Department of Mathematics

Indian Institute of Technology Guwahati

# Real Analysis

Lecture Notes

MA642

Foundations, function spaces, and multivariable analysis

Rajesh K. Srivastava

Department of Mathematics, IIT Guwahati

Graduate Course

Updated: July 7, 2026

# Preface

These lecture notes were prepared for the course *MA642: Real Analysis*. They have been written as a systematic and self-contained text for classroom use, guided reading, study, and examination preparation. Their principal aim is to present the standard syllabus of real analysis in a form that is logically coherent, pedagogically structured, and mathematically precise. In preparing and expanding the material, particular care has been taken to improve the order of exposition, the clarity of definitions, the architecture of proofs, the consistency of notation, and the overall readability of the text. Wherever useful, theorem-like statements have also been supplied with descriptive headings, so that references to theorems, lemmas, propositions, and corollaries indicate not only their number but also their mathematical role in the surrounding discussion.

The notes are intended for a rigorous course in analysis in which the familiar themes of calculus are re-examined from a proof-oriented point of view. They begin with the structure of the real line, including order, completeness, countability, cardinality, and the Cantor set, and then move to the abstract setting of metric and normed linear spaces, where the language of convergence, continuity, compactness, completeness, and connectedness assumes its natural general form. From this foundation, the text returns to classical analysis through continuity and monotone functions on the real line, compactness phenomena in spaces of functions, semicontinuity, approximation, and power series. The later chapters develop differential calculus in one and several variables and culminate in multiple integration, line and surface integrals, and the main integral theorems of vector calculus.

The notes have also been aligned explicitly with the standard syllabus for MA642. The central themes developed here include the least upper bound property of $\mathbb{R}$; countable and uncountable sets; basic topology of metric spaces; normed linear spaces and standard inequalities; Cauchy sequences and completeness; compactness and total boundedness; connectedness and path connectedness; limits and continuity; monotone functions and inverse continuity phenomena; compactness in function spaces; approximation and semicontinuity; power series; differentiation and the mean value theorems; Taylor's theorem; differentiability in $\mathbb{R}^n$; inverse and implicit function theorems; extrema and Lagrange multipliers; and finally multiple integration, change of variables, line integrals, surface integrals, divergence, curl, and the classical theorems of Green, Gauss, and Stokes.

The present notes pursue three complementary goals. First, they aim to make the internal logic of the subject visible: one sees how the completeness of the real numbers leads naturally to compactness arguments, how compactness interacts with continuity, and how differentiability and integration are best understood only after the topological and metric framework has been established. Second, the exposition is designed to be proof-conscious: definitions are introduced in forms suitable for later use, and proofs are written so that both the main idea and the technical mechanism of the argument remain transparent. Third, examples and counterexamples are used throughout to show not only why a theorem is true, but also why each hypothesis is necessary.

A further objective of these notes is to bridge the gap between computational familiarity and theoretical understanding. Many students encounter the basic operations of calculus long before they see the reasons these procedures are valid. Real analysis supplies that missing foundation. It explains why limits behave well under appropriate hypotheses, why compactness produces uniform control from pointwise information, why connectedness restricts the possible shape of image sets, and why differentiability is fundamentally a linear approximation principle. In several variables, these ideas become richer and more delicate, and the inverse and implicit function theorems reveal the local geometric structure hidden behind nonlinear equations. The integral theorems at the end of the course then show how local differential information governs global quantities.

The manuscript has also been organized to support different modes of study. A reader following the course sequentially will find that the later chapters depend naturally on the earlier conceptual framework, while a student preparing for examinations may use the chapter introductions, theorem statements, and exercises as a structured guide to the major ideas. The problem sets are intended not merely as routine drills, but as an essential part of the course: they reinforce definitions, develop technique, test conceptual understanding, and encourage the reader to engage actively with the theory.

**Learning objectives.**

- Develop a rigorous understanding of the structure of the real line and the foundational language of metric and normed spaces.
- Master the basic compactness, completeness, and connectedness principles that govern modern analysis.
- Understand continuity, approximation, and convergence phenomena in both concrete and abstract settings.
- Study differential calculus in one and several variables from a theorem-based point of view.
- Learn the analytic and geometric content of multiple integration and the principal integral theorems of vector calculus.
- Build the proof-writing habits and conceptual discipline necessary for further study in analysis and related areas.

**Key ideas.**

- The exposition has been arranged so that conceptual dependencies appear in a natural and mathematically motivated order.
- Definitions, examples, and proofs are presented with emphasis on clarity, precision, and pedagogical usefulness.
- The notes are designed to function simultaneously as lecture text, study guide, and reference for independent study.

**Prerequisites.** A solid background in single-variable calculus and elementary linear algebra is assumed throughout.

**Notation.** We write $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ for the standard number systems. We use $\subseteq$ for inclusion and $\subset$ for proper inclusion. If $(X, d)$ is a metric space, $x \in X$, and $r > 0$, then

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

denotes the open ball of radius $r$ centred at $x$. Unless explicitly stated otherwise, all vector spaces are over $\mathbb{R}$ or $\mathbb{C}$, according to context.

# Outline of Topics

- **Foundations of analysis:** order and completeness on $\mathbb{R}$, countability, the Cantor set, metric spaces, normed spaces, completeness, compactness, and connectedness.
- **Continuity and function spaces:** limits and continuity on the real line, monotone functions, compactness in spaces of functions, semicontinuity, approximation, and power series.
- **Differential calculus:** one-variable differentiation, mean value principles, Taylor's theorem, and the several-variable theory of differentiability, local extrema, and inverse/implicit mappings.
- **Integration and vector analysis:** multiple Riemann integration, repeated integration, change of variables, line and surface integrals, and the integral theorems of vector calculus.

# Contents

**Preface** i

**Outline of Topics** iv

**Introduction** 1

**1 Foundations on the Real Line** 3
- 1.1 The Real Numbers . . . 3
  - 1.1.1 Preliminaries . . . 3
  - 1.1.2 Completeness Property of $\mathbb{R}$ . . . 4
  - 1.1.3 Archimedean property . . . 5
  - 1.1.4 Nested Interval Theorem . . . 7
- 1.2 Countability and Cardinality . . . 9
  - 1.2.1 Countable and uncountable sets . . . 9
  - 1.2.2 Cantor's diagonal argument . . . 10
  - 1.2.3 Cardinality and comparison of sets . . . 10
- 1.3 A warm-up exercise . . . 11
- 1.4 The Cantor Set . . . 11
  - 1.4.1 Cantor Set . . . 11
  - 1.4.2 Properties of the Cantor Set . . . 12
  - 1.4.3 Ternary representation of the Cantor set . . . 13
- 1.5 Basic Topology on $\mathbb{R}$ . . . 15
  - 1.5.1 Open Sets and Closed Sets . . . 15
  - 1.5.2 Interior of a set . . . 17
  - 1.5.3 Closure of a set . . . 17
  - 1.5.4 Compactness . . . 18
  - 1.5.5 Dense Set . . . 19

**2 Metric and Normed Linear Spaces** 22
- 2.1 Metric Spaces and Topology . . . 22
  - 2.1.1 Metric spaces . . . 22
  - 2.1.2 Continuous maps between metric spaces . . . 31
- 2.2 Normed Linear Spaces and Classical Inequalities . . . 39
  - 2.2.1 Normed Linear Space . . . 40
  - 2.2.2 Convergence of Sequence in Metric Space . . . 42
  - 2.2.3 Cauchy Sequences . . . 42
  - 2.2.4 Young's Inequality . . . 42

2.2.5 Space of Sequences . . . 43
2.2.6 Hölder's Inequality . . . 44
2.2.7 Minkowski's Inequality . . . 44
2.2.8 Geometry of spheres in $(\mathbb{R}^n, \|\cdot\|_p)$ . . . 46
2.2.9 Closed sets in $(X, d)$ . . . 46
2.2.10 Interior in $(X, d)$ . . . 48
2.2.11 Closure in $(X, d)$ . . . 48
2.2.12 Space of Finite Sequences . . . 49
2.2.13 Complete Metric Spaces . . . 51
2.2.14 Sequences of Functions . . . 55
2.2.15 Term-by-term differentiation . . . 61
2.2.16 Uniform continuity . . . 61
2.2.17 Pointwise continuity of $f$ . . . 62

**3 Completeness and Compactness in Metric and Normed Spaces** **68**
3.1 Completeness and Fixed Points . . . 68
3.1.1 Fixed Points . . . 68
3.2 Total Boundedness and Compactness . . . 71
3.2.1 Totally Bounded Set . . . 71
3.2.2 Compact Metric Spaces . . . 76
3.2.3 Separable Metric Spaces . . . 80

**4 Connectedness in Metric and Normed Linear Spaces** **84**
4.1 Connected and Disconnected Sets . . . 84
4.2 Path Connectedness . . . 89
4.2.1 Path Connectedness . . . 89

**5 Continuity and Monotone Functions on the Real Line** **94**
5.1 Limits and Continuity . . . 94
5.1.1 Limits and one-sided limits . . . 94
5.1.2 Continuity . . . 95
5.2 Monotone Functions . . . 95
5.2.1 One-sided limits and discontinuities . . . 95
5.3 Surjections, Inverses, and Continuity . . . 96
5.3.1 Preimages of intervals under monotone maps . . . 97
5.4 Constructing Monotone Functions with Prescribed Jumps . . . 98

**6 Function Spaces, Approximation, and Power Series** **100**
6.1 Compactness in Spaces of Functions . . . 100
6.1.1 Equicontinuity . . . 100
6.1.2 Arzelà-Ascoli Theorem . . . 101

6.1.3 Dini's Theorem . . . 103
6.2 Semicontinuity and Approximation . . . 104
6.2.1 Upper Semicontinuity . . . 104
6.2.2 Weierstrass Approximation Theorem . . . 106
6.3 Power Series . . . 109

**7 Differential Calculus in One Variable 112**
7.1 Differentiability . . . 112
7.2 Mean Value Theorems . . . 113
7.3 Consequences of the MVT . . . 114
7.4 Taylor's theorem . . . 115

**8 Differential Calculus in Several Variables 117**
8.1 Orientation and Chapter Roadmap . . . 117
8.2 Limits and Continuity . . . 117
8.2.1 Notation and basic definitions in Euclidean space . . . 118
8.2.2 Limits in Euclidean space . . . 119
8.2.3 Continuity in Euclidean space . . . 120
8.3 Differentiation in $\mathbb{R}^n$ . . . 122
8.3.1 Partial derivatives . . . 123
8.3.2 Directional derivatives . . . 123
8.3.3 Differentiability . . . 125
8.3.4 Chain rule . . . 128
8.3.5 Taylor's theorem . . . 135
8.3.6 Multi-index notation and the general Taylor formula . . . 136
8.4 Inverse and implicit function theorems . . . 139
8.4.1 Inverse function theorem . . . 140
8.4.2 Implicit function theorem . . . 143
8.5 Extrema, the Hessian test, and Lagrange multipliers . . . 147
8.5.1 Unconstrained extrema and critical points . . . 147
8.5.2 The Hessian and the second derivative test . . . 148
8.5.3 Constrained extrema and the Lagrange multiplier condition . . . 148

**9 Multiple Integration and Vector Calculus 152**
9.1 A brief review of the Riemann integral in one variable . . . 152
9.2 Double integrals . . . 156
9.2.1 Repeated Integrals . . . 161
9.2.2 Bounded functions with discontinuities . . . 163
9.2.3 Double integral over general bounded regions . . . 164
9.2.4 Change of variable . . . 165
9.2.5 Change of variable in single integral . . . 165

9.2.6 Change of variable in double integral . . . 165
9.3 Triple integrals . . . 167
9.3.1 Change of variable in a triple integral . . . 168
9.4 Surface area and surface integrals . . . 169
9.4.1 Area of a parametric surface . . . 170
9.4.2 Surface integrals . . . 172
9.5 Line Integrals . . . 173
9.5.1 Exactness of the line integral . . . 177
9.6 Curl and Divergence . . . 178

**Problem Sets** **184**
Problem Set 1: Metric and Normed Linear Spaces . . . 184
Problem Set 2: Completeness, Compactness, and Function Spaces . . . 189
Problem Set 3: Connectedness and Continuity . . . 192
Problem Set 4: Several-Variable Calculus . . . 193

**References** **197**

# Introduction

Real analysis is the mathematical discipline in which the informal procedures of calculus are recast as rigorous arguments. Its basic notions—limit, continuity, compactness, differentiability, and integration—are familiar from earlier study, but in analysis they are reorganised into a precise theory whose strength lies in the interaction of local and global principles. The least-upper-bound property of $\mathbb{R}$ governs approximation on the real line; compactness converts pointwise or local information into uniform control; connectedness constrains the shape of image sets; and differentiability identifies the correct linear model for nonlinear behaviour. The subject acquires its full depth when these principles are seen not as isolated facts, but as a tightly connected logical system.

The present notes are organised to reflect that internal structure. The opening chapters establish the foundational language of order, completeness, topology, metric geometry, compactness, and connectedness. The next stage develops continuity on the real line and then passes to function spaces, where compactness, approximation, and convergence phenomena interact in a subtler way. The final portions of the text develop differential calculus in one and several variables and culminate in multiple integration and the classical integral theorems of vector analysis. This progression follows the natural movement of the subject from the real line to abstract spaces and then back to concrete analytic applications.

The exposition is intentionally theorem-driven and proof-oriented. Statements are formulated so that hypotheses can be tracked precisely; proofs are written to reveal both the key idea and the technical reduction; examples and counterexamples are used to indicate the limits of the theory. The objective is not merely to record a syllabus, but to present it as a coherent body of mathematics whose methods can be reused beyond the confines of a single course. For that reason the notes aim to function simultaneously as lecture text, study guide, and reference document.

## Chapter-wise organisation

**Chapter 1.** Foundations on the real line: order, completeness, countability, cardinality, and the Cantor set.

**Chapter 2.** Metric and normed linear spaces: metrics, norms, convergence, open and closed sets, and standard models.

**Chapter 3.** Completeness and compactness: Cauchy sequences, fixed points, total boundedness, compactness criteria, and their consequences.

**Chapter 4.** Connectedness: connected and path-connected sets, interval structure in $\mathbb{R}$, and analytic applications.

**Chapter 5.** Continuity and monotone function theory on the real line.

**Chapter 6.** Function spaces, approximation, semicontinuity, and power series.

**Chapter 7.** One-variable differential calculus: differentiability, mean value principles, and Taylor expansion.

**Chapter 8.** Several-variable differential calculus: differentiability, Jacobians, the chain rule, extrema, and inverse/implicit function theorems.

**Chapter 9.** Multiple integration and vector calculus, culminating in repeated integration, change of variables, and the theorems of Green, Gauss, and Stokes.

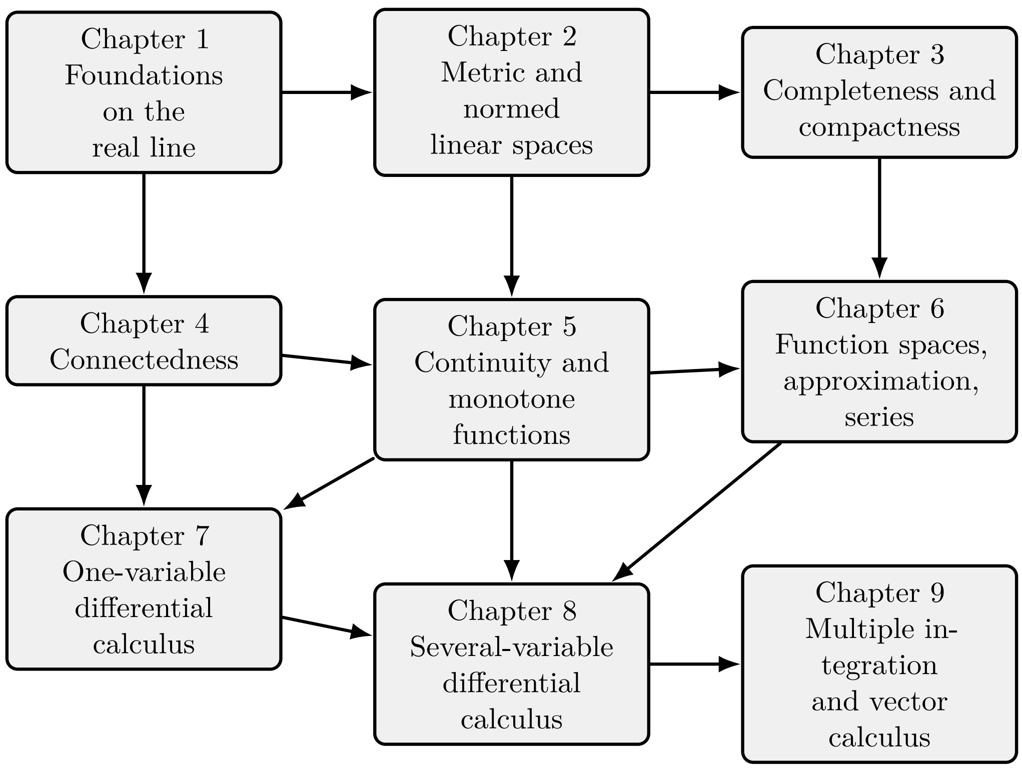



**Figure 1.** Logical progression of the principal themes and dependencies in the course.

Chapter 1

# Foundations on the Real Line

*This chapter develops the foundational structure of the real line from three complementary perspectives: order, completeness, and topology. We establish the least-upper-bound principle and its consequences, including approximation properties, the Archimedean principle, and the nested interval method. We then introduce the elementary topological language of open and closed sets, interior, closure, and density in $\mathbb{R}$, together with basic countability phenomena. The chapter concludes with the Cantor set, whose construction shows in a particularly vivid way that cardinality, topology, and metric size need not coincide.*

**Learning objectives.**

- Understand how order, completeness, and approximation interact on the real line.
- Use the least-upper-bound principle, nested intervals, and Bolzano–Weierstrass as mutually reinforcing forms of completeness.
- Distinguish countable from uncountable phenomena, and use the Cantor set as a guiding example showing that topological size and metric size need not agree.

