mathbb{R} \mid a \leq x\} \quad \text{and} \quad (-\infty, b] = \{x \in \mathbb{R} \mid x \leq b\};$$

- the **unbounded open intervals**

$$(a, \infty) = \{x \in \mathbb{R} \mid a < x\} \quad \text{and} \quad (-\infty, b) = \{x \in \mathbb{R} \mid x < b\}.$$

The intervals $(a, b]$, $[a, b)$, and (a, b) are non-empty exactly when $a < b$, while $[a, b]$ is non-empty exactly when $a \leq b$ (with $[a, a] = \{a\}$ a *degenerate* interval). If an interval is non-empty and bounded, we call a its **left endpoint**, b its **right endpoint**, and $b - a$ its **length**. Intervals of the forms $[a, b]$, $(a, b]$, $[a, b)$, (a, b) are also called **bounded intervals**, to distinguish them from the unbounded ones.

Some texts write open and half-open intervals using inverted square brackets, e.g. $]a, b[$ instead of (a, b) . In these notes we always use round brackets for open endpoints.

SET OPERATIONS

Let P and Q be sets. The **intersection** $P \cap Q$, the **union** $P \cup Q$, the **relative complement** $P \setminus Q$, and the **symmetric difference** $P \Delta Q$ are defined by

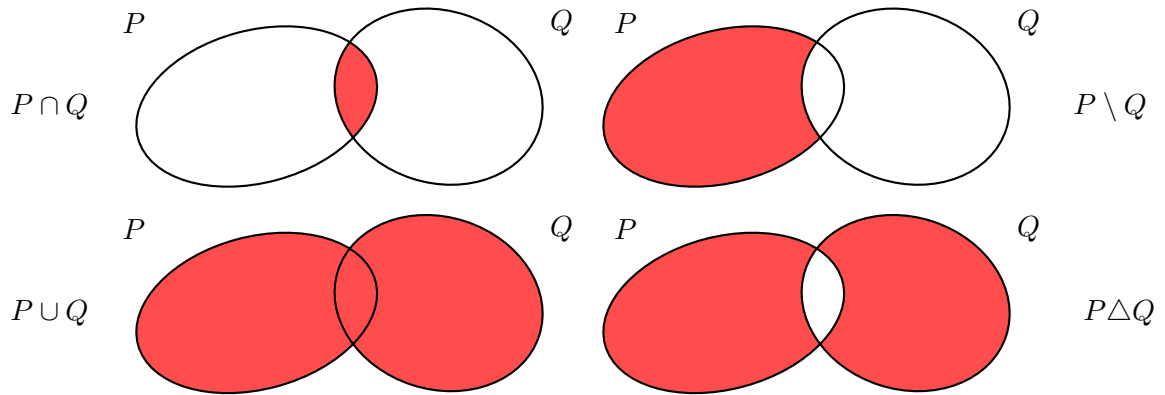
$$P \cap Q = \{x \mid x \in P \text{ and } x \in Q\},$$

$$P \cup Q = \{x \mid x \in P \text{ or } x \in Q\},$$

$$P \setminus Q = \{x \mid x \in P \text{ and } x \notin Q\},$$

$$P \Delta Q = (P \setminus Q) \cup (Q \setminus P) = (P \cup Q) \setminus (P \cap Q).$$

These definitions are illustrated in the following pictures. Sketches of this kind are called **Venn diagrams**.



If it is clear from the context that all sets under consideration are subsets of a given ambient set X , then the **complement** of P in X is denoted by P^c and defined as $P^c = X \setminus P$.

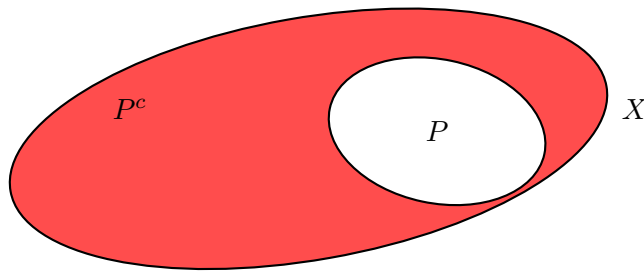


Figure 2.3: The complement $P^c = X \setminus P$ of P in X .

UNION AND INTERSECTION OF SEVERAL SETS

Let \mathcal{A} be a family of sets (i.e., a set whose elements are sets). We define the **union** and the **intersection** of the sets in \mathcal{A} as

$$\bigcup_{A \in \mathcal{A}} A = \{x \mid \exists A \in \mathcal{A} : x \in A\}, \quad \bigcap_{A \in \mathcal{A}} A = \{x \mid \forall A \in \mathcal{A} : x \in A\}.$$

If $\mathcal{A} = \{A_1, A_2, \dots\}$, we also write

$$\bigcup_{n=1}^{\infty} A_n = \{x \mid \exists n \geq 1 : x \in A_n\}, \quad \bigcap_{n=1}^{\infty} A_n = \{x \mid \forall n \geq 1 : x \in A_n\}.$$

EXAMPLE 2.31. — Let $\mathcal{A} = \{[x, \infty) : x \in \mathbb{R}\}$. Then $\bigcup_{A \in \mathcal{A}} A = \mathbb{R}$ and $\bigcap_{A \in \mathcal{A}} A = \emptyset$. Alternatively, consider $\mathcal{A} = \{[x, \infty) : x \in [0, 1]\}$. Then $\bigcup_{A \in \mathcal{A}} A = [0, \infty)$ and $\bigcap_{A \in \mathcal{A}} A = [1, \infty)$.

EXERCISE 2.32. — Let X be a set and let $\mathcal{A} \subseteq \mathcal{P}(X)$ be a collection of subsets of X . Show the *De Morgan laws* for unions and intersections of sets:

$$\left(\bigcup_{A \in \mathcal{A}} A \right)^c = \bigcap_{A \in \mathcal{A}} A^c, \quad \left(\bigcap_{A \in \mathcal{A}} A \right)^c = \bigcup_{A \in \mathcal{A}} A^c$$

- EXERCISE 2.33. — 1. Show that a finite intersection of intervals is again an interval. Describe the endpoints of a non-empty intersection in terms of the endpoints of the given intervals.
2. When is the union of two intervals an interval again? In that case, what can you say when you unite two intervals of the same type (open, closed, half-open)?

DEFINITION 2.34: NEIGHBOURHOODS

Let $x \in \mathbb{R}$. A **neighbourhood** of x is any set that contains an open interval I with $x \in I$. For $\delta > 0$, the open interval $(x - \delta, x + \delta)$ is called the **δ -neighbourhood** of x .

EXAMPLE 2.35. — Both $[-1, 1]$ and $\mathbb{Q} \cup [-1, 1]$ are neighbourhoods of $0 \in \mathbb{R}$ (they contain, say, $(-\frac{1}{2}, \frac{1}{2})$), while $[0, 1]$ is not a neighbourhood of 0.

Note that, for $\delta > 0$ and $x \in \mathbb{R}$,

$$(x - \delta, x + \delta) = \{y \in \mathbb{R} \mid |x - y| < \delta\},$$

so $|x - y|$ can be interpreted as the **distance** from x to y . In particular, since $|x - y| = |y - x|$ for all $x, y \in \mathbb{R}$ (see Lemma 2.19(a)), the distance from x to y is equal to the one from y to x .

DEFINITION 2.36: OPEN AND CLOSED SETS

A subset $U \subseteq \mathbb{R}$ is **open** if for every $x \in U$ there exists an open interval I with $x \in I \subseteq U$. A subset $C \subseteq \mathbb{R}$ is **closed** if its complement $\mathbb{R} \setminus C$ is open.

Open intervals are open, and closed intervals are closed. Intuitively, a set is open if every point x in the set comes with a small interval around x still lying in the set. Note that “open” is *not* the opposite of “closed”: the sets \emptyset and \mathbb{R} are both open and closed (their complements are \mathbb{R} and \emptyset , respectively). By contrast, $\mathbb{Q} \subset \mathbb{R}$ and $[a, b) \subset \mathbb{R}$ are neither open nor closed.

EXERCISE 2.37. — Show that $U \subseteq \mathbb{R}$ is open if and only if for every $x \in U$ there exists $\delta > 0$ such that $(x - \delta, x + \delta) \subseteq U$.

EXERCISE 2.38. — Let \mathcal{U} be a family of open subsets of \mathbb{R} , and let \mathcal{F} be a family of closed subsets of \mathbb{R} . Show that

$$\bigcup_{U \in \mathcal{U}} U \text{ is open,} \quad \text{and} \quad \bigcap_{C \in \mathcal{F}} C \text{ is closed.}$$

(*Optional.*) Prove also: finite intersections of open sets are open, and finite unions of closed sets are closed.

2.2 Complex Numbers

2.2.1 Definition of Complex Numbers

Starting from the field of real numbers \mathbb{R} , we define the set of **complex numbers** as

$$\mathbb{C} = \mathbb{R}^2 = \{(x, y) \mid x, y \in \mathbb{R}\}.$$

We denote elements $z = (x, y) \in \mathbb{C}$ in the form $z = x + iy$, where i is the **imaginary unit**. Here $x \in \mathbb{R}$ is the **real part** of z , written $x = \operatorname{Re}(z)$, and $y \in \mathbb{R}$ is the **imaginary part**, written $y = \operatorname{Im}(z)$. Elements with $\operatorname{Im}(z) = 0$ are called **real**, while those with $\operatorname{Re}(z) = 0$ are **purely imaginary**. Via the injective map $\mathbb{R} \ni x \mapsto x + i0 \in \mathbb{C}$, we identify \mathbb{R} with the subset of real numbers inside \mathbb{C} .

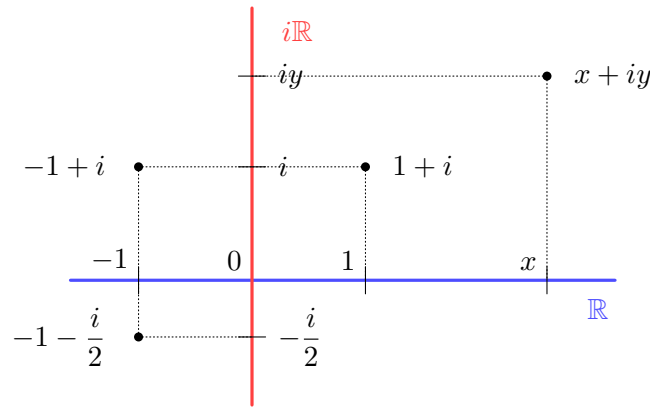


Figure 2.4: The graphical representation of \mathbb{C} is the **complex plane**. In this picture, the real numbers lie on the **real axis** and the purely imaginary numbers on the **imaginary axis**.

As you may expect from previous knowledge, we want i to satisfy $i^2 = -1$. To achieve this, we define addition and multiplication on \mathbb{C} so that it becomes a field. Additionally, we want these operations to coincide with the usual addition and multiplication when considering real numbers.

Since $i^2 = -1$, using commutativity and distributivity we get

$$(x_1 + iy_1)(x_2 + iy_2) = x_1x_2 + ix_1y_2 + iy_1x_2 + i^2y_1y_2 = (x_1x_2 - y_1y_2) + i(x_1y_2 + y_1x_2).$$

This motivates the following definition.

DEFINITION 2.39: ADDITION AND MULTIPLICATION ON \mathbb{C}

On $\mathbb{C} = \mathbb{R} \times \mathbb{R}$ we define **addition** and **multiplication** as follows:

$$\begin{aligned} (x_1, y_1) + (x_2, y_2) &= (x_1 + x_2, y_1 + y_2), \\ (x_1, y_1) \cdot (x_2, y_2) &= (x_1x_2 - y_1y_2, x_1y_2 + x_2y_1). \end{aligned}$$

PROPOSITION 2.40: \mathbb{C} IS A FIELD

With the operations of Definition 2.39, together with the zero element $(0, 0)$ and the unit element $(1, 0)$, the set \mathbb{C} is a field.

Proof. Additive properties follow immediately from the corresponding properties in \mathbb{R} : $(0, 0)$ is the additive identity and $(-x, -y)$ is the additive inverse of (x, y) .

We now check all the properties of multiplication by direct computation. We start from the associativity.

Given $(x_1, y_1), (x_2, y_2), (x_3, y_3) \in \mathbb{C}$, we compute

$$\begin{aligned} ((x_1, y_1) \cdot (x_2, y_2)) \cdot (x_3, y_3) &= (x_1x_2 - y_1y_2, x_1y_2 + y_1x_2) \cdot (x_3, y_3) \\ &= (x_1x_2x_3 - y_1y_2x_3 - x_1y_2y_3 - y_1x_2y_3, x_1y_2x_3 + y_1x_2x_3 + x_1x_2y_3 - y_1y_2y_3). \end{aligned}$$

Analogously, we have

$$\begin{aligned} (x_1, y_1) \cdot ((x_2, y_2) \cdot (x_3, y_3)) &= (x_1, y_1) \cdot (x_2x_3 - y_2y_3, x_2y_3 + y_2x_3) \\ &= (x_1x_2x_3 - y_1y_2x_3 - x_1y_2y_3 - y_1x_2y_3, x_1y_2x_3 + y_1x_2x_3 + x_1x_2y_3 - y_1y_2y_3). \end{aligned}$$

Since the two expressions on the right-hand side coincide, the multiplication is associative.

Commutativity is also easy to check: indeed,

$$(x_1, y_1) \cdot (x_2, y_2) = (x_1x_2 - y_1y_2, x_1y_2 + x_2y_1) = (x_2, y_2) \cdot (x_1, y_1).$$

The neutral element is $(1, 0)$ since

$$(x, y) \cdot (1, 0) = (x \cdot 1 - y \cdot 0, x \cdot 0 + y \cdot 1) = (x, y).$$

Distributivity is verified as follows:

$$\begin{aligned} (x_1, y_1) \cdot ((x_2, y_2) + (x_3, y_3)) &= (x_1, y_1) \cdot (x_2 + x_3, y_2 + y_3) \\ &= (x_1x_2 + x_1x_3 - y_1y_2 - y_1y_3, y_1x_2 + y_1x_3 + x_1y_2 + x_1y_3) \\ &= (x_1x_2 - y_1y_2, y_1x_2 + x_1y_2) + (x_1x_3 - y_1y_3, y_1x_3 + x_1y_3) \\ &= (x_1, y_1) \cdot (x_2, y_2) + (x_1, y_1) \cdot (x_3, y_3), \end{aligned}$$

Finally, it is immediate to check that the multiplication is commutative. Thus, \mathbb{C} is a commutative ring.

To conclude, given $(x, y) \neq (0, 0)$, one can directly verify that its multiplicative inverse is

$$\left(\frac{x}{x^2+y^2}, \frac{-y}{x^2+y^2} \right).$$

Indeed,

$$\begin{aligned} (x, y) \cdot \left(\frac{x}{x^2+y^2}, \frac{-y}{x^2+y^2} \right) &= \left(x \cdot \frac{x}{x^2+y^2} - y \cdot \frac{-y}{x^2+y^2}, y \cdot \frac{x}{x^2+y^2} + x \cdot \frac{-y}{x^2+y^2} \right) \\ &= \left(\frac{x^2+y^2}{x^2+y^2}, \frac{yx-xy}{x^2+y^2} \right) = (1, 0). \end{aligned}$$

This proves that \mathbb{C} is a field. \square

From now on we write $x + iy$ instead of (x, y) , and abbreviate $x + i0$ as x , $0 + iy$ as iy , and $i1$ simply as i . Then $i^2 = -1$, and \mathbb{R} naturally embeds into \mathbb{C} . For $z, w \in \mathbb{C}$ we write zw for their product. If $z \neq 0$, its multiplicative inverse can be denoted by z^{-1} or $\frac{1}{z}$. For example, $i^{-1} = \frac{1}{i} = -i$.

DEFINITION 2.41: COMPLEX CONJUGATION

For $z = x + iy \in \mathbb{C}$ we define its **conjugate** as $\bar{z} = x - iy$. The mapping $\mathbb{C} \ni z \mapsto \bar{z} \in \mathbb{C}$ is called **complex conjugation**.

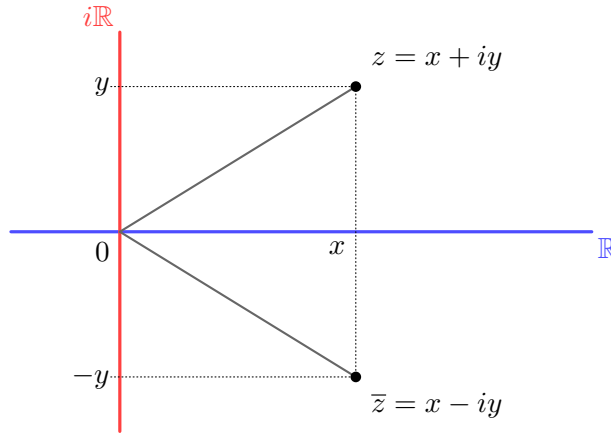


Figure 2.5: A complex number z and its conjugate \bar{z} .

LEMMA 2.42: PROPERTIES OF COMPLEX CONJUGATION

For all $z, w \in \mathbb{C}$:

- (i) $z\bar{z} = x^2 + y^2 \in \mathbb{R}_{\geq 0}$. In particular, $z\bar{z} = 0$ if and only if $z = 0$.
- (ii) $\overline{z + w} = \bar{z} + \bar{w}$.
- (iii) $\overline{zw} = \bar{z}\bar{w}$.

Proof. Property (i) follows from the fact that, for $z = x + iy$, $(x + iy)(x - iy) = x^2 + y^2$. Also, $x^2 + y^2 = 0$ if and only if $x + iy = 0$.

Properties (ii) and (iii) follow from a direct computation, writing $z = x_1 + iy_1$ and $w = x_2 + iy_2$:

$$\overline{z + w} = \overline{(x_1 + x_2) + i(y_1 + y_2)} = (x_1 + x_2) - i(y_1 + y_2) = (x_1 - iy_1) + (x_2 - iy_2) = \bar{z} + \bar{w},$$

$$\begin{aligned}\overline{z \cdot w} &= \overline{(x_1x_2 - y_1y_2) + i(x_1y_2 + y_1x_2)} = (x_1x_2 - y_1y_2) - i(x_1y_2 + y_1x_2) \\ &= (x_1 - iy_1) \cdot (x_2 - iy_2) = \overline{z} \cdot \overline{w}.\end{aligned}$$

□

EXERCISE 2.43. — Show that

$$\operatorname{Re}(z) = \frac{z + \overline{z}}{2}, \quad \operatorname{Im}(z) = \frac{z - \overline{z}}{2i},$$

and deduce that $\mathbb{R} = \{z \in \mathbb{C} \mid z = \overline{z}\}$. Interpret these identities geometrically.

Since $i^2 = -1 < 0$, property (f) in Lemma 2.19 implies that no order compatible with field operations can exist on \mathbb{C} . Nevertheless, calculus can be carried out on \mathbb{C} , and this will be studied in detail in the course on *complex analysis*. The reason is that \mathbb{C} satisfies a suitable extension of the completeness axiom, which you will meet in future courses.

2.2.2 The Absolute Value on the Complex Numbers

Since no ordering on \mathbb{C} makes it an ordered field, we cannot use Definition 2.18 to define an absolute value as we did on \mathbb{R} . Still, we want a notion that extends the real absolute value and preserves as many of its properties as possible. To this end, we use the root function from Exercise 2.28.

5

DEFINITION 2.44: ABSOLUTE VALUE

The **absolute value** (or **norm**) on \mathbb{C} is the map $|\cdot| : \mathbb{C} \rightarrow \mathbb{R}$ given by

$$|z| = \sqrt{z\overline{z}} = \sqrt{x^2 + y^2}, \quad z = x + iy \in \mathbb{C}.$$

For $x \in \mathbb{R}$, we have $\sqrt{x\overline{x}} = \sqrt{x^2} = |x|$, so the complex absolute value extends the real one from Definition 2.18 and the notation is consistent.

Note that $|z| \geq 0$ for all $z \in \mathbb{C}$, with equality if and only if $z = 0$ (see Lemma 2.42(i)). Also, the absolute value is multiplicative:

$$|zw| = \sqrt{zw\overline{zw}} = \sqrt{z\overline{z}}\sqrt{w\overline{w}} = |z||w| \quad \text{for all } z, w \in \mathbb{C}.$$

In particular, for $z \neq 0$,

$$z^{-1} = \frac{\overline{z}}{|z|^2}.$$

These properties follow directly from Lemma 2.42. Also, geometrically, $|z| = \sqrt{x^2 + y^2}$ is the Euclidean length of the segment going from 0 to z .

Next, we show that the absolute value satisfies the triangle inequality.

PROPOSITION 2.45: TRIANGLE INEQUALITY

For all $z, w \in \mathbb{C}$ one has

$$|z + w| \leq |z| + |w|.$$

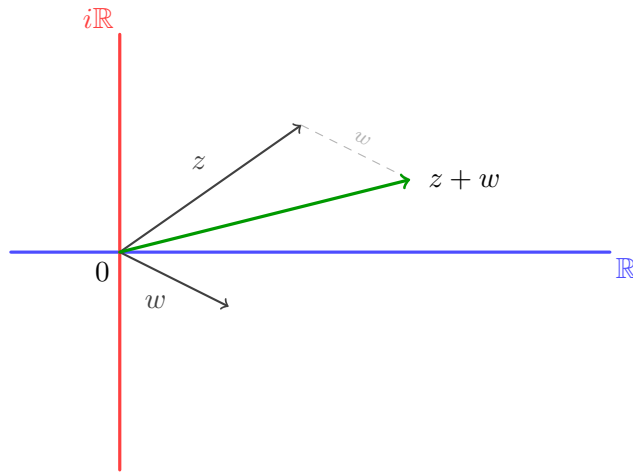


Figure 2.6: A visual representation of the triangle inequality: Placing w at the tip of z forms a triangle with third side $z + w$. The length of this side is at most the sum of the other two, namely, $|z + w| \leq |z| + |w|$.

To prove the proposition, we first need the following result:

LEMMA 2.46: CAUCHY-SCHWARZ INEQUALITY

If $z = x_1 + iy_1$ and $w = x_2 + iy_2$, then

$$x_1x_2 + y_1y_2 \leq |z||w|. \quad (2.3)$$

Proof. We observe that

$$\begin{aligned} |z|^2|w|^2 - (x_1x_2 + y_1y_2)^2 &= (x_1^2 + y_1^2)(x_2^2 + y_2^2) - (x_1x_2 + y_1y_2)^2 \\ &= x_1^2x_2^2 + y_1^2y_2^2 + y_1^2x_2^2 + x_1^2y_2^2 - (x_1^2x_2^2 + y_1^2y_2^2 + 2x_1x_2y_1y_2) \\ &= y_1^2x_2^2 + x_1^2y_2^2 - 2x_1x_2y_1y_2 \\ &= (y_1x_2 - x_1y_2)^2 \geq 0. \end{aligned}$$

This proves that $(x_1x_2 + y_1y_2)^2 \leq |z|^2|w|^2$, so by Exercise 2.29,

$$|x_1x_2 + y_1y_2| \leq |z||w|.$$

Since $x \leq |x|$ for all $x \in \mathbb{R}$, we obtain (2.3). \square

Proof of Proposition 2.45. For $z = x_1 + iy_1$ and $w = x_2 + iy_2$, using Lemma 2.46 we have

$$|z + w|^2 = (x_1 + x_2)^2 + (y_1 + y_2)^2$$

5

$$\begin{aligned}
&= |z|^2 + |w|^2 + 2(x_1x_2 + y_1y_2) \\
&\leq |z|^2 + |w|^2 + 2|z||w| = (|z| + |w|)^2.
\end{aligned}$$

Taking square roots proves the result. \square

Because $|z - w|$ represents the **distance** between z and w , this motivates the following definition:

DEFINITION 2.47: CIRCULAR DISKS

For $z \in \mathbb{C}$ and $r > 0$, we define the **open disk** with radius $r > 0$ around z as

$$B(z, r) = \{w \in \mathbb{C} \mid |z - w| < r\},$$

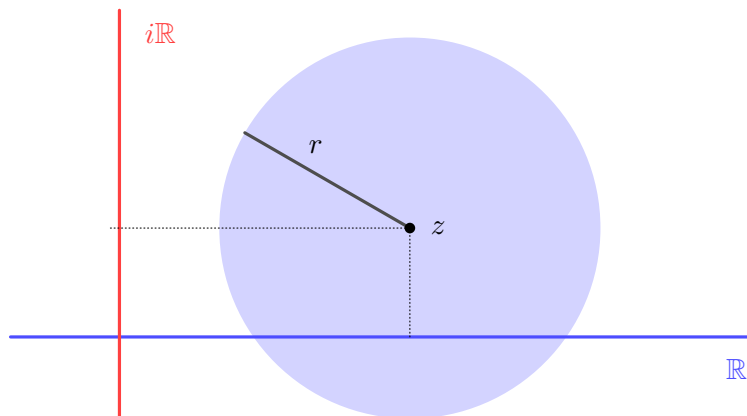
and the **closed disk** with radius $r > 0$ around z as

$$\overline{B(z, r)} = \{w \in \mathbb{C} \mid |z - w| \leq r\}.$$

In other words, the open disk $B(z, r)$ is the set of points at distance strictly less than r from z . We note that this definition is compatible with the one of neighborhood in \mathbb{R} : if $x \in \mathbb{R}$ and $r > 0$, then

$$B(x, r) \cap \mathbb{R} = (x - r, x + r).$$

6



EXERCISE 2.48. — Let $z_1, z_2 \in \mathbb{C}$ and $r_1, r_2 > 0$. Show that, for every $z \in B(z_1, r_1) \cap B(z_2, r_2)$, there exists $r > 0$ such that

$$B(z, r) \subseteq B(z_1, r_1) \cap B(z_2, r_2).$$

Illustrate your choice of r in a picture.

The following definition generalizes that in \mathbb{R} from Exercise 2.37.

DEFINITION 2.49: OPEN AND CLOSED SETS

A set $U \subseteq \mathbb{C}$ is **open** if for every $z \in U$ there exists $r > 0$ such that $B(z, r) \subseteq U$. A set $C \subseteq \mathbb{C}$ is **closed** if its complement $\mathbb{C} \setminus C$ is open.

In particular, by Exercise 2.48, every open disk is open. More generally, arbitrary unions of open sets are open. You will return to these notions in greater generality in your future courses.

2.3 Maximum and Supremum

2.3.1 Existence of the Supremum

DEFINITION 2.50: BOUNDED SETS, MAXIMA, AND MINIMA

Let $X \subseteq \mathbb{R}$ be a subset of the real numbers.

- X is **bounded from above** if there exists $s \in \mathbb{R}$ such that $x \leq s$ for all $x \in X$. Such a number s is called an **upper bound** of X . If s is an upper bound and also an element of X , we say that s is the **maximum** of X and write

$$s = \max(X).$$

- Analogously, X is **bounded from below** if there exists $r \in \mathbb{R}$ such that $r \leq x$ for all $x \in X$. Such a number r is called a **lower bound** of X . If r is a lower bound and also an element of X , we say that r is the **minimum** of X and write

$$r = \min(X).$$

- X is called **bounded** if it is both bounded from above and bounded from below.

REMARK 2.51. — If a set $X \subseteq \mathbb{R}$ has a maximum, then it is unique. Indeed, if $x_1, x_2 \in X$ are both maxima, then $x_1 \leq x_2$ (since x_2 is a maximum) and $x_2 \leq x_1$ (since x_1 is a maximum), so $x_1 = x_2$.

A closed interval $[a, b]$ with $a < b$ has both a minimum and a maximum: $a = \min([a, b])$ and $b = \max([a, b])$. But not all sets have a maximum. For instance, the open interval (a, b) does not have a maximum because the endpoint b , though an upper bound, is not contained in the set. Similarly, \mathbb{R} and unbounded intervals such as $[a, \infty)$ or (a, ∞) have no maximum.

DEFINITION 2.52: SUPREMUM

Let $X \subseteq \mathbb{R}$ be a subset and let

$$A := \{a \in \mathbb{R} \mid x \leq a \text{ for all } x \in X\}$$

be the set of all upper bounds of X . If A has a minimum, we call this minimum the **supremum** of X and write

$$\sup(X) = \min(A).$$

In other words, the supremum of X is the smallest real number that is greater than or equal to every element of X . Note that we can describe the supremum $s = \sup(X)$ as follows:

$$x \leq s \quad \text{for all } x \in X, \quad \text{and} \quad \text{if } t < s, \text{ then } t \text{ is not an upper bound of } X. \quad (2.4)$$

This means that for every $t < s$, there exists some $x \in X$ such that $x > t$. That is,

$$x \leq s \quad \text{for all } x \in X, \quad \text{and} \quad \forall t < s \exists x \in X \text{ such that } x > t. \quad (2.5)$$

The two characterizations (2.4) and (2.5) are equivalent.

REMARK 2.53. — If a set X has a maximum, then this element is also the supremum. Indeed, the maximum is an upper bound of X , and since it lies in X , no smaller upper bound can exist.

EXAMPLE 2.54. — Let

$$X := \left\{1 - \frac{1}{n} \mid n \geq 1\right\} = \left\{0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots\right\}.$$

Then X is a bounded subset of \mathbb{R} with $\sup(X) = 1$.

6

- 1 is an upper bound of X , since $1 - \frac{1}{n} < 1$ for all $n \geq 1$.
- No number smaller than 1 is an upper bound, because for any $t < 1$ we can find n large enough such that $1 - \frac{1}{n} > t$. (We will justify this fact rigorously later using the Archimedean Principle, see Corollary 2.60.)
- Therefore, $\sup(X) = 1$, even though $1 \notin X$, so X has no maximum.



Figure 2.7: The set $X = \left\{1 - \frac{1}{n} \mid n \geq 1\right\}$. The points get arbitrarily close to 1, but never reach it.

Note that not every set has a supremum. If $X = \emptyset$ or if X is unbounded from above, then $\sup(X)$ does not exist. However, for any nonempty and bounded-above subset of \mathbb{R} , the supremum always exists:

THEOREM 2.55: EXISTENCE OF THE SUPREMUM

Let $X \subset \mathbb{R}$ be nonempty and bounded from above. Then $\sup(X)$ exists and is a real number.

Proof. Since X is bounded from above, the set $A := \{a \in \mathbb{R} \mid x \leq a \text{ for all } x \in X\}$ of upper bounds is nonempty. Since $x \leq a$ for any $x \in X$ and $a \in A$, we can apply the completeness

axiom (Definition 2.21) to find $c \in \mathbb{R}$ such that

$$x \leq c \leq a \quad \text{for all } x \in X \text{ and } a \in A.$$

The first inequality implies that c is itself an upper bound (so $c \in A$), while the second inequality tells us that c is smaller than or equal to every other upper bound. Hence, $c = \min(A) = \sup(X)$. \square

PROPOSITION 2.56: SUPREMUM AND SET OPERATIONS

Let X and Y be nonempty subsets of \mathbb{R} that are bounded from above. Define

$$X + Y := \{x + y \mid x \in X, y \in Y\} \quad \text{and} \quad X \cdot Y := \{x \cdot y \mid x \in X, y \in Y\}.$$

Then the sets $X \cup Y$, $X \cap Y$, and $X + Y$ are also bounded from above. Moreover, if $X, Y \subset \mathbb{R}_{\geq 0}$ (that is, $x \geq 0$ and $y \geq 0$ for all $x \in X$ and $y \in Y$), then $X \cdot Y$ is bounded from above as well.

In these cases, the following formulas hold:

- (1) $\sup(X \cup Y) = \max\{\sup(X), \sup(Y)\}$,
- (2) If $X \cap Y \neq \emptyset$, then $\sup(X \cap Y) \leq \min\{\sup(X), \sup(Y)\}$,
- (3) $\sup(X + Y) = \sup(X) + \sup(Y)$,
- (4) If $X, Y \subset \mathbb{R}_{\geq 0}$, then $\sup(X \cdot Y) = \sup(X) \cdot \sup(Y)$.

Proof. We leave (1) and (2) to the reader.

(3) Let $x_0 = \sup(X)$ and $y_0 = \sup(Y)$. For any $z \in X + Y$, there exist $x \in X$ and $y \in Y$ such that $z = x + y$. Since $x \leq x_0$ and $y \leq y_0$, we have

$$z = x + y \leq x_0 + y_0,$$

so $x_0 + y_0$ is an upper bound for $X + Y$. We now want to show that $x_0 + y_0 = \sup(X + Y)$.

Let $z_0 = \sup(X + Y)$ and suppose, by contradiction, that

$$\varepsilon := x_0 + y_0 - z_0 > 0.$$

Since $x_0 = \sup(X)$, by the characterization (2.5) there exists $x \in X$ such that $x > x_0 - \varepsilon/2$. Likewise, there exists $y \in Y$ such that $y > y_0 - \varepsilon/2$. Setting $z = x + y$, we obtain

$$z > x_0 - \frac{\varepsilon}{2} + y_0 - \frac{\varepsilon}{2} = x_0 + y_0 - \varepsilon = z_0,$$

contradicting the assumption that z_0 is an upper bound for $X + Y$. Therefore, $z_0 = x_0 + y_0$.

(4) The proof is analogous. If all elements of X and Y are non-negative, and we set $x_0 = \sup(X)$ and $y_0 = \sup(Y)$, then for any $z = x \cdot y \in XY$, we have

$$z = x \cdot y \leq x_0 \cdot y_0,$$

which shows that $x_0 \cdot y_0$ is an upper bound for $X \cdot Y$. Using a similar “ ε -argument” as done above when proving (3), one shows that this upper bound is sharp, i.e., $x_0 \cdot y_0$ is the least upper bound. \square

If $X \subseteq \mathbb{R}$ is nonempty and bounded from below, the largest lower bound of X is called the **infimum**, denoted by $\inf(X)$. An existence result analogous to Theorem 2.55 holds for infima as well. Moreover, the infimum can be expressed using the supremum:

$$\inf(X) = -\sup\{-x \mid x \in X\}.$$

This means that most results about infima can be deduced directly from those about suprema.

2.3.2 Two-point Compactification

6

In this section, we extend the notions of **supremum** and **infimum** to arbitrary subsets of \mathbb{R} . To do so, we introduce two formal symbols:

$$+\infty \quad \text{and} \quad -\infty,$$

which are not real numbers. We define the **extended real number line** (also called the **two-point compactification** of \mathbb{R}) by

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}.$$

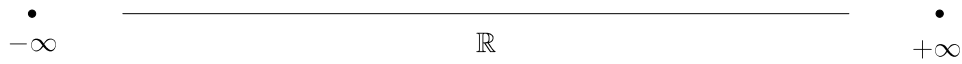


Figure 2.8: The extended real number line $\overline{\mathbb{R}}$, obtained by adding $-\infty$ on the far left and $+\infty$ on the far right of the usual real line.

We extend the usual order relation \leq on \mathbb{R} to $\overline{\mathbb{R}}$ by requiring that

$$-\infty < x < +\infty \quad \text{for all } x \in \mathbb{R}.$$

For simplicity, we often write ∞ instead of $+\infty$.

We now introduce some standard (but informal) computation rules involving these symbols. For all $x \in \mathbb{R}$, we adopt the conventions:

$$\infty + x = \infty + \infty = \infty, \quad -\infty + x = -\infty - \infty = -\infty.$$

If $x > 0$, then

$$x \cdot \infty = \infty \cdot \infty = \infty, \quad x \cdot (-\infty) = \infty \cdot (-\infty) = -\infty,$$

while for $x < 0$ we have

$$x \cdot \infty = -\infty \cdot \infty = -\infty, \quad x \cdot (-\infty) = -\infty \cdot (-\infty) = \infty.$$

These rules are widely used as notational shorthand, but one must handle them with care. Expressions like

$$\infty - \infty, \quad 0 \cdot \infty, \quad \text{or similar}$$

are undefined and should be avoided.

DEFINITION 2.57: SUPREMUM AND INFIMUM IN THE EXTENDED LINE

Let $X \subseteq \mathbb{R}$.

- If X is not bounded from above, we define $\sup(X) = \infty$.
- If $X = \emptyset$, we define $\sup(\emptyset) = -\infty$.
- If X is not bounded from below, we define $\inf(X) = -\infty$.
- If $X = \emptyset$, we define $\inf(\emptyset) = \infty$.

In this context, we refer to ∞ and $-\infty$ as **indefinite values**.

In other words:

- Saying $\sup(X) = \infty$ means that X is not bounded above; i.e.,

$$\forall x_0 \in \mathbb{R} \quad \exists x \in X \text{ such that } x > x_0.$$

- Saying $\sup(X) = -\infty$ means that X is empty.
- Similarly, $\inf(X) = -\infty$ means that X is unbounded below, and $\inf(X) = \infty$ means X is empty.

2.4 Consequences of Completeness

We introduced the root function in Section 2.1 using the completeness axiom, and in Section 2.3 we used the same axiom to prove the existence of suprema. In this section, we discuss further consequences of the completeness axiom.

2.4.1 The Archimedean Principle

The Archimedean principle states that for every real number $x \in \mathbb{R}$ there exists an integer n greater than x . The following theorem, proved using the existence of suprema, gives a precise formulation of this principle.

THEOREM 2.58: ARCHIMEDEAN PRINCIPLE

For every $x \in \mathbb{R}$ there exists exactly one $n \in \mathbb{Z}$ such that

$$n \leq x < n + 1.$$

Proof. We first treat the case $x \geq 0$. Define

$$E = \{n \in \mathbb{Z} \mid n \leq x\}.$$

Since $0 \in E$ and x is an upper bound, E is a non-empty subset of \mathbb{R} bounded from above. Hence, by Theorem 2.55, the supremum $s_0 = \sup(E)$ exists. From the definition of supremum we deduce:

- (i) $s_0 \leq x$ (because x is an upper bound);
- (ii) there exists $n_0 \in E$ with $s_0 - 1 < n_0$ (otherwise $s_0 - 1$ would also be an upper bound).

From (ii) we obtain $s_0 < n_0 + 1$, which implies

- (iii) $n_0 + 1 \notin E$ (otherwise s_0 would not be an upper bound for E).

Moreover, since $m \leq s_0$ for every $m \in E$, we have $m < n_0 + 1$ for all $m \in E$. As all elements of E are integers,

$$m < n_0 + 1 \iff m - n_0 < +1 \iff m - n_0 \leq 0 \iff m \leq n_0.$$

Thus every $m \in E$ is less than or equal to n_0 , and since $n_0 \in E$, we conclude that n_0 is the maximum of E . In particular, by Remark 2.53, the maximum is also the supremum, so $s_0 = n_0$.

Finally, recalling (iii) and the definition of E , we have $n_0 + 1 > x$. Together with (i), this shows

$$n_0 = s_0 \leq x < n_0 + 1,$$

establishing the claim for $x \geq 0$.

Now, if $x < 0$, apply the previous argument to $-x > 0$. Then there exists $m \in \mathbb{Z}$ such that

$$m \leq -x < m + 1,$$

which is equivalent to

$$-m - 1 < x \leq -m.$$

If $x = -m$, then set $n = -m$. If $x < -m$, set $n = -m - 1$. In both cases, we obtain

$$n \leq x < n + 1.$$

Finally, for uniqueness, assume that $n_1, n_2 \in \mathbb{Z}$ both satisfy $n_i \leq x < n_i + 1$. From $n_1 \leq x < n_2 + 1$ we deduce $n_1 < n_2 + 1$, and therefore $n_1 \leq n_2$. Reversing the roles of n_1 and n_2 gives $n_2 \leq n_1$. Hence $n_1 = n_2$. \square

DEFINITION 2.59: INTEGER AND FRACTIONAL PARTS

The **integer part** $\lfloor x \rfloor$ of $x \in \mathbb{R}$ is the integer $n \in \mathbb{Z}$ uniquely determined by Theorem 2.58 such that $n \leq x < n + 1$. The map $x \mapsto \lfloor x \rfloor$ from \mathbb{R} to \mathbb{Z} is called the **rounding function**. The **fractional part** of x is defined as

$$\{x\} = x - \lfloor x \rfloor \in [0, 1).$$

7

COROLLARY 2.60: $\frac{1}{n}$ IS ARBITRARILY SMALL

For every $\varepsilon > 0$ there exists $n \in \mathbb{N}$, with $n \geq 1$, such that

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

Proof. Applying Theorem 2.58 to $x = \frac{1}{\varepsilon} > 0$, we find $m \in \mathbb{Z}$ such that

$$m \leq \frac{1}{\varepsilon} < m + 1.$$

Set $n := m + 1$. In this way we have $0 < \frac{1}{\varepsilon} < n$, which is equivalent to $n > 0$ (therefore, $n \geq 1$) and $\frac{1}{n} < \varepsilon$. \square

COROLLARY 2.61: DENSITY OF \mathbb{Q}

For every $a, b \in \mathbb{R}$ with $a < b$, there exists $r \in \mathbb{Q}$ such that $a < r < b$.

Proof. Set $\varepsilon = b - a$. By Corollary 2.60, there exists $m \in \mathbb{N}$ with $\frac{1}{m} < \varepsilon$. Then, by Theorem 2.58 applies with $x = ma$, there exists $n \in \mathbb{Z}$ with

$$n \leq ma < n + 1,$$

or equivalently,

$$\frac{n}{m} \leq a < \frac{n+1}{m}.$$

Since $\frac{1}{m} < \varepsilon$, by the two inequalities above we obtain

$$a < \frac{n+1}{m} \leq a + \frac{1}{m} < a + \varepsilon = b.$$

Thus $r = \frac{n+1}{m}$ is a rational number between a and b . □

DEFINITION 2.62: DENSE SETS

A subset $X \subset \mathbb{R}$ is called **dense** in \mathbb{R} if every open non-empty interval contains an element of X .

By Corollary 2.61, \mathbb{Q} is dense in \mathbb{R} .

Archimedes' principle can be generalized in several ways. The following generalization will be useful when discussing decimal fractions.

EXERCISE 2.63. — Show the following analogue of Theorem 2.58: *For every $x \in \mathbb{R}$ with $x \geq 1$, there exists exactly one $n \in \mathbb{N}$ such that*

$$10^n \leq x < 10^{n+1}.$$

EXERCISE 2.64. — Show that Corollary 2.60 holds when we replace \mathbb{N} by the powers of 10. In other words: *For every $\varepsilon > 0$ there exists $n \in \mathbb{N}$ such that*

$$10^{-n} < \varepsilon.$$

While the Archimedean principle looks “obvious”, it is actually more subtle than one may think. Indeed, there exist ordered fields for which the Archimedean principle fails, as the following exercise shows.

EXERCISE 2.65 (Advanced). — *ex:non-archimedean* An ordered field is called *non-Archimedean* if the Archimedean property does not hold. Let

$$F = \left\{ \frac{p(x)}{q(x)} : p, q \in \mathbb{R}[x], q(x) \neq 0 \right\},$$

where $\mathbb{R}[x]$ denotes the set of polynomials with real coefficients (e.g. $p(x) = 3x^4 - 5x + 7$, $q(x) = 2x^2 - 1$), and $q(x)$ is different from the polynomial that is identically zero. Inside this field, the set \mathbb{Z} corresponds to the constant polynomial $f(x) \equiv n \in \mathbb{Z}$.

We define the following order: $f > g$ in F if $f(x) - g(x) > 0$ for all sufficiently large $x \in \mathbb{R}$.

1. Show that F is an ordered field.

2. Show that F is non-Archimedean by proving that x is larger than every integer.

2.4.2 Decimal Fraction Expansion and Uncountability

A common way to understand real numbers is through their decimal expansions. Formally, a **decimal fraction** is a sequence of integers

$$a_0, a_1, a_2, a_3, \dots$$

where $a_0 \in \mathbb{Z}$ and $0 \leq a_n \leq 9$ for all $n \geq 1$. To such a sequence, we associate a real number as follows.

Assume first that $a_0 \geq 0$ and define the following approximations:

$$\begin{aligned} x_0 &= a_0, & y_0 &= a_0 + 1, \\ x_1 &= a_0 + \frac{a_1}{10}, & y_1 &= a_0 + \frac{a_1}{10} + \frac{1}{10} = a_0 + \frac{a_1 + 1}{10}, \\ x_2 &= a_0 + \frac{a_1}{10} + \frac{a_2}{100}, & y_2 &= a_0 + \frac{a_1}{10} + \frac{a_2}{100} + \frac{1}{100} = a_0 + \frac{a_1}{10} + \frac{a_2 + 1}{100}, \end{aligned}$$

and, more generally,

$$x_n = a_0 + \frac{a_1}{10} + \dots + \frac{a_n}{10^n} \quad y_n = a_0 + \frac{a_1}{10} + \dots + \frac{a_n + 1}{10^n}. \quad (2.6)$$

7

Note that all the numbers x_n and y_n are rational. Also, since $0 \leq a_k \leq 9$ for $k \geq 1$, one can check that

$$x_0 \leq x_1 \leq \dots \leq x_n \leq x_{n+1} \leq \dots \leq y_{n+1} \leq y_n \leq \dots \leq y_1 \leq y_0.$$

Thus, if we consider the sets $X = \{x_0, x_1, x_2, \dots\}$ and $Y = \{y_0, y_1, y_2, \dots\}$, by the completeness axiom there exists $c \in \mathbb{R}$ such that

$$x_n \leq c \leq y_n \quad \text{for all } n \in \mathbb{N}. \quad (2.7)$$

Also, Exercise 2.64 shows that such a c is uniquely determined: since $y_n - x_n = 10^{-n}$, if $c, d \in \mathbb{R}$ both satisfy (2.7), then

$$|d - c| < 10^{-n} \quad \text{for all } n \in \mathbb{N},$$

which implies $|d - c| = 0$, i.e., $c = d$. We therefore define c to be the real number with decimal expansion

$$c = a_0.a_1a_2a_3\dots$$

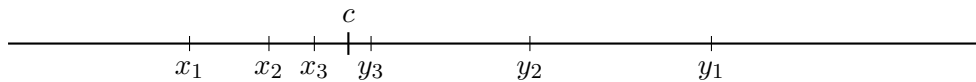


Figure 2.9: The rational numbers x_n give better and better lower estimates for c , while the rational numbers y_n provide increasingly accurate upper estimates. The number c lies between all of them.

REMARK 2.66. — Alternatively, the number c may be defined either as the supremum of the set $X = \{x_0, x_1, x_2, \dots\}$ or as the infimum of the set $Y = \{y_0, y_1, y_2, \dots\}$.

If $a_0 < 0$, we proceed as follows: consider the decimal expansion of the $(-a_0, a_1, a_2, a_3, \dots)$, apply the previous construction, and then define the desired number to be the negative of the result.

EXAMPLE 2.67. — Let

$$a_0 = -33, \quad a_1 = 1, \quad a_2 = 6, \quad a_n = 0 \text{ for } n \geq 3.$$

Then the construction above applied to $(-a_0, a_1, a_2, \dots)$ gives 33.16, so the corresponding real number is

$$c = -33.16.$$

Conversely, we ask: can every $x \in \mathbb{R}$ be written as a decimal expansion? The answer is yes. For $x \geq 0$, set $a_0 := \lfloor x \rfloor$ and define

$$a_n := \lfloor 10^n x \rfloor - 10 \lfloor 10^{n-1} x \rfloor, \quad n \geq 1. \quad (2.8)$$

Then $0 \leq a_n \leq 9$, and one checks that (2.7) holds, so that x has decimal expansion $a_0.a_1a_2a_3\dots$. For $x \leq 0$, apply the same procedure to $-x$ and then change the sign of a_0 .

Note that different decimal expansions can represent the same number, e.g.

$$0.1999\dots = 0.2000\dots = \frac{1}{5}.$$

This phenomenon occurs exactly when the expansion eventually becomes all 9s. To avoid ambiguity, we define a **real decimal fraction** as a sequence

$$a_0, a_1, a_2, a_3, \dots$$

with $0 \leq a_n \leq 9$ for all $n \geq 1$, such that for every $n_0 \geq 1$ there exists $n \geq n_0$ with $a_n \neq 9$.

EXERCISE 2.68. — Let $x \geq 0$ be a real number. Show that the sequence a_0, a_1, a_2, \dots defined by (2.8) is a real decimal fraction. Conclude that this gives a bijection between \mathbb{R} and the set of all real decimal fractions.

CARDINALITY

Let X and Y be sets.

- We say X and Y have the **same cardinality**, written $X \sim Y$, if there is a bijection $f : X \rightarrow Y$.
- We write $X \lesssim Y$ if there is an injection $f : X \rightarrow Y$.
- The empty set has cardinality 0.
- A set X has **finite cardinality** $|X| = n$ if there is a bijection with $\{1, \dots, n\}$.
- A set is **infinite** if it is not finite.
- A set is **countable** if it has a bijection to \mathbb{N} . Its cardinality is denoted \aleph_0 , pronounced Aleph-0.
- A set is **uncountable** if it is infinite but not countable.

REMARK 2.69. — If $X \lesssim Y$ and $Y \lesssim X$, then $X \sim Y$. In other words, if there exist an injective map $f : X \rightarrow Y$ and an injective map $g : Y \rightarrow X$, then one can find a bijective map $h : X \rightarrow Y$. This nontrivial statement is the **Schröder–Bernstein Theorem**.

7

EXAMPLE 2.70 (Some Countability Results). —

1. **\mathbb{N} and the even numbers have the same cardinality.**

Let $\mathbb{N}_{\text{even}} := \{0, 2, 4, 6, \dots\}$. The map

$$f_1 : \mathbb{N} \rightarrow \mathbb{N}_{\text{even}}, \quad f_1(n) = 2n,$$

is a bijection, so $\mathbb{N} \sim \mathbb{N}_{\text{even}}$.

2. **\mathbb{N} and \mathbb{Z} have the same cardinality.**

Define a bijection $f_2 : \mathbb{Z} \rightarrow \mathbb{N}$ by

$$f_2(n) = \begin{cases} 2n & \text{if } n \geq 0, \\ -2n - 1 & \text{if } n < 0. \end{cases}$$

Explicitly:

$$0 \mapsto 0, \quad 1 \mapsto 2, \quad 2 \mapsto 4, \quad 3 \mapsto 6, \quad \dots, \quad -1 \mapsto 1, \quad -2 \mapsto 3, \quad -3 \mapsto 5, \quad \dots$$

This shows that $\mathbb{Z} \sim \mathbb{N}$.

3. **\mathbb{Q} is countable.**

First of all, since $\mathbb{N} \subset \mathbb{Q}$, the inclusion map $n \mapsto n$ is injective. So $\mathbb{N} \lesssim \mathbb{Q}$.

To prove the converse, we first show that $\mathbb{Q} \lesssim \mathbb{N} \times \mathbb{N}$. Indeed, every rational number can be written as a reduced fraction $\frac{p}{q}$ with $p \in \mathbb{Z}$ and $q \in \mathbb{N} \setminus \{0\}$. The map

$$f_3 : \mathbb{Q} \rightarrow \mathbb{Z} \times \mathbb{N} \setminus \{0\}, \quad f_3\left(\frac{p}{q}\right) = (p, q)$$

is injective, so $\mathbb{Q} \lesssim \mathbb{Z} \times \mathbb{N} \setminus \{0\} \lesssim \mathbb{Z} \times \mathbb{N}$. Using the map f_2 from Example 2 we see that that map

$$f_4 : \mathbb{Z} \times \mathbb{N} \rightarrow \mathbb{N} \times \mathbb{N}, \quad f_4(n, m) = (f_2(n), m),$$

is a bijection, therefore $\mathbb{Z} \times \mathbb{N} \sim \mathbb{N} \times \mathbb{N}$. Thus, we have proved that

$$\mathbb{Q} \lesssim \mathbb{Z} \times \mathbb{N} \sim \mathbb{N} \times \mathbb{N}.$$

Next, we show that $\mathbb{N} \times \mathbb{N} \sim \mathbb{N}$. Define the “Cantor pairing function”

$$f_5 : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}, \quad f_5(m, n) = \frac{(m+n)(m+n+1)}{2} + n.$$

One can prove that f_5 is a bijection (although only injectivity is needed here), as one can also convince oneself by looking at Figure 2.10, hence

$$\mathbb{N} \times \mathbb{N} \sim \mathbb{N}.$$

Putting everything together:

$$\mathbb{N} \lesssim \mathbb{Q} \lesssim \mathbb{Z} \times \mathbb{N} \sim \mathbb{N} \times \mathbb{N} \sim \mathbb{N},$$

so it follows from the Schröder–Bernstein Theorem, that $\mathbb{Q} \sim \mathbb{N}$, i.e., the rational numbers are countable.

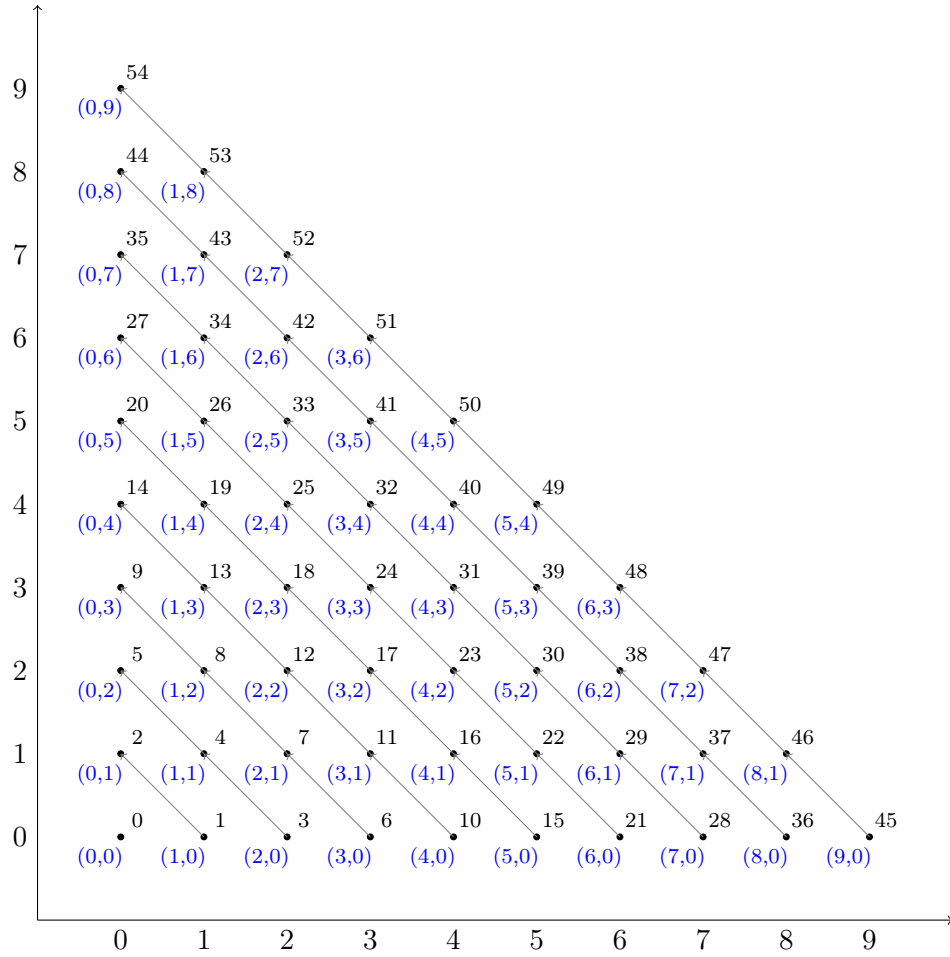


Figure 2.10: A representation of the function f_5 . This map enumerates all elements of $\mathbb{N} \times \mathbb{N}$. The value of $f_5(m, n)$ is in black; the coordinate (m, n) is in blue.

By Example 2.70 we know that \mathbb{Q} is countable, namely, there exists a bijection $f : \mathbb{N} \rightarrow \mathbb{Q}$. Hence, if we define $r_0 = f(0)$, $r_1 = f(1)$, $r_2 = f(2)$, etc., we have proved that there exists a (non-canonical) way of enumerating all the rational numbers. As we shall see now, this is not possible for the real numbers.

PROPOSITION 2.71: UNCOUNTABILITY OF \mathbb{R}

The set \mathbb{R} is uncountable.

Proof. To show that \mathbb{R} is uncountable, we shall actually prove that $[0, 1) \subset \mathbb{R}$ is uncountable. Assume by contradiction $[0, 1)$ is countable. Then there exists a bijective map $f : \mathbb{N} \rightarrow [0, 1)$. In other words, if we define $x_n = f(n)$, we have found a way to enumerate all the real numbers in $[0, 1)$:

$$x_0 = 0.a_{0,1}a_{0,2}a_{0,3}a_{0,4}\dots, \quad x_1 = 0.a_{1,1}a_{1,2}a_{1,3}a_{1,4}\dots,$$

$$x_2 = 0.a_{2,1}a_{2,2}a_{2,3}a_{2,4}\dots, \quad x_3 = 0.a_{3,1}a_{3,2}a_{3,3}a_{3,4}\dots, \quad \dots$$

We now want to construct a new number $x \in [0, 1)$ that is not in the list above. A possible way to do this is the following: write $x = 0.b_1b_2b_3b_4$ and define

$$b_1 = \begin{cases} 5 & \text{if } a_{0,1} \neq 5, \\ 6 & \text{if } a_{0,1} = 5, \end{cases}, \quad b_2 = \begin{cases} 5 & \text{if } a_{1,2} \neq 5, \\ 6 & \text{if } a_{1,2} = 5, \end{cases}, \quad b_3 = \begin{cases} 5 & \text{if } a_{2,3} \neq 5, \\ 6 & \text{if } a_{2,3} = 5, \end{cases},$$

and, more in general,

$$b_i = \begin{cases} 5 & \text{if } a_{i,i+1} \neq 5, \\ 6 & \text{if } a_{i,i+1} = 5. \end{cases} \quad \text{for all } i \geq 1.$$

We now observe that the number $x = 0.b_1b_2b_3\dots$ cannot be in the list, since for every $i \geq 0$ it differs from x_i in the $(i+1)$ -th decimal place. This contradicts the assumption that we have listed all real numbers. Therefore $[0, 1)$ (and hence also \mathbb{R}) is uncountable. \square

Extra material: An alternative proof of the uncountability of \mathbb{R} via power sets

We have seen that sets like \mathbb{Z} and \mathbb{Q} are countable, whereas the set of real numbers \mathbb{R} is fundamentally larger. In this section, we present an alternative proof of this fact and discuss in more detail the question: “How large is \mathbb{R} ?”

POWER SET

Let X be a set. The **power set** $\mathcal{P}(X)$ of X is the set of all subsets of X :

$$\mathcal{P}(X) := \{A \subseteq X\}.$$

EXAMPLE 2.72. — If $X = \{0, 1, 2\}$, then

$$\mathcal{P}(X) = \{\emptyset, \{0\}, \{1\}, \{2\}, \{0, 1\}, \{0, 2\}, \{1, 2\}, \{0, 1, 2\}\}.$$

THEOREM 2.73: CANTOR’S THEOREM

For any set X , the power set $\mathcal{P}(X)$ has strictly larger cardinality than X .

Proof. First of all, since there is an injective map $i : X \rightarrow \mathcal{P}(X)$ given by $i(x) = \{x\}$, we see that $|X| \leq |\mathcal{P}(X)|$.

Now, assume for contradiction that there exists a bijection $f : X \rightarrow \mathcal{P}(X)$ and define the set

$$A := \{x \in X \mid x \notin f(x)\}.$$

Since f is surjective, there exists $a \in X$ such that $f(a) = A$. However we now see that this is impossible. Indeed:

- If $a \in A$, then by the definition of A it follows that $a \notin f(a) = A$, contradiction.
- If $a \notin A$, then by the definition of A it follows that $a \in f(a) = A$, again a contradiction.

Thus, no such bijection exists and $\mathcal{P}(X)$ is strictly larger than X . \square

We now show that the set of real numbers has the same cardinality as $\mathcal{P}(\mathbb{N})$.

PROPOSITION 2.74: THE REALS HAVE THE SAME CARDINALITY AS $\mathcal{P}(\mathbb{N})$

$$|\mathbb{R}| = |\mathcal{P}(\mathbb{N})|.$$

Proof. We first show that $|\mathcal{P}(\mathbb{N})| \lesssim |[0, 1]|$ by constructing an injection

$$\varphi : \mathcal{P}(\mathbb{N}) \hookrightarrow [0, 1].$$

Given a subset $A \subset \mathbb{N}$, define $\varphi(A)$ as the real number with binary expansion

$$\varphi(A) := 0.a_0a_1a_2a_3\ldots \quad \text{where} \quad a_n = \begin{cases} 1 & \text{if } n \in A, \\ 0 & \text{if } n \notin A. \end{cases}$$

If $A \neq B$, let n be the smallest element of $A \triangle B$. Then $\varphi(A)$ and $\varphi(B)$ differ in the n -th digit, so $\varphi(A) \neq \varphi(B)$. Hence φ is injective.

Conversely, we show that $|[0, 1]| \lesssim |\mathcal{P}(\mathbb{N})|$. Every real number $x \in [0, 1)$ has a binary expansion of the form

$$x = 0.b_0b_1b_2\ldots \quad \text{with } b_i \in \{0, 1\}.$$

To such an expansion we can associate the subset

$$A_x := \{n \in \mathbb{N} \mid a_n = 1\} \subseteq \mathbb{N}.$$

This gives a map $[0, 1) \rightarrow \mathcal{P}(\mathbb{N})$. There is only one subtlety: some numbers have two binary expansions (for example, $0.011111\ldots = 0.100000\ldots$). To avoid ambiguity, we agree to always choose the expansion that ends with infinitely many zeros rather than the one ending with infinitely many ones. With this convention, each real number in $[0, 1)$ corresponds uniquely to a subset of \mathbb{N} . Therefore, we have defined an injection $[0, 1) \hookrightarrow \mathcal{P}(\mathbb{N})$, and hence $|[0, 1]| \lesssim |\mathcal{P}(\mathbb{N})|$.

By the Schröder–Bernstein Theorem, we obtain $|[0, 1]| = |\mathcal{P}(\mathbb{N})|$.

Finally, it remains to see that $[0, 1)$ and \mathbb{R} have the same cardinality. An explicit injection is given by

$$f : \mathbb{R} \rightarrow (0, 1), \quad f(x) = \frac{x}{2(1 + |x|)} + \frac{1}{2}.$$

Thus $|\mathbb{R}| \lesssim |(0, 1)| \lesssim |[0, 1]|$. Since trivially $|[0, 1]| \lesssim |\mathbb{R}|$, we conclude that

$$|\mathbb{R}| \sim |[0, 1]|.$$

Combining everything, we obtain $|\mathbb{R}| = |\mathcal{P}(\mathbb{N})|$. \square

The cardinality of $\mathcal{P}(\mathbb{N})$ is denoted \mathfrak{c} and called the **continuum**. Hence, we have proved that

$$|\mathbb{R}| = \mathfrak{c} \quad \text{and} \quad \aleph_0 < \mathfrak{c}.$$

A famous question, posed by Cantor, is whether there exists a set whose cardinality lies strictly between \aleph_0 and \mathfrak{c} . This is the **Continuum Hypothesis**. Remarkably, this question can neither be proved nor disproved from the standard axioms of set theory (ZFC): it is *independent* of them. This was shown by Gödel (1938) and Cohen (1963).

2.5 Sequences of Real Numbers

2.5.1 Convergence of Sequences

Let X be a set. Intuitively, a sequence in X is a list of elements x_0, x_1, x_2, \dots indexed by the natural numbers. In this section we study sequences and their properties in \mathbb{R} . We now give a precise definition.

DEFINITION 2.75: SEQUENCES

A **sequence** in \mathbb{R} is a function $a : \mathbb{N} \rightarrow \mathbb{R}$. For $n \in \mathbb{N}$ we write $a(n) = a_n$ and call a_n the n -th **term** of the sequence. Instead of $a : \mathbb{N} \rightarrow \mathbb{R}$ one often writes $(a_n)_{n \in \mathbb{N}}$, $(a_n)_{n=0}^\infty$, or $(a_n)_{n \geq 0}$.

Since we primarily use the letter x to denote a real number, for sequences of real numbers we shall mostly write $(x_n)_{n \in \mathbb{N}}$, $(x_n)_{n=0}^\infty$, or $(x_n)_{n \geq 0}$.

REMARK 2.76. — Throughout these notes we usually write sequences as $(x_n)_{n=0}^\infty$, but the starting index is not essential. A sequence may equally well be defined from $n = 1$, or from any other integer, without changing any of the notions or results below. In particular, shifting the starting index does not affect the concepts of convergence, boundedness, or monotonicity, which we will discuss later.

8

EXAMPLE 2.77. — Examples of sequences are:

- (i) $(x_n)_{n=0}^\infty$ given by $x_n = (-1)^n$;
- (ii) $(x_n)_{n=1}^\infty$ given by $x_n = \frac{1}{n}$.

DEFINITION 2.78: (EVENTUALLY) CONSTANT SEQUENCES

A sequence $(x_n)_{n=0}^\infty$ is **constant** if $x_n = x_m$ for all $m, n \in \mathbb{N}$. It is **eventually constant** if there exist $N \in \mathbb{N}$ such that $x_n = x_m$ for all $n, m \geq N$.

DEFINITION 2.79: CONVERGENCE OF SEQUENCES

Let $(x_n)_{n=0}^\infty$ be a sequence in \mathbb{R} . We say that $(x_n)_{n=0}^\infty$ **converges** (or is **convergent**) if there exists $A \in \mathbb{R}$ such that

$$\forall \varepsilon > 0 \exists N \in \mathbb{N} \text{ such that } |x_n - A| < \varepsilon \text{ for all } n \geq N.$$

In this case we write

$$\lim_{n \rightarrow \infty} x_n = A \tag{2.9}$$

and call A the **limit** of $(x_n)_{n=0}^\infty$.

A priori it is not clear that a convergent sequence has only one limit. The following lemma shows that the limit is indeed unique, so the notation (2.9) is justified.

LEMMA 2.80: UNIQUENESS OF THE LIMIT

A convergent sequence $(x_n)_{n=0}^{\infty}$ has exactly one limit.

Proof. Let $A, B \in \mathbb{R}$ be limits of $(x_n)_{n=0}^{\infty}$. Fix $\varepsilon > 0$. Then there exist $N_A, N_B \in \mathbb{N}$ such that $|x_n - A| < \varepsilon$ for all $n \geq N_A$ and $|x_n - B| < \varepsilon$ for all $n \geq N_B$. Setting $N = \max\{N_A, N_B\}$, we have

$$|A - B| \leq |A - x_N| + |x_N - B| < \varepsilon + \varepsilon = 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, $|A - B| = 0$ and hence $A = B$. □

EXAMPLE 2.81. — A constant sequence $(x_n)_{n=0}^{\infty}$ with $x_n = A \in \mathbb{R}$ for all n converges to A . Similarly, an eventually constant sequence converges to the value it eventually takes.

EXAMPLE 2.82. — The sequence $(\frac{1}{n})_{n=1}^{\infty}$ converges to 0, i.e., $\lim_{n \rightarrow \infty} \frac{1}{n} = 0$. Indeed, given $\varepsilon > 0$, by Archimedes' principle (Theorem 2.58) there exists $N \in \mathbb{N}$ with $\frac{1}{N} < \varepsilon$. Then, for all $n \geq N$, we have $|\frac{1}{n} - 0| = \frac{1}{n} \leq \frac{1}{N} < \varepsilon$.

8

EXAMPLE 2.83. — The sequence $(x_n)_{n=0}^{\infty}$ defined by $x_n = (-1)^n$ is not convergent, since its terms alternate between 1 and -1 and do not approach any real number.

2.5.2 Convergent Subsequences and Accumulation Points

Let $(x_n)_{n=0}^{\infty}$ be a sequence in \mathbb{R} . A *subsequence* of $(x_n)_{n=0}^{\infty}$ is obtained by keeping only certain elements and discarding the others. For example,

$$x_0, x_1, x_4, x_9, x_{16}, x_{25}, \dots$$

is a subsequence. The formal definition is as follows.

DEFINITION 2.84: SUBSEQUENCES

Let $(x_n)_{n=0}^{\infty}$ be a sequence. A **subsequence** is a sequence of the form $(x_{n_k})_{k=0}^{\infty}$, where $(n_k)_{k=0}^{\infty}$ is a strictly increasing sequence of nonnegative integers, i.e. $n_{k+1} > n_k$ for all $k \in \mathbb{N}$.

REMARK 2.85. — Since $n_{k+1} > n_k$ for all $k \in \mathbb{N}$, it follows by induction that $n_k \geq k$ for every $k \in \mathbb{N}$.

LEMMA 2.86: SUBSEQUENCES OF CONVERGENT SEQUENCES ARE CONVERGENT

Let $(x_n)_{n=0}^{\infty}$ be a sequence converging to $A \in \mathbb{R}$. Then every subsequence $(x_{n_k})_{k=0}^{\infty}$ also converges to A .

Proof. We leave the proof as an exercise. □

8

A sequence can have convergent subsequences without being convergent itself. For example, the sequence $x_n = (-1)^n$ is not convergent, while the subsequences

$$(x_{2n})_{n=0}^{\infty} \quad \text{and} \quad (x_{2n+1})_{n=0}^{\infty}$$

are constant (equal to 1 and -1 , respectively), hence convergent.

DEFINITION 2.87: ACCUMULATION POINTS OF SEQUENCES

Let $(x_n)_{n=0}^{\infty}$ be a sequence in \mathbb{R} . A point $A \in \mathbb{R}$ is an **accumulation point** of $(x_n)_{n=0}^{\infty}$ if for every $\varepsilon > 0$ and every $N \in \mathbb{N}$ there exists $n \geq N$ such that $|x_n - A| < \varepsilon$.

PROPOSITION 2.88: SUBSEQUENCES AND ACCUMULATION POINTS

Let $(x_n)_{n=0}^{\infty}$ be a sequence in \mathbb{R} . A point $A \in \mathbb{R}$ is an accumulation point of $(x_n)_{n=0}^{\infty}$ if and only if there exists a convergent subsequence of $(x_n)_{n=0}^{\infty}$ with limit A .

Proof. Assume first A is an accumulation point. We construct $(n_k)_{k \geq 0}$ recursively:

- first, apply the definition of accumulation point with $N = 1$ and $\varepsilon = 1 = 2^{-0}$ to find $n_0 \geq 1$ with $|x_{n_0} - A| \leq 2^{-0}$;
- second, apply the definition of accumulation point with $N = n_0 + 1$ and $\varepsilon = 2^{-1}$ to find $n_1 \geq n_0 + 1$ with $|x_{n_1} - A| \leq 2^{-1}$;
- more in general, given n_{k-1} , apply the definition of accumulation point with $N = n_{k-1} + 1$ and $\varepsilon = 2^{-k}$ to find $n_k > n_{k-1}$ with $|x_{n_k} - A| \leq 2^{-k}$.

Now, given $\varepsilon > 0$, pick N so that $2^{-N} < \varepsilon$. Then, for all $k \geq N$ we have

$$|x_{n_k} - A| \leq 2^{-k} \leq 2^{-N} < \varepsilon,$$

so $\lim_{k \rightarrow \infty} x_{n_k} = A$.

Conversely, assume that there exists a subsequence $(x_{n_k})_{k=0}^{\infty}$ converging to A . Fix $\varepsilon > 0$ and $N \in \mathbb{N}$. Since $\lim_{k \rightarrow \infty} x_{n_k} = A$, there exists N_0 such that $|x_{n_k} - A| < \varepsilon$ for all $k \geq N_0$. Hence, if we choose $k = \max\{N_0, N\}$, because $n_k \geq k$ (recall Remark 2.85) we have $n_k \geq N$ and $|x_{n_k} - A| < \varepsilon$. Thus A is an accumulation point. □

9

COROLLARY 2.89: INFINITELY MANY TERMS NEAR AN ACCUMULATION POINT

If $A \in \mathbb{R}$ is an accumulation point of $(x_n)_{n=0}^\infty$, then for every $\varepsilon > 0$ there are infinitely many n with $x_n \in (A - \varepsilon, A + \varepsilon)$.

Proof. By Proposition 2.88, there exists a subsequence $(x_{n_k})_{k=0}^\infty$ with $\lim_{k \rightarrow \infty} x_{n_k} = A$. Hence for any $\varepsilon > 0$ there exists K such that $x_{n_k} \in (A - \varepsilon, A + \varepsilon)$ for all $k \geq K$, providing infinitely many elements of the sequence inside the interval $(A - \varepsilon, A + \varepsilon)$. \square

COROLLARY 2.90: ACCUMULATION POINTS OF CONVERGENT SEQUENCES

A convergent sequence has exactly one accumulation point, namely its limit.

Proof. This follows from Lemma 2.86 and Proposition 2.88. \square

EXAMPLE 2.91. — The sequence $(x_n)_{n=0}^\infty$, defined by $x_n = (-1)^n$, has two accumulation points: 1 and -1 . Indeed, the subsequence $(x_{2n})_{n=0}^\infty$ is constantly equal to 1, hence it converges to 1, while the subsequence $(x_{2n+1})_{n=0}^\infty$ is constantly equal to -1 , and therefore converges to -1 .