## 1.1 The Real Numbers

**Section overview.**

- We begin with the order structure of $\mathbb{R}$ and the completeness axiom encoded by sup and inf.
- The main goal is to see how approximation statements, monotone convergence, and nested interval arguments all emerge from this single structural principle.
- Keep track of the difference between an upper bound, the least upper bound, and approximating the supremum from within the set.

### 1.1.1 Preliminaries

Recall that the set of rational numbers is

$$\mathbb{Q} := \left\{ \frac{p}{q} : p, q \in \mathbb{Z},\ q \neq 0 \right\}. \tag{1.1}$$

Not every real number is rational. A classical example is $\sqrt{2}$. Indeed, if $\sqrt{2} = p/q$ with $\gcd(p, q) = 1$, then

$$p^2 = 2q^2.$$

In particular, $p$ is even, so $p = 2m$ for some $m \in \mathbb{Z}$. Substituting into the previous identity gives $q^2 = 2m^2$, so $q$ is also even. This contradicts $\gcd(p, q) = 1$. Therefore $\sqrt{2} \notin \mathbb{Q}$. The elements of $\mathbb{R} \setminus \mathbb{Q}$ are called *irrational numbers*.

*Remark* 1.1. The rational numbers do not satisfy the least upper bound property, and hence they are not complete.

**Definition 1.2** (Upper and lower bounds)**.** Let $A \subseteq \mathbb{R}$. A number $x_0 \in \mathbb{R}$ is an *upper bound* for $A$ if $a \leq x_0$ for every $a \in A$. A number $y_0 \in \mathbb{R}$ is a *lower bound* for $A$ if $y_0 \leq a$ for every $a \in A$.

**Definition 1.3** (Supremum and infimum)**.** Assume that $A \subseteq \mathbb{R}$ is nonempty. An upper bound $x_0$ of $A$ is called the *least upper bound* of $A$ (or the *supremum* of $A$) if $x_0 \leq x$ for every upper bound $x$ of $A$. In this case we write $x_0 = \sup A$. Similarly, a lower bound $y_0$ of $A$ is called the *greatest lower bound* of $A$ (or the *infimum* of $A$) if $y \leq y_0$ for every lower bound $y$ of $A$. In this case we write $y_0 = \inf A$.

**Example 1.4.** $A = \left\{1 - \frac{1}{n} : n \in \mathbb{N}\right\}$. Show that $\inf A = 0$ and $\sup A = 1$.

*Remark* 1.5. The least upper bound principle recorded in the next subsection is the defining completeness axiom for $\mathbb{R}$.

### 1.1.2 Completeness Property of $\mathbb{R}$

**Theorem 1.6** (Least upper bound property)**.** *Every nonempty subset of* $\mathbb{R}$ *that is bounded above has a least upper bound in* $\mathbb{R}$*. Equivalently, every nonempty subset of* $\mathbb{R}$ *that is bounded below has a greatest lower bound in* $\mathbb{R}$*.*

*Remark* 1.7. The statement of Theorem 1.6 is the *completeness* of the real line. For a proof, see Chapter 1 of Rudin's *Principles of Mathematical Analysis*.

**Takeaway.** The least-upper-bound property is the hidden engine behind monotone convergence, nested interval arguments, and the compactness phenomena that first appear on the real line. Later abstract notions of completeness and compactness should be viewed as attempts to recover this same control in broader settings.

*Remark* 1.8 (Extended real conventions). It is often convenient to work in the extended real line $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$. Accordingly, if $A \subseteq \mathbb{R}$ is nonempty and unbounded above we write $\sup A = +\infty$, and if $A \subseteq \mathbb{R}$ is nonempty and unbounded below we write $\inf A = -\infty$. We also adopt the standard conventions

$$\sup \varnothing = -\infty, \qquad \inf \varnothing = +\infty.$$

These choices are consistent with the monotonicity properties of inf and sup under inclusion.

**Proposition 1.9** (Monotonicity of inf and sup)**.** *If* $A \subseteq B \subseteq \mathbb{R}$*, then* $\inf A \geq \inf B$ *and* $\sup A \leq \sup B$ *(with the extended-real conventions above).*

**Example 1.10.**

- If $A \neq \varnothing$ and $A$ is not bounded above, we write $\sup A = +\infty$.

- If $B \neq \varnothing$ and $B$ is not bounded below, we write $\inf B = -\infty$.
- If $A = \varnothing$, we write $\inf A = +\infty$ and $\sup A = -\infty$.

*Remark* 1.11. For instance, since $\varnothing \subseteq \{a\}$, Proposition 1.9 yields $\inf \varnothing \geq \inf\{a\} = a$ for every $a \in \mathbb{R}$, which is consistent with the convention $\inf \varnothing = +\infty$. Similarly, $\sup \varnothing \leq a$ for every $a \in \mathbb{R}$.

**Properties 1.12.**

- If $A \subseteq B \subseteq \mathbb{R}$, then $\inf A \geq \inf B$ and $\sup A \leq \sup B$.
- If $A \neq \varnothing$, then $\inf A \leq \sup A$.
- $\inf \varnothing = +\infty$ and $\sup \varnothing = -\infty$.

### 1.1.3 Archimedean property

**Theorem 1.13** (Archimedean property)**.** *Let $x > 0$ and let $y \in \mathbb{R}$. Then there exists $n \in \mathbb{N}$ such that $nx > y$. Equivalently, for every $\varepsilon > 0$ there exists $n \in \mathbb{N}$ such that $\frac{1}{n} < \varepsilon$.*

*Proof.* Suppose, for contradiction, that $nx \leq y$ for every $n \in \mathbb{N}$. Then $y$ is an upper bound for the set $S \coloneqq \{nx : n \in \mathbb{N}\}$. By Theorem 1.6, the least upper bound $\ell \coloneqq \sup S$ exists in $\mathbb{R}$. Since $\ell - x$ is not an upper bound for $S$, there exists $n \in \mathbb{N}$ such that $\ell - x < nx \leq \ell$. Consequently $(n+1)x > \ell$, contradicting the fact that $\ell$ is an upper bound of $S$. □

*Remark* 1.14. The Archimedean property formalizes the fact that the natural numbers are unbounded in $\mathbb{R}$, and it is a basic tool for constructing rational (and irrational) approximations.

**Exercise 1.15.** Let $A = \{r \in \mathbb{Q} : r^2 < 2,\ r > 0\}$. Show that $\sup A = \sqrt{2}$ (which is not in $\mathbb{Q}$).

**Example 1.16.** Let $x, y \in \mathbb{R}$ with $x < y$. Then $y - x > 0$, so by the Archimedean property there exists $n \in \mathbb{N}$ such that

$$n(y - x) > 1.$$

Hence the interval $(nx, ny)$ has length greater than 1, and therefore it contains an integer $m$. Dividing by $n$ gives

$$x < \frac{m}{n} < y.$$

Thus, between any two distinct real numbers there is a rational number.

To produce an irrational number between $x$ and $y$, choose a rational number $r$ such that

$$\frac{x}{\sqrt{2}} < r < \frac{y}{\sqrt{2}}.$$

Then

$$x < r\sqrt{2} < y,$$

and $r\sqrt{2}$ is irrational unless $r = 0$. Therefore, between any two distinct real numbers there is also an irrational number.

**Example 1.17.** Consider the set

$$A = \left\{ \frac{m}{m+n} : m, n \in \mathbb{N} \right\}.$$

Then

$$\inf A = 0, \qquad \sup A = 1.$$

Indeed, for each $n \in \mathbb{N}$ we have

$$\frac{1}{n+1} \in A,$$

and $\frac{1}{n+1} \to 0$, so $\inf A = 0$. Likewise, for each $m \in \mathbb{N}$,

$$\frac{m}{m+1} \in A,$$

and $\frac{m}{m+1} \to 1$, so $\sup A = 1$.

**Proposition 1.18** (Approximation by nearly extremal points)**.** *Let $A \subseteq \mathbb{R}$ be nonempty, and let $\alpha = \inf A$ and $\beta = \sup A$. Then for every $\varepsilon > 0$ there exist $x_0, y_0 \in A$ such that*

$$x_0 < \alpha + \varepsilon \qquad \textit{and} \qquad y_0 > \beta - \varepsilon.$$

*Proof.* We prove the first statement; the second is analogous. Suppose there exists $\varepsilon > 0$ such that $x \geq \alpha + \varepsilon$ for every $x \in A$. Then $\alpha + \varepsilon$ is a lower bound for $A$, contradicting the fact that $\alpha$ is the greatest lower bound. Hence there exists $x_0 \in A$ with $x_0 < \alpha + \varepsilon$.

Similarly, if there were no point of $A$ exceeding $\beta - \varepsilon$, then $\beta - \varepsilon$ would be an upper bound for $A$, contradicting the definition of $\beta = \sup A$. Thus there exists $y_0 \in A$ with $y_0 > \beta - \varepsilon$. □

**Definition 1.19.** A *sequence* is a function $f\colon \mathbb{N} \to \mathbb{R}$ or $f\colon \mathbb{N} \to \mathbb{C}$. We usually write it as $(f_n)_{n \geq 1}$ or $\{f_n\}$, where $f_n = f(n)$.

**Definition 1.20.** A sequence $(a_n) \subseteq \mathbb{R}$ is said to *converge* to $l \in \mathbb{R}$ if for every $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that

$$n \geq n_0 \quad \Longrightarrow \quad |a_n - l| < \varepsilon.$$

**Example 1.21.** The sequence $a_n = \frac{1}{n}$ converges to 0. Indeed, given $\varepsilon > 0$, choose $n_0 \in \mathbb{N}$ with $n_0 > \frac{1}{\varepsilon}$. Then for every $n \geq n_0$,

$$\left| \frac{1}{n} - 0 \right| = \frac{1}{n} \leq \frac{1}{n_0} < \varepsilon.$$

**Theorem 1.22** (Convergent sequences are bounded)**.** *Every convergent sequence is bounded.*

*Proof.* Let $a_n \to a$. Taking $\varepsilon = 1$, there exists $n_0 \in \mathbb{N}$ such that $|a_n - a| < 1$ for all $n \geq n_0$. Hence $a_n \in (a-1, a+1)$ for all $n \geq n_0$.

Let

$$m = \inf\big((a-1, a+1) \cup \{a_1, \ldots, a_{n_0-1}\}\big), \qquad M = \sup\big((a-1, a+1) \cup \{a_1, \ldots, a_{n_0-1}\}\big).$$

Then $m \le a_n \le M$ for every $n$, so the sequence is bounded. □

**Theorem 1.23** (Monotone convergence theorem). *If $(a_n)$ is increasing and bounded above, then $(a_n)$ converges and*

$$\lim_{n\to\infty} a_n = \sup_{n\ge 1} a_n.$$

*Similarly, a decreasing sequence that is bounded below converges to its infimum.*

*Proof.* Let

$$\alpha = \sup\{a_n : n \ge 1\}.$$

Fix $\varepsilon > 0$. By Proposition 1.18, there exists $n_0$ such that $a_{n_0} > \alpha - \varepsilon$. Since $(a_n)$ is increasing,

$$\alpha - \varepsilon < a_{n_0} \le a_n \le \alpha \qquad \text{for all } n \ge n_0.$$

Hence $|a_n - \alpha| < \varepsilon$ for all $n \ge n_0$, so $a_n \to \alpha$. The decreasing case is analogous. □

### 1.1.4 Nested Interval Theorem

**Theorem 1.24** (Nested interval theorem). *Let $(I_n)_{n\ge 1}$ be a nested sequence of nonempty closed intervals in $\mathbb{R}$, say*

$$I_n = [a_n, b_n], \qquad I_{n+1} \subseteq I_n \quad (n \ge 1).$$

*If $\ell(I_n) = b_n - a_n \to 0$ as $n \to \infty$, then $\bigcap_{n=1}^{\infty} I_n$ consists of exactly one point.*

**Proof strategy.** Track the left endpoints and right endpoints separately. Monotonicity gives convergence of each endpoint sequence, and the hypothesis $\ell(I_n) \to 0$ forces the two limits to coincide.

*Proof.* Since $I_{n+1} \subseteq I_n$, the sequence $(a_n)$ is increasing and bounded above by $b_1$, hence convergent; write $a_n \to a$. Similarly, $(b_n)$ is decreasing and bounded below by $a_1$, hence convergent; write $b_n \to b$. Taking limits in $0 \le b_n - a_n \to 0$ gives $b - a = 0$, so $a = b$.

Let $x := a$. Then $a_n \le x \le b_n$ for every $n$, hence $x \in I_n$ for every $n$, and therefore $x \in \bigcap_{n\ge 1} I_n$.

Conversely, if $y \in \bigcap_{n\ge 1} I_n$, then $a_n \le y \le b_n$ for all $n$. Passing to the limit yields $a \le y \le b$, hence $y = a = b = x$. □

**Definition 1.25.** Let $(x_n)$ be a sequence. If $n_1 < n_2 < \cdots < n_k < \cdots$ is a strictly increasing sequence of natural numbers, then the sequence $(x_{n_k})$ is called a *subsequence* of $(x_n)$.

**Example 1.26.** The sequences $\left(\frac{1}{k^2}\right)$ and $\left(\frac{1}{2^k}\right)$ are subsequences of $\left(\frac{1}{n}\right)$, corresponding respectively to the index choices $n_k = k^2$ and $n_k = 2^k$.

**Theorem 1.27** (Bolzano–Weierstrass). *Every bounded sequence in $\mathbb{R}$ admits a convergent subsequence.*

*Proof.* Let $(x_n)$ be a bounded sequence in $\mathbb{R}$. Choose $a < b$ such that $x_n \in [a, b]$ for all $n \in \mathbb{N}$, and set $I_1 = [a, b]$. Choose $n_1 \in \mathbb{N}$ arbitrarily so that $x_{n_1} \in I_1$.

Suppose that for some $k \geq 1$ we have already chosen a closed interval $I_k = [a_k, b_k]$ containing infinitely many terms of the sequence, together with an index $n_k$ such that $x_{n_k} \in I_k$. Bisect $I_k$ into two closed subintervals of equal length. At least one of these subintervals contains infinitely many terms of the sequence; denote it by $I_{k+1}$. Since $I_{k+1}$ contains infinitely many terms, we may choose $n_{k+1} > n_k$ such that $x_{n_{k+1}} \in I_{k+1}$.

In this way we obtain a nested sequence of closed intervals

$$I_1 \supseteq I_2 \supseteq \cdots ,$$

and a subsequence $(x_{n_k})$ with $x_{n_k} \in I_k$ for every $k$. Moreover,

$$\ell(I_k) = \frac{b-a}{2^{k-1}} \xrightarrow[k\to\infty]{} 0.$$

By Theorem 1.24, the intersection $\bigcap_{k=1}^{\infty} I_k$ consists of a single point, say $\{x\}$.

Now fix $\varepsilon > 0$. Choose $k_0$ such that $\ell(I_{k_0}) < \varepsilon$. Since $x \in I_{k_0}$ and $I_{k_0}$ has length less than $\varepsilon$, it follows that $I_{k_0} \subset (x - \varepsilon, x + \varepsilon)$. For every $k \geq k_0$ we have $x_{n_k} \in I_k \subseteq I_{k_0}$, and therefore $|x_{n_k} - x| < \varepsilon$. Hence $x_{n_k} \to x$ as $k \to \infty$. □

*Remark* 1.28 (Lower and upper limits)*.* Let $(x_n)$ be a bounded sequence in $\mathbb{R}$. For each $k \geq 1$, define

$$u_k := \inf\{x_n : n \geq k\}, \qquad v_k := \sup\{x_n : n \geq k\}.$$

Then $(u_k)$ is increasing and bounded above, while $(v_k)$ is decreasing and bounded below. Hence both sequences converge.

We define

$$\liminf_{n\to\infty} x_n := \lim_{k\to\infty} u_k = \sup_{k\geq 1} \inf_{n\geq k} x_n,$$

and

$$\limsup_{n\to\infty} x_n := \lim_{k\to\infty} v_k = \inf_{k\geq 1} \sup_{n\geq k} x_n.$$

Since $u_k \leq v_k$ for every $k$, it follows that

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

**Example 1.29.** For the sequence $x_n = (-1)^n$, we have

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

**Exercise 1.30.** Show that if $x_n \to x$, then

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

Deduce that a bounded sequence converges if and only if its lower and upper limits agree.

**Example 1.31.** Let $X_n = (x_n, y_n) \in \mathbb{R}^2$ be a bounded sequence. Then both coordinate sequences $(x_n)$ and $(y_n)$ are bounded. By Bolzano–Weierstrass, $(x_n)$ has a convergent subsequence $(x_{n_k})$. The corresponding subsequence $(y_{n_k})$ is still bounded, so it has a convergent subsequence $(y_{n_{k_\ell}})$. Hence

$$(x_{n_{k_\ell}}, y_{n_{k_\ell}}) \to (x, y) \in \mathbb{R}^2$$

for suitable $x, y \in \mathbb{R}$.

*Remark* 1.32. The same diagonal argument works in $\mathbb{R}^n$.

## 1.2 Countability and Cardinality

**Section overview.**

- Countability distinguishes phenomena that can be controlled by sequences from genuinely uncountable behavior.
- The aim is to understand why sets such as $\mathbb{Q}$ and $\mathbb{R}$, though both infinite, are fundamentally different in size.

In analysis one frequently distinguishes sets not only by their topological properties, but also by their *cardinality*. We recall the basic notions and the standard examples.

### 1.2.1 Countable and uncountable sets

**Definition 1.33.** A set $A$ is *countable* if there exists a bijection $A \to \mathbb{N}$. It is *at most countable* if it is finite or countable, equivalently if there exists an injection $A \hookrightarrow \mathbb{N}$. A set that is not at most countable is called *uncountable*.

**Example 1.34.** The sets $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ are countable.

*Proof.* Clearly $\mathbb{N}$ is countable. The map $n \mapsto (-1)^n \lceil n/2 \rceil$ is a bijection $\mathbb{N} \to \mathbb{Z}$, so $\mathbb{Z}$ is countable.

For $\mathbb{Q}$, write every nonzero rational in reduced form $p/q$ with $p \in \mathbb{Z} \setminus \{0\}$ and $q \in \mathbb{N}$. Consider the set $\mathbb{Z} \times \mathbb{N}$ and enumerate it along diagonals:

$$(0,1),\ (1,1),\ (-1,1),\ (0,2),\ (1,2),\ (-1,2),\ (2,1),\ (-2,1), \ldots$$

This gives a surjection $\mathbb{N} \to \mathbb{Z} \times \mathbb{N}$. Composing with $(p, q) \mapsto p/q$ yields a surjection $\mathbb{N} \to \mathbb{Q}$. Since $\mathbb{Q}$ is infinite, it follows that $\mathbb{Q}$ is countable. □

**Theorem 1.35** (Countable unions of countable sets). *A countable union of countable sets is countable. More precisely, if $A_n$ is countable for each $n \in \mathbb{N}$, then $\bigcup_{n\in\mathbb{N}} A_n$ is at most countable.*

*Proof.* For each $n \in \mathbb{N}$, choose an injection $\iota_n \colon A_n \hookrightarrow \mathbb{N}$. To avoid overlaps among the sets $A_n$, define

$$B_1 := A_1, \qquad B_n := A_n \setminus \bigcup_{k=1}^{n-1} A_k \quad (n \geq 2).$$

Then the sets $B_n$ are pairwise disjoint, each $B_n$ is at most countable, and

$$\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} B_n.$$

Now define $F\colon \bigcup_{n\in\mathbb{N}} B_n \to \mathbb{N}\times\mathbb{N}$ by

$$F(x) := (n, \iota_n(x)) \qquad \text{whenever } x \in B_n.$$

Because the family $(B_n)$ is pairwise disjoint, the map $F$ is well defined; since each $\iota_n$ is injective, $F$ is injective as well. Finally, $\mathbb{N}\times\mathbb{N}$ is countable (for example, by diagonal enumeration), so $\bigcup_{n\in\mathbb{N}} A_n$ is at most countable. □

### 1.2.2 Cantor's diagonal argument

**Theorem 1.36** (Cantor)**.** *The interval* $[0,1]$ *is uncountable. Consequently,* $\mathbb{R}$ *is uncountable.*

*Proof.* Assume, for contradiction, that $[0,1]$ is countable, so there exists an enumeration $(x_n)_{n\geq 1}$ of $[0,1]$. Write each $x_n$ in a decimal expansion

$$x_n = 0.d_{n1}d_{n2}d_{n3}\ldots,$$

choosing, when necessary, the expansion that does *not* terminate in an infinite tail of 9's. Define a new number $y \in [0,1]$ by specifying its decimal digits $(e_k)_{k\geq 1}$ via

$$e_k := \begin{cases} 1, & d_{kk} \neq 1, \\ 2, & d_{kk} = 1. \end{cases}$$

Then $y = 0.e_1e_2e_3\ldots$ differs from $x_k$ at the $k$-th digit for every $k$, hence $y \neq x_k$ for all $k$. This contradicts that $(x_n)$ enumerates $[0,1]$. □

### 1.2.3 Cardinality and comparison of sets

**Definition 1.37.** Two sets $A$ and $B$ have the same *cardinality*, written $|A| = |B|$, if there exists a bijection $A \to B$. We write $|A| \leq |B|$ if there exists an injection $A \hookrightarrow B$.

**Theorem 1.38** (Schröder–Bernstein)**.** *If* $|A| \leq |B|$ *and* $|B| \leq |A|$*, then* $|A| = |B|$.

*Proof.* We omit the standard set-theoretic proof; see any text on set theory or the appendix of a real analysis reference. The result will be used only as a conceptual tool for comparing sizes of infinite sets. □

*Remark* 1.39. Cantor's theorem shows that for any set $A$, the power set $\mathcal{P}(A)$ satisfies $|A| < |\mathcal{P}(A)|$. In particular, $|\mathbb{N}| < |\mathcal{P}(\mathbb{N})|$, and one can show that $|\mathbb{R}| = |\mathcal{P}(\mathbb{N})|$ (the *continuum*).

## 1.3 A warm-up exercise

**Exercise 1.40.** Show that the set

$$D := \left\{ \sum_{n=1}^{\infty} \frac{k_n}{2^n} : k_n \in \{0,1\},\ n = 1, 2, \ldots \right\}$$

is dense in $[0,1]$.

*Remark* 1.41. Hint: use binary expansions.

## 1.4 The Cantor Set

**Section overview.**

- The Cantor set is a fundamental example illustrating that closedness, uncountability, and positive length are genuinely distinct notions.
- It will reappear throughout the course as a standard source of counterexamples and structural insight in analysis and topology.

### 1.4.1 Cantor Set

The Cantor set is a distinguished subset of $[0,1]$ that is simultaneously closed, uncountable, and of length zero. For this reason, it occupies a central place in elementary analysis, topology, and measure theory.
Set $C_0 = [0,1]$. Remove the open middle third

$$J_1 \coloneqq \left(\tfrac{1}{3}, \tfrac{2}{3}\right)$$

and define

$$C_1 \coloneqq [0, \tfrac{1}{3}] \cup [\tfrac{2}{3}, 1].$$

At the next stage, remove the open middle third of each component interval of $C_1$, namely

$$J_2 \coloneqq \left(\tfrac{1}{9}, \tfrac{2}{9}\right) \cup \left(\tfrac{7}{9}, \tfrac{8}{9}\right),$$

and set

$$C_2 \coloneqq [0, \tfrac{1}{9}] \cup [\tfrac{2}{9}, \tfrac{1}{3}] \cup [\tfrac{2}{3}, \tfrac{7}{9}] \cup [\tfrac{8}{9}, 1].$$

Proceeding inductively, one obtains a decreasing sequence $\{C_n\}_{n\geq 0}$ in which $C_n$ is the union of $2^n$ pairwise disjoint closed intervals, each of length $3^{-n}$. The *middle-third Cantor set* is defined by the nested intersection

$$C \coloneqq \bigcap_{n=0}^{\infty} C_n.$$

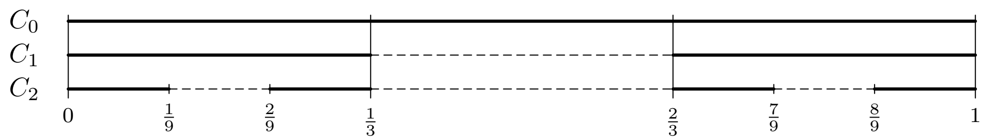



**Figure 1.1.** The initial stages of the Cantor-set construction. Dashed segments indicate the open intervals removed at each step.

### 1.4.2 Properties of the Cantor Set

We now summarize the principal structural properties of the Cantor set

$$C = \bigcap_{n=0}^{\infty} C_n.$$

1. Each $C_n$ is a nonempty compact subset of $[0,1]$, and the family is nested:

$$C_{n+1} \subseteq C_n \qquad (n \geq 0).$$

   Consequently, $C$ is a nonempty closed subset of $[0,1]$.

2. Every endpoint of every interval removed during the construction belongs to $C$.
3. If $\{J_n\}_{n\geq 1}$ denotes the collection of open intervals removed in the construction, then

$$C = [0,1] \setminus \bigcup_{n=1}^{\infty} J_n.$$

4. The Cantor set has length zero. Indeed, since $C \subseteq C_n$ for every $n$, we have

$$\ell(C) \leq \ell(C_n) = 2^n \cdot \frac{1}{3^n} = \left(\frac{2}{3}\right)^n \longrightarrow 0.$$

   It follows that $\ell(C) = 0$.

5. The set $C$ is nowhere dense. Suppose, to the contrary, that $C^\circ \neq \varnothing$. Then there exists an interval $(x-\varepsilon, x+\varepsilon) \subseteq C$. Hence

$$0 < 2\varepsilon = \ell((x-\varepsilon, x+\varepsilon)) \leq \ell(C) = 0,$$

   a contradiction. Therefore $C^\circ = \varnothing$. Since $C$ is closed, its closure is itself, and thus $C$ is nowhere dense.

6. The set $C$ is totally disconnected; that is, every connected subset of $C$ consists of a single point.

7. Every point of $C$ is a limit point of $C$. Indeed, let $x \in C$. For each $n$, the point $x$ belongs to one of the closed intervals comprising $C_n$; denote this interval by $[x_n, y_n]$. Then

$$x \in [x_n, y_n], \qquad y_n - x_n = \frac{1}{3^n},$$

so that

$$|x_n - x| \le \frac{1}{3^n} \to 0.$$

Both endpoints $x_n$ and $y_n$ belong to $C$, and for all sufficiently large $n$ at least one of them is distinct from $x$. Hence every neighborhood of $x$ contains a point of $C$ different from $x$. Thus $C$ is perfect.

If $E$ denotes the set of all endpoints of the deleted intervals, then $\overline{E} = C$. Since $E$ is countable, it follows that the Cantor set is separable.

### 1.4.3 Ternary representation of the Cantor set

We begin with the familiar identities

$$\frac{1}{3} = \frac{0}{3} + \frac{2}{3^2} + \frac{2}{3^3} + \cdots = (0.0222\ldots)_3, \qquad \frac{2}{3} = (0.2)_3.$$

More generally, every endpoint of a deleted interval admits a ternary expansion involving only the digits 0 and 2.

Let $E$ again denote the set of all endpoints of deleted intervals. Since points of $E$ possess two ternary expansions, it is convenient to exclude them when discussing uniqueness.

**Proposition 1.42** (Ternary description of the removed set)**.** *Let*

$$F := \left\{ x \in [0,1] : x = \sum_{i=1}^{\infty} \frac{a_i}{3^i},\ a_i \in \{0,1,2\} \right\} \cap E^c.$$

*Then each $x \in F$ has a unique ternary expansion. Moreover,*

$$C = \left\{ x \in [0,1] : x = \sum_{i=1}^{\infty} \frac{a_i}{3^i},\ a_i \in \{0,2\} \right\}.$$

*Proof.* Let $x \in F$ and write

$$x = \sum_{i=1}^{\infty} \frac{a_i}{3^i}, \qquad a_i \in \{0,1,2\}.$$

Observe first that $a_1 = 1$ if and only if $x \in (\frac{1}{3}, \frac{2}{3})$, equivalently, $x \notin C_1$. More generally, once the first $i-1$ digits are fixed, the condition $a_i = 1$ means precisely that $x$ lies in one of the middle-third intervals removed at stage $i$. Hence, if some ternary digit of $x$ is equal to 1, then $x \notin C$.

Conversely, if $x \notin C$, then $x \notin C_i$ for some $i$, so $x$ belongs to one of the open middle-third intervals deleted at stage $i$. This implies that the $i$-th ternary digit of $x$ is equal to 1.

Therefore $x \in C$ if and only if its ternary expansion contains only the digits 0 and 2. $\square$

**Proposition 1.43** (Uniqueness of ternary representation on $C$)**.** *If*

$$x = \sum_{i=1}^{\infty} \frac{a_i}{3^i} = \sum_{i=1}^{\infty} \frac{b_i}{3^i}, \qquad a_i, b_i \in \{0, 2\},$$

*then $a_i = b_i$ for every $i$.*

*Proof.* Assume, for contradiction, that the two sequences are distinct, and let $i_0$ be the smallest index for which $a_{i_0} \neq b_{i_0}$. Then $a_i = b_i$ for all $i < i_0$. Without loss of generality, assume that $a_{i_0} = 0$ and $b_{i_0} = 2$. After subtracting the common initial part and multiplying by $3^{i_0 - 1}$, we may reduce to the case $i_0 = 1$. Then

$$x \leq \frac{0}{3} + \frac{2}{3^2} + \frac{2}{3^3} + \cdots = \frac{1}{3}, \qquad x \geq \frac{2}{3},$$

which is impossible. Hence the representation is unique. $\square$

**Proposition 1.44** (Uncountability of the Cantor set)**.** *The Cantor set is uncountable.*

*Proof.* Define a map $f \colon C \to [0, 1]$ by the rule

$$x = \sum_{i=1}^{\infty} \frac{a_i}{3^i}, \qquad a_i \in \{0, 2\},$$

$$f(x) := \sum_{i=1}^{\infty} \frac{a_i/2}{2^i}.$$

Since $a_i/2 \in \{0, 1\}$, we indeed have $f(x) \in [0, 1]$. The preceding uniqueness result for ternary expansions on $C$ shows that $f$ is well defined.

Now let

$$y = \sum_{i=1}^{\infty} \frac{b_i}{2^i} \in [0, 1], \qquad b_i \in \{0, 1\}.$$

Define

$$x = \sum_{i=1}^{\infty} \frac{2b_i}{3^i}.$$

Then $x \in C$ and $f(x) = y$. Thus $f$ is surjective. Since $[0, 1]$ is uncountable, it follows that the Cantor set is uncountable. $\square$

**Proposition 1.45** (Monotonicity of the Cantor map)**.** *The map $f : C \to [0, 1]$ defined above is monotone increasing.*

*Proof.* Let $x, y \in C$ with $x < y$. Write

$$x = \sum_{i=1}^{\infty} \frac{a_i}{3^i}, \qquad y = \sum_{i=1}^{\infty} \frac{b_i}{3^i}, \qquad a_i, b_i \in \{0, 2\}.$$

By uniqueness of ternary representation on $C$, there exists a smallest index $n$ such that $a_n \neq b_n$. Then $a_i = b_i$ for $1 \leq i \leq n-1$, and necessarily $a_n = 0$ and $b_n = 2$. Consequently, the first binary digit at which $f(x)$ and $f(y)$ differ is 0 for $f(x)$ and 1 for $f(y)$. Therefore $f(x) < f(y)$, and hence $f$ is monotone increasing. □

*Remark* 1.46. The map $f$ is not injective, because binary expansions are not unique. For example,

$$f\Big(\frac{1}{3}\Big) = f((0.0222\ldots)_3) = (0.0111\ldots)_2 = (0.1)_2 = \frac{1}{2} = f\Big(\frac{2}{3}\Big).$$

More generally, the values of $f$ agree at the two endpoints of each deleted middle-third interval.

*Remark* 1.47 (The Cantor function). The map $f$ constructed above is the starting point for the Cantor function. One extends it from $C$ to the whole interval $[0,1]$ by declaring it to be constant on each deleted interval. The resulting function is continuous, nondecreasing, and maps $[0,1]$ onto $[0,1]$.

## 1.5 Basic Topology on $\mathbb{R}$

**Section overview.**

- Topological language organizes the local and global structure behind limits, continuity, and compactness.
- On the real line these ideas admit especially concrete descriptions in terms of intervals and closures.

### 1.5.1 Open Sets and Closed Sets

**Definition 1.48** (Open set in $\mathbb{R}$)**.** A set $O \subseteq \mathbb{R}$ is **open** if for every $x \in O$ there exists $\varepsilon > 0$ such that

$$(x-\varepsilon, x+\varepsilon) \subseteq O.$$

Equivalently, every point of $O$ is an interior point of $O$.

*Remark* 1.49. A (finite or countable) union of open intervals is an open set. Conversely, every open subset of $\mathbb{R}$ can be written as a countable union of pairwise disjoint open intervals.

**Theorem 1.50** (Decomposition of open sets in $\mathbb{R}$)**.** *Let $O \subset \mathbb{R}$ be open. Then there exists a countable family of pairwise disjoint open intervals $\{I_n\}_{n\geq 1}$ such that*

$$O = \dot{\bigcup_{n=1}^{\infty}} I_n.$$

*Moreover, this representation is unique up to a permutation of the intervals.*

*Proof.* Fix $x \in O$. Since $O$ is open, there exists an open interval $(a,b)$ with $x \in (a,b) \subset O$. Define

$$a_x := \inf\{a \in \mathbb{R} : (a,x] \subset O\}, \qquad b_x := \sup\{b \in \mathbb{R} : [x,b) \subset O\},$$

and set $I_x := (a_x, b_x)$.