EXERCISE 2.92. — Let $(x_n)_{n=0}^\infty$ be a sequence in \mathbb{R} , and let $E \subseteq \mathbb{R}$ be the set of its accumulation points. Show that E is closed.

EXERCISE 2.93. — Construct a sequence $(x_n)_{n=0}^\infty$ in \mathbb{R} whose set of accumulation points is the entire interval $[0, 1]$.

Hint: Enumerate the rational numbers in $[0, 1]$.

2.5.3 Addition, Multiplication, and Inequalities

Given two sequences $(x_n)_{n=0}^\infty$ and $(y_n)_{n=0}^\infty$, one can combine them through **addition** and **multiplication**, and study how these operations affect their limits. Moreover, one can compare sequences using inequalities and relate these inequalities to the corresponding limits.

PROPOSITION 2.94: LIMITS AND OPERATIONS

Let $(x_n)_{n=0}^\infty$ and $(y_n)_{n=0}^\infty$ be sequences converging to $A, B \in \mathbb{R}$, respectively. Then:

- (1) The sequence $(x_n + y_n)_{n=0}^\infty$ converges to $A + B$.
- (2) The sequence $(x_n y_n)_{n=0}^\infty$ converges to AB .
- (3) Given $\alpha \in \mathbb{R}$, the sequence $(\alpha x_n)_{n=0}^\infty$ converges to αA .
- (4) Suppose $x_n \neq 0$ for all $n \in \mathbb{N}$ and $A \neq 0$. Then the sequence $(x_n^{-1})_{n=0}^\infty$ converges to A^{-1} .

Proof. We only prove (1) and (4), leaving (2) and (3) as exercises for the reader. To prove (2), see Exercise 2.101 below.

(1) Fix $\varepsilon > 0$. Because $\lim_{n \rightarrow \infty} x_n = A$ and $\lim_{n \rightarrow \infty} y_n = B$, there exist $N_A, N_B \in \mathbb{N}$ such that

$$|x_n - A| < \frac{\varepsilon}{2} \quad \forall n \geq N_A, \quad |y_n - B| < \frac{\varepsilon}{2} \quad \forall n \geq N_B.$$

Then, for all $n \geq N = \max\{N_A, N_B\}$,

$$|(x_n + y_n) - (A + B)| \leq |x_n - A| + |y_n - B| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

(4) Since $A \neq 0$, we can apply the definition of limit with $\varepsilon = \frac{|A|}{2}$ to find $N_0 \in \mathbb{N}$ such that for all $n \geq N_0$,

$$|x_n - A| < \frac{|A|}{2}.$$

Using the reverse triangle inequality (Lemma 2.19(h)), this gives

$$|x_n| \geq |A| - |x_n - A| > \frac{|A|}{2} \quad \forall n \geq N_0,$$

so

$$\frac{1}{|x_n|} < \frac{2}{|A|} \quad \forall n \geq N_0.$$

9

Hence, we have proved that

$$|x_n^{-1} - A^{-1}| = \left| \frac{x_n - A}{x_n A} \right| = \frac{|x_n - A|}{|x_n| |A|} < \frac{2}{|A|^2} |x_n - A| \quad \forall n \geq N_0.$$

Now, given $\varepsilon > 0$, choose $N_1 \in \mathbb{N}$ such that

$$|x_n - A| < \frac{|A|^2}{2} \varepsilon \quad \forall n \geq N_1.$$

Then, if we define $N = \max\{N_0, N_1\}$, we have

$$|x_n^{-1} - A^{-1}| < \frac{2}{|A|^2} |x_n - A| < \varepsilon \quad \forall n \geq N.$$

Thus the sequence $(x_n^{-1})_{n=0}^{\infty}$ converges to A^{-1} . □

PROPOSITION 2.95: LIMITS AND INEQUALITIES

Let $(x_n)_{n=0}^{\infty}$ and $(y_n)_{n=0}^{\infty}$ be sequences converging to $A, B \in \mathbb{R}$, respectively.

1. If $A < B$, then there exists $N \in \mathbb{N}$ such that $x_n < y_n$ for all $n \geq N$.
2. If there exists $N \in \mathbb{N}$ such that $x_n \leq y_n$ for all $n \geq N$, then $A \leq B$.

Proof. We first prove (1). Let $\varepsilon = \frac{1}{3}(B - A) > 0$, so that $A + \varepsilon < B - \varepsilon$. We know that there exist $N_A, N_B \in \mathbb{N}$ such that

$$n \geq N_A \Rightarrow |x_n - A| < \varepsilon, \quad n \geq N_B \Rightarrow |y_n - B| < \varepsilon.$$

Thus, for $N = \max\{N_A, N_B\}$, we have

$$x_n < A + \varepsilon < B - \varepsilon < y_n \quad \forall n \geq N,$$

proving (1).

For (2), suppose by contradiction that $A > B$. Then (1) implies the existence of N_0 such that $x_n > y_n$ for all $n \geq N_0$, contradicting the assumption $x_n \leq y_n$ for large n . \square

REMARK 2.96. — In Proposition 2.95(2), even if $x_n < y_n$ for all n , one cannot conclude $A < B$. For instance, take

$$x_n = -\frac{1}{n}, \quad y_n = \frac{1}{n}.$$

Then $x_n < y_n$ for all n , but both sequences converge to 0.

LEMMA 2.97: SANDWICH LEMMA

Let $(x_n)_{n=0}^\infty, (y_n)_{n=0}^\infty, (z_n)_{n=0}^\infty$ be sequences such that, for some $N \in \mathbb{N}$,

$$x_n \leq y_n \leq z_n \quad \forall n \geq N.$$

Suppose that $(x_n)_{n=0}^\infty$ and $(z_n)_{n=0}^\infty$ both converge to the same limit. Then $(y_n)_{n=0}^\infty$ also converges, and

$$\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} y_n = \lim_{n \rightarrow \infty} z_n.$$

Proof. The proof is left as an exercise. \square

EXERCISE 2.98. — Compute the following limits, if they exist:

$$\lim_{n \rightarrow \infty} \frac{7n^4 + 15}{3n^4 + n^3 + n - 1}, \quad \lim_{n \rightarrow \infty} \frac{n^2 + 5}{n^3 + n + 1}, \quad \lim_{n \rightarrow \infty} \frac{n^5 - 10}{n^2 + 1}.$$

2.5.4 Bounded and Monotone Sequences

DEFINITION 2.99: BOUNDED SEQUENCES

A sequence $(x_n)_{n=0}^{\infty}$ in \mathbb{R} is called **bounded** if there exists a real number $M \geq 0$ such that

$$|x_n| \leq M \quad \forall n \in \mathbb{N}.$$

LEMMA 2.100: CONVERGENT SEQUENCES ARE BOUNDED

Every convergent sequence is bounded.

Proof. Let $(x_n)_{n=0}^{\infty}$ be a convergent sequence with limit $A \in \mathbb{R}$. Choosing $\varepsilon = 1$ in the definition of limit, there exists $N \in \mathbb{N}$ such that $|x_n - A| \leq 1$ for all $n \geq N$. By the triangle inequality (see Lemma 2.19(g)),

$$|x_n| = |(x_n - A) + A| \leq |x_n - A| + |A| \leq 1 + |A| \quad \forall n \geq N.$$

Set

$$M = \max\{1 + |A|, |x_0|, |x_1|, \dots, |x_{N-1}|\}.$$

Then $|x_n| \leq M$ for all $n \in \mathbb{N}$, as required. \square

EXERCISE 2.101. — Prove statement (2) in Proposition 2.94.

Hint: Note that

$$|x_n y_n - AB| = |x_n y_n - x_n B + x_n B - AB| \leq |y_n - B| |x_n| + |x_n - A| |B|.$$

Since $(x_n)_{n=0}^{\infty}$ converges, Lemma 2.100 guarantees that $|x_n| \leq M$ for some $M \in \mathbb{R}$. Use this to conclude.

We will see later that every bounded sequence has at least one accumulation point, or equivalently, a convergent subsequence. Before that, we introduce the notion of monotonicity, which will allow us to identify a large class of automatically convergent sequences.

DEFINITION 2.102: MONOTONE SEQUENCES

A sequence $(x_n)_{n=0}^{\infty}$ is called:

- **(monotonically) increasing** if $m > n \implies x_m \geq x_n$,
- **strictly (monotonically) increasing** if $m > n \implies x_m > x_n$,
- **(monotonically) decreasing** if $m > n \implies x_m \leq x_n$,
- **strictly (monotonically) decreasing** if $m > n \implies x_m < x_n$.

If a sequence is increasing or decreasing, we call it **monotone**; if it is strictly increasing or strictly decreasing, we call it **strictly monotone**.

REMARK 2.103. — An equivalent formulation of monotone sequences can be given using only successive terms:

- $(x_n)_{n=0}^{\infty}$ is increasing if $x_{n+1} \geq x_n$ for all n ;
- $(x_n)_{n=0}^{\infty}$ is strictly increasing if $x_{n+1} > x_n$ for all n ;
- $(x_n)_{n=0}^{\infty}$ is decreasing if $x_{n+1} \leq x_n$ for all n ;
- $(x_n)_{n=0}^{\infty}$ is strictly decreasing if $x_{n+1} < x_n$ for all n .

Monotone bounded sequences are always convergent. This is proved in the next theorem and illustrated in Figure 2.11.

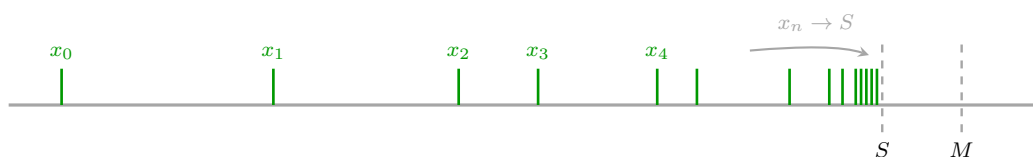


Figure 2.11: A monotonically increasing sequence $(x_n)_{n=0}^{\infty}$ represented along a single axis. Each term x_n is shown as a vertical segment; the sequence moves rightward, approaching the supremum S without ever reaching the upper bound M . The limit of the sequence equals S .

THEOREM 2.104: CONVERGENCE OF MONOTONE SEQUENCES

A monotone sequence $(x_n)_{n=0}^{\infty}$ converges if and only if it is bounded.

More precisely, let $X = \{x_0, x_1, x_2, x_3, \dots\} \subset \mathbb{R}$ denote the set of points in the sequence.

If $(x_n)_{n=0}^{\infty}$ is increasing, then $\lim_{n \rightarrow \infty} x_n = \sup(X)$;

If $(x_n)_{n=0}^{\infty}$ is decreasing, then $\lim_{n \rightarrow \infty} x_n = \inf(X)$.

Proof. If $(x_n)_{n=0}^{\infty}$ converges, Lemma 2.100 shows that it is bounded.

Conversely, let $(x_n)_{n=0}^\infty$ be a bounded monotone sequence. Without loss of generality, assume it is increasing (otherwise consider $(-x_n)_{n=0}^\infty$). Since $(x_n)_{n=0}^\infty$ is bounded from above, the set $X = \{x_0, x_1, x_2, x_3, \dots\}$ has a supremum, say $A = \sup(X)$.

By definition of A :

- (i) $x_n \leq A$ for all n ;
- (ii) for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $x_N > A - \varepsilon$.

Then, for all $n \geq N$, using (ii) and monotonicity we deduce that $x_n \geq x_N > A - \varepsilon$. Combining this with (i), we conclude that

$$A - \varepsilon < x_n < A + \varepsilon \quad \text{for all } n \geq N,$$

proving $\lim_{n \rightarrow \infty} x_n = A$. □

REMARK 2.105. — If $(x_n)_{n=0}^\infty$ is monotone and admits a bounded subsequence $(x_{n_k})_{k=0}^\infty$, then the whole sequence is bounded and, therefore, convergent by Theorem 2.104.

Indeed, assume for instance that $(x_n)_{n=0}^\infty$ is increasing and $(x_{n_k})_{k=0}^\infty$ is bounded above by M . Then, by monotonicity and Remark 2.85,

$$x_0 \leq x_k \leq x_{n_k} \leq M \quad \forall k \in \mathbb{N},$$

so $(x_n)_{n=0}^\infty$ is bounded. The case of a decreasing sequence is analogous.

EXERCISE 2.106. — Let $(x_n)_{n=0}^\infty$ be defined by $x_0 = 1$ and

$$x_n = \frac{2}{3} \left(x_{n-1} + \frac{1}{x_{n-1}} \right) \quad \text{for } n \geq 1.$$

Show that $(x_n)_{n=0}^\infty$ converges and determine its limit.

Hint: First, prove that the sequence converges to a nonnegative limit A . Then, show that the limit satisfies the relation $A = \frac{2}{3} \left(A + \frac{1}{A} \right)$ and use it to identify A .

EXERCISE 2.107. — Let $(x_n)_{n=0}^\infty$ be monotonically increasing and $(y_n)_{n=0}^\infty$ monotonically decreasing, with $x_n \leq y_n$ for all $n \in \mathbb{N}$. Show that both sequences converge and that

$$\lim_{n \rightarrow \infty} x_n \leq \lim_{n \rightarrow \infty} y_n.$$

Illustrate your reasoning with a diagram similar to Figure 2.11.

2.5.5 Superior and Inferior Limits

Let $(x_n)_{n=0}^\infty$ be a bounded sequence. To study its behaviour for large n , it is useful to consider its *tails*:

$$X_{\geq n} = \{x_n, x_{n+1}, x_{n+2}, \dots\} = \{x_k \mid k \geq n\} \subset \mathbb{R}.$$

The concept of limit can be restated in terms of these tails: the sequence $(x_n)_{n=0}^\infty$ converges to A if and only if, for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$X_{\geq N} \subset (A - \varepsilon, A + \varepsilon).$$

However, since not every sequence has a limit, we now introduce a related notion (the **superior** and **inferior limits**), which always exist for bounded sequences.

For each $n \in \mathbb{N}$, define

$$s_n = \sup(X_{\geq n}) = \sup_{k \geq n} x_k, \quad i_n = \inf(X_{\geq n}) = \inf_{k \geq n} x_k.$$

Since $X_{\geq m} \subset X_{\geq n}$ whenever $m > n$, we have

$$i_n \leq i_m \leq s_m \leq s_n \quad \text{for all } m > n.$$

Thus, $(s_n)_{n=0}^\infty$ is a monotonically decreasing sequence, while $(i_n)_{n=0}^\infty$ is monotonically increasing. Moreover, since $(x_n)_{n=0}^\infty$ is bounded, both $(s_n)_{n=0}^\infty$ and $(i_n)_{n=0}^\infty$ are bounded as well. Hence, by Theorem 2.104, both sequences converge. Their limits will be called the *superior limit* and the *inferior limit* of $(x_n)_{n=0}^\infty$, respectively.

REMARK 2.108. — In the sequel, we may use the shorthand

$$\sup_{k \geq n} x_k \quad \text{for} \quad \sup\{x_k \mid k \geq n\}, \quad \inf_{k \geq n} x_k \quad \text{for} \quad \inf\{x_k \mid k \geq n\}.$$

That is, the index range “ $k \geq n$ ” indicates that we are taking the supremum or infimum over all terms of the sequence with index greater than or equal to n . With this notation,

$$s_n = \sup_{k \geq n} x_k, \quad i_n = \inf_{k \geq n} x_k.$$

Note that, since $x_n \in X_{\geq n}$,

$$i_n \leq x_n \leq s_n \quad \forall n \in \mathbb{N}. \quad (2.10)$$

DEFINITION 2.109: SUPERIOR AND INFERIOR LIMITS

Let $(x_n)_{n=0}^{\infty}$ be a bounded sequence in \mathbb{R} . The numbers

$$\limsup_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \left(\sup_{k \geq n} x_k \right), \quad \liminf_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \left(\inf_{k \geq n} x_k \right)$$

are called the **superior limit** and **inferior limit** of $(x_n)_{n=0}^{\infty}$, respectively.

From (2.10) and Proposition 2.95, we have

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

EXAMPLE 2.110. — Let $(x_n)_{n=1}^{\infty}$ be defined by $x_n = (-1)^n + \frac{1}{n}$. We compute the corresponding sequences (s_n) and (i_n) :

n	1	2	3	4	5	6	7	8	...
x_n	0	$\frac{3}{2}$	$-\frac{2}{3}$	$\frac{5}{4}$	$-\frac{4}{5}$	$\frac{7}{6}$	$-\frac{6}{7}$	$\frac{9}{8}$...
s_n	$\frac{3}{2}$	$\frac{3}{2}$	$\frac{5}{4}$	$\frac{5}{4}$	$\frac{7}{6}$	$\frac{7}{6}$	$\frac{9}{8}$	$\frac{9}{8}$...
i_n	-1	-1	-1	-1	-1	-1	-1	-1	...

Here $s_n = x_n$ for even n and $s_n = x_{n+1}$ for odd n , hence

$$\limsup_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \left((-1)^{2n} + \frac{1}{2n} \right) = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{2n} \right) = 1.$$

Instead, since $x_n \geq -1$ and $\lim_{n \rightarrow \infty} x_{2n+1} = -1$, we get $i_n = -1$ for all $n \geq 1$, therefore $\liminf_{n \rightarrow \infty} x_n = -1$.

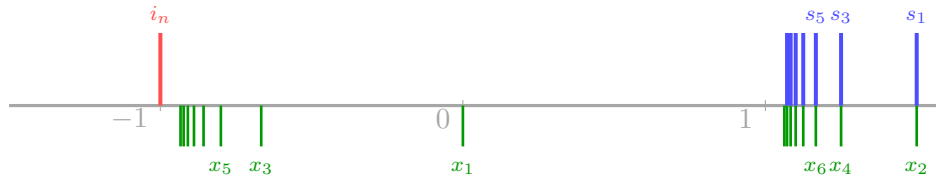


Figure 2.12: Representation of the sequence $x_n = (-1)^n + \frac{1}{n}$. Green bars below the axis show the first few values of x_n : even indices approach 1, odd indices approach -1. Blue bars above the axis illustrate the superior bounds $s_n = \sup_{k \geq n} x_k$: here $s_1(= s_2)$, $s_3(= s_4)$, and $s_5(= s_6)$ are labeled, while the unlabeled bars suggest the continued decrease toward 1. The red bar at -1 marks $i_n = \inf_{k \geq n} x_k = -1$ for all n .

LEMMA 2.111: CONVERGENCE AND SUPERIOR/INFERIOR LIMITS

A bounded sequence $(x_n)_{n=0}^{\infty}$ in \mathbb{R} converges if and only if

$$\limsup_{n \rightarrow \infty} x_n = \liminf_{n \rightarrow \infty} x_n.$$

Proof. For every $n \in \mathbb{N}$, define

$$i_n = \inf_{k \geq n} x_k, \quad s_n = \sup_{k \geq n} x_k,$$

and set

$$I = \lim_{n \rightarrow \infty} i_n = \liminf_{n \rightarrow \infty} x_n, \quad S = \lim_{n \rightarrow \infty} s_n = \limsup_{n \rightarrow \infty} x_n.$$

First, suppose that $I = S$. Since $i_n \leq x_n \leq s_n$ (see (2.10)), the Sandwich Lemma 2.97 implies that the sequence $(x_n)_{n=0}^\infty$ converges, and its limit equals $I = S$.

Conversely, assume that $(x_n)_{n=0}^\infty$ converges to $A \in \mathbb{R}$. Given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$A - \varepsilon < x_n < A + \varepsilon \quad \forall n \geq N.$$

Then, for all $n \geq N$, the same inequalities holds for s_n and i_n :

$$A - \varepsilon \leq i_n \leq s_n \leq A + \varepsilon.$$

Taking limits and using Proposition 2.95, we obtain

$$A - \varepsilon \leq I \leq S \leq A + \varepsilon.$$

10

Since $\varepsilon > 0$ is arbitrary, it follows that $A = I = S$, which proves the result. \square

THEOREM 2.112: SUPERIOR AND INFERIOR LIMITS ARE ACCUMULATION POINTS

Let $(x_n)_{n=0}^\infty$ be a bounded sequence and let

$$A = \limsup_{n \rightarrow \infty} x_n.$$

Then A is an accumulation point of $(x_n)_{n=0}^\infty$, and for every $\varepsilon > 0$ the following hold:

- (1) only finitely many elements satisfy $x_n \geq A + \varepsilon$;
- (2) infinitely many elements satisfy $A - \varepsilon < x_n < A + \varepsilon$.

An analogous statement holds for the inferior limit.

Proof. Recall the notation

$$s_n = \sup_{k \geq n} x_k, \quad n \in \mathbb{N}.$$

Since the sequence $(s_n)_{n=0}^\infty$ is monotonically decreasing and converges to A , given $\varepsilon > 0$ there exists $N_0 \in \mathbb{N}$ such that

$$A \leq s_n < A + \varepsilon \quad \forall n \geq N_0. \quad (2.11)$$

We first prove that A is an accumulation point.

Fix $N \in \mathbb{N}$ and set $N_1 = \max\{N, N_0\}$. Since $s_{N_1} = \sup_{k \geq N_1} x_k$, there exists $n_1 \geq N_1 \geq N_0$ such that

$$s_{N_1} - \varepsilon < x_{n_1} \leq s_{N_1}.$$

Thus, combining this bound with (2.11) we obtain

$$A - \varepsilon \leq s_{N_1} - \varepsilon < x_{n_1} \leq s_{N_1} < A + \varepsilon.$$

This construction shows that for any $\varepsilon > 0$ and any $N \in \mathbb{N}$, there exists $n_1 \geq N$ such that $A - \varepsilon < x_{n_1} < A + \varepsilon$. Thus A is an accumulation point of $(x_n)_{n=0}^\infty$.

We now prove (1) and (2). From (2.11) we have $x_n < A + \varepsilon$ for all $n \geq N_0$, so only finitely many terms satisfy $x_n \geq A + \varepsilon$. This shows (1).

Also, since A is an accumulation point, it follows from Corollary 2.89 that infinitely many terms of the sequence lie within any interval $(A - \varepsilon, A + \varepsilon)$. \square

COROLLARY 2.113: BOUNDED SEQUENCES HAVE CONVERGENT SUBSEQUENCES

Every bounded sequence has at least one accumulation point and therefore possesses a convergent subsequence.

Proof. By Theorem 2.112, the number

$$A = \limsup_{n \rightarrow \infty} x_n$$

is always an accumulation point of $(x_n)_{n=0}^\infty$. Moreover, by Proposition 2.88, every accumulation point is the limit of a convergent subsequence. Hence every bounded sequence admits at least one convergent subsequence. \square

EXERCISE 2.114. — Let $(x_n)_{n=0}^\infty$ be a bounded sequence in \mathbb{R} , and let $E \subseteq \mathbb{R}$ be the set of accumulation points of $(x_n)_{n=0}^\infty$. Show that

$$\limsup_{n \rightarrow \infty} x_n = \max(E), \quad \liminf_{n \rightarrow \infty} x_n = \min(E).$$

EXERCISE 2.115. — Let $(a_n)_{n=0}^\infty$, $(b_n)_{n=0}^\infty$, $(c_n)_{n=0}^\infty$ be convergent sequences with limits $A, B, C \in \mathbb{R}$, respectively. Define a new sequence $(x_n)_{n=0}^\infty$ by

$$x_n = \begin{cases} a_n, & \text{if } n = 3k, k \in \mathbb{N}, \\ b_n, & \text{if } n = 3k + 1, k \in \mathbb{N}, \\ c_n, & \text{if } n = 3k + 2, k \in \mathbb{N}. \end{cases}$$

Compute $\limsup_{n \rightarrow \infty} x_n$, $\liminf_{n \rightarrow \infty} x_n$, and describe the set of accumulation points of $(x_n)_{n=0}^\infty$.

EXERCISE 2.116. — Let $(x_n)_{n=0}^{\infty}$ be a bounded sequence such that $(x_{n+1} - x_n)_{n=0}^{\infty}$ converges to 0. Set

$$A = \liminf_{n \rightarrow \infty} x_n, \quad B = \limsup_{n \rightarrow \infty} x_n.$$

Show that the set of accumulation points of $(x_n)_{n=0}^{\infty}$ is the interval $[A, B]$. Construct an example of such a sequence with $A = 0$ and $B = 1$.

2.5.6 Cauchy Sequences

DEFINITION 2.117: CAUCHY SEQUENCES

A sequence $(x_n)_{n=0}^{\infty}$ in \mathbb{R} is called a **Cauchy sequence** if for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$|x_n - x_m| < \varepsilon \quad \forall m, n \geq N.$$

LEMMA 2.118: CAUCHY SEQUENCES ARE BOUNDED

Every Cauchy sequence is bounded.

Proof. By definition, there exists $N \in \mathbb{N}$ such that

$$|x_n - x_N| \leq 1 \quad \forall n \geq N.$$

Hence, for $n \geq N$, we have $|x_n| \leq |x_N| + 1$.

Now, define

$$M = \max\{|x_0|, |x_1|, \dots, |x_{N-1}|, |x_N| + 1\}.$$

Then $|x_n| \leq M$ for all $n \in \mathbb{N}$, so $(x_n)_{n=0}^{\infty}$ is bounded. \square

EXERCISE 2.119. — Show that a Cauchy sequence $(x_n)_{n=0}^{\infty}$ converges if and only if it has a convergent subsequence.

THEOREM 2.120: CONVERGENCE AND CAUCHY SEQUENCES

A sequence $(x_n)_{n=0}^{\infty}$ of real numbers converges if and only if it is a Cauchy sequence.

Proof. Suppose first that $(x_n)_{n=0}^{\infty}$ converges to some $A \in \mathbb{R}$, and let us prove that $(x_n)_{n=0}^{\infty}$ is a Cauchy sequence. Given $\varepsilon > 0$, choose $N \in \mathbb{N}$ such that

$$|x_n - A| < \frac{\varepsilon}{2} \quad \forall n \geq N.$$

Then for all $m, n \geq N$,

$$|x_n - x_m| \leq |x_n - A| + |x_m - A| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

hence $(x_n)_{n=0}^{\infty}$ is a Cauchy sequence.

Viceversa, let $(x_n)_{n=0}^{\infty}$ be a Cauchy sequence. Since it is bounded (by Lemma 2.118), Corollary 2.113 implies that there exists a subsequence $(x_{n_k})_{k=0}^{\infty}$ converging to some limit $A \in \mathbb{R}$.

Given $\varepsilon > 0$, choose $N_0 \in \mathbb{N}$ such that

$$|x_n - x_m| < \frac{\varepsilon}{2} \quad \forall m, n \geq N_0,$$

and choose $N_1 \in \mathbb{N}$ such that

$$|x_{n_k} - A| < \frac{\varepsilon}{2} \quad \forall k \geq N_1.$$

Let $N = \max\{N_0, N_1\}$. Since $n_N \geq N$ (see Remark 2.85), for all $n \geq N$ we have

$$|x_n - A| \leq |x_n - x_{n_N}| + |x_{n_N} - A| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

11

Thus $(x_n)_{n=0}^{\infty}$ converges to A . □

EXAMPLE 2.121. — Consider the condition

$$\forall \varepsilon > 0 \exists N \in \mathbb{N} \text{ such that } |x_{n+1} - x_n| < \varepsilon \quad \forall n \geq N,$$

which is equivalent to saying that $\lim_{n \rightarrow \infty} |x_{n+1} - x_n| = 0$. We note that this condition is *not* equivalent to convergence. Indeed, consider the sequence

$$0, 1, 1 + \frac{1}{2}, 2, 2 + \frac{1}{3}, 2 + \frac{2}{3}, 3, 3 + \frac{1}{4}, 3 + \frac{2}{4}, 3 + \frac{3}{4}, 4, 4 + \frac{1}{5}, 4 + \frac{2}{5}, 4 + \frac{3}{5}, 4 + \frac{4}{5}, 5, 5 + \frac{1}{6}, \dots$$

which progresses between consecutive integers $n - 1$ and n in steps of size $\frac{1}{n}$. This sequence is unbounded and hence not convergent, but the distances between successive elements tend to zero. Therefore, the condition $\lim_{n \rightarrow \infty} |x_{n+1} - x_n| = 0$ is insufficient for convergence.

2.5.7 Improper Limits

We now extend the notion of limit to allow the **improper limit values** $+\infty$ (often abbreviated as ∞) and $-\infty$.

DEFINITION 2.122: IMPROPER LIMITS

Let $(x_n)_{n=0}^{\infty}$ be a sequence in \mathbb{R} .

We say that $(x_n)_{n=0}^{\infty}$ **diverges to** $+\infty$, and we write

$$\lim_{n \rightarrow \infty} x_n = +\infty,$$

if for every $M > 0$ there exists $N \in \mathbb{N}$ such that $x_n > M$ for all $n \geq N$.

Similarly, $(x_n)_{n=0}^{\infty}$ **diverges to** $-\infty$ if for every $M > 0$ there exists $N \in \mathbb{N}$ such that $x_n < -M$ for all $n \geq N$.

In both cases, we say that $(x_n)_{n=0}^{\infty}$ has an **improper limit**.

An unbounded sequence need not diverge to $+\infty$ or $-\infty$. For instance, the sequence

$$0, -1, 2, -3, 4, -5, 6, -7, 8, -9, \dots,$$

that is, $x_n = (-1)^n n$, is unbounded but neither diverges to $+\infty$ nor to $-\infty$.

EXERCISE 2.123. — Let $(x_n)_{n=0}^{\infty}$ be an unbounded sequence of real numbers. Show that there exists a subsequence which diverges either to $+\infty$ or to $-\infty$.

11

The notion of improper limit allows us to extend the definitions of superior and inferior limits to *unbounded* sequences. If $(x_n)_{n=0}^{\infty}$ is not bounded from above, then

$$\sup_{k \geq n} x_k = +\infty \quad \forall n \in \mathbb{N},$$

and we write

$$\limsup_{n \rightarrow \infty} x_n = +\infty.$$

If $(x_n)_{n=0}^{\infty}$ is bounded from above but not from below, then we define

$$\limsup_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \left(\sup_{k \geq n} x_k \right),$$

where the right-hand side is a real limit if the decreasing sequence $\sup_{k \geq n} x_k$ is bounded, and the improper limit $-\infty$ otherwise. The definition of \liminf extends analogously.

EXERCISE 2.124. — (Prove the sandwich principle for improper limits.) Let $(x_n)_{n=0}^{\infty}$ and $(y_n)_{n=0}^{\infty}$ be two sequences with $x_n \leq y_n$ for all $n \in \mathbb{N}$. Show that:

$$\lim_{n \rightarrow \infty} x_n = +\infty \implies \lim_{n \rightarrow \infty} y_n = +\infty,$$

$$\lim_{n \rightarrow \infty} y_n = -\infty \implies \lim_{n \rightarrow \infty} x_n = -\infty.$$

2.6 Sequences of Complex Numbers

Informally, a **sequence of complex numbers** is just like a sequence of real numbers, except that each term is a complex number instead of a real one. Thus, we study ordered lists

$$z_0, z_1, z_2, \dots$$

where each element z_n belongs to \mathbb{C} . As in the real case, we are mainly interested in their convergence, divergence, and limit behavior.

To analyze sequences in \mathbb{C} , it is often sufficient to consider separately the corresponding sequences of real and imaginary parts in \mathbb{R} .

DEFINITION 2.125: SEQUENCES OF COMPLEX NUMBERS

A sequence of complex numbers $(z_n)_{n=0}^\infty$, where

$$z_n = x_n + iy_n,$$

is said to **converge** to a limit $A + iB \in \mathbb{C}$ if the two sequences of real numbers $(x_n)_{n=0}^\infty$ and $(y_n)_{n=0}^\infty$ converge to A and B , respectively. In this case, we write

$$\lim_{n \rightarrow \infty} z_n = A + iB.$$

We say that $(z_n)_{n=0}^\infty$ **diverges to ∞** if the sequence of moduli $(|z_n|)_{n=0}^\infty$ diverges to $+\infty$, that is,

$$\lim_{n \rightarrow \infty} |z_n| = \lim_{n \rightarrow \infty} \sqrt{x_n^2 + y_n^2} = +\infty.$$

REMARK 2.126. — As for sequences of real numbers, one can consider subsequences of sequences in \mathbb{C} . Given a strictly increasing sequence of nonnegative integers $(n_k)_{k=0}^\infty$, the corresponding subsequence is

$$(z_{n_k})_{k=0}^\infty = (x_{n_k} + iy_{n_k})_{k=0}^\infty.$$

EXERCISE 2.127. — Let $(z_n)_{n=0}^\infty$ be a convergent sequence in \mathbb{C} . Show that the sequence of moduli $(|z_n|)_{n=0}^\infty$ converges, and determine its limit. Conversely, does the convergence of $(|z_n|)_{n=0}^\infty$ imply the convergence of $(z_n)_{n=0}^\infty$?

EXERCISE 2.128. — Given a complex number $z \in \mathbb{C}$, consider the **geometric sequence** $(z_n)_{n=0}^\infty$ defined by

$$z_n = z^n = \underbrace{z \cdot \dots \cdot z}_{n \text{ times}}.$$

Determine the set of all complex numbers z for which the sequence $(z^n)_{n=0}^{\infty}$ converges.

11

REMARK 2.129. — Most of the properties of convergent sequences in \mathbb{R} remain valid for complex sequences. In particular, limits in \mathbb{C} are unique, and all standard limit rules (linearity, product, quotient, etc.) hold. Each statement can be proved by applying the corresponding property separately to the real and imaginary parts.

Chapter 3

Functions of one Real Variable

In this chapter we study real-valued functions defined on subsets of \mathbb{R} , typically intervals. The central concept is *continuity*.

3.1 Real-valued Functions

3.1.1 Boundedness and Monotonicity

We begin with two elementary properties already encountered for sequences: boundedness and monotonicity. A **real-valued** function is any function with values in \mathbb{R} . We assume the domain is a nonempty subset of \mathbb{R} ; informally, we speak of functions of one real **variable**.

For a nonempty set $D \subseteq \mathbb{R}$, the set of **real-valued** functions on D is

$$\mathcal{F}(D) = \{f \mid f : D \rightarrow \mathbb{R}\}.$$

For $f_1, f_2 \in \mathcal{F}(D)$, $\alpha \in \mathbb{R}$, and $x \in D$ we define

$$(f_1 + f_2)(x) = f_1(x) + f_2(x), \quad (\alpha f_1)(x) = \alpha f_1(x), \quad (f_1 f_2)(x) = f_1(x)f_2(x).$$

11

Given $a \in \mathbb{R}$, we write $f \equiv a$ for the constant function $x \mapsto a$ on D .

REMARK 3.1. — With the operations above, $\mathcal{F}(D)$ is a commutative ring (the additive identity is $f \equiv 0$ and the multiplicative identity is $f \equiv 1$).

A point $x \in D$ is a **zero** of $f \in \mathcal{F}(D)$ if $f(x) = 0$. The **zero set** of f is $\{x \in D \mid f(x) = 0\}$. We order $\mathcal{F}(D)$ pointwise: for $f_1, f_2 \in \mathcal{F}(D)$,

$$f_1 \leq f_2 \iff f_1(x) \leq f_2(x) \quad \forall x \in D,$$

$$f_1 < f_2 \iff f_1(x) < f_2(x) \quad \forall x \in D.$$

We say that $f \in \mathcal{F}(D)$ is **nonnegative** if $f \geq 0$, and **positive** if $f > 0$.

EXERCISE 3.2. — Let $N_1, N_2 \subseteq D$ be the zero sets of $f_1, f_2 \in \mathcal{F}(D)$, respectively. What is the zero set of $f_1 f_2$?

EXERCISE 3.3. — Verify that the relation \leq defined above on $\mathcal{F}(D)$ is an order relation.

DEFINITION 3.4: BOUNDED FUNCTIONS

Let $D \neq \emptyset$ and $f : D \rightarrow \mathbb{R}$. We say that f is **bounded from above** if there exists $M > 0$ such that

$$f(x) \leq M \quad \forall x \in D.$$

We say that f is **bounded from below** if there exists $M > 0$ such that

$$f(x) \geq -M \quad \forall x \in D.$$

We say that f is **bounded** if it is both bounded from above and from below. Equivalently, f is bounded if there exists $M > 0$ such that

$$|f(x)| \leq M \quad \forall x \in D.$$

DEFINITION 3.5: MONOTONE FUNCTIONS

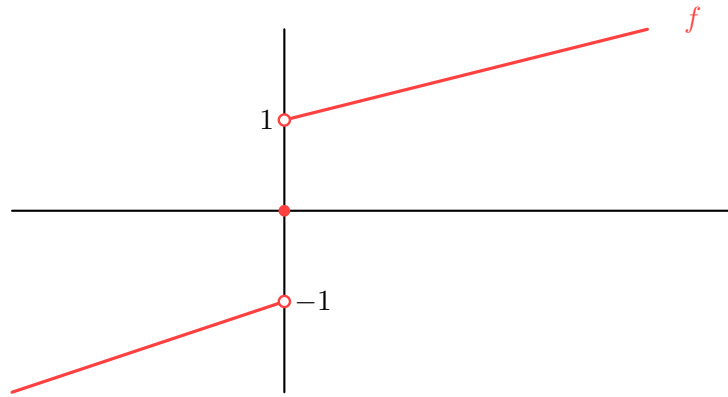
Let $D \subseteq \mathbb{R}$ and $f : D \rightarrow \mathbb{R}$. The function f is:

1. **increasing** if $x < y$ implies $f(x) \leq f(y)$ for all $x, y \in D$;
2. **strictly increasing** if $x < y$ implies $f(x) < f(y)$ for all $x, y \in D$;
3. **decreasing** if $x < y$ implies $f(x) \geq f(y)$ for all $x, y \in D$;
4. **strictly decreasing** if $x < y$ implies $f(x) > f(y)$ for all $x, y \in D$.

We call f **monotone** if it is increasing or decreasing, and **strictly monotone** if it is strictly increasing or strictly decreasing.

EXAMPLE 3.6. — • Let $D = [a, b]$ and $f(x) = x^2$. Then f is strictly increasing if $a \geq 0$, strictly decreasing if $b \leq 0$, and not monotone if $a < 0 < b$.

- For any $D \subseteq \mathbb{R}$ and any odd integer $n \geq 1$, the map $x \mapsto x^n$ on D is strictly increasing.
- The rounding function $\lfloor \cdot \rfloor : \mathbb{R} \rightarrow \mathbb{R}$ (see Definition 2.59) is increasing but not strictly increasing.
- A constant function is both increasing and decreasing. Conversely, a function on $D \subseteq \mathbb{R}$ that is both increasing and decreasing is constant.



11

Figure 3.1: A strictly monotone function is always injective but need not be surjective. For example, $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = \frac{1}{8}x + \operatorname{sgn}(x)$, is strictly increasing but not surjective (e.g. $\frac{1}{2}$ is not in the image).

EXERCISE 3.7. — Let $D \subseteq \mathbb{R}$, and let $f_1, f_2 \in \mathcal{F}(D)$ be strictly increasing. Show that:

- (i) $f_1 + f_2$ is strictly increasing;
- (ii) for $a \in \mathbb{R}$, the function af_1 is strictly increasing if $a > 0$, and strictly decreasing if $a < 0$;
- (iii) if $f_1 > 0$ and $f_2 > 0$, then $f_1 f_2$ is strictly increasing.

3.1.2 Continuity

DEFINITION 3.8: CONTINUOUS FUNCTIONS

Let $D \subseteq \mathbb{R}$ and $f : D \rightarrow \mathbb{R}$. We say that f is **continuous at** $x_0 \in D$ if for all $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\forall x \in D, \quad |x - x_0| < \delta \implies |f(x) - f(x_0)| < \varepsilon.$$

We say that f is **continuous on** D if it is continuous at every point of D .

12

REMARK 3.9. — It suffices to verify the implication above for small ε . Precisely:
Assume there exists $\varepsilon_0 > 0$ such that for every $\varepsilon \in (0, \varepsilon_0]$ there is a $\delta > 0$ with

$$|x - x_0| < \delta \implies |f(x) - f(x_0)| < \varepsilon \quad (\forall x \in D).$$

Then f is continuous at x_0 .

Indeed, for $\varepsilon > \varepsilon_0$ we can choose the number $\delta > 0$ corresponding to ε_0 to get

$$\forall x \in D, \quad |x - x_0| < \delta \implies |f(x) - f(x_0)| < \varepsilon_0 < \varepsilon.$$

In other words, if δ works for ε_0 , then it works for all $\varepsilon > \varepsilon_0$.

The next figure shows a continuous function on $D = [a, b) \cup (c, d] \cup \{e\}$. For every $x_0 \in D$ and every $\varepsilon > 0$, there exists $\delta > 0$ such that all $x \in D$ with $|x - x_0| < \delta$ satisfy $|f(x) - f(x_0)| < \varepsilon$.

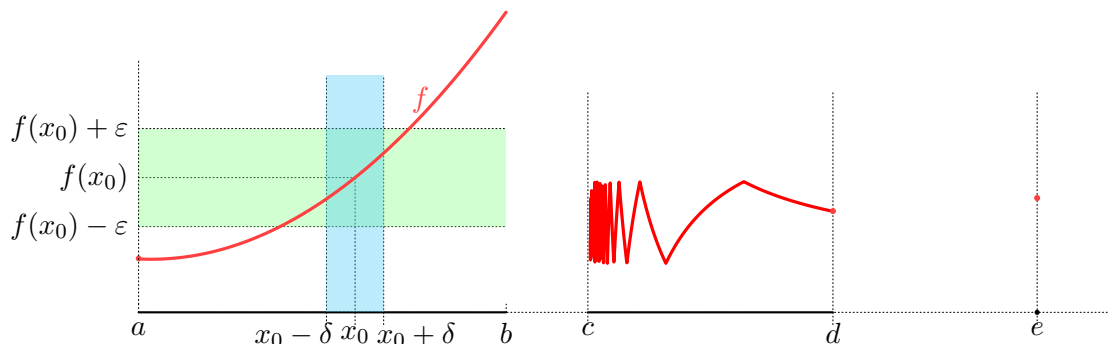


Figure 3.2: A continuous function with domain $D = [a, b) \cup (c, d] \cup \{e\}$.

EXAMPLE 3.10. — • For $a, b \in \mathbb{R}$, the affine function $f(x) = ax + b$ is continuous. Indeed, for $x_0 \in \mathbb{R}$ and $\varepsilon > 0$, set $\delta = \varepsilon/|a|$ (if $a = 0$ then f is constant and we can simply choose $\delta = 1$). Then $|x - x_0| < \delta$ implies

$$|f(x) - f(x_0)| = |(ax + b) - (ax_0 + b)| = |a(x - x_0)| = |a||x - x_0| < \varepsilon.$$

- The absolute value $f(x) = |x|$ is continuous. Indeed, give $\varepsilon > 0$ simply choose $\delta = \varepsilon$. Then $|x - x_0| < \delta$ implies $|f(x) - f(x_0)| = ||x| - |x_0|| \leq |x - x_0| < \varepsilon$ by inverse triangle inequality.
- The rounding function $f(x) = \lfloor x \rfloor$ is not continuous at integers. Indeed, if $x_0 \in \mathbb{Z}$, then for any small $\delta > 0$,

$$|\lfloor x_0 - \frac{\delta}{2} \rfloor - \lfloor x_0 \rfloor| = 1,$$

so the ε - δ condition fails for $\varepsilon < 1$.

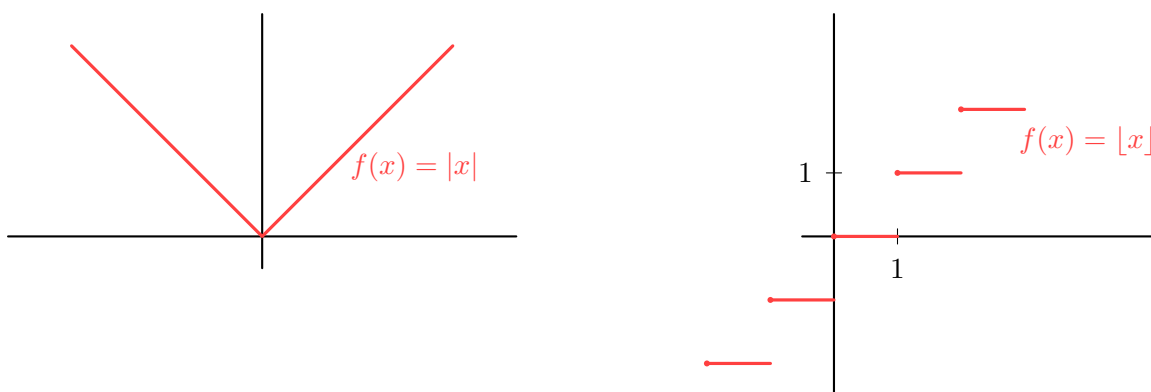


Figure 3.3: The absolute value function (left) and the floor function (right).

EXERCISE 3.11. — Show that $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = x^2$, and $g : [0, \infty) \rightarrow \mathbb{R}$, $g(x) = \sqrt{x}$, are both continuous.

RESTRICTION

Let $D \subseteq \mathbb{R}$ and $f : D \rightarrow \mathbb{R}$. For any $D' \subseteq D$ the **restriction** of f to D' is the function $f|_{D'} : D' \rightarrow \mathbb{R}$ defined by

$$f|_{D'}(x) = f(x) \quad \forall x \in D'.$$

We regard $f|_{D'}$ and f as different functions unless $D' = D$.

EXERCISE 3.12. — Let $D \subseteq \mathbb{R}$ and $f : D \rightarrow \mathbb{R}$ be continuous. If $D' \subseteq D$, show that $f|_{D'}$ is continuous.

PROPOSITION 3.13: COMBINATION OF CONTINUOUS FUNCTIONS

Let $D \subseteq \mathbb{R}$, and let $f_1, f_2 : D \rightarrow \mathbb{R}$ be continuous at $x_0 \in D$. Then $f_1 + f_2$, $f_1 f_2$, and $a f_1$ (for any $a \in \mathbb{R}$) are continuous at x_0 .

Proof. We first prove the result for the sum. Let $\varepsilon > 0$. Since f_1 and f_2 are continuous at x_0 , there exist $\delta_1, \delta_2 > 0$ such that

$$|x - x_0| < \delta_1 \Rightarrow |f_1(x) - f_1(x_0)| < \frac{\varepsilon}{2}, \quad |x - x_0| < \delta_2 \Rightarrow |f_2(x) - f_2(x_0)| < \frac{\varepsilon}{2}.$$

So, choosing $\delta = \min\{\delta_1, \delta_2\}$, for $|x - x_0| < \delta$ we get

$$|(f_1 + f_2)(x) - (f_1 + f_2)(x_0)| \leq |f_1(x) - f_1(x_0)| + |f_2(x) - f_2(x_0)| < \varepsilon,$$

which shows that $f_1 + f_2$ is continuous at x_0 .

For the product, note that

$$\begin{aligned} |f_1(x)f_2(x) - f_1(x_0)f_2(x_0)| &= |f_1(x)f_2(x) - f_1(x_0)f_2(x) + f_1(x_0)f_2(x) - f_1(x_0)f_2(x_0)| \\ &\leq |f_1(x)f_2(x) - f_1(x_0)f_2(x)| + |f_1(x_0)f_2(x) - f_1(x_0)f_2(x_0)| \\ &= |f_2(x)| |f_1(x) - f_1(x_0)| + |f_1(x_0)| |f_2(x) - f_2(x_0)|. \end{aligned}$$

Now, first choose $\delta_0 > 0$ so that $|x - x_0| < \delta_0$ implies $|f_2(x) - f_2(x_0)| < 1$, so that

$$|x - x_0| < \delta_0 \Rightarrow |f_2(x)| < 1 + |f_2(x_0)|.$$

Then choose $\delta_1, \delta_2 > 0$ so that

$$|x - x_0| < \delta_1 \Rightarrow |f_1(x) - f_1(x_0)| < \frac{\varepsilon}{2(1 + |f_2(x_0)|)},$$

$$|x - x_0| < \delta_2 \Rightarrow |f_2(x) - f_2(x_0)| < \frac{\varepsilon}{2(1 + |f_1(x_0)|)}.$$

So, choosing $\delta = \min\{\delta_0, \delta_1, \delta_2\}$, for $|x - x_0| < \delta$ we get

$$\begin{aligned} |f_1(x)f_2(x) - f_1(x_0)f_2(x_0)| &< |f_2(x)| \frac{\varepsilon}{2(1 + |f_2(x_0)|)} + |f_1(x_0)| \frac{\varepsilon}{2(1 + |f_1(x_0)|)} \\ &< (1 + |f_2(x_0)|) \frac{\varepsilon}{2(1 + |f_2(x_0)|)} + |f_1(x_0)| \frac{\varepsilon}{2(1 + |f_1(x_0)|)} \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \end{aligned}$$

thus f_1f_2 is continuous at x_0 .

Finally, the statement about af_1 follows by choosing $f_2 \equiv a$ (a constant function) and using the product case proved above: since f_1 and f_2 are continuous at x_0 , their product $f_1f_2 = af_1$ is continuous at x_0 . \square

SUM AND PRODUCT NOTATION

Let $n \in \mathbb{N}$ and $a_0, a_1, \dots, a_n \in \mathbb{R}$. We use the notation

$$\sum_{j=0}^n a_j = a_0 + a_1 + a_2 + \cdots + a_n, \quad \prod_{j=0}^n a_j = a_0 a_1 a_2 \cdots a_n.$$

Here a_j is a **summand** in the sum and a **factor** in the product; j is the **index** (or **running variable**).

If J is a finite set and numbers $(a_j)_{j \in J}$ are given, we write

$$\sum_{j \in J} a_j \quad \text{and} \quad \prod_{j \in J} a_j.$$

By convention, for the empty index set \emptyset ,

$$\sum_{j \in \emptyset} a_j = 0, \quad \prod_{j \in \emptyset} a_j = 1.$$

EXAMPLE 3.14. — A polynomial is a function of the form

$$p(x) = a_0 + a_1x + \cdots + a_nx^n = \sum_{j=0}^n a_jx^j, \quad a_j \in \mathbb{R}.$$

Since $x \mapsto x$ is continuous and products/sums of continuous functions are continuous (Proposition 3.13), every polynomial is continuous.

COMPOSITION OF FUNCTIONS

Let $f : X \rightarrow Y$ and $g : Y \rightarrow Z$. The **composition** is $g \circ f : X \rightarrow Z$, defined by $(g \circ f)(x) = g(f(x))$ for all $x \in X$.

Associativity. If $f : W \rightarrow X$, $g : X \rightarrow Y$, and $h : Y \rightarrow Z$, then

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

Indeed, for all $w \in W$,

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

Therefore, we may omit parentheses and write $h \circ g \circ f : W \rightarrow Z$.

PROPOSITION 3.15: COMPOSITION OF CONTINUOUS FUNCTIONS

Let $D_1, D_2 \subseteq \mathbb{R}$, $x_0 \in D_1$, and $f : D_1 \rightarrow D_2$ be continuous at x_0 . If $g : D_2 \rightarrow \mathbb{R}$ is continuous at $f(x_0)$, then $g \circ f : D_1 \rightarrow \mathbb{R}$ is continuous at x_0 . In particular, the composition of continuous functions is continuous.

Proof. Let $\varepsilon > 0$. By continuity of g at $f(x_0)$, there exists $\eta > 0$ such that

$$\forall y \in D_2, \quad |y - f(x_0)| < \eta \Rightarrow |g(y) - g(f(x_0))| < \varepsilon.$$

By continuity of f at x_0 , there exists $\delta > 0$ such that

$$\forall x \in D_1, \quad |x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \eta.$$

Combining the implications gives, for $x \in D_1$,

$$|x - x_0| < \delta \implies |f(x) - f(x_0)| < \eta \implies |g(f(x)) - g(f(x_0))| < \varepsilon.$$

□

REMARK 3.16. — Applying Proposition 3.15 with $g(y) = |y|$ (see Example 3.10), we see that if $f : D \rightarrow \mathbb{R}$ is continuous then $x \mapsto |f(x)|$ is continuous.

EXERCISE 3.17. — Show that $f : \mathbb{R} \setminus \{0\} \rightarrow \mathbb{R}$, $x \mapsto \frac{1}{x}$, is continuous. Deduce that if $g : D \rightarrow \mathbb{R}$ is continuous and has no zeros on D , then $x \mapsto \frac{1}{g(x)}$ is continuous on D . Conclude that $x \mapsto \frac{h(x)}{g(x)}$ is continuous on D whenever $h, g : D \rightarrow \mathbb{R}$ are continuous and g has no zeros on D .

EXERCISE 3.18. — Let $a < b < c$ and let $f_1 : [a, b] \rightarrow \mathbb{R}$, $f_2 : [b, c] \rightarrow \mathbb{R}$ be continuous. Define $f : [a, c] \rightarrow \mathbb{R}$ by

$$f(x) = \begin{cases} f_1(x), & x \in [a, b), \\ f_2(x), & x \in [b, c]. \end{cases}$$

Show that f is continuous if and only if $f_1(b) = f_2(b)$.

EXERCISE 3.19. — Let $I \subset \mathbb{R}$ be an open interval and let $f : I \rightarrow \mathbb{R}$ be a function. Show that f is continuous if and only if $f^{-1}(U)$ is open for every open set $U \subset \mathbb{R}$.

EXERCISE 3.20. — Let $D \subseteq \mathbb{R}$ and $f : D \rightarrow \mathbb{R}$.

1. If f is continuous at $x_0 \in D$, then there exist an open neighbourhood U of x_0 and $M > 0$ such that $|f(x)| \leq M$ for all $x \in D \cap U$.
2. If f is continuous at $x_0 \in D$ and $f(x_0) \neq 0$, then there exists an open neighbourhood U of x_0 such that $f(x)f(x_0) > 0$ for all $x \in D \cap U$ (in other words, $f(x)$ and $f(x_0)$ have the same sign).

3.1.3 Sequential Continuity

Continuity admits a sequential characterization: a function $f : D \rightarrow \mathbb{R}$ is continuous if and only if it sends every convergent sequence in D to a convergent sequence with the corresponding limit. This is called **sequential continuity**. We first fix a notation.

NOTATION FOR LIMITS OF SEQUENCES

Let $(x_n)_{n=0}^{\infty} \subseteq \mathbb{R}$ and $\bar{x} \in \mathbb{R}$. We write

$$x_n \rightarrow \bar{x} \quad \text{or} \quad x_n \xrightarrow[n \rightarrow \infty]{} \bar{x}$$

to mean

$$\lim_{n \rightarrow \infty} x_n = \bar{x}.$$

THEOREM 3.21: CONTINUITY = SEQUENTIAL CONTINUITY

Let $D \subseteq \mathbb{R}$, $f : D \rightarrow \mathbb{R}$, and $\bar{x} \in D$. Then f is continuous at \bar{x} if and only if for every sequence $(x_n)_{n=0}^{\infty} \subset D$ with $x_n \rightarrow \bar{x}$ we have $f(x_n) \rightarrow f(\bar{x})$.

Proof. Assume that f is continuous at \bar{x} . Then, given $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\forall x \in D, \quad |x - \bar{x}| < \delta \Rightarrow |f(x) - f(\bar{x})| < \varepsilon.$$

Also, since $x_n \rightarrow \bar{x}$, there exists $N \in \mathbb{N}$ such that

$$n \geq N \implies |x_n - \bar{x}| < \delta.$$

Thus

$$n \geq N \implies |f(x_n) - f(\bar{x})| < \varepsilon,$$

which implies that the sequence $(f(x_n))_{n=0}^{\infty}$ converges to $f(\bar{x})$.

To prove the converse, assume that f is not continuous at \bar{x} . This means that there exists $\varepsilon > 0$ such that, for every $\delta > 0$, there is $x \in D$ with

$$|x - \bar{x}| < \delta \quad \text{and} \quad |f(x) - f(\bar{x})| \geq \varepsilon.$$

12

Now, for every $n \in \mathbb{N}$, we apply this property with $\delta = 2^{-n}$ to find a point $x_n \in D$ such that

$$|x_n - \bar{x}| < 2^{-n} \quad \text{and} \quad |f(x_n) - f(\bar{x})| \geq \varepsilon.$$

Then the sequence constructed in this way satisfies $x_n \rightarrow \bar{x}$ but $f(x_n) \not\rightarrow f(\bar{x})$. \square

REMARK 3.22. — The proof above shows the following:

If $f : D \rightarrow \mathbb{R}$ is not continuous at \bar{x} , then there exists $\varepsilon > 0$ and a sequence $(x_n)_{n=0}^{\infty} \subset D$ with $x_n \rightarrow \bar{x}$ such that $|f(x_n) - f(\bar{x})| \geq \varepsilon$ for all $n \in \mathbb{N}$.

EXERCISE 3.23. — Let $D \subseteq \mathbb{R}$ and $f : D \rightarrow \mathbb{R}$ be continuous. Suppose $(x_n)_{n=0}^{\infty}$ is a sequence in D such that $(f(x_n))_{n=0}^{\infty}$ converges. Must $(x_n)_{n=0}^{\infty}$ converge?

3.2 Continuous Functions

3.2.1 The Intermediate Value Theorem

In this section we prove a fundamental theorem that formalises the idea that the graph of a continuous function on an interval is a continuous curve, and thus cannot make any jumps. We show that a continuous function f on an interval $[a, b]$ contained in its domain takes all intermediate values between $f(a)$ and $f(b)$. As we shall see, the proof relies on the existence of the supremum, and therefore on the completeness axiom.

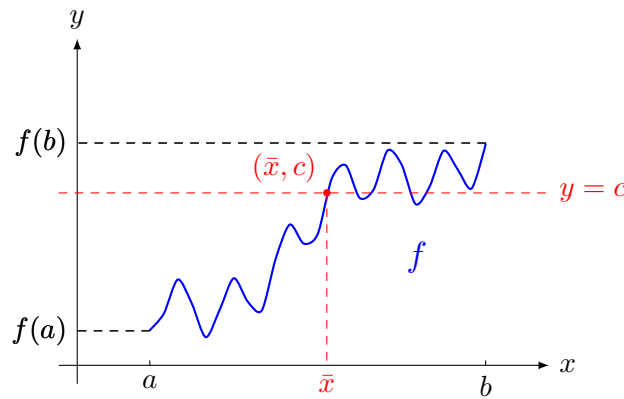


Figure 3.4: A continuous function $f : [a, b] \rightarrow \mathbb{R}$ must cross every horizontal line between $f(a)$ and $f(b)$.

THEOREM 3.24: INTERMEDIATE VALUE THEOREM

Let $f : [a, b] \rightarrow \mathbb{R}$ be a continuous function with $f(a) \leq f(b)$. Then, for every real number c with $f(a) \leq c \leq f(b)$, there exists $\bar{x} \in [a, b]$ such that $f(\bar{x}) = c$.

Proof. Fix $c \in [f(a), f(b)]$, and define

$$X = \{x \in [a, b] \mid f(x) \leq c\}.$$

Since $a \in X$ and $X \subseteq [a, b]$, the set X is nonempty and bounded from above. By Theorem 2.55, its supremum

$$\bar{x} = \sup(X) \in [a, b]$$

exists. We now use the continuity of f at \bar{x} to show that $f(\bar{x}) = c$.

Since \bar{x} is the supremum of X , for each $n \geq 0$ we can find a point $x_n \in X \cap [\bar{x} - 2^{-n}, \bar{x}]$. Then $|x_n - \bar{x}| \leq 2^{-n}$, hence $x_n \rightarrow \bar{x}$. Also, by the definition of X , $f(x_n) \leq c$. Thus, by Theorem 3.21 (continuity of f along sequences),

$$f(\bar{x}) = \lim_{n \rightarrow \infty} f(x_n),$$

and Proposition 2.95 yields $\lim_{n \rightarrow \infty} f(x_n) \leq c$. Therefore $f(\bar{x}) \leq c$.

Suppose, by contradiction, that $f(\bar{x}) < c$ and set $\varepsilon := c - f(\bar{x}) > 0$. By continuity at \bar{x} , there exists $\delta > 0$ such that for all $x \in [a, b]$,

$$|x - \bar{x}| < \delta \implies |f(x) - f(\bar{x})| < \varepsilon,$$

hence $f(x) < f(\bar{x}) + \varepsilon = c$. Therefore, by the definition of X ,

$$(\bar{x} - \delta, \bar{x} + \delta) \cap [a, b] \subset X.$$

Moreover, since $f(\bar{x}) < c \leq f(b)$, we cannot have $\bar{x} = b$; hence $\bar{x} < b$.

Because $\bar{x} < b$, the interval $(\bar{x}, \bar{x} + \delta) \cap [a, b] \subset X$ is nonempty. Pick

$$y \in (\bar{x}, \bar{x} + \delta) \cap [a, b] \subset X.$$

Then $y \in X$ and $y > \bar{x}$, which contradicts the defining property of the supremum: \bar{x} is an upper bound of X , so X cannot contain elements larger than \bar{x} . This contradiction shows that $f(\bar{x}) \geq c$.

Together with $f(\bar{x}) \leq c$ proved above, we conclude $f(\bar{x}) = c$, as desired. \square

REMARK 3.25. — If $f : [a, b] \rightarrow \mathbb{R}$ is continuous with $f(a) \geq f(b)$, the theorem still holds in the following form:

For every real number c with $f(a) \geq c \geq f(b)$ there exists $\bar{x} \in [a, b]$ such that $f(\bar{x}) = c$.

To prove this, one can:

1. either repeat the previous proof with $X = \{x \in [a, b] \mid f(x) \geq c\}$;
2. or apply Theorem 3.24 to the function $g = -f$.

EXERCISE 3.26. — Let I be a non-empty interval and $f : I \rightarrow \mathbb{R}$ a continuous injective function. Show that f is strictly monotone.

3.2.2 Inverse Function Theorem

IDENTITY AND INVERSE FUNCTION

Given a set X , the **identity function** $\text{id}_X : X \rightarrow X$ is defined by

$$\text{id}_X(x) = x \quad \text{for every } x \in X.$$

If $f : X \rightarrow Y$ is bijective, there exists a unique function $g : Y \rightarrow X$ such that, for each $y \in Y$, the value $g(y)$ is the unique element $x \in X$ satisfying $f(x) = y$. With this definition,

$$g \circ f = \text{id}_X \quad \text{and} \quad f \circ g = \text{id}_Y.$$

The function g is called the **inverse function** (or **inverse mapping**) of f , and is denoted by f^{-1} .

REMARK 3.27. — A function $f : X \rightarrow Y$ is bijective if and only if there exists a function $g : Y \rightarrow X$ such that $g \circ f = \text{id}_X$ and $f \circ g = \text{id}_Y$.

In this subsection we show that every continuous strictly monotone function has an inverse function that is also continuous.

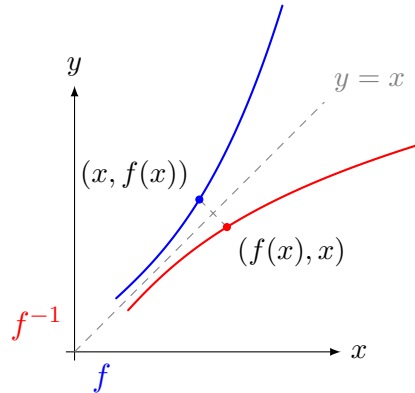


Figure 3.5: The graphs of a strictly increasing function f and its inverse f^{-1} are reflections across the line $y = x$. The points $(x, f(x))$ and $(f(x), x)$ are mirror images.

THEOREM 3.28: INVERSE FUNCTION THEOREM

Let I be an interval and $f : I \rightarrow \mathbb{R}$ a continuous strictly monotone function. Then $f(I)$ is an interval, and the mapping $f : I \rightarrow f(I)$ has a continuous strictly monotone inverse function $f^{-1} : f(I) \rightarrow I$.

Proof. We may assume that I is non-empty and not a single point. Also, without loss of generality, suppose f is strictly increasing (otherwise replace f with $-f$).

Let $J = f(I)$. Since f is strictly monotone, it is injective. Also, since by definition $J = f(I)$, it is surjective, hence bijective. Therefore there exists a unique inverse $g = f^{-1} : J \rightarrow I$.

Because f is strictly increasing, we have

$$x_1 < x_2 \iff f(x_1) < f(x_2) \quad \forall x_1, x_2 \in I. \quad (3.1)$$

Defining $y_1 = f(x_1)$ and $y_2 = f(x_2)$, this is equivalent to

$$y_1 < y_2 \iff g(y_1) < g(y_2) \quad \forall y_1, y_2 \in J.$$

Thus g is strictly increasing.

To show that J is an interval, $y_1, y_2 \in J$, and assume without loss of generality that $y_1 < y_2$. Since $J = f(I)$, (3.1) implies that $y_1 = f(x_1)$ and $y_2 = f(x_2)$ for some $x_1, x_2 \in I$ with $x_1 < x_2$. Now, by the Intermediate Value Theorem 3.24 applied to $f : [x_1, x_2] \rightarrow \mathbb{R}$, we have that all values $c \in [y_1, y_2]$ are in the image of $f : [x_1, x_2] \rightarrow \mathbb{R}$, that is

$$[y_1, y_2] \subset f([x_1, x_2]) \subset J.$$

Since $y_1 < y_2$ were two arbitrary points in J , this proves that J is an interval.

It remains to show that $g = f^{-1}$ is continuous. Fix $\bar{y} \in J$, and suppose by contradiction that g is not continuous at \bar{y} . Then, by Remark 3.22, there exist $\varepsilon > 0$ and a sequence $(y_n)_{n=0}^\infty \subset J$ such that

$$y_n \rightarrow \bar{y} \quad \text{but} \quad |g(y_n) - g(\bar{y})| \geq \varepsilon \text{ for all } n \in \mathbb{N}. \quad (3.2)$$

Set $x_n = g(y_n) \in I$ and $\bar{x} = g(\bar{y})$. Then for every $n \in \mathbb{N}$,

$$\text{either } x_n \leq \bar{x} - \varepsilon \quad \text{or} \quad x_n \geq \bar{x} + \varepsilon.$$

In particular, at least one of these cases must occur infinitely often. Without loss of generality, assume $x_n \leq \bar{x} - \varepsilon$ for infinitely many n , and extract a subsequence $(x_{n_k})_{k=0}^\infty$ with $x_{n_k} \leq \bar{x} - \varepsilon$ for all k . Since I is an interval, $\bar{x} - \varepsilon \in I$, and by strict monotonicity of f we obtain

$$y_{n_k} = f(x_{n_k}) \leq f(\bar{x} - \varepsilon) < f(\bar{x}) = \bar{y}.$$

Taking the limit and using Proposition 2.95 gives (recall that $y_n \rightarrow \bar{y}$, see (3.2))

$$\bar{y} = \lim_{k \rightarrow \infty} y_{n_k} \leq f(\bar{x} - \varepsilon) < f(\bar{x}) = \bar{y},$$

a contradiction. Hence g is continuous. □

EXAMPLE 3.29. — Let $n \in \mathbb{N}$, $n \geq 1$. The function $f : [0, \infty) \rightarrow [0, \infty)$ defined by $f(x) = x^n$ is continuous, strictly increasing, and surjective. By the Inverse Function Theorem, there exists a continuous strictly increasing inverse $[0, \infty) \rightarrow [0, \infty)$, denoted by

$$x \mapsto \sqrt[n]{x},$$

called the **n -th root**. Furthermore, for $m, n \in \mathbb{N}$ with $n \geq 1$, we define

$$x^{\frac{m}{n}} = \underbrace{\sqrt[n]{x} \cdot \dots \cdot \sqrt[n]{x}}_{m \text{ times}} \quad \text{for } x \in [0, \infty),$$

and, using Exercise 3.17,

$$x^{-\frac{m}{n}} = \frac{1}{x^{\frac{m}{n}}} \quad \text{for } x \in (0, \infty).$$

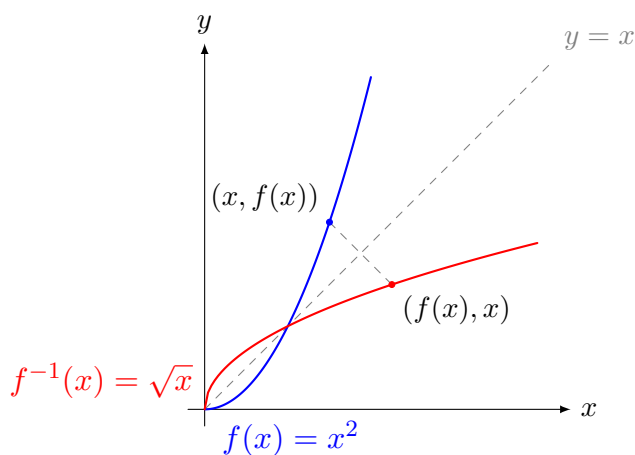


Figure 3.6: The function $f(x) = x^2$ and its inverse $f^{-1}(x) = \sqrt{x}$ are reflections of each other across the line $y = x$.

3.3 Continuous Functions on Compact Intervals

In this section we show that continuous functions on **bounded closed** intervals — called **compact intervals** — enjoy special properties.

3.3.1 Boundedness and Extrema

The key property of compact intervals, which will be used several times, is the following:

LEMMA 3.30: COMPACTNESS

Let $[a, b]$ be a compact interval, and let $(x_n)_{n=0}^{\infty}$ be a sequence contained in $[a, b]$. Then there exists a subsequence $(x_{n_k})_{k=0}^{\infty}$ such that

$$\lim_{k \rightarrow \infty} x_{n_k} = \bar{x} \quad \text{for some } \bar{x} \in [a, b].$$

Proof. Since $(x_n)_{n=0}^{\infty}$ is bounded (as it lies in $[a, b]$), Corollary 2.113 ensures the existence of a convergent subsequence $(x_{n_k})_{k=0}^{\infty}$. Let \bar{x} denote its limit. Because $a \leq x_{n_k} \leq b$ for all k , Proposition 2.95 yields $a \leq \bar{x} \leq b$. \square

THEOREM 3.31: BOUNDEDNESS

Let $[a, b]$ be a compact interval, and let $f : [a, b] \rightarrow \mathbb{R}$ be continuous. Then f is bounded.

Proof. Assume by contradiction that f is unbounded. Then, for every $n \in \mathbb{N}$, there exists $x_n \in [a, b]$ such that $|f(x_n)| \geq n$. By Lemma 3.30, there is a subsequence $(x_{n_k})_{k=0}^{\infty}$ converging to some $\bar{x} \in [a, b]$.

Since f is continuous, so is $|f|$ (recall Remark 3.16), therefore $|f(x_{n_k})| \rightarrow |f(\bar{x})| \in \mathbb{R}$. This contradicts $|f(x_{n_k})| \geq n_k \rightarrow \infty$, so f must be bounded. \square

EXERCISE 3.32. — Find examples of:

1. a continuous but unbounded function on a bounded *open* interval;
2. a continuous but unbounded function on an *unbounded closed* interval;
3. a function unbounded on a compact interval but discontinuous at only one point.

DEFINITION 3.33: EXTREME VALUES

Let $D \subset \mathbb{R}$ and $f : D \rightarrow \mathbb{R}$.

- We say that f takes its **maximum value** at $x_0 \in D$ if $f(x) \leq f(x_0)$ for all $x \in D$. Then $f(x_0)$ is the **maximum** of f .
- We say that f takes its **minimum value** at $x_0 \in D$ if $f(x) \geq f(x_0)$ for all $x \in D$. Then $f(x_0)$ is the **minimum** of f .

Maxima and minima are called **extreme values** or **extrema**.

THEOREM 3.34: EXTREME VALUE THEOREM

Let $[a, b]$ be a compact interval, and let $f : [a, b] \rightarrow \mathbb{R}$ be continuous. Then f attains both its maximum and its minimum.

Proof. Theorem 3.31 guarantees that f is bounded, or equivalently, that $f([a, b]) \subset \mathbb{R}$ is a bounded subset of \mathbb{R} . Thus, Theorem 2.55 implies that

$$S := \sup f([a, b])$$

exists. By definition of the supremum, for each $n \in \mathbb{N}$ there exists $y_n \in f([a, b])$ such that $S - 2^{-n} \leq y_n \leq S$. Hence $y_n \rightarrow S$. Also, since $y_n \in f([a, b])$, there exists $x_n \in [a, b]$ such that $f(x_n) = y_n$.

Now, by Lemma 3.30, we can find subsequence $(x_{n_k})_{k=0}^{\infty}$ such that $x_{n_k} \rightarrow \bar{x} \in [a, b]$. By continuity of f ,

$$f(\bar{x}) = \lim_{k \rightarrow \infty} f(x_{n_k}) = \lim_{k \rightarrow \infty} y_{n_k} = S,$$

so f attains its maximum at \bar{x} .

Applying the same reasoning to $-f$ shows that f also attains its minimum. \square

EXERCISE 3.35. — Does every continuous function f on the open interval $(0, 1)$ attain its maximum?

3.3.2 Uniform Continuity**DEFINITION 3.36: UNIFORM CONTINUITY**

Let $D \subseteq \mathbb{R}$. A function $f : D \rightarrow \mathbb{R}$ is **uniformly continuous** if, for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|x - y| < \delta \implies |f(x) - f(y)| < \varepsilon \quad \text{for all } x, y \in D.$$

REMARK 3.37. — The difference between the usual definition of continuity and the one of uniform continuity lies in how the choice of δ depends on the points considered.