*Step 1:* $I_x \subset O$. Let $z \in (a_x, b_x)$. Choose $\eta > 0$ so small that $a_x + \eta < z < b_x - \eta$. By the definition of $a_x$ as an infimum, there exists $a < a_x + \eta$ such that $(a, x] \subset O$; hence $(a_x + \eta, x] \subset O$. Similarly, by the definition of $b_x$ as a supremum, there exists $b > b_x - \eta$ such that $[x, b) \subset O$; hence $[x, b_x - \eta) \subset O$. Therefore $(a_x + \eta, b_x - \eta) \subset O$, and in particular $z \in O$. Thus $I_x \subset O$.

*Step 2: maximality and disjointness.* By construction, $I_x$ is an open interval containing $x$ and contained in $O$. Moreover, it is maximal with respect to these properties. Indeed, if $J$ is any open interval with $x \in J \subset O$, then every left endpoint of $J$ belongs to the set whose infimum defines $a_x$, and every right endpoint of $J$ belongs to the set whose supremum defines $b_x$; hence $J \subset I_x$. Now if $x, y \in O$ and $I_x \cap I_y \neq \varnothing$, then $I_x \cup I_y$ is again an open interval contained in $O$. By maximality, this forces $I_x = I_y$. Consequently, the family $\{I_x : x \in O\}$ consists of pairwise disjoint open intervals and

$$O = \bigcup_{x \in O} I_x.$$

*Step 3: countability.* Every nonempty open interval contains a rational number. Choose $q_x \in I_x \cap \mathbb{Q}$ for each $x \in O$. If $I_x \neq I_y$, then $I_x \cap I_y = \varnothing$, so necessarily $q_x \neq q_y$. Thus the assignment $I_x \mapsto q_x$ is injective into $\mathbb{Q}$. Since $\mathbb{Q}$ is countable, the family $\{I_x : x \in O\}$ is countable. Renaming these intervals as $\{I_n\}_{n \geq 1}$ gives

$$O = \dot{\bigcup_{n=1}^{\infty}} I_n.$$

*Step 4: uniqueness.* Suppose also that

$$O = \dot{\bigcup_{n=1}^{\infty}} I_n = \dot{\bigcup_{m=1}^{\infty}} J_m,$$

where both families consist of pairwise disjoint open intervals. Fix $n$. Since $I_n \subset O$, we have

$$I_n = I_n \cap O = \dot{\bigcup_{m=1}^{\infty}} (I_n \cap J_m).$$

Each set $I_n \cap J_m$ is an open subset of the interval $I_n$, and the family is pairwise disjoint. Because $I_n$ is connected, at most one of these intersections can be nonempty. Hence $I_n \subset J_{m_0}$ for some $m_0$. Applying the same argument with the roles of $I_n$ and $J_{m_0}$ reversed yields $J_{m_0} \subset I_n$. Therefore $I_n = J_{m_0}$. This proves uniqueness up to a permutation of the intervals. $\square$

**Definition 1.51** (Closed set in $\mathbb{R}$)**.** A set $F \subseteq \mathbb{R}$ is **closed** if it contains the limits of all convergent sequences of its points, that is, whenever $(x_n) \subseteq F$ and $x_n \to x$ in $\mathbb{R}$, we have $x \in F$.

**Theorem 1.52** (Closed sets and open complements)**.** *A set $F \subseteq \mathbb{R}$ is closed if and only if its complement $F^c$ is open.*

*Proof.* Assume first that $F$ is closed, and let $x \in F^c$. If $F^c$ were not open at $x$, then for every $n \in \mathbb{N}$ we could choose

$$x_n \in \left(x - \frac{1}{n},\, x + \frac{1}{n}\right) \cap F.$$

Then $x_n \to x$ with $x_n \in F$ for all $n$, so closedness of $F$ would imply $x \in F$, a contradiction. Thus $F^c$ is open.

Conversely, assume $F^c$ is open and let $(x_n) \subseteq F$ with $x_n \to x$. If $x \notin F$, then $x \in F^c$, so there exists $r > 0$ such that $(x - r, x + r) \subseteq F^c$. For $n$ large enough we have $x_n \in (x - r, x + r) \subseteq F^c$, contradicting $x_n \in F$. Therefore $x \in F$, and hence $F$ is closed. □

*Remark* 1.53. The same definitions make sense in $\mathbb{R}^n$ (and, more generally, in metric spaces).

**Example 1.54.** Let

$$A := \{(x, y) \in \mathbb{R}^2 : y = \sin(1/x),\ x \neq 0\}.$$

Then $A$ is neither open nor closed in $\mathbb{R}^2$ (with the Euclidean metric). Indeed, the sequence

$$\left(\frac{1}{n\pi}, 0\right) \in A$$

converges to $(0, 0) \notin A$, so $A$ is not closed. On the other hand, no Euclidean ball centered at a point of $A$ can lie entirely in $A$, because $A$ is the graph of a function on $\mathbb{R} \setminus \{0\}$ and therefore has empty interior in $\mathbb{R}^2$. Hence $A$ is not open.

### 1.5.2 Interior of a set

**Definition 1.55** (Interior)**.** Let $A \subseteq \mathbb{R}$. The **interior** of $A$, denoted $A^\circ$, is the union of all open sets contained in $A$:

$$A^\circ := \bigcup \{\, O \subseteq A : O \text{ is open in } \mathbb{R} \,\}.$$

Equivalently, $A^\circ$ is the largest open subset of $A$, in the sense that $A^\circ$ is open, $A^\circ \subseteq A$, and whenever $O$ is open with $O \subseteq A$, we have $O \subseteq A^\circ$.

**Example 1.56.** We have

$$\mathbb{N}^\circ = \varnothing, \qquad \mathbb{Q}^\circ = \varnothing, \qquad (\mathbb{R} \setminus \mathbb{Q})^\circ = \varnothing.$$

Moreover, in $\mathbb{R}^2$ equipped with the Euclidean metric,

$$\big\{(x, y) \in \mathbb{R}^2 : y = \sin(1/x),\, x \neq 0\big\}^\circ = \varnothing.$$

### 1.5.3 Closure of a set

**Definition 1.57** (Closure)**.** Let $A \subseteq \mathbb{R}$. The **closure** of $A$, denoted $\overline{A}$, is the smallest closed set containing $A$, that is,

$$\overline{A} := \bigcap \{\, F \subseteq \mathbb{R} : F \text{ is closed and } A \subseteq F \,\}.$$

Equivalently (in a metric space), $\overline{A}$ is the set of all limits of sequences in $A$:

$$\overline{A} = \{\, x \in \mathbb{R} : \exists (x_n) \subseteq A \text{ with } x_n \to x \,\}.$$

**Example 1.58.** Let

$$A := \{(x, \sin(1/x)) \in \mathbb{R}^2 : x \neq 0\}.$$

Then

$$\overline{A} = A \cup (\{0\} \times [-1, 1]).$$

*Proof.* Let $(x_n, \sin(1/x_n)) \in A$ be a convergent sequence in $\mathbb{R}^2$ and write

$$(x_n, \sin(1/x_n)) \to (x, y).$$

If $x \neq 0$, then $x_n \to x$ with $x_n \neq 0$ for all sufficiently large $n$, and $t \mapsto \sin(1/t)$ is continuous at $x$. Hence $\sin(1/x_n) \to \sin(1/x)$, so $(x, y) = (x, \sin(1/x)) \in A$.

If $x = 0$, then automatically $y \in [-1, 1]$ because each $\sin(1/x_n) \in [-1, 1]$ and $[-1, 1]$ is closed. Thus every limit point of $A$ lies in $A \cup (\{0\} \times [-1, 1])$, proving

$$\overline{A} \subseteq A \cup (\{0\} \times [-1, 1]).$$

For the reverse inclusion, fix $y \in [-1, 1]$ and choose $\theta \in [-\pi/2, \pi/2]$ such that $\sin\theta = y$. Define

$$x_n := \frac{1}{2\pi n + \theta} \qquad (n \geq 1).$$

Then $x_n \neq 0$, $x_n \to 0$, and

$$\sin(1/x_n) = \sin(2\pi n + \theta) = \sin\theta = y.$$

Hence $(x_n, \sin(1/x_n)) \to (0, y)$, showing $(0, y) \in \overline{A}$. Since $A \subseteq \overline{A}$, we conclude $A \cup (\{0\} \times [-1, 1]) \subseteq \overline{A}$. □

### 1.5.4 Compactness

**Definition 1.59** (Open cover)**.** Let $K \subseteq \mathbb{R}^n$. A family $\{U_i\}_{i \in I}$ of open subsets of $\mathbb{R}^n$ is called an **open cover** of $K$ if

$$K \subseteq \bigcup_{i \in I} U_i.$$

**Definition 1.60** (Compact set)**.** A set $K \subseteq \mathbb{R}^n$ is **compact** if every open cover of $K$ admits a finite subcover.

**Theorem 1.61** (Heine–Borel theorem)**.** *A subset $K \subseteq \mathbb{R}^n$ (with the Euclidean metric) is compact if and only if it is closed and bounded.*

**Idea.** In Euclidean space, compactness has two complementary interpretations: a set must prevent sequences from escaping to infinity, and it must also retain the limits of sequences that remain trapped. Boundedness addresses the first issue, closedness the second.

*Remark* 1.62. In metric spaces, compactness (Definition 1.60) is equivalent to *sequential compactness* (every sequence has a convergent subsequence with limit in the set). In $\mathbb{R}^n$, Theorem 1.61 provides the convenient characterization "compact $\iff$ closed and bounded".

*Remark* 1.63. In $\mathbb{R}$ one may equivalently say that $K \subseteq \mathbb{R}$ is compact if and only if every open cover of $K$ by open intervals admits a finite subcover. This equivalence follows from the general theory of compact metric spaces together with the Bolzano–Weierstrass theorem. Similar statements hold for compact subsets of $\mathbb{R}^n$.

**Exercise 1.64.** Show that the set

$$\overline{\{(x, y) \in \mathbb{R}^2 : y = \sin(1/x),\ x \neq 0\}}$$

is closed but not bounded.

**Example 1.65.** A subset $F \subseteq \mathbb{R}$ is closed if and only if it contains all of its adherent points; equivalently,

$$\big((x-\varepsilon, x+\varepsilon) \cap F \neq \varnothing \text{ for every } \varepsilon > 0\big) \implies x \in F.$$

*Proof.* Assume first that $F$ is closed, and suppose that every interval $(x-\varepsilon, x+\varepsilon)$ meets $F$. For each $n \in \mathbb{N}$, choose

$$x_n \in \left(x - \frac{1}{n}, x + \frac{1}{n}\right) \cap F.$$

Then $|x_n - x| < 1/n$ for every $n$, so $x_n \to x$. Since $F$ is closed and each $x_n$ belongs to $F$, we conclude that $x \in F$.

Conversely, assume that $F$ satisfies the displayed condition, and let $(x_n) \subseteq F$ be a sequence with $x_n \to x$. Fix $\varepsilon > 0$. For sufficiently large $n$ we have $x_n \in (x-\varepsilon, x+\varepsilon)$, so $(x-\varepsilon, x+\varepsilon) \cap F \neq \varnothing$. By hypothesis, this implies $x \in F$. Therefore $F$ is closed. □

### 1.5.5 Dense Set

**Definition 1.66.** A subset $A \subseteq \mathbb{R}$ is called *dense in* $\mathbb{R}$ if $\overline{A} = \mathbb{R}$. Equivalently, $A$ is dense in $\mathbb{R}$ if every nonempty open interval in $\mathbb{R}$ contains a point of $A$.

**Example 1.67.** Let $x \in \mathbb{R}$. Choose a decimal expansion

$$x = x_0 + \frac{x_1}{10} + \frac{x_2}{10^2} + \cdots, \qquad x_0 \in \mathbb{Z},\quad x_k \in \{0, 1, \ldots, 9\}\ (k \geq 1). \tag{1.2}$$

For $n \geq 0$, define the rational truncations

$$S_n \coloneqq x_0 + \sum_{k=1}^{n} \frac{x_k}{10^k} \in \mathbb{Q}.$$

Then $S_n \to x$ as $n \to \infty$, so $\overline{\mathbb{Q}} = \mathbb{R}$.

Similarly, $\mathbb{R} \setminus \mathbb{Q}$ is dense in $\mathbb{R}$. Indeed, fix $x \in \mathbb{R}$ and define

$$u_n := x + \frac{\sqrt{2}}{n}, \qquad v_n := x + \frac{\sqrt{3}}{n} \qquad (n \geq 1).$$

For each $n$, at least one of $u_n$ or $v_n$ is irrational: if both were rational, then their difference $(\sqrt{2} - \sqrt{3})/n$ would be rational, which is impossible. Choosing, for each $n$, an irrational number among $\{u_n, v_n\}$ produces a sequence of irrationals converging to $x$. Hence $\overline{\mathbb{R} \setminus \mathbb{Q}} = \mathbb{R}$.

Finally, the representation (1.2) is not unique; for instance, $0.5 = 0.4999\ldots$.

**Theorem 1.68** (Base-$p$ expansion on $[0,1]$)**.** *Let $p \in \mathbb{Z}$ with $p \geq 2$, and let $x \in [0,1]$. Then there exists a sequence of integers $(a_n)_{n \geq 1}$ such that $0 \leq a_n \leq p-1$ for every $n$ and*

$$x = \sum_{n=1}^{\infty} \frac{a_n}{p^n}.$$

*Proof.* We construct the digits inductively. Suppose that integers $a_1, \ldots, a_{n-1}$ have already been chosen so that

$$0 \leq x - \sum_{k=1}^{n-1} \frac{a_k}{p^k} \leq \frac{1}{p^{n-1}}.$$

Multiply this inequality by $p^n$ and set

$$r_n := p^n \left( x - \sum_{k=1}^{n-1} \frac{a_k}{p^k} \right).$$

Then $0 \leq r_n \leq p$. Choose $a_n \in \{0, 1, \ldots, p-1\}$ so that

$$a_n \leq r_n < a_n + 1.$$

Such a choice is possible because the interval $[0,p]$ is covered by the half-open intervals $[j, j+1)$ for $j = 0, 1, \ldots, p-1$, together with the endpoint $p$, for which we may take $a_n = p-1$. It follows that

$$0 \leq r_n - a_n < 1,$$

and hence, after dividing by $p^n$,

$$0 \leq x - \sum_{k=1}^{n} \frac{a_k}{p^k} < \frac{1}{p^n}.$$

Thus the partial sums

$$S_n := \sum_{k=1}^{n} \frac{a_k}{p^k}$$

satisfy $0 \leq x - S_n < p^{-n}$ for every $n$. Since $p^{-n} \to 0$, we obtain $S_n \to x$. Therefore

$$x = \sum_{n=1}^{\infty} \frac{a_n}{p^n},$$

as required. $\square$

**Exercise 1.69.** Show that $\left\{\frac{k}{2^n} : k = 0, 1, 2, \ldots, 2^n; n = 1, 2, \ldots\right\}$ is dense in $[0, 1]$.
*(Hint: Use binary expansion)*

## Exercises

**Exercise 1.70.** Prove the Archimedean property: for every $x \in \mathbb{R}$ there exists $n \in \mathbb{N}$ with $n > x$. Deduce that for every $\varepsilon > 0$ there exists $n \in \mathbb{N}$ such that $\frac{1}{n} < \varepsilon$.

**Exercise 1.71.** Let $(I_n)_{n\geq 1}$ be a nested sequence of nonempty closed intervals with $I_{n+1} \subset I_n$ and $\ell(I_n) \to 0$. Prove that $\bigcap_{n=1}^{\infty} I_n$ consists of a single point (Nested Interval Theorem).

**Exercise 1.72.** Show that a subset $A \subset \mathbb{R}$ is open if and only if it is a (possibly empty) countable disjoint union of open intervals.

**Exercise 1.73.** Let $A \subset \mathbb{R}$. Prove that $x \in \overline{A}$ if and only if every open interval containing $x$ meets $A$.

**Exercise 1.74.** Show that $\mathbb{Q}$ is dense in $\mathbb{R}$, and that $\mathbb{R} \setminus \mathbb{Q}$ is also dense in $\mathbb{R}$.

**Exercise 1.75.** Let $C$ be the middle-third Cantor set. Prove that $C$ is closed and has empty interior.

**Exercise 1.76.** Prove that every $x \in C$ admits a ternary expansion using only digits 0 and 2, and conversely. Deduce that $C$ is uncountable.

**Exercise 1.77.** Show that $C$ is perfect: it is closed and every point of $C$ is a limit point of $C$.

**Exercise 1.78.** Prove that if $A \subset \mathbb{R}$ is closed and bounded, then $A$ is compact (Heine–Borel in $\mathbb{R}$).

**Exercise 1.79.** Let $A \subset \mathbb{R}$ be nonempty and bounded above. Prove that $\sup A$ is characterized by: (i) $a \leq \sup A$ for all $a \in A$; (ii) for every $\varepsilon > 0$ there exists $a \in A$ with $\sup A - \varepsilon < a \leq \sup A$.

Chapter 2

# Metric and Normed Linear Spaces

*This chapter passes from the concrete setting of the real line to the abstract frameworks of metric spaces and normed linear spaces. We introduce metrics, open balls, and the topology generated by a metric, and we formulate convergence, continuity, and Cauchy conditions in that setting. We then develop the geometry of normed spaces and the classical inequalities—Young, Hölder, and Minkowski—that govern $\ell^p$ and $L^p$ analysis. These structures provide the language in which the later theory of completeness, compactness, and function spaces is naturally expressed.*

**Learning objectives.**

- Move fluently between metric language, topological language, and norm-induced geometry.
- Prove continuity and convergence statements directly from the definitions and compare them across inequivalent metrics.
- Recognize the role of classical inequalities in turning linear structure into quantitative analytic control.

## 2.1 Metric Spaces and Topology

**Section overview.**

- This section builds the abstract language in which convergence, continuity, and compactness will later be phrased.
- The key task is to learn which arguments depend only on the metric and which use additional linear structure.
- As you read, compare every new definition with its familiar version on $\mathbb{R}^n$.

*We introduce metric spaces as the common framework for convergence, continuity, and compactness. After developing the topology induced by a metric (open and closed sets, closure, density), we study continuous maps between metric spaces and the basic invariance properties they enjoy.*

### 2.1.1 Metric spaces

#### 2.1.1.1 Metrics and basic examples

**Definition 2.1** (Metric space)**.** Let $X$ be a nonempty set. A function $d\colon X \times X \to [0, \infty)$ is called a *metric* on $X$ if, for all $x, y, z \in X$,

(i) $d(x, y) = 0$ if and only if $x = y$.

(ii) $d(x, y) = d(y, x)$ (symmetry).

(iii) $d(x,z) \le d(x,y) + d(y,z)$ (triangle inequality).

The pair $(X,d)$ is called a *metric space*.

**Example 2.2** (Standard metrics on $\mathbb{R}^n$)**.** For $x = (x_1,\dots,x_n)$ and $y = (y_1,\dots,y_n)$ in $\mathbb{R}^n$, define

$$\begin{aligned} d_1(x,y) &:= \sum_{i=1}^n |x_i - y_i|, \\ d_2(x,y) &:= \Big(\sum_{i=1}^n |x_i - y_i|^2\Big)^{1/2}, \\ d_\infty(x,y) &:= \max_{1\le i\le n} |x_i - y_i|. \end{aligned}$$

Each of these functions defines a metric on $\mathbb{R}^n$.

**Exercise 2.3.** Let $(X,d)$ be a metric space. Prove that $d'(x,y) := \min\{1, d(x,y)\}$ defines a metric on $X$.

**Example 2.4** (Supremum metric on $C[0,1]$)**.** Let $X = C[0,1]$, the space of continuous real-valued functions on $[0,1]$. For $f,g \in X$, set

$$d_\infty(f,g) := \sup_{0\le t\le 1} |f(t) - g(t)|.$$

Since every continuous function on $[0,1]$ is bounded, the supremum is finite. The triangle inequality follows from

$$|f(t) - h(t)| \le |f(t) - g(t)| + |g(t) - h(t)|, \qquad t \in [0,1].$$

Thus $d_\infty$ is a metric on $C[0,1]$.

**Example 2.5** (Discrete metric)**.** Let $X \neq \varnothing$. Define, for $x,y \in X$,

$$d_0(x,y) := \begin{cases} 1, & x \neq y, \\ 0, & x = y. \end{cases}$$

Then $d_0$ is a metric on $X$. The metric space $(X, d_0)$ is called the *discrete metric space*. In particular, every nonempty set admits at least one metric.

*Remark* 2.6. For the discrete metric $d_0$, the triangle inequality is immediate. Indeed, if $x = z$, then $d_0(x,z) = 0$. If $x \neq z$, then $d_0(x,z) = 1$, while $d_0(x,y) + d_0(y,z) \ge 1$. Thus $d_0(x,z) \le d_0(x,y) + d_0(y,z)$ in every case.

**Example 2.7.** If $(X,d)$ is a metric space, then

$$\widetilde{d}(x,y) := \frac{d(x,y)}{1 + d(x,y)}$$

is also a metric on $X$.

*Proof.* Consider the function $f(t) = \frac{t}{1+t}$ on $[0, \infty)$. It is increasing, satisfies $f(0) = 0$, and one checks that

$$f(s+t) \le f(s) + f(t) \qquad (s, t \ge 0).$$

Therefore $\widetilde{d} = f \circ d$ is a metric by Proposition 2.8. □

**Proposition 2.8** (Changing a metric by a subadditive function)**.** *Let $(X, d)$ be a metric space and let $f\colon [0, \infty) \to [0, \infty)$ be an increasing function such that*

$$f(s+t) \le f(s) + f(t) \qquad \text{for all } s, t \ge 0, \tag{2.1}$$

*and $f(t) = 0$ if and only if $t = 0$. Then $d_f := f \circ d$ defines a metric on $X$.*

*Proof.* Non-negativity and symmetry are immediate. If $d_f(x, y) = 0$, then $f(d(x, y)) = 0$, hence $d(x, y) = 0$ and $x = y$. Finally, by the triangle inequality for $d$ and the monotonicity and subadditivity (2.1) of $f$,

$$\begin{aligned} d_f(x, z) &= f(d(x, z)) \\ &\le f(d(x, y) + d(y, z)) \\ &\le f(d(x, y)) + f(d(y, z)) \\ &= d_f(x, y) + d_f(y, z). \end{aligned}$$

□

**Example 2.9.** Let $H^\infty$ denote the set of all sequences $x = (x_n)$ with $|x_n| \le 1$ for every $n$. Define

$$d(x, y) := \sum_{n=1}^{\infty} \frac{|x_n - y_n|}{2^n}.$$

Then $d$ is a metric on $H^\infty$.

*Proof.* Since $|x_n - y_n| \le 2$ for every $n$, the series converges absolutely and

$$d(x, y) \le \sum_{n=1}^{\infty} \frac{2}{2^n} < \infty.$$

Positivity and symmetry are immediate.

If $d(x, y) = 0$, then each summand must be zero, so $x_n = y_n$ for all $n$; hence $x = y$.

Finally, for $x, y, z \in H^\infty$ and each $k$,

$$\sum_{n=1}^{k} \frac{|x_n - z_n|}{2^n} \le \sum_{n=1}^{k} \frac{|x_n - y_n|}{2^n} + \sum_{n=1}^{k} \frac{|y_n - z_n|}{2^n}.$$

Letting $k \to \infty$ gives

$$d(x, z) \le d(x, y) + d(y, z).$$

Therefore $d$ is a metric. $\square$

**Exercise 2.10.** Prove that $d(x,y) = \left|\frac{1}{x} - \frac{1}{y}\right|$ defines a metric on $(0,\infty)$.

#### 2.1.1.2 Open balls and open sets

**Definition 2.11** (Open and closed balls)**.** Let $(X,d)$ be a metric space, let $x_0 \in X$, and let $r > 0$. The set

$$B_r(x_0) := \{y \in X : d(x_0,y) < r\}$$

is called the *open ball* of radius $r$ centred at $x_0$. The set

$$\overline{B_r(x_0)} := \{y \in X : d(x_0,y) \le r\}$$

is called the *closed ball* of radius $r$ centred at $x_0$.

**Definition 2.12** (Open sets)**.** A subset $O \subseteq X$ is called *open* if for every $x \in O$ there exists $r > 0$ such that

$$B_r(x) \subseteq O.$$

**Proposition 2.13** (Stability properties of open sets)**.** *Let $\{O_i : i \in I\}$ be a family of open subsets of a metric space $(X,d)$. Then:*

*(i) $\bigcup_{i\in I} O_i$ is open.*

*(ii) $\bigcap_{i=1}^{n} O_i$ is open for every finite collection $O_1,\ldots,O_n$.*

*Arbitrary intersections of open sets need not be open.*

**Example 2.14.** In $(\mathbb{R},|\cdot|)$,

$$\bigcap_{n=1}^{\infty}\left(-\frac{1}{n},\frac{1}{n}\right) = \{0\},$$

and the singleton $\{0\}$ is not open.

**Example 2.15.** Let $f\colon \mathbb{R}\to\mathbb{R}$ be continuous. Then the set

$$A := \{x \in \mathbb{R} : f(x) > 0\}$$

is open.

*Proof.* Fix $x \in A$. Then $f(x) > 0$. By continuity of $f$ at $x$, with $\varepsilon = f(x)$, there exists $\delta > 0$ such that

$$|y - x| < \delta \quad \Longrightarrow \quad |f(y) - f(x)| < f(x).$$

For such $y$ we obtain

$$f(y) > f(x) - |f(y) - f(x)| > 0.$$

Hence $(x-\delta, x+\delta) \subseteq A$. Since $x \in A$ was arbitrary, $A$ is open. $\square$

***Open Sets in $\mathbb{R}$:***

**Theorem 2.16** (Decomposition of open sets in $\mathbb{R}$)**.** *Let $O \subset \mathbb{R}$ be open. Then there exists a countable family of pairwise disjoint open intervals, possibly unbounded, $\{I_n\}_{n\geq 1}$ such that*

$$O = \dot{\bigcup_{n=1}^{\infty}} I_n.$$

*If $O = \varnothing$, this family is empty. Moreover, this representation is unique up to a permutation of the intervals.*

*Proof.* If $O = \varnothing$, there is nothing to prove. Assume $O \neq \varnothing$.

For each $x \in O$, define

$$I_x := \bigcup \{J \subseteq O : J \text{ is an open interval and } x \in J\}.$$

This union is nonempty because $O$ is open, so there exists $\delta > 0$ such that $(x-\delta, x+\delta) \subseteq O$. Since $I_x$ is a union of open sets, it is open; clearly $x \in I_x \subseteq O$.

*Step 1: $I_x$ is an interval.* Let $u, v \in I_x$ with $u < v$. Then there exist open intervals $J_u, J_v \subseteq O$ such that

$$x, u \in J_u, \qquad x, v \in J_v.$$

Because both intervals contain $x$, they intersect. Hence $J_u \cup J_v$ is an open interval contained in $O$. Since $u, v \in J_u \cup J_v$, every $t$ with $u < t < v$ also lies in $J_u \cup J_v \subseteq I_x$. Therefore $I_x$ is an open interval.

*Step 2: maximality and disjointness.* The interval $I_x$ is maximal among open intervals contained in $O$ and containing $x$: if $J \subseteq O$ is an open interval with $x \in J$, then $J$ appears in the defining union, so $J \subseteq I_x$. If $x, y \in O$ and $I_x \cap I_y \neq \varnothing$, then $I_x \cup I_y$ is an open interval contained in $O$ that contains both $x$ and $y$. By maximality, $I_x \subseteq I_y$ and $I_y \subseteq I_x$, hence $I_x = I_y$. Consequently, the distinct intervals among $\{I_x : x \in O\}$ are pairwise disjoint, and

$$O = \bigcup_{x\in O} I_x.$$

*Step 3: countability.* Every nonempty open interval in $\mathbb{R}$ contains a rational number. For each $q \in \mathbb{Q} \cap O$, consider the interval $I_q$. Since $\mathbb{Q} \cap O$ is countable, the family $\{I_q : q \in \mathbb{Q} \cap O\}$ is at most countable. Conversely, if $I_x$ is one of the intervals above, choose $q \in I_x \cap \mathbb{Q}$. Then $q \in O$ and $I_q = I_x$. Hence every distinct interval occurs as $I_q$ for some rational $q \in O$. Therefore the collection of distinct intervals is countable. Renaming these intervals as $\{I_n\}_{n\geq 1}$ gives the desired representation.

*Step 4: uniqueness.* Suppose

$$O = \dot{\bigcup_{n=1}^{\infty}} I_n = \dot{\bigcup_{m=1}^{\infty}} J_m,$$

where $\{I_n\}$ and $\{J_m\}$ are pairwise disjoint families of open intervals. Fix $n$. Then

$$I_n = I_n \cap O = \dot{\bigcup_{m=1}^{\infty}} (I_n \cap J_m).$$

Each set $I_n \cap J_m$ is open in the relative topology of $I_n$, and the family is pairwise disjoint. Since $I_n$ is an interval, it is connected. Hence only one of these intersections can be nonempty. Thus $I_n \subseteq J_{m_0}$ for some $m_0$. By the same argument with the roles reversed, $J_{m_0} \subseteq I_n$, and therefore $I_n = J_{m_0}$. This proves uniqueness up to a permutation. □

#### 2.1.1.3 Sequences in metric spaces

**Definition 2.17** (Convergent sequence)**.** A sequence $(x_n)$ in a metric space $(X, d)$ converges to $x_0 \in X$ if for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$n \geq N \quad \Longrightarrow \quad d(x_n, x_0) < \varepsilon.$$

**Definition 2.18** (Cauchy sequence)**.** A sequence $(x_n)$ in $(X, d)$ is a *Cauchy sequence* if for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$m, n \geq N \quad \Longrightarrow \quad d(x_n, x_m) < \varepsilon.$$

**Example 2.19.** In the metric space $((0, 1), |\cdot|)$, the sequence $x_n = \frac{1}{n}$ is Cauchy, but it does not converge in $(0, 1)$, because its limit in $\mathbb{R}$ is $0 \notin (0, 1)$.

Every convergent sequence is Cauchy.

**Definition 2.20** (Bounded set)**.** A subset $A \subseteq X$ is called *bounded* if there exist $x_0 \in X$ and $M > 0$ such that

$$d(a, x_0) \leq M \qquad \text{for all } a \in A.$$

**Example 2.21.** The set

$$\left\{ \left(x, \sin \frac{1}{x}\right) : 0 < |x| \leq 1 \right\} \cup (\{0\} \times [-1, 1])$$

is bounded in $\mathbb{R}^2$.

**Proposition 2.22** (Cauchy sequences are bounded)**.** *Every Cauchy sequence is bounded.*

*Proof.* Let $(x_n)$ be a Cauchy sequence. Taking $\varepsilon = 1$, there exists $N$ such that

$$d(x_n, x_m) < 1 \qquad \text{whenever } m, n \geq N.$$

In particular,

$$d(x_n, x_N) < 1 \qquad \text{for all } n \geq N.$$

Set

$$M := \max\{1, d(x_1, x_N), \dots, d(x_{N-1}, x_N)\}.$$

Then $d(x_n, x_N) \le M$ for every $n$, so the sequence is bounded. $\square$

The converse need not hold. For example, the sequence $(-1)^n$ is bounded in $\mathbb{R}$ but not Cauchy.

**Proposition 2.23** (A Cauchy sequence with a convergent subsequence converges)**.** *Let $(x_n)$ be a Cauchy sequence in $(X, d)$. If a subsequence $(x_{n_k})$ converges to $x \in X$, then $x_n \to x$.*

*Proof.* Fix $\varepsilon > 0$. Since $(x_n)$ is Cauchy, there exists $N_1$ such that

$$d(x_n, x_m) < \frac{\varepsilon}{2} \qquad \text{for all } m, n \ge N_1.$$

Since $x_{n_k} \to x$, there exists $K$ such that

$$d(x_{n_k}, x) < \frac{\varepsilon}{2} \qquad \text{for all } k \ge K.$$

Choose $k$ so large that $n_k \ge N_1$ and $k \ge K$. Then for every $n \ge N_1$,

$$d(x_n, x) \le d(x_n, x_{n_k}) + d(x_{n_k}, x) < \varepsilon.$$

Hence $x_n \to x$. $\square$

In particular, a Cauchy sequence can have at most one limit.

#### 2.1.1.4 Closed sets in metric spaces

**Definition 2.24.** A subset $F \subseteq X$ is called *closed* if its complement $F^c = X \setminus F$ is open. Equivalently, $F$ is closed if whenever $x \in X$ satisfies

$$B_\varepsilon(x) \cap F \neq \varnothing \qquad \text{for every } \varepsilon > 0,$$

then necessarily $x \in F$.

**Example 2.25.** The set

$$A = \left\{(x, y) : y = \sin\frac{1}{x},\ x \neq 0\right\} \subset \mathbb{R}^2$$

is neither open nor closed. It is not closed because

$$\left(\frac{1}{n\pi}, 0\right) \in A \qquad \text{and} \qquad \left(\frac{1}{n\pi}, 0\right) \to (0,0) \notin A.$$

It is not open because no Euclidean ball centred at a point of $A$ can be contained in the graph.

**Theorem 2.26** (Sequential characterization of closed sets)**.** *Let $(X, d)$ be a metric space and let $F \subseteq X$. The following are equivalent:*

*(i) $F$ is closed.*

*(ii) For every $x \in X$, if $B_\varepsilon(x) \cap F \neq \varnothing$ for every $\varepsilon > 0$, then $x \in F$.*

*(iii) Whenever $(x_n) \subseteq F$ and $x_n \to x$ in $X$, we have $x \in F$.*

*Proof. (i) $\Rightarrow$ (ii).* Assume $F$ is closed, so $F^c$ is open. If $x \notin F$, then $x \in F^c$, and therefore there exists $\varepsilon_0 > 0$ such that

$$B_{\varepsilon_0}(x) \subseteq F^c.$$

Hence $B_{\varepsilon_0}(x) \cap F = \varnothing$, contradicting (ii). So $x \in F$.