For a function that is continuous at each $x_0 \in D$, the δ in the definition may depend on both ε and x_0 : for every $\varepsilon > 0$ and each x_0 , we can find a $\delta = \delta(\varepsilon, x_0)$ that works *near* x_0 .

Uniform continuity is stronger: there exists a single $\delta = \delta(\varepsilon)$ that works *simultaneously* for all $x, y \in D$. In other words, the control on the variation of f does not deteriorate as we move along the domain. This property is automatically satisfied on compact intervals for continuous functions, as we will prove below.

EXAMPLE 3.38. — We know that the function $f(x) = x^2$ is continuous on \mathbb{R} . We now prove that it is not uniformly continuous. Indeed, fix $\varepsilon = 1$. We will show that for every $\delta > 0$ there exist $x_\delta, y_\delta \in \mathbb{R}$ with

$$|x_\delta - y_\delta| < \delta \quad \text{and} \quad |f(x_\delta) - f(y_\delta)| \geq 1.$$

Set

$$y_\delta = \frac{1}{\delta}, \quad x_\delta = y_\delta + \frac{\delta}{2}.$$

Then $|x_\delta - y_\delta| = \frac{\delta}{2} < \delta$, and

$$|f(x_\delta) - f(y_\delta)| = x_\delta^2 - y_\delta^2 = \left(\frac{1}{\delta} + \frac{\delta}{2}\right)^2 - \frac{1}{\delta^2} = 1 + \frac{\delta^2}{4} \geq 1.$$

This shows that no δ works for $\varepsilon = 1$, so f is not uniformly continuous on \mathbb{R} .

14

THEOREM 3.39: UNIFORM CONTINUITY ON COMPACT INTERVALS

Let $[a, b]$ be a compact interval, and $f : [a, b] \rightarrow \mathbb{R}$ continuous. Then f is uniformly continuous.

Proof. Assume, by contradiction, that f is not uniformly continuous on $[a, b]$. Then there exists $\varepsilon > 0$ such that for every $\delta > 0$ one can find $x, y \in [a, b]$ with

$$|x - y| < \delta \quad \text{and} \quad |f(x) - f(y)| \geq \varepsilon.$$

Taking $\delta = 2^{-n}$ for each $n \in \mathbb{N}$, we obtain sequences $(x_n)_{n=0}^\infty$ and $(y_n)_{n=0}^\infty$ in $[a, b]$ with

$$|x_n - y_n| < 2^{-n} \quad \text{and} \quad |f(x_n) - f(y_n)| \geq \varepsilon. \quad (3.3)$$

By Lemma 3.30, the sequence $(x_n)_{n=0}^\infty$ has a subsequence $(x_{n_k})_{k=0}^\infty$ converging to some $\bar{x} \in [a, b]$. Then

$$|y_{n_k} - \bar{x}| \leq |y_{n_k} - x_{n_k}| + |x_{n_k} - \bar{x}| < 2^{-n_k} + |x_{n_k} - \bar{x}| \xrightarrow[k \rightarrow \infty]{} 0,$$

so $y_{n_k} \rightarrow \bar{x}$ as well. Thus, by continuity of f and Theorem 3.21,

$$\lim_{k \rightarrow \infty} f(x_{n_k}) = \lim_{k \rightarrow \infty} f(y_{n_k}) = f(\bar{x}),$$

therefore

$$|f(x_{n_k}) - f(y_{n_k})| \leq |f(x_{n_k}) - f(\bar{x})| + |f(\bar{x}) - f(y_{n_k})| \xrightarrow[k \rightarrow \infty]{} 0,$$

which contradicts (3.3). Hence, f is uniformly continuous on $[a, b]$. \square

EXERCISE 3.40. — Does Theorem 3.39 remain true for continuous functions on the open interval $(0, 1)$?

EXERCISE 3.41. — In this exercise, we introduce another notion of continuity: **Lipschitz continuity**.

14

1. A function $f : D \rightarrow \mathbb{R}$ on a set $D \subset \mathbb{R}$ is **Lipschitz continuous** if there exists $L \geq 0$ such that

$$|f(x) - f(y)| \leq L|x - y| \quad \forall x, y \in D.$$

Give examples of Lipschitz continuous functions, and show that any such function is uniformly continuous.

2. Let $f(x) = \sqrt{x}$ on $\mathbb{R}_{\geq 0}$. Show that:

- (i) $f_{[0,1]} : [0, 1] \rightarrow \mathbb{R}$ is not Lipschitz continuous;
- (ii) $f_{[1,\infty)} : [1, \infty) \rightarrow \mathbb{R}$ is Lipschitz continuous;
- (iii) $f : [0, \infty) \rightarrow \mathbb{R}$ is uniformly continuous.

3.4 Example: Exponential and Logarithmic Functions

In this section, we will use the notion of convergence and limits of sequences to define the exponential function and prove some of its basic properties, such as continuity.

3.4.1 Definition of the Exponential Function

We start with a preliminary lemma that will be used repeatedly in this section.

LEMMA 3.42: BERNOULLI'S INEQUALITY

For all $a \in \mathbb{R}$ with $a \geq -1$ and all $n \in \mathbb{N}$ with $n \geq 1$, it holds that

$$(1 + a)^n \geq 1 + na.$$

Proof. We proceed by induction. For $n = 1$ we have $(1 + a)^1 = 1 + a = 1 + 1 \cdot a$.

Now assume that the inequality holds for some $n \geq 1$. Since $1 + a \geq 0$ by assumption, we find

$$(1 + a)^{n+1} = (1 + a)^n(1 + a) \geq (1 + na)(1 + a) = 1 + na + a + na^2 \geq 1 + (n + 1)a,$$

which establishes the induction step and completes the proof. \square

14

REMARK 3.43. — Using Bernoulli's inequality and Archimedes' principle one can prove the following statement: *For all $x > 1$ and all $y \in \mathbb{R}$, there exists $n \in \mathbb{N}$ such that $x^n \geq y$.*

Indeed, assume $y > 0$ (otherwise the inequality $x^n \geq y$ is true for every $n \in \mathbb{N}$) and apply Bernoulli's inequality with $a = x - 1 > 0$ to obtain $(1 + a)^n \geq 1 + na \geq na$. Choosing $n \geq \frac{y}{a}$ (using Archimedes' principle), it follows that $x^n \geq y$.

PROPOSITION 3.44: EXISTENCE OF THE EXPONENTIAL

Let $x \in \mathbb{R}$. The sequence $(a_n)_{n=1}^{\infty}$ defined by

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

is convergent, and its limit is a positive real number.

To prove this result, we first establish monotonicity.

LEMMA 3.45: MONOTONICITY

Given $x \in \mathbb{R}$, let $n_0 \in \mathbb{N}$ satisfy $n_0 \geq 1$ and $n_0 > -x$. Then the sequence $(a_n)_{n=n_0}^{\infty}$ defined in Proposition 3.44 is increasing.

Proof. Note that

$$\begin{aligned}\frac{a_{n+1}}{a_n} &= \frac{\left(1 + \frac{x}{n+1}\right)^{n+1}}{\left(1 + \frac{x}{n}\right)^n} = \left(1 + \frac{x}{n}\right) \left(\frac{1 + \frac{x}{n+1}}{1 + \frac{x}{n}}\right)^{n+1} \\ &= \frac{n+x}{n} \left(\frac{n(n+1+x)}{(n+1)(n+x)}\right)^{n+1} = \frac{n+x}{n} \left(1 - \frac{x}{(n+1)(n+x)}\right)^{n+1}.\end{aligned}$$

Also, for $n \geq n_0$ we have $n+x > 0$, hence

$$\frac{x}{(n+1)(n+x)} \leq \frac{x+n}{(n+1)(n+x)} = \frac{1}{n+1} \leq 1,$$

or equivalently

$$-\frac{x}{(n+1)(n+x)} \geq -1 \quad \forall n \geq n_0.$$

Hence, by Bernoulli's inequality (Lemma 3.42) applied with $a = -\frac{x}{(n+1)(n+x)}$ and $n+1$ in place of n , we get

$$\begin{aligned}\frac{a_{n+1}}{a_n} &= \frac{n+x}{n} \left(1 - \frac{x}{(n+1)(n+x)}\right)^{n+1} \geq \frac{n+x}{n} \left(1 - (n+1) \frac{x}{(n+1)(n+x)}\right) \\ &= \frac{n+x}{n} \left(1 - \frac{x}{n+x}\right) = 1,\end{aligned}$$

therefore $a_n \leq a_{n+1}$ for all $n \geq n_0$. □

14

Proof of Proposition 3.44. Fix $x \in \mathbb{R}$, and let $n_0 \in \mathbb{N}$ satisfy $n_0 \geq 1$ and $n_0 > -x$. By Lemma 3.45, the sequence $(a_n)_{n=n_0}^\infty$ is increasing. If we can show that it is bounded, Theorem 2.104 ensures convergence.

Case 1: $x \leq 0$. In this case, $-x = |x|$ and $n_0 > |x|$. Then, for all $n \geq n_0 > |x|$,

$$0 < 1 + \frac{x}{n} \leq 1,$$

so

$$0 < \left(1 + \frac{x}{n}\right)^n \leq 1.$$

Hence $(a_n)_{n=n_0}^\infty$ is an increasing sequence bounded above by 1, and therefore convergent. In particular,

$$\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = \sup_{n \geq n_0} \left(1 + \frac{x}{n}\right)^n > 0.$$

Case 2: $x > 0$. For $n > x$ we have

$$\left(1 + \frac{x}{n}\right)^n \left(1 - \frac{x}{n}\right)^n = \left(1 - \frac{x^2}{n^2}\right)^n \leq 1,$$

so

$$1 \leq \left(1 + \frac{x}{n}\right)^n \leq \left(1 - \frac{x}{n}\right)^{-n} = \left(1 + \frac{(-x)}{n}\right)^{-n}.$$

In other words, if we define $b_n = \left(1 + \frac{(-x)}{n}\right)^n$, then

$$1 \leq a_n \leq \frac{1}{b_n} \quad \forall n > x. \quad (3.4)$$

Since b_n converges to a positive limit by Case 1, Proposition 2.94(4) implies that the sequence $\left(\frac{1}{b_n}\right)_{n>x}$ also converges, and therefore it is bounded (Lemma 2.100). Recalling (3.4), we conclude that the increasing sequence $(a_n)_{n=1}^\infty$ is also bounded, and therefore convergent. \square

DEFINITION 3.46: EXPONENTIAL FUNCTION

The **exponential function** $\exp : \mathbb{R} \rightarrow \mathbb{R}_{>0}$ is defined by

$$\exp(x) = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n \quad \forall x \in \mathbb{R}.$$

The **Euler number** is defined as

$$e = \exp(1) = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n. \quad (3.5)$$

Its numerical value is

$$e = 2.71828\,18284\,59045\,23536\,02874\,71352\,66249\,77572\,47093\,69995\ldots$$

A useful consequence of Lemma 3.45 is the following bound.

COROLLARY 3.47: GROWTH OF THE EXPONENTIAL

Given $n \in \mathbb{N}$ with $n \geq 1$, the exponential function satisfies

$$\exp(x) \geq \left(1 + \frac{x}{n}\right)^n \quad \forall x > -n.$$

Proof. By Lemma 3.45 and Definition 3.46, for $x > -n$ we have

$$a_n \leq a_{n+1} \leq \cdots \leq \exp(x).$$

\square

3.4.2 Properties of the Exponential Function

THEOREM 3.48: PROPERTIES OF THE EXPONENTIAL FUNCTION

The exponential function $\exp : \mathbb{R} \rightarrow \mathbb{R}_{>0}$ is bijective, strictly increasing, and continuous. Moreover,

$$\exp(0) = 1, \quad (3.6)$$

$$\exp(-x) = \exp(x)^{-1}, \quad (3.7)$$

$$\exp(x + y) = \exp(x) \exp(y), \quad (3.8)$$

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

Proof. We first verify the identities (3.6), (3.7), and (3.8).

1. *Proof of (3.6).* By definition,

$$\exp(0) = \lim_{n \rightarrow \infty} \left(1 + \frac{0}{n}\right)^n = \lim_{n \rightarrow \infty} 1^n = 1.$$

2. *Proof of (3.7).* Using Proposition 2.94(2) and the definition of \exp ,

$$\exp(x) \exp(-x) = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n \cdot \lim_{n \rightarrow \infty} \left(1 - \frac{x}{n}\right)^n = \lim_{n \rightarrow \infty} \left(1 - \frac{x^2}{n^2}\right)^n.$$

For $n \geq |x|$ we have $-\frac{x^2}{n^2} \geq -1$, so by Bernoulli's inequality (Lemma 3.42),

$$1 - \frac{x^2}{n} \leq \left(1 - \frac{x^2}{n^2}\right)^n \leq 1 \quad \forall n \geq |x|.$$

Since the left-hand side tends to 1, by Lemma 2.97 we get

$$\lim_{n \rightarrow \infty} \left(1 - \frac{x^2}{n^2}\right)^n = 1,$$

which implies $\exp(x) \exp(-x) = 1$ and therefore $\exp(-x) = \exp(x)^{-1}$.

3. *Proof of (3.8).* For $n \geq 1$,

$$\left(1 - \frac{x}{n}\right) \left(1 - \frac{y}{n}\right) \left(1 + \frac{x+y}{n}\right) = 1 + \frac{c_n}{n^2},$$

where

$$c_n = -(x^2 + y^2) - xy + xy \frac{x+y}{n}.$$

For $n \geq |x| + |y|$ we have $\left|\frac{x+y}{n}\right| \leq 1$, hence

$$-2|xy| \leq -xy + xy \frac{x+y}{n} \leq 2|xy|.$$

Using $(|x| - |y|)^2 \geq 0$ we obtain $2|xy| \leq x^2 + y^2$, so for $n \geq |x| + |y|$,

$$-2(x^2 + y^2) \leq c_n \leq 0,$$

and in particular $\frac{c_n}{n^2} \geq -\frac{2(x^2+y^2)}{n^2} \geq -\frac{2(x^2+y^2)}{n} \geq -1$ for all sufficiently large n . Therefore, by Bernoulli's inequality,

$$1 - 2\frac{x^2 + y^2}{n} \leq 1 + \frac{c_n}{n} \leq \left(1 + \frac{c_n}{n^2}\right)^n \leq 1 \quad \text{for } n \text{ large,}$$

so Lemma 2.97 yields

$$\lim_{n \rightarrow \infty} \left(1 + \frac{c_n}{n^2}\right)^n = 1.$$

Now,

$$\frac{\exp(x+y)}{\exp(x)\exp(y)} = \lim_{n \rightarrow \infty} \left(1 - \frac{x}{n}\right)^n \left(1 - \frac{y}{n}\right)^n \left(1 + \frac{x+y}{n}\right)^n = \lim_{n \rightarrow \infty} \left(1 + \frac{c_n}{n^2}\right)^n = 1,$$

which proves $\exp(x+y) = \exp(x)\exp(y)$.

We next establish the continuity, monotonicity, and bijectivity of \exp .

We begin with two basic estimates that we will use repeatedly:

$$\exp(x) \geq 1 + x \quad \forall x \in \mathbb{R}, \quad (3.9)$$

which follows from Corollary 3.47 for $x > -1$, and is trivial for $x \leq -1$ (since $\exp(x) > 0$). Using (3.7) together with (3.9) we obtain

$$\exp(x) = \frac{1}{\exp(-x)} \leq \frac{1}{1-x} \quad \forall x < 1. \quad (3.10)$$

A further consequence is that

$$\exp\left(-\frac{1}{x}\right) < x < \exp(x) \quad \forall x > 0. \quad (3.11)$$

Indeed, the right inequality $x < \exp(x)$ follows from (3.9). For the left inequality, apply (3.9) with $y = \frac{1}{x} > 0$:

$$\exp\left(\frac{1}{x}\right) \geq 1 + \frac{1}{x} > \frac{1}{x} \implies \exp\left(-\frac{1}{x}\right) = \exp\left(\frac{1}{x}\right)^{-1} < \left(\frac{1}{x}\right)^{-1} = x.$$

1. *Strict monotonicity.* If $x < y$, then (3.9) yields $\exp(y-x) \geq 1 + (y-x) > 1$, hence

$$\exp(y) = \exp(x)\exp(y-x) > \exp(x).$$

2. *Continuity.* First, continuity at 0. Fix $\delta \in (0, 1)$ and consider $x \in (-\delta, \delta)$.

(i) If $x \in [0, \delta)$, then by the monotonicity of \exp we have $\exp(x) > \exp(0) = 1$, so it

follows from (3.10) that

$$|\exp(x) - \exp(0)| = \exp(x) - 1 \leq \frac{1}{1-x} - 1 < \frac{1}{1-\delta} - 1 = \frac{\delta}{1-\delta}.$$

(ii) If $x \in (-\delta, 0]$ we now have $\exp(x) < \exp(0) = 1$ (again by the monotonicity of \exp), so it follows from (3.9) that

$$|\exp(x) - \exp(0)| = 1 - \exp(x) \leq 1 - (1+x) = -x < \delta \leq \frac{\delta}{1-\delta}.$$

In other words, we proved that $|\exp(x) - \exp(0)| < \frac{\delta}{1-\delta}$ for all $x \in (-\delta, \delta)$.

Now, given $\varepsilon > 0$, choose $\delta = \frac{\varepsilon}{1+\varepsilon}$; with this choice it follows that $\frac{\delta}{1-\delta} = \varepsilon$, therefore

$$x \in (-\delta, \delta) \implies |\exp(x) - 1| < \frac{\delta}{1-\delta} = \varepsilon,$$

proving the continuity of \exp at 0.

To show continuity at an arbitrary point $\bar{x} \in \mathbb{R}$, write $\exp(x) = \exp(\bar{x})\exp(x - \bar{x})$, and let $(x_n)_{n=0}^\infty$ be a sequence with $x_n \rightarrow \bar{x}$. Then, since $x_n - \bar{x} \rightarrow 0$ and \exp is continuous at 0, it follows from Theorem 3.21 that

$$\lim_{n \rightarrow \infty} \exp(x_n) = \exp(\bar{x}) \lim_{n \rightarrow \infty} \exp(x_n - \bar{x}) = \exp(\bar{x})\exp(0) = \exp(\bar{x}).$$

Since $(x_n)_{n=0}^\infty$ is an arbitrary sequence converging to \bar{x} , Theorem 3.21 implies that \exp is continuous at x_0 .

3. *Bijectivity.* First of all, strict monotonicity implies injectivity. For surjectivity, fix $a > 0$ and set $x_0 = -a^{-1}$ and $x_1 = a$. Then, thanks to (3.11),

$$\exp(x_0) < a < \exp(x_1).$$

Hence, by the continuity of \exp and the Intermediate Value Theorem 3.24 applied on $[x_0, x_1]$, there exists $x \in [x_0, x_1]$ with $\exp(x) = a$. This shows surjectivity and concludes the proof. □

3.4.3 The Natural Logarithm

DEFINITION 3.49: LOGARITHM

The unique inverse function

$$\log : \mathbb{R}_{>0} \rightarrow \mathbb{R}$$

of the bijective mapping $\exp : \mathbb{R} \rightarrow \mathbb{R}_{>0}$ is called the **logarithm**.

COROLLARY 3.50: PROPERTIES OF THE LOGARITHM

The logarithm $\log : \mathbb{R}_{>0} \rightarrow \mathbb{R}$ is strictly increasing, continuous, and bijective. Moreover,

$$\log(1) = 0, \quad (3.12)$$

$$\log(a^{-1}) = -\log(a), \quad (3.13)$$

$$\log(ab) = \log(a) + \log(b), \quad (3.14)$$

for all $a, b > 0$.

Proof. This follows directly from Theorem 3.48 and the Inverse Function Theorem 3.28. Equations (3.12), (3.13), and (3.14) follow from the corresponding properties of the exponential, choosing $x = \log a$ and $y = \log b$. \square

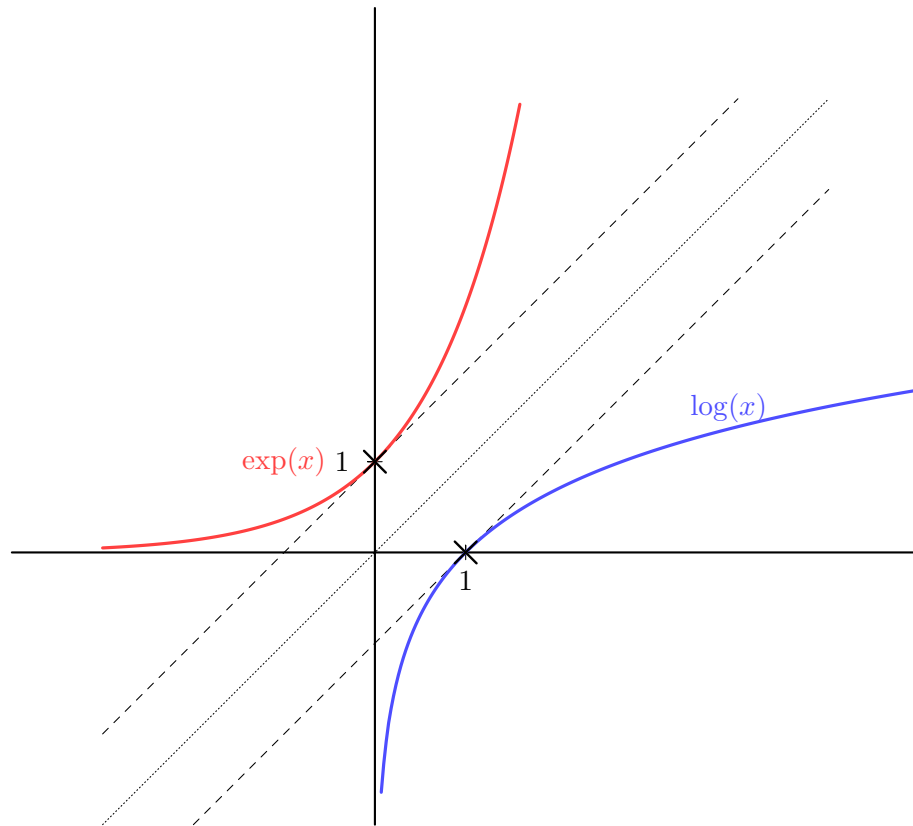


Figure 3.7: Graphs of the exponential function and the logarithm. The auxiliary dashed lines show that $\exp(x) \geq x + 1$ and $\log(x) \leq x - 1$.

The logarithm defined here is also called the **natural logarithm** to distinguish it from logarithms with another **base** $a > 1$ (for instance $a = 10$ or $a = 2$). For any $a > 1$, we define

$$\log_a(x) = \frac{\log x}{\log a} \quad \forall x > 0.$$

For example, $\log_{10}(10^n) = n$ for all $n \in \mathbb{Z}$. Unless stated otherwise, $\log(x)$ always denotes the natural logarithm, i.e., the logarithm to base e .

We can now define powers with arbitrary real exponents. For $a > 0$ and $x \in \mathbb{R}$ we set

$$a^x = \exp(x \log a).$$

In particular, $e^x = \exp(x \log e) = \exp(x)$ for all $x \in \mathbb{R}$. Similarly, for $x > 0$ and $a \in \mathbb{R}$,

$$x^a = \exp(a \log x).$$

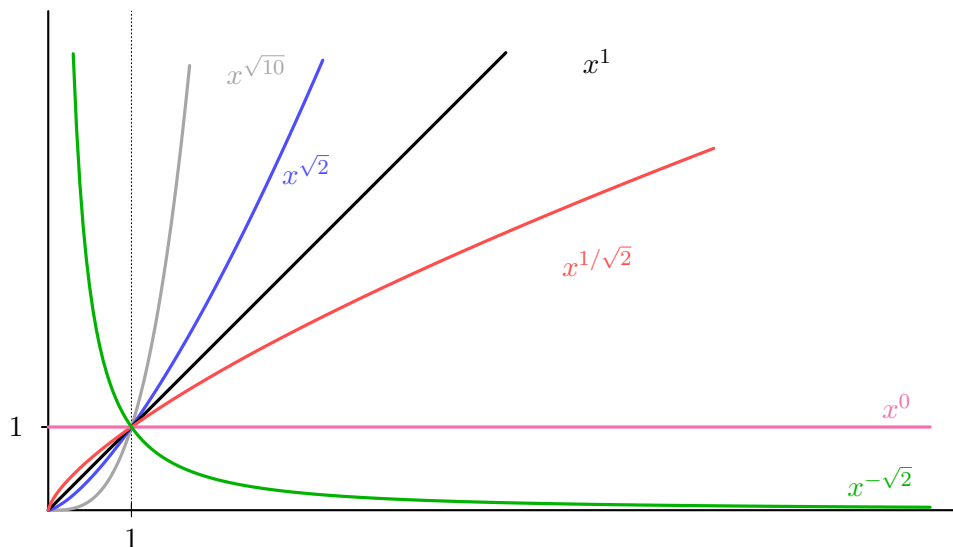


Figure 3.8: Graphs of $a \mapsto a^x$ for various real exponents x .

EXERCISE 3.51. — Show that for $x \in \mathbb{Q}$ and $a > 0$, this definition agrees with the one for rational powers from Example 3.29. Verify also the rules

$$\log(a^x) = x \log(a), \quad a^x a^y = a^{x+y}, \quad (a^x)^y = a^{xy}$$

for all $a > 0$ and $x, y \in \mathbb{R}$.

EXERCISE 3.52. — Let $a > 0$. Show that there exists a constant $C_a > 0$ such that $\log(x) \leq C_a x^a$ for all $x > 0$.

EXERCISE 3.53. — Given $a \in \mathbb{R}$, consider the sequence $(x_n)_{n=1}^\infty$ given by $x_n = \sqrt[n]{n^a}$. Show that this sequence converges, with

$$\lim_{n \rightarrow \infty} \sqrt[n]{n^a} = 1.$$

EXERCISE 3.54. — In this exercise, we introduce another notion of continuity (compare with Exercise 3.41).

1. Let $D \subset \mathbb{R}$ and $\alpha \in (0, 1]$. A real-valued function f on D is called **α -Hölder continuous** if there exists $L \geq 0$ such that

$$|f(x) - f(y)| \leq L|x - y|^\alpha \quad \forall x, y \in D.$$

Show that any α -Hölder continuous function is uniformly continuous.

(Note: for $\alpha = 1$, this reduces to the Lipschitz condition.)

15

2. Given $\alpha \in (0, 1]$, consider $f : [0, \infty) \rightarrow \mathbb{R}$, $f(x) = x^\alpha$. Show that f is α -Hölder continuous.

First observe that $t^\alpha \geq t$ for all $t \in [0, 1]$. Applying this fact with $t = \frac{x}{x+y}$ and $t = \frac{y}{x+y}$ with $x, y > 0$, deduce that

$$(x + y)^\alpha \leq x^\alpha + y^\alpha \quad \forall x, y \geq 0, \alpha \in (0, 1]. \quad (3.15)$$

Finally, use (3.15) to prove that f is α -Hölder continuous.

3.5 Limits of Functions

We consider functions $f : D \rightarrow \mathbb{R}$ defined on a subset $D \subset \mathbb{R}$, and we wish to define the limit of $f(x)$ as $x \in D$ approaches a point $x_0 \in \mathbb{R}$. Typical examples include $D = \mathbb{R}$, $D = [0, 1]$, or $D = (0, 1)$, with $x_0 = 0$ in each case.

3.5.1 Limit in the Vicinity of a Point

Let $D \subset \mathbb{R}$ be non-empty, and let $x_0 \in \mathbb{R}$ be such that

$$D \cap (x_0 - \delta, x_0 + \delta) \neq \emptyset \quad (3.16)$$

for all $\delta > 0$. Whenever this holds, we say that x_0 is an **accumulation point** of D . Note that if $x_0 \in D$, then (3.16) is automatically satisfied.

Condition (3.16) ensures that there exists a sequence of points in D converging to x_0 .

DEFINITION 3.55: LIMIT OF A FUNCTION

Let $f : D \rightarrow \mathbb{R}$, and x_0 be an accumulation point of D .

A number $L \in \mathbb{R}$ is called the **limit of $f(x)$ as $x \rightarrow x_0$** if, for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\forall x \in D, \quad |x - x_0| < \delta \implies |f(x) - L| < \varepsilon.$$

15

In general, the limit of $f(x)$ as $x \rightarrow x_0$ may not exist. However, if it exists, it is uniquely determined. Hence, we speak of *the* limit and write

$$\lim_{x \rightarrow x_0} f(x) = L$$

to indicate that the limit exists and is equal to L . Informally, this means that the function values $f(x)$ are arbitrarily close to L whenever $x \in D$ is sufficiently close to x_0 .

The limit of a function satisfies properties analogous to those of Proposition 2.94. More precisely:

If f and g are functions on D such that

$$\lim_{x \rightarrow x_0} f(x) = L_1 \quad \text{and} \quad \lim_{x \rightarrow x_0} g(x) = L_2,$$

then

$$\lim_{x \rightarrow x_0} (f + g)(x) = L_1 + L_2, \quad \lim_{x \rightarrow x_0} (fg)(x) = L_1 L_2.$$

Moreover, $f \leq g$ implies $L_1 \leq L_2$, and the sandwich lemma holds: if $f \leq h \leq g$ and $L_1 = L_2$, then $\lim_{x \rightarrow x_0} h(x) = L_1 = L_2$.

REMARK 3.56. — Let $f : D \rightarrow \mathbb{R}$ be a function. If $x_0 \in D$, then f is continuous at x_0 if and only if $\lim_{x \rightarrow x_0} f(x) = f(x_0)$.

Suppose that $x_0 \in D$ is an accumulation point of $D \setminus \{x_0\}$. Let $f : D \rightarrow \mathbb{R}$, and consider its restriction $f|_{D \setminus \{x_0\}}$. It may happen that f is discontinuous at x_0 , but that the limit

$$L = \lim_{x \rightarrow x_0} f|_{D \setminus \{x_0\}}(x) \quad (3.17)$$

nevertheless exists. In this case, the point x_0 is called a **removable discontinuity** of f , and one also writes

$$L = \lim_{\substack{x \rightarrow x_0 \\ x \neq x_0}} f(x). \quad (3.18)$$

If we now define

$$\tilde{f}(x) = \begin{cases} f(x), & x \in D \setminus \{x_0\}, \\ L, & x = x_0, \end{cases} \quad (3.19)$$

then \tilde{f} is continuous at x_0 . In other words, we can remove the discontinuity of f by redefining its value at x_0 to be L .

If instead $x_0 \notin D$ but the limit in (3.18) exists, we call the function \tilde{f} defined in (3.19) the **continuous extension** of f to $D \cup \{x_0\}$.

Arguing as in the proof of Theorem 3.21, we obtain the following result.

LEMMA 3.57: LIMIT AND SEQUENCES

Let $f : D \rightarrow \mathbb{R}$. Then $L = \lim_{x \rightarrow \bar{x}} f(x)$ if and only if, for every sequence $(x_n)_{n=0}^{\infty} \subset D$ converging to \bar{x} , one has $\lim_{n \rightarrow \infty} f(x_n) = L$.

EXERCISE 3.58. — Prove Lemma 3.57.

We now state a result describing the behaviour of limits under composition with a continuous function.

PROPOSITION 3.59: LIMIT AND COMPOSITION

Let $E \subset \mathbb{R}$, and let $f : D \rightarrow E$ be such that the limit $L = \lim_{x \rightarrow \bar{x}} f(x)$ exists and belongs to E . If $g : E \rightarrow \mathbb{R}$ is continuous at L , then

$$\lim_{x \rightarrow \bar{x}} g(f(x)) = g(L).$$

Proof. Let $(x_n)_{n=0}^{\infty} \subset D$ be a sequence converging to \bar{x} . By Lemma 3.57 we have $\lim_{n \rightarrow \infty} f(x_n) = L$. Since g is continuous at L , Theorem 3.21 gives $\lim_{n \rightarrow \infty} g(f(x_n)) = g(L)$. Because $(x_n)_{n=0}^{\infty}$ was arbitrary, using Lemma 3.57 again, we conclude that $\lim_{x \rightarrow \bar{x}} g(f(x)) = g(L)$. \square

We now introduce conventions for improper limits of functions, in analogy with improper limits for sequences.

DEFINITION 3.60: IMPROPER LIMITS

Let $f : D \rightarrow \mathbb{R}$, and let x_0 be an accumulation point of D .

We say that f **diverges to $+\infty$ as $x \rightarrow x_0$** , and write

$$\lim_{x \rightarrow x_0} f(x) = +\infty,$$

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

$$\forall x \in D, \quad |x - x_0| < \delta \implies f(x) \geq M.$$

Analogously, f **diverges to $-\infty$ as $x \rightarrow x_0$** and we write $\lim_{x \rightarrow x_0} f(x) = -\infty$, if for every $M > 0$ there exists $\delta > 0$ such that

$$\forall x \in D, \quad |x - x_0| < \delta \implies f(x) \leq -M.$$

16

3.5.2 One-sided Limits

It is often useful to consider limits taken from one side only and to allow x_0 to be $\pm\infty$ as well. To this end, let $x_0 \in \mathbb{R}$ be such that

$$D \cap (x_0, x_0 + \delta) \neq \emptyset \tag{3.20}$$

for every $\delta > 0$. In this case, we say that x_0 is a **right-hand accumulation point** of D . Analogously, if

$$D \cap (x_0 - \delta, x_0) \neq \emptyset \tag{3.21}$$

for every $\delta > 0$, we say that x_0 is an **left-hand accumulation point** of D .

DEFINITION 3.61: ONE-SIDED LIMITS

Let $f : D \rightarrow \mathbb{R}$, and let $x_0 \in \mathbb{R}$ be a right-hand accumulation point of D .

A number $L \in \mathbb{R}$ is called the **right-hand limit** of f at x_0 if, for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$x \in D \cap (x_0, x_0 + \delta) \implies |f(x) - L| < \varepsilon.$$

In this case we write $L = \lim_{x \rightarrow x_0^+} f(x)$.

We also allow improper one-sided limits. We say that

$$\lim_{x \rightarrow x_0^+} f(x) = +\infty$$

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

$$x \in D \cap (x_0, x_0 + \delta) \implies f(x) \geq M.$$

Similarly, $\lim_{x \rightarrow x_0^+} f(x) = -\infty$ means: for every $M > 0$ there exists $\delta > 0$ such that

$$x \in D \cap (x_0, x_0 + \delta) \implies f(x) \leq -M.$$

The **left-hand limit** is defined analogously, considering a left-hand accumulation point of D and writing $\lim_{x \rightarrow x_0^-} f(x)$.

Next, we define the notion of limit at infinity.

DEFINITION 3.62: LIMITS AT INFINITY

Let $f : D \rightarrow \mathbb{R}$, and assume that $D \cap (R, \infty) \neq \emptyset$ for every $R > 0$.

A number $L \in \mathbb{R}$ is called the **limit of f as $x \rightarrow +\infty$** if, for every $\varepsilon > 0$, there exists $R > 0$ such that

$$x \in D \cap (R, \infty) \implies |f(x) - L| < \varepsilon.$$

We say that f **diverges to $+\infty$ as $x \rightarrow +\infty$** if, for every $M > 0$, there exists $R > 0$ such that

$$x \in D \cap (R, \infty) \implies f(x) \geq M.$$

The corresponding definitions for $x \rightarrow -\infty$ and divergence to $-\infty$ are analogous.

Limits at $+\infty$ can be converted into right-hand limits at 0 via inversion. Given $f : D \rightarrow \mathbb{R}$ as above, define

$$E = \{x > 0 : x^{-1} \in D\}, \quad g : E \rightarrow \mathbb{R}, \quad g(x) = f(x^{-1}).$$

Then

$$\lim_{x \rightarrow +\infty} f(x) = \lim_{x \rightarrow 0^+} g(x),$$

so one limit exists if and only if the other does.

DEFINITION 3.63: ONE-SIDED CONTINUITY AND JUMPS

Let $f : D \rightarrow \mathbb{R}$ and $x_0 \in D$. If $\lim_{x \rightarrow x_0^+} f(x)$ exists and equals $f(x_0)$, then f is **continuous from the right** at x_0 . **Continuity from the left** is defined similarly.

We call x_0 a **jump point** if both one-sided limits exist but are different:

$$L_- := \lim_{x \rightarrow x_0^-} f(x) \in \mathbb{R}, \quad L_+ := \lim_{x \rightarrow x_0^+} f(x) \in \mathbb{R}, \quad L_- \neq L_+.$$

The following graph represents a function with three points of discontinuity x_1, x_2, x_3 .

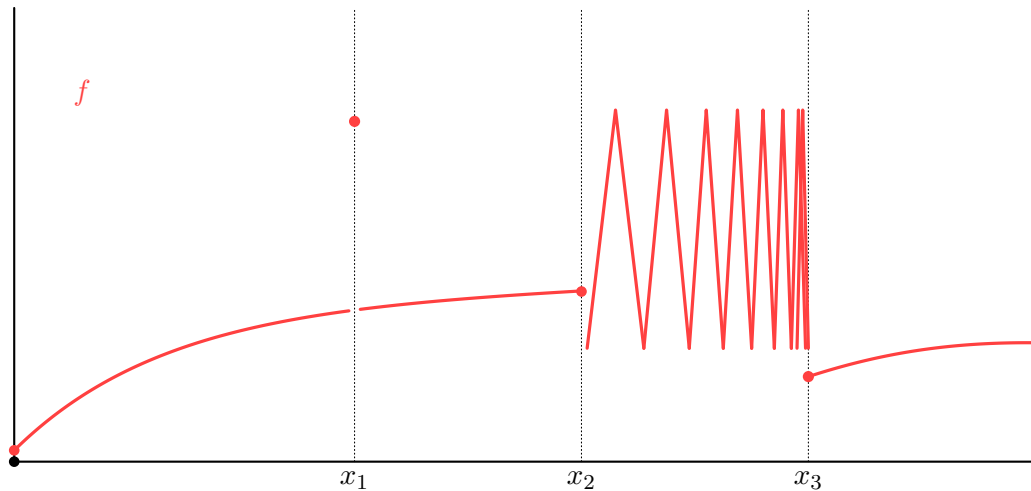


Figure 3.9: At x_1 the discontinuity is removable (both one-sided limits exist and are equal, but $f(x_1)$ differs). At x_2 the function is continuous from the left but not from the right, and x_2 is a jump point. At x_3 the function is continuous from the right; the left-hand limit does not exist, hence x_3 is not a jump point.

EXAMPLE 3.64. — Let $D = (0, \infty)$ and $f(x) = x^x = \exp(x \log x)$. We want to compute $\lim_{x \rightarrow 0^+} f(x)$. We split the argument into three steps.

- (i) $\lim_{y \rightarrow \infty} ye^{-y} = 0$. Indeed, by Corollary 3.47, for $y > 0$ one has $e^y \geq (1 + \frac{y}{2})^2$, hence

$$0 \leq ye^{-y} \leq \frac{y}{(1 + \frac{y}{2})^2} \leq \frac{4}{y} \xrightarrow{y \rightarrow \infty} 0$$

by the sandwich lemma.

- (ii) $\lim_{x \rightarrow 0^+} x \log x = 0$. For this, given $\varepsilon > 0$, choose $R > 0$ so that $|ye^{-y}| < \varepsilon$ for all $y > R$ (this is possible thanks to point (i) above). Now, set $\delta = e^{-R}$ and take $x \in (0, \delta)$. Then

$y = -\log x > R$, which implies that

$$|x \log x| = |e^{-y} y| < \varepsilon \quad \forall x \in (0, \delta).$$

(iii) $\lim_{x \rightarrow 0^+} x^x = 1$. Indeed, thanks to (ii) and the continuity of the exponential function, we can apply Proposition 3.59 to get

$$\lim_{x \rightarrow 0^+} x^x = \lim_{x \rightarrow 0^+} \exp(x \log x) = \exp(0) = 1.$$

EXERCISE 3.65. — Let $a \in \mathbb{R}$. Compute, where defined,

$$\lim_{x \rightarrow 2} \frac{x^3 - x^2 - x - 2}{x - 2}, \quad \lim_{x \rightarrow \infty} \frac{3e^{2x} + e^x + 1}{2e^{2x} - 1}, \quad \lim_{x \rightarrow \infty} \frac{e^x}{x^a}, \quad \lim_{x \rightarrow \infty} \frac{\log x}{x^a}.$$

Specify a suitable domain D (so that the functions are well defined) in each case.

3.5.3 Landau Notation

16

We introduce two standard notations that compare the asymptotic behaviour of a function to that of another function (often called *relative asymptotics*). They are named after Edmund Landau (1877–1938).

DEFINITION 3.66: BIG-O AT A POINT

Let $f, g : D \rightarrow \mathbb{R}$, and let x_0 be an accumulation point of D . We write

$$f(x) = O(g(x)) \quad \text{as } x \rightarrow x_0$$

if there exist $M > 0$ and $\delta > 0$ such that

$$x \in D \cap (x_0 - \delta, x_0 + \delta) \implies |f(x)| \leq M |g(x)|.$$

We then say that f is a **big-O** of g as $x \rightarrow x_0$.

If $g(x) \neq 0$ for all x sufficiently close to x_0 (with $x \in D$), then

$$f(x) = O(g(x)) \text{ as } x \rightarrow x_0 \iff \frac{f(x)}{g(x)} \text{ is bounded near } x_0.$$

DEFINITION 3.67: BIG-O AT INFINITY

Let $f, g : D \rightarrow \mathbb{R}$, and assume that $D \cap (R, \infty) \neq \emptyset$ for every $R > 0$. We write

$$f(x) = O(g(x)) \quad \text{as } x \rightarrow +\infty$$

if there exist $M > 0$ and $R > 0$ such that

$$x \in D \cap (R, \infty) \implies |f(x)| \leq M |g(x)|.$$

The definition for $x \rightarrow -\infty$ is analogous.

The big-O notation hides the precise bound by an *implicit constant* M , which is often irrelevant for the argument that one is interested in.

EXAMPLE 3.68. — • If f and g are bounded and continuous near x_0 with $g(x_0) \neq 0$, then $f(x) = O(g(x))$ as $x \rightarrow x_0$.

• As $x \rightarrow 0$, one has $x^2 = O(x)$, but $x \neq O(x^2)$ (since x/x^2 is unbounded near 0).

• As $x \rightarrow +\infty$, $\frac{3x^3}{x^3 + 3} = O(1)$, but $\frac{3x^3}{x^3 + 3} \neq O(x^\alpha)$ for $\alpha < 0$.

As discussed above, the big-O means that f is bounded by a multiple of g . One may also consider a stronger condition, namely that f is asymptotically negligible with respect to g . This leads to the following definition.

DEFINITION 3.69: LITTLE-O AT A POINT

Let $f, g : D \rightarrow \mathbb{R}$, and let x_0 be an accumulation point of D . We write

$$f(x) = o(g(x)) \quad \text{as } x \rightarrow x_0$$

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

$$x \in D \cap (x_0 - \delta, x_0 + \delta) \implies |f(x)| \leq \varepsilon |g(x)|.$$

We then say that f is a **little-o** of g as $x \rightarrow x_0$.

If $g(x) \neq 0$ for all x near x_0 (with $x \in D$), then

$$f(x) = o(g(x)) \text{ as } x \rightarrow x_0 \iff \lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = 0.$$

Moreover, $f(x) = o(g(x)) \implies f(x) = O(g(x))$.

DEFINITION 3.70: LITTLE-O AT INFINITY

Let $f, g : D \rightarrow \mathbb{R}$, and assume that $D \cap (R, \infty) \neq \emptyset$ for every $R > 0$. We write

$$f(x) = o(g(x)) \quad \text{as } x \rightarrow +\infty$$

if, for every $\varepsilon > 0$, there exists $R > 0$ such that

$$x \in D \cap (R, \infty) \implies |f(x)| \leq \varepsilon |g(x)|.$$

The definition for $x \rightarrow -\infty$ is analogous.

EXAMPLE 3.71. — • $x = o(x^2)$ as $x \rightarrow +\infty$, and $x^2 = o(x)$ as $x \rightarrow 0$.

• For any $\alpha < 1$,

$$\frac{3x^3}{2x^2 + x^{10}} = o(|x|^\alpha) \quad \text{as } x \rightarrow 0,$$

but not for $\alpha \geq 1$. Indeed,

$$\left| \frac{3x^3}{|x|^\alpha(2x^2 + x^{10})} \right| = |x|^{1-\alpha} \frac{3}{2 + x^8} \longrightarrow 0 \quad \text{as } x \rightarrow 0,$$

whenever $\alpha < 1$.

In computations, one often uses Landau symbols as placeholders. Writing

$$f(x) + o(g(x)) \quad \text{as } x \rightarrow x_0$$

means there is a function $h : D \rightarrow \mathbb{R}$ with $h(x) = o(g(x))$ as $x \rightarrow x_0$. Similarly for big-O.

EXERCISE 3.72. — Let $p > 1$, $a \in \mathbb{R}$, and $b > 0$. Show:

1. $x^p = o(x)$ as $x \rightarrow 0$;
2. $x = o(x^p)$ as $x \rightarrow \infty$;
3. $x^a = o(e^x)$ as $x \rightarrow \infty$;
4. $\log x = o(x^b)$ as $x \rightarrow \infty$.

EXAMPLE 3.73. — Polynomial division gives, as $x \rightarrow \infty$,

$$\frac{x^3 - 7x^2 + 6x + 2}{x^2} = x - 7 + O\left(\frac{1}{x}\right) = x - 7 + o(1) = x + O(1) = x + o(x).$$

It may seem surprising that all four expressions are correct (and even useful) in different contexts. Depending on the desired precision, one may use the more accurate form with error term $-7 + O\left(\frac{1}{x}\right)$ or the coarser estimate involving $o(x)$.

EXERCISE 3.74. — Let $f_1, f_2, g : D \rightarrow \mathbb{R}$. If $f_1(x) = o(g(x))$ and $f_2(x) = o(g(x))$ as $x \rightarrow x_0$, prove that for all $\alpha_1, \alpha_2 \in \mathbb{R}$,

$$\alpha_1 f_1(x) + \alpha_2 f_2(x) = o(g(x)) \quad \text{as } x \rightarrow x_0.$$

Formulate and prove the analogous statement for big-O.

EXERCISE 3.75. — Let $f_1, f_2, g_1, g_2 : D \rightarrow \mathbb{R}$. Show:

- If $f_1(x) = o(g_1(x))$ and $f_2(x) = o(g_2(x))$ as $x \rightarrow x_0$, then $f_1(x)f_2(x) = o(g_1(x)g_2(x))$.
- If $f_1(x) = o(g_1(x))$ and $f_2(x) = O(g_2(x))$ as $x \rightarrow x_0$, then $f_1(x)f_2(x) = o(g_1(x)g_2(x))$.
- If $f_1(x) = O(g_1(x))$ and $f_2(x) = O(g_2(x))$ as $x \rightarrow x_0$, then $f_1(x)f_2(x) = O(g_1(x)g_2(x))$.

16

EXAMPLE 3.76. — Let $f(x) = x + x^3 + 4x^4 + x^7$ and $g(x) = x + \frac{3x^2}{1+x}$. As $x \rightarrow 0$ we have

$$f(x) = x + o(x^2) \quad (\text{since } x^3, x^4, x^7 = o(x^2)), \quad g(x) = x + O(x^2) \quad (\text{since } \frac{3x^2}{1+x} = 3x^2 + O(x^3)).$$

Hence

$$f(x)g(x) = (x + o(x^2))(x + O(x^2)) = x^2 + x o(x^2) + x O(x^2) + o(x^2) O(x^2).$$

By the product for $o(\cdot)$ and $O(\cdot)$ (Exercise 3.75),

$$x o(x^2) = o(x^3), \quad x O(x^2) = O(x^3), \quad o(x^2) O(x^2) = o(x^4).$$

Also, by the sum rule (Exercise 3.74),

$$o(x^3) + O(x^3) + o(x^4) = O(x^3).$$

Therefore,

$$f(x)g(x) = x^2 + O(x^3).$$

3.6 Sequences of Functions

3.6.1 Pointwise Convergence

DEFINITION 3.77: SEQUENCES OF FUNCTIONS

A **sequence** of real-valued functions on a subset $D \subset \mathbb{R}$ is a family of functions $f_n : D \rightarrow \mathbb{R}$ indexed by \mathbb{N} . The function f_n is called the n -th **element** of the sequence. One often writes $(f_n)_{n \in \mathbb{N}}$, $(f_n)_{n=0}^\infty$, or $(f_n)_{n \geq 0}$ for a sequence of functions.

As for sequences of real numbers, the starting index of a sequence of functions is not essential and may differ from 0.

DEFINITION 3.78: POINTWISE CONVERGENCE

Let $D \subset \mathbb{R}$, and let $(f_n)_{n=0}^\infty$ be a sequence of functions $f_n : D \rightarrow \mathbb{R}$. Let $f : D \rightarrow \mathbb{R}$ be another function. We say that $(f_n)_{n=0}^\infty$ **converges pointwise** to f if for every $x \in D$, the sequence of real numbers $(f_n(x))_{n=0}^\infty$ converges to $f(x)$. In this case, f is called the **pointwise limit** of the sequence $(f_n)_{n=0}^\infty$.

17

EXERCISE 3.79. — Show that the pointwise limit of a sequence of functions, if it exists, is uniquely determined.

In the following example we show that, in general, continuity is not preserved under pointwise convergence.

EXAMPLE 3.80. — Let $D = [0, 1]$ and, given $n \geq 1$, define $f_n : D \rightarrow \mathbb{R}$ by $f_n(x) = x^n$. Then the sequence of continuous functions $(f_n)_{n=1}^\infty$ converges pointwise to the function $f : D \rightarrow \mathbb{R}$ given by

$$f(x) = \lim_{n \rightarrow \infty} f_n(x) = \begin{cases} 0 & \text{if } x < 1, \\ 1 & \text{if } x = 1. \end{cases}$$

The limit function f is not continuous on $[0, 1]$.

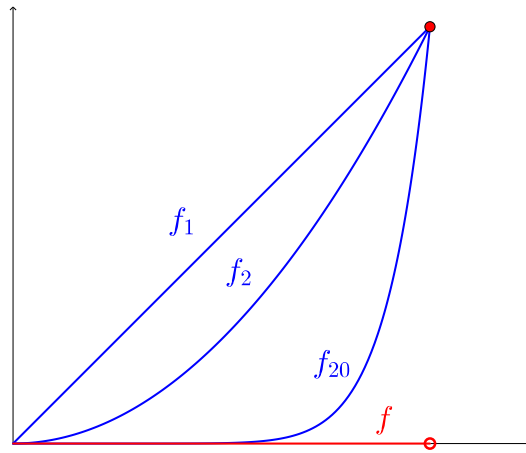


Figure 3.10: The sequence $f_n(x) = x^n$ converges pointwise on $[0, 1]$ to the function $f(x) = 0$ for $x < 1$ and $f(1) = 1$, which is not continuous at $x = 1$.

3.6.2 Uniform Convergence

As we have seen in the previous section, pointwise convergence of functions is not sufficient to preserve continuity. We now introduce a stronger notion of convergence that ensures this property. To motivate the definition, let us first rewrite the notion of pointwise convergence using quantifiers.

Recall that $(f_n)_{n=0}^{\infty}$ converges *pointwise* to f on D if, for every $x \in D$ and every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$|f_n(x) - f(x)| < \varepsilon \quad \text{for all } n \geq N.$$

In this definition we first fix $x \in D$ and $\varepsilon > 0$, and then find an index N that may depend on both x and ε . In contrast, in the definition of *uniform convergence* below, we change the order of quantifiers to ensure that N depends only on ε , and not on the particular choice of $x \in D$.

DEFINITION 3.81: UNIFORM CONVERGENCE

Let $D \subset \mathbb{R}$, and let $(f_n)_{n=0}^{\infty}$ be a sequence of functions $f_n : D \rightarrow \mathbb{R}$. Let $f : D \rightarrow \mathbb{R}$ be another function. We say that $(f_n)_{n=0}^{\infty}$ **converges uniformly** to f on D if, for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$|f_n(x) - f(x)| < \varepsilon \quad \text{for all } n \geq N \text{ and all } x \in D.$$

The condition $|f_n(x) - f(x)| < \varepsilon$ is equivalent to

$$f(x) - \varepsilon \leq f_n(x) \leq f(x) + \varepsilon.$$

This gives a useful geometric interpretation of uniform convergence: (f_n) converges uniformly to f if, for every $\varepsilon > 0$, the graph of f_n eventually lies entirely within the ε -tube around the graph of f (see Figure 3.11).

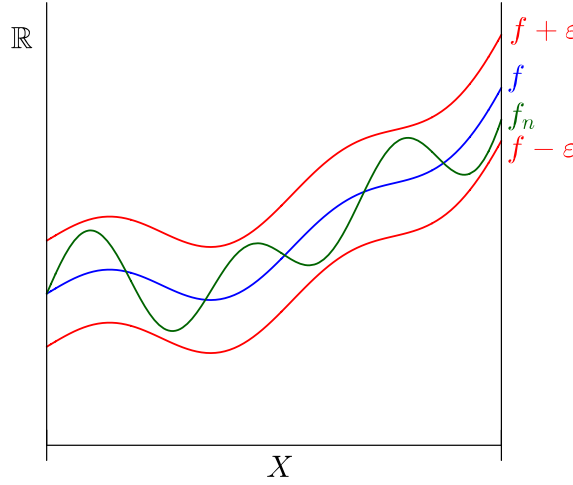


Figure 3.11: Illustration of uniform convergence: for all sufficiently large n , the graph of f_n lies entirely within the ε -tube around the graph of f .

EXERCISE 3.82. — Let $(f_n)_{n=0}^{\infty}$ be a sequence of functions $f_n : D \rightarrow \mathbb{R}$. Show that if $(f_n)_{n=0}^{\infty}$ converges uniformly to a function f , then $(f_n)_{n=0}^{\infty}$ also converges pointwise to f .

THEOREM 3.83: CONTINUITY UNDER UNIFORM CONVERGENCE

Let $D \subset \mathbb{R}$, and let $(f_n)_{n=0}^{\infty}$ be a sequence of continuous functions $f_n : D \rightarrow \mathbb{R}$ converging uniformly to a function $f : D \rightarrow \mathbb{R}$. Then f is continuous.

Proof. To prove that f is continuous, we fix $\bar{x} \in D$ and show that f is continuous at \bar{x} .

Given $\varepsilon > 0$, the uniform convergence of f_N to f provides $N \in \mathbb{N}$ such that

$$|f_N(y) - f(y)| < \frac{\varepsilon}{3} \quad \text{for all } y \in D.$$

Also, since f_N is continuous at \bar{x} , there exists $\delta > 0$ such that

$$|x - \bar{x}| < \delta \implies |f_N(x) - f_N(\bar{x})| < \frac{\varepsilon}{3}.$$

Then, for $|x - \bar{x}| < \delta$, we have

$$\begin{aligned} |f(x) - f(\bar{x})| &\leq |f(x) - f_N(x)| + |f_N(x) - f_N(\bar{x})| + |f_N(\bar{x}) - f(\bar{x})| \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon, \end{aligned}$$

which shows that f is continuous at \bar{x} . Since \bar{x} is arbitrary, f is continuous on D . \square

REMARK 3.84. — Intuitively, uniform convergence allows us to *exchange* the order of taking limits. More precisely, assume that $(f_n)_{n=0}^\infty$ is a sequence of continuous functions converging pointwise to f . Then, by the pointwise convergence and the continuity of the functions f_n , we have

$$f(\bar{x}) = \lim_{n \rightarrow \infty} f_n(\bar{x}), \quad f_n(\bar{x}) = \lim_{x \rightarrow \bar{x}} f_n(x), \quad f(x) = \lim_{n \rightarrow \infty} f_n(x) \quad \text{for all } x \in D.$$

Hence,

$$f(\bar{x}) = \lim_{n \rightarrow \infty} f_n(\bar{x}) = \lim_{n \rightarrow \infty} \left(\lim_{x \rightarrow \bar{x}} f_n(x) \right), \quad \lim_{x \rightarrow \bar{x}} f(x) = \lim_{x \rightarrow \bar{x}} \left(\lim_{n \rightarrow \infty} f_n(x) \right).$$

Note that the function f is continuous at \bar{x} if and only if $f(\bar{x}) = \lim_{x \rightarrow \bar{x}} f(x)$, which by the identities above is equivalent to

$$\lim_{x \rightarrow \bar{x}} \left(\lim_{n \rightarrow \infty} f_n(x) \right) = \lim_{n \rightarrow \infty} \left(\lim_{x \rightarrow \bar{x}} f_n(x) \right).$$

As we have seen, for pointwise convergence this interchange of limits may fail because f need not be continuous. However, Theorem 3.83 ensures that this equality holds under uniform convergence.

The next exercise reformulates this idea in terms of sequences.

17

EXERCISE 3.85. — Let $(f_n)_{n=0}^\infty$ be a sequence of uniformly continuous functions $f_n : D \rightarrow \mathbb{R}$ converging uniformly to $f : D \rightarrow \mathbb{R}$. Let $(x_n)_{n=0}^\infty$ be a sequence in D converging to $\bar{x} \in D$. Show that

$$\lim_{n \rightarrow \infty} f_n(x_n) = f(\bar{x}). \quad (3.22)$$

Find an example showing that pointwise convergence of $(f_n)_{n=0}^\infty$ to f is not sufficient to guarantee (3.22).

EXERCISE 3.86. — Let $(f_n)_{n=0}^\infty$ be a sequence of functions $f_n : D \rightarrow \mathbb{R}$, and let $f : D \rightarrow \mathbb{R}$. Suppose $D = D_1 \cup D_2$ for two subsets such that $(f_n|_{D_1})$ converges uniformly to $f|_{D_1}$ and $(f_n|_{D_2})$ converges uniformly to $f|_{D_2}$. Show that $(f_n)_{n=0}^\infty$ converges uniformly to f on D .

EXERCISE 3.87. — Let $(f_n)_{n=0}^\infty$ be a sequence of bounded functions $f_n : D \rightarrow \mathbb{R}$. Show that if $(f_n)_{n=0}^\infty$ converges uniformly to a function $f : D \rightarrow \mathbb{R}$, then f is bounded. Also, give an example of a sequence (f_n) of bounded functions converging pointwise to an unbounded function.

EXERCISE 3.88. — Let $(f_n)_{n=0}^\infty$ be a sequence of uniformly continuous real-valued functions on D converging uniformly to $f : D \rightarrow \mathbb{R}$. Show that f is uniformly continuous.

Extra material: Equicontinuity and the Arzelà–Ascoli Theorem

We have seen that any bounded sequence of points $(x_n)_{n=0}^\infty$ admits a convergent subsequence. Is there an analogous result for sequences of continuous functions? The Ascoli–Arzelà Theorem addresses this question. We first need some definitions.

DEFINITION 3.89: EQUIBOUNDEDNESS AND EQUICONTINUITY

Let $(f_n)_{n=0}^\infty$ be a sequence of functions $f_n : D \rightarrow \mathbb{R}$. We say that the sequence is **equibounded** on D if there exists $M > 0$ such that

$$|f_n(x)| \leq M \quad \text{for all } x \in D, n \in \mathbb{N}.$$

Also, the sequence is **equicontinuous** on D if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$|x - y| < \delta \implies |f_n(x) - f_n(y)| < \varepsilon \quad \text{for all } x, y \in D, n \in \mathbb{N}.$$

THEOREM 3.90: ASCOLI–ARZELÀ

Let $[a, b]$ be a compact interval, and let $(f_n)_{n=1}^\infty$ be a sequence of continuous functions $f_n : [a, b] \rightarrow \mathbb{R}$ that is equibounded and equicontinuous and pointwise bounded. Then there exist a subsequence $(f_{n_k})_{k=1}^\infty$ and a continuous function $f : [a, b] \rightarrow \mathbb{R}$ such that $(f_{n_k})_{k=1}^\infty$ converges uniformly to f on $[a, b]$.

REMARK 3.91. — • By Theorem 3.83, the uniform limit in Arzelà–Ascoli is automatically continuous; the theorem guarantees the existence of such a subsequence.

- The compactness of $[a, b]$ is essential; without it, equiboundedness and equicontinuity do not force any uniformly convergent subsequence on all of D (see Exercise 3.93).

EXERCISE 3.92. — Following Example 3.80, show that the family $\{x^n \mid n \in \mathbb{N}\}$ is *not* equicontinuous on $[0, 1]$.

Hint: Fix $\varepsilon = \frac{1}{2}$. Then, for any $\delta > 0$, choose $x = 1$, $y = 1 - \frac{\delta}{2}$, and take n large enough so that $(1 - \frac{\delta}{2})^n < \frac{1}{2}$.

EXERCISE 3.93. — Let $g(x) = \max\{(1 - |x|), 0\}$ and define $f_n(x) = g(x - n)$ on \mathbb{R} . Show that $(f_n)_{n=0}^\infty$ is equibounded and equicontinuous on \mathbb{R} , $f_n(x) \rightarrow 0$ for each fixed x , but no subsequence converges uniformly on \mathbb{R} . Conclude that compactness of the domain cannot be dropped from Theorem 3.90.

Chapter 4

Series and Power Series

In this chapter we study series (infinite sums). They provide a framework to define many classical functions; in particular, we will use series to define the trigonometric functions.

4.1 Series of Real Numbers

DEFINITION 4.1: CONVERGENT AND DIVERGENT SERIES

Let $(a_n)_{n=0}^{\infty}$ be a sequence of real numbers, and let $A \in \mathbb{R}$. We say that the series $\sum_{k=0}^{\infty} a_k$ **converges** to A if

$$A = \lim_{n \rightarrow \infty} \sum_{k=0}^n a_k.$$

In other words, computing the infinite sum $\sum_{k=0}^{\infty} a_k$ means finding (if it exists) the limit of the **partial sums**

$$s_n = \sum_{k=0}^n a_k, \quad n \in \mathbb{N}.$$

We call a_n the **n -th term** (or **n -th summand**) of the series. If the limit exists, its value A is the **sum of the series**.

If the limit does not exist, the series is said to be **not convergent**. In particular, if the sequence of partial sums $(s_n)_{n=0}^{\infty}$ diverges to $+\infty$ (respectively, to $-\infty$), we say that the series **diverges to** $+\infty$ (respectively, **to** $-\infty$). This situation is therefore a specific case of a series that does not converge.

REMARK 4.2. — Unless otherwise specified, all series will consist of real numbers.

PROPOSITION 4.3: NECESSARY CONDITION FOR CONVERGENCE

If the series $\sum_{k=0}^{\infty} a_k$ converges, then $a_n \rightarrow 0$ as $n \rightarrow \infty$.

Proof. By assumption the partial sums $s_n = \sum_{k=0}^n a_k$ satisfy $s_n \rightarrow A \in \mathbb{R}$. Then, for $n \geq 1$,

$$a_n = s_n - s_{n-1} \xrightarrow{n \rightarrow \infty} A - A = 0. \quad \square$$

EXAMPLE 4.4 (Geometric Series). — For $q \in \mathbb{R}$, the geometric series $\sum_{n=0}^{\infty} q^n$ converges if and only if $|q| < 1$, and in this case

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}.$$

Indeed, if the series converges, then by Proposition 4.3 we must have $q^n \rightarrow 0$, hence $|q| < 1$. Conversely, for $|q| < 1$ one proves by induction that

$$s_n = \sum_{k=0}^n q^k = \frac{1 - q^{n+1}}{1 - q} \quad \forall n \in \mathbb{N}, q \neq 1.$$

Also, since $|q| < 1$, $q^{n+1} \rightarrow 0$. Thus,

$$s_n = \frac{1 - q^{n+1}}{1 - q} \xrightarrow{n \rightarrow \infty} \frac{1}{1 - q}.$$

17

EXAMPLE 4.5 (Harmonic Series). — The converse of Proposition 4.3 fails: the **harmonic series** $\sum_{k=1}^{\infty} \frac{1}{k}$ does not converge. To see this, consider $n = 2^\ell$ with $\ell \in \mathbb{N}$. Grouping terms gives

$$\begin{aligned} \sum_{k=1}^{2^\ell} \frac{1}{k} &= 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \cdots + \frac{1}{8}\right) + \cdots + \left(\frac{1}{2^{\ell-1}+1} + \cdots + \frac{1}{2^\ell}\right) \\ &\geq 1 + \frac{1}{2} + \underbrace{\frac{1}{4} + \frac{1}{4}}_{=\frac{1}{2}} + \underbrace{\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}}_{=\frac{1}{2}} + \cdots + \underbrace{\frac{1}{2^\ell} + \cdots + \frac{1}{2^\ell}}_{=\frac{1}{2}} \\ &= 1 + \underbrace{\frac{1}{2} + \cdots + \frac{1}{2}}_{\ell \text{ times}} = 1 + \frac{\ell}{2}, \end{aligned}$$

which is unbounded as $\ell \rightarrow \infty$.

EXERCISE 4.6. — Let $\sum_{k=0}^{\infty} a_k$ and $\sum_{k=0}^{\infty} b_k$ be convergent series, and let $\alpha, \beta \in \mathbb{R}$. Show that $\sum_{k=0}^{\infty} (\alpha a_k + \beta b_k)$ converges and

$$\sum_{k=0}^{\infty} (\alpha a_k + \beta b_k) = \alpha \sum_{k=0}^{\infty} a_k + \beta \sum_{k=0}^{\infty} b_k.$$

LEMMA 4.7: CONVERGENCE OF THE TAIL

Let $\sum_{k=0}^{\infty} a_k$ be a series and fix $N \in \mathbb{N}$. Then $\sum_{k=0}^{\infty} a_k$ is convergent if and only if $\sum_{k=N}^{\infty} a_k$ is convergent, and in that case

$$\sum_{k=0}^{\infty} a_k = \sum_{k=0}^{N-1} a_k + \sum_{k=N}^{\infty} a_k.$$

The same equivalence holds for divergence to $+\infty$ or $-\infty$.

Proof. For every $n \geq N$,

$$\sum_{k=0}^n a_k = \sum_{k=0}^{N-1} a_k + \sum_{k=N}^n a_k.$$

Thus the partial sums of $\sum_{k=0}^{\infty} a_k$ converge if and only if those of $\sum_{k=N}^{\infty} a_k$ do, and the identity in the statement follows by letting $n \rightarrow \infty$. The divergence case is analogous. \square

17

4.1.1 Series with Nonnegative Elements**PROPOSITION 4.8: NONNEGATIVE SERIES: CONVERGENCE VS. DIVERGENCE**

Let $\sum_{k=0}^{\infty} a_k$ be a series with nonnegative terms $a_k \geq 0$ for all $k \in \mathbb{N}$. Then the partial sums $s_n = \sum_{k=0}^n a_k$ form an increasing sequence. If $(s_n)_{n=0}^{\infty}$ is bounded, the series $\sum_{k=0}^{\infty} a_k$ converges; otherwise it diverges to $+\infty$.

Proof. Since $a_{n+1} \geq 0$, we have $s_{n+1} = s_n + a_{n+1} \geq s_n$ for all $n \in \mathbb{N}$, so $(s_n)_{n=0}^{\infty}$ is increasing.

If the sequence $(s_n)_{n=0}^{\infty}$ is bounded, then it converges by Theorem 2.104. If the partial sums are not bounded, then they diverge to $+\infty$. \square

REMARK 4.9. — If $\sum_{k=0}^{\infty} a_k$ has nonnegative terms, then $(s_n)_{n=0}^{\infty}$ is bounded if and only if it has a bounded subsequence $(s_{n_k})_{k=0}^{\infty}$ (see Remark 2.105).

COROLLARY 4.10: COMPARISON TEST (MAJORANT/MINORANT)

Let $\sum_{k=0}^{\infty} a_k$ and $\sum_{k=0}^{\infty} b_k$ be series with $0 \leq a_k \leq b_k$ for all $k \in \mathbb{N}$. Then

$$0 \leq \sum_{k=0}^{\infty} a_k \leq \sum_{k=0}^{\infty} b_k,$$

and in particular

$$\begin{aligned} \sum_{k=0}^{\infty} b_k \text{ convergent} &\implies \sum_{k=0}^{\infty} a_k \text{ convergent}, \\ \sum_{k=0}^{\infty} a_k \text{ divergent to } +\infty &\implies \sum_{k=0}^{\infty} b_k \text{ divergent to } +\infty. \end{aligned}$$

These implications remain true if the inequalities $0 \leq a_n \leq b_n$ hold only for all $n \geq N$, for some $N \in \mathbb{N}$.

Proof. From $a_k \leq b_k$ we get $\sum_{k=0}^n a_k \leq \sum_{k=0}^n b_k$ for all $n \in \mathbb{N}$. Therefore

$$\sum_{k=0}^{\infty} a_k = \lim_{n \rightarrow \infty} \sum_{k=0}^n a_k \leq \lim_{n \rightarrow \infty} \sum_{k=0}^n b_k = \sum_{k=0}^{\infty} b_k.$$

18

The last part of the statement follows from Lemma 4.7. □

Under the assumptions of the corollary, $\sum_{k=0}^{\infty} b_k$ is called a *majorant* of $\sum_{k=0}^{\infty} a_k$, and $\sum_{k=0}^{\infty} a_k$ a *minorant* of $\sum_{k=0}^{\infty} b_k$. Hence the names **majorant** and **minorant criterion**.

EXAMPLE 4.11. — The series $\sum_{k=1}^{\infty} \frac{1}{k^2}$ is convergent. Indeed, for $k \geq 2$,

$$a_k = \frac{1}{k^2} \leq \frac{1}{k(k-1)} = b_k,$$

and since $\frac{1}{k(k-1)} = \frac{1}{k-1} - \frac{1}{k}$,

$$\sum_{k=2}^n b_k = \sum_{k=2}^n \left(\frac{1}{k-1} - \frac{1}{k} \right) = 1 - \frac{1}{n} \xrightarrow{n \rightarrow \infty} 1.$$

By the comparison test, $\sum_{k=2}^{\infty} \frac{1}{k^2} \leq \sum_{k=2}^{\infty} b_k = 1$, hence $\sum_{k=1}^{\infty} \frac{1}{k^2}$ converges.

EXAMPLE 4.12. — Consider $a_n = \frac{2n-10}{n^3-10n+100}$. Since $\lim_{n \rightarrow \infty} n^2 a_n = 2$, there exists $N \in \mathbb{N}$ such that for all $n \geq N$, $0 \leq n^2 a_n \leq 3$, or equivalently

$$0 \leq a_n \leq \frac{3}{n^2} \quad \forall n \geq N.$$

By Corollary 4.10 and Example 4.11, the series $\sum_{n=0}^{\infty} a_n$ converges.

PROPOSITION 4.13: CAUCHY CONDENSATION TEST

Let $(a_k)_{k=0}^{\infty}$ be a decreasing sequence of nonnegative numbers. Then

$$\sum_{k=0}^{\infty} a_k \text{ converges} \iff \sum_{k=0}^{\infty} 2^k a_{2^k} \text{ converges.}$$

Proof. Consider the partial sum of the series $\sum_{k=0}^{\infty} a_k$ starting from $k = 2$ up to an index that is a power of 2. Since the terms a_k are decreasing, the following inequalities hold:

$$\begin{aligned} \sum_{k=2}^{2^{n+1}} a_k &= a_2 + (a_3 + a_4) + (a_5 + \cdots + a_8) + \cdots + (a_{2^n+1} + \cdots + a_{2^{n+1}}) \\ &\leq \underbrace{a_1}_{=1 \cdot a_1} + \underbrace{(a_2 + a_2)}_{=2 \cdot a_2} + \underbrace{(a_4 + \cdots + a_4)}_{=4 \cdot a_4} + \cdots + \underbrace{(a_{2^n} + \cdots + a_{2^n})}_{=2^n a_{2^n}} \\ &= a_1 + 2a_2 + 4a_4 + \cdots + 2^n a_{2^n} = \sum_{k=0}^n 2^k a_{2^k}, \end{aligned}$$

and similarly,

$$\begin{aligned} \sum_{k=2}^{2^{n+1}} a_k &= a_2 + (a_3 + a_4) + (a_5 + \cdots + a_8) + \cdots + (a_{2^n+1} + \cdots + a_{2^{n+1}}) \\ &\geq \underbrace{a_2}_{=1 \cdot a_2} + \underbrace{(a_4 + a_4)}_{=2 \cdot a_4} + \underbrace{(a_8 + \cdots + a_8)}_{=4 \cdot a_8} + \cdots + \underbrace{(a_{2^n+1} + \cdots + a_{2^{n+1}})}_{=2^n a_{2^n+1}} \\ &= \frac{1}{2} (2a_2 + 4a_4 + \cdots + 2^{n+1} a_{2^{n+1}}) = \frac{1}{2} \sum_{k=1}^{n+1} 2^k a_{2^k}. \end{aligned}$$

In other words,

$$\sum_{k=0}^n 2^k a_{2^k} \geq \sum_{j=2}^{2^{n+1}} a_j \geq \frac{1}{2} \sum_{k=1}^{n+1} 2^k a_{2^k}.$$

By Remark 4.9 and Corollary 4.10, the partial sums of one series are bounded if and only if those of the other are. Hence, the two series converge or diverge together. \square

EXAMPLE 4.14. — For $p \in \mathbb{R}$, the series $\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges exactly when $p > 1$. Indeed:

- If $p \leq 0$, then $\frac{1}{n^p} \geq 1$, so the series diverges (e.g., by Proposition 4.3).
- If $p > 0$, the sequence $(\frac{1}{n^p})_{n \geq 1}$ is decreasing; so, it follows from Proposition 4.13,

$$\sum_{n=1}^{\infty} \frac{1}{n^p} \text{ converges} \iff \sum_{k=0}^{\infty} 2^k \frac{1}{(2^k)^p} = \sum_{k=0}^{\infty} (2^{1-p})^k \text{ converges.}$$

Recalling Example 4.4, this holds exactly when $2^{1-p} < 1$, i.e. $p > 1$.