*(ii) $\Rightarrow$ (iii).* Let $(x_n) \subseteq F$ and suppose $x_n \to x$. Given $\varepsilon > 0$, there exists $N$ such that $x_n \in B_\varepsilon(x)$ for all $n \geq N$. Thus $B_\varepsilon(x) \cap F \neq \varnothing$ for every $\varepsilon > 0$, and (ii) implies $x \in F$.

*(iii) $\Rightarrow$ (i).* Suppose $F$ is not closed. Then $F^c$ is not open, so there exists $x \in F^c$ such that for every $n \in \mathbb{N}$ one can choose $x_n \in F$ with

$$d(x_n, x) < \frac{1}{n}.$$

Then $x_n \to x$, and by (iii) we conclude that $x \in F$, a contradiction. Therefore $F$ is closed. $\square$

**Example 2.27.** If $f\colon \mathbb{R} \to \mathbb{R}$ is continuous, then the zero set

$$A = \{x \in \mathbb{R} : f(x) = 0\}$$

is closed. Indeed, if $x_n \in A$ and $x_n \to x$, then continuity of $f$ gives

$$f(x) = \lim_{n\to\infty} f(x_n) = 0,$$

so $x \in A$.

#### 2.1.1.5 Interior points and interior of a set

Let $A \subseteq X$. The *interior* of $A$, denoted $A^\circ$ or $\mathrm{Int}(A)$, is the largest open set contained in $A$:

$$A^\circ = \bigcup\{O \subseteq X : O \text{ is open and } O \subseteq A\}.$$

Equivalently, $x \in A^\circ$ if and only if there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq A$.

#### 2.1.1.6 Closure and limit points

Let $A \subseteq (X, d)$. The *closure* of $A$, denoted $\overline{A}$, is the smallest closed set containing $A$:

$$\overline{A} = \bigcap\{F \subseteq X : F \text{ is closed and } A \subseteq F\}.$$

**Definition 2.28** (Limit point / accumulation point)**.** A point $x \in X$ is called a *limit point* of $A$ if for every $r > 0$,

$$(B(x,r) \setminus \{x\}) \cap A \neq \varnothing.$$

The set of all limit points of $A$ is denoted by $A'$.

*Remark* 2.29. If $x$ is a limit point of $A$, then for each $n \in \mathbb{N}$ one can choose

$$x_n \in (B(x,1/n) \setminus \{x\}) \cap A.$$

Then $x_n \to x$.

**Exercise 2.30.** Show that every finite subset of a metric space has no limit points.

**Exercise 2.31.** Let

$$A = \left\{ \left(n, \frac{1}{n}\right) : n \in \mathbb{N} \right\} \subset \mathbb{R}^2.$$

Show that $\overline{A} = A$ and $A^\circ = \varnothing$.

**Example 2.32.** 1. If

$$A = \{(x,y) : |x| < 1,\ |y| < 1\},$$

then

$$\overline{A} = \{(x,y) : |x| \leq 1,\ |y| \leq 1\}.$$

2. If

$$A = \left\{ (x,y) : y = \sin\left(\frac{1}{x}\right),\ x \neq 0 \right\},$$

then

$$\overline{A} = \left\{ (x,y) : y = \sin\left(\frac{1}{x}\right),\ x \neq 0 \right\} \cup (\{0\} \times [-1,1]).$$

**Proposition 2.33** (Closure detected by balls)**.** *Let $A \subseteq (X,d)$. Then $x \in \overline{A}$ if and only if*

$$B_\varepsilon(x) \cap A \neq \varnothing \qquad \textit{for every } \varepsilon > 0.$$

*Proof.* If $x \in \overline{A}$ and there existed $\varepsilon_0 > 0$ such that $B_{\varepsilon_0}(x) \cap A = \varnothing$, then $A \subseteq (B_{\varepsilon_0}(x))^c$. Since $(B_{\varepsilon_0}(x))^c$ is closed, minimality of $\overline{A}$ would imply

$$\overline{A} \subseteq (B_{\varepsilon_0}(x))^c,$$

contradicting $x \in \overline{A}$.

Conversely, if every ball centred at $x$ meets $A$, then every ball centred at $x$ meets the closed set $\overline{A}$. Hence $x \in \overline{A}$. □

**Proposition 2.34** (Closure as a set together with its limit points)**.** *For every subset $A \subseteq X$,*

$$\overline{A} = A \cup A'.$$

*Proof.* If $x \in A'$, then every ball centred at $x$ meets $A$, so $x \in \overline{A}$ by Proposition 2.33. Thus $A \cup A' \subseteq \overline{A}$.

Conversely, let $x \in \overline{A}$. If $x \in A$, then clearly $x \in A \cup A'$. If $x \notin A$, then every ball centred at $x$ meets $A$ at a point different from $x$, so $x \in A'$. Hence $\overline{A} \subseteq A \cup A'$. □

**Proposition 2.35** (Sequential characterization of closure)**.** *A point $x$ belongs to $\overline{A}$ if and only if there exists a sequence $(x_n)$ in $A$ such that $x_n \to x$.*

*Proof.* If $x \in \overline{A}$, then for each $n$ we may choose $x_n \in B_{1/n}(x) \cap A$. Then $d(x_n, x) < 1/n$, so $x_n \to x$.

Conversely, if $x_n \in A$ and $x_n \to x$, then every neighborhood of $x$ contains $x_n$ for all sufficiently large $n$. Thus every ball centred at $x$ meets $A$, so $x \in \overline{A}$. □

#### 2.1.1.7 Dense subsets

A set $A \subseteq X$ is called *dense* in $X$ if $\overline{A} = X$. Equivalently, $A$ is dense in $X$ if every nonempty open ball meets $A$.

**Example 2.36.** The set $\mathbb{Q}$ is dense in $\mathbb{R}$. Indeed, every real number can be approximated by its finite decimal truncations, which are rational.

**Example 2.37.** If $1 \le p < \infty$, then $\overline{c_{00}} = \ell^p$. For $x = (x_1, x_2, \dots) \in \ell^p$, define

$$X^{(n)} := (x_1, \dots, x_n, 0, 0, \dots) \in c_{00}.$$

Then

$$\|x - X^{(n)}\|_p^p = \sum_{k=n+1}^{\infty} |x_k|^p \to 0.$$

Hence $X^{(n)} \to x$ in $\ell^p$.

**Example 2.38.** We also have $\overline{c_{00}} = c_0$ in the supremum norm. If $x = (x_n) \in c_0$, define

$$X^{(n)} := (x_1, \dots, x_n, 0, 0, \dots).$$

Since $x_n \to 0$,

$$\|x - X^{(n)}\|_\infty = \sup_{k \ge n+1} |x_k| \to 0.$$

Therefore $X^{(n)} \to x$ in $c_0$.

*Remark* 2.39. Thus $\overline{c_{00}} = c_0 \subsetneq \ell^\infty$, so $c_{00}$ is not dense in $\ell^\infty$.

### 2.1.2 Continuous maps between metric spaces

A function $f\colon (X, d) \to (\mathbb{R}, u)$ is said to be *continuous at* $x_0 \in X$ if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$d(x_0, y) < \delta \quad \Longrightarrow \quad |f(y) - f(x_0)| < \varepsilon.$$

Equivalently,

$$f(B_\delta(x_0)) \subseteq (f(x_0) - \varepsilon,\ f(x_0) + \varepsilon).$$

**Example 2.40.** Fix $x_0 \in X$ and define $f(y) = d(x_0, y)$. Then $f$ is continuous on $X$. Indeed, the triangle inequality gives

$$|d(x_0, y) - d(x_0, z)| \leq d(y, z) \qquad \text{for all } y, z \in X.$$

Thus $f$ is in fact Lipschitz continuous with constant 1.

**Theorem 2.41** (Equivalent characterizations of continuity)**.** *Let $f\colon (X, d) \to (\mathbb{R}, u)$. The following statements are equivalent:*

*(i) $f$ is continuous on $X$ in the $\varepsilon$–$\delta$ sense.*

*(ii) Whenever $(x_n)$ is a sequence in $X$ with $x_n \to x$, one has $f(x_n) \to f(x)$.*

*(iii) For every open set $O \subseteq \mathbb{R}$, the preimage $f^{-1}(O)$ is open in $X$.*

*(iv) For every closed set $F \subseteq \mathbb{R}$, the preimage $f^{-1}(F)$ is closed in $X$.*

*Proof.* Recall that $f$ is continuous at $x \in X$ if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$d(x, y) < \delta \quad \Longrightarrow \quad |f(y) - f(x)| < \varepsilon.$$

**(i) ⇒ (ii).** Assume that $f$ is continuous on $X$, and let $(x_n) \subset X$ satisfy $x_n \to x$. Fix $\varepsilon > 0$. By continuity of $f$ at $x$, there exists $\delta > 0$ such that

$$d(x, y) < \delta \quad \Longrightarrow \quad |f(y) - f(x)| < \varepsilon.$$

Since $x_n \to x$, there exists $N \in \mathbb{N}$ such that $d(x_n, x) < \delta$ for all $n \geq N$. Hence $|f(x_n) - f(x)| < \varepsilon$ for all $n \geq N$, and therefore $f(x_n) \to f(x)$.

**(ii) ⇒ (i).** Assume (ii), fix $x \in X$, and suppose for contradiction that $f$ is not continuous at $x$. Then there exists $\varepsilon_0 > 0$ such that for every $\delta > 0$ one can find $y \in X$ with

$$d(x, y) < \delta \qquad \text{and} \qquad |f(y) - f(x)| \geq \varepsilon_0.$$

For each $n \in \mathbb{N}$, choose $y_n \in X$ with

$$d(x, y_n) < \frac{1}{n} \qquad \text{and} \qquad |f(y_n) - f(x)| \geq \varepsilon_0.$$

Then $y_n \to x$, whereas $f(y_n) \not\to f(x)$, contradicting (ii). Thus $f$ is continuous at $x$. Since $x$ was arbitrary, $f$ is continuous on $X$.

**(i) ⇒ (iii).** Assume that $f$ is continuous on $X$, and let $O \subseteq \mathbb{R}$ be open. Take $x \in f^{-1}(O)$, so $f(x) \in O$. Because $O$ is open, there exists $\varepsilon > 0$ such that

$$(f(x) - \varepsilon,\ f(x) + \varepsilon) \subseteq O.$$

By continuity of $f$ at $x$, there exists $\delta > 0$ such that

$$d(x, y) < \delta \quad \Longrightarrow \quad |f(y) - f(x)| < \varepsilon.$$

Thus $f(y) \in O$ whenever $d(x, y) < \delta$, and therefore

$$B_\delta(x) \subseteq f^{-1}(O).$$

Hence $f^{-1}(O)$ is open in $X$.

**(iii) ⇒ (i).** Assume (iii), fix $x \in X$, and let $\varepsilon > 0$. Set

$$O := (f(x) - \varepsilon,\ f(x) + \varepsilon).$$

Then $O$ is open in $\mathbb{R}$, so $f^{-1}(O)$ is open in $X$ by (iii). Since $x \in f^{-1}(O)$, there exists $\delta > 0$ such that

$$B_\delta(x) \subseteq f^{-1}(O).$$

It follows that $d(x, y) < \delta$ implies $f(y) \in O$, or equivalently,

$$|f(y) - f(x)| < \varepsilon.$$

Thus $f$ is continuous at $x$. Since $x$ was arbitrary, $f$ is continuous on $X$.

**(iii) ⇒ (iv).** If $F \subseteq \mathbb{R}$ is closed, then $\mathbb{R} \setminus F$ is open. By (iii), $f^{-1}(\mathbb{R} \setminus F)$ is open in $X$. Since

$$f^{-1}(\mathbb{R} \setminus F) = X \setminus f^{-1}(F),$$

the set $f^{-1}(F)$ is closed in $X$.

**(iv) ⇒ (iii).** If $O \subseteq \mathbb{R}$ is open, then $\mathbb{R} \setminus O$ is closed. By (iv), the set $f^{-1}(\mathbb{R} \setminus O)$ is closed in $X$. But

$$f^{-1}(\mathbb{R} \setminus O) = X \setminus f^{-1}(O),$$

so $f^{-1}(O)$ is open. □

*Remark* 2.42. More generally, if $(X, d)$ and $(Y, \rho)$ are metric spaces, then a map $f\colon X \to Y$ is continuous at $x \in X$ if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$d(x, y) < \delta \quad \Longrightarrow \quad \rho\big(f(x), f(y)\big) < \varepsilon.$$

This is the usual $\varepsilon$–$\delta$ formulation of continuity in metric spaces.

#### 2.1.2.1 Uniform continuity

**Definition 2.43.** A function $f\colon A \subseteq (X,d) \to (Y,\rho)$ is said to be *uniformly continuous* on $A$ if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$d(x,y) < \delta \quad \Longrightarrow \quad \rho(f(x), f(y)) < \varepsilon \qquad \text{for all } x, y \in A.$$

The essential point is that $\delta$ depends only on $\varepsilon$, not on the location of the points.

**Example 2.44.** For a fixed $x_0 \in X$, the function $f(x) = d(x, x_0)$ is uniformly continuous on $X$, because

$$|f(x) - f(y)| = |d(x,x_0) - d(y,x_0)| \le d(x,y).$$

**Example 2.45.** If $A \subseteq X$, define

$$d(x,A) := \inf\{d(x,a) : a \in A\}.$$

Then the map $x \mapsto d(x,A)$ is uniformly continuous, since

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

**Example 2.46.** The function $f(x) = 1/x$ is continuous on $(0,1)$ but not uniformly continuous there. Indeed, if

$$x_n = \frac{1}{n}, \qquad y_n = \frac{1}{n+1},$$

then $|x_n - y_n| \to 0$, but

$$|f(x_n) - f(y_n)| = 1$$

for every $n$.

**Theorem 2.47** (Sequential criterion for uniform continuity)**.** *Let $f\colon A \subseteq (X,d) \to \mathbb{R}$. Then $f$ is uniformly continuous on $A$ if and only if for every pair of sequences $(x_n)$ and $(y_n)$ in $A$ satisfying $d(x_n, y_n) \to 0$, one has*

$$|f(x_n) - f(y_n)| \to 0.$$

*Proof.* Assume first that $f$ is uniformly continuous. Given $\varepsilon > 0$, choose $\delta > 0$ such that

$$d(x,y) < \delta \quad \Longrightarrow \quad |f(x) - f(y)| < \varepsilon$$

for all $x, y \in A$. If $d(x_n, y_n) \to 0$, then $d(x_n, y_n) < \delta$ for all sufficiently large $n$, and hence $|f(x_n) - f(y_n)| < \varepsilon$ eventually. Therefore $|f(x_n) - f(y_n)| \to 0$.

Conversely, suppose the sequential criterion holds, but $f$ is not uniformly continuous. Then there exists $\varepsilon_0 > 0$ such that for every $n \in \mathbb{N}$ one can choose $x_n, y_n \in A$ with

$$d(x_n, y_n) < \frac{1}{n} \qquad \text{and} \qquad |f(x_n) - f(y_n)| \geq \varepsilon_0.$$

Thus $d(x_n, y_n) \to 0$, while $|f(x_n) - f(y_n)|$ does not tend to 0, a contradiction. □

**Example 2.48.** A uniformly continuous function sends Cauchy sequences to Cauchy sequences. Indeed, if $(x_n)$ is Cauchy in $X$ and $f$ is uniformly continuous, then for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$d(x, y) < \delta \quad \Longrightarrow \quad |f(x) - f(y)| < \varepsilon.$$

Since $(x_n)$ is Cauchy, we have $d(x_n, x_m) < \delta$ for all sufficiently large $m, n$, and therefore

$$|f(x_n) - f(x_m)| < \varepsilon$$

for all sufficiently large $m, n$. Hence $(f(x_n))$ is Cauchy in $\mathbb{R}$.

**Theorem 2.49** (Heine–Cantor theorem on an interval)**.** *Let $f\colon [a, b] \to \mathbb{R}$ be continuous. Then $f$ is uniformly continuous on $[a, b]$.*

*Proof.* If $f$ were not uniformly continuous, then there would exist $\varepsilon_0 > 0$ and sequences $(x_n)$, $(y_n)$ in $[a, b]$ such that

$$|x_n - y_n| < \frac{1}{n} \qquad \text{but} \qquad |f(x_n) - f(y_n)| \geq \varepsilon_0.$$

By compactness of $[a, b]$, a subsequence $(x_{n_k})$ converges to some $x \in [a, b]$. Since $|x_{n_k} - y_{n_k}| \to 0$, we also have $y_{n_k} \to x$. By continuity,

$$f(x_{n_k}) \to f(x) \qquad \text{and} \qquad f(y_{n_k}) \to f(x),$$

so $|f(x_{n_k}) - f(y_{n_k})| \to 0$, contradicting $|f(x_{n_k}) - f(y_{n_k})| \geq \varepsilon_0$. □

**Theorem 2.50** (Uniform continuity on compact sets)**.** *Let $K$ be compact and let $f : K \to Y$ be continuous between metric spaces. Then $f$ is uniformly continuous on $K$.*

**Proof strategy.** Negate uniform continuity to produce two nearby sequences whose images stay uniformly separated. Compactness lets us extract a convergent subsequence; continuity then forces the two image sequences to collide, giving the contradiction.