REMARK 4.15. — The argument in Example 4.14 provides another proof that the harmonic series diverges (cf. Example 4.5).

EXERCISE 4.16. — For $p \in \mathbb{R}$, show that the series $\sum_{n=2}^{\infty} \frac{1}{n (\log n)^p}$ converges exactly when $p > 1$.

Hint: for $p \leq 0$, compare with the harmonic series; for $p > 0$, use Proposition 4.13 and Example 4.14.

EXERCISE 4.17. — Decide whether the series $\sum_{n=3}^{\infty} \frac{1}{n \log n \log \log n}$ converges or diverges.

18

4.1.2 Conditional Convergence

DEFINITION 4.18: ABSOLUTE AND CONDITIONAL CONVERGENCE

A series $\sum_{k=0}^{\infty} a_k$ is **absolutely convergent** if $\sum_{k=0}^{\infty} |a_k|$ converges. It is **conditionally convergent** if $\sum_{k=0}^{\infty} a_k$ converges but $\sum_{k=0}^{\infty} |a_k|$ diverges.

A striking feature of conditionally convergent series is that their terms can be rearranged to obtain any prescribed limit.

THEOREM 4.19: RIEMANN REARRANGEMENT THEOREM

Let $\sum_{n=0}^{\infty} a_n$ be a conditionally convergent series and let $A \in \mathbb{R}$. There exists a bijection $\varphi : \mathbb{N} \rightarrow \mathbb{N}$ such that

$$A = \sum_{n=0}^{\infty} a_{\varphi(n)}.$$

Extra material: Proof of Theorem 4.19

Sketch of proof. Since $\sum_{n=0}^{\infty} a_n$ is conditionally convergent, we have $a_n \rightarrow 0$ as $n \rightarrow \infty$ (by Proposition 4.3), while $\sum_{n=0}^{\infty} |a_n| = \infty$ by assumption.

Let

$$P = \{n \in \mathbb{N} : a_n \geq 0\}, \quad N = \{n \in \mathbb{N} : a_n < 0\},$$

and enumerate them in increasing order:

$$P = \{p_0 < p_1 < \dots\}, \quad N = \{n_0 < n_1 < \dots\}.$$

Because the series is conditionally convergent, one can check that

$$\sum_{k=0}^{\infty} a_{p_k} = +\infty, \quad \sum_{k=0}^{\infty} (-a_{n_k}) = +\infty;$$

otherwise, the original series would either diverge or be absolutely convergent.

Now, the idea of the construction is as follows: given $A \in \mathbb{R}$, we rearrange the terms by alternately appending the least unused positive terms until the partial sum first exceeds A , then the least unused negative terms until it drops below A , and so on. The divergence of the two subseries ensures that this process never terminates and that every index is eventually used.

More in detail, we define the bijection $\varphi : \mathbb{N} \rightarrow \mathbb{N}$ recursively. If $A < 0$, set $\varphi(0) = n_1$; if $A \geq 0$, set $\varphi(0) = p_1$. Assume $\varphi(0), \dots, \varphi(n)$ have been chosen, and let

$$s_n = \sum_{k=0}^n a_{\varphi(k)}.$$

Then define

$$\varphi(n+1) = \begin{cases} \min(P \setminus \{\varphi(0), \dots, \varphi(n)\}), & \text{if } s_n < A, \\ \min(N \setminus \{\varphi(0), \dots, \varphi(n)\}), & \text{if } s_n \geq A. \end{cases}$$

The map φ is injective by construction, and surjective because both subseries of positive and negative terms diverge to $+\infty$ (hence no index remains unused).

Since $a_n \rightarrow 0$ as $n \rightarrow \infty$, the successive corrections become arbitrarily small, and the sequence of partial sums $(s_n)_{n=0}^{\infty}$ converges to A . This establishes the theorem. \square

EXERCISE 4.20. — Fill in the details omitted in the proof of Theorem 4.19. Show also that one can obtain $A = \pm\infty$ by a suitable rearrangement.

4.1.3 Leibniz Criterion for Alternating Series

DEFINITION 4.21: ALTERNATING SERIES

If $(a_k)_{k=0}^{\infty}$ is a sequence of nonnegative numbers, the series

$$\sum_{k=0}^{\infty} (-1)^k a_k$$

is called the **alternating series** associated with the sequence $(a_k)_{k=0}^{\infty}$.

PROPOSITION 4.22: LEIBNIZ CRITERION

Let $(a_k)_{k=0}^{\infty}$ be a monotonically decreasing sequence of nonnegative numbers with $a_k \rightarrow 0$. Then the alternating series $\sum_{k=0}^{\infty} (-1)^k a_k$ converges, and for all $n \in \mathbb{N}$,

$$\sum_{k=0}^{2n+1} (-1)^k a_k \leq \sum_{k=0}^{\infty} (-1)^k a_k \leq \sum_{k=0}^{2n} (-1)^k a_k. \quad (4.1)$$

Proof. Let $s_n = \sum_{k=0}^n (-1)^k a_k$. Since the sequence $(a_n)_{n=0}^{\infty}$ is decreasing and non-negative, we have

$$\begin{aligned} s_{2n+2} &= s_{2n} - \underbrace{a_{2n+1} + a_{2n+2}}_{\leq 0} \leq s_{2n}, \\ s_{2n+1} &= s_{2n-1} + \underbrace{a_{2n} - a_{2n+1}}_{\geq 0} \geq s_{2n-1}, \\ s_{2n+2} &= s_{2n+1} + \underbrace{a_{2n+2}}_{\geq 0} \geq s_{2n+1}. \end{aligned}$$

for all $n \in \mathbb{N}$. In other words,

$$s_1 \leq s_3 \leq \dots \leq s_{2n-1} \leq s_{2n+1} \leq \dots \leq s_{2n+2} \leq s_{2n} \leq \dots \leq s_2 \leq s_0.$$

This implies that the sequence $(s_{2n})_{n=0}^{\infty}$ is decreasing and bounded below, while the sequence $(s_{2n+1})_{n=0}^{\infty}$ is increasing and bounded above. Thus, both limits $A = \lim_{n \rightarrow \infty} s_{2n+1}$ and $B = \lim_{n \rightarrow \infty} s_{2n}$ exist and satisfy

$$s_1 \leq s_3 \leq \dots \leq s_{2n-1} \leq s_{2n+1} \leq A \leq B \leq s_{2n+2} \leq s_{2n} \leq \dots \leq s_2 \leq s_0. \quad (4.2)$$

In particular

$$0 \leq B - A \leq s_{2n+2} - s_{2n+1} = a_{2n+2} \quad \forall n \in \mathbb{N},$$

and because $a_{2n+2} \rightarrow 0$, we deduce that $A = B$.

Also, (4.2) yields $s_{2n+1} \leq A = B \leq s_{2n}$, which corresponds exactly to (4.1). \square

EXAMPLE 4.23 (Alternating Harmonic Series). — The series

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

converges by Proposition 4.22, whereas $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges (Example 4.5); hence it is only conditionally convergent.

4.2 Absolute Convergence

In this section we will look at absolutely convergent series and prove some convergence criteria. As before, unless otherwise specified, all sequences consist of real numbers.

4.2.1 Criteria for Absolute Convergence

We begin by restating the concept of a Cauchy sequence in the context of convergent series.

THEOREM 4.24: CAUCHY CRITERION FOR SERIES

The series $\sum_{k=0}^{\infty} a_k$ converges if and only if, for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n > m \geq N$,

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

Proof. By definition, the series $\sum_{k=0}^{\infty} a_k$ converges if and only if the sequence of partial sums

$$s_n = \sum_{k=0}^n a_k$$

18

converges. By Theorem 2.120, this occurs if and only if $(s_n)_{n=0}^{\infty}$ is a Cauchy sequence, that is, $|s_n - s_m| < \varepsilon$ for all $n, m \geq N$. Since $s_n - s_m = 0$ when $n = m$, and the expression is symmetric in n and m , it suffices to consider the case $n > m$. In this case,

$$s_n - s_m = \sum_{k=m+1}^n a_k,$$

which proves the claim. □

EXAMPLE 4.25. — To see once more that the harmonic series diverges, note that for any $N \in \mathbb{N}$,

$$\sum_{k=N+1}^{2N} \frac{1}{k} \geq N \cdot \frac{1}{2N} = \frac{1}{2}.$$

Thus, the Cauchy condition fails for $\varepsilon = \frac{1}{2}$, and the series cannot converge.

We can now prove that absolutely convergent series do indeed converge.

PROPOSITION 4.26: ABSOLUTE CONVERGENCE IMPLIES CONVERGENCE

If a series $\sum_{n=0}^{\infty} a_n$ converges absolutely, then it converges and satisfies the generalized triangle inequality

$$\left| \sum_{n=0}^{\infty} a_n \right| \leq \sum_{n=0}^{\infty} |a_n|.$$

Proof. Since $\sum_{n=0}^{\infty} |a_n|$ converges, by the Cauchy criterion (Theorem 4.24) there exists $N \in \mathbb{N}$ such that, for all $n > m \geq N$,

$$\sum_{k=m+1}^n |a_k| < \varepsilon.$$

By the triangle inequality,

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

so $\sum_{n=0}^{\infty} a_n$ also satisfies the Cauchy criterion and therefore converges.

Moreover, again by the triangle inequality,

$$\left| \sum_{k=0}^n a_k \right| \leq \sum_{k=0}^n |a_k| \leq \sum_{k=0}^{\infty} |a_k| \quad \forall n \in \mathbb{N},$$

and taking the limit as $n \rightarrow \infty$ gives the desired inequality. \square

We now establish two classical criteria guaranteeing absolute convergence. In their proofs, we repeatedly use the following fact:

REMARK 4.27. — If a sequence $(x_n)_{n=0}^{\infty}$ converges to $\alpha \in \mathbb{R}$, then Proposition 2.95 implies the following facts:

- (i) for any $q > \alpha$ there exists $N \in \mathbb{N}$ such that $x_n < q$ for all $n \geq N$;
- (ii) for any $r < \alpha$ there exists $N \in \mathbb{N}$ such that $x_n > r$ for all $n \geq N$.

PROPOSITION 4.28: CAUCHY ROOT CRITERION

Given a sequence $(a_n)_{n=0}^{\infty}$, define

$$\alpha = \limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|} \in \mathbb{R} \cup \{\infty\}.$$

Then

$$\alpha < 1 \implies \sum_{n=0}^{\infty} a_n \text{ converges absolutely,} \quad \alpha > 1 \implies \sum_{n=0}^{\infty} a_n \text{ does not converge.}$$

Proof. Suppose $\alpha < 1$ and set $q = \frac{1+\alpha}{2}$, so that $q \in (\alpha, 1)$. By definition,

$$\limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|} = \lim_{n \rightarrow \infty} \sup_{k \geq n} \sqrt[k]{|a_k|}.$$

Thus $x_n = \sup_{k \geq n} \sqrt[k]{|a_k|} \rightarrow \alpha$. Since $\alpha < q$, Remark 4.27(i) implies the existence of $N \in \mathbb{N}$ such that

$$x_n = \sup_{k \geq n} \sqrt[k]{|a_k|} < q \quad \forall n \geq N,$$

therefore

$$|a_k| < q^k \quad \forall k \geq N,$$

Since $q < 1$, $\sum_{k=N}^{\infty} |a_k|$ converges by comparison with the geometric series (Example 4.4), so $\sum_{n=0}^{\infty} a_n$ converges absolutely.

If $\alpha > 1$, since the limsup is an accumulation point (Theorem 2.112), Proposition 2.88 implies the existence of a subsequence $(a_{n_k})_{k=0}^{\infty}$ such that $\lim_{k \rightarrow \infty} \sqrt[n_k]{|a_{n_k}|} = \alpha$. Hence, thanks to Remark 4.27(ii) with $r = 1$, $\sqrt[n_k]{|a_{n_k}|} > 1$ for all k large, or equivalently, $|a_{n_k}| > 1$ for large k . In particular, the sequence $(a_n)_{n=0}^{\infty}$ does not converge to 0. Recalling Proposition 4.3, this implies that the series $\sum_{n=0}^{\infty} a_n$ does not converge. \square

REMARK 4.29. — If $\alpha = 1$, the root criterion is inconclusive:

- $\sqrt[n]{1/n} \rightarrow 1$ (Exercise 3.53) but $\sum_{n=1}^{\infty} 1/n$ diverges (Example 4.5);
- $\sqrt[n]{1/n^2} \rightarrow 1$ (Exercise 3.53) but $\sum_{n=1}^{\infty} 1/n^2$ converges (Example 4.11).

PROPOSITION 4.30: D'ALEMBERT'S QUOTIENT CRITERION

Given a sequence $(a_n)_{n=0}^{\infty}$ with $a_n \neq 0$ for all n , assume that

$$\lim_{n \rightarrow \infty} \frac{|a_{n+1}|}{|a_n|} = \alpha \in [0, \infty).$$

Then

$$\alpha < 1 \implies \sum_{n=0}^{\infty} a_n \text{ converges absolutely,} \quad \alpha > 1 \implies \sum_{n=0}^{\infty} a_n \text{ does not converge.}$$

Proof. The proof parallels that of the root criterion.

If $\alpha < 1$, set $q = \frac{1+\alpha}{2} \in (\alpha, 1)$. Since $\frac{|a_{n+1}|}{|a_n|} \rightarrow \alpha$ and $\alpha < q$, by Remark 4.27(i) there exists $N \in \mathbb{N}$ such that

$$\frac{|a_{k+1}|}{|a_k|} < q \quad \forall k \geq N.$$

This gives

$$|a_k| = \frac{|a_k|}{|a_{k-1}|} \cdot \frac{|a_{k-1}|}{|a_{k-2}|} \cdots \frac{|a_{N+1}|}{|a_N|} \cdot |a_N| < q^{k-N} |a_N| = \frac{|a_N|}{q^N} q^k \quad \forall k \geq N.$$

Since $q < 1$, the geometric comparison test shows that $\sum_{n=0}^{\infty} |a_n|$ converges absolutely.

If $\alpha > 1$, then Remark 4.27(ii) with $r = 1$ implies the existence of $N \in \mathbb{N}$ such that

$$\frac{|a_{k+1}|}{|a_k|} > 1 \quad \forall k \geq N.$$

In particular,

$$|a_k| = \frac{|a_k|}{|a_{k-1}|} \cdot \frac{|a_{k-1}|}{|a_{k-2}|} \cdots \frac{|a_{N+1}|}{|a_N|} \cdot |a_N| > |a_N| \quad \forall k \geq N.$$

Hence $(a_n)_{n=0}^{\infty}$ does not tend to 0, and by Proposition 4.3 the series does not converge. \square

EXERCISE 4.31 (Generalization of the Quotient Criterion). — Let $(a_n)_{n=0}^{\infty}$ be a sequence of nonzero numbers, and define

$$\alpha_+ = \limsup_{n \rightarrow \infty} \frac{|a_{n+1}|}{|a_n|}, \quad \alpha_- = \liminf_{n \rightarrow \infty} \frac{|a_{n+1}|}{|a_n|}.$$

Show that

$$\alpha_+ < 1 \implies \sum_{n=0}^{\infty} a_n \text{ converges absolutely,} \quad \alpha_- > 1 \implies \sum_{n=0}^{\infty} a_n \text{ does not converge.}$$

19

Is the second implication still true if α_- is replaced by α_+ ?

4.2.2 Reordering Series

THEOREM 4.32: REARRANGEMENT OF ABSOLUTELY CONVERGENT SERIES

Let $\sum_{n=0}^{\infty} a_n$ be an absolutely convergent series, and let $\varphi : \mathbb{N} \rightarrow \mathbb{N}$ be a bijection. Then $\sum_{n=0}^{\infty} a_{\varphi(n)}$ is absolutely convergent, and

$$\sum_{n=0}^{\infty} a_n = \sum_{n=0}^{\infty} a_{\varphi(n)}. \quad (4.3)$$

Proof. Fix $\varepsilon > 0$. Since $\sum_{n=0}^{\infty} |a_n|$ converges, there exists $N \in \mathbb{N}$ such that

$$\sum_{k=N+1}^{\infty} |a_k| < \frac{\varepsilon}{2}.$$

Let

$$M = \max \{ \varphi^{-1}(0), \dots, \varphi^{-1}(N) \}.$$

Equivalently, $M \in \mathbb{N}$ is the smallest number such that

$$\{a_0, \dots, a_N\} \subset \{a_{\varphi(0)}, \dots, a_{\varphi(M)}\}.$$

Then

$$\{a_0, \dots, a_N\} \subset \{a_{\varphi(0)}, \dots, a_{\varphi(n)}\} \quad \forall n \geq M,$$

therefore

$$\sum_{\ell=0}^n a_{\varphi(\ell)} - \sum_{k=0}^N a_k = \sum_{\substack{0 \leq \ell \leq n \\ \varphi(\ell) > N}} a_{\varphi(\ell)}.$$

Moreover, since all indices $\varphi(\ell) > N$ with $0 \leq \ell \leq n$ correspond to terms among $\{|a_k| \mid k \geq N+1\}$, we have

$$\sum_{\substack{0 \leq \ell \leq n \\ \varphi(\ell) > N}} |a_{\varphi(\ell)}| \leq \sum_{k=N+1}^{\infty} |a_k|.$$

This implies that, for $n \geq M$, we can estimate

$$\begin{aligned} \left| \sum_{\ell=0}^n a_{\varphi(\ell)} - \sum_{k=0}^{\infty} a_k \right| &= \left| \sum_{\ell=0}^n a_{\varphi(\ell)} - \sum_{k=0}^N a_k - \sum_{k=N+1}^{\infty} a_k \right| = \left| \sum_{\substack{0 \leq \ell \leq n \\ \varphi(\ell) > N}} a_{\varphi(\ell)} - \sum_{k=N+1}^{\infty} a_k \right| \\ &\leq \sum_{\substack{0 \leq \ell \leq n \\ \varphi(\ell) > N}} |a_{\varphi(\ell)}| + \sum_{k=N+1}^{\infty} |a_k| \leq 2 \sum_{k=N+1}^{\infty} |a_k| < \varepsilon. \end{aligned}$$

This shows that

$$\sum_{\ell=0}^n a_{\varphi(\ell)} \rightarrow \sum_{k=0}^{\infty} a_k \quad \text{as } n \rightarrow \infty,$$

which proves the identity (4.3). Applying the same reasoning to $\sum_{n=0}^{\infty} |a_n|$ shows that $\sum_{n=0}^{\infty} |a_{\varphi(n)}| = \sum_{n=0}^{\infty} |a_n| < \infty$, hence $\sum a_{\varphi(n)}$ is absolutely convergent. \square

4.2.3 Products of Series

Our next goal is to define the product of two absolutely convergent series $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty} b_n$. The following result shows that, in this case, the product can be obtained by summing all possible products $a_j b_k$ of their terms in *any* order (that is, by choosing an arbitrary bijection between \mathbb{N} and $\mathbb{N} \times \mathbb{N}$).

THEOREM 4.33: PRODUCT THEOREM

Let $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty} b_n$ be absolutely convergent series, and let $\alpha : \mathbb{N} \rightarrow \mathbb{N} \times \mathbb{N}$ be a bijection. Writing $\alpha(n) = (\alpha_1(n), \alpha_2(n))$, one has

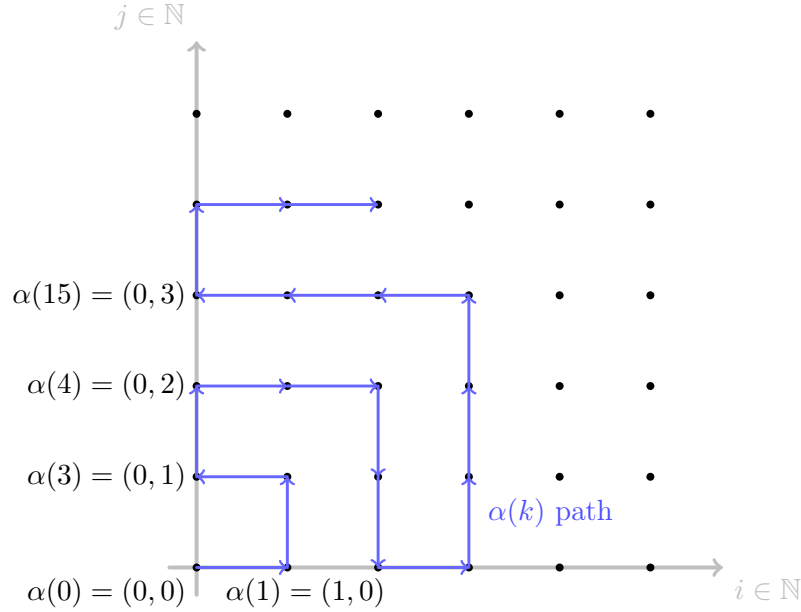
$$\left(\sum_{n=0}^{\infty} a_n \right) \left(\sum_{n=0}^{\infty} b_n \right) = \sum_{n=0}^{\infty} a_{\alpha_1(n)} b_{\alpha_2(n)}, \quad (4.4)$$

and the series on the right converges absolutely.

Proof. Consider first a bijection $\alpha : \mathbb{N} \rightarrow \mathbb{N} \times \mathbb{N}$, written $\alpha(n) = (\alpha_1(n), \alpha_2(n))$, such that

$$\{\alpha(k) : 0 \leq k < n^2\} = \{0, 1, \dots, n-1\}^2 \quad \forall n \in \mathbb{N}.$$

For example, $(\alpha(n))_{n=0}^\infty$ could traverse the grid $\mathbb{N} \times \mathbb{N}$ as shown in Figure 4.1 below.



19

Figure 4.1: Illustration of the bijection $\alpha : \mathbb{N} \rightarrow \mathbb{N} \times \mathbb{N}$, which enumerates all lattice points in \mathbb{N}^2 by following the blue path. For each n , the first n^2 elements $\{\alpha(k) \mid 0 \leq k < n^2 - 1\}$ exactly cover the square $\{0, 1, \dots, n-1\}^2$.

Then, for every $n \in \mathbb{N}$,

$$\sum_{k=0}^{n^2-1} |a_{\alpha_1(k)}| |b_{\alpha_2(k)}| = \left(\sum_{\ell=0}^{n-1} |a_\ell| \right) \left(\sum_{m=0}^{n-1} |b_m| \right).$$

Since the right-hand side is bounded by

$$\left(\sum_{\ell=0}^{\infty} |a_\ell| \right) \left(\sum_{m=0}^{\infty} |b_m| \right),$$

we have

$$\sup_{n \in \mathbb{N}} \sum_{k=0}^{n^2-1} |a_{\alpha_1(k)}| |b_{\alpha_2(k)}| \leq \left(\sum_{\ell=0}^{\infty} |a_\ell| \right) \left(\sum_{m=0}^{\infty} |b_m| \right) < \infty.$$

This implies that the series $\sum_{k=0}^{\infty} a_{\alpha_1(k)} b_{\alpha_2(k)}$ converges absolutely. In particular, since it converges, its value can be computed along every subsequence, therefore

$$\sum_{k=0}^{\infty} a_{\alpha_1(k)} b_{\alpha_2(k)} = \lim_{n \rightarrow \infty} \sum_{k=0}^{n^2-1} a_{\alpha_1(k)} b_{\alpha_2(k)}$$

Now, writing the identity

$$\sum_{k=0}^{n^2-1} a_{\alpha_1(k)} b_{\alpha_2(k)} = \left(\sum_{\ell=0}^{n-1} a_{\ell} \right) \left(\sum_{m=0}^{n-1} b_m \right),$$

and taking the limit as $n \rightarrow \infty$, Proposition 2.94(2) gives

$$\begin{aligned} \sum_{k=0}^{\infty} a_{\alpha_1(k)} b_{\alpha_2(k)} &= \lim_{n \rightarrow \infty} \sum_{k=0}^{n^2-1} a_{\alpha_1(k)} b_{\alpha_2(k)} \\ &= \left(\lim_{n \rightarrow \infty} \sum_{\ell=0}^{n-1} a_{\ell} \right) \left(\lim_{n \rightarrow \infty} \sum_{m=0}^{n-1} b_m \right) = \left(\sum_{n=0}^{\infty} a_n \right) \left(\sum_{n=0}^{\infty} b_n \right), \end{aligned}$$

which proves (4.4) for this specific bijection α .

For an arbitrary bijection $\beta : \mathbb{N} \rightarrow \mathbb{N}^2$, define $\varphi = \alpha^{-1} \circ \beta : \mathbb{N} \rightarrow \mathbb{N}$. Then $\beta = \alpha \circ \varphi$, and writing $\beta(n) = (\beta_1(n), \beta_2(n)) = (\alpha_1(\varphi(n)), \alpha_2(\varphi(n)))$, the Rearrangement Theorem 4.32 yields

$$\sum_{n=0}^{\infty} a_{\beta_1(n)} b_{\beta_2(n)} = \sum_{n=0}^{\infty} a_{\alpha_1(\varphi(n))} b_{\alpha_2(\varphi(n))} = \sum_{n=0}^{\infty} a_{\alpha_1(n)} b_{\alpha_2(n)} = \left(\sum_{n=0}^{\infty} a_n \right) \left(\sum_{n=0}^{\infty} b_n \right).$$

□

COROLLARY 4.34: CAUCHY PRODUCT

If $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty} b_n$ are absolutely convergent, then

$$\left(\sum_{n=0}^{\infty} a_n \right) \left(\sum_{n=0}^{\infty} b_n \right) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^n a_{n-k} b_k \right),$$

and the series on the right converges absolutely.

Proof. Consider the bijection $\alpha : \mathbb{N} \rightarrow \mathbb{N} \times \mathbb{N}$ defined as the inverse of the function $f_5 : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$ from Example 2.70(3), represented in Figure 2.10. Explicitly,

$$\begin{aligned} \alpha(0) &= (0, 0), & \alpha(1) &= (1, 0), & \alpha(2) &= (0, 1), & \alpha(3) &= (2, 0), & \alpha(4) &= (1, 1), & \dots \\ \alpha(20) &= (0, 5), & \dots, & \alpha(31) &= (4, 3), & \dots, & \alpha(49) &= (5, 4), & \dots, & \text{etc.} \end{aligned}$$

By Theorem 4.33,

$$\left(\sum_{n=0}^{\infty} a_n \right) \left(\sum_{n=0}^{\infty} b_n \right) = \sum_{n=0}^{\infty} a_{\alpha_1(n)} b_{\alpha_2(n)}.$$

Listing the terms explicitly and grouping them by diagonals as in Figure 4.2, we obtain

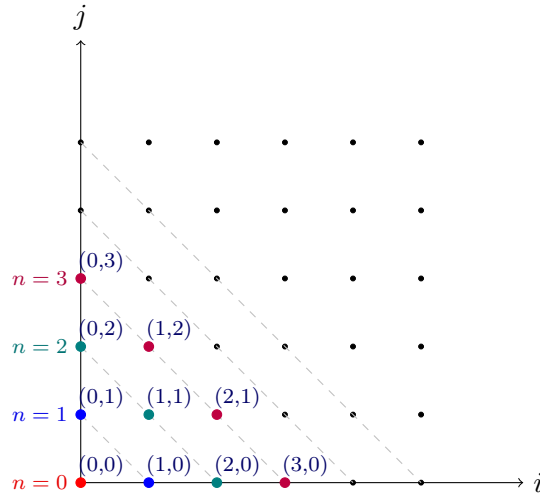


Figure 4.2: Grouping by diagonals $j + k = n$ in $\mathbb{N} \times \mathbb{N}$. Each diagonal corresponds to the set of terms $a_{n-k}b_k$ appearing in the n -th partial sum of the Cauchy product.

$$\begin{aligned} \sum_{n=0}^{\infty} a_{\alpha_1(n)} b_{\alpha_2(n)} &= a_0 b_0 + (a_0 b_1 + a_1 b_0) + (a_2 b_0 + a_1 b_1 + a_0 b_2) \\ &\quad + (a_3 b_0 + a_2 b_1 + a_1 b_2 + a_0 b_3) + \cdots \\ &= \sum_{n=0}^{\infty} \left(\sum_{\substack{j,k \geq 0 \\ j+k=n}} a_j b_k \right) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^n a_{n-k} b_k \right). \end{aligned}$$

Finally, absolute convergence follows from the triangle inequality and Theorem 4.33:

$$\sum_{n=0}^{\infty} \left| \sum_{k=0}^n a_{n-k} b_k \right| \leq \sum_{n=0}^{\infty} \sum_{k=0}^n |a_{n-k} b_k| = \sum_{n=0}^{\infty} |a_{\alpha_1(n)}| |b_{\alpha_2(n)}| < \infty.$$

□

EXAMPLE 4.35. — Let $q \in \mathbb{R}$ with $|q| < 1$. Then $\sum_{n=0}^{\infty} q^n$ converges absolutely (see Example 4.4). Applying the Cauchy product to this series with itself yields

$$\frac{1}{(1-q)^2} = \left(\sum_{n=0}^{\infty} q^n \right)^2 = \sum_{n=0}^{\infty} \sum_{k=0}^n q^{n-k} q^k = \sum_{n=0}^{\infty} (n+1) q^n.$$

This gives us an explicit formula for the value of the series $\sum_{n=0}^{\infty} n q^n$:

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

4.3 Series of Complex Numbers

To define the notion of a convergent series in \mathbb{C} , it is sufficient to consider separately the corresponding series of its real and imaginary parts in \mathbb{R} .

DEFINITION 4.36: SERIES OF COMPLEX NUMBERS

Let $(z_n)_{n=0}^\infty = (x_n + iy_n)_{n=0}^\infty$ be a sequence of complex numbers, and let $Z = A + iB \in \mathbb{C}$. The series $\sum_{n=0}^\infty z_n$ is said to **converge** to Z if both real series $\sum_{n=0}^\infty x_n$ and $\sum_{n=0}^\infty y_n$ converge, with limits A and B , respectively:

$$\sum_{n=0}^\infty x_n = A, \quad \sum_{n=0}^\infty y_n = B.$$

We say that $\sum_{n=0}^\infty z_n$ **converges absolutely** if the series of moduli $\sum_{n=0}^\infty |z_n|$ converges.

Whenever the series $\sum_{n=0}^\infty z_n$ and $\sum_{n=0}^\infty w_n$ converge absolutely, their sum and product are given (exactly as in the real case) by the following formulas:

$$\begin{aligned} \sum_{n=0}^\infty z_n + \sum_{n=0}^\infty w_n &= \sum_{n=0}^\infty (z_n + w_n), \\ \left(\sum_{n=0}^\infty z_n \right) \left(\sum_{n=0}^\infty w_n \right) &= \sum_{n=0}^\infty \left(\sum_{k=0}^n z_{n-k} w_k \right). \end{aligned} \tag{4.5}$$

REMARK 4.37. — Let $(z_n)_{n=0}^\infty = (x_n + iy_n)_{n=0}^\infty$ be a sequence of complex numbers, and assume that the series $\sum_{n=0}^\infty |z_n|$ converges. Since

$$0 \leq |x_n| \leq |z_n|, \quad 0 \leq |y_n| \leq |z_n| \quad \forall n \in \mathbb{N},$$

the Majorant Criterion (Corollary 4.10) implies that both $\sum_{n=0}^\infty |x_n|$ and $\sum_{n=0}^\infty |y_n|$ converge. Hence, the series of real and imaginary parts are absolutely convergent.

Conversely, since $|z_n| \leq |x_n| + |y_n|$, the absolute convergence of $\sum_{n=0}^\infty x_n$ and $\sum_{n=0}^\infty y_n$ also implies the absolute convergence of $\sum_{n=0}^\infty z_n$. Therefore, absolute convergence in \mathbb{C} is equivalent to absolute convergence of the real and imaginary parts.

4.4 Power Series

Our next goal is to investigate power series. These are series where the terms are powers of the variable $x \in \mathbb{R}$ (or $z \in \mathbb{C}$, if one considers complex power series) multiplied by coefficients.

4.4.1 Radius of Convergence

DEFINITION 4.38: POWER SERIES

A **power series** with real coefficients is a series of the form

$$\sum_{n=0}^{\infty} a_n x^n,$$

where $(a_n)_{n=0}^{\infty}$ is a sequence in \mathbb{R} and $x \in \mathbb{R}$. Here, x is the **variable**, and $a_n \in \mathbb{R}$ is the **coefficient** of x^n .

By convention, we set $x^0 = 1$ for all $x \in \mathbb{R}$, including $x = 0$. In other words, the first term of the power series is always a_0 .

20

Addition and multiplication of power series are given by

$$\begin{aligned} \sum_{n=0}^{\infty} a_n x^n + \sum_{n=0}^{\infty} b_n x^n &= \sum_{n=0}^{\infty} (a_n + b_n) x^n, \\ \left(\sum_{n=0}^{\infty} a_n x^n \right) \left(\sum_{n=0}^{\infty} b_n x^n \right) &= \sum_{n=0}^{\infty} \left(\sum_{k=0}^n a_{n-k} b_k \right) x^n, \end{aligned}$$

where the product formula follows from Corollary 4.34:

$$\left(\sum_{n=0}^{\infty} a_n x^n \right) \left(\sum_{n=0}^{\infty} b_n x^n \right) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^n a_{n-k} x^{n-k} b_k x^k \right) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^n a_{n-k} b_k \right) x^n.$$

A power series is a polynomial whenever only finitely many of its coefficients are *nonzero*.

The convergence of a power series depends on the coefficients $(a_n)_{n=0}^{\infty}$ and is characterized in Theorem 4.41.

DEFINITION 4.39: RADIUS OF CONVERGENCE

Let $\sum_{n=0}^{\infty} a_n x^n$ be a power series, and define

$$\rho := \limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|}.$$

The **radius of convergence** is defined as

$$R = \begin{cases} 0, & \rho = \infty, \\ \rho^{-1}, & 0 < \rho < \infty, \\ \infty, & \rho = 0. \end{cases}$$

(Equivalently, $R = \frac{1}{\rho}$ with the conventions $1/0 = \infty$ and $1/\infty = 0$.)

In the following, when we write $R \in [0, \infty]$, we mean that R is either a nonnegative real number or $R = \infty$.

EXERCISE 4.40. — For each $R \in [0, \infty]$, find a power series $\sum_{n=0}^{\infty} a_n x^n$ with radius of convergence R .

20

THEOREM 4.41: CONVERGENCE OF POWER SERIES

Let $\sum_{n=0}^{\infty} a_n x^n$ be a power series with radius of convergence $R \in (0, \infty]$. Then $\sum_{n=0}^{\infty} a_n x^n$ converges absolutely for all $x \in \mathbb{R}$ with $|x| < R$, and does not converge for all $x \in \mathbb{R}$ with $|x| > R$. In particular, for $x \in (-R, R)$ we can define the function $f(x) = \sum_{n=0}^{\infty} a_n x^n$.

Proof. Let $x \in \mathbb{R}$, and write $\rho = \limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|}$ as in Definition 4.39. Then

$$\limsup_{n \rightarrow \infty} \sqrt[n]{|a_n x^n|} = \left(\limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|} \right) |x| = \rho |x|.$$

By the root criterion (see Proposition 4.28), the series $\sum_{n=0}^{\infty} a_n x^n$ converges absolutely if $\rho |x| < 1$, and does not converge if $\rho |x| > 1$ (in particular, if $\rho = 0$, then it converges absolutely for all $x \in \mathbb{R}$). Since $R = \frac{1}{\rho}$, the result follows. \square

THEOREM 4.42: CONTINUITY OF POWER SERIES

Let $\sum_{n=0}^{\infty} a_n x^n$ be a power series with radius of convergence $R \in (0, \infty]$, and define the polynomials $f_n(x) = \sum_{k=0}^n a_k x^k$. For any $r \in (0, R)$, the sequence $(f_n)_{n=0}^{\infty}$ converges uniformly to f on $[-r, r]$. In particular, the power series defines a continuous function $f : (-R, R) \rightarrow \mathbb{R}$.

Proof. By Theorem 4.41 with $x = r$, the series $\sum_{n=0}^{\infty} |a_n| r^n$ converges. Hence, for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $\sum_{k=N+1}^{\infty} |a_k| r^k < \varepsilon$. Thus, for all $x \in [-r, r]$ and all $n \geq N$,

$$|f_n(x) - f(x)| = \left| \sum_{k=n+1}^{\infty} a_k x^k \right| \leq \sum_{k=n+1}^{\infty} |a_k| |x|^k \leq \sum_{k=N+1}^{\infty} |a_k| r^k < \varepsilon.$$

This shows that $(f_n)_{n=0}^{\infty}$ converges uniformly to f on $[-r, r]$. Since each f_n is continuous (being a polynomial), Theorem 3.83 implies that f is continuous on $[-r, r]$. As $r < R$ is arbitrary, f is continuous on $(-R, R)$. \square

EXAMPLE 4.43. — In general, the partial sums $f_n(x) = \sum_{k=0}^n a_k x^k$ do *not* converge uniformly to $f(x) = \sum_{k=0}^{\infty} a_k x^k$ on the whole interval $(-R, R)$.

To see this, consider the geometric series $\sum_{n=0}^{\infty} x^n$. Its radius of convergence is $R = 1$, and on $(-1, 1)$ we have $f(x) = \sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$ (see Example 4.4). If the convergence on $(-1, 1)$ were uniform, then applying the notion of uniform convergence with $\varepsilon = 1$ would give $N \in \mathbb{N}$ such that, for all $n \geq N$ and $x \in (-1, 1)$,

$$\left| \sum_{k=0}^n x^k - \frac{1}{1-x} \right| < 1.$$

Taking $n = N$ and using the triangle inequality, we would get

$$\left| \frac{1}{1-x} \right| < 1 + \left| \sum_{k=0}^N x^k \right| \leq 1 + \sum_{k=0}^N |x|^k \leq 1 + (N+1) = N+2 \quad \forall x \in (-1, 1),$$

a contradiction since $\lim_{x \rightarrow 1^-} \frac{1}{1-x} = \infty$.

EXERCISE 4.44. — Find the radius of convergence R of the power series

$$\sum_{n=1}^{\infty} \frac{(\sqrt{n^2+n} - \sqrt{n^2+1})^n}{n^2} x^n,$$

and study convergence at $x = R$ and $x = -R$.

EXERCISE 4.45. — Let $\sum_{n=0}^{\infty} a_n x^n$ be a power series with $a_n \neq 0$ for all $n \in \mathbb{N}$, and assume that the limit $\lim_{n \rightarrow \infty} \frac{|a_n|}{|a_{n+1}|}$ exists. Then

$$R = \lim_{n \rightarrow \infty} \frac{|a_n|}{|a_{n+1}|}.$$

Hint: Define $L = \lim_{n \rightarrow \infty} \frac{|a_n|}{|a_{n+1}|}$. Combining Proposition 4.30 and Proposition 2.94(4), deduce that $\sum_{n=0}^{\infty} a_n x^n$ converges absolutely for all $x \in \mathbb{R}$ with $|x| < L$, and does not converge for all $x \in \mathbb{R}$ with $|x| > L$. Comparing this statement with Theorem 4.41, conclude that $L = R$.

PROPOSITION 4.46: RADIUS OF CONVERGENCE OF SUM AND PRODUCT

Let $R \geq 0$, and let $\sum_{n=0}^{\infty} a_n x^n$ and $\sum_{n=0}^{\infty} b_n x^n$ be power series with radius of convergence at least R . Then their sum and their Cauchy product also have radius of convergence at least R .

Proof. By linearity and Corollary 4.34, the absolute convergence of $\sum a_n x^n$ and $\sum b_n x^n$ for $|x| < R$ implies that

$$\sum_{n=0}^{\infty} (a_n + b_n) x^n \quad \text{and} \quad \left(\sum_{n=0}^{\infty} a_n x^n \right) \left(\sum_{n=0}^{\infty} b_n x^n \right) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^n a_{n-k} b_k \right) x^n$$

both converge absolutely for $|x| < R$. Since a power series cannot converge for $|x|$ larger than its radius of convergence, each has radius of convergence at least R . \square

EXAMPLE 4.47. — If $\sum_{n=0}^{\infty} a_n x^n$ has radius of convergence at least 1, then

$$\frac{1}{1-x} \sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} (a_0 + \cdots + a_n) x^n \quad \forall x \in (-1, 1). \quad (4.6)$$

20

Indeed, since $\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$ for $x \in (-1, 1)$ and has radius 1, (4.6) follows from Proposition 4.46.

EXERCISE 4.48. — Compute $\sum_{n=1}^{\infty} n 2^{-n}$.

4.4.2 Complex Power Series

Analogously to the real case, we can consider series with complex coefficients and a complex variable $z \in \mathbb{C}$.

DEFINITION 4.49: COMPLEX POWER SERIES

A **complex power series** with complex coefficients is a series of the form

$$\sum_{n=0}^{\infty} a_n z^n,$$

where $(a_n)_{n=0}^{\infty}$ is a sequence in \mathbb{C} and $z \in \mathbb{C}$.

Again, by convention, we set $z^0 = 1$ for all $z \in \mathbb{C}$, including $z = 0$.

Addition and multiplication are defined as in the real case:

$$\begin{aligned} \sum_{n=0}^{\infty} a_n z^n + \sum_{n=0}^{\infty} b_n z^n &= \sum_{n=0}^{\infty} (a_n + b_n) z^n, \\ \left(\sum_{n=0}^{\infty} a_n z^n \right) \left(\sum_{n=0}^{\infty} b_n z^n \right) &= \sum_{n=0}^{\infty} \left(\sum_{k=0}^n a_{n-k} b_k \right) z^n. \end{aligned} \quad (4.7)$$

The **radius of convergence** is defined exactly as in Definition 4.39, replacing $|x|$ by $|z|$. With these definitions, the following theorem holds, and the proof is identical to the real case.

THEOREM 4.50: CONVERGENCE OF COMPLEX POWER SERIES

Let $\sum_{n=0}^{\infty} a_n z^n$ be a power series with radius of convergence $R \in (0, \infty]$. Then the series $\sum_{n=0}^{\infty} a_n z^n$ converges absolutely for all $z \in \mathbb{C}$ with $|z| < R$, and diverges for all $z \in \mathbb{C}$ with $|z| > R$. In particular, for $|z| < R$ one can define the (complex-valued) function

$$f(z) = \sum_{n=0}^{\infty} a_n z^n.$$

Extra Material: Continuity of Complex Power Series

Before stating the analogue of Theorem 4.42, we recall the notions of continuity and uniform convergence for functions defined on subsets of \mathbb{C} .

DEFINITION 4.51: CONTINUITY FOR COMPLEX FUNCTIONS

Let $D \subset \mathbb{C}$ and $f : D \rightarrow \mathbb{C}$. We say that f is **continuous at a point** $z_0 \in D$ if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\forall z \in D, \quad |z - z_0| < \delta \implies |f(z) - f(z_0)| < \varepsilon.$$

If f is continuous at every $z_0 \in D$, then f is said to be **continuous on** D .

DEFINITION 4.52: UNIFORM CONVERGENCE FOR COMPLEX FUNCTIONS

Let $D \subset \mathbb{C}$ and let $(f_n)_{n=0}^{\infty}$ be a sequence of functions $f_n : D \rightarrow \mathbb{C}$ and let $f : D \rightarrow \mathbb{C}$. We say that $(f_n)_{n=0}^{\infty}$ **converges uniformly** to f on D if for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n \geq N$ and all $z \in D$,

$$|f_n(z) - f(z)| < \varepsilon.$$

These definitions are identical in form to the real case, with the only difference that $|\cdot|$ now denotes the modulus on \mathbb{C} . With these definition, some proofs about continuous functions can be repeated verbatim. In particular, polynomial functions are continuous; also, Theorem 3.83

remains valid: the uniform limit of continuous complex-valued functions is continuous. As a consequence, we get the following:

THEOREM 4.53: CONTINUITY OF COMPLEX POWER SERIES

Let $\sum_{n=0}^{\infty} a_n z^n$ be a complex power series with radius of convergence $R \in (0, \infty]$, and define the partial sums

$$f_n(z) = \sum_{k=0}^n a_k z^k.$$

For every $r \in (0, R)$, the sequence $(f_n)_{n=0}^{\infty}$ converges uniformly to f for $|z| \leq r$. In particular, the power series defines a continuous function for $|z| < R$.

Proof. The proof is identical to the real case. Since $\sum_{n=0}^{\infty} |a_n| r^n < \infty$ for every $r < R$, the sequence of partial sums $(f_n)_{n=0}^{\infty}$ converges uniformly for $|z| \leq r$. Since each f_n is a polynomial in z , it is continuous on \mathbb{C} . Since the uniform limit of continuous functions is continuous, f is continuous for $|z| \leq r$. Finally, since $r < R$ is arbitrary, f is continuous for $|z| < R$. \square

4.5 Example: Exponential and Trigonometric Functions

4.5.1 The Exponential Map as a Power Series

In Section 3.4 we introduced the real exponential function and established its main properties. We now show that the exponential map can alternatively be defined via the **exponential series**

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}, \quad (4.8)$$

where

$$0! = 1, \quad n! = 1 \cdot 2 \cdot \dots \cdot n.$$

Since $\frac{n!}{(n+1)!} = \frac{1}{n+1} \rightarrow 0$ as $n \rightarrow \infty$, it follows directly from the quotient criterion (see Exercise 4.45) that this series has infinite radius of convergence. Hence, Theorem 4.42 implies that the right-hand side of (4.8) defines a continuous function on \mathbb{R} .

REMARK 4.54. — Alternatively, given $N \in \mathbb{N}$, we note that

$$n! \geq \underbrace{n \cdot (n-1) \cdot \dots \cdot (N+1) \cdot N}_{n-N+1 \text{ terms}} \geq N^{n-N+1};$$

therefore, since $\frac{n-N+1}{n} \rightarrow 1$ as $n \rightarrow \infty$

$$\rho = \limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|} \leq \lim_{n \rightarrow \infty} \frac{1}{N^{\frac{n-N+1}{n}}} = \frac{1}{N}.$$

Since N can be chosen arbitrarily large, we conclude that $\rho = 0$, therefore $R = \infty$.

The representation of the exponential function as a power series is, in many ways, more flexible than its definition as a limit. Moreover, as we shall see, its complex version will naturally connect with the sine and cosine functions.

Before discussing these connections, we first show that the two representations of the exponential function (via series and via limit) coincide.

PROPOSITION 4.55: EXPONENTIAL MAP AS POWER SERIES

For every $x \in \mathbb{R}$,

$$\sum_{k=0}^{\infty} \frac{x^k}{k!} = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n.$$

Extra material: Proof of Proposition 4.55

Proof. For every $n \geq 0$, we have

$$\left(1 + \frac{x}{n}\right)^n = \sum_{k=0}^n \binom{n}{k} \frac{x^k}{n^k} = \sum_{k=0}^n \frac{x^k}{k!} \prod_{\ell=0}^{k-1} \left(1 - \frac{\ell}{n}\right).$$

Fix $x \in \mathbb{R}$ and $\varepsilon > 0$. Since $\sum_{k=0}^{\infty} \frac{|x|^k}{k!} < \infty$, there exists $N \in \mathbb{N}$ such that

$$\sum_{k=N+1}^{\infty} \frac{|x|^k}{k!} < \frac{\varepsilon}{2}.$$

In particular,

$$\left| \sum_{k=0}^N \frac{x^k}{k!} - \sum_{k=0}^{\infty} \frac{x^k}{k!} \right| \leq \sum_{k=N+1}^{\infty} \frac{|x|^k}{k!} < \frac{\varepsilon}{2}. \quad (4.9)$$

Moreover, for $n \geq N$,

$$\begin{aligned} \left| \sum_{k=0}^N \frac{x^k}{k!} - \sum_{k=0}^n \frac{x^k}{k!} \prod_{\ell=0}^{k-1} \left(1 - \frac{\ell}{n}\right) \right| &\leq \sum_{k=0}^N \frac{|x|^k}{k!} \left(1 - \prod_{\ell=0}^{k-1} \left(1 - \frac{\ell}{n}\right)\right) + \underbrace{\sum_{k=N+1}^n \frac{|x|^k}{k!} \prod_{\ell=0}^{k-1} \left(1 - \frac{\ell}{n}\right)}_{\leq 1} \\ &\leq \sum_{k=0}^N \frac{|x|^k}{k!} \left(1 - \prod_{\ell=0}^{k-1} \left(1 - \frac{\ell}{n}\right)\right) + \sum_{k=N+1}^{\infty} \frac{|x|^k}{k!}. \end{aligned}$$

Hence, we proved that, for every $n \geq N$,

$$\left| \sum_{k=0}^N \frac{x^k}{k!} - \left(1 + \frac{x}{n}\right)^n \right| \leq \sum_{k=0}^N \frac{|x|^k}{k!} \left(1 - \prod_{\ell=0}^{k-1} \left(1 - \frac{\ell}{n}\right)\right) + \sum_{k=N+1}^{\infty} \frac{|x|^k}{k!}.$$

Since for each fixed k ,

$$\lim_{n \rightarrow \infty} \left(1 - \prod_{\ell=0}^{k-1} \left(1 - \frac{\ell}{n}\right)\right) = 0,$$

letting $n \rightarrow \infty$ yields

$$\begin{aligned} \left| \sum_{k=0}^N \frac{x^k}{k!} - \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n \right| &= \lim_{n \rightarrow \infty} \left| \sum_{k=0}^N \frac{x^k}{k!} - \left(1 + \frac{x}{n}\right)^n \right| \\ &\leq \lim_{n \rightarrow \infty} \sum_{k=0}^N \frac{|x|^k}{k!} \left(1 - \prod_{\ell=0}^{k-1} \left(1 - \frac{\ell}{n}\right)\right) + \sum_{k=N+1}^{\infty} \frac{|x|^k}{k!} \\ &= \sum_{k=N+1}^{\infty} \frac{|x|^k}{k!}. \end{aligned}$$

By the triangle inequality and recalling (4.9), this gives

$$\begin{aligned} \left| \sum_{k=0}^{\infty} \frac{x^k}{k!} - \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n \right| &\leq \left| \sum_{k=0}^N \frac{x^k}{k!} - \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n \right| + \left| \sum_{k=N+1}^{\infty} \frac{x^k}{k!} \right| \\ &\leq \sum_{k=N+1}^{\infty} \frac{|x|^k}{k!} + \left| \sum_{k=N+1}^{\infty} \frac{x^k}{k!} \right| \leq 2 \sum_{k=N+1}^{\infty} \frac{|x|^k}{k!} < \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, the identity (4.8) follows. \square

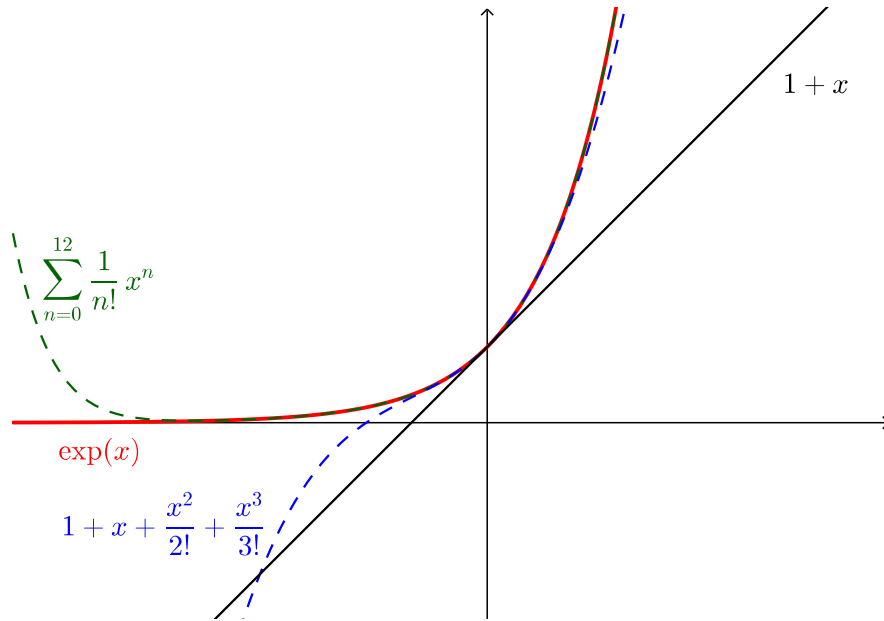


Figure 4.3: The exponential function and some of its partial sums.

21

DEFINITION 4.56: THE COMPLEX EXPONENTIAL MAP

The **complex exponential map** is the function $\exp : \mathbb{C} \rightarrow \mathbb{C}$ defined by

$$\exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}, \quad z \in \mathbb{C}.$$

For a positive real number $a > 0$ and $z \in \mathbb{C}$, we define

$$a^z := \exp(z \log a), \quad \text{in particular } e^z = \exp(z).$$

Before proving the main properties of the exponential function, recall the **binomial formula**: for all $z, w \in \mathbb{C}$ and $n \in \mathbb{N}$,

$$(z + w)^n = \sum_{k=0}^n \binom{n}{k} z^k w^{n-k}, \quad \binom{n}{k} = \frac{n!}{k!(n-k)!}. \quad (4.10)$$

THEOREM 4.57: PROPERTIES OF THE COMPLEX EXPONENTIAL

The complex exponential map $\exp : \mathbb{C} \rightarrow \mathbb{C}$ is continuous, and for all $z, w \in \mathbb{C}$,

$$e^{z+w} = e^z e^w, \quad |e^z| = e^{\operatorname{Re}(z)}. \quad (4.11)$$

In particular, $|e^{ix}| = 1$ for all $x \in \mathbb{R}$.

Proof. Since the series $\sum_{n=0}^{\infty} \frac{z^n}{n!}$ has infinite radius of convergence, Theorem 4.53 implies that $\exp : \mathbb{C} \rightarrow \mathbb{C}$ is continuous.

For $z, w \in \mathbb{C}$, using (4.5) and (4.10), we obtain

$$\begin{aligned} e^z e^w &= \left(\sum_{n=0}^{\infty} \frac{z^n}{n!} \right) \left(\sum_{n=0}^{\infty} \frac{w^n}{n!} \right) = \sum_{n=0}^{\infty} \sum_{k=0}^n \frac{z^k w^{n-k}}{k!(n-k)!} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^n \binom{n}{k} z^k w^{n-k} = \sum_{n=0}^{\infty} \frac{(z+w)^n}{n!} = e^{z+w}. \end{aligned}$$

To compute the modulus, note that complex conjugation is continuous and satisfies $\overline{z^k} = \overline{z}^k$ (see Lemma 2.42(3)), therefore

$$\overline{e^z} = \overline{\lim_{n \rightarrow \infty} \sum_{k=0}^n \frac{z^k}{k!}} = \lim_{n \rightarrow \infty} \overline{\sum_{k=0}^n \frac{z^k}{k!}} = \lim_{n \rightarrow \infty} \sum_{k=0}^n \frac{\overline{z}^k}{k!} = e^{\overline{z}}.$$

Since $|w|^2 = w\overline{w}$ and $w + \overline{w} = 2\operatorname{Re}(w)$ for all $w \in \mathbb{C}$, we have

$$|e^z|^2 = e^z \overline{e^z} = e^z e^{\overline{z}} = e^{z+\overline{z}} = e^{2\operatorname{Re}(z)},$$

hence $|e^z| = e^{\operatorname{Re}(z)}$. In particular, $|e^{ix}| = e^0 = 1$ for all $x \in \mathbb{R}$. \square

REMARK 4.58. — Note that the proof of Proposition 4.55 can be repeated verbatim to show that $e^z = \lim_{n \rightarrow \infty} \left(1 + \frac{z}{n}\right)^n$ for all $z \in \mathbb{C}$.

4.5.2 Sine and Cosine

Given $x \in \mathbb{R}$, we split the power series of e^{ix} into its even and odd terms:

$$e^{ix} = \sum_{n=0}^{\infty} \frac{i^n}{n!} x^n = \sum_{n=0}^{\infty} \frac{i^{2n}}{(2n)!} x^{2n} + \sum_{n=0}^{\infty} \frac{i^{2n+1}}{(2n+1)!} x^{2n+1}.$$

Since $i^{2n} = (-1)^n$ and $i^{2n+1} = i(-1)^n$, we obtain

$$e^{ix} = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} + i \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}.$$

This motivates the following definitions of the **sine** and **cosine** functions:

$$\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}, \quad \cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n}, \quad (4.12)$$

so that the identity

$$e^{ix} = \cos(x) + i \sin(x)$$

holds for all $x \in \mathbb{R}$.

As for the exponential series, the radius of convergence of the power series in (4.12) is infinite. Therefore, by Theorems 4.41 and 4.42, \sin and \cos are continuous on \mathbb{R} .

DEFINITION 4.59: EVEN AND ODD FUNCTIONS

Let $D \subset \mathbb{R}$ be a set satisfying $D = -D$, that is,

$$x \in D \implies -x \in D.$$

A function $f : D \rightarrow \mathbb{R}$ is called

- **even** if $f(-x) = f(x)$ for all $x \in D$,
- **odd** if $f(-x) = -f(x)$ for all $x \in D$.

Geometrically, an even function is symmetric with respect to the y -axis, while an odd function is symmetric with respect to the origin.

Since $(-x)^{2n+1} = -x^{2n+1}$ and $(-x)^{2n} = x^{2n}$ for all $n \in \mathbb{N}$, it follows directly from (4.12) that

$$\begin{aligned} \sin(-x) &= -\sin(x) \quad (\sin \text{ is an } \mathbf{odd} \text{ function}), \\ \cos(-x) &= \cos(x) \quad (\cos \text{ is an } \mathbf{even} \text{ function}). \end{aligned}$$

EXERCISE 4.60. — Prove that, for all $x \in \mathbb{R}$,

$$|\sin(x)| = |\sin(|x|)| \quad \text{and} \quad |\cos(x)| = |\cos(|x|)|.$$

THEOREM 4.61: FROM THE COMPLEX EXPONENTIAL TO SINE AND COSINE

For all $x \in \mathbb{R}$, the following relations hold:

$$e^{ix} = \cos(x) + i \sin(x), \quad \sin(x) = \frac{e^{ix} - e^{-ix}}{2i}, \quad \cos(x) = \frac{e^{ix} + e^{-ix}}{2}.$$

For all $x, y \in \mathbb{R}$, the trigonometric addition formulas are:

$$\begin{aligned} \sin(x + y) &= \sin(x) \cos(y) + \cos(x) \sin(y), \\ \cos(x + y) &= \cos(x) \cos(y) - \sin(x) \sin(y). \end{aligned} \tag{4.13}$$

Proof. For $x \in \mathbb{R}$ we have

$$e^{ix} = \cos(x) + i \sin(x), \quad e^{-ix} = \cos(-x) + i \sin(-x) = \cos(x) - i \sin(x).$$

Adding and subtracting these two identities gives the formulas for $\cos(x)$ and $\sin(x)$ in terms of e^{ix} and e^{-ix} .

To prove the addition formulas, we recall that

$$e^{i(x+y)} = e^{ix} e^{iy}$$

(see (4.11)), therefore

$$\begin{aligned} \cos(x + y) + i \sin(x + y) &= e^{i(x+y)} = e^{ix} e^{iy} \\ &= (\cos(x) + i \sin(x))(\cos(y) + i \sin(y)) \\ &= (\cos(x) \cos(y) - \sin(x) \sin(y)) + i(\sin(x) \cos(y) + \cos(x) \sin(y)). \end{aligned}$$

Comparing real and imaginary parts yields the identities in (4.13). \square

In particular, setting $y = x$ gives the **angle-doubling formulas**:

$$\sin(2x) = 2 \sin(x) \cos(x), \quad \cos(2x) = \cos^2(x) - \sin^2(x). \tag{4.14}$$

Recalling that $|e^{ix}| = 1$, we also obtain the **circle identity**:

$$\cos^2(x) + \sin^2(x) = 1 \quad \forall x \in \mathbb{R}.$$

4.5.3 The Circle Number

THEOREM 4.62: EXISTENCE OF π AS THE FIRST POSITIVE ZERO OF SINE

There exists exactly one number $\pi \in (0, 4)$ such that $\sin(\pi) = 0$. For this number it holds that

$$e^{i\frac{\pi}{2}} = i, \quad e^{i\pi} = -1, \quad e^{i2\pi} = 1.$$

Proof. For $x \in (0, 2]$, the sequence $(\frac{x^{2n+1}}{(2n+1)!})_{n=0}^{\infty}$ is monotonically decreasing. Hence, by the Leibniz criterion for alternating series (Proposition 4.22), the following estimates hold:

$$x - \frac{x^3}{3!} \leq \sin(x) \leq x - \frac{x^3}{3!} + \frac{x^5}{5!} \quad \forall x \in [0, 2]. \quad (4.15)$$

Analogously, the sequence $(\frac{x^{2n}}{(2n)!})_{n=0}^{\infty}$ is monotonically decreasing for $x \in [0, 1]$, therefore

$$1 - \frac{x^2}{2} \leq \cos(x) \leq 1 - \frac{x^2}{2} + \frac{x^4}{24} \quad \forall x \in [0, 1]. \quad (4.16)$$

Note that $\sin(0) = 0$. Also, thanks to (4.15), for $x = 1$ we have

$$\sin(1) \geq 1 - \frac{1}{6} > \frac{1}{\sqrt{2}}.$$

Thus, because \sin is continuous, the Intermediate Value Theorem (Theorem 3.24) implies the existence of $p \in (0, 1)$ such that $\sin(p) = \frac{1}{\sqrt{2}}$.

Because $\sin^2(p) + \cos^2(p) = 1$ and $\cos(x) \geq 1 - \frac{1}{2}x^2 > 0$ for $x \in [0, 1]$ (see (4.16)), this implies that

$$\cos(p) = \sqrt{1 - \sin^2(p)} = \frac{1}{\sqrt{2}}.$$

In other words,

$$e^{ip} = \cos(p) + i \sin(p) = \frac{1+i}{\sqrt{2}}.$$

Now, if we define $\pi = 4p \in (0, 4)$, we obtain

$$e^{i\frac{\pi}{2}} = e^{i2p} = (e^{ip})^2 = \frac{(1+i)^2}{2} = i, \quad e^{i\pi} = (e^{i\frac{\pi}{2}})^2 = i^2 = -1, \quad e^{i2\pi} = (-1)^2 = 1.$$

In particular, since $\cos(\pi) + i \sin(\pi) = e^{i\pi} = -1$, we deduce that

$$\sin(\pi) = 0, \quad \cos(\pi) = -1.$$

This proves that we have found a number $\pi \in (0, 4)$ that satisfies all the desired properties. It remains to prove uniqueness.

Assume there exists $s \in (0, 4)$, with $s \neq \pi$, satisfying $\sin(s) = 0$. From the estimate

$$\sin(x) \geq x - \frac{x^3}{3!} = x \left(1 - \frac{x^2}{6}\right) > 0 \quad \text{for } x \in (0, 2],$$

we see that $\sin(x)$ has no zeros in $(0, 2]$. Hence, we deduce that $\pi, s \in (2, 4)$.

Now, define $r = |\pi - s|$, so that $r \in (0, 2)$. Then, by Exercise 4.60 and the addition formula (4.13), we get

$$|\sin(r)| = |\sin(\pi - s)| = \left| \underbrace{\sin(\pi)}_{=0} \cos(s) - \cos(\pi) \underbrace{\sin(s)}_{=0} \right| = 0.$$

However, this is impossible since \sin has no zeros in $(0, 2)$. So, such a number s cannot exist, and $\pi \in (0, 4)$ is uniquely determined by the condition $\sin(\pi) = 0$. \square

COROLLARY 4.63: PERIODICITY OF SINE AND COSINE

$$\begin{aligned}\sin(x + \tfrac{\pi}{2}) &= \cos(x), & \cos(x + \tfrac{\pi}{2}) &= -\sin(x), \\ \sin(x + \pi) &= -\sin(x), & \cos(x + \pi) &= -\cos(x), \\ \sin(x + 2\pi) &= \sin(x), & \cos(x + 2\pi) &= \cos(x).\end{aligned}$$

Proof. From Theorem 4.62 we know that

$$\sin(\tfrac{\pi}{2}) = 1, \quad \cos(\tfrac{\pi}{2}) = 0, \quad \sin(\pi) = 0, \quad \cos(\pi) = -1, \quad \sin(2\pi) = 0, \quad \cos(2\pi) = 1.$$

Using these values together with the addition formulas (4.13), the identities follow directly.

Alternatively, one can observe that

$$e^{i(x+\frac{\pi}{2})} = e^{i\frac{\pi}{2}} e^{ix} = ie^{ix} \implies \cos(x + \tfrac{\pi}{2}) + i \sin(x + \tfrac{\pi}{2}) = i \cos(x) - \sin(x),$$

which implies that $\sin(x + \frac{\pi}{2}) = \cos(x)$ and $\cos(x + \frac{\pi}{2}) = -\sin(x)$. The other identities can be proved analogously. \square

From Corollary 4.63 it follows that both \sin and \cos are periodic functions with period 2π . To determine $\sin(x)$ or $\cos(x)$ for any real x , it suffices to know their values on the interval $[0, \frac{\pi}{2}]$.

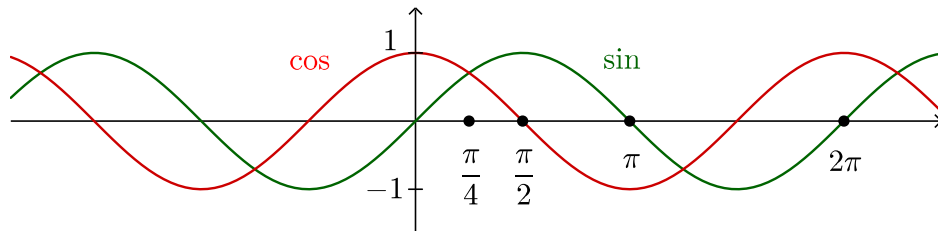


Figure 4.4: Graphs of the sine and cosine functions.

EXERCISE 4.64. — Show that the zeros of $\sin : \mathbb{R} \rightarrow \mathbb{R}$ are precisely the points in $\pi\mathbb{Z} \subset \mathbb{R}$, and the zeros of $\cos : \mathbb{R} \rightarrow \mathbb{R}$ are exactly the points in $\pi\mathbb{Z} + \frac{\pi}{2}$. Also, show that $\cos(x) = 1$ if and only if $x = 2n\pi$ with $n \in \mathbb{Z}$.

EXERCISE 4.65. — Show that

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

for all $x, y \in \mathbb{R}$. Use this to show that $\sin : [-\frac{\pi}{2}, \frac{\pi}{2}] \rightarrow [-1, 1]$ is strictly increasing and hence bijective.

4.5.4 Polar Coordinates and Multiplication of Complex Numbers

Using the complex exponential function, we can express complex numbers in **polar coordinates**, that is, in the form

$$z = re^{i\theta} = r \cos(\theta) + ir \sin(\theta),$$

where $r = |z|$ is the distance of z from the origin, and θ is the angle between the positive real axis $\mathbb{R}_{\geq 0}$ and the segment from 0 to z . In other words, if $z = x + iy$, then

$$x = r \cos(\theta), \quad y = r \sin(\theta), \quad r = \sqrt{x^2 + y^2}.$$

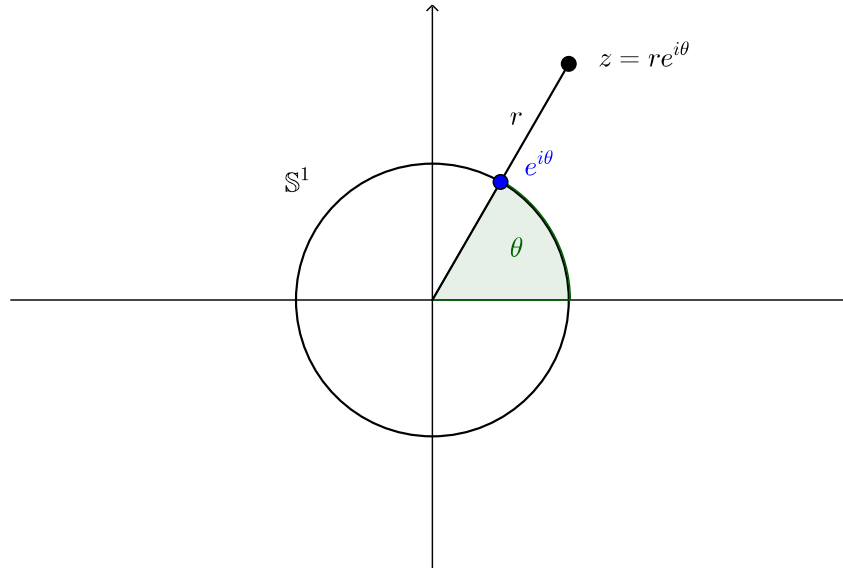


Figure 4.5: Polar representation of a complex number.

If $z \neq 0$, the angle θ is uniquely determined and is called the **argument** of z , denoted $\theta = \arg(z)$. The set of all complex numbers with absolute value one is

$$\mathbb{S}^1 = \{z \in \mathbb{C} \mid |z| = 1\} = \{e^{i\theta} \mid \theta \in [0, 2\pi)\},$$

and is called the **unit circle** in \mathbb{C} .

PROPOSITION 4.66: EXISTENCE OF POLAR COORDINATES

For every $z \in \mathbb{C} \setminus \{0\}$ there exist uniquely determined real numbers $r > 0$ and $\theta \in [0, 2\pi)$ such that $z = re^{i\theta}$.

Extra material: Proof of Proposition 4.66

Proof. Let $r = |z|$ and define $w = \frac{z}{r}$. Then $|w| = \frac{|z|}{r} = 1$. We must show that there exists a unique $\theta \in [0, 2\pi)$ such that $w = e^{i\theta}$. We first prove the existence of θ , distinguishing two cases according to the sign of $\text{Im}(w)$, and then establish uniqueness.

Existence when $\text{Im}(w) \geq 0$. Since $\text{Re}(w)^2 + \text{Im}(w)^2 = 1$, we have $\text{Re}(w) \in [-1, 1]$. Hence, since $\cos(0) = 1$ and $\cos(\pi) = -1$, the Intermediate Value Theorem 3.24 ensures the existence of $\theta \in [0, \pi]$ such that $\text{Re}(w) = \cos(\theta)$. Noticing that, for such θ , both $\text{Im}(w)$ and $\sin(\theta)$ are nonnegative, we get

$$\sin(\theta) = \sqrt{1 - \cos^2(\theta)} = \sqrt{1 - \text{Re}(w)^2} = \text{Im}(w),$$

therefore $w = e^{i\theta}$.

Existence when $\text{Im}(w) < 0$. Since $-w$ satisfies $\text{Im}(-w) > 0$, by the previous case there exists $\vartheta \in (0, \pi)$ such that $-w = e^{i\vartheta}$. Since $e^{i\pi} = -1$, we obtain

$$w = e^{i\pi} e^{i\vartheta} = e^{i(\pi+\vartheta)},$$

where $\theta = \pi + \vartheta \in (\pi, 2\pi)$.

Uniqueness. Let $\theta, \theta' \in [0, 2\pi)$ satisfy $e^{i\theta} = e^{i\theta'}$. Then $e^{i(\theta-\theta')} = 1$, that is,

$$\sin(\theta - \theta') = 0, \quad \cos(\theta - \theta') = 1.$$

Since $\theta - \theta' \in (-2\pi, 2\pi)$, Theorem 4.62 and Corollary 4.63 imply $\theta - \theta' = 0$. Hence $\theta = \theta'$, proving uniqueness. \square

In polar coordinates, multiplication of complex numbers takes a simple geometric form: if $z = re^{i\varphi}$ and $w = se^{i\psi}$, then

$$zw = rse^{i(\varphi+\psi)}.$$

Thus, when multiplying two complex numbers, their magnitudes multiply and their arguments add.