*Proof.* Assume $f$ is not uniformly continuous. Then there exists $\varepsilon_0 > 0$ and sequences $x_n, y_n \in K$ such that

$$d(x_n, y_n) \to 0 \quad \text{and} \quad \rho\big(f(x_n), f(y_n)\big) \geq \varepsilon_0 \text{ for all } n.$$

By compactness, there exists a subsequence $x_{n_k} \to x \in K$. Since $d(x_{n_k}, y_{n_k}) \to 0$, it follows that $y_{n_k} \to x$. By continuity, $f(x_{n_k}) \to f(x)$ and $f(y_{n_k}) \to f(x)$, contradicting $\rho(f(x_{n_k}), f(y_{n_k})) \geq \varepsilon_0$. Hence $f$ is uniformly continuous. □

**Proposition 2.51** (Gluing uniformly continuous functions at an endpoint)**.** *Let $a < b < c$, and let $f\colon (a,c) \to \mathbb{R}$ be a function whose restrictions to $(a,b]$ and to $[b,c)$ are uniformly continuous. Then $f$ is uniformly continuous on $(a,c)$.*

*Proof.* Fix $\varepsilon > 0$. By uniform continuity on $(a,b]$ and on $[b,c)$, choose $\delta_1, \delta_2 > 0$ such that

$$x, y \in (a,b],\ |x-y| < \delta_1 \quad \Longrightarrow \quad |f(x) - f(y)| < \varepsilon,$$

and

$$x, y \in [b,c),\ |x-y| < \delta_2 \quad \Longrightarrow \quad |f(x) - f(y)| < \varepsilon.$$

Apply these two estimates with one point equal to $b$. After decreasing $\delta_1$ and $\delta_2$ if necessary, we also have

$$x \in (a,b],\ |x-b| < \delta_1 \quad \Longrightarrow \quad |f(x) - f(b)| < \frac{\varepsilon}{2},$$

and

$$y \in [b,c),\ |y-b| < \delta_2 \quad \Longrightarrow \quad |f(y) - f(b)| < \frac{\varepsilon}{2}.$$

Set $\delta = \min\{\delta_1, \delta_2\}$. If $x, y$ lie on the same side of $b$, the corresponding uniform-continuity estimate gives $|f(x) - f(y)| < \varepsilon$. If $x < b < y$ and $|x-y| < \delta$, then $|x-b| < \delta$ and $|y-b| < \delta$, and therefore

$$|f(x) - f(y)| \leq |f(x) - f(b)| + |f(b) - f(y)| < \varepsilon.$$

Thus $f$ is uniformly continuous on $(a,c)$. □

*Remark* 2.52. The endpoint compatibility is essential. Uniform continuity on $(a,b]$ and on $(b,c)$ alone does not suffice, since the two one-sided behaviours at $b$ may be incompatible.

**Proposition 2.53** (Continuous functions vanishing at infinity)**.** *Let $f\colon \mathbb{R} \to \mathbb{R}$ be continuous and suppose that*

$$\lim_{|x| \to \infty} f(x) = 0.$$

*Then $f$ is uniformly continuous on $\mathbb{R}$.*

*Proof.* Fix $\varepsilon > 0$. Choose $R > 0$ such that

$$|x| > R \quad \Longrightarrow \quad |f(x)| < \frac{\varepsilon}{3}.$$

Since $f$ is continuous on the compact interval $[-R-1, R+1]$, it is uniformly continuous there. Hence there exists $\delta_0 > 0$ such that

$$x, y \in [-R-1, R+1], \quad |x-y| < \delta_0 \quad \Longrightarrow \quad |f(x) - f(y)| < \varepsilon.$$

Let $\delta = \min\{1, \delta_0\}$. If $|x - y| < \delta$, then either both $x$ and $y$ belong to $[-R-1, R+1]$, in which case the compact-interval estimate applies, or else both lie outside $[-R, R]$ on the same tail. In the latter case,

$$|f(x) - f(y)| \le |f(x)| + |f(y)| < \frac{2\varepsilon}{3} < \varepsilon.$$

Thus $f$ is uniformly continuous on $\mathbb{R}$. $\square$

Consequently, every function in $C_0(\mathbb{R})$ is uniformly continuous. By contrast, boundedness alone does not imply uniform continuity on $\mathbb{R}$.

**Takeaway.** Compactness of the domain is the decisive hypothesis. Boundedness of the range, by itself, places almost no local restriction on the oscillation of the function.

**Example 2.54.** $f(x) = \sin x^2$, which is continuous and bounded but not uniformly continuous on $\mathbb{R}$. (Hint: Take $x^2 = n\pi$ and $y^2 = n\pi + \frac{1}{2}\pi$.)

**Example 2.55.** Let $f\colon \mathbb{R} \to \mathbb{R}$ be differentiable and suppose that $f'$ is bounded. Then $f$ is uniformly continuous on $\mathbb{R}$. For any $x, y \in \mathbb{R}$, by the Mean Value Theorem,

$$|f(x) - f(y)| = |f'(t)(x - y)| \le M|x - y|,$$

where $t$ is between $x$ and $y$, and $M$ is an upper bound for $|f'(t)|$. However, $f(x) = \sqrt{x}$ for $x \in (0, \infty)$ is uniformly continuous, although its derivative $f'(x) = \frac{1}{2\sqrt{x}}$ is not bounded.

**Proposition 2.56** (Bounded monotone continuous functions)**.** *Let $f\colon \mathbb{R} \to \mathbb{R}$ be bounded, continuous, and monotone. Then $f$ is uniformly continuous on $\mathbb{R}$.*

*Proof.* It suffices to consider the case in which $f$ is increasing; the decreasing case follows by applying the argument to $-f$. Since $f$ is bounded and increasing, the finite limits

$$L_- = \lim_{x \to -\infty} f(x), \qquad L_+ = \lim_{x \to +\infty} f(x)$$

exist. Fix $\varepsilon > 0$. Choose $A < B$ so that

$$x \le A \Longrightarrow |f(x) - L_-| < \frac{\varepsilon}{3}, \qquad x \ge B \Longrightarrow |f(x) - L_+| < \frac{\varepsilon}{3}.$$

By continuity on the compact interval $[A-1, B+1]$, the function $f$ is uniformly continuous there. Hence there exists $\delta_0 > 0$ such that

$$x, y \in [A-1, B+1], \quad |x - y| < \delta_0 \quad \Longrightarrow \quad |f(x) - f(y)| < \varepsilon.$$

Let $\delta = \min\{1, \delta_0\}$. If $|x - y| < \delta$, then either both points lie in $[A-1, B+1]$, or both lie in the same tail $(-\infty, A]$ or $[B, \infty)$. On the left tail,

$$|f(x) - f(y)| \le |f(x) - L_-| + |f(y) - L_-| < \frac{2\varepsilon}{3},$$

and the right tail is handled similarly using $L_+$. Thus $|f(x) - f(y)| < \varepsilon$ whenever $|x - y| < \delta$, proving uniform continuity. □

We observe that a uniformly continuous function can be extended uniformly to the closure of the set.

**Theorem 2.57** (Uniformly continuous extension to the closure)**.** *Let $(X, d)$ be a metric space, let $A \subset X$, and let $(Y, \rho)$ be a* complete *metric space. If $f\colon A \to Y$ is uniformly continuous, then there exists a unique uniformly continuous map $\widetilde{f}\colon \overline{A} \to Y$ such that $\widetilde{f}|_A = f$.*

*Proof.* Fix $x \in \overline{A}$. Choose a sequence $(x_n) \subset A$ with $x_n \to x$. Then $(x_n)$ is Cauchy in $X$, hence $(f(x_n))$ is Cauchy in $Y$ because $f$ is uniformly continuous. Since $Y$ is complete, the limit $\lim_{n\to\infty} f(x_n)$ exists. Define

$$\widetilde{f}(x) := \lim_{n\to\infty} f(x_n).$$

*Well-definedness.* If $(y_n) \subset A$ is another sequence with $y_n \to x$, then $d(x_n, y_n) \to 0$. By uniform continuity, $\rho(f(x_n), f(y_n)) \to 0$. Since both $(f(x_n))$ and $(f(y_n))$ converge, their limits must coincide, so $\widetilde{f}$ is well defined. *Extension property.* If $x \in A$, choose the constant sequence $x_n = x$. Then $\widetilde{f}(x) = f(x)$. *Uniform continuity of $\widetilde{f}$.* Let $\varepsilon > 0$. By uniform continuity of $f$, there exists $\delta > 0$ such that

$$d(u, v) < \delta \implies \rho\big(f(u), f(v)\big) < \varepsilon/3 \qquad (u, v \in A).$$

Set $\delta' := \delta/3$. Take $x, y \in \overline{A}$ with $d(x, y) < \delta'$. Choose sequences $(x_n), (y_n) \subset A$ such that $x_n \to x$ and $y_n \to y$. Pick $N$ so large that

$$d(x_N, x) < \delta', \qquad d(y_N, y) < \delta', \qquad \rho\big(\widetilde{f}(x), f(x_N)\big) < \varepsilon/3, \qquad \rho\big(\widetilde{f}(y), f(y_N)\big) < \varepsilon/3.$$

Then

$$d(x_N, y_N) \le d(x_N, x) + d(x, y) + d(y, y_N) < \delta' + \delta' + \delta' = \delta,$$

hence $\rho(f(x_N), f(y_N)) < \varepsilon/3$. Therefore,

$$\rho\big(\widetilde{f}(x), \widetilde{f}(y)\big) \le \rho\big(\widetilde{f}(x), f(x_N)\big) + \rho\big(f(x_N), f(y_N)\big) + \rho\big(f(y_N), \widetilde{f}(y)\big) < \varepsilon.$$

This proves that $\widetilde{f}$ is uniformly continuous on $\overline{A}$. *Uniqueness.* If $g\colon \overline{A} \to Y$ is another uniformly continuous extension of $f$, then $g$ is continuous. For $x \in \overline{A}$ and any sequence $(x_n) \subset A$ with $x_n \to x$, continuity gives

$$g(x) = \lim_{n\to\infty} g(x_n) = \lim_{n\to\infty} f(x_n) = \widetilde{f}(x),$$

so $g = \widetilde{f}$. □

**Exercise 2.58** (Why completeness is needed)**.** Show that completeness of the target space cannot be omitted in Theorem 2.57. More precisely, consider $A = \mathbb{Q} \subset \mathbb{R}$ with the usual metric and the function $f\colon \mathbb{Q} \to \mathbb{Q}$ given by $f(q) = q$. Prove that $f$ is uniformly continuous on $\mathbb{Q}$, but there is no continuous (hence no uniformly continuous) extension $\widetilde{f}\colon \mathbb{R} \to \mathbb{Q}$.

Next, we show that a uniformly continuous function has at most linear growth.

**Theorem 2.59** (Linear growth of uniformly continuous functions on $\mathbb{R}$)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be uniformly continuous. Then there exist constants $A, B \geq 0$ such that $|f(x)| \leq A|x| + B$ for all $x \in \mathbb{R}$.*

*Proof.* For $\varepsilon = 1$, there exists $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < 1$. We estimate $f$ separately on a bounded interval and outside it. Let $a > 0$. Since $f$ is continuous, it is bounded on the compact interval $[-a, a]$; thus $|f(x)| \leq A$ for some constant $A < \infty$ and all $x \in [-a, a]$. Now consider the restriction of $f$ to $[a, \infty)$. For $x \in [a, \infty)$, choose $n \in \mathbb{N}$ such that $x \in [a + n\delta,\, a + (n+1)\delta]$. Then

$$f(x) - f(a) = f(x) - f(a + n\delta) + f(a + n\delta) - f(a)$$

$$= f(x) - f(a + n\delta) + \sum_{j=1}^{n} [f(a + j\delta) - f(a + (j+1)\delta)]$$

$$\Rightarrow |f(x)| < 1 + n + |f(a)|$$

$$\Rightarrow \left|\frac{f(x)}{x}\right| < \frac{(n+1) + |f(a)|}{a + n\delta} < \frac{(n+1) + |f(a)|}{n\delta} < \left(1 + \frac{1}{n}\right)\frac{1}{\delta} + \frac{|f(a)|}{n\delta} \leq B < \infty.$$

Notice that $B$ is independent of $n$, hence $B$ is independent of $x$. That is, $|f(x)| \leq B|x|$ if $x > a$. Therefore, $|f(x)| \leq B|x| + A \quad$ for all $x \in \mathbb{R}$. □

**Example 2.60.** Notice that $f(x) = x^2$ is not uniformly continuous on $\mathbb{R}$, as it does not satisfy the conclusion of the above theorem. On the other hand, for $x_n = n$ and $y_n = n + \frac{1}{n}$, we have $|f(x_n) - f(y_n)| = 2$.

## 2.2 Normed Linear Spaces and Classical Inequalities

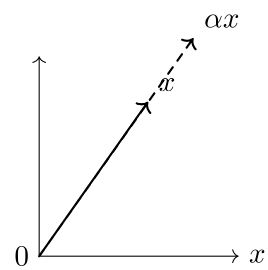
**Section overview.**

- Norms quantify size and distance in linear settings, while classical inequalities provide the estimates that make the theory effective.
- This section supplies much of the analytic toolkit used later in sequence spaces and function spaces.

### 2.2.1 Normed Linear Space

A *normed linear space* is a vector space equipped with a notion of length that is compatible with its linear structure.

Let $(X, +, \cdot)$ be a vector space over the field $F$ (that is, $F = \mathbb{R}$ or $F = \mathbb{C}$). Our aim is to endow $X$ with a topology that reflects the algebraic operations naturally. In other words, we seek a notion of distance on $X$ that interacts well with vector addition and scalar multiplication.




The linear structure of $X$ is encoded by the two basic operations

$$\begin{aligned} &\text{(i)} \quad (x, y) \mapsto x + y \qquad (X \times X \to X), \\ &\text{(ii)} \quad (\alpha, x) \mapsto \alpha x \qquad (F \times X \to X). \end{aligned}$$

A topology on $X$ is compatible with the vector-space structure when these two maps are continuous. A vector space endowed with such a topology is called a *topological vector space.*

If $J$ denotes the topology on $X$ and $U$ the usual topology on $F$, then the relevant product topologies are $J \times J$ on $X \times X$ and $U \times J$ on $F \times X$. Open sets in these product spaces are unions of basic rectangles of the form $O_1 \times O_2$.

In normed spaces, the topology comes from a function that measures the size of a vector. Motivated by linearity and homogeneity, we seek a notion of distance from the origin satisfying the following properties:

$$\begin{aligned} (i) &\quad \operatorname{dist}(0, \alpha x) = |\alpha| \operatorname{dist}(0, x) \\ (ii) &\quad \operatorname{dist}(0, x + y) \leq \operatorname{dist}(0, x) + \operatorname{dist}(0, y) \\ (iii) &\quad \text{when } \alpha = 0, \ \operatorname{dist}(0, 0) = 0 \end{aligned}$$

Let $p := \operatorname{dist} : X \to [0, \infty)$ be defined by $p(x) = \operatorname{dist}(0, x)$. Then

$$\begin{aligned} (i) &\quad p(x) = 0 \quad \text{for } x = 0 \\ (ii) &\quad p(\alpha x) = |\alpha| \, p(x) \quad \text{(absolute homogeneity)} \\ (iii) &\quad p(x + y) \leq p(x) + p(y) \quad \text{(triangle inequality)} \end{aligned}$$

Such a function $p$ is called a *seminorm.* It satisfies the basic algebraic properties of a norm, except that nonzero vectors may still have seminorm 0.

**Example 2.61.**

$$p : \mathbb{R}^2 \to [0, \infty), \quad p(x_1, x_2) = |x_1|.$$

Then $p$ is a seminorm and $p(0, 1) = 0$.

Thus every point on the $y$-axis has seminorm 0. This shows precisely why a seminorm need not define an honest metric on the underlying vector space.

Let $\|\cdot\| : X \to [0, \infty)$ be a map such that

(i) $\|x\| \geq 0$ for each $x \in X$, and $\|x\| = 0$ if and only if $x = 0$.

(ii) $\|\alpha x\| = |\alpha|\, \|x\|$ for each $(\alpha, x) \in F \times X$ (absolute homogeneity).

(iii) $\|x + y\| \leq \|x\| + \|y\|$ for each $x, y \in X$ (triangle inequality).

The map $\|\cdot\|$ is called a *norm* on $X$.

Every norm induces a metric on $X$ by

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

This metric in turn determines a topology on $X$. For $r > 0$ and $x \in X$, the corresponding open ball is

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

Open sets are then defined in the usual metric sense.

Not every metric on a vector space arises from a norm. For example, the discrete metric on a vector space is not induced by any norm, since it fails the absolute homogeneity property.

For $x, y \in X$, define

$$d_0(x, y) = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases}$$

If we write $\|x\| = d(0, x)$, then for $\alpha \in \mathbb{F}$, $\|\alpha x\| \neq |\alpha|\|x\|$ $(x \neq 0)$ unless $|\alpha| = 1$.