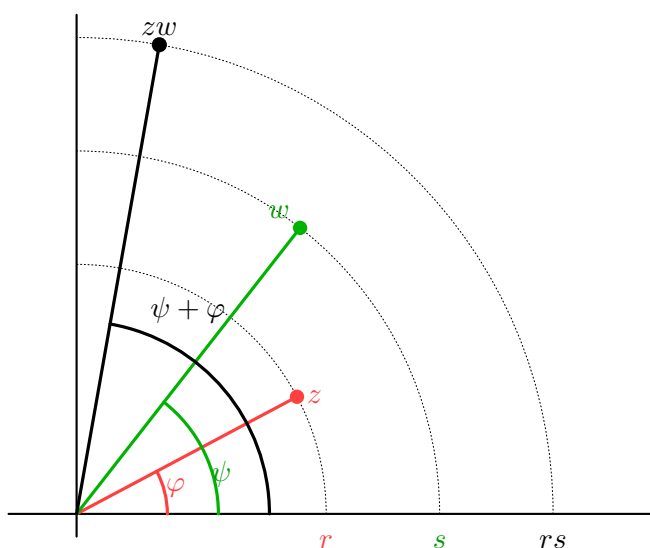


Figure 4.6: Multiplication of complex numbers in polar coordinates.

EXERCISE 4.67. — Let $w = re^{i\theta} \neq 0$. Show that the n -th roots of w (i.e. the solutions $z \in \mathbb{C}$ of $z^n = w$) are given by

$$\left\{ \sqrt[n]{r} e^{i\left(\frac{\theta+2\pi k}{n}\right)} \mid k = 0, 1, \dots, n-1 \right\}.$$

In particular, for $w = 1$ the n -th roots are

$$\{e^{i2\pi k/n} \mid k = 0, 1, \dots, n-1\},$$

and are called the n -th **roots of unity**.

EXERCISE 4.68. — For every integer $n \geq 2$, show that

$$\sum_{k=0}^{n-1} e^{i2\pi k/n} = 0.$$

21

4.5.5 The Complex Logarithm

We defined the real logarithm as the inverse of the bijective mapping $\exp : \mathbb{R} \rightarrow \mathbb{R}_{>0}$. For complex numbers, however, the exponential map $\exp : \mathbb{C} \rightarrow \mathbb{C}$ is *not* injective, since $\exp(ix) = 1$ for all $x = 2\pi n$ with $n \in \mathbb{Z}$. To define a complex logarithm, one must therefore restrict the exponential function to a suitable subset $D \subset \mathbb{C}$ such that the restricted map $\exp|_D : D \rightarrow \mathbb{C}^\times$ becomes bijective. There are many possible choices for such a domain D , but a detailed discussion of this topic lies beyond the scope of this course.

4.5.6 Other Trigonometric and Hyperbolic Functions

In addition to the exponential, sine, and cosine functions, several related **trigonometric functions** are defined.

The **tangent** and **cotangent** functions are defined by

$$\tan(x) = \frac{\sin(x)}{\cos(x)}, \quad \cot(x) = \frac{\cos(x)}{\sin(x)},$$

for all $x \in \mathbb{R}$ such that the denominators are nonzero.

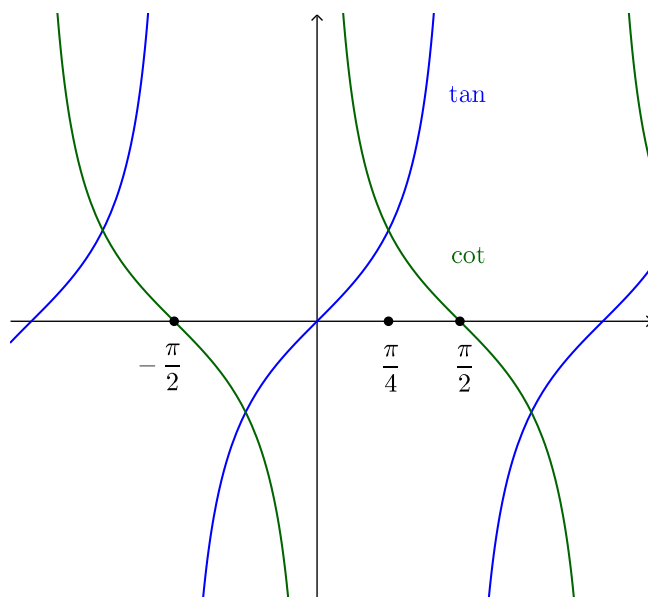


Figure 4.7: Graphs of the tangent and cotangent functions.

EXERCISE 4.69. — Show that, for $x, y \in \mathbb{R}$ where both sides are defined,

$$\tan(x + y) = \frac{\tan(x) + \tan(y)}{1 - \tan(x)\tan(y)}.$$

21

Find and prove an analogous addition formula for the cotangent function.

The **hyperbolic sine** and **hyperbolic cosine** are defined by the power series

$$\sinh(x) = \sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+1)!}, \quad \cosh(x) = \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!}.$$

Equivalently,

$$\sinh(x) = \frac{e^x - e^{-x}}{2}, \quad \cosh(x) = \frac{e^x + e^{-x}}{2},$$

and hence $e^x = \cosh(x) + \sinh(x)$ for all $x \in \mathbb{R}$.

The **hyperbolic tangent** and **hyperbolic cotangent** are defined by

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \quad \coth(x) = \frac{\cosh(x)}{\sinh(x)} = \frac{e^x + e^{-x}}{e^x - e^{-x}},$$

where $\coth(x)$ is defined for all $x \in \mathbb{R} \setminus \{0\}$ (since $\sinh(x) \neq 0$ for $x \neq 0$).

The functions \sinh and \tanh are odd, while \cosh is even. Also, they satisfy the addition formulas

$$\begin{aligned} \sinh(x + y) &= \sinh(x) \cosh(y) + \cosh(x) \sinh(y), \\ \cosh(x + y) &= \cosh(x) \cosh(y) + \sinh(x) \sinh(y), \end{aligned}$$

and the **hyperbolic identity**

$$\cosh^2(x) - \sinh^2(x) = 1 \quad \forall x \in \mathbb{R}.$$

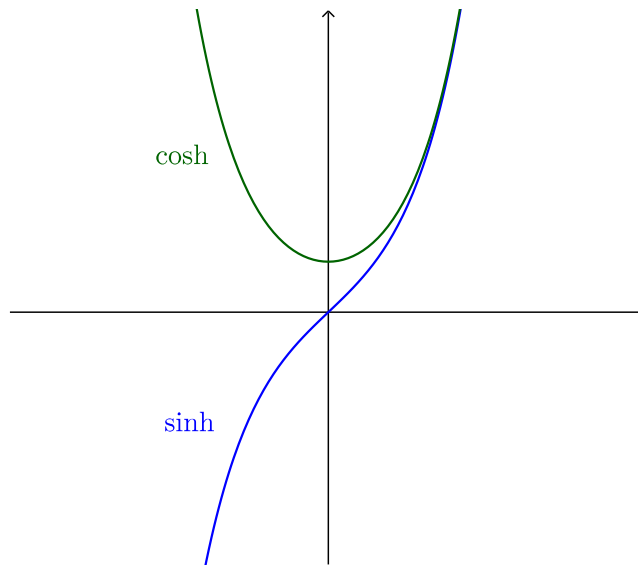


Figure 4.8: Graphs of the hyperbolic sine and cosine.

EXERCISE 4.70. — Starting from the definitions of \sinh and \cosh , prove the above identities.

Chapter 5

Differential Calculus

In this chapter we deal with differential calculus in one variable. This is of fundamental importance for understanding functions on \mathbb{R} .

5.1 The Derivative

5.1.1 Definition and Geometrical Interpretation

In this section, $D \subseteq \mathbb{R}$ denotes a nonempty set with no isolated points; that is, every $x \in D$ is an accumulation point of $D \setminus \{x\}$. A typical example is a nonempty interval containing more than one point.

DEFINITION 5.1: DERIVATIVE

Let $f : D \rightarrow \mathbb{R}$ be a function and $x_0 \in D$. We say that f is **differentiable** at x_0 if the limit

$$f'(x_0) = \lim_{\substack{x \rightarrow x_0 \\ x \neq x_0}} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{\substack{h \rightarrow 0 \\ h \neq 0}} \frac{f(x_0 + h) - f(x_0)}{h} \quad (5.1)$$

exists. In this case we call $f'(x_0)$ the **derivative** of f at x_0 . If f is differentiable at every point of D , then we also say that f is **differentiable** on D , and we call the resulting function $f' : D \rightarrow \mathbb{R}$ the **derivative** of f .

To simplify notation, we will often write

$$f'(x_0) = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h},$$

without explicitly mentioning that $x \neq x_0$ or $h \neq 0$. Note that the condition

$$f'(x_0) = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

can be rewritten as

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{x - x_0} = 0,$$

or equivalently, using the little- o notation from Definition 3.69,

$$f(x) - f(x_0) - f'(x_0)(x - x_0) = o(x - x_0). \quad (5.2)$$

REMARK 5.2. — If $f : D \rightarrow \mathbb{R}$ is differentiable at x_0 , then it is also continuous at x_0 . Indeed, using (5.2),

$$\lim_{x \rightarrow x_0} f(x) = \lim_{x \rightarrow x_0} (f(x_0) + f'(x_0)(x - x_0) + o(x - x_0)) = f(x_0),$$

hence f is continuous at x_0 .

An alternative notation for the derivative of f is $\frac{df}{dx}$. If $x_0 \in D$ is a right accumulation point of D , then f is **differentiable from the right** at x_0 if the **right derivative**

$$f'_+(x_0) = \lim_{x \rightarrow x_0^+} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{h \rightarrow 0^+} \frac{f(x_0 + h) - f(x_0)}{h}$$

exists. **Differentiability from the left** and the **left derivative** $f'_-(x_0)$ are defined analogously using the limit $x \rightarrow x_0^-$.

AFFINE FUNCTIONS

An **affine function** is a function of the form $x \mapsto sx + r$, for real numbers s and r . The graph of an affine function is a nonvertical **line** in \mathbb{R}^2 . The parameter s in the equation $y = sx + r$ is called the **slope** of the line.

If $f : D \rightarrow \mathbb{R}$ is differentiable at $x_0 \in D$, the function $x \mapsto f(x_0) + f'(x_0)(x - x_0)$ is called the **affine approximation** of f at x_0 .

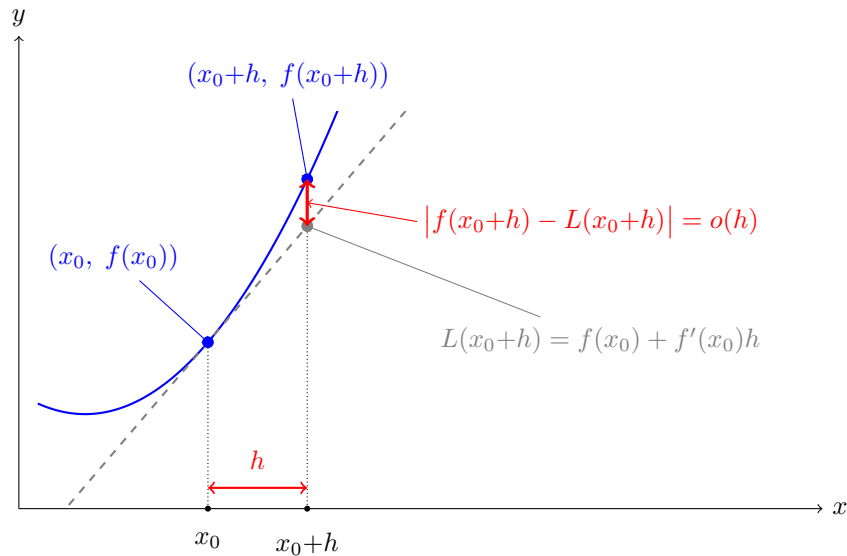


Figure 5.1: The figure shows the graph of f (blue) and its tangent at x_0 (gray dashed), which corresponds to the graph of the affine approximation $L(x) = f(x_0) + f'(x_0)(x - x_0)$. Note that, when we move a distance h from x_0 to $x_0 + h$, the vertical error between $f(x_0 + h)$ and $L(x_0 + h)$ is $o(h)$.

22

EXAMPLE 5.3. — • Constant functions are differentiable everywhere and have the zero function as their derivative.

- The identity function $f(x) = x$ is differentiable, and its derivative is the constant function 1. Indeed,

$$f'(x_0) = \lim_{x \rightarrow x_0} \frac{x - x_0}{x - x_0} = 1 \quad \forall x_0 \in \mathbb{R}.$$

EXAMPLE 5.4. — The exponential function $\exp : \mathbb{R} \rightarrow \mathbb{R}_{>0}$ is differentiable and its derivative is again the exponential function. Indeed, for $x \in \mathbb{R}$, using that $e^{x+h} = e^x e^h$ we get

$$\frac{e^{x+h} - e^x}{h} = e^x \frac{e^h - 1}{h} = e^x \frac{\sum_{k=0}^{\infty} \frac{h^k}{k!} - 1}{h} = e^x \frac{\sum_{k=1}^{\infty} \frac{h^k}{k!}}{h} = e^x \sum_{k=1}^{\infty} \frac{h^{k-1}}{k!} = e^x \sum_{n=0}^{\infty} \frac{h^n}{(n+1)!}.$$

We now observe that the power series $x \mapsto \sum_{n=0}^{\infty} \frac{x^n}{(n+1)!}$ has infinite radius of convergence (see, e.g., Exercise 4.45); in particular, the function $g(x) = \sum_{n=0}^{\infty} \frac{x^n}{(n+1)!}$ is continuous on \mathbb{R} and $g(0) = \frac{1}{1!} = 1$. Hence

$$(e^x)' = \lim_{h \rightarrow 0} \frac{e^{x+h} - e^x}{h} = e^x \lim_{h \rightarrow 0} g(h) = e^x g(0) = e^x.$$

More generally, let $\alpha \in \mathbb{C}$ and define $f : \mathbb{R} \rightarrow \mathbb{C}$ by $f(x) = e^{\alpha x}$. Then, arguing as before, we get

$$\begin{aligned} (e^{\alpha x})' &= \lim_{h \rightarrow 0} \frac{e^{\alpha x + \alpha h} - e^{\alpha x}}{h} = e^{\alpha x} \lim_{h \rightarrow 0} \frac{e^{\alpha h} - 1}{h} = e^{\alpha x} \lim_{h \rightarrow 0} \sum_{k=1}^{\infty} \frac{\alpha^k h^{k-1}}{k!} \\ &= \alpha e^{\alpha x} \lim_{h \rightarrow 0} \sum_{k=1}^{\infty} \frac{(\alpha h)^{k-1}}{k!} = \alpha e^{\alpha x} \lim_{h \rightarrow 0} \sum_{n=0}^{\infty} \frac{(\alpha h)^n}{(n+1)!} = \alpha e^{\alpha x}. \end{aligned}$$

EXAMPLE 5.5. — Let $f : \mathbb{R} \setminus \{0\} \rightarrow \mathbb{R}$ be given by $f(x) = \frac{1}{x}$. Then f is differentiable and $f'(x) = -\frac{1}{x^2}$ for all $x \in \mathbb{R} \setminus \{0\}$:

$$f'(x) = \lim_{h \rightarrow 0} \frac{\frac{1}{x+h} - \frac{1}{x}}{h} = \lim_{h \rightarrow 0} \frac{x - (x+h)}{(x+h)xh} = -\lim_{h \rightarrow 0} \frac{1}{(x+h)x} = -\frac{1}{x^2}.$$

DEFINITION 5.6: HIGHER DERIVATIVES

Let $f : D \rightarrow \mathbb{R}$ be a function. We define the **higher derivatives** of f , if they exist, by

$$f^{(0)} = f, \quad f^{(1)} = f', \quad f^{(2)} = f'', \quad \dots, \quad f^{(n+1)} = (f^{(n)})'$$

for all $n \in \mathbb{N}$. If $f^{(n)}$ exists, we say that f is **n -times differentiable**. If the n th derivative $f^{(n)}$ is also continuous, f is called **n -times continuously differentiable**. We denote the set of n -times continuously differentiable functions on D by $C^n(D)$.

Equivalently, $C^0(D)$ is the set of real-valued continuous functions on D , and $C^1(D)$ is the set of all differentiable functions whose derivative is continuous (these are called **continuously differentiable** or of **class C^1**). Recursively, for $n \geq 1$,

$$C^n(D) = \{f : D \rightarrow \mathbb{R} \mid f \text{ is differentiable and } f' \in C^{n-1}(D)\},$$

and we say that $f \in C^n(D)$ is of **class C^n** .

EXAMPLE 5.7. — Consider the function $f : \mathbb{R} \rightarrow \mathbb{R}$ defined by $f(x) = \operatorname{sgn}(x)x^2$, or equivalently

$$f(x) = \begin{cases} x^2, & x \geq 0, \\ -x^2, & x < 0. \end{cases}$$

For $x > 0$ and $x < 0$, we can compute the derivative directly:

$$f'(x) = \begin{cases} 2x, & x > 0, \\ -2x, & x < 0. \end{cases}$$

For $x = 0$, we have

$$\frac{f(x) - f(0)}{x - 0} = \frac{\operatorname{sgn}(x) x^2}{x} = \operatorname{sgn}(x) x \longrightarrow 0 \quad \text{as } x \rightarrow 0,$$

so $f'(0) = 0$. Altogether,

$$f'(x) = \begin{cases} 2x, & x > 0, \\ 0, & x = 0, \\ -2x, & x < 0, \end{cases} \quad \text{that is, } f'(x) = 2|x|.$$

Hence f is continuously differentiable, i.e. $f \in C^1(\mathbb{R})$. However, since $f'(x) = 2|x|$ is not differentiable at $x = 0$, the function f is not of class C^2 .

DEFINITION 5.8: SMOOTH FUNCTIONS

We define

$$C^\infty(D) = \bigcap_{n=0}^{\infty} C^n(D) = \{ f : D \rightarrow \mathbb{R} \mid f \text{ is differentiable infinitely many times} \},$$

and call functions $f \in C^\infty(D)$ **smooth** or of **class** C^∞ .

22

EXAMPLE 5.9. — The exponential function $\exp : \mathbb{R} \rightarrow \mathbb{R}$ is smooth.

5.1.2 Differentiation Rules

As with continuous functions, we rarely reprove differentiability from first principles for each new example; instead, we use general rules that reduce differentiability of compound expressions to that of simpler ones.

PROPOSITION 5.10: DERIVATIVE OF SUM AND PRODUCT

Let $D \subseteq \mathbb{R}$ and let $x_0 \in D$ be an accumulation point of $D \setminus \{x_0\}$. Let $f, g : D \rightarrow \mathbb{R}$ be differentiable at x_0 . Then $f + g$ and $f \cdot g$ are differentiable at x_0 , and

$$(f + g)'(x_0) = f'(x_0) + g'(x_0), \tag{5.3}$$

$$(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0). \tag{5.4}$$

In particular, for any $\alpha \in \mathbb{R}$, the scalar multiple αf is differentiable at x_0 and $(\alpha f)'(x_0) = \alpha f'(x_0)$.

Proof. Using the properties of limit discussed in Section 3.5.1, we have

$$\begin{aligned}\lim_{x \rightarrow x_0} \frac{(f+g)(x) - (f+g)(x_0)}{x - x_0} &= \lim_{x \rightarrow x_0} \left(\frac{f(x) - f(x_0)}{x - x_0} + \frac{g(x) - g(x_0)}{x - x_0} \right) \\ &= \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} + \lim_{x \rightarrow x_0} \frac{g(x) - g(x_0)}{x - x_0} \\ &= f'(x_0) + g'(x_0),\end{aligned}$$

and

$$\begin{aligned}\lim_{x \rightarrow x_0} \frac{(fg)(x) - (fg)(x_0)}{x - x_0} &= \lim_{x \rightarrow x_0} \frac{(f(x) - f(x_0))g(x) + f(x_0)(g(x) - g(x_0))}{x - x_0} \\ &= \lim_{x \rightarrow x_0} \left(\frac{f(x) - f(x_0)}{x - x_0} g(x) \right) + f(x_0) \lim_{x \rightarrow x_0} \frac{g(x) - g(x_0)}{x - x_0} \\ &= \left(\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} \right) \cdot \left(\lim_{x \rightarrow x_0} g(x) \right) + f(x_0) \lim_{x \rightarrow x_0} \frac{g(x) - g(x_0)}{x - x_0} \\ &= f'(x_0)g(x_0) + f(x_0)g'(x_0),\end{aligned}$$

where we used that g is continuous at x_0 (see Remark 5.2) to conclude $\lim_{x \rightarrow x_0} g(x) = g(x_0)$. \square

COROLLARY 5.11: HIGHER ORDER DERIVATIVES OF THE PRODUCT

Let $f, g : D \rightarrow \mathbb{R}$ be n -times differentiable. Then $f + g$ and $f \cdot g$ are also n -times differentiable, with $(f + g)^{(n)} = f^{(n)} + g^{(n)}$ and

$$(fg)^{(n)} = \sum_{k=0}^n \binom{n}{k} f^{(k)} g^{(n-k)}.$$

In particular, for every $\alpha \in \mathbb{R}$, $(\alpha f)^{(n)} = \alpha f^{(n)}$.

Proof. For $n = 1$ this is Proposition 5.10. The general case follows by induction on $n \geq 1$. \square

COROLLARY 5.12: DERIVATIVES OF POLYNOMIALS

Polynomial functions are differentiable on all of \mathbb{R} . Moreover, $(1)' = 0$ and $(x^n)' = nx^{n-1}$ for all $n \geq 1$.

Proof. We argue by induction. The cases $n = 0$ and $n = 1$ were covered in Example 5.3. For $n > 1$, assume $(x^n)' = nx^{n-1}$. Then, by (5.4), since $x^{n+1} = x \cdot x^n$,

$$(x^{n+1})' = (x \cdot x^n)' = 1 \cdot x^n + x \cdot (nx^{n-1}) = (n+1)x^n.$$

This proves the inductive step and establishes the result. Finally, the linearity of the derivative (see (5.3)) yields the differentiability of any polynomial. \square

EXAMPLE 5.13. — From Example 5.4 with $\alpha = \pm 1$ and $\alpha = \pm i$,

$$(e^x)' = e^x, \quad (e^{-x})' = -e^{-x}, \quad (e^{ix})' = ie^{ix}, \quad (e^{-ix})' = -ie^{-ix}.$$

By Theorem 4.61,

$$\sin'(x) = \frac{(e^{ix})' - (e^{-ix})'}{2i} = \frac{e^{ix} + e^{-ix}}{2} = \cos(x),$$

and analogously $\cos'(x) = -\sin(x)$. Similarly, $\sinh'(x) = \cosh(x)$ and $\cosh'(x) = \sinh(x)$.

THEOREM 5.14: CHAIN RULE

Let $D, E \subseteq \mathbb{R}$, and let $x_0 \in D$ be an accumulation point of $D \setminus \{x_0\}$. Let $f : D \rightarrow E$ be differentiable at x_0 such that $y_0 = f(x_0)$ is an accumulation point of $E \setminus \{y_0\}$, and let $g : E \rightarrow \mathbb{R}$ be differentiable at y_0 . Then $g \circ f : D \rightarrow \mathbb{R}$ is differentiable at x_0 and

$$(g \circ f)'(x_0) = g'(f(x_0)) f'(x_0).$$

REMARK 5.15. — Heuristically, we would like to compute the derivative of the composition by writing

$$\frac{g(f(x)) - g(f(x_0))}{x - x_0} = \frac{g(f(x)) - g(f(x_0))}{f(x) - f(x_0)} \cdot \frac{f(x) - f(x_0)}{x - x_0},$$

and then take the limit as $x \rightarrow x_0$ to get

$$(g \circ f)'(x_0) = g'(f(x_0)) f'(x_0).$$

This argument proves the result if $f(x) \neq f(x_0)$ for x near x_0 , but it is not a proof as stated: if $f(x) = f(x_0)$ for infinitely many x close to x_0 , the fraction $\frac{g(f(x)) - g(f(x_0))}{f(x) - f(x_0)}$ is undefined, and one faces a $0/0$ indeterminacy. In the proof below, we will provide a rigorous argument, but it is helpful to keep the above heuristic in mind as the underlying intuition for the chain rule.

Proof. Observe that we can write

$$\begin{aligned} g(y) &= g(y_0) + [g(y) - g(y_0)] \\ &= g(y_0) + g'(y_0)(y - y_0) + [g(y) - g(y_0) - g'(y_0)(y - y_0)] \\ &= g(y_0) + g'(y_0)(y - y_0) + \omega(y)(y - y_0), \end{aligned}$$

where the function $\omega : E \rightarrow \mathbb{R}$ is defined as

$$\omega(y) = \begin{cases} \frac{g(y) - g(y_0)}{y - y_0} - g'(y_0) & \text{for } y \in E \setminus \{y_0\}, \\ 0 & \text{for } y = y_0. \end{cases}$$

Since g is differentiable at y_0 , it follows that $\omega(y) \rightarrow 0$ as $y \rightarrow y_0$; hence, the function ω is continuous at y_0 . Substituting $y = f(x)$ and using $y_0 = f(x_0)$,

$$g(f(x)) = g(f(x_0)) + g'(f(x_0)) [f(x) - f(x_0)] + \omega(f(x)) [f(x) - f(x_0)],$$

therefore

$$\begin{aligned} \lim_{x \rightarrow x_0} \frac{g(f(x)) - g(f(x_0))}{x - x_0} &= \lim_{x \rightarrow x_0} \left(g'(f(x_0)) \frac{f(x) - f(x_0)}{x - x_0} + \omega(f(x)) \frac{f(x) - f(x_0)}{x - x_0} \right) \\ &= g'(f(x_0)) f'(x_0) + \underbrace{\omega(f(x_0))}_{=0} f'(x_0) = g'(f(x_0)) f'(x_0), \end{aligned}$$

where we used the continuity of ω at $y_0 = f(x_0)$ to deduce that $\omega(f(x)) \rightarrow \omega(f(x_0)) = 0$ as $x \rightarrow x_0$. \square

REMARK 5.16. — By a nontrivial induction argument, if $f : D \rightarrow E$ and $g : E \rightarrow \mathbb{R}$ are n -times differentiable, then $g \circ f : D \rightarrow \mathbb{R}$ is n -times differentiable, and one can express the n th derivative of $g \circ f$ in terms of sums and products of $g^{(k)} \circ f$ and $f^{(j)}$ with $1 \leq k, j \leq n$. This is known as Faà di Bruno's formula, but we will not explore it here.

COROLLARY 5.17: QUOTIENT RULE

Let $D \subseteq \mathbb{R}$, let $x_0 \in D$ be an accumulation point of $D \setminus \{x_0\}$, and let $f, g : D \rightarrow \mathbb{R}$ be differentiable at x_0 . If $g(x_0) \neq 0$, then $\frac{f}{g}$ is differentiable at x_0 and

$$\left(\frac{f}{g} \right)'(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{g(x_0)^2}.$$

Proof. Consider the function $\psi : \mathbb{R} \setminus \{0\} \rightarrow \mathbb{R}$ given by $\psi(y) = \frac{1}{y}$. This function is differentiable by Example 5.5, with $\psi'(y) = -\frac{1}{y^2}$. Then, by the chain rule (Theorem 5.14), $\frac{1}{g} = \psi \circ g$ is differentiable at x_0 , with

$$\left(\frac{1}{g} \right)'(x_0) = \psi'(g(x_0))g'(x_0) = -\frac{g'(x_0)}{g(x_0)^2}.$$

Applying now the product rule (Proposition 5.10), $\frac{f}{g} = f \cdot \frac{1}{g}$ is differentiable at x_0 , and

$$\left(\frac{f}{g} \right)'(x_0) = \left(f \cdot \frac{1}{g} \right)'(x_0) = f'(x_0) \frac{1}{g(x_0)} - f(x_0) \frac{g'(x_0)}{g(x_0)^2} = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{g(x_0)^2}.$$

\square

EXAMPLE 5.18. — We want to compute the derivative of the function $f : \mathbb{R} \rightarrow \mathbb{R}$ given by

$f(x) = \exp(\sin(\sin(x^2)))$. For this, let $g(x) = \sin(\sin(x^2))$, $h(x) = \sin(x^2)$, and $k(x) = x^2$. Then, since

$$f(x) = \exp(g(x)), \quad g(x) = \sin(h(x)), \quad h(x) = \sin(k(x)),$$

applying the chain rule repeatedly and using that $\exp' = \exp$ and $\sin' = \cos$, we get

$$f'(x) = \exp(g(x))g'(x), \quad g'(x) = \cos(h(x))h'(x), \quad h'(x) = \cos(k(x))k'(x), \quad k'(x) = 2x,$$

therefore

$$f'(x) = \exp(\sin(\sin(x^2))) \cos(\sin(x^2)) \cos(x^2) 2x \quad \forall x \in \mathbb{R}.$$

EXERCISE 5.19. — Determine the derivative of the function $x \mapsto \cos(\sin^3(\exp(x)))$.

In the next theorem, given $f : D \rightarrow E$ a continuous bijection, we investigate the derivative of the inverse of a function. Notice that, for the derivative of f^{-1} to be well-defined at a point $\bar{y} = f(\bar{x})$, we must ensure that if $\bar{x} \in D$ is an accumulation point of $D \setminus \{\bar{x}\}$, then $f(\bar{x})$ is an accumulation point of $E \setminus \{f(\bar{x})\}$. This follows from continuity and injectivity: if $(x_n)_{n=0}^\infty \subset D \setminus \{\bar{x}\}$ satisfies $x_n \rightarrow \bar{x}$, then $(f(x_n))_{n=0}^\infty \subset E \setminus \{f(\bar{x})\}$ and $f(x_n) \rightarrow f(\bar{x})$.

THEOREM 5.20: DERIVATIVE OF THE INVERSE

Let $D, E \subseteq \mathbb{R}$, and let $f : D \rightarrow E$ be a continuous bijection whose inverse $f^{-1} : E \rightarrow D$ is also continuous. Let $\bar{x} \in D$ be an accumulation point of $D \setminus \{\bar{x}\}$, and assume that f is differentiable at \bar{x} with $f'(\bar{x}) \neq 0$. Then f^{-1} is differentiable at $\bar{y} = f(\bar{x})$ and

$$(f^{-1})'(\bar{y}) = \frac{1}{f'(\bar{x})} = \frac{1}{f'(f^{-1}(\bar{y}))}.$$

Proof. To compute $(f^{-1})'(\bar{y})$, take a sequence $(y_n)_{n=0}^\infty \subset E \setminus \{\bar{y}\}$ with $y_n \rightarrow \bar{y}$ and set $x_n = f^{-1}(y_n)$. Then

$$\frac{f^{-1}(y_n) - f^{-1}(\bar{y})}{y_n - \bar{y}} = \frac{x_n - \bar{x}}{f(x_n) - f(\bar{x})} = \left(\frac{f(x_n) - f(\bar{x})}{x_n - \bar{x}} \right)^{-1}.$$

By the continuity of f^{-1} , we have $x_n \rightarrow \bar{x} = f^{-1}(\bar{y})$. Thus, since f is differentiable at \bar{x} with $f'(\bar{x}) \neq 0$, Proposition 2.94(4) implies

$$\left(\frac{f(x_n) - f(\bar{x})}{x_n - \bar{x}} \right)^{-1} \longrightarrow \frac{1}{f'(\bar{x})},$$

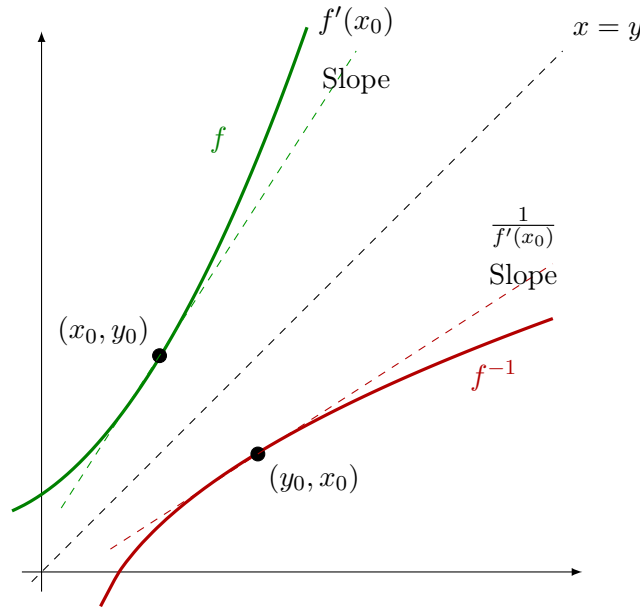
proving that

$$\lim_{n \rightarrow \infty} \frac{f^{-1}(y_n) - f^{-1}(\bar{y})}{y_n - \bar{y}} = \frac{1}{f'(\bar{x})}.$$

Since the sequence $(y_n)_{n=0}^\infty$ was arbitrary, Lemma 3.57 gives

$$\lim_{y \rightarrow \bar{y}} \frac{f^{-1}(y) - f^{-1}(\bar{y})}{y - \bar{y}} = \frac{1}{f'(\bar{x})},$$

as desired. □



23

Figure 5.2: An intuitive representation of Theorem 5.20. Let $y_0 = f(x_0)$. Reflecting the graph of f and the tangent line at (x_0, y_0) across the line $x = y$ in \mathbb{R}^2 yields the graph of f^{-1} and, as asserted, the tangent line at (y_0, x_0) . A short calculation shows that the reflection of a line with slope m across $x = y$ has slope $1/m$.

EXAMPLE 5.21. — The function $g : \mathbb{R} \setminus \{0\} \rightarrow \mathbb{R}$ defined by $g(y) = \log |y|$ is differentiable, with $g'(y) = \frac{1}{y}$ for all $y \in \mathbb{R} \setminus \{0\}$. Indeed, let $f = \exp : \mathbb{R} \rightarrow \mathbb{R}_{>0}$, so that $\log = f^{-1}$. By Theorem 5.20,

$$\log'(y) = \frac{1}{f'(\log(y))}.$$

Since $\exp' = \exp$, for $y > 0$ we have

$$\log'(y) = \frac{1}{\exp(\log(y))} = \frac{1}{y}.$$

For $y < 0$, since $g(y) = \log(-y)$, by chain rule (Theorem 5.14) and applying the case above to $-y > 0$, we get

$$g'(y) = -\log'(-y) = -\frac{1}{-y} = \frac{1}{y}.$$

EXAMPLE 5.22. — Given $x > 0$ and $\alpha \in \mathbb{R}$, we can compute the derivative of x^α via

$$x^\alpha = \exp(\alpha \log x) \implies (x^\alpha)' = \exp'(\alpha \log x) \alpha \log'(x) = \exp(\alpha \log x) \frac{\alpha}{x} = \alpha x^{\alpha-1}.$$

This generalizes Corollary 5.12.

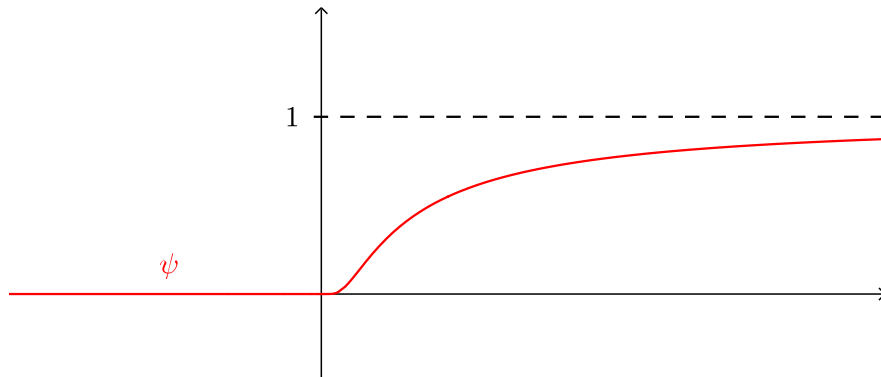
EXAMPLE 5.23. — The logarithm $f = \log : \mathbb{R}_{>0} \rightarrow \mathbb{R}$ is smooth. Indeed, by the example above, $f'(x) = \frac{1}{x}$. By induction (using the Leibniz rule) one gets

$$f''(x) = -\frac{1}{x^2}, \quad f^{(3)}(x) = \frac{2}{x^3}, \quad \text{and in general} \quad f^{(n)}(x) = (-1)^{n-1} \frac{(n-1)!}{x^n} \quad \forall n \geq 1.$$

EXERCISE 5.24. — Consider the function $\psi : \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$\psi(x) = \begin{cases} \exp\left(-\frac{1}{x}\right), & x > 0, \\ 0, & x \leq 0. \end{cases}$$

Show that ψ is smooth on \mathbb{R} , and that all its derivatives at 0 vanish.



Hint: Show first by induction that, for all $n \in \mathbb{N}$,

$$\psi^{(n)}(x) = \exp\left(-\frac{1}{x}\right) f_n\left(\frac{1}{x}\right) \quad (x > 0), \quad (5.5)$$

where f_n is a polynomial. Then, using that $\exp(-1/x)$ tends to 0 faster than any power of x as $x \rightarrow 0^+$ (equivalently, $\exp(y)$ dominates every polynomial as $y \rightarrow +\infty$, by Corollary 3.47), prove that

$$\psi^{(n+1)}(0) = \lim_{x \rightarrow 0} \frac{\psi^{(n)}(x) - \psi^{(n)}(0)}{x} = \lim_{x \rightarrow 0^+} \frac{\psi(x) f_n(\frac{1}{x})}{x} = \lim_{x \rightarrow 0^+} \psi(x) f_n(\frac{1}{x}) \frac{1}{x} = 0 \quad \forall n \in \mathbb{N}.$$

5.2 Main Theorems of Differential Calculus

5.2.1 Local Extrema

DEFINITION 5.25: LOCAL EXTREMA

Let $D \subseteq \mathbb{R}$ and $x_0 \in D$. We say that a function $f : D \rightarrow \mathbb{R}$ has a **local maximum** at x_0 if there exists $\delta > 0$ such that

$$f(x) \leq f(x_0) \quad \forall x \in D \cap (x_0 - \delta, x_0 + \delta).$$

If the inequality is strict (i.e. $f(x) < f(x_0)$ for all $x \in D \cap (x_0 - \delta, x_0 + \delta) \setminus \{x_0\}$), then f has a **strict local maximum** at x_0 .

A **(strict) local minimum** is defined analogously.

We call x_0 a **local extremum** of f if f has either a local minimum or a local maximum at x_0 .

PROPOSITION 5.26: LOCAL EXTREMA VS. FIRST DERIVATIVE

Let $D \subseteq \mathbb{R}$ and $f : D \rightarrow \mathbb{R}$. Suppose $x_0 \in D$ is a local extremum of f , that f is differentiable at x_0 , and that x_0 is both a right-hand and a left-hand accumulation point of D . Then

$$f'(x_0) = 0.$$

Proof. Without loss of generality, assume f has a local maximum at x_0 (otherwise replace f by $-f$). We first note that, for x close to x_0 and to the right of it, we have $f(x) - f(x_0) \leq 0$ and $x - x_0 > 0$. Hence,

$$f'_+(x_0) = \lim_{x \rightarrow x_0^+} \frac{f(x) - f(x_0)}{x - x_0} \leq 0.$$

Similarly, for x close to x_0 and to the left of it, we have $f(x) - f(x_0) \leq 0$ and $x - x_0 < 0$, so

$$f'_-(x_0) = \lim_{x \rightarrow x_0^-} \frac{f(x) - f(x_0)}{x - x_0} \geq 0,$$

Since f is differentiable at x_0 , the two one-sided derivatives coincide, that is,

$$f'(x_0) = f'_+(x_0) = f'_-(x_0),$$

and therefore $f'(x_0) = 0$. □

An immediate consequence of the previous result is the following:

COROLLARY 5.27: LOCAL EXTREMA IN AN INTERVAL

Let $I \subseteq \mathbb{R}$ be an interval and $f : I \rightarrow \mathbb{R}$. If $x_0 \in I$ is a local extremum of f , then at least one of the following holds:

1. x_0 is an endpoint of I ;
2. f is not differentiable at x_0 ;
3. f is differentiable at x_0 and $f'(x_0) = 0$.

In particular, all local extrema of a differentiable function on an open interval are zeros of the derivative.

EXERCISE 5.28. — Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be the polynomial $f(x) = x^3 - x$.

- (i) Find all local extrema of f .
- (ii) Find all local extrema of the function $|f|$ on $[-3, 3]$.

5.2.2 The Mean Value Theorem

We now turn to general theorems of differential calculus and their consequences. Our first question is whether the derivative of a differentiable function attains the slope of certain secants; the next theorem is the starting point.

THEOREM 5.29: ROLLE'S THEOREM

Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b) . If $f(a) = f(b)$, then there exists $\xi \in (a, b)$ with $f'(\xi) = 0$.

Proof. By Theorem 3.34, f attains its minimum and maximum on $[a, b]$ at some points $x_0, x_1 \in [a, b]$. By Proposition 5.26, any interior extremum has zero derivative. Thus, we consider two cases:

- (i) If either x_0 or x_1 lies in (a, b) , we are done.
- (ii) If both x_0 and x_1 are endpoints, since $f(a) = f(b)$ then $\min f = \max f = f(a) = f(b)$, therefore f is constant. In particular, $f'(x) = 0$ for all $x \in (a, b)$ and the result follows also in this case. \square

An immediate consequence of this result is the following:

COROLLARY 5.30: NON-VANISHING DERIVATIVE IMPLIES DIFFERENT ENDPOINTS VALUES

Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b) . If $f'(x) \neq 0$ for all $x \in (a, b)$, then $f(a) \neq f(b)$.

Proof. If $f(a) = f(b)$, then Rolle's Theorem would imply the existence of $\xi \in (a, b)$ with $f'(\xi) = 0$, contradicting the assumption that f' never vanishes. \square

THEOREM 5.31: MEAN VALUE THEOREM

Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b) . Then there exists $\xi \in (a, b)$ such that

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

Proof. Define $g : [a, b] \rightarrow \mathbb{R}$ by

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

Then g is continuous on $[a, b]$, differentiable on (a, b) , and satisfies

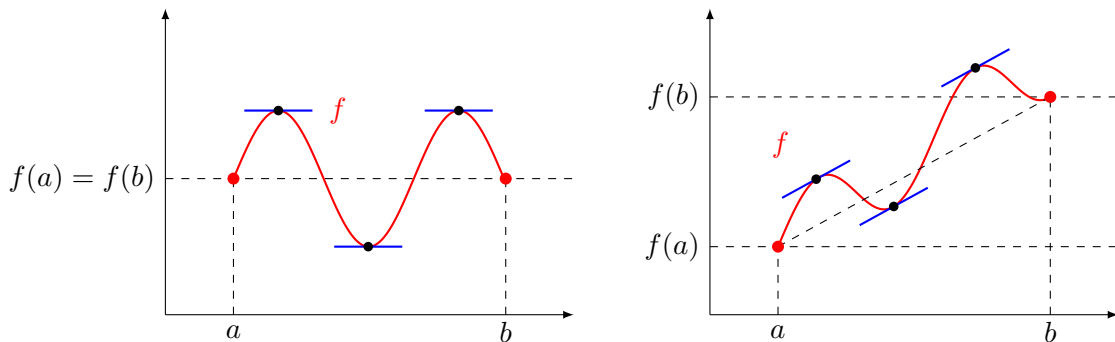
$$g(a) = f(a), \quad g(b) = f(b) - (f(b) - f(a)) = f(a).$$

By Rolle's Theorem (Theorem 5.29), there exists $\xi \in (a, b)$ with

$$0 = g'(\xi) = f'(\xi) - \frac{f(b) - f(a)}{b - a},$$

proving the result. \square

In words, Rolle's Theorem states that if a differentiable function on an interval takes the same value at both endpoints, then its slope must be zero somewhere in between (left image). The Mean Value Theorem, on the other hand, asserts that for any differentiable function on an interval, there exists a point where the slope equals the average slope (right image). Moreover, the Mean Value Theorem can be reduced to Rolle's Theorem by subtracting from f a suitable linear function so that the endpoint values coincide.



Using the Mean Value Theorem, we can show that for differentiable functions on an interval, the notion of Lipschitz continuity introduced in Exercise 3.41 is equivalent to having a bounded derivative.

COROLLARY 5.32: LIPSCHITZ CONTINUITY VS BOUNDED DERIVATIVE

Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b) . Then f is Lipschitz continuous on $[a, b]$ if and only if f' is bounded on (a, b) .

Proof. Suppose first that f is Lipschitz on $[a, b]$ with constant L . This implies that, given $x, x_0 \in (a, b)$ with $x \neq x_0$,

$$\left| \frac{f(x) - f(x_0)}{x - x_0} \right| \leq L.$$

Taking the limit as $x \rightarrow x_0$ gives $|f'(x_0)| \leq L$, so f' is bounded on (a, b) .

Conversely, suppose f' is bounded on (a, b) , say $|f'| \leq M$ for all $z \in (a, b)$. Then, given $x, y \in [a, b]$ with $x < y$, the Mean Value Theorem (Theorem 5.31) applied on the interval $[x, y]$ yields $\xi \in (x, y) \subset (a, b)$ with

$$f(y) - f(x) = f'(\xi)(y - x),$$

therefore

$$|f(y) - f(x)| = |f'(\xi)| |y - x| \leq M |y - x|.$$

Since $x, y \in [a, b]$ are arbitrary, this shows that f is Lipschitz on $[a, b]$ with Lipschitz constant M . \square

24

EXERCISE 5.33. — Let $[a, b]$ be a compact interval and $f : [a, b] \rightarrow \mathbb{R}$ be continuously differentiable. Show that f is Lipschitz continuous. What happens if compactness is dropped?

EXAMPLE 5.34. — Let $f : [0, 2\pi] \rightarrow \mathbb{C}$ be given by $f(x) = e^{ix} = \cos x + i \sin x$. Then $f(0) = f(2\pi) = 1$, but

$$f'(x) = ie^{ix} \neq 0 \quad \forall x \in [0, 2\pi].$$

Thus Rolle's theorem and the Mean Value Theorem fail for complex-valued functions.

THEOREM 5.35: CAUCHY MEAN VALUE THEOREM

Let $f, g : [a, b] \rightarrow \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b) . Then there exists $\xi \in (a, b)$ such that

$$g'(\xi)(f(b) - f(a)) = f'(\xi)(g(b) - g(a)). \quad (5.6)$$

If, in addition, $g'(x) \neq 0$ for all $x \in (a, b)$, then $g(a) \neq g(b)$ and

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

Proof. Define the function $F : [a, b] \rightarrow \mathbb{R}$ as

$$F(x) = g(x)(f(b) - f(a)) - f(x)(g(b) - g(a)).$$

Then

$$F(a) = g(a)(f(b) - f(a)) - f(a)(g(b) - g(a)) = g(a)f(b) - f(a)g(b),$$

$$F(b) = g(b)(f(b) - f(a)) - f(b)(g(b) - g(a)) = g(a)f(b) - f(a)g(b).$$

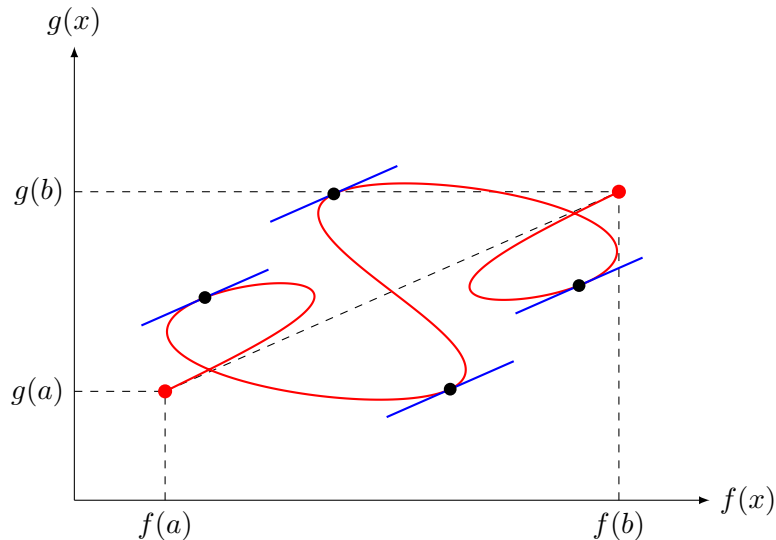
Thus, by Rolle's Theorem (Theorem 5.29), there exists $\xi \in (a, b)$ with

$$0 = F'(\xi) = g'(\xi)(f(b) - f(a)) - f'(\xi)(g(b) - g(a)),$$

which is (5.6).

If $g'(x) \neq 0$ for all x , then Corollary 5.30 yields $g(a) \neq g(b)$. Dividing (5.6) by $g'(\xi)(g(b) - g(a))$ yields the second formula. \square

As with the Mean Value Theorem 5.31, Cauchy's theorem has a geometric interpretation: under the stated assumptions, the curve $t \mapsto (f(t), g(t))$ has a tangent parallel to the line through $(f(a), g(a))$ and $(f(b), g(b))$.



The goal of the next exercises is to show that derivatives satisfy the *intermediate value property*: although f' need not be continuous, it cannot have jump discontinuities. This result is known as **Darboux's Theorem**.

EXERCISE 5.36. — Follow the steps below to prove the following statement:

Let $f : (a, b) \rightarrow \mathbb{R}$ be differentiable, and let $x_0 < x_1$ be points in (a, b) with $f'(x_0) \neq f'(x_1)$. Show that, for every α between $f'(x_0)$ and $f'(x_1)$, there exists $c \in (x_0, x_1)$ such that $f'(c) = \alpha$.

1. Without loss of generality, assume $f'(x_0) < f'(x_1)$ and fix $\alpha \in (f'(x_0), f'(x_1))$.
2. Consider the auxiliary function $g(x) = f(x) - \alpha x$ and show that g attains a minimum at some $c \in [x_0, x_1]$.
3. Prove that c cannot be one of the endpoints:

(i) Since

$$\frac{g(x_0 + h) - g(x_0)}{h} \xrightarrow{h \rightarrow 0^+} g'(x_0) = f'(x_0) - \alpha,$$

and $f'(x_0) < \alpha$, conclude that $c \neq x_0$.

(ii) Similarly, using that

$$\frac{g(x_1) - g(x_1 - h)}{h} \xrightarrow{h \rightarrow 0^+} g'(x_1) = f'(x_1) - \alpha,$$

and $f'(x_1) > \alpha$, conclude that $c \neq x_1$.

4. Deduce that $c \in (x_0, x_1)$, and use Proposition 5.26 to show that $g'(c) = 0$, hence $f'(c) = \alpha$.

24

5.2.3 L'Hôpital's Rule

The results collectively known as *L'Hôpital's rule* (also spelled L'Hospital) are named after Guillaume François Antoine, Marquis de l'Hôpital (1661–1704). The rule likely goes back to Johann Bernoulli, but was published by de l'Hôpital in his *Analyse des Infiniment Petits pour l'Intelligence des Lignes Courbes*, the first systematic treatment of infinitesimal calculus. However his approach was thoroughly geometric, as he had neither a rigorous notion of limit nor of differentiability.

THEOREM 5.37: L'HÔPITAL'S RULE

Let $f, g : (a, b) \rightarrow \mathbb{R}$ be differentiable. Suppose:

1. $g(x) \neq 0$ and $g'(x) \neq 0$ for all $x \in (a, b)$;

2. $\lim_{x \rightarrow a^+} f(x) = \lim_{x \rightarrow a^+} g(x) = 0$;

3. the limit $L = \lim_{x \rightarrow a^+} \frac{f'(x)}{g'(x)}$ exists.

Then $\lim_{x \rightarrow a^+} \frac{f(x)}{g(x)}$ exists and equals L .

Proof. By (2), we can extend f and g continuously to $[a, b]$ by setting $f(a) = g(a) = 0$. Fix $\varepsilon > 0$. By (3), there exists $\delta > 0$ such that

$$\frac{f'(\xi)}{g'(\xi)} \in (L - \varepsilon, L + \varepsilon) \quad \forall \xi \in (a, a + \delta).$$

Now, for any $x \in (a, a + \delta)$, we can apply Cauchy's Mean Value Theorem (Theorem 5.35) to f and g on $[a, x]$ to obtain some $\xi_x \in (a, x)$ with

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f'(\xi_x)}{g'(\xi_x)}.$$

Since $\xi_x \in (a, x) \subset (a, a + \delta)$, it follows that

$$\frac{f(x)}{g(x)} = \frac{f'(\xi_x)}{g'(\xi_x)} \in (L - \varepsilon, L + \varepsilon) \quad \text{for all } x \in (a, a + \delta).$$

24

Because $\varepsilon > 0$ is arbitrary, this proves that $\lim_{x \rightarrow a^+} \frac{f(x)}{g(x)} = L$. □

Theorem 5.37 is one of several versions of L'Hôpital's rule. For instance, one can allow both limits in (2) to be improper, i.e., $\lim_{x \rightarrow a^+} g(x) = \pm\infty$ and $\lim_{x \rightarrow a^+} f(x) = \pm\infty$ (with arbitrary signs). More precisely:

THEOREM 5.38: L'HÔPITAL'S RULE FOR IMPROPER LIMITS

Let $f, g : (a, b) \rightarrow \mathbb{R}$ be differentiable. Suppose:

1. $g(x) \neq 0$ and $g'(x) \neq 0$ for all $x \in (a, b)$;

2. $\lim_{x \rightarrow a^+} |f(x)| = \lim_{x \rightarrow a^+} |g(x)| = \infty$;

3. the limit $L = \lim_{x \rightarrow a^+} \frac{f'(x)}{g'(x)}$ exists.

Then $\lim_{x \rightarrow a^+} \frac{f(x)}{g(x)}$ exists and equals L .

Extra material: Proof of Theorem 5.38

Proof. Fix $\varepsilon > 0$. By (3) there exists $\delta > 0$ such that

$$\frac{f'(\xi)}{g'(\xi)} \in (L - \varepsilon, L + \varepsilon) \quad \forall \xi \in (a, a + \delta).$$

For $x \in (a, a + \delta)$, apply Cauchy's Mean Value Theorem (Theorem 5.35) on $[x, a + \delta]$ to obtain $\xi_x \in (x, a + \delta)$ with

$$\frac{f(x) - f(a + \delta)}{g(x) - g(a + \delta)} = \frac{f'(\xi_x)}{g'(\xi_x)} \in (L - \varepsilon, L + \varepsilon).$$

Hence

$$\frac{f(x) - f(a + \delta)}{g(x) - g(a + \delta)} \in (L - \varepsilon, L + \varepsilon) \quad \forall x \in (a, a + \delta). \quad (5.7)$$

We now observe that

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(a + \delta)}{g(x) - g(a + \delta)} \cdot \frac{1 - \frac{g(a + \delta)}{g(x)}}{1 - \frac{f(a + \delta)}{f(x)}}. \quad (5.8)$$

Since $|f(x)|, |g(x)| \rightarrow \infty$ as $x \rightarrow a^+$, the second factor in (5.8) tends to 1. Thus, combining (5.7) and (5.8), there exists $\eta \in (0, \delta)$ such that

$$\frac{f(x)}{g(x)} \in (L - 2\varepsilon, L + 2\varepsilon) \quad \forall x \in (a, a + \eta).$$

Since $\varepsilon > 0$ is arbitrary, $\lim_{x \rightarrow a^+} \frac{f(x)}{g(x)} = L$. □

EXAMPLE 5.39. — We illustrate the use of l'Hôpital's rule (applied repeatedly) by computing

$$\lim_{x \rightarrow 0} \frac{\sin x - x}{x^3}.$$

First, note that

$$\lim_{x \rightarrow 0} (\sin x - x) = 0, \quad \lim_{x \rightarrow 0} x^3 = 0,$$

so we have an indeterminate form of type $\frac{0}{0}$ and l'Hôpital's rule applies. Therefore

$$\lim_{x \rightarrow 0} \frac{\sin x - x}{x^3} = \lim_{x \rightarrow 0} \frac{(\sin x - x)'}{(x^3)'} = \lim_{x \rightarrow 0} \frac{\cos x - 1}{3x^2}.$$

At this stage we still have

$$\lim_{x \rightarrow 0} (\cos x - 1) = 0, \quad \lim_{x \rightarrow 0} 3x^2 = 0,$$

so the expression is again of type $\frac{0}{0}$ and we may apply l'Hôpital's rule a second time:

$$\lim_{x \rightarrow 0} \frac{\cos x - 1}{3x^2} = \lim_{x \rightarrow 0} \frac{(\cos x - 1)'}{(3x^2)'} = \lim_{x \rightarrow 0} \frac{-\sin x}{6x}.$$

Since

$$\lim_{x \rightarrow 0} (-\sin x) = 0, \quad \lim_{x \rightarrow 0} 6x = 0,$$

we still have a $\frac{0}{0}$ form, and we apply l'Hôpital's rule a third time:

$$\lim_{x \rightarrow 0} \frac{-\sin x}{6x} = \lim_{x \rightarrow 0} \frac{-\cos x}{6} = -\frac{1}{6}.$$

Thus

$$\lim_{x \rightarrow 0} \frac{\sin x - x}{x^3} = -\frac{1}{6}.$$

This example shows how l'Hôpital's rule can be iterated, provided that at each step the new quotient is still of indeterminate form $\frac{0}{0}$ (or $\frac{\infty}{\infty}$) and the hypotheses of the theorem remain satisfied.

Instead of the one-sided limits $x \rightarrow a^+$ or $x \rightarrow b^-$, one can also consider $x \rightarrow -\infty$ or $x \rightarrow \infty$:

THEOREM 5.40: L'HÔPITAL'S RULE AT INFINITY

Let $R > 0$ and $f, g : (R, \infty) \rightarrow \mathbb{R}$ be differentiable. Suppose:

1. $g(x) \neq 0$ and $g'(x) \neq 0$ for all $x \in (R, \infty)$;
2. either $\lim_{x \rightarrow \infty} f(x) = \lim_{x \rightarrow \infty} g(x) = 0$ or $\lim_{x \rightarrow \infty} |f(x)| = \lim_{x \rightarrow \infty} |g(x)| = \infty$;
3. the limit $L = \lim_{x \rightarrow \infty} \frac{f'(x)}{g'(x)}$ exists.

Then $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)}$ exists and equals L .

Proof. If $\lim_{x \rightarrow \infty} f(x) = \lim_{x \rightarrow \infty} g(x) = 0$, apply Theorem 5.37 on $(0, \frac{1}{R})$ to the functions $x \mapsto f(\frac{1}{x})$ and $x \mapsto g(\frac{1}{x})$. If $\lim_{x \rightarrow \infty} |f(x)| = \lim_{x \rightarrow \infty} |g(x)| = \infty$, apply instead Theorem 5.38 on $(0, \frac{1}{R})$ to the functions $x \mapsto f(\frac{1}{x})$ and $x \mapsto g(\frac{1}{x})$. \square

REMARK 5.41. — As discussed in the following exercise, the proofs of Theorems 5.37, 5.38, and 5.40 also apply (with small modifications) when $\frac{f'(x)}{g'(x)}$ diverges to $+\infty$ or $-\infty$. In this case, one concludes that $\frac{f(x)}{g(x)}$ diverges to the same infinite limit as $\frac{f'(x)}{g'(x)}$.

EXERCISE 5.42. — Prove Theorem 5.37 in the case where $\lim_{x \rightarrow a^+} \frac{f'(x)}{g'(x)} = +\infty$.

Hint: Repeat the argument in the proof of Theorem 5.37, replacing the condition $\frac{f'(\xi)}{g'(\xi)} \in (L - \varepsilon, L + \varepsilon)$ with inequalities of the form $\frac{f'(\xi)}{g'(\xi)} > M$ for arbitrary $M > 0$.

EXERCISE 5.43. — Compute the following limits using L'Hôpital's rule:

$$(a) \lim_{x \rightarrow 0^+} \frac{\sin x - x}{x^2 \sin x}, \quad (b) \lim_{x \rightarrow 0} \frac{e^x - x - 1}{\cos x - 1}, \quad (c) \lim_{x \rightarrow 2} \frac{x^4 - 4^x}{\sin(\pi x)}, \quad (d) \lim_{x \rightarrow -\infty} x^3 e^x.$$

EXERCISE 5.44. — Let $a < b$ and let $f : [a, b] \rightarrow \mathbb{R}$ be continuous. Suppose $x_0 \in [a, b]$, that f is differentiable on $[a, b] \setminus \{x_0\}$, and that $\lim_{x \rightarrow x_0} f'(x)$ exists. Show that f is differentiable at x_0 and that f' is continuous at x_0 .

EXERCISE 5.45. — Let $I = (a, b) \subseteq \mathbb{R}$ and let $f : I \rightarrow \mathbb{R}$ be twice differentiable. Applying L'Hôpital's rule twice, show that

$$f''(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} \quad \forall x \in I.$$

5.2.4 Monotonicity and Convexity via Differential Calculus

The mean value theorem allows us to characterize some properties of functions via the derivative. In what follows, $f' \geq 0$ means $f'(x) \geq 0$ for all $x \in I$. Here I always denotes a nontrivial interval (nonempty and not a single point).

PROPOSITION 5.46: MONOTONICITY VS. FIRST DERIVATIVE

Let $I \subseteq \mathbb{R}$ be an interval, and let $f : I \rightarrow \mathbb{R}$ be differentiable. Then

$$f' \geq 0 \iff f \text{ is increasing.}$$

Proof. If f is increasing, then $f(x+h) - f(x) \geq 0$ for $h > 0$, and $f(x+h) - f(x) \leq 0$ for $h < 0$. Hence, in both cases, $\frac{f(x+h)-f(x)}{h} \geq 0$, therefore

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \geq 0.$$

Conversely, suppose f is not increasing. Then there exist $x_1 < x_2$ with $f(x_2) < f(x_1)$. By the Mean Value Theorem 5.31, there exists $\xi \in (x_1, x_2)$ such that

$$f'(\xi) = \frac{f(x_2) - f(x_1)}{x_2 - x_1} < 0,$$

so $f' \not\geq 0$ on I . □

REMARK 5.47. — If $f' > 0$, the same argument shows that f is strictly increasing. However the converse fails: the function $f(x) = x^3$ is strictly increasing, but $f'(0) = 0$.

COROLLARY 5.48: CONSTANT FUNCTIONS VS. FIRST DERIVATIVE

Let $I \subseteq \mathbb{R}$ be an interval and $f : I \rightarrow \mathbb{R}$. Then f is constant if and only if f is differentiable and $f'(x) = 0$ for all $x \in I$.

Proof. The derivative of a constant is 0.

Conversely, if $f' = 0$, then $f' \geq 0$ and $-f' \geq 0$, so by Proposition 5.46 both f and $-f$ are increasing, hence f is constant. □

EXERCISE 5.49. — Let $I \subseteq \mathbb{R}$ and $f : I \rightarrow \mathbb{R}$. Show that f is a polynomial if and only if f is smooth and there exists $n \in \mathbb{N}$ such that $f^{(n)} \equiv 0$.

DEFINITION 5.50: CONVEX FUNCTIONS

Let $I \subseteq \mathbb{R}$ and $f : I \rightarrow \mathbb{R}$. We call f **convex** if, for all $a, b \in I$ with $a < b$ and all $t \in (0, 1)$,

$$f((1-t)a + tb) \leq (1-t)f(a) + tf(b). \quad (5.9)$$

We call f **strictly convex** if the inequality in (5.9) is strict. A function $g : I \rightarrow \mathbb{R}$ is **(strictly) concave** if $-g$ is (strictly) convex.

Geometrically, (5.9) says that on every interval $[a, b]$ the graph of f lies below the secant through $(a, f(a))$ and $(b, f(b))$.

An equivalent definition of a convex function is the following:

$f : I \rightarrow \mathbb{R}$ is convex if for all $a, b \in I$ with $a < b$ and all $x \in (a, b)$,

$$\frac{f(x) - f(a)}{x - a} \leq \frac{f(b) - f(x)}{b - x}, \quad (5.10)$$

and strictly convex if the inequality is strict.

Geometrically, this definition corresponds to saying that the “right” secant is steeper than the “left” one.

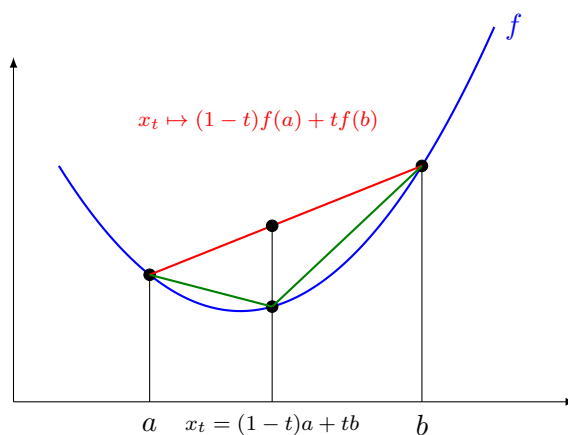


Figure 5.3: In blue, a convex function. In red, the secant through $(a, f(a))$ and $(b, f(b))$, which lies above the function f . In green, the secants between $(a, f(a))$, $(x_t, f(x_t))$ and $(x_t, f(x_t))$, $(b, f(b))$, the latter being steeper than the former.

EXERCISE 5.51. — Show that (5.9) for all $t \in (0, 1)$ is equivalent to (5.10) for all $x \in (a, b)$.

PROPOSITION 5.52: CONVEXITY VS. MONOTONICITY OF THE FIRST DERIVATIVE

Let $I \subseteq \mathbb{R}$ and let $f : I \rightarrow \mathbb{R}$ be differentiable. Then f is convex if and only if f' is increasing.

Proof. Assume f' is increasing. Then, for $a < b$ and $x \in (a, b)$, the Mean Value Theorem 5.31 applied on the intervals $[a, x]$ and $[x, b]$ yields $\xi \in (a, x)$ and $\zeta \in (x, b)$ such that

$$f'(\xi) = \frac{f(x) - f(a)}{x - a}, \quad f'(\zeta) = \frac{f(b) - f(x)}{b - x}.$$

Since f' is increasing we have $f'(\xi) \leq f'(\zeta)$, so (5.10) follows. Since $a < b \in I$ and $x \in (a, b)$ are arbitrary, f is convex.

Conversely, assume f is convex. Given $a < b$, consider $h > 0$ small enough so that $a + h < b - h$ and apply (5.10) twice: first, applying it on the interval $(a, b - h)$ with $x = a + h$ we get

$$\frac{f(a + h) - f(a)}{h} \leq \frac{f(b - h) - f(a + h)}{(b - h) - (a + h)};$$

then, applying it on the interval $(a + h, b)$ with $x = b - h$ we obtain

$$\frac{f(b - h) - f(a + h)}{(b - h) - (a + h)} \leq \frac{f(b) - f(b - h)}{h}.$$

Combining these two inequalities we deduce that, for all sufficiently small $h > 0$,

$$\frac{f(a + h) - f(a)}{h} \leq \frac{f(b) - f(b - h)}{h}. \quad (5.11)$$

Letting $h \rightarrow 0^+$ gives $f'(a) \leq f'(b)$. Since $a < b$ are arbitrary, f' is increasing. \square

EXERCISE 5.53. — Under the assumptions of Proposition 5.52, prove that f is strictly convex if and only if f' is strictly increasing.

COROLLARY 5.54: CONVEXITY VS. SECOND DERIVATIVE

Let $I \subseteq \mathbb{R}$ and let $f : I \rightarrow \mathbb{R}$ be twice differentiable. Then f is convex if and only if $f'' \geq 0$.

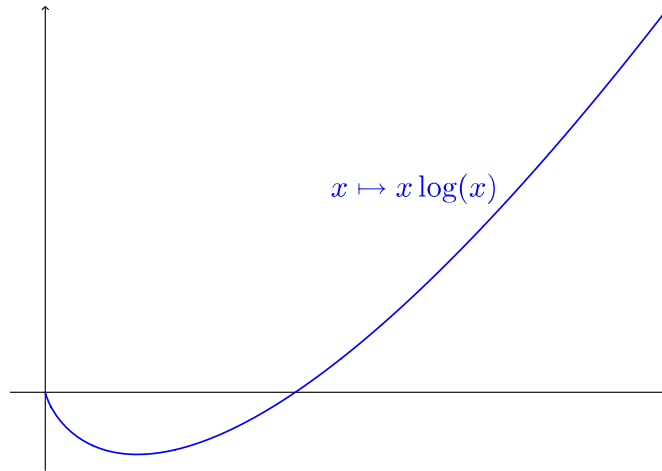
Proof. By Proposition 5.52, f is convex if and only if f' is increasing. Applying Proposition 5.46 to f' , we have that f' increasing if and only if $f'' \geq 0$. \square

EXERCISE 5.55. — Under the assumptions of Corollary 5.54, show that if $f''(x) > 0$ for all $x \in I$, then f is strictly convex. Is the converse true?

EXERCISE 5.56. — The function $f : (0, \infty) \rightarrow \mathbb{R}$, $x \mapsto x \log x$, is strictly convex. Indeed, f is smooth and

$$f'(x) = \log x + 1, \quad f''(x) = \frac{1}{x} > 0 \quad \forall x > 0,$$

so Exercise 5.55 applies. Moreover, $\lim_{x \rightarrow 0^+} x \log x = 0$ (see Example 3.64) and $\lim_{x \rightarrow 0^+} f'(x) = -\infty$, as reflected in the graph below.



25

EXERCISE 5.57 (Minima of Convex Functions). — Let $I \subseteq \mathbb{R}$ and $f : I \rightarrow \mathbb{R}$ be convex. Show that every local minimum of f is a global minimum.

EXERCISE 5.58. — Show that for all real numbers $x \geq -1$ and $p \geq 1$, the continuous Bernoulli inequality holds:

$$(1+x)^p \geq 1+px.$$

Hint: Define $f : [-1, \infty) \rightarrow \mathbb{R}$ as $f(x) = (1+x)^p - 1 - px$. Prove that f is convex and satisfies $f(0) = f'(0) = 0$. Use these facts to deduce that $f \geq 0$ on $[-1, \infty)$.

EXERCISE 5.59. — Given $\alpha \in (0, 1]$, show that the function $x \mapsto x^\alpha$ is concave on $(0, \infty)$, and use this fact to provide an alternative proof of (3.15).

5.3 Example: Differentiation of Trigonometric Functions

In this section we study the derivative and monotonicity properties of the trigonometric functions.

5.3.1 Sine and Arc Sine

Recalling Exercise 5.13, the functions $\sin : \mathbb{R} \rightarrow \mathbb{R}$ and $\cos : \mathbb{R} \rightarrow \mathbb{R}$ are smooth and satisfy

$$\sin'(x) = \cos(x), \quad \cos'(x) = -\sin(x).$$

By Theorem 4.62 and Exercise 4.64, the zeros of $\cos : \mathbb{R} \rightarrow \mathbb{R}$ are the points $\{\frac{\pi}{2} + k\pi \mid k \in \mathbb{Z}\}$, and $\cos(0) = 1$. By the Intermediate Value Theorem (Theorem 3.24), it follows that $\sin'(x) = \cos(x) > 0$ for all $x \in (-\frac{\pi}{2}, \frac{\pi}{2})$. Hence, by Remark 5.47, the function

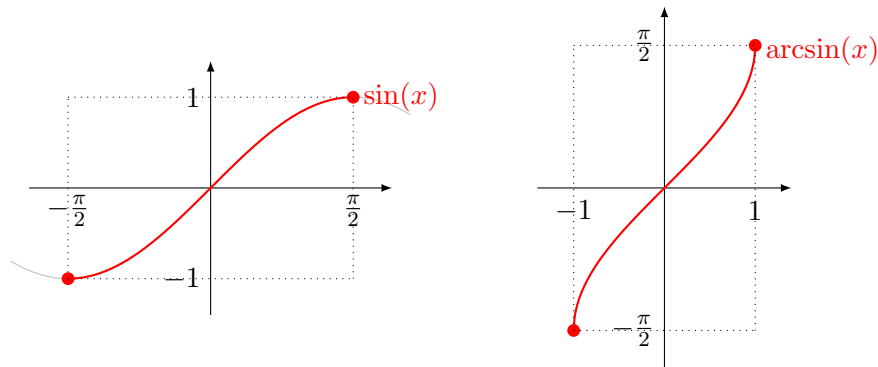
$$\sin : [-\frac{\pi}{2}, \frac{\pi}{2}] \rightarrow [-1, 1] \quad (5.12)$$

is strictly increasing and bijective (recall that $\sin(-\frac{\pi}{2}) = -1$ and $\sin(\frac{\pi}{2}) = 1$). Consequently, the restriction of the sine function to $[-\frac{\pi}{2}, \frac{\pi}{2}]$ has an inverse, denoted

$$\arcsin : [-1, 1] \rightarrow [-\frac{\pi}{2}, \frac{\pi}{2}],$$

called the **arcsine**.

The following figure shows the graphs of the sine function on $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and of its inverse.



REMARK 5.60. — Since $\sin'' = -\sin$, it follows that \sin is convex on $(-\frac{\pi}{2}, 0)$ and concave on $(0, \frac{\pi}{2})$.

By Theorem 5.20, the arcsine is differentiable at s whenever the derivative of the sine at $x = \arcsin(s)$ is nonzero. Since $\sin' = \cos$ vanishes only at the endpoints of $[-\frac{\pi}{2}, \frac{\pi}{2}]$, for $x \in (-\frac{\pi}{2}, \frac{\pi}{2})$ and $s = \sin(x)$ we obtain

$$\arcsin'(s) = \frac{1}{\cos(x)} = \frac{1}{\sqrt{1 - \sin^2(x)}} = \frac{1}{\sqrt{1 - s^2}},$$

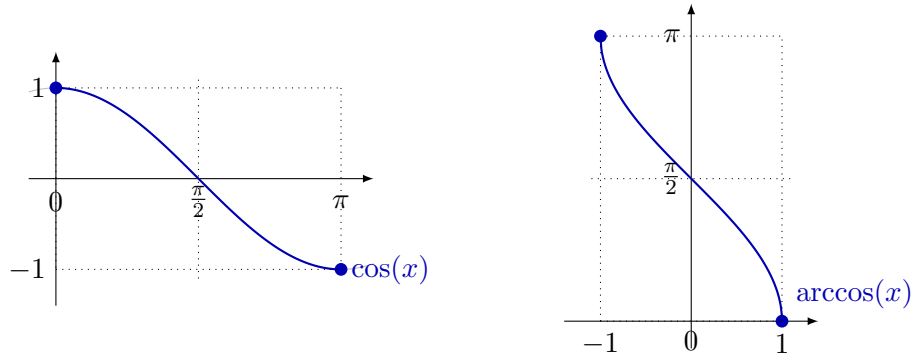
where we used that $\cos(x) > 0$ for $x \in (-\frac{\pi}{2}, \frac{\pi}{2})$, thus $\cos(x) = \sqrt{1 - \sin^2(x)}$.

5.3.2 Cosine and Arc Cosine

A similar discussion applies to the cosine function. The cosine is strictly decreasing on the interval $[0, \pi]$, and satisfies $\cos(0) = 1$ and $\cos(\pi) = -1$. Hence,

$$\cos : [0, \pi] \rightarrow [-1, 1]$$

is bijective.



The inverse function is called the **arccosine** and is denoted by

$$\arccos : [-1, 1] \rightarrow [0, \pi].$$

Applying the differentiation rule for the inverse function, for $x \in (0, \pi)$ and $s = \cos(x)$ we obtain

$$\arccos'(s) = \frac{1}{-\sin(x)} = -\frac{1}{\sqrt{1 - \cos^2(x)}} = -\frac{1}{\sqrt{1 - s^2}}.$$

REMARK 5.61. — Since $\cos'' = -\cos$, it follows that \cos is concave on $(0, \frac{\pi}{2})$ and convex on $(\frac{\pi}{2}, \pi)$.

5.3.3 Tangent and Arc Tangent

Consider the restriction $\tan : (-\frac{\pi}{2}, \frac{\pi}{2}) \rightarrow \mathbb{R}$ of the tangent function. Using the quotient rule (Corollary 5.17) we find

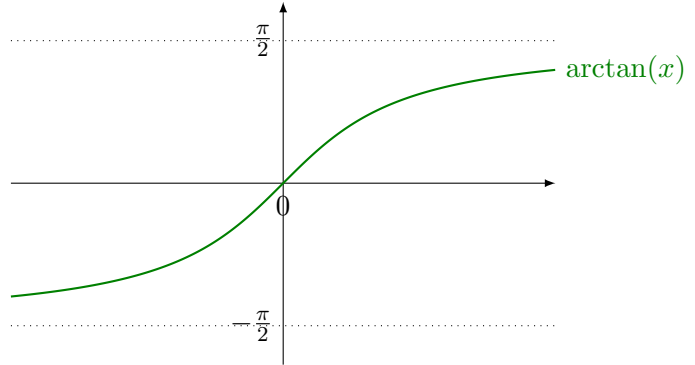
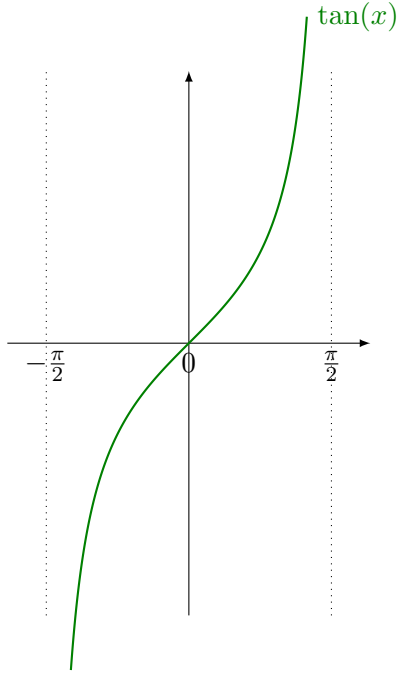
$$\tan'(x) = \left(\frac{\sin(x)}{\cos(x)} \right)' = \frac{\cos(x)\cos(x) - \sin(x)(-\sin(x))}{\cos^2(x)} = \frac{1}{\cos^2(x)}$$

for all $x \in (-\frac{\pi}{2}, \frac{\pi}{2})$. Hence \tan is strictly increasing on $(-\frac{\pi}{2}, \frac{\pi}{2})$. Moreover,

$$\lim_{x \rightarrow (\frac{\pi}{2})^-} \tan(x) = \lim_{x \rightarrow (\frac{\pi}{2})^-} \frac{\sin(x)}{\cos(x)} = +\infty, \quad \lim_{x \rightarrow (-\frac{\pi}{2})^+} \tan(x) = \lim_{x \rightarrow (-\frac{\pi}{2})^+} \frac{\sin(x)}{\cos(x)} = -\infty.$$

Hence, by the Intermediate Value Theorem, the tangent function is bijective:

$$\tan : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \rightarrow \mathbb{R}.$$



25

Its inverse

$$\arctan : \mathbb{R} \rightarrow \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$$

is called the **arctangent**. By Theorem 5.20, for $x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ and $s = \tan(x)$ we obtain

$$\arctan'(s) = \frac{1}{\frac{1}{\cos^2(x)}} = \cos^2(x).$$

Since

$$s^2 = \tan^2(x) = \frac{\sin^2(x)}{\cos^2(x)} = \frac{1 - \cos^2(x)}{\cos^2(x)} = \frac{1}{\cos^2(x)} - 1,$$

it follows that $1 + s^2 = \frac{1}{\cos^2(x)}$, therefore

$$\arctan'(s) = \frac{1}{1 + s^2} \quad \forall s \in \mathbb{R}.$$

The cotangent behaves similarly. The restriction $\cot|_{(0,\pi)} : (0, \pi) \rightarrow \mathbb{R}$ is strictly decreasing and bijective. Its inverse

$$\operatorname{arccot} : \mathbb{R} \rightarrow (0, \pi)$$

is called the **arccotangent** and satisfies

$$\operatorname{arccot}'(s) = -\frac{1}{1 + s^2} \quad \forall s \in \mathbb{R}.$$

5.3.4 Hyperbolic Functions

We now perform the analogous analysis for the hyperbolic trigonometric functions:

$$\sinh(x) = \frac{e^x - e^{-x}}{2}, \quad \cosh(x) = \frac{e^x + e^{-x}}{2}, \quad \tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

We have $\sinh'(x) = \cosh(x) > 0$ for all $x \in \mathbb{R}$, so by Proposition 5.46 the hyperbolic sine is strictly increasing. Since $\lim_{x \rightarrow \infty} \sinh(x) = +\infty$ and $\lim_{x \rightarrow -\infty} \sinh(x) = -\infty$, the function

$$\sinh : \mathbb{R} \rightarrow \mathbb{R}$$

is bijective. Its inverse

$$\operatorname{arsinh} : \mathbb{R} \rightarrow \mathbb{R}$$

is called the **inverse hyperbolic sine**. By Theorem 5.20, for $x \in \mathbb{R}$ and $s = \sinh(x)$ we obtain

$$\operatorname{arsinh}'(s) = \frac{1}{\cosh(x)} = \frac{1}{\sqrt{1 + \sinh^2(x)}} = \frac{1}{\sqrt{1 + s^2}}.$$

Moreover, the inverse hyperbolic sine has a closed formula. Indeed, starting from $\sinh(x) = s$, we get

$$\frac{e^x - e^{-x}}{2} = s \quad \implies \quad e^{2x} - 2se^x - 1 = 0.$$

Setting $y = e^x$, this becomes $y^2 - 2sy - 1 = 0$, yielding

$$y = s \pm \sqrt{1 + s^2}.$$

Since $y = e^x > 0$, the admissible root is $y = s + \sqrt{1 + s^2}$, hence

$$e^x = y = s + \sqrt{1 + s^2} \quad \implies \quad x = \log(s + \sqrt{1 + s^2}).$$

In other words,

$$\operatorname{arsinh}(s) = \log(s + \sqrt{1 + s^2}), \quad s \in \mathbb{R}.$$

The hyperbolic cosine satisfies $\cosh'(x) = \sinh(x)$ and $\cosh''(x) = \cosh(x) > 0$ for all $x \in \mathbb{R}$. Thus, \cosh is strictly convex (Corollary 5.54) and has a global minimum at $x = 0$, since $\cosh'(0) = \sinh(0) = 0$. For $x > 0$, $\cosh'(x) > 0$, so \cosh is strictly increasing on $[0, \infty)$. As $\cosh(0) = 1$ and $\lim_{x \rightarrow \infty} \cosh(x) = +\infty$, we deduce

$$\cosh : [0, \infty) \rightarrow [1, \infty)$$

is bijective. Its inverse

$$\operatorname{arcosh} : [1, \infty) \rightarrow [0, \infty)$$

is called the **inverse hyperbolic cosine**. It is differentiable on $[1, \infty)$ and satisfies, for $s = \cosh(x)$ with $x > 0$ (thus $s > 1$),

$$\operatorname{arcosh}'(s) = \frac{1}{\sinh(x)} = \frac{1}{\sqrt{\cosh(x)^2 - 1}} = \frac{1}{\sqrt{s^2 - 1}}.$$

Furthermore, arguing as done before for the inverse hyperbolic sine,

$$\operatorname{arcosh}(s) = \log(s + \sqrt{s^2 - 1}) \quad \forall s \geq 1.$$

25

The **inverse hyperbolic tangent**,

$$\operatorname{artanh} : (-1, 1) \rightarrow \mathbb{R}, \quad \operatorname{artanh}(s) = \frac{1}{2} \log\left(\frac{1+s}{1-s}\right)$$

is the inverse function of the strictly increasing bijection $\tanh : \mathbb{R} \rightarrow (-1, 1)$, and it satisfies

$$\operatorname{artanh}'(s) = \frac{1}{1-s^2} \quad \forall s \in (-1, 1).$$

EXERCISE 5.62. — Verify all assertions made in Paragraphs 5.3.4, 5.3.4, and 5.3.4.

Chapter 6

The Riemann Integral

In this chapter, we take the idea from Section 1.1 and extend it to the notion of the Riemann integral, using the concepts of supremum and infimum.

6.1 Step Functions and their Integral

6.1.1 Decompositions and Step Functions

PARTITIONS

Two sets A, B are called **disjoint** if $A \cap B = \emptyset$. For a collection \mathcal{A} of sets, we say that the sets in \mathcal{A} are **pairwise disjoint** if for all $A_1, A_2 \in \mathcal{A}$ with $A_1 \neq A_2$ it holds that $A_1 \cap A_2 = \emptyset$.

Let X be a set. A **partition** of X is a family \mathcal{P} of non-empty pairwise disjoint subsets of X such that

$$X = \bigcup_{P \in \mathcal{P}} P.$$

In other words, each $P \in \mathcal{P}$ is non-empty, and every element of X belongs to exactly one $P \in \mathcal{P}$.

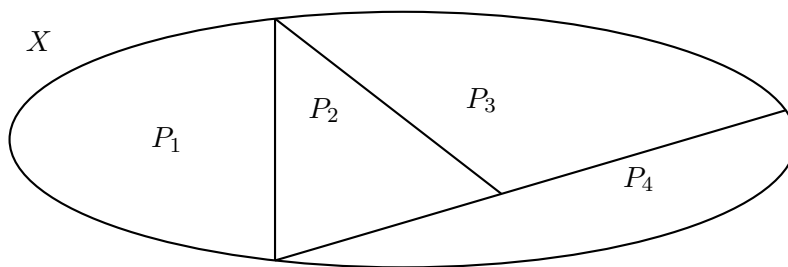


Figure 6.1: Schematic representation of a partition $\mathcal{P} = \{P_1, \dots, P_4\}$ of a set X .

For the following discussion, we fix two real numbers $a < b$ and work with the compact interval $[a, b] \subset \mathbb{R}$.

DEFINITION 6.1: DECOMPOSITION OF AN INTERVAL

A **decomposition** of $[a, b]$ is a finite sequence of points

$$a = x_0 < x_1 < \dots < x_{n-1} < x_n = b,$$

with $n \in \mathbb{N}$. The points x_0, \dots, x_n are called the **division points** of the decomposition.

Formally, a decomposition of $[a, b]$ is a finite subset of $[a, b]$ containing a and b , together with the ordering of its elements. Each decomposition induces a natural partition of $[a, b]$:

$$[a, b] = \{x_0\} \cup (x_0, x_1) \cup \{x_1\} \cup \dots \cup (x_{n-1}, x_n) \cup \{x_n\},$$

which we will use implicitly from now on.

A decomposition

$$a = y_0 < y_1 < \dots < y_m = b$$

is called a **refinement** of the decomposition

$$a = x_0 < x_1 < \dots < x_n = b$$

if

$$\{x_0, x_1, \dots, x_n\} \subseteq \{y_0, y_1, \dots, y_m\}.$$

The notion of refinement defines a partial order on the set of all decompositions of $[a, b]$. Note that any two decompositions of $[a, b]$ admit a common refinement given by the union of all division points.

DEFINITION 6.2: STEP FUNCTIONS

A function $f : [a, b] \rightarrow \mathbb{R}$ is called a **step function** if there exists a decomposition

$$a = x_0 < x_1 < \dots < x_n = b$$

such that, for each $k = 1, \dots, n$, the restriction of f to the open interval (x_{k-1}, x_k) is constant. In this case, we say that f is a step function *with respect to* the decomposition $a = x_0 < x_1 < \dots < x_n = b$.

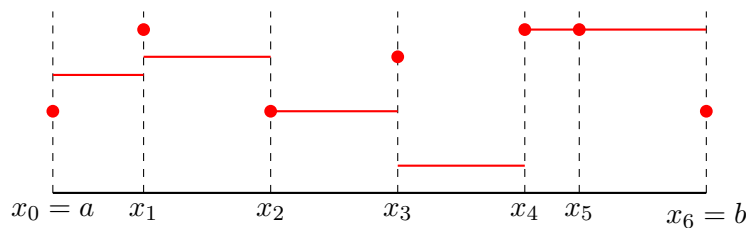


Figure 6.2: The graph of a step function on the interval $[a, b]$. Note that the values of f at the division points of the decomposition are irrelevant for the definition of a step function.

PROPOSITION 6.3: LINEARITY OF THE SPACE OF STEP FUNCTIONS

Let $f, g : [a, b] \rightarrow \mathbb{R}$ be step functions, and $\alpha, \beta \in \mathbb{R}$. Then $\alpha f + \beta g$ is also a step function.

Proof. Let f be a step function with respect to the decomposition $a = x_0 < x_1 < \dots < x_n = b$, and let g be a step function with respect to the decomposition $a = y_0 < y_1 < \dots < y_m = b$. The union of all division points $\{x_0, \dots, x_n\} \cup \{y_0, \dots, y_m\}$ defines a new decomposition

$$a = z_0 < z_1 < \dots < z_N = b$$

that is a common refinement of the two. Since both f and g are constant on each open interval (z_{k-1}, z_k) , so is the function $\alpha f + \beta g$. Thus $\alpha f + \beta g$ is a step function with respect to this decomposition. \square

EXAMPLE 6.4. — Constant functions are step functions.

REMARK 6.5. — As in the proof of Proposition 6.3, one can show that the product of two step functions is again a step function. Moreover, step functions are bounded, since they take only finitely many values.

26

6.1.2 The Integral of a Step Function**DEFINITION 6.6: INTEGRAL OF A STEP FUNCTION**

Let $f : [a, b] \rightarrow \mathbb{R}$ be a step function with respect to a decomposition

$$a = x_0 < x_1 < \dots < x_n = b.$$

We define the **integral** of f on $[a, b]$ as the real number

$$\int_a^b f(x) dx = \sum_{k=1}^n c_k (x_k - x_{k-1}), \quad (6.1)$$

where c_k denotes the constant value of f on the interval (x_{k-1}, x_k) .

REMARK 6.7. — For non-negative step functions $f \geq 0$, the value of the integral (6.1) can be interpreted geometrically as the total area of the rectangles of height c_k and base length $x_k - x_{k-1}$. In general, the integral represents the *signed net area* enclosed between the graph of f and the x -axis.

At this stage, in (6.1), the symbols \int and dx should be regarded as purely formal. Historically, the symbol \int is an elongated S for “sum,” and dx indicates an “infinitesimal length,” i.e. $x_k - x_{k-1}$ in the limit of an infinitely fine decomposition. This notation was introduced by Leibniz (1646–1716).

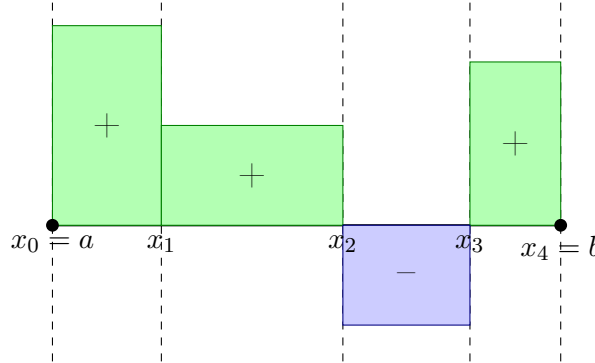


Figure 6.3: For a non-negative step function $f \geq 0$, the integral in (6.1) represents the area of the set $\{(x, y) \in \mathbb{R}^2 \mid a \leq x \leq b, 0 \leq y \leq f(x)\}$, and in general the signed area.

REMARK 6.8. — Definition 6.6 raises a potential issue: a priori, the right-hand side of (6.1) might depend on the chosen decomposition. To show that this is not the case, let

$$a = y_0 < \cdots < y_m = b$$

be another decomposition of $[a, b]$ with respect to which f is a step function. We must prove that

$$\sum_{k=1}^n c_k(x_k - x_{k-1}) = \sum_{k=1}^m d_k(y_k - y_{k-1}), \quad (6.2)$$

where d_k denotes the constant value of f on (y_{k-1}, y_k) . We argue in three steps:

1. If the decomposition $a = y_0 < \cdots < y_m = b$ is a refinement of $a = x_0 < \cdots < x_n = b$ differing by a single additional division point $y_\ell \in (x_{\ell-1}, x_\ell)$, then the two sums in (6.2) coincide, since $c_\ell = d_\ell = d_{\ell+1}$ and

$$c_\ell(x_\ell - x_{\ell-1}) = d_\ell(y_\ell - y_{\ell-1}) + d_{\ell+1}(y_{\ell+1} - y_\ell).$$

2. By induction on the number of additional division points, (6.2) holds for any refinement.
3. As shown in the proof of Proposition 6.3, any two decompositions admit a common refinement. Comparing both sums with the sum corresponding to this common refinement proves (6.2) in full generality.

Therefore, the value of the integral of a step function is *independent* of the chosen decomposition.

PROPOSITION 6.9: LINEARITY OF THE INTEGRAL OF STEP FUNCTIONS

Let $f, g : [a, b] \rightarrow \mathbb{R}$ be step functions, and let $\alpha, \beta \in \mathbb{R}$. Then

$$\int_a^b (\alpha f + \beta g)(x) dx = \alpha \int_a^b f(x) dx + \beta \int_a^b g(x) dx.$$

Proof. As in the proof of Proposition 6.3, we can find a decomposition $a = x_0 < \dots < x_n = b$ such that both f and g (and hence $\alpha f + \beta g$) are constant on each interval (x_{k-1}, x_k) . If f takes the value c_k and g the value d_k on (x_{k-1}, x_k) , then $\alpha f + \beta g$ takes the value $\alpha c_k + \beta d_k$. Thus

$$\begin{aligned} \int_a^b (\alpha f + \beta g)(x) dx &= \sum_{k=1}^n (\alpha c_k + \beta d_k)(x_k - x_{k-1}) \\ &= \alpha \sum_{k=1}^n c_k(x_k - x_{k-1}) + \beta \sum_{k=1}^n d_k(x_k - x_{k-1}) \\ &= \alpha \int_a^b f(x) dx + \beta \int_a^b g(x) dx, \end{aligned}$$

as claimed. \square

PROPOSITION 6.10: MONOTONICITY OF THE INTEGRAL OF STEP FUNCTIONS

Let $f, g : [a, b] \rightarrow \mathbb{R}$ be step functions such that $f \leq g$. Then

$$\int_a^b f(x) dx \leq \int_a^b g(x) dx.$$

Proof. As in the proofs of Proposition 6.3, we can find a decomposition $a = x_0 < \dots < x_n = b$ such that both f and g are constant on each interval (x_{k-1}, x_k) . Writing c_k and d_k for their respective values, the assumption $f \leq g$ implies $c_k \leq d_k$ for all $k = 1, \dots, n$. Hence

$$\int_a^b f(x) dx = \sum_{k=1}^n c_k(x_k - x_{k-1}) \leq \sum_{k=1}^n d_k(x_k - x_{k-1}) = \int_a^b g(x) dx.$$

\square

Applying Proposition 6.10 with $g \equiv 0$, we deduce the following:

COROLLARY 6.11: POSITIVITY OF THE INTEGRAL OF STEP FUNCTIONS

If $f : [a, b] \rightarrow \mathbb{R}$ is a step function such that $f(x) \geq 0$ for all $x \in [a, b]$, then

$$\int_a^b f(x) dx \geq 0.$$

EXERCISE 6.12. — Let $[a, b]$ and $[b, c]$ be two compact intervals, and let $f_1 : [a, b] \rightarrow \mathbb{R}$ and $f_2 : [b, c] \rightarrow \mathbb{R}$ be step functions. Define

$$f : [a, c] \rightarrow \mathbb{R}, \quad f(x) = \begin{cases} f_1(x), & x \in [a, b), \\ f_2(x), & x \in [b, c]. \end{cases}$$

26

1. Show that f is a step function on $[a, c]$.

2. Prove that

$$\int_a^c f(x) dx = \int_a^b f_1(x) dx + \int_b^c f_2(x) dx.$$

3. Show that every step function on $[a, c]$ is of the form described above.

6.2 Definition and First Properties of the Riemann Integral

As in the last section, we consider functions on a compact interval $[a, b] \subset \mathbb{R}$. To alleviate notation, we write \mathcal{SF} for the set of step functions on $[a, b]$. Also, we often write $\int_a^b f dx$ in place of $\int_a^b f(x) dx$.

6.2.1 Integrability of Real-valued Functions

Before defining lower and upper sums, we recall a simple but useful property of the supremum and infimum of two related sets. We will use this fact several times in what follows.

RELATION BETWEEN SUPREMUM AND INFIMUM

Let $A, B \subset \mathbb{R}$ be nonempty sets such that $s \leq t$ for all $s \in A$ and $t \in B$. Then

$$\sup A \leq \inf B. \quad (6.3)$$

Moreover,

$$\sup A = \inf B \iff \text{for every } \varepsilon > 0 \text{ there exist } s \in A \text{ and } t \in B \text{ s.t. } t - s < \varepsilon. \quad (6.4)$$

The following definition of integrability is a variant of Riemann's definition, which goes back to the French mathematician Jean-Gaston Darboux (1842–1917).

DEFINITION 6.13: LOWER AND UPPER SUMS

Let $f : [a, b] \rightarrow \mathbb{R}$ be a function. Define the sets of **lower sums** $\mathcal{L}(f) \subset \mathbb{R}$ and **upper sums** $\mathcal{U}(f) \subset \mathbb{R}$ by

$$\mathcal{L}(f) = \left\{ \int_a^b \ell dx \mid \ell \in \mathcal{SF} \text{ and } \ell \leq f \right\}, \quad \mathcal{U}(f) = \left\{ \int_a^b u dx \mid u \in \mathcal{SF} \text{ and } f \leq u \right\}.$$

If f is bounded, then these sets are non-empty. Indeed, if $|f| \leq M$, then the constant step functions

$$\ell(x) = -M \quad \forall x \in [a, b], \quad u(x) = M \quad \forall x \in [a, b],$$

satisfy $\ell \in \mathcal{L}(f)$ and $u \in \mathcal{U}(f)$.

For $\ell, u \in \mathcal{SF}$ with $\ell \leq f \leq u$, Proposition 6.10 gives

$$\int_a^b \ell dx \leq \int_a^b u dx.$$

This implies that $s \leq t$ for all $s \in \mathcal{L}(f)$ and $t \in \mathcal{U}(f)$, so (6.3) yields

$$\sup \mathcal{L}(f) \leq \inf \mathcal{U}(f).$$

DEFINITION 6.14: RIEMANN INTEGRAL

A bounded function $f : [a, b] \rightarrow \mathbb{R}$ is **Riemann integrable** if $\sup \mathcal{L}(f) = \inf \mathcal{U}(f)$. In this case, this common value is called the **Riemann integral** of f , and we write

$$\int_a^b f \, dx = \sup \mathcal{L}(f) = \inf \mathcal{U}(f).$$

We call a the **lower (integration) limit** and b the **upper (integration) limit**, and the function f the **integrand** of the integral $\int_a^b f \, dx$. If $f \geq 0$ is Riemann integrable, we interpret the number $\int_a^b f \, dx$ as the **area** of the set

$$\{(x, y) \in \mathbb{R}^2 \mid a \leq x \leq b, 0 \leq y \leq f(x)\}.$$

REMARK 6.15. — For now we only discuss Riemann integrability and the Riemann integral, so we will simply say “integrable” and “integral”. Note, however, that there is another fundamental theory, the **Lebesgue integral**, which we will not cover in this course.

PROPOSITION 6.16: RIEMANN INTEGRABILITY CONDITION

Let $f : [a, b] \rightarrow \mathbb{R}$ be bounded. Then f is Riemann integrable if and only if for every $\varepsilon > 0$ there exist step functions $\ell, u \in \mathcal{SF}$ such that

$$\ell \leq f \leq u \quad \text{and} \quad \int_a^b (u - \ell) \, dx < \varepsilon.$$

In this case,

$$\left| \int_a^b f \, dx - \int_a^b \ell \, dx \right| < \varepsilon, \quad \left| \int_a^b u \, dx - \int_a^b f \, dx \right| < \varepsilon.$$

Proof. By (6.4) applied with $A = \mathcal{L}(f)$ and $B = \mathcal{U}(f)$ we obtain

$$\begin{aligned} f \text{ is Riemann integrable} &\iff \sup \mathcal{L}(f) = \inf \mathcal{U}(f) \\ &\iff \forall \varepsilon > 0 \exists s \in \mathcal{L}(f), t \in \mathcal{U}(f) : t - s < \varepsilon \\ &\iff \forall \varepsilon > 0 \exists \ell, u \in \mathcal{SF} : \ell \leq f \leq u \text{ and } \int_a^b u \, dx - \int_a^b \ell \, dx < \varepsilon \\ &\iff \forall \varepsilon > 0 \exists \ell, u \in \mathcal{SF} : \ell \leq f \leq u \text{ and } \int_a^b (u - \ell) \, dx < \varepsilon, \end{aligned}$$

where we used Proposition 6.9 to deduce that

$$\int_a^b u \, dx - \int_a^b \ell \, dx = \int_a^b (u - \ell) \, dx.$$

Finally, the concluding inequalities follow from

$$\int_a^b \ell \, dx \leq \int_a^b f \, dx \leq \int_a^b u \, dx \quad \text{and} \quad \int_a^b (u - \ell) \, dx < \varepsilon.$$

□

It is useful to note that the Riemann integral extends the integral of step functions; in particular, we can speak of the Riemann integral of a step function, see Exercise 6.17.

EXERCISE 6.17. — Let $f : [a, b] \rightarrow \mathbb{R}$ be a step function. Show that f is Riemann integrable and that its Riemann integral equals its step-function integral.

EXERCISE 6.18. — Repeat the proof of Proposition 1.1 and show, in the language of this section, that $f : [0, 1] \rightarrow \mathbb{R}$, $x \mapsto x^2$, is Riemann integrable with $\int_0^1 x^2 \, dx = \frac{1}{3}$. Also,

$$\mathcal{L}(f) = \left(-\infty, \frac{1}{3}\right) \quad \text{and} \quad \mathcal{U}(f) = \left(\frac{1}{3}, \infty\right).$$

EXAMPLE 6.19. — Not all functions are Riemann integrable. Indeed, consider the function $f : [0, 1] \rightarrow \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1, & x \in \mathbb{Q}, \\ 0, & x \notin \mathbb{Q}. \end{cases}$$

We claim that f is not Riemann integrable.

Let $u \in \mathcal{SF}$ with $f \leq u$, and let $0 = x_0 < \cdots < x_n = 1$ be a decomposition such that u is constant c_k on (x_{k-1}, x_k) . Since \mathbb{Q} is dense in \mathbb{R} , there exists $x \in (x_{k-1}, x_k) \cap \mathbb{Q}$, hence $1 = f(x) \leq u(x) = c_k$, so $c_k \geq 1$. Therefore

$$\int_0^1 u(x) \, dx = \sum_{k=1}^n c_k (x_k - x_{k-1}) \geq \sum_{k=1}^n (x_k - x_{k-1}) = x_n - x_0 = 1.$$

Thus $\inf \mathcal{U}(f) \geq 1$, and taking $u \equiv 1$ gives $\inf \mathcal{U}(f) = 1$. A similar argument with lower sums shows that $\sup \mathcal{L}(f) = 0$. Hence f is not Riemann integrable.

6.2.2 Linearity and Monotonicity of the Riemann Integral

THEOREM 6.20: LINEARITY OF THE RIEMANN INTEGRAL

If $f, g : [a, b] \rightarrow \mathbb{R}$ are integrable and $\alpha, \beta \in \mathbb{R}$, then $\alpha f + \beta g$ is integrable and

$$\int_a^b (\alpha f + \beta g) \, dx = \alpha \int_a^b f \, dx + \beta \int_a^b g \, dx.$$

Proof. Given $\varepsilon > 0$, Proposition 6.16 yields step functions ℓ_1, ℓ_2, u_1, u_2 with

$$\ell_1 \leq f \leq u_1, \quad \ell_2 \leq g \leq u_2, \quad \int_a^b (u_1 - \ell_1) dx < \varepsilon, \quad \int_a^b (u_2 - \ell_2) dx < \varepsilon,$$

and

$$\left| \int_a^b f dx - \int_a^b \ell_1 dx \right| < \varepsilon, \quad \left| \int_a^b g dx - \int_a^b \ell_2 dx \right| < \varepsilon.$$

Assume first $\alpha, \beta \geq 0$. Then

$$\alpha \ell_1 + \beta \ell_2 \leq \alpha f + \beta g \leq \alpha u_1 + \beta u_2,$$

and

$$\int_a^b [(\alpha u_1 + \beta u_2) - (\alpha \ell_1 + \beta \ell_2)] dx = \alpha \int_a^b (u_1 - \ell_1) dx + \beta \int_a^b (u_2 - \ell_2) dx < (\alpha + \beta)\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this proves that $\alpha f + \beta g$ is integrable. Moreover, by the triangle inequality and Proposition 6.9,

$$\begin{aligned} \left| \int_a^b (\alpha f + \beta g) dx - \alpha \int_a^b f dx - \beta \int_a^b g dx \right| &\leq \left| \int_a^b (\alpha f + \beta g) dx - \int_a^b (\alpha \ell_1 + \beta \ell_2) dx \right| \\ &\quad + \underbrace{\left| \int_a^b (\alpha \ell_1 + \beta \ell_2) dx - \alpha \int_a^b \ell_1 dx - \beta \int_a^b \ell_2 dx \right|}_{=0} \\ &\quad + \alpha \left| \int_a^b \ell_1 dx - \int_a^b f dx \right| + \beta \left| \int_a^b \ell_2 dx - \int_a^b g dx \right| \\ &\leq (\alpha + \beta)\varepsilon + \alpha\varepsilon + \beta\varepsilon = 2(\alpha + \beta)\varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, the linearity identity follows.

The case when one of α, β is negative is analogous, but one needs to reverse the corresponding inequalities. For instance, if $\alpha \geq 0$ and $\beta < 0$, then

$$\alpha \ell_1 + \beta u_2 \leq \alpha f + \beta g \leq \alpha u_1 + \beta \ell_2,$$

and

$$\int_a^b [(\alpha u_1 + \beta \ell_2) - (\alpha \ell_1 + \beta u_2)] dx = \alpha \int_a^b (u_1 - \ell_1) dx + |\beta| \int_a^b (u_2 - \ell_2) dx < (\alpha + |\beta|)\varepsilon.$$

This implies again that $\alpha f + \beta g$ is integrable, and the linearity identity holds similarly. \square

EXERCISE 6.21. — Let $f : [a, b] \rightarrow \mathbb{R}$ be Riemann integrable, and let $f^* : [a, b] \rightarrow \mathbb{R}$ be obtained by changing the value of f at finitely many points. Show that f^* is Riemann integrable and has the same Riemann integral as f .

PROPOSITION 6.22: MONOTONICITY OF THE RIEMANN INTEGRAL

Let $f, g : [a, b] \rightarrow \mathbb{R}$ be integrable. If $f \leq g$, then $\int_a^b f \, dx \leq \int_a^b g \, dx$.

Proof. Since $f \leq g$, for any step function ℓ with $\ell \leq f$ we have $\ell \leq g$. This implies that $\mathcal{L}(f) \subseteq \mathcal{L}(g)$, therefore

$$\int_a^b f \, dx = \sup \mathcal{L}(f) \leq \sup \mathcal{L}(g) = \int_a^b g \, dx.$$

□

POSITIVE AND NEGATIVE PARTS

Given a function $f : D \rightarrow \mathbb{R}$, we define its **positive part** $f^+ : D \rightarrow \mathbb{R}$ and **negative part** $f^- : D \rightarrow \mathbb{R}$ by

$$f^+(x) = \max\{0, f(x)\}, \quad f^-(x) = -\min\{0, f(x)\}.$$

These satisfy

$$f = f^+ - f^-, \quad |f| = f^+ + f^-, \quad f^+ = \frac{|f| + f}{2}, \quad f^- = \frac{|f| - f}{2}.$$

Moreover, for any functions $f, g : D \rightarrow \mathbb{R}$,

$$f \leq g \implies f^+ \leq g^+, \quad f \leq g \implies f^- \geq g^-.$$

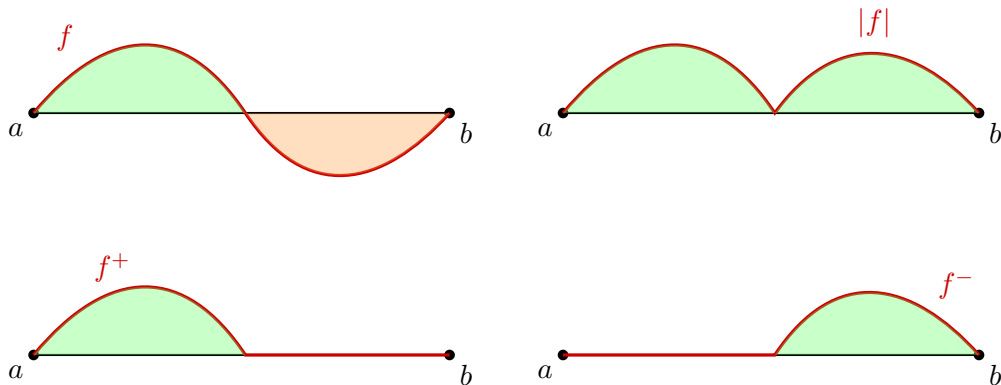


Figure 6.4: Top left: graph of f . Top right: graph of $|f|$. Bottom left: graph of f^+ . Bottom right: graph of f^- . The integral $\int_a^b f \, dx$ is the signed area (positive minus negative), whereas $\int_a^b |f| \, dx$ is the total area. Moreover, $\int_a^b f \, dx = \int_a^b f^+ \, dx - \int_a^b f^- \, dx$ and $\int_a^b |f| \, dx = \int_a^b f^+ \, dx + \int_a^b f^- \, dx$.

REMARK 6.23. — For any real numbers $z_1, z_2 \in \mathbb{R}$, one has

$$(z_1 - z_2)^+ \geq z_1^+ - z_2^+. \quad (6.5)$$

Indeed, since $z^+ \geq z$ and $z^+ \geq 0$ for all $z \in \mathbb{R}$, by applying these inequalities with $z = z_1 - z_2$ and $z = z_2$ we obtain

$$z_1 = (z_1 - z_2) + z_2 \leq (z_1 - z_2)^+ + z_2^+ \quad \text{and} \quad 0 \leq (z_1 - z_2)^+ + z_2^+.$$

Hence $(z_1 - z_2)^+ + z_2^+$ is greater than or equal to both z_1 and 0, and therefore

$$z_1^+ = \max\{z_1, 0\} \leq (z_1 - z_2)^+ + z_2^+,$$

which yields (6.5) after rearranging.

THEOREM 6.24: TRIANGLE INEQUALITY FOR THE RIEMANN INTEGRAL

Let $f : [a, b] \rightarrow \mathbb{R}$ be integrable. Then f^+ , f^- , and $|f|$ are integrable, and

$$\left| \int_a^b f \, dx \right| \leq \int_a^b |f| \, dx.$$

Proof. Fix $\varepsilon > 0$. Since f is integrable, there exist step functions $\ell \leq f \leq u$ with $\int_a^b (u - \ell) \, dx < \varepsilon$. Then ℓ^+ and u^+ are step functions with $\ell^+ \leq f^+ \leq u^+$.

Since $u - \ell \geq 0$, we have $(u - \ell) = (u - \ell)^+$. Moreover, applying (6.5) with $z_1 = u(x)$ and $z_2 = \ell(x)$, we obtain

$$(u(x) - \ell(x))^+ \geq u(x)^+ - \ell(x)^+ \quad \forall x \in [a, b].$$

Hence

$$\int_a^b (u^+ - \ell^+) \, dx \leq \int_a^b (u - \ell)^+ \, dx = \int_a^b (u - \ell) \, dx < \varepsilon,$$

so f^+ is integrable. By Theorem 6.20, also $f^- = f^+ - f$ and $|f| = 2f^+ - f$ are integrable. Finally,

$$\left| \int_a^b f \, dx \right| = \left| \int_a^b f^+ \, dx - \int_a^b f^- \, dx \right| \leq \int_a^b f^+ \, dx + \int_a^b f^- \, dx = \int_a^b |f| \, dx.$$

□

EXERCISE 6.25. — Let $a < b < c$. Show that $f : [a, c] \rightarrow \mathbb{R}$ is integrable if and only if $f|_{[a, b]}$ and $f|_{[b, c]}$ are integrable, and in that case

$$\int_a^c f \, dx = \int_a^b f|_{[a, b]} \, dx + \int_b^c f|_{[b, c]} \, dx.$$

EXERCISE 6.26. — Let $f : [a, b] \rightarrow \mathbb{R}$ be integrable and $\lambda > 0$. Define $g : [\lambda a, \lambda b] \rightarrow \mathbb{R}$ by $g(x) = f(\lambda^{-1}x)$. Show that g is integrable and

$$\lambda \int_a^b f \, dx = \int_{\lambda a}^{\lambda b} g \, dx.$$

EXERCISE 6.27. — Let $f : [a, b] \rightarrow \mathbb{R}$ be integrable. Show that the function $F : [a, b] \rightarrow \mathbb{R}$ given by

$$F(x) = \int_a^x f(t) \, dt$$

is continuous.

EXERCISE 6.28. — Let $f : [0, 1] \rightarrow \mathbb{R}$ be integrable and $\varepsilon > 0$. Show that there exists a continuous function $g : [0, 1] \rightarrow \mathbb{R}$ such that

$$\int_0^1 |f(x) - g(x)| \, dx < \varepsilon.$$

6.3 Integrability Theorems

6.3.1 Integrability of Monotone Functions

As before we work on a compact interval $[a, b] \subset \mathbb{R}$. Note that every monotone function $f : [a, b] \rightarrow \mathbb{R}$ is bounded; for instance, if f is increasing then $f(a)$ is a lower bound and $f(b)$ is an upper bound.

THEOREM 6.29: MONOTONE FUNCTIONS ARE INTEGRABLE

Every monotone function $f : [a, b] \rightarrow \mathbb{R}$ is Riemann integrable.

Proof. Without loss of generality, f is increasing (otherwise replace f by $-f$ and use Proposition 6.20). We want to apply Proposition 6.16: given $\varepsilon > 0$, we need to construct step functions $\ell, u \in \mathcal{SF}$ such that $\ell \leq f \leq u$ and $\int_a^b (u - \ell) dx < \varepsilon$.

Fix $n \in \mathbb{N}$ (to be chosen later) and the uniform partition

$$a = x_0 < x_1 < \dots < x_n = b, \quad x_k = a + \frac{k}{n}(b - a).$$

Define the step functions $\ell, u : [a, b] \rightarrow \mathbb{R}$ as

$$\begin{aligned} \ell(x) &= f(x_{k-1}) \quad \text{and} \quad u(x) = f(x_k) && \text{for } x \in (x_{k-1}, x_k), \quad k = 1, \dots, n, \\ \ell(x) &= u(x) = f(x) && \text{for } x \in \{x_0, \dots, x_n\}. \end{aligned}$$

Note that, since f is increasing, $\ell \leq f \leq u$. Moreover, for each k we have $u - \ell = f(x_k) - f(x_{k-1})$ on (x_{k-1}, x_k) . Recalling that $x_k - x_{k-1} = \frac{b-a}{n}$, this yields

$$\begin{aligned} \int_a^b (u - \ell) dx &= \sum_{k=1}^n (f(x_k) - f(x_{k-1})) (x_k - x_{k-1}) = \frac{b-a}{n} \sum_{k=1}^n (f(x_k) - f(x_{k-1})) \\ &= \frac{b-a}{n} (f(x_n) - f(x_0)) = \frac{b-a}{n} (f(b) - f(a)). \end{aligned}$$

Choosing $n \in \mathbb{N}$ so large that $\frac{b-a}{n} (f(b) - f(a)) < \varepsilon$, Proposition 6.16 implies that f is Riemann integrable. \square

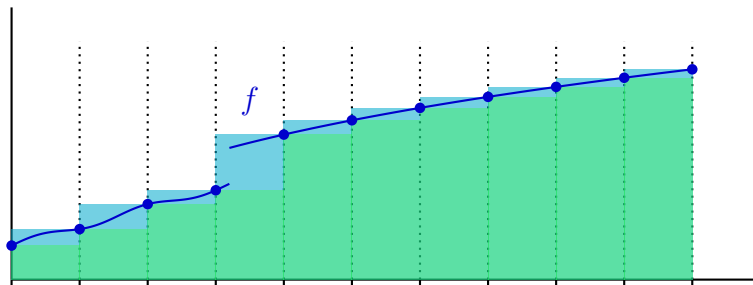


Figure 6.5: A monotone function, with upper and lower sums arbitrarily close.

Using the additivity property in Exercise 6.25, Theorem 6.29 extends to functions that are only piecewise monotone.

DEFINITION 6.30: PIECEWISE MONOTONE FUNCTIONS

A function $f : [a, b] \rightarrow \mathbb{R}$ is **piecewise monotone** if there exists a decomposition

$$a = x_0 < x_1 < \dots < x_n = b$$

such that $f|_{(x_{k-1}, x_k)}$ is monotone for every $k = 1, \dots, n$.

COROLLARY 6.31: PIECEWISE MONOTONE FUNCTIONS ARE INTEGRABLE

Every bounded piecewise monotone function $f : [a, b] \rightarrow \mathbb{R}$ is Riemann integrable.

Proof. Combine Theorem 6.29 with Exercises 6.21 and 6.25. □

6.3.2 Integrability of Continuous Functions

Using boundedness and uniform continuity on compact intervals (Theorems 3.31 and 3.39), we can prove that continuous functions are integrable.

THEOREM 6.32: CONTINUOUS FUNCTIONS ARE INTEGRABLE

Every continuous function $f : [a, b] \rightarrow \mathbb{R}$ is Riemann integrable.

Proof. Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous and fix $\varepsilon > 0$. By uniform continuity (Theorem 3.39), there exists $\delta > 0$ such that

$$|x - y| < \delta \implies |f(x) - f(y)| < \varepsilon \quad \forall x, y \in [a, b]. \quad (6.6)$$

Choose a partition $a = x_0 < \dots < x_n = b$ with $x_k - x_{k-1} < \delta$. For each k set

$$c_k = \min\{f(x) \mid x_{k-1} \leq x \leq x_k\}, \quad d_k = \max\{f(x) \mid x_{k-1} \leq x \leq x_k\},$$

which exist by Theorem 3.34, and let $y_k, z_k \in [x_{k-1}, x_k]$ satisfy $f(y_k) = c_k$ and $f(z_k) = d_k$. Then, since $|y_k - z_k| \leq x_k - x_{k-1} < \delta$, (6.6) yields $d_k - c_k < \varepsilon$.

Define now the step functions $\ell, u : [a, b] \rightarrow \mathbb{R}$ as

$$\begin{aligned} \ell(x) = c_k \quad \text{and} \quad u(x) = d_k & \quad \text{for } x \in (x_{k-1}, x_k), \quad k = 1, \dots, n, \\ \ell(x) = u(x) = f(x) & \quad \text{for } x \in \{x_0, \dots, x_n\}. \end{aligned}$$

Then $\ell \leq f \leq u$ and $u - \ell = d_k - c_k$ on (x_{k-1}, x_k) , hence

$$\int_a^b (u - \ell) dx = \sum_{k=1}^n (d_k - c_k)(x_k - x_{k-1}) < \varepsilon \sum_{k=1}^n (x_k - x_{k-1}) = \varepsilon(b - a).$$

Since $\varepsilon > 0$ is arbitrary, f is integrable. \square

By Exercise 6.25, the previous theorem extends to piecewise continuous functions.

DEFINITION 6.33: PIECEWISE CONTINUOUS FUNCTIONS

A function $f : [a, b] \rightarrow \mathbb{R}$ is **piecewise continuous** if there exists a decomposition

$$a = x_0 < x_1 < \dots < x_n = b$$

such that $f|_{(x_{k-1}, x_k)}$ is continuous for all k and both one-sided limits $\lim_{x \rightarrow x_{k-1}^+} f(x)$ and $\lim_{x \rightarrow x_k^-} f(x)$ exist. Equivalently, each $f|_{(x_{k-1}, x_k)}$ extends to a continuous function on $[x_{k-1}, x_k]$.

COROLLARY 6.34: PIECEWISE CONTINUOUS FUNCTIONS ARE INTEGRABLE

Every piecewise continuous function $f : [a, b] \rightarrow \mathbb{R}$ is Riemann integrable.

27

Proof. Apply Theorem 6.32 to the continuous extensions on each subinterval, and use Exercises 6.21 and 6.25. \square

Most “common” functions are piecewise continuous or piecewise monotone, hence integrable by Theorems 6.29 and 6.32. Note also that there are continuous functions that are not monotone on any open subinterval.

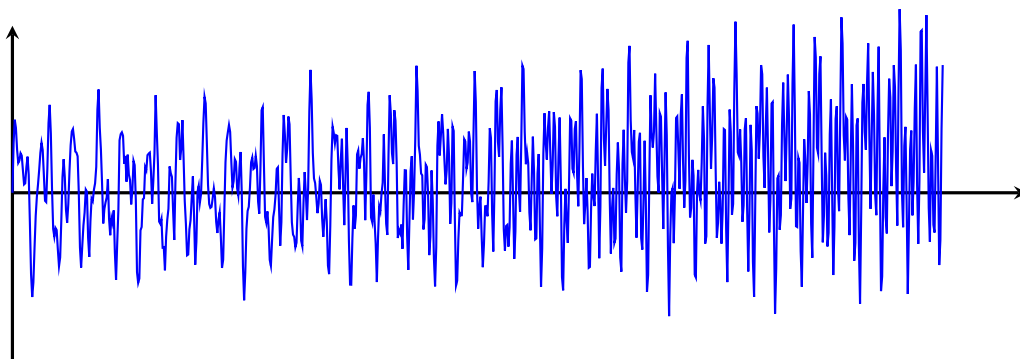


Figure 6.6: Approximate representation of a continuous function which is not monotone on any open subinterval.

EXERCISE 6.35. — Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous. Prove that

$$f \equiv 0 \iff \int_a^b |f(x)| dx = 0.$$

EXERCISE 6.36. — Let \mathcal{C} be the space of continuous functions on $[a, b]$ and define $I : \mathcal{C} \rightarrow \mathbb{R}$ by

$$I(f) = \int_a^b f \, dx.$$

Show that I is continuous with respect to the uniform norm; that is, for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\sup_{x \in [a, b]} |f(x) - g(x)| < \delta \implies |I(f) - I(g)| \leq \varepsilon.$$

Hint: Take $\delta = \frac{\varepsilon}{b-a}$.

6.3.3 Integration and Sequences of Functions

Let $(f_n)_{n=0}^\infty$, with $f_n : [a, b] \rightarrow \mathbb{R}$, be a sequence of integrable functions. Assume that f_n converges pointwise or uniformly to $f : [a, b] \rightarrow \mathbb{R}$. Is f integrable? And if so, does

$$\lim_{n \rightarrow \infty} \int_a^b f_n \, dx = \int_a^b f \, dx$$

hold?

In general, the pointwise limit of integrable functions need not be integrable. Also, as the following example shows, even when the pointwise limit f is integrable, one may have that $\lim_n \int f_n \neq \int f$.

EXAMPLE 6.37. — Let $D = [0, 1]$ and define $f_n : D \rightarrow \mathbb{R}$ by

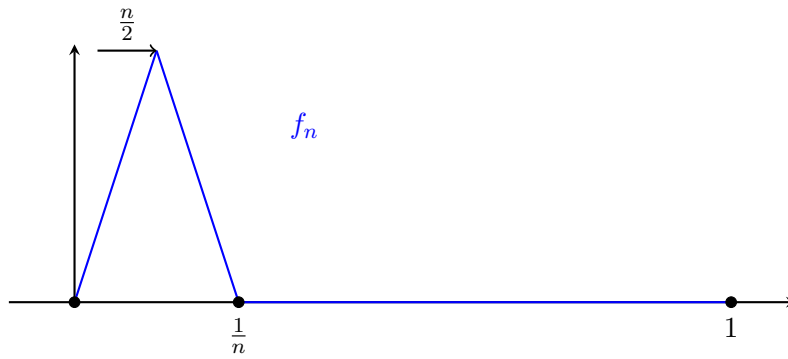
$$f_n(x) = \begin{cases} n^2 x & \text{if } 0 \leq x \leq \frac{1}{2n}, \\ n^2 \left(\frac{1}{n} - x\right) & \text{if } \frac{1}{2n} \leq x \leq \frac{1}{n}, \\ 0 & \text{if } \frac{1}{n} \leq x \leq 1. \end{cases}$$

Each f_n is continuous (hence integrable). Also, its graph is a triangle of base $\frac{1}{n}$ and height $\frac{n}{2}$, so

$$\int_0^1 f_n(x) \, dx = \frac{1}{2} \cdot \frac{1}{n} \cdot \frac{n}{2} = \frac{1}{4}.$$

Moreover, $f_n(0) = 0$ for all n , and for every $x > 0$ we have $f_n(x) = 0$ for all $n > 1/x$, hence $f_n(x) \rightarrow 0$. Thus f_n converges pointwise to the constant function $f = 0$, but

$$\int_0^1 f_n \, dx = \frac{1}{4} \neq 0 = \int_0^1 f \, dx.$$



On the other hand, as the next result shows, uniform convergence is sufficient for both integrability of the limit and interchange of limit and integral.

THEOREM 6.38: UNIFORM CONVERGENCE AND RIEMANN INTEGRALS COMMUTE

Let $(f_n)_{n=0}^\infty$, with $f_n : [a, b] \rightarrow \mathbb{R}$, be a sequence of integrable functions converging uniformly to $f : [a, b] \rightarrow \mathbb{R}$. Then f is integrable and

$$\int_a^b f \, dx = \lim_{n \rightarrow \infty} \int_a^b f_n \, dx. \quad (6.7)$$

28

Proof. Fix $\varepsilon > 0$. By uniform convergence, there exists N such that $|f_n - f| \leq \varepsilon$ on $[a, b]$ for all $n \geq N$.

Since f_N is integrable, there exist step functions ℓ, u with $\ell \leq f_N \leq u$ and $\int_a^b (u - \ell) \, dx < \varepsilon$. Set $\hat{\ell} = \ell - \varepsilon$ and $\hat{u} = u + \varepsilon$. Then $\hat{\ell}, \hat{u} \in \mathcal{SF}$. Also, since $|f_N - f| < \varepsilon$,

$$\hat{\ell} = \ell - \varepsilon \leq f_N - \varepsilon \leq f \leq f_N + \varepsilon \leq u + \varepsilon = \hat{u}$$

and (because $\hat{u} - \hat{\ell} = u - \ell + 2\varepsilon$)

$$\int_a^b (\hat{u} - \hat{\ell}) \, dx = \int_a^b (u - \ell) \, dx + 2\varepsilon(b - a) < \varepsilon + 2\varepsilon(b - a).$$

As $\varepsilon > 0$ is arbitrary, Proposition 6.16 yields that f is integrable.

Moreover, using monotonicity (Proposition 6.22) and the triangle inequality for the Riemann integral (Theorem 6.24),

$$\left| \int_a^b f \, dx - \int_a^b f_n \, dx \right| = \left| \int_a^b (f - f_n) \, dx \right| \leq \int_a^b |f - f_n| \, dx \leq \varepsilon(b - a) \quad \forall n \geq N,$$

proving (6.7). □

Chapter 7

The Derivative and the Riemann Integral

In this chapter we study the interplay between the Riemann integral (Chapter 6) and differentiation (Chapter 5). These connections are fundamental for the developments that follow.

7.1 The Fundamental Theorem of Calculus

Throughout this section we fix a compact interval $I \subseteq \mathbb{R}$ that is nonempty and contains more than one point. For brevity, we write *integrable* for *Riemann integrable*.

7.1.1 The Fundamental Theorem

DEFINITION 7.1: PRIMITIVE FUNCTION

Let $I \subset \mathbb{R}$ be an interval and $f : I \rightarrow \mathbb{R}$ a function. Any differentiable function $F : I \rightarrow \mathbb{R}$ such that $F' = f$ is called a **primitive** (or **antiderivative**) of f .

28

REMARK 7.2. — As the following exercise shows, a primitive may not always exist.

EXERCISE 7.3. — Show that there is no differentiable function $F : \mathbb{R} \rightarrow \mathbb{R}$ with $F'(x) = \operatorname{sgn}(x)$ for all $x \in \mathbb{R}$.

Hint: use Darboux's Theorem (see Exercise 5.36).

The next result is known as the **Fundamental Theorem of (Integral and Differential) Calculus**, going back to Leibniz, Newton, and Barrow

THEOREM 7.4: FUNDAMENTAL THEOREM OF CALCULUS

Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous. Then:

(i) For every $C \in \mathbb{R}$, the function $F : [a, b] \rightarrow \mathbb{R}$ defined by

$$F(x) = \int_a^x f(t) dt + C \quad (7.1)$$

is a primitive of f .

(ii) Every primitive $F : [a, b] \rightarrow \mathbb{R}$ of f has the form (7.1) for some constant C .

Proof. By Theorem 6.32, f is integrable.

Let F be defined as in (7.1). To prove (i), we fix $x_0 \in [a, b]$ and we want to show that $F'(x_0) = f(x_0)$. To this aim, fix $\varepsilon > 0$. By continuity, there exists $\delta > 0$ such that

$$z \in [a, b], \quad |z - x_0| < \delta \implies |f(z) - f(x_0)| < \varepsilon \quad (7.2)$$

Now, given $x \in (x_0, x_0 + \delta) \cap [a, b]$, it follows from Exercise 6.25 that

$$\begin{aligned} \left| \frac{F(x) - F(x_0)}{x - x_0} - f(x_0) \right| &= \left| \frac{1}{x - x_0} \left(\int_a^x f(t) dt - \int_a^{x_0} f(t) dt \right) - f(x_0) \right| \\ &= \left| \frac{1}{x - x_0} \int_{x_0}^x f(t) dt - f(x_0) \right|. \end{aligned}$$

Also,

$$f(x_0) = f(x_0) \frac{1}{x - x_0} \int_{x_0}^x dt = \frac{1}{x - x_0} \int_{x_0}^x f(x_0) dt.$$

Combining these two equations and using Theorem 6.24, we get

$$\begin{aligned} \left| \frac{F(x) - F(x_0)}{x - x_0} - f(x_0) \right| &= \left| \frac{1}{x - x_0} \int_{x_0}^x f(t) dt - \frac{1}{x - x_0} \int_{x_0}^x f(x_0) dt \right| \\ &= \left| \frac{1}{x - x_0} \int_{x_0}^x (f(t) - f(x_0)) dt \right| \\ &\leq \frac{1}{x - x_0} \int_{x_0}^x |f(t) - f(x_0)| dt. \end{aligned}$$

Note now that, in the last integral, $t \in [x_0, x] \subset [x_0, x_0 + \delta) \cap [a, b]$. Hence, it follows from (7.2) that $|f(t) - f(x_0)| < \varepsilon$, therefore

$$\left| \frac{F(x) - F(x_0)}{x - x_0} - f(x_0) \right| < \frac{1}{x - x_0} \int_{x_0}^x \varepsilon dt = \varepsilon.$$

Similarly, if $x \in (x_0 - \delta, x_0) \cap [a, b]$, then

$$\left| \frac{F(x) - F(x_0)}{x - x_0} - f(x_0) \right| = \left| \frac{1}{x_0 - x} \int_x^{x_0} (f(t) - f(x_0)) dt \right| \leq \frac{1}{x_0 - x} \int_x^{x_0} |f(t) - f(x_0)| dt < \varepsilon.$$

In summary, we proved that

$$x \in [a, b], \quad |x - x_0| < \delta \implies \left| \frac{F(x) - F(x_0)}{x - x_0} - f(x_0) \right| < \varepsilon,$$

therefore

$$F'(x_0) = \lim_{x \rightarrow x_0} \frac{F(x) - F(x_0)}{x - x_0} = f(x_0),$$

as desired.

We now prove (ii). Let F be a primitive of f . Then, since $(\int_a^x f(t) dt)' = f(x)$ (by (i)),

$$\left(F(x) - \int_a^x f(t) dt \right)' = F'(x) - f(x) = f(x) - f(x) = 0 \quad \forall x \in (a, b).$$

By Corollary 5.48, this implies that $F(x) - \int_a^x f(t) dt$ is constant on $[a, b]$, concluding the proof of (ii). \square

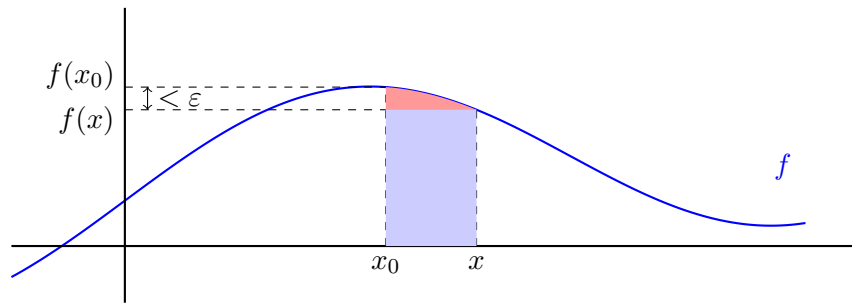


Figure 7.1: This figure illustrates the key estimate in the proof: $F(x) - F(x_0)$ equals the rectangle $f(x_0)(x - x_0)$ plus a red area with absolute value $< \varepsilon|x - x_0|$, hence $\frac{F(x) - F(x_0)}{x - x_0} \rightarrow f(x_0)$ as $x \rightarrow x_0$.

COROLLARY 7.5: INTEGRAL VS. DERIVATIVE

If $F : [a, b] \rightarrow \mathbb{R}$ is continuously differentiable, then for all $x \in [a, b]$,

$$F(x) = F(a) + \int_a^x F'(t) dt.$$

Proof. Since F is a primitive of F' , Theorem 7.4 yields $F(x) = \int_a^x F'(t) dt + C$. Evaluating at $x = a$ gives $C = F(a)$. \square

COROLLARY 7.6: RIEMANN INTEGRAL AND PRIMITIVES

If $f : [a, b] \rightarrow \mathbb{R}$ is continuous and F is a primitive of f , then

$$\int_a^b f(t) dt = F(b) - F(a).$$

Proof. Apply Corollary 7.5 with $F' = f$ and $x = b$. \square

EXAMPLE 7.7. — For all $a < b$:

1. $\int_a^b e^x dx = e^b - e^a,$
2. $\int_a^b \sin x dx = -\cos b + \cos a,$
3. $\int_a^b \cos x dx = \sin b - \sin a,$
4. $\int_a^b \sinh x dx = \cosh b - \cosh a,$
5. $\int_a^b \cosh x dx = \sinh b - \sinh a,$
6. $\int_a^b x^\alpha dx = \frac{b^{1+\alpha} - a^{1+\alpha}}{1+\alpha}$ for $1+\alpha \neq 0$ and $0 < a < b$,
7. $\int_a^b \frac{dx}{x} = \log b - \log a$ for $0 < a < b$.

EXERCISE 7.8. — Let $f : [a, b] \rightarrow \mathbb{R}$ be discontinuous at at most finitely many points. Show that $F(x) = \int_a^x f(t) dt$ is continuous on $[a, b]$, differentiable at every continuity point of f , and satisfies $F'(x) = f(x)$ at those points.

28

EXERCISE 7.9. — Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous. Show that there exists $\xi \in (a, b)$ with

$$\int_a^b f(x) dx = f(\xi) (b - a).$$

7.1.2 Integration by Parts and by Substitution

Given a function $h : [a, b] \rightarrow \mathbb{R}$, we use the notation $[h(x)]_a^b := h(b) - h(a)$.

THEOREM 7.10: INTEGRATION BY PARTS

If $f, g : [a, b] \rightarrow \mathbb{R}$ are continuously differentiable, then

$$\int_a^b f(x) g'(x) dx = [f(x)g(x)]_a^b - \int_a^b f'(x) g(x) dx.$$

Proof. By Proposition 5.10, $(fg)' = f'g + fg'$. Rearranging and integrating, thanks to Corollary 7.6 we get

$$\int_a^b fg' dx = \int_a^b (fg)' dx - \int_a^b f'g dx = [fg]_a^b - \int_a^b f'g dx.$$

□

As a convention, for any $h : [a, b] \rightarrow \mathbb{R}$,

$$\int_b^a h(x) dx = - \int_a^b h(x) dx. \quad (7.3)$$

THEOREM 7.11: INTEGRATION BY SUBSTITUTION, 1ST FORM

Let $I, J \subset \mathbb{R}$ be intervals, $f : I \rightarrow J$ be continuously differentiable, and $g : J \rightarrow \mathbb{R}$ be continuous. For any $[a, b] \subset I$,

$$\int_a^b g(f(x)) f'(x) dx = \int_{f(a)}^{f(b)} g(y) dy.$$

Proof. Fix $y_0 \in J$ and set $G(y) = \int_{y_0}^y g(t) dt$. Since $G' = g$, by the chain rule (see Theorem 5.14) we get $(G \circ f)' = G'(f) f' = g(f) f'$. Integrating this identity and using Corollary 7.6 yields

$$\begin{aligned} \int_a^b g(f(x)) f'(x) dx &= \int_a^b (G \circ f)'(x) dx = G(f(b)) - G(f(a)) \\ &= \int_{y_0}^{f(b)} g(t) dt - \int_{y_0}^{f(a)} g(t) dt = \int_{f(a)}^{f(b)} g(t) dt. \end{aligned}$$

□

28

Before stating the next result, we note the following: If $h : [a, b] \rightarrow \mathbb{R}$ is continuously differentiable with $h' \neq 0$, then h' has constant sign on $[a, b]$, so h is strictly monotone and invertible; h^{-1} is continuous by Theorem 3.28 and differentiable on $(h(a), h(b))$ by Theorem 5.20. Furthermore, since $(h^{-1})' = \frac{1}{h' \circ h^{-1}}$, also $(h^{-1})'$ is continuous.

THEOREM 7.12: INTEGRATION BY SUBSTITUTION, 2ND FORM

Let $I, J \subset \mathbb{R}$ be intervals, $f : I \rightarrow J$ be C^1 , and $g : J \rightarrow \mathbb{R}$ be continuous. Let $[a, b] \subset I$ and assume $f'(x) \neq 0$ on $[a, b]$. If $f^{-1} : [f(a), f(b)] \rightarrow \mathbb{R}$ denotes the inverse of $f|_{[a, b]}$, then

$$\int_a^b g(f(x)) dx = \int_{f(a)}^{f(b)} g(y) (f^{-1})'(y) dy.$$

Proof. In order to apply Theorem 7.11, we first observe that

$$\int_a^b g(f(x)) dx = \int_a^b \frac{g(f(x))}{f'(x)} f'(x) dx = \int_a^b \frac{g(f(x))}{f' \circ f^{-1}(f(x))} f'(x) dx.$$

So we can apply Theorem 7.11 with $\frac{g}{f' \circ f^{-1}}$ in place of g to get

$$\int_a^b g(f(x)) dx = \int_{f(a)}^{f(b)} \frac{g(y)}{f'(f^{-1}(y))} dy.$$

28 Since $\frac{1}{f' \circ f^{-1}} = (f^{-1})'$ (recall Theorem 5.20), the result follows. \square

7.1.3 Improper Integrals

A function $f : I \rightarrow \mathbb{R}$ is **locally integrable** if $f|_{[a,b]}$ is integrable for every compact $[a,b] \subset I$.

DEFINITION 7.13: IMPROPER INTEGRALS

Let $I \subseteq \mathbb{R}$ be a nonempty interval and $f : I \rightarrow \mathbb{R}$ be locally integrable. Set $c = \inf I \in \mathbb{R} \cup \{-\infty\}$ and $d = \sup I \in \mathbb{R} \cup \{\infty\}$, and fix $x_0 \in I$. We define the **improper integral** of f on I by

$$\int_c^d f(x) dx := \lim_{a \rightarrow c^+} \int_a^{x_0} f(x) dx + \lim_{b \rightarrow d^-} \int_{x_0}^b f(x) dx,$$

whenever both limits exist and the sum is well-defined (we do not allow the indeterminate form $\infty - \infty$). Here the first limit is taken over $a \in I$ with $c < a < x_0$ (interpreting $a \rightarrow -\infty$ if $c = -\infty$) and the second over $b \in I$ with $x_0 < b < d$ (interpreting $b \rightarrow +\infty$ if $d = \infty$). If the value is finite we say the integral *converges*; if it is $\pm\infty$ we say it *diverges to $\pm\infty$* ; otherwise, it *does not converge*. When defined, the value is independent of the choice of x_0 .

29

EXAMPLE 7.14. — Consider the integral of $f(x) = \frac{1}{x}$ over the interval $I = (0, 1)$. The function f is continuous, and hence locally integrable, on I . However, f is unbounded near 0, so it is not integrable on $[0, 1]$ in the usual Riemann sense, and the integral over $(0, 1)$ must be understood as an improper integral.

Since the only problem is at 0 and the integrand is bounded and continuous up to 1, we can write

$$\int_0^1 \frac{1}{x} dx = \lim_{a \rightarrow 0^+} \int_a^1 \frac{1}{x} dx.$$

Using the primitive $\log x$, we compute

$$\int_a^1 \frac{1}{x} dx = [\log x]_a^1 = \log 1 - \log a = -\log a,$$

therefore

$$\int_0^1 \frac{1}{x} dx = \lim_{a \rightarrow 0^+} (-\log a) = +\infty.$$

Thus the improper integral over $(0, 1)$ diverges to $+\infty$.

EXAMPLE 7.15. — Recalling that $\arctan'(x) = \frac{1}{1+x^2}$, we have

$$\int_0^\infty \frac{dx}{1+x^2} = \lim_{b \rightarrow \infty} [\arctan x]_0^b = \lim_{b \rightarrow \infty} \arctan b = \frac{\pi}{2}.$$

EXAMPLE 7.16. — For $\alpha \in \mathbb{R}$, we want to compute $\int_1^\infty x^{-\alpha} dx$. Note that

$$\int_1^b x^{-\alpha} dx = \left[\frac{x^{1-\alpha}}{1-\alpha} \right]_1^b = \frac{b^{1-\alpha} - 1}{1-\alpha} \quad \text{for } \alpha \neq 1, \quad \int_1^b \frac{dx}{x} = \log b \quad \text{for } \alpha = 1,$$

so letting $b \rightarrow \infty$ we obtain

$$\int_1^\infty x^{-\alpha} dx = \begin{cases} \frac{1}{\alpha-1}, & \alpha > 1, \\ \infty, & \alpha \leq 1. \end{cases}$$

In particular, the integral converges if and only if $\alpha > 1$.

As we now show, when $f \geq 0$ the improper integral over an interval always exists in the extended sense: it either converges to a finite value or diverges to $+\infty$.

LEMMA 7.17: IMPROPER INTEGRAL OF NONNEGATIVE FUNCTIONS

Let $I \subseteq \mathbb{R}$ be a nonempty interval with $c = \inf I$ and $d = \sup I$, and let $f : I \rightarrow [0, \infty)$ be locally integrable. Fix any $x_0 \in I$. Then the one-sided limits

$$L_- := \lim_{a \rightarrow c^+} \int_a^{x_0} f(x) dx, \quad L_+ := \lim_{b \rightarrow d^-} \int_{x_0}^b f(x) dx$$

exist in $[0, \infty]$, and

$$\int_c^d f(x) dx = L_- + L_+ = \sup_{c < \alpha < \beta < d} \int_\alpha^\beta f(x) dx \in [0, \infty].$$

In particular, the improper integral over I always exists in the extended sense and equals $+\infty$ whenever the supremum is $+\infty$.

Proof. Since $f \geq 0$, the maps $a \mapsto \int_a^{x_0} f$ (for $a < x_0$) and $b \mapsto \int_{x_0}^b f$ (for $b > x_0$) are monotone, hence the limits L_-, L_+ exist in $[0, \infty]$. Also, for any $c < a < x_0 < b < d$,

$$\int_a^b f = \int_a^{x_0} f + \int_{x_0}^b f.$$

Taking suprema gives

$$\sup_{c < \alpha < \beta < d} \int_\alpha^\beta f = \sup_{a < x_0} \int_a^{x_0} f + \sup_{b > x_0} \int_{x_0}^b f = L_- + L_+,$$

which equals the definition of $\int_c^d f$ above. \square

EXAMPLE 7.18. — Consider

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \int_{-\infty}^{-1} e^{-x^2} dx + \int_{-1}^1 e^{-x^2} dx + \int_1^{\infty} e^{-x^2} dx.$$

Since $x^2 \geq x$ for $x \geq 1$, we have $e^{-x^2} \leq e^{-x}$ on $[1, \infty)$, hence

$$\int_1^{\infty} e^{-x^2} dx \leq \int_1^{\infty} e^{-x} dx = \lim_{b \rightarrow \infty} [-e^{-x}]_1^b = e^{-1} < \infty.$$

By symmetry, $\int_{-\infty}^{-1} e^{-x^2} dx = \int_1^{\infty} e^{-x^2} dx < \infty$. Finally, $\int_{-1}^1 e^{-x^2} dx \leq \int_{-1}^1 1 dx = 2$. Thus the improper integral converges.

THEOREM 7.19: INTEGRAL TEST FOR SERIES

Let $f : [0, \infty) \rightarrow [0, \infty)$ be monotone decreasing. Then, for every $N \in \mathbb{N}$,

$$\sum_{n=1}^{N+1} f(n) \leq \int_0^{N+1} f(x) dx \leq \sum_{n=0}^N f(n).$$

In particular,

$$\sum_{n=1}^{\infty} f(n) \text{ converges} \iff \int_0^{\infty} f(x) dx \text{ converges}.$$

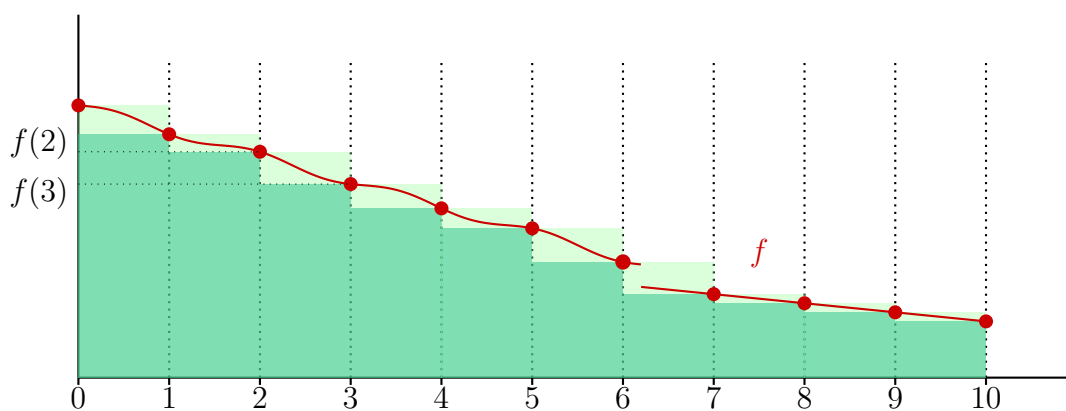
Proof. By monotonicity, f is locally integrable. Define step functions on $[0, \infty)$ by

$$u(x) = f(\lfloor x \rfloor), \quad \ell(x) = f(\lceil x \rceil),$$

where $\lfloor x \rfloor$ is the rounding function (i.e. the largest integer $\leq x$), while $\lceil x \rceil$ denotes the smallest integer $\geq x$. Then $\ell \leq f \leq u$, and for $N \geq 1$,

$$\sum_{n=1}^{N+1} f(n) = \int_0^{N+1} \ell(x) dx \leq \int_0^{N+1} f(x) dx \leq \int_0^{N+1} u(x) dx = \sum_{n=0}^N f(n).$$

The result follows by taking the limit as $N \rightarrow \infty$. \square



EXAMPLE 7.20. — The harmonic series can be written as $\{f(n)\}_{n=0}^{\infty}$ with $f(x) = \frac{1}{1+x}$, hence it diverges since

$$\int_0^{\infty} \frac{dx}{1+x} = \infty.$$

Moreover, Theorem 7.19 allows us to estimate its partial sums. Let $H_N = \sum_{n=1}^N \frac{1}{n}$ be the N -th partial sum. Applying the integral test with $f(x) = \frac{1}{1+x}$, we obtain

$$\sum_{n=1}^{N+1} \frac{1}{1+n} \leq \int_0^{N+1} \frac{dx}{1+x} \leq \sum_{n=0}^N \frac{1}{1+n}.$$

The left-hand side is

$$\sum_{n=1}^{N+1} \frac{1}{1+n} = \sum_{n=2}^{N+2} \frac{1}{n} = H_{N+2} - 1,$$

the integral is

$$\int_0^{N+1} \frac{dx}{1+x} = [\log(1+x)]_0^{N+1} = \log(N+2),$$

and the right-hand side is

$$\sum_{n=0}^N \frac{1}{1+n} = \sum_{n=1}^{N+1} \frac{1}{n} = H_{N+1}.$$

Thus, for every $N \geq 1$,

$$H_{N+2} - 1 \leq \log(N+2) \leq H_{N+1}.$$

In particular, $H_N \rightarrow \infty$ as $N \rightarrow \infty$, and these inequalities show that the harmonic series grows like $\log N$.

In contrast, $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges because

$$\int_0^{\infty} \frac{dx}{(1+x)^2} < \infty.$$

7.2 Integration and Differentiation of Power Series

From Example 7.7(6), for $n \geq 0$ we have $\int_0^x t^n dt = \frac{x^{n+1}}{n+1}$, i.e. $\frac{x^{n+1}}{n+1}$ is a primitive of x^n . Also, by Corollary 5.12, $(x^n)' = nx^{n-1}$. These formulas allow us to integrate and differentiate polynomials. We now address integrating and differentiating power series.

We recall the limit

$$\lim_{n \rightarrow \infty} \sqrt[n]{\frac{1}{n}} = 1 \quad (7.4)$$

see Exercise 3.53. Also, we shall use the following fact.

REMARK 7.21. — Let $(a_n)_{n=0}^\infty$ be a sequence of nonnegative numbers with

$$L = \limsup_{n \rightarrow \infty} a_n < \infty,$$

and let $(b_n)_{n=0}^\infty$ and $(\gamma_n)_{n=0}^\infty$ satisfy $b_n \rightarrow 1$ and $\gamma_n \rightarrow 1$. Then

$$\limsup_{n \rightarrow \infty} a_n^{\gamma_n} b_n = L.$$

29

In other words, multiplying by a factor that tends to 1, or raising to an exponent that tends to 1, does not change the value of the lim sup.

EXERCISE 7.22. — Prove Remark 7.21.

Hint: Fix $\varepsilon \in (0, 1)$. Since $b_n, \gamma_n \rightarrow 1$ as $n \rightarrow \infty$, there exists N such that

$$1 - \varepsilon \leq b_n, \gamma_n \leq 1 + \varepsilon \quad \text{for all } n \geq N.$$

Observe now that, given $a \geq 0$, for every $s \in [1 - \varepsilon, 1 + \varepsilon]$ it holds

$$\min \{a^{1-\varepsilon}, a^{1+\varepsilon}\} \leq a^s \leq \max \{a^{1-\varepsilon}, a^{1+\varepsilon}\},$$

where the minimum/maximum depends on whether $a \leq 1$ or $a \geq 1$. Apply this with $a = a_n$ and $s = \gamma_n$, and combine with $1 - \varepsilon \leq b_n \leq 1 + \varepsilon$, to deduce

$$(1 - \varepsilon) \min \{a_n^{1-\varepsilon}, a_n^{1+\varepsilon}\} \leq a_n^{\gamma_n} b_n \leq (1 + \varepsilon) \max \{a_n^{1-\varepsilon}, a_n^{1+\varepsilon}\} \quad \text{for all } n \geq N.$$

Then take the lim sup in these inequalities and finally let $\varepsilon \rightarrow 0$.

THEOREM 7.23: INTEGRATION OF POWER SERIES

Let $f(x) = \sum_{n=0}^{\infty} a_n x^n$ have radius of convergence $R > 0$. Then

$$F(x) = \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1}$$

has the same radius of convergence R and is a primitive of f on $(-R, R)$.

Proof. Set $\rho = \limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|}$, so that $R = \rho^{-1}$. Define $c_0 = 0$ and $c_n = \frac{a_{n-1}}{n}$ for $n \geq 1$, so $F(x) = \sum_{n=0}^{\infty} c_n x^n$. Noticing that

$$\sqrt[n]{|c_n|} = \sqrt[n]{\frac{1}{n}} \left(\sqrt[n-1]{|a_{n-1}|} \right)^{\frac{n-1}{n}},$$

it follows from (7.4) and Remark 7.21 (applied with $b_n = \sqrt[n]{\frac{1}{n}}$ and $\gamma_n = \frac{n-1}{n}$) that

$$\limsup_{n \rightarrow \infty} \sqrt[n]{|c_n|} = \limsup_{n \rightarrow \infty} \sqrt[n-1]{|a_{n-1}|} = \rho,$$

hence also F has radius of convergence R .

We now want to prove that $F' = f$. Fix $[-r, r] \subset (-R, R)$ and define the polynomials $f_n(x) = \sum_{k=0}^n a_k x^k$. Then

$$\int_0^x f_n(t) dt = \sum_{k=0}^n \frac{a_k}{k+1} x^{k+1}.$$

By Theorem 4.42, the sequence of functions $(f_n)_{n=0}^{\infty}$ converge uniformly to f on $[-r, r]$, so Theorem 6.38 yields

$$\int_0^x f(t) dt = \lim_{n \rightarrow \infty} \int_0^x f_n(t) dt \quad \forall x \in [-r, r].$$

On the other hand, again by Theorem 4.42,

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n \frac{a_k}{k+1} x^{k+1} = F(x) \quad \forall x \in [-r, r].$$

This proves that $F(x) = \int_0^x f(t) dt$ on $[-r, r]$, so Theorem 7.4 implies that $F'(x) = f(x)$ on $[-r, r]$. Since $[-r, r] \subset (-R, R)$ is arbitrary, we proved that $F' = f$ on $(-R, R)$. \square

COROLLARY 7.24: DIFFERENTIATION OF POWER SERIES

Let $f(x) = \sum_{n=0}^{\infty} a_n x^n$ have radius of convergence $R > 0$. Then f is differentiable on $(-R, R)$ with

$$f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1} \quad \forall x \in (-R, R),$$

and the series on the right has radius of convergence R .

Proof. Let $c_n = (n+1)a_{n+1}$ and $g(x) = \sum_{n=1}^{\infty} na_n x^{n-1} = \sum_{n=0}^{\infty} c_n x^n$. Let \bar{R} be the radius of g . Then Theorem 7.23 implies that the power series

$$G(x) = \sum_{n=0}^{\infty} \frac{c_n}{n+1} x^{n+1}$$

has radius \bar{R} and is a primitive of g . But

$$G(x) = \sum_{n=0}^{\infty} \frac{(n+1)a_{n+1}}{n+1} x^{n+1} = \sum_{n=0}^{\infty} a_{n+1} x^{n+1} = \sum_{n=1}^{\infty} a_n x^n = f(x) - a_0.$$

This implies that G and f have the same radius of convergence (so $\bar{R} = R$) and that $g = G' = (f - a_0)' = f'$. \square

EXERCISE 7.25. — Let $f(x) = \sum_{n=0}^{\infty} a_n x^n$ have radius of convergence $R > 0$. Show that f is C^∞ on $(-R, R)$ and express $f^{(n)}$ as a power series for each $n \in \mathbb{N}$.

EXERCISE 7.26. — Let $f(x) = \sum_{n=0}^{\infty} a_n x^n$ and $g(x) = \sum_{n=0}^{\infty} b_n x^n$ with radii $R_f, R_g > 0$, and set $R = \min\{R_f, R_g\}$. Prove that if $f(x) = g(x)$ for all $x \in (-R, R)$, then $a_n = b_n$ for all n . In particular, $R_f = R_g$.

29

EXERCISE 7.27. — Let $\alpha \in \mathbb{R}$. The goal of this exercise is to show that

$$(1+x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n \quad \forall x \in (-1, 1), \quad (7.5)$$

where $\binom{\alpha}{n}$ is defined as $\binom{\alpha}{n} = \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}$.

(a) Show that, for $\alpha \notin \mathbb{N}$, the series $f(x) = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n$ has radius of convergence 1.

(b) Compute f' and show that

$$f'(x) = \alpha \frac{f(x)}{1+x} \quad \forall x \in (-1, 1). \quad (7.6)$$

(c) Let $g(x) = (1+x)^\alpha$ and use (7.6) to prove $(\frac{f}{g})' = 0$ on $(-1, 1)$. Conclude the validity of (7.5) from the fact that $f(0) = g(0) = 1$.

EXAMPLE 7.28. — We have already seen in Example 4.23 that, as a consequence of the Leibniz criterion (Proposition 4.22), the alternating harmonic series converges. However, we

were unable to determine the value of the series. We now show that

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = \log(2).$$

Using the geometric series and the fundamental theorem of calculus, for $x \in (-1, 1)$ we have

$$(\log(1+x))' = \frac{1}{1+x} = \frac{1}{1-(-x)} = \sum_{n=0}^{\infty} (-1)^n x^n,$$

hence, since $\log(1) = 0$, Theorem 7.23 yields

$$\log(1+x) = \int_0^x \frac{1}{1+t} dt = \sum_{n=0}^{\infty} \frac{(-1)^n}{n+1} x^{n+1} = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} x^k.$$

For $x \in [0, 1]$, the terms $a_k = \frac{x^k}{k}$ are nonnegative, decreasing, and converge to 0, so by Proposition 4.22,

$$\sum_{k=1}^{2n} \frac{(-1)^{k+1}}{k} x^k \leq \log(1+x) \leq \sum_{k=1}^{2n+1} \frac{(-1)^{k+1}}{k} x^k \quad \forall n \in \mathbb{N}.$$

Letting $x \rightarrow 1^-$ yields that, for each $n \in \mathbb{N}$,

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

Now, using again Proposition 4.22, we can let $n \rightarrow \infty$ to get the desired result.

EXAMPLE 7.29. — Similarly to what we did in the previous exercise, we note that

$$\arctan'(x) = \frac{1}{1+x^2} = \sum_{k=0}^{\infty} (-1)^k x^{2k} \quad \forall x \in (-1, 1),$$

hence

$$\arctan x = \int_0^x \frac{1}{1+t^2} dt = \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} x^{2k+1}.$$

Using the Leibniz criterion as before yields

$$\sum_{k=0}^{2n+1} \frac{(-1)^k}{2k+1} x^{2k+1} \leq \arctan(x) \leq \sum_{k=0}^{2n} \frac{(-1)^k}{2k+1} x^{2k+1} \quad \forall n \in \mathbb{N}.$$

Letting first $x \rightarrow 1^-$ gives

$$\sum_{k=0}^{2n+1} \frac{(-1)^k}{2k+1} \leq \arctan(1) = \frac{\pi}{4} \leq \sum_{k=0}^{2n} \frac{(-1)^k}{2k+1} \quad \forall n \in \mathbb{N},$$

and letting $n \rightarrow \infty$ gives

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

Sometimes the primitive of a function cannot be expressed in terms of the “standard” elementary functions.

EXAMPLE 7.30 (Integral Sine). — The **integral sine** is the primitive function $\text{Si} : \mathbb{R} \rightarrow \mathbb{R}$ of the continuous function

$$x \in \mathbb{R} \mapsto \begin{cases} \frac{\sin(x)}{x} & \text{if } x \neq 0 \\ 1 & \text{if } x = 0 \end{cases}$$

with the normalisation $\text{Si}(0) = 0$, that is $\text{Si}(x) = \int_0^x \frac{\sin(t)}{t} dt$. Thanks to Theorem 7.23, the function Si can be expressed as a power series:

$$\text{Si}(x) = \int_0^x \frac{\sin(t)}{t} dt = \int_0^x \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} t^{2n} dt = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!(2n+1)} x^{2n+1}$$

for all $x \in \mathbb{R}$.

7.3 Integration Methods

Let $I \subseteq \mathbb{R}$ be an interval, and $f : I \rightarrow \mathbb{R}$ a function. The notation

$$\int f(x) dx = F(x) + C$$

means that F is a primitive function (antiderivative) of f . In the expression $F(x) + C$, C is read as an indefinite constant, usually called **integration constant**. Since the domain I of f is an interval, two primitive functions of f differ by a constant, which makes the notation meaningful. One calls $F(x) + C$ the **indefinite integral** of f . Indefinite integrals of special functions can be found in tables or by means of computer algebra systems. In this section, we show general methods to determine indefinite integrals.

Throughout this section, $I \subseteq \mathbb{R}$ denotes a non-empty interval that is not a single point. Also, all functions in this section are real-valued functions with domain I that are integrable on any compact interval $[a, b] \subseteq I$.

7.3.1 Integration by Parts and by Substitution in Leibniz Notation

In the computation of indefinite integrals, it is convenient to use Leibniz notation. This notation allows us to reformulate, in a natural formalism, both integration by parts and by substitution (see Section 7.1.2). We recall that the derivative of a function h is denoted by h' or by $\frac{dh}{dx}$. In this section, the second notation (called Leibniz notation) will be useful.

Integration by Parts. Let f and g be functions with primitives F and G , respectively. Recall that, from the product rule for the derivative in Proposition 5.10, it follows that $(FG)' = fG + Fg$. This implies the **integration by parts** formula

$$\int F(x) g(x) dx = F(x)G(x) - \int f(x) G(x) dx + C. \quad (7.7)$$

In Leibniz notation, $f = \frac{dF}{dx}$ and $g = \frac{dG}{dx}$. This leads to the notation $f dx = dF$ and $g dx = dG$, and integration by parts is sometimes written as

$$\int F dG = FG - \int G dF + C,$$

which should be understood as a short form of formula (7.7).

Integration by Substitution. Let J be an interval and let $f : I \rightarrow J$ be a differentiable function. If $G : J \rightarrow \mathbb{R}$ is a primitive of g then, by the chain rule in Theorem 5.14, $[G(f(x))]' = g(f(x)) f'(x)$ for all $x \in I$. From this it follows that

$$\int g(f(x)) f'(x) dx = G(f(x)) + C.$$

Since $G(u) = \int g(u) du + C$, we obtain

$$\int g(f(x)) f'(x) dx = \int g(u) du + C \quad (7.8)$$

where we used the change of variables $u = f(x)$. The substitution rule is also called **change of variable**, as one has replaced the variable u in $\int g(u) du$ by $u = f(x)$. In Leibniz notation this is very natural: if $u = f(x)$ then $du = f'(x) dx$, and (7.8) follows.

We also recall the second form of the substitution rule: if $f' \neq 0$ we can set $x = f^{-1}(u)$ so that $\frac{dx}{du} = (f^{-1})'(u)$, and obtain

$$\int g(f(x)) dx = \int g(u) \frac{dx}{du} du + C, \quad (7.9)$$

see Section 7.1.2.

7.3.2 Integration by Parts: Examples

EXAMPLE 7.31. — We want to calculate the indefinite integral $\int x e^x dx$. Since $e^x = (e^x)'$, using (7.7) we get

$$\int x e^x dx = \int x (e^x)' dx = x e^x - \int x' \cdot e^x dx + C = x e^x - \int e^x dx + C.$$

30

Since $\int e^x dx = e^x + C$, we conclude that

$$\int x e^x dx = x e^x - e^x + C.$$

We note that it is sufficient to use only one integration constant C in such calculations, since several such constants can be combined into one.

EXAMPLE 7.32. — We calculate the integral $\int \log(x) dx$:

$$\begin{aligned} \int \log(x) dx &= \int \log(x) \cdot 1 dx = \int \log(x) \cdot x' dx \\ &= \log(x) \cdot x - \int \log'(x) x dx + C \\ &= \log(x) \cdot x - \int \frac{1}{x} x dx + C \\ &= \log(x) \cdot x - \int 1 dx + C = x \log(x) - x + C. \end{aligned}$$

Suggestion: To ensure that the final result is correct, differentiate the result and check if you get the original function. For instance, in this case, one can easily check that

$$(x \log(x) - x + C)' = \log(x).$$

EXERCISE 7.33. — Give a recursive formula for calculating the indefinite integrals

$$\int x^n e^x dx, \quad \int x^n \sin(x) dx, \quad \int x^n \cos(x) dx$$

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

EXERCISE 7.34. — Calculate

$$\int x^s \log(x) dx, \quad \int e^{ax} \sin(bx) dx$$

for all $s, a, b \in \mathbb{R}$. Note that the case $s = -1$ needs to be treated separately, in analogy with Example 7.7(6)-(7).

7.3.3 Integration by Substitution: Examples

EXAMPLE 7.35. — We want to compute $\int \frac{x}{1+x^2} dx$. Let $u = f(x) = 1 + x^2$, so that $du = f'(x) dx = 2x dx$. Then we find

$$\int \frac{x}{1+x^2} dx = \frac{1}{2} \int \frac{1}{1+x^2} (2x dx) = \frac{1}{2} \int \frac{1}{u} du = \frac{1}{2} \log |u| = \frac{1}{2} \log(1+x^2) + C.$$

30

EXAMPLE 7.36. — Given $r > 0$, we want to compute the indefinite integral $\int \sqrt{r^2 - x^2} dx$. Due to the trigonometric identity $\sqrt{r^2 - r^2 \sin^2(\theta)} = r \cos(\theta)$ it is convenient to use the change of variable $x = r \sin(\theta)$, $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$. With this choice we have $dx = r \cos(\theta) d\theta$, therefore

$$\int \sqrt{r^2 - x^2} dx = r \int \sqrt{r^2 - r^2 \sin^2(\theta)} \cos(\theta) d\theta = r^2 \int \cos^2(\theta) d\theta.$$

To compute $\int \cos^2(\theta) d\theta$ we use integration by parts as follows:

$$\begin{aligned} \int \cos^2(\theta) d\theta &= \int \cos(\theta) \sin'(\theta) d\theta \\ &= \cos(\theta) \sin(\theta) - \int \cos'(\theta) \sin(\theta) d\theta + C \\ &= \cos(\theta) \sin(\theta) + \int \sin^2(\theta) d\theta + C. \end{aligned}$$

Since $\sin^2(\theta) = 1 - \cos^2(\theta)$, we get

$$\begin{aligned} \int \cos^2(\theta) d\theta &= \cos(\theta) \sin(\theta) + \int 1 d\theta - \int \cos^2(\theta) d\theta + C \\ &= \cos(\theta) \sin(\theta) + \theta - \int \cos^2(\theta) d\theta + C, \end{aligned}$$

therefore

$$2 \int \cos^2(\theta) d\theta = \cos(\theta) \sin(\theta) + \theta + C \quad \implies \quad \int \cos^2(\theta) d\theta = \frac{1}{2} (\cos(\theta) \sin(\theta) + \theta) + C.$$

(Note that, since $C \in \mathbb{R}$ is arbitrary, in the last formula we still write C in place of $\frac{C}{2}$.) This proves that

$$\int \sqrt{r^2 - x^2} dx = r^2 \int \cos^2(\theta) d\theta = \frac{r^2}{2} (\sin(\theta) \cos(\theta) + \theta) + C.$$

(Again, we write C in place of Cr^2 .) Recalling that $x = r \sin(\theta)$ with $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$, it follows that $\theta = \arcsin(\frac{x}{r})$ and $\cos(\theta) = \sqrt{1 - \frac{x^2}{r^2}}$, therefore

$$\int \sqrt{r^2 - x^2} dx = \frac{1}{2} x \sqrt{r^2 - x^2} + \frac{r^2}{2} \arcsin\left(\frac{x}{r}\right) + C.$$

Substitutions like in Example 7.39 are called **trigonometric substitutions**. We will not always argue carefully in these calculations and will rather trust the Leibniz notation, but recall that, to apply (7.9), there must be invertibility of the function when we express the old variable by the new variable.

For the following list of trigonometric substitutions, let $n \in \mathbb{Z}$.

- In expressions of the form $(a^2 - x^2)^{\frac{n}{2}}$ for $a > 0$, as already seen in the example above, one considers the substitution $x = a \sin(\theta)$ with $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$, giving $dx = a \cos(\theta) d\theta$ and $(a^2 - x^2)^{\frac{1}{2}} = a \cos(\theta)$.
- In expressions of the form $(a^2 + x^2)^{\frac{n}{2}}$ for $a > 0$, the substitution $x = a \tan(\theta)$ with $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$ yields $dx = \frac{a}{\cos^2(\theta)} d\theta$ and $(a^2 + x^2)^{\frac{n}{2}} = \left(\frac{a}{\cos(\theta)}\right)^n$.
- Although this is not a trigonometric substitution, we still note the following: For the expression $x(a^2 - x^2)^{\frac{n}{2}}$ or the expression $x(a^2 + x^2)^{\frac{n}{2}}$, the substitutions $u = a^2 - x^2$ and $u = a^2 + x^2$, respectively, allow us to compute the indefinite integrals.

EXAMPLE 7.37. — (i) Given $a > 0$, using the substitution $x = a \tan(\theta)$, recalling that $(a^2 + x^2)^{\frac{1}{2}} = \frac{a}{\cos(\theta)}$ and $dx = \frac{a}{\cos^2(\theta)} d\theta$ (recall that $\tan'(\theta) = \frac{1}{\cos^2(\theta)}$), we get

$$\begin{aligned} \int \frac{1}{(a^2 + x^2)^{\frac{3}{2}}} dx &= \int \frac{\cos^3(\theta)}{a^3} \frac{a}{\cos^2(\theta)} d\theta = \frac{1}{a^2} \int \cos(\theta) d\theta = \frac{1}{a^2} \sin(\theta) + C \\ &= \frac{1}{a^2} \tan(\theta) \cos(\theta) + C = \frac{x}{a^2 \sqrt{a^2 + x^2}} + C. \end{aligned}$$

(ii) Choosing $u = 1 - x^2$ (so that $du = -2x dx$), we have

$$\int x \sqrt{1 - x^2} dx = -\frac{1}{2} \int u^{1/2} du = -\frac{1}{3} u^{3/2} + C = -\frac{1}{3} (1 - x^2)^{3/2} + C.$$

Certain indefinite integrals can be computed with hyperbolic substitutions. For instance, for expressions of the form $(x^2 - a^2)^{\frac{n}{2}}$ with $a \in \mathbb{R}$, the substitution $x = a \cosh(u)$ yields $dx = a \sinh(u) du$ and $(x^2 - a^2)^{\frac{1}{2}} = a \sinh(u)$.

EXAMPLE 7.38. — Using the substitution $x = \cosh(u)$ (so $dx = \sinh(u) du$), we compute

$$\int \sqrt{x^2 - 1} dx = \int \sqrt{\cosh^2(u) - 1} \sinh(u) du = \int \sinh^2(u) du.$$

In analogy to the argument used in Example 7.36, we compute $\int \sinh^2(u) du$ as follows:

$$\begin{aligned} \int \sinh^2(u) du &= \cosh(u) \sinh(u) - \int \cosh^2(u) du + C \\ &= \cosh(u) \sinh(u) - \int (1 + \sinh^2(u)) du + C \\ &= \cosh(u) \sinh(u) - u - \int \sinh^2(u) du + C, \end{aligned}$$

This yields

$$2 \int \sinh^2(u) du = \cosh(u) \sinh(u) - u + C \quad \implies \quad \int \sinh^2(u) du = \frac{\cosh(u) \sinh(u) - u}{2} + C,$$

hence

$$\int \sqrt{x^2 - 1} dx = \frac{\cosh(u) \sinh(u) - u}{2} + C = \frac{x\sqrt{x^2 - 1} - \operatorname{arcosh}(x)}{2} + C.$$

Another method that we would like to mention briefly here is the so-called **half-angle method** (or **Weierstrass substitution**). This is useful for the integral of expressions like $\frac{1}{\sin(x)}$ or $\frac{\cos^2(x) + \cos(x) + \sin(x)}{1 + \sin(x)}$, see also Remark 7.42 below. We show this method in detail in the next example.

EXAMPLE 7.39. — We want to compute $\int \frac{1}{\sin(x)} dx$, and we consider the change of variable $u = \tan\left(\frac{x}{2}\right)$. We can note that, by the doubling angle formulas for sine and cosine (see (4.14)), it follows that

$$\sin(x) = 2 \sin\left(\frac{x}{2}\right) \cos\left(\frac{x}{2}\right) = \frac{2 \frac{\sin(\frac{x}{2})}{\cos(\frac{x}{2})}}{1 + \frac{\sin^2(\frac{x}{2})}{\cos^2(\frac{x}{2})}} = \frac{2 \tan\left(\frac{x}{2}\right)}{1 + \tan^2\left(\frac{x}{2}\right)} = \frac{2u}{1 + u^2}$$

and analogously $\cos(x) = \frac{1 - u^2}{1 + u^2}$.

Furthermore, the relation $u = \tan\left(\frac{x}{2}\right)$ implies that $x = 2 \arctan(u)$, therefore $dx = \frac{2}{1 + u^2} du$ (recall that $\arctan'(s) = \frac{1}{1 + s^2}$).

Using these formulas, we get

$$\int \frac{1}{\sin(x)} dx = \int \frac{1 + u^2}{2u} \frac{2}{1 + u^2} du = \int \frac{1}{u} du = \log |u| + C = \log \left| \tan\left(\frac{x}{2}\right) \right| + C.$$

7.3.4 Integration of Rational Functions

A function of the form $f(x) = \frac{p(x)}{q(x)}$ for polynomials p and $q \neq 0$ is called a **rational function**. In this section, we show a procedure for computing the indefinite integral of a rational function $f = \frac{p}{q}$ on an interval I on which q has no zeros. By polynomial division with remainder, one can always write $f = \frac{p}{q}$ in the form $f = g + \frac{r}{q}$, where g and r are polynomials with $\deg r < \deg q$. The polynomial function g is easy to integrate. Therefore, we always assume that the degree of p is smaller than the degree of q .

We start by integrating some elementary rational functions. Let $a \in \mathbb{R}$ and $n \in \mathbb{N}$ with $n \geq 2$. Then:

$$\int \frac{1}{x-a} dx = \log|x-a| + C \quad (7.10)$$

$$\int \frac{1}{(x-a)^n} dx = \frac{1}{1-n}(x-a)^{1-n} + C \quad (7.11)$$

$$\int \frac{1}{a^2+x^2} dx = \frac{1}{a} \arctan\left(\frac{x}{a}\right) + C \quad (7.12)$$

$$\int \frac{x}{a^2+x^2} dx = \frac{1}{2} \log(a^2+x^2) + C \quad (7.13)$$

$$\int \frac{x}{(a^2+x^2)^n} dx = \frac{1}{2(1-n)}(a^2+x^2)^{1-n} + C \quad (7.14)$$

The integrals (7.10) and (7.11) are calculated with substitution $u = x-a$, for (7.12) substitute $u = \frac{x}{a}$, for (7.13) and (7.14) substitute $u = a^2+x^2$.

To integrate a general rational function, we use what is called the **partial fraction decomposition** of rational functions. Let p, q be polynomials without nontrivial common divisors such that $q \neq 0$ and $\deg p < \deg q$.

First, factorize the polynomial q into linear and quadratic factors

$$q(x) = (x-a_1)^{k_1} \cdots (x-a_n)^{k_n} (x^2+b_1x+c_1)^{\ell_1} \cdots (x^2+b_mx+c_m)^{\ell_m}$$

Then, the rational function $\frac{p(x)}{q(x)}$ can be rewritten as a linear combination of rational functions of the form

$$\frac{1}{(x-a_i)^k}, \quad \frac{1}{(x^2+b_jx+c_j)^\ell}, \quad \frac{x}{(x^2+b_jx+c_j)^\ell},$$

for some $k \leq k_i$ and $\ell \leq \ell_j$, and then one needs to integrate each of these individual terms.

EXAMPLE 7.40. — We want to calculate the indefinite integral $\int \frac{x^4+1}{x^2(x+1)} dx$. We first rewrite the denominator:

$$\frac{x^4+1}{x^2(x+1)} = \frac{x^4+1}{x^3+x^2}.$$

We now perform polynomial division with remainder. We want to write

$$x^4+1 = (x^3+x^2)q(x) + r(x),$$

where $\deg r < 3$. The leading term of $x^4 + 1$ is x^4 , and the leading term of $x^3 + x^2$ is x^3 , so the first term of the quotient must be $\frac{x^4}{x^3} = x$:

$$x \cdot (x^3 + x^2) = x^4 + x^3.$$

Subtracting this from $x^4 + 1$ gives

$$(x^4 + 1) - (x^4 + x^3) = -x^3 + 1.$$

Now the leading term of the remainder is $-x^3$, so the next term in the quotient is $\frac{-x^3}{x^3} = -1$:

$$-1 \cdot (x^3 + x^2) = -x^3 - x^2.$$

Subtracting this from $-x^3 + 1$ gives

$$(-x^3 + 1) - (-x^3 - x^2) = x^2 + 1.$$

The remainder is now $x^2 + 1$, which has degree $2 < 3$, so the division stops. Thus

$$x^4 + 1 = (x^3 + x^2)(x - 1) + (x^2 + 1),$$

and therefore

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

We now look for a partial fraction decomposition of

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

The denominator factors as $x^2(x + 1)$, where:

- x^2 is a repeated linear factor,
- $x + 1$ is a simple linear factor.

For a factor x^2 , the general rule is that the numerator should be a polynomial of degree at most 1 (one less than the power of the factor), so we take a numerator of the form $ax + b$. For a simple linear factor $x + 1$, the numerator is a polynomial of degree at most 0, that is, a constant c . Thus the most general form of the decomposition is

$$\frac{x^2 + 1}{x^2(x + 1)} = \frac{ax + b}{x^2} + \frac{c}{x + 1}$$

for some real numbers a, b, c to be determined. To determine a, b, c we multiply both sides by $x^2(x + 1)$, which gives

$$x^2 + 1 = ax(x + 1) + b(x + 1) + cx^2 = ax^2 + ax + bx + b + cx^2.$$

Comparing the coefficients of x^2 , x , and the constant term, we obtain the linear system

$$a + c = 1, \quad a + b = 0, \quad b = 1,$$

so $a = -1$, $b = 1$, and $c = 2$. In summary,

$$\frac{x^4 + 1}{x^2(x + 1)} = x - 1 + \frac{x^2 + 1}{x^2(x + 1)} = x - 1 - \frac{1}{x} + \frac{1}{x^2} + \frac{2}{x + 1},$$

and therefore

$$\begin{aligned} \int \frac{x^4 + 1}{x^2(x + 1)} dx &= \int x dx - \int 1 dx - \int \frac{1}{x} dx + \int \frac{1}{x^2} dx + 2 \int \frac{1}{x + 1} dx \\ &= \frac{x^2}{2} - x - \log |x| - \frac{1}{x} + 2 \log |x + 1| + C. \end{aligned}$$

EXAMPLE 7.41. — We calculate the indefinite integral $\int \frac{1}{x(x^2 + 2x + 2)} dx$. Note that the polynomial $x^2 + 2x + 2$ has no real zeros. For the partial fraction decomposition, we look for $a, b, c \in \mathbb{R}$ such that

$$\frac{1}{x(x^2 + 2x + 2)} = \frac{a}{x} + \frac{bx + c}{x^2 + 2x + 2}.$$

To find a, b, c we multiply both sides by $x(x^2 + 2x + 2)$ and get

$$1 = a(x^2 + 2x + 2) + (bx + c)x = ax^2 + 2ax + 2a + bx^2 + cx,$$

thus

$$a + b = 0, \quad 2a + c = 0, \quad 2a = 1,$$

which gives $a = \frac{1}{2}$, $b = -\frac{1}{2}$, and $c = -1$. It follows that

$$\begin{aligned} \int \frac{1}{x(x^2 + 2x + 2)} dx &= \frac{1}{2} \int \frac{1}{x} dx - \frac{1}{2} \int \frac{x + 2}{x^2 + 2x + 2} dx \\ &= \frac{1}{2} \log |x| - \frac{1}{2} \int \frac{x + 2}{(x + 1)^2 + 1} dx + C \\ &= \frac{1}{2} \log |x| - \frac{1}{2} \int \frac{u + 1}{u^2 + 1} du + C \quad (u = x + 1) \\ &= \frac{1}{2} \log |x| - \frac{1}{2} \int \frac{u}{u^2 + 1} du - \frac{1}{2} \int \frac{1}{u^2 + 1} du + C \\ &= \frac{1}{2} \log |x| - \frac{1}{4} \log(u^2 + 1) - \frac{1}{2} \arctan(u) + C \\ &= \frac{1}{2} \log |x| - \frac{1}{4} \log((x + 1)^2 + 1) - \frac{1}{2} \arctan(x + 1) + C, \end{aligned}$$

where we used (7.13) and (7.14).

In some cases, the above procedure may also lead to compute integrals of the form $\int \frac{1}{(a^2 + x^2)^n} dx$ for an $a \in \mathbb{R}$ and $n \geq 2$, which (as explained previously) we can handle with the

trigonometric substitution $\tan(u) = \frac{x}{a}$.

REMARK 7.42. — Now that we know how to integrate rational functions, we can rediscuss the **half-angle method** introduced before. This allows one to compute the integral of rational expressions in sine and cosine. In fact, with the substitution $u = \tan\left(\frac{x}{2}\right)$, using that

$$\sin(x) = \frac{2u}{1+u^2}, \quad \cos(x) = \frac{1-u^2}{1+u^2}, \quad dx = \frac{2}{1+u^2} du,$$

(see Example 7.39), one ends up with the integral of a rational function in u .

EXERCISE 7.43. — Calculate the indefinite integral $\int \frac{\cos(x)}{2+\sin(x)} dx$ using the substitution $u = \tan\left(\frac{x}{2}\right)$.

REMARK 7.44. — Sometimes, one substitution or the other is carried out because there is a nested function in the function to be integrated, and there is simply no other method available. For example, in the integral $\int \sin(\sqrt{x}) dx$, none of the mentioned methods are available, but one is tempted to set $u = \sqrt{x}$, and this indeed leads to an integral that one can solve. Similarly, in an integral of the form $\int \frac{1}{1+e^x} dx$, one sets $u = e^x$.

7.3.5 Definite Integrals

Using the methods developed in the previous sections, we can now calculate both integrals on compact intervals $[a, b]$ and improper integrals (recall Definition 7.13). We now discuss some examples of the latter.

EXAMPLE 7.45. — We compute the improper integral $\int_0^1 \log(x) dx$ (this is improper since $x \mapsto \log(x)$ is unbounded as $x \rightarrow 0^+$). Recalling (7.32) and Example 3.64, we have

$$\begin{aligned} \int_0^1 \log(x) dx &= \lim_{a \rightarrow 0^+} \int_a^1 \log(x) dx = \lim_{a \rightarrow 0^+} [x \log(x) - x]_a^1 \\ &= \lim_{a \rightarrow 0^+} (\log(1) - 1 - a \log(a) + a) = -1. \end{aligned}$$

EXERCISE 7.46. — Calculate $\int_0^{\frac{\pi}{2}} \tan(x) dx$.

EXERCISE 7.47. — Decide for which $p \in \mathbb{R}_{\geq 0}$ the improper integral $\int_0^\infty x \sin(x^p) dx$ converges.

7.3.6 The Gamma Function

The **Gamma-function** Γ is defined, for $s \in (0, \infty)$, by the improper integral

$$\Gamma(s) = \int_0^\infty x^{s-1} e^{-x} dx. \quad (7.15)$$

To verify that this improper integral indeed converges, we examine the integration limits 0 and ∞ separately. For $0 < a < b$ we find, using integration by parts,

$$\int_a^b x^{s-1} e^{-x} dx = \frac{1}{s} \int_a^b (x^s)' e^{-x} dx = \frac{1}{s} [x^s e^{-x}]_a^b + \frac{1}{s} \int_a^b x^s e^{-x} dx. \quad (7.16)$$

We obtain

$$\begin{aligned} \int_0^b x^{s-1} e^{-x} dx &= \lim_{a \rightarrow 0} \left(\frac{1}{s} [x^s e^{-x}]_a^b + \frac{1}{s} \int_a^b x^s e^{-x} dx \right) \\ &= \frac{1}{s} b^s \exp(-b) + \frac{1}{s} \int_0^b x^s e^{-x} dx, \end{aligned}$$

where the integral on the right is an actual Riemann integral since the function $x^s e^{-x}$ is continuous on $[0, b]$. To investigate the upper limit of integration, we note that there exists $R > 0$ such that $e^x > x^{s+2}$ holds for all $x > R$. Thus

$$\int_0^\infty x^s e^{-x} dx \leq \int_0^R x^s e^{-x} dx + \int_R^\infty x^{-2} dx < \infty$$

which shows that (7.16) converges as $b \rightarrow \infty$. Specifically, we obtain

$$\int_0^\infty x^{s-1} e^{-x} dx = \lim_{b \rightarrow \infty} \left(\frac{1}{s} b^s \exp(-b) + \frac{1}{s} \int_0^b x^s e^{-x} dx \right) = \frac{1}{s} \int_0^\infty x^s e^{-x} dx.$$

This shows that the gamma function satisfies the relation

$$\Gamma(s+1) = s \Gamma(s) \quad (7.17)$$

for all $s \in (0, \infty)$, from which one can deduce that the Gamma function extends the factorial function from \mathbb{N} to $(0, \infty)$. In fact

$$\Gamma(1) = \int_0^\infty x^0 e^{-x} dx = \int_0^\infty e^{-x} dx = [e^{-x}]_0^\infty = e^0 = 1,$$

therefore (7.17) implies that

$$\Gamma(n+1) = n\Gamma(n) = n(n-1)\Gamma(n-1) = \dots = n!\Gamma(1) = n! \quad \forall n \in \mathbb{N}.$$

7.4 Taylor Series

7.4.1 Taylor Approximation

As we have seen in Chapter 5, given a differentiable function $f : D \rightarrow \mathbb{R}$, the derivative $f'(x_0)$ gives the slope of the tangent to the graph of f at x_0 . Moreover, the corresponding affine function

$$x \mapsto f(x_0) + f'(x_0)(x - x_0)$$

approximates the function f within an error $o(|x - x_0|)$ as $x \rightarrow x_0$, see (5.2). The idea behind Taylor's theorem is that the "quality" of the approximation can be increased by considering higher-order polynomials instead of affine approximations.

In this section, it will be convenient to use the following abuse of notation: given $a, b \in \mathbb{R}$, irrespective of the order between a and b , $[a, b]$ denotes the interval between them. In other words, for all $a, b \in \mathbb{R}$, $[a, b]$ and $[b, a]$ denote the same interval.

We also recall that, if $a < b$, then

$$\left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx,$$

31

see Theorem 6.24. If instead $b < a$, then a minus sign appears (recall (7.3)) and we get

$$\left| \int_a^b f(x) dx \right| = \left| \int_b^a f(x) dx \right| \leq \int_b^a |f(x)| dx = - \int_a^b |f(x)| dx = \left| \int_a^b |f(x)| dx \right|.$$

In conclusion, independently of the order of a and b , we always have

$$\left| \int_a^b f(x) dx \right| \leq \left| \int_a^b |f(x)| dx \right|.$$

Let $I \subseteq \mathbb{R}$ be an open interval, and $f : I \rightarrow \mathbb{R}$ be an n times differentiable function. The n -th **Taylor approximation** of f around a point $x_0 \in I$ is the polynomial function

$$P_n(x) = \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k. \quad (7.18)$$

Note that, with this choice of the coefficients, $P^{(k)}(x_0) = f^{(k)}(x_0)$ for $k \in \{0, \dots, n\}$.

We will state and prove different versions of Taylor's Theorem. We begin with this first version:

THEOREM 7.48: TAYLOR EXPANSION TO ORDER n WITH INTEGRAL REMAINDER

Let $n \geq 1$, $f : [a, b] \rightarrow \mathbb{R}$ an n -times continuously differentiable function, and fix $x_0 \in [a, b]$. Then, for all $x \in [a, b]$,

$$f(x) = P_{n-1}(x) + \int_{x_0}^x f^{(n)}(t) \frac{(x-t)^{n-1}}{(n-1)!} dt, \quad (7.19)$$

where P_{n-1} is the $(n-1)$ -th Taylor approximation of f defined in (7.18).

REMARK 7.49. — In the above theorem, the assumption that f is an n -times continuously differentiable function guarantees that the integral of the continuous function $t \mapsto f^{(n)}(t) \frac{(x-t)^{n-1}}{(n-1)!}$ exists.

Proof. The proof follows by induction on n and integration by parts.

If $n = 1$ then $f : [a, b] \rightarrow \mathbb{R}$ is continuously differentiable and, by Corollary 7.5, we get

$$f(x) = f(x_0) + \int_{x_0}^x f'(t) dt = P_0(x) + \int_{x_0}^x f^{(1)}(t) \frac{(x-t)^0}{0!} dt,$$

as desired.

To explain the idea behind the inductive step, assume first $n = 2$ (so f is twice continuously differentiable). Then, in the integral above, we can apply integration by parts to the functions $f'(t)$ and $g(t) = t - x$. Indeed, since $g' = 1$ and $g(x) = 0$, we get

$$\begin{aligned} f(x) &= f(x_0) + \int_{x_0}^x f'(t)g'(t) dt \\ &= f(x_0) + [f'(t)g(t)]_{x_0}^x - \int_{x_0}^x f''(t)g(t) dt \\ &= f(x_0) + f'(x_0)(x - x_0) + \int_{x_0}^x f''(t)(x - t) dt \\ &= P_1(x) + \int_{x_0}^x f^{(2)}(t) \frac{(x-t)^1}{1!} dt. \end{aligned}$$

This proves the case $n = 2$.

More generally, assume that the statement of the theorem is true for some $n \geq 1$ and that $f : [a, b] \rightarrow \mathbb{R}$ is an $(n+1)$ -times continuously differentiable function. Then, by the induction hypothesis,

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k + \int_{x_0}^x f^{(n)}(t) \frac{(x-t)^{n-1}}{(n-1)!} dt$$

for all $x \in [a, b]$. If we set $g(t) = -\frac{(x-t)^n}{n!}$, then $g'(t) = \frac{(x-t)^{n-1}}{(n-1)!}$ and it follows from integration by parts that

$$\begin{aligned} f(x) &= \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k + \left[f^{(n)}(t)g(t) \right]_{x_0}^x - \int_{x_0}^x f^{(n+1)}(t)g(t) dt \\ &= \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k + f^{(n)}(x_0) \frac{(x - x_0)^n}{n!} + \int_{x_0}^x f^{(n+1)}(t) \frac{(x - t)^n}{n!} dt \\ &= \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k + \int_{x_0}^x f^{(n+1)}(t) \frac{(x - t)^n}{n!} dt. \end{aligned}$$

This proves the induction step, and hence the result. \square

Next, we prove the following alternative version of Taylor's Theorem. Here, we do not require $f^{(n)}$ to be continuous, but only to exist. Note that in the case $n = 1$, this result corresponds to the Mean Value Theorem 5.31.

THEOREM 7.50: TAYLOR EXPANSION TO ORDER n WITH LAGRANGE REMAINDER

Let $n \geq 1$, $f : [a, b] \rightarrow \mathbb{R}$ an n -times differentiable function, and fix $x_0 \in [a, b]$. Then, for all $x \in [a, b]$ with $x \neq x_0$ there exists $\xi_L \in (x_0, x)$ such that

$$f(x) = P_{n-1}(x) + \frac{f^{(n)}(\xi_L)}{n!} (x - x_0)^n. \quad (7.20)$$

REMARK 7.51. — If $x = x_0$, the formula holds trivially: one may take $\xi_L = x_0$, since

$$f(x_0) = P_{n-1}(x_0) \quad \text{and} \quad \frac{f^{(n)}(\xi_L)}{n!} (x_0 - x_0)^n = 0.$$

Proof. Fix $x \in (a, b)$. Without loss of generality, assume $x > x_0$ (the case $x < x_0$ is analogous) and consider the function $g : (a, b) \rightarrow \mathbb{R}$ defined as

$$g(t) = f(t) + f^{(1)}(t)(x - t) + \dots + \frac{f^{(n-1)}(t)}{(n-1)!} (x - t)^{n-1} = \sum_{k=0}^{n-1} \frac{f^{(k)}(t)}{k!} (x - t)^k. \quad (7.21)$$

Then $g(x) = f(x)$ and $g(x_0) = P_{n-1}(x)$. Also, its derivative is given by

$$\begin{aligned} g'(t) &= \sum_{k=0}^{n-1} \frac{f^{(k+1)}(t)}{k!} (x - t)^k - \sum_{k=0}^{n-1} \frac{f^{(k)}(t)}{k!} k(x - t)^{k-1} \\ &= \sum_{k=0}^{n-1} \frac{f^{(k+1)}(t)}{k!} (x - t)^k - \sum_{k=1}^{n-1} \frac{f^{(k)}(t)}{(k-1)!} (x - t)^{k-1} \\ &= \sum_{k=0}^{n-1} \frac{f^{(k+1)}(t)}{k!} (x - t)^k - \sum_{k=0}^{n-2} \frac{f^{(k+1)}(t)}{k!} (x - t)^k = \frac{f^{(n)}(t)}{(n-1)!} (x - t)^{n-1}. \end{aligned}$$

Hence, applying the Cauchy Mean Value Theorem 5.35 in the interval $[x_0, x]$ to the functions $g(t)$ and $h(t) = -(x - t)^n$, we deduce the existence of a point $\xi_L \in (x_0, x)$ such that

31

$$\frac{f(x) - P_{n-1}(x)}{(x - x_0)^n} = \frac{g(x) - g(x_0)}{h(x) - h(x_0)} = \frac{g'(\xi_L)}{h'(\xi_L)} = \frac{\frac{f^{(n)}(\xi_L)}{(n-1)!}(x - \xi_L)^{n-1}}{n(x - \xi_L)^{n-1}} = \frac{f^{(n)}(\xi_L)}{n!}.$$

This implies (7.20) and concludes the proof. \square

We can now state our two versions of Taylor's approximation, using the little-o and the big-O notation.

COROLLARY 7.52: TAYLOR APPROXIMATION WITH LITTLE-O

Let $n \geq 1$, $f : [a, b] \rightarrow \mathbb{R}$ an n -times continuously differentiable function, and fix $x_0 \in [a, b]$. Then, for all $x \in [a, b]$,

$$f(x) = P_n(x) + o(|x - x_0|^n) \quad \text{as } x \rightarrow x_0. \quad (7.22)$$

Proof. Thanks to Theorem 7.48,

$$\begin{aligned} f(x) &= P_{n-1}(x) + \int_{x_0}^x f^{(n)}(t) \frac{(x-t)^{n-1}}{(n-1)!} dt \\ &= P_{n-1}(x) + \int_{x_0}^x f^{(n)}(x_0) \frac{(x-t)^{n-1}}{(n-1)!} dt + \int_{x_0}^x (f^{(n)}(t) - f^{(n)}(x_0)) \frac{(x-t)^{n-1}}{(n-1)!} dt. \end{aligned}$$

Also, using the change of variable $s = x - t$, we note that

32

$$\int_{x_0}^x \frac{(x-t)^{n-1}}{(n-1)!} dt = \frac{1}{(n-1)!} \int_0^{x-x_0} s^{n-1} ds = \frac{1}{(n-1)!} \frac{(x-x_0)^n}{n} = \frac{(x-x_0)^n}{n!}, \quad (7.23)$$

therefore

$$\int_{x_0}^x f^{(n)}(x_0) \frac{(x-t)^{n-1}}{(n-1)!} dt = f^{(n)}(x_0) \frac{(x-x_0)^n}{n!}.$$

Hence, we can write

$$\begin{aligned} f(x) &= P_{n-1}(x) + f^{(n)}(x_0) \frac{(x-x_0)^n}{n!} + \int_{x_0}^x (f^{(n)}(t) - f^{(n)}(x_0)) \frac{(x-t)^{n-1}}{(n-1)!} dt \\ &= P_n(x) + \int_{x_0}^x (f^{(n)}(t) - f^{(n)}(x_0)) \frac{(x-t)^{n-1}}{(n-1)!} dt. \end{aligned} \quad (7.24)$$

Now, given $\varepsilon > 0$, it follows from the continuity of $f^{(n)}$ at x_0 that there exists $\delta > 0$ such that $|f^{(n)}(x) - f^{(n)}(x_0)| < \varepsilon$ for all $x \in (x_0 - \delta, x_0 + \delta) \cap [a, b]$. Therefore, if $x \in (x_0 - \delta, x_0 + \delta) \cap [a, b]$, we can bound the integrand in the last integral by

$$\left| (f^{(n)}(t) - f^{(n)}(x_0)) \frac{(x-t)^{n-1}}{(n-1)!} \right| \leq |f^{(n)}(t) - f^{(n)}(x_0)| \frac{|x-t|^{n-1}}{(n-1)!}$$

$$< \varepsilon \frac{|x - t|^{n-1}}{(n-1)!} \quad \forall t \in [x_0, x].$$

Hence, using (7.23) again, we get

$$\begin{aligned} |f(x) - P_n(x)| &\leq \left| \int_{x_0}^x (f^{(n)}(t) - f^{(n)}(x_0)) \frac{(x-t)^{n-1}}{(n-1)!} dt \right| \\ &< \varepsilon \left| \int_{x_0}^x \frac{|x-t|^{n-1}}{(n-1)!} dt \right| = \varepsilon \frac{|x-x_0|^n}{n!}, \end{aligned}$$

which shows that $f(x) - P_n(x) = o(|x - x_0|^n)$ as $x \rightarrow x_0$. \square

COROLLARY 7.53: TAYLOR APPROXIMATION WITH BIG-O

Let $n \geq 1$, $f : [a, b] \rightarrow \mathbb{R}$ an n -times differentiable function, and fix $x_0 \in [a, b]$. Assume that there exists $M > 0$ such that $|f^{(n)}(x)| \leq M$ for all $x \in [a, b]$. Then

$$f(x) = P_{n-1}(x) + O(|x - x_0|^n) \quad \text{as } x \rightarrow x_0. \quad (7.25)$$

Proof. Given $x \in [a, b]$, we apply (7.20) to find a point $\xi_L \in [x_0, x]$ such that

$$f(x) - P_{n-1}(x) = \frac{f^{(n)}(\xi_L)}{n!} (x - x_0)^n.$$

32

Since $|f^{(n)}(\xi_L)| \leq M$, this implies

$$|f(x) - P_{n-1}(x)| \leq \frac{M}{n!} |x - x_0|^n,$$

therefore $f(x) - P_{n-1}(x) = O(|x - x_0|^n)$ as $x \rightarrow x_0$, as desired. \square

EXAMPLE 7.54. — As mentioned earlier, if f is differentiable at x_0 , then

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + o(|x - x_0|) \quad \text{as } x \rightarrow x_0.$$

Taylor's Theorem allows us to obtain more accurate approximations when f has higher regularity.

1. If f is twice differentiable and f'' is bounded in a neighborhood of x_0 , then Corollary 7.53 (with $n = 2$) gives

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + O(|x - x_0|^2) \quad \text{as } x \rightarrow x_0.$$

If in addition f'' is continuous (equivalently, $f \in C^2$), then Corollary 7.52 yields

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 + o(|x - x_0|^2) \quad \text{as } x \rightarrow x_0.$$

2. If f is smooth, then Corollary 7.52 yields

$$f(x) = P_n(x) + o(|x - x_0|^n) \quad \text{as } x \rightarrow x_0,$$

while Corollary 7.53 applied with $n + 1$ in place of n gives

$$f(x) = P_n(x) + O(|x - x_0|^{n+1}) \quad \text{as } x \rightarrow x_0.$$

Hence, in the case when f is smooth, the bound on $f - P_n$ provided by Corollary 7.53 is often more convenient.

EXAMPLE 7.55. — We illustrate how Taylor expansions can be used to compute limits and to obtain expansions of more complicated functions.

1. *Compute the limit $\lim_{x \rightarrow 0} \frac{\sin(x) - x}{x^3}$.*

We consider the Taylor polynomial of \sin at 0 of degree 3. Since

$$\sin(0) = 0, \quad \sin'(0) = 1, \quad \sin''(0) = 0, \quad \sin^{(3)}(0) = -1,$$

the Taylor polynomial of degree 3 at 0 is

$$P_3(x) = 0 + 1 \cdot x + \frac{0}{2} \cdot x^2 + \frac{-1}{6} \cdot x^3 = x - \frac{x^3}{6}.$$

Taylor's Theorem gives

$$\sin(x) = P_3(x) + O(x^4) = x - \frac{x^3}{6} + O(x^4) \quad \text{as } x \rightarrow 0.$$

We stop the expansion at order 3 because the limit involves x^3 in the denominator; any term of order ≥ 4 will disappear in the limit.

Then

$$\sin(x) - x = \left(x - \frac{x^3}{6} + O(x^4)\right) - x = -\frac{x^3}{6} + O(x^4),$$

and thus

$$\frac{\sin(x) - x}{x^3} = \frac{-\frac{x^3}{6} + O(x^4)}{x^3} = -\frac{1}{6} + O(x) \quad \text{as } x \rightarrow 0.$$

Hence

$$\lim_{x \rightarrow 0} \frac{\sin(x) - x}{x^3} = -\frac{1}{6}.$$

2. *Compute the Taylor expansion of $\sin^2(x)$ near 0.*

Using the expansion above,

$$\sin(x) = x - \frac{x^3}{6} + O(x^4) \quad \text{as } x \rightarrow 0,$$

we obtain $\sin^2(x)$ by squaring:

$$\begin{aligned}\sin^2(x) &= \left(x - \frac{x^3}{6} + O(x^4)\right)^2 \\ &= x^2 - \frac{x^4}{3} + \frac{x^6}{36} + 2x \cdot O(x^4) - \frac{x^3}{2} \cdot O(x^4) + O(x^4) \cdot O(x^4).\end{aligned}$$

We now observe that

$$2x \cdot O(x^4) = O(x^5), \quad x^3 \cdot O(x^4) = O(x^7), \quad O(x^4) \cdot O(x^4) = O(x^8).$$

All these terms are of order at least x^5 , so they may be merged into a single remainder term $O(x^5)$. The term $\frac{x^6}{36}$ can also be absorbed into the remainder for the same reason. This proves that

$$\sin^2(x) = x^2 - \frac{x^4}{3} + O(x^5),$$

and the Taylor polynomial of \sin^2 at 0 of degree 4 is

$$P_4(x) = x^2 - \frac{x^4}{3}.$$

This illustrates a general strategy: once we know the expansion of a function (here \sin), we can obtain expansions of combinations of that function (here \sin^2) by algebraic operations on the corresponding Taylor polynomials.

32

REMARK 7.56. — In the previous example we showed that

$$\sin(x) = x - \frac{x^3}{6} + O(x^4) \quad \text{as } x \rightarrow 0.$$

These coefficients agree with those appearing in the power series expansion of \sin . In Section 7.4.2 below, we will see that \sin is an *analytic* function, meaning that its Taylor series around 0 actually converges to $\sin(x)$. For such functions, the Taylor polynomial of degree n is simply the truncation of an infinite series that represents the function itself.

EXAMPLE 7.57. — We can use the Taylor approximation to refine the discussion in Section 5.2.1. Let $f : (a, b) \rightarrow \mathbb{R}$ be an n -times continuously differentiable function. Suppose $x_0 \in (a, b)$ satisfies

$$f'(x_0) = \dots = f^{(n-1)}(x_0) = 0.$$

Then the following implications hold:

- If $f^{(n)}(x_0) < 0$ and n is even, then f has an isolated local maximum in x_0 .
- If $f^{(n)}(x_0) > 0$ and n is even, then f has an isolated local minimum in x_0 .
- If $f^{(n)}(x_0) \neq 0$ and n is odd, then x_0 is not a local extremum of f .

All three statements follow from (7.20), which, in this case, takes the form

$$f(x) = f(x_0) + \frac{f^{(n)}(\xi_L)}{n!}(x - x_0)^n, \quad \xi_L \in (x_0, x).$$

Indeed, if $f^{(n)}(x_0) > 0$, by continuity there exists $\delta > 0$ such that $f^{(n)}(\xi_L) > 0$ for $\xi_L \in (x_0, x) \subset (x_0 - \delta, x_0 + \delta)$. If n is even, then $(x - x_0)^n > 0$ for $x \neq x_0$ and we deduce that $f(x) > f(x_0)$ for $x \in (x_0 - \delta, x_0 + \delta)$ with $x \neq x_0$. If n is odd, then $(x - x_0)^n$ changes sign when considering $x > x_0$ and $x < x_0$, so x_0 is not a local extremum of f .

On the other hand, if $f^{(n)}(x_0) < 0$ and n is even, the same argument as above shows that $f(x) < f(x_0)$ for $x \in (x_0 - \delta, x_0 + \delta)$ with $x \neq x_0$, while in the case n odd x_0 is not a local extremum of f .

7.4.2 Analytic Functions

Motivated by Taylor's Theorem, one might expect that if f is smooth and we replace the finite Taylor polynomial P_n by the full **Taylor series**

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k,$$

then this infinite series should converge to $f(x)$. Unfortunately, this is not true in general: only rather special smooth functions enjoy this property.

Note that the Taylor series is centered at x_0 instead of 0 (i.e., x^n is replaced with $(x - x_0)^n$). Hence, all theorems about power series from Section 4.4 still hold, but taking into account that now x_0 plays the role of the center. In particular, if the series has radius of convergence $R > 0$, then it converges for all $x \in (x_0 - R, x_0 + R)$, while it diverges for $|x - x_0| > R$.

DEFINITION 7.58: ANALYTIC FUNCTIONS

Let $I \subseteq \mathbb{R}$ be an interval and $x_0 \in I$. A smooth function $f : I \rightarrow \mathbb{R}$ is called *analytic* at x_0 if the Taylor series of f around x_0 has radius of convergence $R > 0$ and there exists $\delta \in (0, R)$ such that

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n \quad \forall x \in (x_0 - \delta, x_0 + \delta) \cap I.$$

We say f is analytic in I if f is analytic at all points in I .

In other words, analytic functions $f : I \rightarrow \mathbb{R}$ are characterized by the fact that, for every point $x_0 \in I$, there exists a power series that converges to f in a neighborhood of x_0 .

As the next example shows, there are smooth functions f whose Taylor series converges to a function different from f .

EXAMPLE 7.59. — Consider the function $\psi : \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$\psi : x \in \mathbb{R} \mapsto \begin{cases} \exp\left(-\frac{1}{x}\right) & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}.$$

As shown in Exercise 5.24, ψ is smooth on \mathbb{R} and satisfies $\psi^{(n)}(0) = 0$ for all $n \in \mathbb{N}$. Hence, the Taylor series of the function ψ at the point $x_0 = 0$ is the zero series:

$$\sum_{n=0}^{\infty} \frac{\psi^{(n)}(0)}{n!} x^n = \sum_{n=0}^{\infty} \frac{0}{n!} x^n = 0.$$

This series has an infinite radius of convergence and converges to the function 0. Since $\psi(x) > 0$ holds for all $x > 0$, the Taylor series does not converge to ψ , and so ψ is not analytic at the point $x_0 = 0$.

The next result provides a criterion that guarantees that the Taylor series of f converges to f in a neighborhood of x_0 .

THEOREM 7.60: A CRITERION FOR ANALYTICITY AT x_0

Let $I \subseteq \mathbb{R}$ be an interval and $f : I \rightarrow \mathbb{R}$ a smooth function. Given $x_0 \in I$, assume that there exist constants $r, C_0, A > 0$ such that

$$|f^{(n)}(x)| \leq C_0 A^n n! \quad \text{for all } x \in (x_0 - r, x_0 + r) \cap I, \quad n \in \mathbb{N}.$$

Then f is analytic at x_0 .

Proof. We first estimate the radius of convergence R of the Taylor series. If we define $a_n = \frac{f^{(n)}(x_0)}{n!}$, then the Taylor series is equal to $\sum_{n=0}^{\infty} a_n (x - x_0)^n$. Thus, thanks to our assumption on the size of $|f^{(n)}|$, it follows that

$$|a_n| \leq \frac{C_0 A^n n!}{n!} = C_0 A^n.$$

This implies that

$$\limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|} \leq \limsup_{n \rightarrow \infty} \sqrt[n]{C_0 A^n} = \limsup_{n \rightarrow \infty} \sqrt[n]{C_0} A = A,$$

therefore, by the definition of radius of convergence (see Definition 4.39), $R \geq \frac{1}{A}$.

Now, fix $\delta < \min \left\{ r, \frac{1}{A} \right\}$. Given $x \in (x_0 - \delta, x_0 + \delta) \cap I$, we apply (7.20) and our assumption on the size of $|f^{(n)}|$ to deduce that

$$|f(x) - P_{n-1}(x)| \leq \frac{|f^{(n)}(\xi_L)|}{n!} |x - x_0|^n \leq C_0 A^n |x - x_0|^n \leq C_0 (A\delta)^n.$$

Since $A\delta < 1$ (because $\delta < \frac{1}{A}$), letting $n \rightarrow \infty$ we conclude that

$$f(x) = \lim_{n \rightarrow \infty} P_{n-1}(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!} (x - x_0)^n \quad \forall x \in (x_0 - \delta, x_0 + \delta) \cap I,$$

as desired. □

As a direct consequence of Theorem 7.60, we immediately deduce the following:

COROLLARY 7.61: A CRITERION FOR ANALYTICITY

Let $f : [a, b] \rightarrow \mathbb{R}$ be a smooth function, and assume there exist constants $C_0, A > 0$ such that

$$|f^{(n)}(x)| \leq C_0 A^n n! \quad \text{for all } x \in [a, b], \quad n \in \mathbb{N}. \quad (7.26)$$

Then f is analytic on $[a, b]$.

- EXERCISE 7.62. —
1. Show that the functions \exp, \sin, \sinh satisfy the property (7.26) on any interval $[a, b] \subset \mathbb{R}$.
 2. Show that the function \log satisfies (7.26) on any interval $[a, b] \subset (0, \infty)$.
 3. Let $f, g : [a, b] \rightarrow \mathbb{R}$ be functions satisfying (7.26). Show that $f + g$ and $f \cdot g$ also satisfy this property (possibly with different constants C_0 and A).

Chapter 8

Ordinary Differential Equations

Setting up and solving differential equations is one of the main practical uses of calculus. Such equations are used in physics, chemistry, biology, and many other sciences. They are also central in areas like structural engineering, modern economics, and information technology.

8.1 Ordinary Differential Equations (ODEs)

In this section we study *ordinary differential equations* (ODEs). They describe how an unknown quantity depends on one real variable (often denoted by x or t) and how this dependence is constrained by relations involving its derivatives.

It is convenient to fix some notation from the beginning:

- we usually denote the *unknown function* by u ;
- the independent variable is denoted by x (or by t when it represents time);
- letters like f , g , a_0 , a_1 will typically denote given functions or constants appearing in the equation (data of the problem).

Although derivatives are usually denoted using $'$ (so u' , u'' , etc.), it is common to use a dot to denote derivatives with respect to time (so \dot{u} , \ddot{u} , etc.).

DEFINITION 8.1: ODEs

An **ordinary differential equation (ODE)** is a relation involving a function $u : \mathbb{R} \rightarrow \mathbb{R}$ of a real variable $x \in \mathbb{R}$ and its derivatives. The general form of an n -th order ODE is

$$G(x, u(x), u'(x), u''(x), \dots, u^{(n)}(x)) = 0, \quad (8.1)$$

where $G : \mathbb{R}^{n+2} \rightarrow \mathbb{R}$ is a given function.

In many examples the independent variable is time t and the equation describes the evolution of a system, but we keep the generic notation x unless we want to stress the time interpretation.

ODEs can be classified according to several criteria:

1. *Order*: An ODE is of order n if $u^{(n)}$ is the highest derivative appearing in the equation. For instance:

(a) $u'' + u = 0 \rightsquigarrow$ second order.

(b) $u^{(3)} = x^2u + x \rightsquigarrow$ third order.

(c) $(u')^2 + u - x^3 = 0 \rightsquigarrow$ first order.

2. *Linearity*: An ODE is *linear* if it is linear in u and its derivatives. Otherwise, it is *nonlinear*. Here “linear” means that u, u', u'', \dots appear only to the first power and are not multiplied with each other.

(a) $u'' + u = 0 \rightsquigarrow$ linear.

(b) $u'' + u^2 = 0 \rightsquigarrow$ nonlinear (because of u^2).

(c) $u'' + u'u = 0 \rightsquigarrow$ nonlinear (because of the product $u'u$).

(d) $u^{(3)} = x^2u + x \rightsquigarrow$ linear.

(e) $(u')^2 + u - x^3 = 0 \rightsquigarrow$ nonlinear (because of $(u')^2$).

3. *Homogeneity (for linear ODEs)*: For a linear ODE, we say it is *homogeneous* if all terms involve the function or its derivatives. Equivalently, if u is a solution then Au is a solution for all $A \in \mathbb{R}$. If there is an additional term that does not depend on u (a “forcing term”), the equation is *non-homogeneous*.

(a) $u'' + u = 0 \rightsquigarrow$ homogeneous.

(b) $u^{(3)} = x^2u + x \rightsquigarrow$ non-homogeneous (because of the term $+x$).

(c) $u^{(3)} = x^2u \rightsquigarrow$ homogeneous.

EXAMPLE 8.2. — We now present some classic examples of ODEs and their applications. In each case we indicate what plays the role of the unknown u .

1. *Newton’s Law of Cooling*: This law states that

“The rate of heat loss of a body is proportional to the difference between its temperature and the temperature of the surrounding environment.”

If $T(t)$ denotes the temperature of the object at time t , the law gives

$$\dot{T}(t) = -k(T(t) - T_{\text{env}}),$$

where:

- T_{env} is the ambient temperature (assumed constant);
- $k > 0$ is a constant describing the heat transfer.

Here the unknown function is $u(t) = T(t)$, and the given data are k and T_{env} . This is a linear, non-homogeneous, first-order ODE.

2. *Harmonic Oscillator:* Consider a mass attached to a spring. Let $x(t)$ denote its displacement from equilibrium. The restoring force is proportional to x :

$$F_{\text{spring}} = -kx,$$

with $k > 0$ (Hooke's law). Newton's law $m\ddot{x} = F$ gives

$$m\ddot{x}(t) = -kx(t),$$

or equivalently

$$\ddot{x}(t) + \omega^2 x(t) = 0,$$

where $\omega = \sqrt{\frac{k}{m}} > 0$ is the angular frequency. Here $u(t) = x(t)$ is the unknown, and ω is a given parameter. This is linear, homogeneous, and second order.

If we also take into account friction proportional to the velocity \dot{x} , we obtain the damped harmonic oscillator:

$$\ddot{x}(t) + 2\zeta\omega\dot{x}(t) + \omega^2 x(t) = 0,$$

where $\zeta \geq 0$ is the damping ratio. This equation models many vibrating systems in physics and engineering.

3. *Logistic Population Growth:* In population dynamics, a common model for a population with limited resources is the logistic equation

$$\dot{P}(t) = rP(t) \left(1 - \frac{P(t)}{K} \right),$$

where:

- $P(t)$ is the population at time t ;
- $r > 0$ is the intrinsic growth rate;
- $K > 0$ is the carrying capacity (maximum sustainable population).

Here the role of u is played by P , while r and K are given parameters. This is a nonlinear first-order ODE.

4. *Bessel Equation*: The *Bessel equation* with parameter $\alpha \in \mathbb{R}$ is

$$x^2 u''(x) + xu'(x) + (x^2 - \alpha^2)u(x) = 0.$$

This is linear, homogeneous, and second order. The unknown is $u(x)$, and α is a given parameter. Its solutions are the *Bessel functions*, which appear, for instance, in problems of heat conduction and wave propagation in cylindrical geometries, and in quantum mechanics.

5. *Airy Equation*: The *Airy equation* is

$$u''(x) - \alpha^2 xu(x) = 0,$$

where $\alpha > 0$ is a given constant. The corresponding solutions are called *Airy functions*. They arise, for example, in quantum mechanics when studying a particle in a linear or triangular potential.

So far, we have only considered single equations, but one can also study *systems* of ODEs with several unknown functions u_1, \dots, u_n . We will not go into this now, but many ideas are similar.

33

In addition, solutions are often required to satisfy extra conditions such as $u(0) = 0$ (prescribed position at time 0) and/or $u'(0) = 1$ (prescribed velocity at time 0). When these conditions are imposed at a single time (typically $t = 0$), they are called **initial conditions**. More general conditions (for instance at two different points, such as $u(0) = 0$ and $u(1) = 1$) are called **boundary conditions**.

Later we shall see that, under suitable assumptions on the data (for example on a function f appearing in the equation), prescribing initial conditions often leads to a unique solution. This is the content of the Cauchy–Lipschitz (or Picard–Lindelöf) theorem.

8.1.1 Linear First Order ODEs

We now consider linear first-order ODEs. Throughout this subsection we fix a non-empty interval $I \subseteq \mathbb{R}$ that is not a single point, and we study equations of the form

$$u'(x) + f(x)u(x) = g(x),$$

where f and g are given continuous functions on I , and u is the unknown.

We start with the homogeneous case $g \equiv 0$.

THEOREM 8.3: HOMOGENEOUS LINEAR 1ST ORDER ODES

Let $f : I \rightarrow \mathbb{R}$ be continuous and consider the homogeneous first-order linear ODE

$$u'(x) + f(x)u(x) = 0 \quad \forall x \in I. \quad (8.2)$$

Let $F : I \rightarrow \mathbb{R}$ be a primitive of f . Then all C^1 solutions $u : I \rightarrow \mathbb{R}$ of (8.2) are of the form

$$u(x) = Ae^{-F(x)}, \quad A \in \mathbb{R}.$$

In other words, the set of solutions of (8.2) forms a one-dimensional linear subspace of $C^1(I)$.

Proof. Given $A \in \mathbb{R}$, define $u(x) = Ae^{-F(x)}$. Then

$$u'(x) = -F'(x)Ae^{-F(x)} = -f(x)Ae^{-F(x)} = -f(x)u(x) \quad \forall x \in I,$$

so u solves the ODE.

Conversely, let $u \in C^1(I)$ solve (8.2) and set $v(x) = e^{F(x)}u(x)$. Then

$$v'(x) = (e^{F(x)})'u(x) + e^{F(x)}u'(x) = e^{F(x)}(f(x)u(x) - f(x)u(x)) = 0 \quad \forall x \in I.$$

By Corollary 5.48, we deduce that $v(x) = A$ for some $A \in \mathbb{R}$, hence $u(x) = Ae^{-F(x)}$. \square

REMARK 8.4. — In the previous result, solutions are written using a primitive F of f . Since primitives are defined up to an additive constant, we can replace F by $F + C$ for any $C \in \mathbb{R}$. This amounts to replacing $Ae^{-F(x)}$ by $Ae^{-C}e^{-F(x)}$, and since $A \in \mathbb{R}$ is arbitrary, this does not change the set of solutions.

EXAMPLE 8.5. — 1. *Constant-coefficient homogeneous equation.* Consider the differential equation

$$u'(x) + 3u(x) = 0.$$

This is of the form $u'(x) + f(x)u(x) = 0$ with $f(x) = 3$, and a primitive of f is $F(x) = 3x$. By Theorem 8.3, all solutions are of the form

$$u(x) = Ae^{-F(x)} = Ae^{-3x}, \quad A \in \mathbb{R}.$$

2. *Variable-coefficient homogeneous equation.* Consider the differential equation

$$u'(x) - \frac{2}{x}u(x) = 0, \quad x \in (0, \infty).$$

Here $f(x) = -\frac{2}{x}$, so a primitive is $F(x) = -2 \log x = -\log(x^2)$. Therefore

$$u(x) = Ae^{\log(x^2)} = Ax^2, \quad A \in \mathbb{R}.$$

3. *Trigonometric coefficient.* Consider the differential equation

$$u'(x) + \sin x u(x) = 0.$$

Here $f(x) = \sin x$, so a primitive is $F(x) = -\cos x$. Hence

$$y(x) = Ae^{-F(x)} = Ae^{\cos x}, \quad A \in \mathbb{R}.$$

Next, we consider the non-homogeneous linear first-order ODE

$$u'(x) + f(x)u(x) = g(x) \quad \forall x \in I, \quad (8.3)$$

where $f, g : I \rightarrow \mathbb{R}$ are given continuous functions.

To motivate the solution formula, we use the method of **variation of constants**. For the homogeneous equation we know that every solution has the form $Ae^{-F(x)}$ with $A \in \mathbb{R}$. The idea is to *replace this constant by a function* and look for a solution of the non-homogeneous equation of the form

$$u(x) = H(x)e^{-F(x)},$$

for some function $H \in C^1(I)$. With this choice,

$$u'(x) = H'(x)e^{-F(x)} - F'(x)H(x)e^{-F(x)} = H'(x)e^{-F(x)} - f(x)u(x).$$

Thus u solves (8.3) if and only if

$$H'(x)e^{-F(x)} = g(x),$$

which means that H must be a primitive of the function $g(x)e^{F(x)}$.

This motivates the general solution formula:

THEOREM 8.6: NON-HOMOGENEOUS LINEAR 1ST ORDER ODES

Let $f, g : I \rightarrow \mathbb{R}$ be continuous, and consider the non-homogeneous first-order linear ODE (8.3). Let $F : I \rightarrow \mathbb{R}$ be a primitive of f , and let $H : I \rightarrow \mathbb{R}$ be a primitive of ge^F . Then every C^1 solution $u : I \rightarrow \mathbb{R}$ of (8.3) is of the form

$$u(x) = H(x)e^{-F(x)} + Ae^{-F(x)}, \quad A \in \mathbb{R}.$$

In particular, the set of solutions of (8.3) is a one-dimensional affine subspace of $C^1(I)$.

Proof. If H is a primitive of ge^F , then $H + A$ is also a primitive for any constant A . Hence, by the same computation as the one performed above, it follows that

$$u(x) = (H(x) + A)e^{-F(x)}$$

solves (8.3). Indeed

$$\begin{aligned} u'(x) &= (H(x) + A)'e^{-F(x)} - F'(x)(H(x) + A)e^{-F(x)} \\ &= H'(x)e^{-F(x)} - f(x)u(x) = g(x) - f(x)u(x). \end{aligned}$$

Conversely, let u be any solution of (8.3), and set $v(x) = u(x) - H(x)e^{-F(x)}$. Then

$$\begin{aligned} v'(x) &= u'(x) - H'(x)e^{-F(x)} - F'(x)H(x)e^{-F(x)} \\ &= -f(x)u(x) + g(x) - g(x)e^{F(x)}e^{-F(x)} + f(x)H(x)e^{-F(x)} \\ &= -f(x)u(x) + g(x) - g(x) + f(x)(u(x) - v(x)) = -f(x)v(x). \end{aligned}$$

Thus v solves the homogeneous equation (8.2). By Proposition 8.3, we have $v(x) = Ae^{-F(x)}$ for some constant A . Therefore,

$$u(x) = v(x) + H(x)e^{-F(x)} = Ae^{-F(x)} + H(x)e^{-F(x)},$$

which proves the result. \square

33

The previous results give explicit formulas to solve every linear first-order ODE of the form $u' + fu = g$. In concrete situations, the difficulty lies in computing a primitive F of f and then a primitive of $g(x)e^{F(x)}$.

As we have seen, solutions depend on a free parameter $A \in \mathbb{R}$. This allows us to impose an initial condition of the form $u(x_0) = u_0$, which uniquely determines A .

EXAMPLE 8.7. — We solve the ODE

$$u'(x) - 2xu(x) = e^{x^2}, \quad u(0) = 1, \quad (8.4)$$

on \mathbb{R} . Here u is the unknown, and $f(x) = -2x$, $g(x) = e^{x^2}$ are given. According to Theorem 8.6, we first find a primitive of f :

$$F(x) = -x^2.$$

Then we consider

$$g(x)e^{F(x)} = e^{x^2}e^{-x^2} = 1,$$

whose primitive is $H(x) = x$. Thus u must be of the form

$$u(x) = (x + A)e^{x^2}.$$

Imposing the condition $u(0) = 1$ gives $A = 1$, hence

$$u(x) = (x + 1)e^{x^2}. \quad (8.5)$$

REMARK 8.8. — If one forgets the formula from Theorem 8.6, it is enough to remember the following procedure for solving (8.3). We start from

$$u'(x) + f(x)u(x) = g(x),$$

and multiply both sides by a function $e^{w(x)}$:

$$u'(x)e^{w(x)} + f(x)u(x)e^{w(x)} = g(x)e^{w(x)}.$$

We look for w such that the left-hand side is the derivative of $u(x)e^{w(x)}$, that is

$$(u(x)e^{w(x)})' = u'(x)e^{w(x)} + w'(x)u(x)e^{w(x)}.$$

So we require

$$w'(x) = f(x).$$

If we choose $w = F$ to be any primitive of f (the additive constant does not matter), then

$$(u(x)e^{F(x)})' = g(x)e^{F(x)},$$

and therefore

$$u(x)e^{F(x)} = \int ge^F + A,$$

for some $A \in \mathbb{R}$. Thus, if H is a primitive of ge^F , this reads

$$u(x)e^{F(x)} = H(x) + A \quad \implies \quad u(x) = H(x)e^{-F(x)} + Ae^{-F(x)}.$$

EXAMPLE 8.9. — We revisit Example 8.7: we want to solve the ODE

$$u'(x) - 2xu(x) = e^{x^2}, \quad u(0) = 1,$$

using only the integrating-factor method described in Remark 8.8.

We multiply both sides by $e^{w(x)}$ and choose w such that

$$(u(x)e^{w(x)})' = u'(x)e^{w(x)} + w'(x)u(x)e^{w(x)}$$

reproduces the left-hand side of the equation. Thus we require $w'(x) = f(x) = -2x$, and therefore we may take $w(x) = -x^2$. Multiplying the ODE by e^{-x^2} gives

$$(u(x)e^{-x^2})' = e^{-x^2}(u'(x) - 2xu(x)) = e^{-x^2}e^{x^2} = 1.$$

Integrating, we get

$$u(x)e^{-x^2} = x + A \implies u(x) = (x + A)e^{x^2}.$$

Imposing $u(0) = 1$ gives $A = 1$, so

$$u(x) = (x + 1)e^{x^2}.$$

EXERCISE 8.10. — Find a solution of

$$u'(x) - \left(\frac{4}{x} + 1\right)u(x) = x^4, \quad u(1) = 1,$$

in the interval $(0, \infty)$.

8.1.2 Autonomous First Order ODEs

We next study *autonomous* first-order ODEs, where the rate of change of the unknown depends only on its current value, and not explicitly on x :

$$u'(x) = f(u(x)), \tag{8.6}$$

where $f : \mathbb{R} \rightarrow \mathbb{R}$ is a given continuous function, and u is the unknown. The function f tells us how u should change depending on its current value.

A standard way to solve such equations is the method of **separation of variables**. If $u(x) = C \in \mathbb{R}$ for some C such that $f(C) = 0$, then the constant function $u = C$ is a solution of the ODE. Otherwise, if $f(u(x)) \neq 0$, we can divide both sides by $f(u(x))$:

$$\frac{u'(x)}{f(u(x))} = 1.$$

Integrating and using the substitution formula (7.8), we obtain

$$\int \frac{1}{f(u)} du = \int \frac{1}{f(u(x))} u'(x) dx = \int 1 dx = x + A, \tag{8.7}$$

where A is a constant of integration.

If H is a primitive of $1/f$, this reads

$$H(u(x)) = x + A \implies u(x) = H^{-1}(x + A).$$

Since we assumed $f(u(x)) \neq 0$ on the interval under consideration, we have $H' = \frac{1}{f} \neq 0$ there, so H is strictly monotone and therefore invertible on that interval.

EXAMPLE 8.11. — We study the logistic growth model, treating it as an autonomous first-order ODE and solving it explicitly.

Consider

$$u'(x) = r u(x) \left(1 - \frac{u(x)}{K} \right), \quad (8.8)$$

where $r > 0$ is the growth rate and $K > 0$ is the *carrying capacity*. Here $u(x)$ denotes the population at “time” x . From the modelling point of view one is usually interested in solutions with

$$0 \leq u(x) \leq K,$$

since $u < 0$ has no meaning as a population, and K represents an upper bound (the maximal sustainable population).

Mathematically, the right-hand side vanishes at $u = 0$ and $u = K$, so the constant functions $u(x) = 0$ and $u(x) = K$ are solutions. To solve the ODE by separation of variables we assume $0 < u(x) < K$ on the interval under consideration (so that we may safely divide by u and $K - u$).

We rearrange and integrate:

$$\frac{K u'(x)}{u(x)(K - u(x))} = r \quad \Longrightarrow \quad \int \frac{K}{u(K - u)} du = \int r dx = rx + A.$$

34

We decompose the rational function

$$\frac{K}{u(K - u)} = \frac{1}{u} + \frac{1}{K - u},$$

so that

$$\int \frac{K}{u(K - u)} du = \int \frac{1}{u} du + \int \frac{1}{K - u} du = \log u - \log(K - u) = \log \left(\frac{u}{K - u} \right).$$

Thus

$$\begin{aligned} \log \left(\frac{u(x)}{K - u(x)} \right) = rx + A &\Longrightarrow \frac{u(x)}{K - u(x)} = e^{rx+A} \\ &\Longrightarrow u(x) = K e^{rx+A} - u(x) e^{rx+A} \\ &\Longrightarrow (1 + e^{rx+A}) u(x) = K e^{rx+A} \\ &\Longrightarrow u(x) = K \frac{e^{rx+A}}{1 + e^{rx+A}}. \end{aligned}$$

If $u_0 \in (0, K)$ denotes the initial population at $x = 0$, then

$$u_0 = K \frac{e^A}{1 + e^A} \quad \Longrightarrow \quad e^A = \frac{u_0}{K - u_0}.$$

Substituting back gives the explicit solution

$$u(x) = \frac{Ku_0}{u_0 + (K - u_0)e^{-rx}}.$$

It is useful to briefly discuss the different initial conditions:

- If $u(0) = 0$, then $u(x) = 0$ is a (constant) solution. In other words, if the population is zero at the initial time, then it remains zero for all times.
- If $u(0) = K$, then $u(x) = K$ is a (constant) solution. In this case the population stays exactly at the carrying capacity.
- If $0 < u(0) < K$, the above formula applies and one checks that $0 < u(x) < K$ for all x , with $u(x) \rightarrow K$ as $x \rightarrow \infty$. The population grows and approaches the carrying capacity from below.
- If $u(0) > K$, then $u'(0) < 0$, so u decreases. One can check that the same explicit formula holds also in this case (with $u_0 > K$). In particular, $u(x) > K$ for all x , and $u(x) \rightarrow K$ as $x \rightarrow \infty$: the population decreases towards the carrying capacity from above.

The method of separation of variables applies also to equations of the form

$$u'(x) = f(u(x))g(x),$$

where $f, g : \mathbb{R} \rightarrow \mathbb{R}$ are continuous. Assuming $f(u(x)) \neq 0$ on a suitable interval, we divide:

$$\frac{u'(x)}{f(u(x))} = g(x).$$

Integrating,

$$\int \frac{1}{f(u)} du = \int g(x) dx + A.$$

If H is a primitive of $1/f$ and G a primitive of g , then

$$H(u(x)) = G(x) + A \implies u(x) = H^{-1}(G(x) + A).$$

We summarize this discussion in the next:

THEOREM 8.12: SEPARABLE EQUATIONS

Let $I, J \subseteq \mathbb{R}$ be intervals, let $f : J \rightarrow \mathbb{R}$ and $g : I \rightarrow \mathbb{R}$ be continuous, and assume $f(y) \neq 0$ for all $y \in J$. Let H be a primitive of $1/f$ on J , and let G be a primitive of g on I . Then every C^1 solutions $u : I \rightarrow J$ of

$$u'(x) = f(u(x)) g(x) \quad (8.9)$$

is of the form

$$u(x) = H^{-1}(G(x) + A), \quad A \in \mathbb{R}.$$

Proof. Since f never vanishes, $H' = 1/f$ has constant sign. Hence H is strictly monotone and therefore invertible.

If $u(x) = H^{-1}(G(x) + A)$, then by the chain rule (Theorem 5.14) and the formula for the derivative of the inverse function (Theorem 5.20), we get

$$u'(x) = (H^{-1})'(G(x) + A) G'(x) = \frac{1}{H' \circ H^{-1}(G(x) + A)} g(x) = \frac{1}{H'(u(x))} g(x) = f(u(x)) g(x),$$

so u solves (8.9)

Conversely, suppose u solves (8.9). Then

$$(H \circ u(x))' = H'(u(x)) u'(x) = \frac{1}{f(u(x))} f(u(x)) g(x) = g(x) = G'(x).$$

Thus $H(u(x)) - G(x)$ has derivative zero, and is therefore constant:

$$H(u(x)) = G(x) + A \quad \text{for some } A \in \mathbb{R}.$$

Applying H^{-1} gives

$$u(x) = H^{-1}(G(x) + A),$$

which completes the proof. \square

EXAMPLE 8.13. — We solve the separable equation

$$u'(x) = x u(x)^2$$

on an interval where u does not vanish. We rewrite the equation by separating variables:

$$\frac{u'}{u^2} = x.$$

Integrating both sides gives

$$\int \frac{du}{u^2} = \int x \, dx, \quad \implies \quad -\frac{1}{u} = \frac{x^2}{2} + A, \quad A \in \mathbb{R}.$$

Solving for u yields the family of solutions

$$u(x) = -\frac{1}{\frac{x^2}{2} + A}, \quad A \in \mathbb{R}.$$

If an initial condition $u(x_0) = u_0$ (with $u_0 \neq 0$) is prescribed, we determine the constant A by inserting $x = x_0$:

$$-\frac{1}{u_0} = \frac{x_0^2}{2} + A \implies A = -\frac{1}{u_0} - \frac{x_0^2}{2}.$$

Thus the unique solution with $u(x_0) = u_0$ is

$$u(x) = -\frac{1}{\frac{x^2}{2} - \frac{1}{u_0} - \frac{x_0^2}{2}}.$$

8.1.3 Homogeneous Linear Second Order ODEs with Constant Coefficients

We now move to second-order linear ODEs. These are considerably more difficult to solve than first-order ones in general, so we start with the simplest case: homogeneous equations with constant coefficients,

$$u''(x) + a_1 u'(x) + a_0 u(x) = 0, \quad (8.10)$$

where $a_0, a_1 \in \mathbb{R}$ are given constants and u is the unknown. Such equations already cover many important applications (for instance, oscillations and damped vibrations).

EXAMPLE 8.14. — • For $u'' = 0$, affine functions are solutions: $u(x) = Ax + B$, with $A, B \in \mathbb{R}$.

- For $u'' - u = 0$, e^x and e^{-x} are solutions, therefore

$$u(x) = Ae^x + Be^{-x}, \quad A, B \in \mathbb{R},$$

is a solution.

- For $u'' + u = 0$, \sin and \cos are solutions, therefore

$$u(x) = A \sin(x) + B \cos(x), \quad A, B \in \mathbb{R}.$$

is a solution. Since sine and cosine can be written in terms of $e^{\pm ix}$, one can also rewrite the solutions above as

$$u(x) = Ce^{ix} + De^{-ix}, \quad C, D \in \mathbb{C},$$

and then re-express them in terms of real-valued sine and cosine (recall that we are interested in real-valued functions).

EXERCISE 8.15. — Check the assertions in Example 8.14.

The last two examples suggest looking for solutions of (8.10) of the form

$$u(x) = e^{\alpha x}, \quad \alpha \in \mathbb{C}.$$

With this choice,

$$u''(x) + a_1 u'(x) + a_0 u(x) = (\alpha^2 + a_1 \alpha + a_0)u(x),$$

so u is a solution if and only if

$$\alpha^2 + a_1 \alpha + a_0 = 0.$$

The quadratic polynomial

$$p(t) = t^2 + a_1 t + a_0$$

is called the **characteristic polynomial**. Its roots determine the shape of the solutions. We distinguish three cases according to the discriminant $\Delta = a_1^2 - 4a_0$.

- Case 1: $\Delta > 0$. The polynomial $p(t)$ has two distinct real roots

$$\alpha = \frac{-a_1 + \sqrt{\Delta}}{2}, \quad \beta = \frac{-a_1 - \sqrt{\Delta}}{2}. \quad (8.11)$$

Then $x \mapsto e^{\alpha x}$ and $x \mapsto e^{\beta x}$ are two linearly independent solutions, and therefore

$$u(x) = Ae^{\alpha x} + Be^{\beta x}, \quad A, B \in \mathbb{R},$$

is a solution of (8.10).

- Case 2: $\Delta < 0$. Then $p(t)$ has two complex-conjugate roots

$$\alpha + i\beta = -\frac{a_1}{2} + i\frac{\sqrt{-\Delta}}{2}, \quad \alpha - i\beta = -\frac{a_1}{2} - i\frac{\sqrt{-\Delta}}{2}, \quad (8.12)$$

with $\beta > 0$. The complex-valued functions $x \mapsto e^{(\alpha \pm i\beta)x}$ solve (8.10), and hence their real and imaginary parts are real solutions. This gives

$$u(x) = Ae^{\alpha x} \sin(\beta x) + Be^{\alpha x} \cos(\beta x), \quad A, B \in \mathbb{R}.$$

- Case 3: $\Delta = 0$. Then $p(t)$ has a double real root

$$\alpha = -\frac{a_1}{2}, \quad (8.13)$$

so $x \mapsto e^{\alpha x}$ is a solution of (8.10). To find another independent solution, recall the special case $u'' = 0$, where two linearly independent solutions are 1 and x , which can be written as $e^{\gamma x}$ and $xe^{\gamma x}$ with $\gamma = 0$.

This suggests that $x \mapsto xe^{\alpha x}$ might be a solution. Indeed,

$$(xe^{\alpha x})'' + a_1(xe^{\alpha x})' + a_0xe^{\alpha x} = \underbrace{(\alpha^2 + a_1\alpha + a_0)}_{=0}xe^{\alpha x} + \underbrace{(2\alpha + a_1)}_{=0}e^{\alpha x} = 0,$$

where the first term vanishes because α is a root of p , and the second vanishes by (8.13).

Hence

$$u(x) = Ae^{\alpha x} + Bxe^{\alpha x}, \quad A, B \in \mathbb{R},$$

solves (8.10).

For second-order ODEs it is customary to prescribe both the value of u and the value of its derivative at some point (for instance, $u(0) = 1$ and $u'(0) = 0$). The two constants A, B in the formulas above are precisely what we need in order to satisfy two such conditions.

We now want to prove that the families of solutions described in Paragraph 8.1.3 indeed give *all* solutions of (8.10).

THEOREM 8.16: EXISTENCE AND UNIQUENESS: THE HOMOGENEOUS CASE

Given $a_0, a_1 \in \mathbb{R}$, let $\Delta = a_1^2 - 4a_0$ and consider the following solutions of (8.10):

$$\begin{array}{lll} \underline{\Delta > 0}: & u_1(x) = e^{\alpha x}, & u_2(x) = e^{\beta x}, & \alpha, \beta \text{ as in (8.11),} \\ \underline{\Delta < 0}: & u_1(x) = e^{\alpha x} \sin(\beta x), & u_2(x) = e^{\alpha x} \cos(\beta x), & \alpha, \beta \text{ as in (8.12),} \\ \underline{\Delta = 0}: & u_1(x) = e^{\alpha x}, & u_2(x) = xe^{\alpha x}, & \alpha \text{ as in (8.13).} \end{array}$$

If $u \in C^2(I)$ solves (8.10), then there exist $A, B \in \mathbb{R}$ such that

$$u = Au_1 + Bu_2.$$

In other words, the set of solutions of (8.10) forms a two-dimensional linear subspace of $C^2(I)$.

Proof. We treat the case $\Delta > 0$ (the other cases are similar). Assume for simplicity that $0 \in I$ (otherwise, fix $x_0 \in I$ and repeat the argument with x_0 in place of 0).

Observing that

$$u_1(0) = u_2(0) = 1, \quad u_1'(0) = \alpha > \beta = u_2'(0),$$

by defining

$$v_1(x) = \frac{\alpha u_2(x) - \beta u_1(x)}{\alpha - \beta}, \quad v_2(x) = \frac{u_1(x) - u_2(x)}{\alpha - \beta},$$

it follows that v_1 and v_2 are solutions of (8.10) that satisfy

$$v_1(0) = 1, \quad v_1'(0) = 0, \quad v_2(0) = 0, \quad v_2'(0) = 1.$$

Now, given $u \in C^2(I)$ solving (8.10), set

$$w(x) = u(x) - u(0)v_1(x).$$

Then, w is a solution satisfying $w(0) = 0$. Our goal is to show that w is a multiple of u_2 . For this, consider the function

$$W(x) = w(x)v_2'(x) - w'(x)v_2(x),$$

called the *Wronskian* of w and v_2 , and compute $W'(x)$. Differentiating gives

$$W'(x) = w'(x)v_2'(x) + w(x)v_2''(x) - w''(x)v_2(x) - w'(x)v_2'(x) = w(x)v_2''(x) - w''(x)v_2(x).$$

Since both w and v_2 solve (8.10), we have

$$w'' = -a_1w' - a_0w, \quad v_2'' = -a_1v_2' - a_0v_2.$$

Substituting into the expression for W' gives

$$\begin{aligned} W'(x) &= w(-a_1v_2' - a_0v_2) - (-a_1w' - a_0w)v_2 \\ &= -a_1wv_2' - a_0wv_2 + a_1w'v_2 + a_0wv_2 \\ &= a_1(w'v_2 - wv_2') = -a_1W(x). \end{aligned}$$

35

Thus W satisfies the first-order linear ODE

$$W'(x) + aW(x) = 0, \quad W(0) = 0.$$

By Theorem 8.3 $W(x) = Ae^{-a_1x}$, and the condition $W(0) = 0$ implies that $W = 0$ on I . In other words,

$$w(x)v_2'(x) - w'(x)v_2(x) = 0 \quad \forall x \in I.$$

Now, on any interval J where v_2 does not vanish, we can rewrite this as

$$w'(x) = \frac{v_2'(x)}{v_2(x)}w(x) \quad \forall x \in I \cap J.$$

Thus, since $\frac{v_2'}{v_2} = (\log |v_2|)'$, we can apply Theorem 8.3 to deduce the existence of a constant $A \in \mathbb{R}$ such that

$$w(x) = Ae^{\log |v_2(x)|} = A|v_2(x)| \quad \forall x \in I \cap J.$$

Since $|v_2|$ is either equal to v_2 or to $-v_2$ on J (because v_2 does not vanish), we proved that on each interval J where v_2 does not vanish, there exists a constant $A_J \in \mathbb{R}$ such that

$$w(x) = A_J v_2(x).$$

In our situation, one can check that v_2 vanishes only at 0, so there exist $A_-, A_+ \in \mathbb{R}$ such that

$$w = A_- v_2 \quad \text{on } (-\infty, 0) \cap I, \quad w = A_+ v_2 \quad \text{on } (0, \infty) \cap I.$$

If either $(-\infty, 0) \cap I$ or $(0, \infty) \cap I$ are empty, we have proved that w is a multiple of v_2 on I . Otherwise we observe that, by the continuity of w and v_2 , recalling that $v_2(0) = 0$ we get

$$w(0) = \lim_{x \rightarrow 0^-} w(x) = \lim_{x \rightarrow 0^-} A_- v_2(x) = A_- v_2(0) = 0,$$

therefore,

$$w = A_- v_2 \quad \text{on } (-\infty, 0] \cap I, \quad w = A_+ v_2 \quad \text{on } [0, \infty) \cap I.$$

Thus, recalling that $v_2'(0) = 1$,

$$w'_-(0) = \lim_{h \rightarrow 0^+} \frac{w(-h) - w(0)}{h} = \lim_{h \rightarrow 0^+} \frac{A_- v_2(-h) - A_- v_2(0)}{h} = A_- (v_2)'_-(0) = A_-,$$

and analogously $w'_+(0) = A_+ v_2'(0) = A_+$. Since $w \in C^2(I)$ we have that $w'_-(0) = w'_+(0)$, so $A_- = A_+$, and therefore $w(x) = A_+ v_2(x)$ on all of I .

In conclusion

$$u(x) = u(0)v_1(x) + A_+ v_2(x) = u(0) \frac{\alpha u_2(x) - \beta u_1(x)}{\alpha - \beta} + A_+ \frac{u_1(x) - u_2(x)}{\alpha - \beta},$$

which is a linear combination of u_1 and u_2 , as claimed. \square

REMARK 8.17 (Zero initial data). — From the explicit formulas in Theorem 8.16 one can easily check that the only solution $u \in C^2(I)$ of the homogeneous equation $u'' + a_1 u' + a_0 u = 0$ satisfying $u(x_0) = u'(x_0) = 0$ for some $x_0 \in I$ is the zero solution.

For instance, in the case $\Delta > 0$ we have

$$u(x) = A e^{\alpha x} + B e^{\beta x}, \quad \alpha \neq \beta,$$

so

$$u(x_0) = 0, \quad u'(x_0) = 0 \quad \implies \quad \begin{cases} A e^{\alpha x_0} + B e^{\beta x_0} = 0, \\ A \alpha e^{\alpha x_0} + B \beta e^{\beta x_0} = 0. \end{cases}$$

Multiplying the first equation by α and subtracting from the second gives $(\beta - \alpha) B e^{\beta x_0} = 0$, therefore $B = 0$. Going back to the first equation, we also get $A = 0$.

The cases $\Delta = 0$ and $\Delta < 0$ are handled similarly. In conclusion, no nontrivial solution can vanish together with its derivative at a point.

As a consequence of the previous remark, we obtain the following.

COROLLARY 8.18: WRONSKIAN AND LINEAR DEPENDENCE

Let $v_1, v_2 \in C^2(I)$ be solutions of (8.10) on an interval $I \subseteq \mathbb{R}$, and let

$$W(x) = v_1(x)v_2'(x) - v_1'(x)v_2(x)$$

be their Wronskian. If $W(x_0) = 0$ for some $x_0 \in I$, then v_1 and v_2 are linearly dependent on I .

Proof. Consider the 2×2 matrix

$$M(x_0) = \begin{pmatrix} v_1(x_0) & v_2(x_0) \\ v_1'(x_0) & v_2'(x_0) \end{pmatrix}.$$

The condition $W(x_0) = 0$ means precisely that $\det M(x_0) = 0$, so there exists a nonzero vector $(A, B) \in \mathbb{R}^2$ such that

$$A v_1(x_0) + B v_2(x_0) = 0, \quad A v_1'(x_0) + B v_2'(x_0) = 0.$$

35

Define

$$w(x) = A v_1(x) + B v_2(x).$$

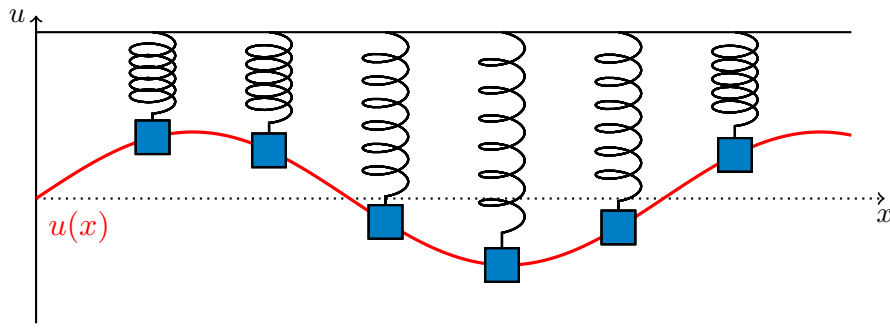
Then w solves (8.10) and satisfies

$$w(x_0) = 0, \quad w'(x_0) = 0.$$

By Remark 8.17, the only solution with these initial data is the trivial one, so $w = 0$ on I . This proves that $A v_1 + B v_2 = 0$, thus v_1 and v_2 are linearly dependent on I . \square

EXAMPLE 8.19. — We consider a vertical spring whose upper end is attached to a fixed ceiling, a weight of mass m attached to its lower end, and the mass can move only in the vertical direction. There is a unique position where the spring force balances gravity; we call this the *equilibrium position* and denote it by $u = 0$. If x denotes time, we write $u(x)$ for the vertical displacement of the mass from this equilibrium position at time x (with upward displacements taken as positive).

By Newton's law, the mass m multiplied by the acceleration $u''(x)$ equals the total force acting on the mass.



A first contribution to the total force comes from the spring: according to Hooke's law, the restoring force is $-ku$, where $k > 0$ is the spring constant. In addition, friction acts against the motion; we assume this friction force is proportional to the velocity and given by $-du'$, where $d \geq 0$ is the damping constant.

The equation of motion is therefore

$$mu'' = -du' - ku,$$

or equivalently

$$u'' + \frac{d}{m}u' + \frac{k}{m}u = 0.$$

This is a homogeneous linear ODE of second order with constant coefficients, with unknown u and given constants m, d, k . Setting

$$\omega = \sqrt{\frac{k}{m}}, \quad \zeta = \frac{d}{2m\omega},$$

we can rewrite it as

$$u'' + 2\zeta\omega u' + \omega^2 u = 0$$

(see also Example 8.2(2)). The characteristic polynomial is

$$p(t) = t^2 + 2\zeta\omega t + \omega^2$$

with discriminant $\Delta = 4(\zeta^2 - 1)\omega^2$.

If $\Delta < 0$ (equivalently $\zeta < 1$), the friction is small compared to the spring strength and the system oscillates:

$$u(x) = e^{-\zeta\omega x} (A \sin(\gamma x) + B \cos(\gamma x)), \quad \gamma = \sqrt{1 - \zeta^2} \omega.$$

The constants A, B are determined by the initial position $u(0)$ and the initial velocity $u'(0)$. In the special case $\zeta = 0$ there is no damping, and u is periodic.

If friction is large compared to the spring strength ($\zeta \geq 1$, so $\Delta \geq 0$), the oscillations disappear, and the weight returns monotonically to equilibrium:

- If $\zeta > 1$, then

$$u(x) = Ae^{-\lambda_1 x} + Be^{-\lambda_2 x}, \quad \lambda_{1,2} = \left(\zeta \pm \sqrt{\zeta^2 - 1}\right) \omega.$$

- If $\zeta = 1$, then

$$u(x) = Ae^{-\omega x} + Bxe^{-\omega x}.$$

One can check that $\zeta - \sqrt{\zeta^2 - 1} < 1$ for all $\zeta > 1$, so the fastest exponential convergence is achieved when $\zeta = 1$. This type of behavior is desirable, for example, in door-closing mechanisms. Here again the constants A, B are fixed by prescribing suitable initial conditions.

8.1.4 Non-Homogeneous Linear Second Order ODEs with Constant Coefficients

We now add a forcing term and consider the non-homogeneous linear second-order ODE with constant coefficients

$$u''(x) + a_1 u'(x) + a_0 u(x) = g(x), \quad (8.14)$$

where $a_0, a_1 \in \mathbb{R}$ are constants and $g \in C^0(I)$ is a given function. The unknown is again u .

With the notation of Paragraph 8.1.3, let u_1, u_2 be two linearly independent solutions of the homogeneous equation (8.10):

$$\begin{aligned} \underline{\Delta > 0}: \quad & u_1(x) = e^{\alpha x}, & u_2(x) = e^{\beta x}, & \alpha, \beta \text{ as in (8.11),} \\ \underline{\Delta < 0}: \quad & u_1(x) = e^{\alpha x} \sin(\beta x), & u_2(x) = e^{\alpha x} \cos(\beta x), & \alpha, \beta \text{ as in (8.12),} \\ \underline{\Delta = 0}: \quad & u_1(x) = e^{\alpha x}, & u_2(x) = xe^{\alpha x}, & \alpha \text{ as in (8.13).} \end{aligned}$$

We look for a solution u of (8.14) of the form

$$u(x) = H_1(x)u_1(x) + H_2(x)u_2(x),$$

where H_1, H_2 are unknown functions. This is the method of **variation of constants** in the second-order setting: we know the solutions of the homogeneous equation, and we allow the coefficients in the linear combination to depend on x .

We compute

$$u' = (H_1' u_1 + H_2' u_2) + (H_1 u_1' + H_2 u_2'),$$

and

$$u'' = (H_1' u_1 + H_2' u_2)' + (H_1 u_1' + H_2 u_2')' + (H_1 u_1'' + H_2 u_2'').$$

Hence

$$\begin{aligned} u'' + a_1 u' + a_0 u &= (H_1' u_1 + H_2' u_2)' + (H_1 u_1' + H_2 u_2')' + (H_1 u_1'' + H_2 u_2'') \\ &\quad + a_1 (H_1' u_1 + H_2' u_2) + a_1 (H_1 u_1' + H_2 u_2') + a_0 (H_1 u_1 + H_2 u_2). \end{aligned}$$

Since u_1 and u_2 solve the homogeneous equation, we have

$$H_1(u_1'' + a_1u_1' + a_0u_1) = 0, \quad H_2(u_2'' + a_1u_2' + a_0u_2) = 0,$$

therefore

$$u'' + a_1u' + a_0u = (H_1'u_1 + H_2'u_2)' + (H_1'u_1' + H_2'u_2') + a_1(H_1'u_1 + H_2'u_2).$$

For u to solve the non-homogeneous ODE (8.14), we can choose H_1' and H_2' so that

$$H_1'(x)u_1(x) + H_2'(x)u_2(x) = 0, \quad H_1'(x)u_1'(x) + H_2'(x)u_2'(x) = g(x). \quad (8.15)$$

From the first equation, we can write

$$H_2' = -\frac{u_1}{u_2}H_1'. \quad (8.16)$$

Substituting into the second equation gives

$$H_1'u_1' - \frac{u_1}{u_2}u_2'H_1' = g \quad \implies \quad H_1' = \frac{u_2 g}{u_1'u_2 - u_2'u_1}.$$

Inserting this into (8.16) yields

$$H_2' = -\frac{u_1}{u_2} \frac{u_2 g}{u_1'u_2 - u_2'u_1} = \frac{u_1 g}{u_2'u_1 - u_1'u_2}.$$

Note that the denominator

$$u_1'u_2 - u_2'u_1 = -(u_1u_2' - u_1'u_2)$$

is (up to sign) the Wronskian of u_1 and u_2 . Since u_1 and u_2 are linearly independent solutions of the homogeneous equation, Corollary 8.18 implies that this Wronskian never vanishes on I . In particular, the above formulas for H_1' and H_2' are well-defined on all of I . Therefore,

$$H_1 = \int \frac{u_2 g}{u_1'u_2 - u_2'u_1} dx, \quad H_2 = \int \frac{u_1 g}{u_2'u_1 - u_1'u_2} dx.$$

This shows that if H_1 and H_2 are primitives of the functions above, then

$$u = H_1u_1 + H_2u_2$$

is a particular solution of (8.14). Finally, the general solution is obtained by adding any solution of the homogeneous equation:

$$u = Au_1 + Bu_2 + H_1u_1 + H_2u_2, \quad A, B \in \mathbb{R}.$$

We summarize this in the following proposition.

THEOREM 8.20: EXISTENCE AND UNIQUENESS: THE NON-HOMOGENEOUS CASE

Given $a_0, a_1 \in \mathbb{R}$, let $\Delta = a_1^2 - 4a_0$ and consider the following solutions of (8.10):

$$\begin{aligned} \underline{\Delta > 0}: & \quad u_1(x) = e^{\alpha x}, & u_2(x) = e^{\beta x}, & \quad \alpha, \beta \text{ as in (8.11),} \\ \underline{\Delta < 0}: & \quad u_1(x) = e^{\alpha x} \sin(\beta x), & u_2(x) = e^{\alpha x} \cos(\beta x), & \quad \alpha, \beta \text{ as in (8.12),} \\ \underline{\Delta = 0}: & \quad u_1(x) = e^{\alpha x}, & u_2(x) = xe^{\alpha x}, & \quad \alpha \text{ as in (8.13).} \end{aligned}$$

Let H_1 and H_2 be primitives of $\frac{u_2 g}{u_1' u_2 - u_2' u_1}$ and $\frac{u_1 g}{u_2' u_1 - u_1' u_2}$, respectively. If $u \in C^2(I)$ solves (8.14), then there exist $A, B \in \mathbb{R}$ such that

$$u = Au_1 + Bu_2 + H_1 u_1 + H_2 u_2.$$

In other words, the set of solutions of (8.14) forms a two-dimensional affine subspace of $C^2(I)$.

Proof. First we show existence. By the computation above, if H_1 and H_2 are primitives of

$$\frac{u_2 g}{u_1' u_2 - u_2' u_1} \quad \text{and} \quad \frac{u_1 g}{u_2' u_1 - u_1' u_2},$$

then the function

$$u_p = H_1 u_1 + H_2 u_2$$

satisfies

$$u_p'' + a_1 u_p' + a_0 u_p = g,$$

so u_p is a particular solution of (8.14). Since the equation is linear, for any $A, B \in \mathbb{R}$ the function

$$u = Au_1 + Bu_2 + u_p$$

also solves (8.14). This proves existence of solutions of the stated form.

For uniqueness, let $u \in C^2(I)$ be any solution of (8.14), and define

$$v = u - u_p = u - (H_1 u_1 + H_2 u_2).$$

Then v solves the homogeneous equation (8.10) and therefore, by Theorem 8.16, there exist $A, B \in \mathbb{R}$ such that

$$v = Au_1 + Bu_2.$$

Hence $u = v + u_p = Au_1 + Bu_2 + H_1 u_1 + H_2 u_2$, as desired. □

Although this method is very general, in practice the integrals defining H_1 and H_2 can be quite complicated. In some (very special) situations it is easier to *guess* a particular solution

by trying functions of the form

$$p(x)e^{\gamma x}, \quad p(x)e^{\gamma x} \cos(\eta x), \quad p(x)e^{\gamma x} \sin(\eta x),$$

where p is a polynomial and $\gamma, \eta > 0$ are chosen to reflect the structure of g .

EXAMPLE 8.21. — Solve

$$u''(x) + u(x) = 1, \quad u(0) = 0, \quad u'(0) = 1.$$

The homogeneous equation $u'' + u = 0$ has solutions $A \cos(x) + B \sin(x)$, so we look for a solution of the form

$$u(x) = H_1(x) \cos(x) + H_2(x) \sin(x),$$

where H_1, H_2 are unknown. This is exactly the method of variation of constants specialized to this case.

The computation leads to the system

$$H_1'(x) \cos(x) + H_2'(x) \sin(x) = 0, \quad -H_1'(x) \sin(x) + H_2'(x) \cos(x) = 1.$$

Solving for H_1' and H_2' as in the general method gives

$$H_1 = - \int \frac{\sin(x)}{\sin^2(x) + \cos^2(x)} dx = - \int \sin(x) dx,$$

$$H_2 = \int \frac{\cos(x)}{\cos^2(x) + \sin^2(x)} dx = \int \cos(x) dx.$$

Hence we may take $H_1 = \cos(x)$ and $H_2 = \sin(x)$, which gives a particular solution

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

(In this case, one could also have tried to guess this particular solution!) The general solution is therefore

$$u(x) = 1 + A \cos(x) + B \sin(x).$$

Imposing $u(0) = 0$ and $u'(0) = 1$ (two initial conditions for the unknown u) yields

$$u(x) = 1 - \cos(x) + \sin(x).$$

EXAMPLE 8.22. — Solve

$$u''(x) + u(x) = \sin(x), \quad u(0) = 0, \quad u'(0) = 1.$$

Again, the homogeneous solutions are $A \cos(x) + B \sin(x)$. The method of variation of constants gives

$$H_1 = - \int \frac{\sin(x) \sin(x)}{\sin^2(x) + \cos^2(x)} dx = - \int \sin^2(x) dx,$$

$$H_2 = \int \frac{\cos(x) \sin(x)}{\cos^2(x) + \sin^2(x)} dx = \int \cos(x) \sin(x) dx.$$

One possible choice is

$$H_1 = \frac{1}{2} (\cos(x) \sin(x) - x), \quad H_2 = -\frac{1}{2} \cos^2(x).$$

This gives the particular solution

$$u(x) = H_1(x) \cos(x) + H_2(x) \sin(x) = -\frac{1}{2} x \cos(x).$$

Hence the general solution is

$$u(x) = -\frac{x \cos(x)}{2} + A \cos(x) + B \sin(x).$$

Imposing $u(0) = 0$ and $u'(0) = 1$, we find

$$u(x) = -\frac{1}{2} x \cos(x) + \frac{3}{2} \sin(x).$$

35

REMARK 8.23. — In Example 8.22, notice the appearance of the factor x in front of $\cos(x)$ in the particular solution. This occurs because $\sin(x)$ and $\cos(x)$ already solve the homogeneous equation, so a particular solution cannot be obtained by taking a simple linear combination of them.

The situation from Example 8.22 illustrates a general principle behind the *method of undetermined coefficients*, which provides a systematic way to guess a particular solution of a linear ODE with constant coefficients. The procedure can be summarised as follows:

- Begin by making a guess that has the same functional form as the forcing term $g(x)$.
 - If $g(x)$ is a polynomial of degree n , guess a general polynomial of degree n .
 - If $g(x) = e^{\alpha x}$, guess $Ae^{\alpha x}$.
 - If $g(x) = \sin(\beta x)$ or $\cos(\beta x)$, guess $A \cos(\beta x) + B \sin(\beta x)$.
 - If $g(x)$ is a sum of such terms, combine the corresponding guesses.
- Check whether this first guess lies in the space of solutions of the homogeneous equation. If it does, then substituting it into the differential equation produces zero on the left-hand side, so it cannot match the non-zero forcing term $g(x)$.

- If the guess lies in the homogeneous solution space, multiply it by x . For a second-order equation, if this new guess still belongs to the homogeneous space, multiply by x^2 . At this point the guess will not solve the homogeneous equation, so no further powers are needed.

This explains why Example 8.22 requires a particular solution of the form $x \cos(x)$: the functions $\sin(x)$ and $\cos(x)$ solve the homogeneous equation, so we should look for a particular solution of the form $Ax \cos(x) + Bx \sin(x)$. A similar phenomenon occurs when $g(x) = e^{\alpha x}$ and α is a root of the characteristic equation. In that case one should try $Axe^{\alpha x}$. However, when $\Delta = 0$, then both $e^{\alpha x}$ and $xe^{\alpha x}$ solve the homogeneous equation, so the correct guess for a particular solution is $Ax^2e^{\alpha x}$.

35

EXERCISE 8.24. — Solve the following ODEs (the unknown function is always u):

1. $u''(x) + u'(x) + u(x) = \sin(2x), \quad u(0) = 0, \quad u'(0) = 1.$

Hint: Look for a particular solution of the form $a \sin(2x) + b \cos(2x)$.

2. $u''(x) + 4u(x) = \cos(2x), \quad u(0) = 1, \quad u'(0) = 0.$

Hint: Look for a particular solution of the form $ax \cos(2x) + bx \sin(2x)$.

3. $u''(x) + u'(x) - 2u(x) = x^2, \quad u(0) = 2, \quad u'(0) = 1.$

Hint: Look for a particular solution of the form $ax^2 + bx + c$.

4. $u''(x) + 2u'(x) - 3u(x) = \cos(x) + x, \quad u(0) = 1, \quad u'(0) = 1.$

Hint: Look for a particular solution of the form $a \sin(x) + b \cos(x) + cx + d$.

8.2 Existence and Uniqueness for ODEs

8.2.1 Existence and Uniqueness for First Order ODEs

Our goal now is to present the general theory of first-order ODEs for real-valued functions on the real line. In general, a first-order ODE is an equation relating x , $u(x)$, and $u'(x)$,

$$G(x, u(x), u'(x)) = 0.$$

In this section we restrict to equations for which one can “isolate” u' , so that one can write the ODE in *normal form*:

$$u'(x) = f(x, u(x)).$$

DEFINITION 8.25: FIRST-ORDER ODES IN NORMAL FORM

A first-order ODE in *normal form* is an equation of the type

$$u'(x) = f(x, u(x)),$$

where $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a given function and $u : \mathbb{R} \rightarrow \mathbb{R}$ is the unknown.

The Cauchy–Lipschitz Theorem, also known as the Picard–Lindelöf Theorem, is a fundamental result in the theory of ODEs. It ensures the existence and uniqueness of solutions under suitable conditions on f .

In the next theorem we need to assume that f is continuous as a function of the two variables x and y . This means that, for any point (x_0, y_0) in the domain and for any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|x - x_0| < \delta \quad \text{and} \quad |y - y_0| < \delta \quad \implies \quad |f(x, y) - f(x_0, y_0)| < \varepsilon.$$

THEOREM 8.26: CAUCHY-LIPSCHITZ: GLOBAL VERSION

Let $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ satisfy the following conditions:

1. f is continuous in $\mathbb{R} \times \mathbb{R}$;
2. f is Lipschitz with respect to the second variable; that is, there exists a constant $L > 0$ such that

$$|f(x, y_1) - f(x, y_2)| \leq L|y_1 - y_2| \quad \forall x \in \mathbb{R}, y_1, y_2 \in \mathbb{R}.$$

Then, for any point $(x_0, y_0) \in \mathbb{R} \times \mathbb{R}$ there exists a unique C^1 function $u : \mathbb{R} \rightarrow \mathbb{R}$ such that

$$\begin{cases} u'(x) = f(x, u(x)) & \text{for all } x \in \mathbb{R}, \\ u(x_0) = y_0. \end{cases} \quad (8.17)$$

As we shall see in Section 8.2.2 below, the proof is based on the method of successive approximations, also known as Picard iterations. It involves constructing a sequence of continuous functions that converge to the solution of the differential equation. Before diving into the proof of this important theorem, we first discuss some examples and generalizations.

Theorem 8.26 guarantees that solutions to the first order ODE (8.17) are unique when f is Lipschitz in the second variable. This assumption is crucial, as the next example shows.

EXAMPLE 8.27. — Consider the ODE

$$u'(x) = |u(x)|^\alpha, \quad u(0) = 0, \quad (8.18)$$

with $\alpha \in (0, 1]$.

- For $\alpha = 1$ the function $f(y) = |y|$ is Lipschitz, since

$$|f(y_1) - f(y_2)| = ||y_1| - |y_2|| \leq |y_1 - y_2|.$$

Hence, Theorem 8.26 guarantees that the solution is unique. Since the constant function $u = 0$ is a solution, this is the unique solution.

- For $\alpha < 1$ the function $f(y) = |y|^\alpha$ is not Lipschitz. Indeed, if this function were Lipschitz, then there would exist a constant $L > 0$ such that

$$|f(y) - f(0)| = |y|^\alpha \leq L|y| \quad \forall y \in \mathbb{R}.$$

This would imply

$$1 \leq L|y|^{1-\alpha} \quad \forall y \in \mathbb{R} \setminus \{0\},$$

but this is false since $\lim_{y \rightarrow 0} |y|^{1-\alpha} = 0$ (recall that $\alpha < 1$).

Note that, also in this case, the constant function $u = 0$ is a solution. We now try to use the method of separation of variables (recall Section 8.1.2) to find a second solution that is not zero, say with $u(x) > 0$ somewhere:

$$u'(x) = u(x)^\alpha \quad \implies \quad \frac{u'(x)}{u(x)^\alpha} = 1,$$

so, by integration, we obtain

$$\int \frac{du}{u^\alpha} = \int dx = x + A \quad \implies \quad \frac{u(x)^{1-\alpha}}{1-\alpha} = x + A.$$

Choosing $A = 0$ (which is compatible with $u(0) = 0$) we get

$$u(x) = ((1-\alpha)x)^{\frac{1}{1-\alpha}},$$

which is positive for $x > 0$. Hence, the function

$$u(x) = \begin{cases} 0 & \text{for } x \leq 0, \\ ((1-\alpha)x)^{\frac{1}{1-\alpha}} & \text{for } x > 0, \end{cases}$$

is a second solution of (8.18). Actually, given any value $x_0 \geq 0$, all the functions

$$u_{x_0}(x) = \begin{cases} 0 & \text{for } x \leq x_0, \\ ((1-\alpha)(x-x_0))^{\frac{1}{1-\alpha}} & \text{for } x > x_0, \end{cases}$$

solve (8.18), so there are infinitely many solutions.

Motivated by the previous example, one may wonder if the solution of (8.18) is unique for $\alpha > 1$. We begin with the following observation, stated as an exercise.

EXERCISE 8.28. — Let $\alpha > 1$. Prove that the function $f : \mathbb{R} \rightarrow \mathbb{R}$ given by $f(y) = |y|^\alpha$ is locally Lipschitz (i.e., it is Lipschitz in every compact interval $[a, b]$), but it is not Lipschitz on the whole \mathbb{R} .

Hint: Use Corollary 5.32.

By the previous exercise, we see that Theorem 8.26 does not apply to (8.18) when $\alpha > 1$. Still, since this function is locally Lipschitz, one may hope that some existence and uniqueness theorem still holds. This is indeed the case, as implied by the local version of the Cauchy–Lipschitz Theorem stated below. As we shall discuss later, since now the function f is only assumed to be locally Lipschitz, in general we cannot find a solution u defined on the whole \mathbb{R} .

THEOREM 8.29: CAUCHY-LIPSCHITZ: LOCAL VERSION

Let $I \subset \mathbb{R}$ be an interval, and let $f : I \times \mathbb{R} \rightarrow \mathbb{R}$ satisfy the following conditions:

1. f is continuous in $I \times \mathbb{R}$;
2. f is locally Lipschitz with respect to the second variable; that is, for every pair of compact intervals $[a, b] \subset I$ and $[c, d] \subset \mathbb{R}$ there exists a constant $L > 0$ such that

$$|f(x, y_1) - f(x, y_2)| \leq L|y_1 - y_2| \quad \forall x \in [a, b], y_1, y_2 \in [c, d].$$

Then, for any point $(x_0, y_0) \in I \times \mathbb{R}$ there exist an interval $I' \subset I$ containing x_0 and a unique C^1 function $u : I' \rightarrow \mathbb{R}$ such that

$$\begin{cases} u'(x) = f(x, u(x)) & \text{for all } x \in I', \\ u(x_0) = y_0. \end{cases} \quad (8.19)$$

In other words, under a local Lipschitz assumption, one can only guarantee the existence and uniqueness of a solution on some interval around x_0 . Moreover, as long as the solution $u(x)$ remains bounded within I' , one can continue applying Theorem 8.29 to extend the interval I' as much as possible.

36

To better understand why solutions are defined only on some interval $I' \subset I$, we consider the following example.

EXAMPLE 8.30. — Consider the ODE

$$u'(x) = u(x)^2, \quad u(0) = 1.$$

The function $f(y) = y^2$ is locally Lipschitz, so Theorem 8.29 applies.

To find the solution, we use separation of variables. More precisely, since $u(0) > 0$, by continuity u will be positive in a neighborhood of 0 and we get

$$u'(x) = u(x)^2 \quad \implies \quad \frac{u'(x)}{u(x)^2} = 1,$$

so, by integration, we obtain

$$\int \frac{du}{u^2} = \int dx = x + A \quad \implies \quad -\frac{1}{u(x)} = x + A.$$

Choosing $x = 0$ this implies $A = -1$ and we get

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

Note that this function solves the ODE on $(-\infty, 1)$, but $\lim_{x \rightarrow 1^-} u(x) = \infty$, so we cannot extend this solution beyond $x = 1$.

EXERCISE 8.31. — Consider the ODE

$$u'(x) = u(x)^\alpha, \quad u(0) = 1,$$

36

with $\alpha > 1$. Show that Theorem 8.29 applies and find the unique solution (which again blows up in finite time).

Although Theorem 8.29 guarantees that most nonlinear ODEs have a unique solution (at least locally in x), nonlinear ODEs are very difficult to solve and there are no general techniques to tackle such problems, neither in practice nor in theory. Therefore, in applications, one often resorts to numerical methods.

8.2.2 Extra material: Proof of Theorem 8.26

REMARK 8.32. — In the proof, we shall use the following fact: if $v : \mathbb{R} \rightarrow \mathbb{R}$ is a continuous function, then also the function $s \mapsto f(s, v(s))$ is continuous. This is a consequence of the continuity of f and the fact that the composition of continuous functions is continuous. Although we did not prove this fact in this specific setting where f depends on two variables, this can be proved in the same way as in Proposition 3.15.

Proof of Theorem 8.26. Let $L > 0$ be a Lipschitz constant for f with respect to the second variable, as in Theorem 8.26. We first prove local existence on a short interval around x_0 , then uniqueness on that interval, and finally extend the solution to the whole \mathbb{R} .

• **Step 1: An equivalent integral equation.** We first show that $u : \mathbb{R} \rightarrow \mathbb{R}$ is a C^1 solution to (8.17) if and only if u is a continuous function satisfying

$$u(x) = y_0 + \int_{x_0}^x f(s, u(s)) ds \quad \forall x \in \mathbb{R}. \quad (8.20)$$

Indeed, if u solves the ODE, then by integration (see Corollary 7.5) we deduce the validity of (8.20).

Conversely, if u is a continuous function satisfying (8.20), then Theorem 7.4 and Remark 8.32 imply that

$$x \mapsto \int_{x_0}^x f(s, u(s)) ds$$

is a primitive of the continuous function $f(x, u(x))$. Hence $u'(x) = f(x, u(x))$ for all x , so u is C^1 . Finally, choosing $x = x_0$ in (8.20) we deduce that $u(x_0) = y_0$.

Therefore, to prove existence, it suffices to construct a solution to (8.20). This will be accomplished by constructing what are known as Picard approximations, that is, a sequence of functions that converge to a solution of (8.20).

• **Step 2: Construction of Picard approximations.** First, we define the continuous function $u_0 : \mathbb{R} \rightarrow \mathbb{R}$ as

$$u_0(x) = y_0 \quad \forall x \in \mathbb{R}.$$

Then we define $u_1 : \mathbb{R} \rightarrow \mathbb{R}$ as

$$u_1(x) = y_0 + \int_{x_0}^x f(s, u_0(s)) ds.$$

The integral is well-defined since u_0 is continuous and therefore $s \mapsto f(s, u_0(s))$ is continuous (see Remark 8.32). We also observe that u_1 is the primitive of a continuous function, so it is C^1 (and, in particular, continuous).

More generally, given $n \in \mathbb{N}$, once the continuous function $u_n : \mathbb{R} \rightarrow \mathbb{R}$ is constructed, we define

$$u_{n+1}(x) = y_0 + \int_{x_0}^x f(s, u_n(s)) ds.$$

Again, since u_n is continuous, also $s \mapsto f(s, u_n(s))$ is continuous, and therefore u_{n+1} is C^1 (and in particular continuous).

• **Step 3: Convergence of Picard approximations on a short interval.** Set

$$\tau = \frac{1}{2L} \quad (\text{so that } L\tau = \tfrac{1}{2}).$$

We now prove the uniform convergence of the sequence of Picard approximations u_n on the interval $[x_0 - \tau, x_0 + \tau]$ by showing that this sequence corresponds to the partial sums of a uniformly convergent series of functions.

Define $v_k = u_k - u_{k-1}$ for $k \geq 1$, so that

$$u_n(x) = y_0 + \sum_{k=1}^n v_k(x).$$

We want to prove that the series

$$u_\infty(x) = y_0 + \sum_{k=1}^{\infty} v_k(x)$$

converges absolutely for every $x \in [x_0 - \tau, x_0 + \tau]$, so that the function $u_\infty : [x_0 - \tau, x_0 + \tau] \rightarrow \mathbb{R}$ is well-defined, and that the sequence of functions $\{u_n\}_{n=0}^{\infty}$ converges uniformly to u_∞ on $[x_0 - \tau, x_0 + \tau]$.

To this end, we observe that

$$v_{n+1}(x) = u_{n+1}(x) - u_n(x) = \int_{x_0}^x (f(s, u_n(s)) - f(s, u_{n-1}(s))) ds \quad \forall x \in \mathbb{R},$$

therefore, by the Lipschitz regularity of f in the second variable,

$$\begin{aligned} |v_{n+1}(x)| &\leq \int_{x_0}^x |f(s, u_n(s)) - f(s, u_{n-1}(s))| ds \\ &\leq L \int_{x_0}^x |u_n(s) - u_{n-1}(s)| ds = L \int_{x_0}^x |v_n(s)| ds \end{aligned} \quad (8.21)$$

for every $x \in \mathbb{R}$ and $n \geq 1$.

For $n \geq 1$, define

$$a_n = \max_{x \in [x_0 - \tau, x_0 + \tau]} |v_n(x)|.$$

Given $x \in [x_0 - \tau, x_0 + \tau]$, we have $[x_0, x] \subset [x_0 - \tau, x_0 + \tau]$, hence

$$\int_{x_0}^x |v_n(s)| ds \leq a_n |x - x_0| \leq a_n \tau,$$

and combining this with (8.21) yields

$$|v_{n+1}(x)| \leq L a_n \tau = \frac{a_n}{2}.$$

Since x is arbitrary in $[x_0 - \tau, x_0 + \tau]$, this proves that

$$a_{n+1} \leq L a_n \tau = \frac{a_n}{2} \quad \forall n \geq 1.$$

By induction we deduce

$$a_{n+1} \leq 2^{-n} a_1 \quad \forall n \geq 1.$$

Thus, for every $x \in [x_0 - \tau, x_0 + \tau]$ and every $k \geq 1$,

$$|v_k(x)| \leq 2^{-(k-1)} a_1.$$

By the majorant criterion, the series $\sum_{k=1}^{\infty} v_k(x)$ converges absolutely for every $x \in [x_0 - \tau, x_0 + \tau]$.

Moreover, for $x \in [x_0 - \tau, x_0 + \tau]$ and $n \geq 0$,

$$\begin{aligned} |u_{\infty}(x) - u_n(x)| &= \left| \sum_{k=n+1}^{\infty} v_k(x) \right| \leq \sum_{k=n+1}^{\infty} |v_k(x)| \leq a_1 \sum_{k=n+1}^{\infty} 2^{-(k-1)} \\ &= a_1 2^{-n} \xrightarrow{n \rightarrow \infty} 0, \end{aligned} \quad (8.22)$$

so the sequence of functions $\{u_n\}_{n=0}^{\infty}$ converges uniformly to u_{∞} on $[x_0 - \tau, x_0 + \tau]$.

• **Step 4: The limit function u_{∞} solves (8.20).** We now take the limit as $n \rightarrow \infty$ in

$$u_{n+1}(x) = y_0 + \int_{x_0}^x f(s, u_n(s)) ds, \quad \forall x \in [x_0 - \tau, x_0 + \tau].$$

The left-hand side converges to $u_{\infty}(x)$ uniformly on $[x_0 - \tau, x_0 + \tau]$ by Step 3.

For the right-hand side, recalling (8.22) and that $L\tau = 1/2$, we estimate

$$\left| \int_{x_0}^x (f(s, u_\infty(s)) - f(s, u_n(s))) ds \right| \leq L \int_{x_0}^x |u_\infty(s) - u_n(s)| ds \leq L|x - x_0|a_1 2^{-n} \leq a_1 2^{-(n+1)},$$

which proves that

$$\int_{x_0}^x f(s, u_n(s)) ds \rightarrow \int_{x_0}^x f(s, u_\infty(s)) ds \quad \text{as } n \rightarrow \infty.$$

So we conclude that u_∞ solves (8.20) on $[x_0 - \tau, x_0 + \tau]$, which shows the existence of a solution u_∞ on $[x_0 - \tau, x_0 + \tau]$.

• **Step 5: Local uniqueness on $[x_0 - \tau, x_0 + \tau]$.** Let $u_1, u_2 : [x_0 - \tau, x_0 + \tau] \rightarrow \mathbb{R}$ be two solutions of (8.17), and therefore of (8.20). Then, for each $x \in [x_0 - \tau, x_0 + \tau]$,

$$u_1(x) - u_2(x) = \int_{x_0}^x (f(s, u_1(s)) - f(s, u_2(s))) ds,$$

so

$$|u_1(x) - u_2(x)| \leq \int_{x_0}^x |f(s, u_1(s)) - f(s, u_2(s))| ds \leq L \int_{x_0}^x |u_1(s) - u_2(s)| ds.$$

Define

$$a = \max_{x \in [x_0 - \tau, x_0 + \tau]} |u_1(x) - u_2(x)|.$$

For $x \in [x_0 - \tau, x_0 + \tau]$ we then have

$$|u_1(x) - u_2(x)| \leq La|x - x_0| \leq L\tau a = \frac{a}{2},$$

so taking the maximum over $x \in [x_0 - \tau, x_0 + \tau]$ yields

$$a \leq \frac{a}{2}.$$

This implies that $a = 0$, that is, $u_1 = u_2$ on $[x_0 - \tau, x_0 + \tau]$.

• **Step 6: Global existence and uniqueness.** We now extend the solution uniquely to the whole \mathbb{R} by iterating the local existence and uniqueness argument.

Set $x_1 = x_0 + \tau$ and $y_1 = u_\infty(x_1)$. Applying the same Picard construction with initial data (x_1, y_1) gives a unique solution $u^{(1)}$ on $[x_1 - \tau, x_1 + \tau] = [x_0, x_0 + 2\tau]$. On the overlap $[x_0, x_1]$ both u_∞ and $u^{(1)}$ solve (8.17) with the same value at x_1 , so by Step 5 they coincide there. Hence we can glue them and obtain a single solution on $[x_0 - \tau, x_1 + \tau] = [x_0 - \tau, x_0 + 2\tau]$.

Iterating this construction to the right, we obtain a unique solution on $[x_0 - \tau, \infty)$. A completely analogous construction to the left (starting from $x_0 - \tau$ and moving to the left by steps of length τ) yields a unique solution on $(-\infty, x_0 + \tau]$. Gluing the left and right pieces together (again using local uniqueness on overlaps) gives a unique solution $u : \mathbb{R} \rightarrow \mathbb{R}$ satisfying (8.17). This completes the proof of Theorem 8.26. \square

8.2.3 Higher Order ODEs

Before considering general higher-order ODEs, we begin with a familiar second-order linear equation and show how it can be rewritten as a system of first-order ODEs.

EXAMPLE 8.33 (Linear second-order equation as a first-order system). — Let $a_0, a_1 : I \rightarrow \mathbb{R}$ be continuous functions and consider the linear homogeneous second-order ODE

$$u''(x) + a_1(x)u'(x) + a_0(x)u(x) = 0 \quad \forall x \in I. \quad (8.23)$$

Define

$$U_1(x) = u(x), \quad U_2(x) = u'(x).$$

Then

$$U_1'(x) = U_2(x), \quad U_2'(x) = u''(x) = -a_1(x)U_2(x) - a_0(x)U_1(x),$$

so (8.23) is equivalent to the first-order system

$$\begin{cases} U_1'(x) &= U_2(x), \\ U_2'(x) &= -a_1(x)U_2(x) - a_0(x)U_1(x). \end{cases}$$

36

Thus, prescribing the values of $u(x_0)$ and $u'(x_0)$ at some $x_0 \in I$ is equivalent to prescribing the initial condition $(U_1(x_0), U_2(x_0))$ for the associated first-order system.

The same idea extends to arbitrary higher-order ODEs. More precisely, suppose we are given an n -th order ODE of the form

$$G(x, u(x), u'(x), u''(x), \dots, u^{(n)}(x)) = 0,$$

and assume that the highest derivative can be isolated and written as

$$u^{(n)}(x) = f(x, u(x), u'(x), \dots, u^{(n-1)}(x)), \quad (8.24)$$

where $f : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}$ is a given function.

We introduce the variables

$$U_1 = u, \quad U_2 = u', \quad U_3 = u'', \quad \dots, \quad U_n = u^{(n-1)}.$$

By definition, we have

$$U_1' = U_2, \quad U_2' = U_3, \quad \dots, \quad U_{n-1}' = U_n,$$

and equation (8.24) becomes

$$U'_n = u^{(n)} = f(x, U_1, U_2, \dots, U_n).$$

Therefore, the n -th order equation (8.24) is equivalent to the first-order system

$$\begin{cases} U'_1 = U_2, \\ U'_2 = U_3, \\ \vdots \\ U'_{n-1} = U_n, \\ U'_n = f(x, U_1, U_2, \dots, U_n). \end{cases}$$

The Cauchy–Lipschitz Theorem (both its global and local versions) extends to systems of first-order ODEs and ensures existence and uniqueness of solutions whenever the right-hand side is continuous and (locally) Lipschitz with respect to the variables (U_1, \dots, U_n) . In particular, once the initial conditions

$$U_1(x_0) = u(x_0), \quad U_2(x_0) = u'(x_0), \quad \dots, \quad U_n(x_0) = u^{(n-1)}(x_0)$$

are prescribed at some $x_0 \in I$, there exists a unique (local) solution to the system, and hence a unique solution to the original n -th order ODE.

As an application, we obtain a classical structure result for linear second-order equations.

THEOREM 8.34: EXISTENCE AND UNIQUENESS FOR LINEAR SECOND-ORDER ODES

Let $a_0, a_1 : I \rightarrow \mathbb{R}$ be continuous and bounded functions. Then the set of solutions of the linear homogeneous equation

$$u''(x) + a_1(x)u'(x) + a_0(x)u(x) = 0 \quad \forall x \in I$$

is a two-dimensional linear subspace of $C^2(I)$.

Proof. Fix $x_0 \in I$ and define $U_1 = u$, $U_2 = u'$. As shown in Example 8.33, the equation can be rewritten as the first-order system

$$\begin{cases} U'_1 = U_2, \\ U'_2 = -a_1(x)U_2 - a_0(x)U_1. \end{cases}$$

The right-hand side is linear in (U_1, U_2) and continuous and bounded in x , so it is Lipschitz in both U_1 and U_2 . Thus, the Cauchy–Lipschitz Theorem for systems ensures the existence and uniqueness of solutions once $(U_1(x_0), U_2(x_0))$ is prescribed. Equivalently, the original second-order equation has a unique solution once $u(x_0)$ and $u'(x_0)$ are given.

Let u_1 be the unique solution satisfying $u_1(x_0) = 1$ and $u_1'(x_0) = 0$, and let u_2 be the unique solution satisfying $u_2(x_0) = 0$ and $u_2'(x_0) = 1$. By linearity, $Au_1 + Bu_2$ is a solution for every $A, B \in \mathbb{R}$.

Conversely, if u is any solution and we set $A = u(x_0)$ and $B = u'(x_0)$, then the function $v = u - Au_1 - Bu_2$ solves the equation and satisfies $v(x_0) = v'(x_0) = 0$. By uniqueness, v must be identically zero, and therefore $u = Au_1 + Bu_2$. \square

The higher-dimensional version of the Cauchy–Lipschitz Theorem for systems will be studied in Analysis 2. Understanding the proof of the one-dimensional case presented in Section 8.2.2 provides a solid foundation for that more general theory.

Index

- absolute value, 38
- accumulation point, 60
- affine
 - function, 154, 224
- Airy equation, 237
- arc tangent, 179
- arccotangent, 179
- area, 189
- argument, 148
- Arzelà–Ascoli theorem, 116
- associativity, 82
- base
 - logarithm, 100
- Bessel equation, 237
- bijection, 27
- bijective, 27
- boundary conditions, 237
- bounded, 42, 64
- bounded function, 77
- Cauchy sequence, 71
- Cauchy–Schwarz inequality, 39
- change of variable, 215
- closed, 34
- closed disk, 40
- coefficient, 134
- completeness axiom, 26
- complex conjugation, 37
- complex numbers, 35
- continuous
 - piecewise, 197
- continuous extension, 104
- continuum, 57
- decimal fraction, 50
- decomposition, 183
- decreasing, 77
- dense, 49
- distance, 34
- equibounded, 116
- equicontinuity, 116
- Euler’s number, 96
- eventually constant, 58
- exponential function, 96
- exponential series, 140
- factor, 81
- fractional part, 48
- geometric sequence, 74
- higher derivative, 156
- Hölder continuous, 102
- identity function, 87
- image, 27
- imaginary part, 35
- imaginary unit, 35
- improper limit, 73
- increasing, 77
- indefinite values, 46
- index, 81
- inferior limit, 68
- Infimum, 45
- injection, 27
- injective, 27
- integrand, 189
- integration constant, 214
- integration limit, 189
- inverse function, 87
- Inverse Hyperbolic Cosine, 181
- Inverse Hyperbolic Sine, 180

- inverse mapping, 87
- Lebesgue integral, 189
- limit, 58
- line, 154
- Lipschitz continuous, 93
- lower bound, 42
- lower sums, 188
- maximum, 42
- minimum, 42
- monotone, 77
 - piecewise, 196
- natural logarithm, 100
- neighbourhood, 33
- norm, 38
- number line, 26
- open, 34
- open disk, 40
- partial fraction decomposition, 219
- partition, 182, 183
- piecewise continuous, 197
- piecewise monotone, 196
- pointwise convergence, 112
- polar coordinates, 148
- power series, 134, 137
- preimage, 27
- radius of convergence, 135
- rational function, 219
- real decimal fraction, 51
- real part, 35
- real-valued, 76
- removable discontinuity, 104
- restriction, 80
- root, 89
- rounding function, 48
- running variable, 81
- sequence, 58, 112
- sequential continuity, 83
- smooth function, 157
- square root function, 28
- strictly monotone, 77
- subsequence, 59
- summand, 81
- superior limit, 68
- supremum, 42
- surjection, 27
- surjective, 27
- two-point compactification, 45
- uniform convergence, 113, 138
- uniformly continuous, 91
- unit circle, 148
- upper bound, 42
- upper sums, 188
- variable, 76, 134
- zero, 76

Bibliography

- [ACa2003] N. A'Campo, *A natural construction for the real numbers* arXiv preprint 0301015, (2003)
- [Apo1983] T. Apostol, *A proof that Euler missed: Evaluating $\zeta(2)$ the easy way* The Mathematical Intelligencer **5** no.3, p. 59–60 (1983)
- [Aig2014] M. Aigner and G. M. Ziegler, *Das BUCH der Beweise* Springer, (2014)
- [Amm2006] H. Amann und J. Escher, *Analysis I*, 3. Auflage, Grundstudium Mathematik, Birkhäuser Basel, (2006)
- [Bla2003] C. Blatter, *Analysis I* ETH Skript, <https://people.math.ethz.ch/~blatter/dlp.html> (2003)
- [Bol1817] B. Bolzano, *Rein analytischer Beweis des Lehrsatzes, daß zwischen je zwei Werthen, die ein entgegengesetztes Resultat gewähren, wenigstens eine reelle Wurzel der Gleichung liege*, Haase Verl. Prag (1817)
- [Boo1847] G. Boole, *The mathematical analysis of logic* Philosophical library, (1847)
- [Can1895] G. Cantor, *Beiträge zur Begründung der transfiniten Mengenlehre* Mathematische Annalen **46** no.4, 481–512 (1895)
- [Cau1821] A.L. Cauchy, *Cours d'analyse de l'école royale polytechnique* L'Imprimerie Royale, Debure frères, Libraires du Roi et de la Bibliothèque du Roi. Paris, (1821)
- [Ded1872] R. Dedekind, *Stetigkeit und irrationale Zahlen* Friedrich Vieweg und Sohn, Braunschweig (1872)
- [Die1990] J. Dieudonné, *Elements d'analyse* Editions Jacques Gabay (1990)
- [Hat02] A. Hatcher, *Algebraic Topology* Cambridge University Press (2002)
- [Hil1893] D. Hilbert, *Über die Transzendenz der Zahlen e und π* Mathematische Annalen **43**, 216–219 (1893)
- [Hos1715] G.F.A. Marquis de l'Hôpital, *Analyse des Infiniment Petits pour l'Intelligence des Lignes Courbes* 2nde Edition, F. Montalant, Paris (1715)

- [Lin1894] E. Lindelöf, *Sur l'application des méthodes d'approximations successives à l'étude des intégrales réelles des équations différentielles ordinaires* Journal de mathématiques pures et appliquées **10** no.4, 117–128 (1894)
- [Rus1903] B. Russell, *The principles of mathematics* WW Norton & Company, (1903)
- [Rot88] J. J. Rotman, *An introduction to Algebraic Topology* Graduate Texts in Mathematics **119** Springer 1988
- [Smu1978] R. Smullyan, *What is the name of this book?* Prentice-Hall, (1978)
- [Zag1990] D. Zagier, *A one-sentence proof that every prime $p \equiv 1 \pmod{4}$ is a sum of two squares.* Amer. Math. Monthly **97**, no.2, p. 144 (1990)