However, if $d$ is a metric on a linear space $X$ such that $d(x, y) = d(x - y, 0)$ and $d(\alpha x, \alpha y) = |\alpha| d(x, y)$, then $d(x, 0) = \|x\|$ defines a norm on $X$.

1. $\|x\| = 0 \iff d(x, 0) = 0 \iff x = 0$.
2. $\|\alpha x\| = d(\alpha x, 0) = |\alpha| d(x, 0) = |\alpha|\|x\|$.
3. $\|x + y\| = d(x + y, 0) = d(x, -y) \leq d(x, 0) + d(-y, 0) = \|x\| + \|y\|$.

A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be convex if

$$f(t_1 x_1 + \cdots + t_n x_n) \leq t_1 f(x_1) + \cdots + t_n f(x_n)$$

where $0 \leq t_i \leq 1$ and $x_i \in \mathbb{R}^n$.

**Example 2.62.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function satisfying $f(\alpha x) = \alpha f(x)$ for all $\alpha \in \mathbb{R}$, for all $x \in \mathbb{R}^n$. Prove that

(i) $f(x + y) \leq f(x) + f(y)$.

(ii) $f(0) = 0$.

(iii) $f(-x) \geq -f(x)$.

(iv) $f(t_1 x_1 + \cdots + t_n x_n) \leq t_1 f(x_1) + \cdots + t_n f(x_n)$.

Further, what requires to make $f$ a norm on $\mathbb{R}^n$?

### 2.2.2 Convergence of Sequence in Metric Space

A sequence $(x_n)$ in a metric space $(X, d)$ is said to be converging to $x \in X$, if for any $\epsilon > 0$, $\exists N_0 \in \mathbb{N}$ such that $n \geq N_0 \implies d(x_n, x) < \epsilon$.

**Example 2.63.** Let $X = (0, \infty)$ and $d(x, y) = \left|\frac{1}{x} - \frac{1}{y}\right|$. Then $x_n = n$ does not converge to any point of $X$. However, this sequence is not so bad as $x_n = n \to 0$, which is not in $X$. Such sequences can be classified as Cauchy sequences.

### 2.2.3 Cauchy Sequences

**Definition 2.64.** A sequence $(x_n)$ in $(X, d)$ is said to be a *Cauchy sequence* if for any $\epsilon > 0$, there exist $N_0 \in \mathbb{N}$ such that $\forall m, n \geq N_0$, $d(x_n, x_m) < \epsilon$.

**Example 2.65.** Show that every Cauchy sequence in a metric space is bounded.

*Proof.* Let $(x_n)$ be a Cauchy sequence in $(X, d)$. Choose $n_0 \in \mathbb{N}$ such that

$$d(x_n, x_m) < 1, \qquad m, n \geq n_0.$$

In particular, $d(x_n, x_{n_0}) < 1$ for every $n \geq n_0$, so the tail of the sequence lies in $B_1(x_{n_0})$. Set

$$r := \max\{1, d(x_{n_0}, x_1), \ldots, d(x_{n_0}, x_{n_0 - 1})\}.$$

Then $d(x_n, x_{n_0}) \leq r$ for every $n \in \mathbb{N}$, and hence $\{x_n : n \in \mathbb{N}\} \subseteq B_r(x_{n_0})$. Thus every Cauchy sequence is bounded. $\square$

### 2.2.4 Young's Inequality

**Proposition 2.66** (Young's inequality)**.** *Let $1 < p < \infty$, and let $q$ be the conjugate exponent defined by $\frac{1}{p} + \frac{1}{q} = 1$. Then for every $a, b \geq 0$,*

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}. \tag{2.2}$$

*Equality holds if and only if $a^p = b^q$.*

*Proof.* Consider the increasing function $\phi(x) = x^{p-1}$ on $[0, \infty)$. Its inverse is $\phi^{-1}(y) = y^{q-1}$, since $(p-1)(q-1) = 1$. The rectangle of area $ab$ is contained in the union of the regions under the graph of $\phi$ on $[0, a]$ and under the graph of $\phi^{-1}$ on $[0, b]$; hence

$$ab \leq \int_0^a x^{p-1}\, dx + \int_0^b y^{q-1}\, dy = \frac{a^p}{p} + \frac{b^q}{q}.$$

Equality holds precisely when the two regions meet on the boundary, that is, when $b = \phi(a) = a^{p-1}$, or equivalently when $a^p = b^q$. $\square$

**Example 2.67.** Let $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$. Write

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

Then $(\mathbb{R}^n, \|\cdot\|_1)$ is a normed linear space (n.l.s.). If

$$\|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}$$

then by Cauchy-Schwarz inequality, $(\mathbb{R}^n, \|\cdot\|_2)$ is a n.l.s. For

$$\|x\|_\infty = \sup_i |x_i|$$

$(\mathbb{R}^n, \|\cdot\|_\infty)$ is a normed linear space.
For $1 \le p < \infty$, write

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

Then $\ell_n^p := (\mathbb{R}^n, \|\cdot\|_p)$ will be a normed linear space.

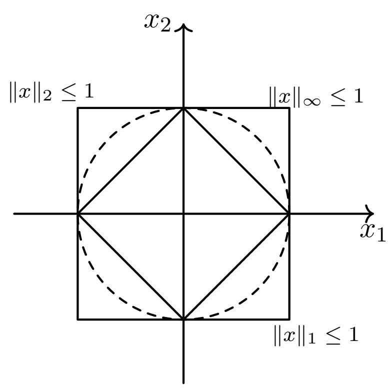



**Figure 2.1.** Unit balls in $\mathbb{R}^2$ for the norms $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$.

### 2.2.5 Space of Sequences

Let $1 \le p < \infty$ and let $\ell^p$ denote the space of all sequences that satisfy

$$\sum_{i=1}^{\infty} |x_i|^p < \infty; \qquad x = (x_1, x_2, \ldots, x_n, \ldots)$$

Then $(\ell^p, \|\cdot\|_p)$ or simply $\ell^p$, will be a normed linear space.

If $p = \infty$,

$$\|x\|_\infty = \sup_{1 \le i < \infty} |x_i| < \infty,$$

then $(\ell^\infty, \|\cdot\|_\infty)$ is a normed linear space (follows from definition of supremum).
For $1 \le p < \infty$, showing $\ell^p$ is a normed linear space required the following inequalities.

### 2.2.6 Hölder's Inequality

**Proposition 2.68** (Hölder's inequality)**.** *Let $1 \le p \le \infty$, and let $q$ be the conjugate exponent, so that $\frac{1}{p} + \frac{1}{q} = 1$. If $x = (x_j) \in \ell^p$ and $y = (y_j) \in \ell^q$, then $x \cdot y = (x_j y_j) \in \ell^1$ and*

$$\|x \cdot y\|_1 \le \|x\|_p \|y\|_q.$$

*Proof.* If $p = 1$ and $q = \infty$, then

$$\sum_{j=1}^{\infty} |x_j y_j| \le \Big(\sup_{j\ge 1} |y_j|\Big) \sum_{j=1}^{\infty} |x_j| = \|x\|_1 \|y\|_\infty.$$

Assume now that $1 < p < \infty$, so $1 < q < \infty$. If $x = 0$ or $y = 0$, there is nothing to prove. For each $j$, apply Young's inequality to

$$a_j = \frac{|x_j|}{\|x\|_p}, \qquad b_j = \frac{|y_j|}{\|y\|_q}.$$

Then

$$\frac{|x_j y_j|}{\|x\|_p \|y\|_q} \le \frac{|x_j|^p}{p\|x\|_p^p} + \frac{|y_j|^q}{q\|y\|_q^q}.$$

Summing from $j = 1$ to $n$ gives

$$\sum_{j=1}^{n} \frac{|x_j y_j|}{\|x\|_p \|y\|_q} \le \frac{1}{p} \sum_{j=1}^{n} \frac{|x_j|^p}{\|x\|_p^p} + \frac{1}{q} \sum_{j=1}^{n} \frac{|y_j|^q}{\|y\|_q^q} \le \frac{1}{p} + \frac{1}{q} = 1.$$

Hence

$$\sum_{j=1}^{n} |x_j y_j| \le \|x\|_p \|y\|_q \qquad (n \in \mathbb{N}).$$

Passing to the limit in $n$ yields

$$\|x \cdot y\|_1 \le \|x\|_p \|y\|_q.$$

□

### 2.2.7 Minkowski's Inequality

Let $1 \le p \le \infty$. Then, for $x, y \in \ell^p$, we have $x + y \in \ell^p$ and

$$\|x + y\|_p \le \|x\|_p + \|y\|_p.$$

*Proof.* For $p = 1$ or $p = \infty$, the inequality is immediate. Assume $1 < p < \infty$. Then

$$\begin{aligned} \|x+y\|_p &= \left(\sum_{i=1}^{\infty} |x_i + y_i|^p\right)^{1/p} \\ &\leq \left(\sum_{i=1}^{\infty} (|x_i| + |y_i|)^p\right)^{1/p} \end{aligned} \tag{2.3}$$

Since

$$(|x_i| + |y_i|)^p = (|x_i| + |y_i|)(|x_i| + |y_i|)^{p-1}$$

By Hölder's inequality,

$$\sum (|x_i| + |y_i|)^{p-1}|x_i| \leq \left(\sum (|x_i| + |y_i|)^{(p-1)q}\right)^{1/q} \left(\sum |x_i|^p\right)^{1/p}$$

Thus,

$$\sum (|x_i| + |y_i|)^p \leq \left(\sum (|x_i| + |y_i|)^p\right)^{1/q} (\|x\|_p + \|y\|_p)$$

That is

$$\left(\sum (|x_i| + |y_i|)^p\right)^{1-\frac{1}{q}} \leq \|x\|_p + \|y\|_p$$

From (2.3), we get

$$\|x+y\|_p \leq \left(\sum (|x_i| + |y_i|)^p\right)^{1/p} \leq \|x\|_p + \|y\|_p$$

$\square$

*Remark* 2.69. Equality in $\|x+y\|_p \leq \|x\|_p + \|y\|_p$ holds if and only if $x = \frac{\|x\|_p}{\|y\|_p} y$.
(*Hint.* Consider $\|x\|_p = 1 = \|y\|_p$ etc.)

**Example 2.70.** Recall that every convergent sequence is bounded. It follows that the space $c$ of all convergent sequences is a normed linear space under the norm

$$\|x\| = \sup |x_i| < \infty;$$

where $x = (x_1, x_2, \ldots, x_n, \ldots)$.

Further, the space $c_0$ of all sequences converging to zero is also a normed linear space. That is, $x = (x_1, x_2, \ldots, x_n, \ldots)$,

$$\lim_{n \to \infty} |x_n| = 0.$$

Thus, $(c_0, \|\cdot\|_\infty)$ is a linear subspace of $(c, \|\cdot\|_\infty)$.

**Exercise 2.71.** Show that the following strict inclusions hold:

$$\ell^1 \subsetneq \ell^2 \subsetneq c_0 \subsetneq c \subsetneq \ell^\infty$$

(*Hint.* $x = (x_n) \in \ell^1$, then $\lim x_n = 0 \implies x \in \ell^\infty$, $\sum |x_n|^2 \leq \sum \|x\|_\infty |x_n| \implies \|x\|_2^2 \leq \|x\|_\infty \|x\|_1$.)

**Exercise 2.72.** For $x = (x_1, ..., x_n) \in \mathbb{R}^n$ (or $\mathbb{C}^n$), show that:

$$\|x\|_\infty \le \|x\|_1 \le \sqrt{n}\|x\|_2 \le n\|x\|_\infty$$

### 2.2.8 Geometry of spheres in $(\mathbb{R}^n, \|\cdot\|_p)$

For $0 < p \le \infty$ and $x \in \mathbb{R}^n$, define

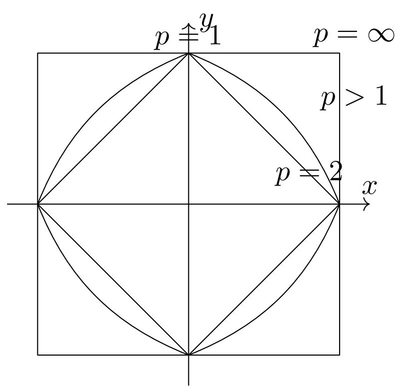
$$\|x\|_p = \left(\sum |x_i|^p\right)^{1/p}.$$

Then $\|\cdot\|_p$ is a norm for $1 \le p < \infty$. For $0 < p < 1$, the quantity $\|x\|_p^p = d_p(0, x)$, where

$$d_p(x, y) = \|x - y\|_p^p,$$

defines a metric. We return to this point later.

Let $S_1^p(0) = \{x : d_p(0, x) = 1\}$. The following figure illustrates the corresponding unit spheres for several values of $p$ ($0 < p < \infty$ and $p = \infty$).




Shapes for $0 < p < 1$ would look like star-shaped curves (not shown).

### 2.2.9 Closed sets in $(X, d)$

**Definition 2.73.** A subset $F \subseteq X$ is said to be *closed* if its complement $X \setminus F$ is open.

**Theorem 2.74** (Characterizations of closed sets)**.** *Let $(X, d)$ be a metric space and let $F \subseteq X$. The following are equivalent:*

*(i) $F$ is closed.*

*(ii) Whenever $x \in X$ satisfies $B_\epsilon(x) \cap F \ne \varnothing$ for every $\epsilon > 0$, one has $x \in F$.*

*(iii) For every sequence $(x_n)$ in $F$, if $x_n \to x$ in $X$, then $x \in F$.*

*Proof. (i) $\Rightarrow$ (ii).* Assume that $F$ is closed and let $x \in X$ satisfy $B_\epsilon(x) \cap F \ne \varnothing$ for every $\epsilon > 0$. If $x \notin F$, then $x \in X \setminus F$, and since $X \setminus F$ is open there exists $\epsilon_0 > 0$ such that

$$B_{\epsilon_0}(x) \subseteq X \setminus F.$$

This contradicts $B_{\epsilon_0}(x) \cap F \neq \varnothing$. Hence $x \in F$.

*(ii) ⇒ (iii).* Let $(x_n) \subseteq F$ and suppose that $x_n \to x$. Fix $\epsilon > 0$. Since $x_n \to x$, there exists $N$ such that $x_n \in B_\epsilon(x)$ for all $n \geq N$. In particular, $B_\epsilon(x) \cap F \neq \varnothing$. By (ii), it follows that $x \in F$.

*(iii) ⇒ (i).* Let $x \in X \setminus F$. If every ball centered at $x$ met $F$, then for each $n \in \mathbb{N}$ we could choose

$$x_n \in B_{1/n}(x) \cap F.$$

Then $x_n \to x$, and (iii) would imply $x \in F$, a contradiction. Therefore there exists $\epsilon_0 > 0$ such that $B_{\epsilon_0}(x) \cap F = \varnothing$, equivalently

$$B_{\epsilon_0}(x) \subseteq X \setminus F.$$

This shows that $X \setminus F$ is open, so $F$ is closed. □

**Proposition 2.75** (Sequential characterization of continuity)**.** *Let $f : (X, d) \to \mathbb{R}$ and let $x \in X$. Then $f$ is continuous at $x$ if and only if for every sequence $(x_n)$ in $X$ with $x_n \to x$, one has $f(x_n) \to f(x)$.*

*Proof.* Assume first that $f$ is continuous at $x$, and let $(x_n)$ be a sequence in $X$ such that $x_n \to x$. Given $\epsilon > 0$, continuity at $x$ yields $\delta > 0$ such that

$$d(y, x) < \delta \implies |f(y) - f(x)| < \epsilon.$$

Since $x_n \to x$, there exists $N \in \mathbb{N}$ such that $d(x_n, x) < \delta$ for all $n \geq N$. Hence $|f(x_n) - f(x)| < \epsilon$ for all $n \geq N$, so $f(x_n) \to f(x)$.

Conversely, suppose that every sequence $(x_n)$ with $x_n \to x$ satisfies $f(x_n) \to f(x)$. If $f$ were not continuous at $x$, then there would exist $\epsilon_0 > 0$ such that for every $\delta > 0$ one could find $y \in X$ with

$$d(x, y) < \delta \qquad \text{and} \qquad |f(y) - f(x)| \geq \epsilon_0.$$

Choosing $y_n$ with $d(x, y_n) < 1/n$ and $|f(y_n) - f(x)| \geq \epsilon_0$, we obtain a sequence $y_n \to x$ for which $f(y_n) \not\to f(x)$, a contradiction. □

**Exercise 2.76.** If $f : (X, d) \to \mathbb{R}$ is continuous and $f(x_0) \neq 0$ for some $x_0 \in X$, then $\exists \delta > 0$ such that

$$f(x) \neq 0 \ \forall x \in B_\delta(x).$$

(*Hint.* take $\epsilon_0 = \frac{1}{2}|f(x_0)| > 0$, $\exists \delta > 0$ etc.)

**Example 2.77.** Show that if $f : (X, d) \to \mathbb{R}$ is continuous, then $A = \{x : f(x) > 0\}$ is open, without using the fact that its complement is closed.
(*Hint.* Let $x \in A$, then for $\epsilon = \frac{1}{2}f(x) > 0$, $\exists \delta > 0$ such that $d(x, y) < \delta \implies |f(x) - f(y)| < \epsilon$)

### 2.2.10 Interior in $(X, d)$

Let $A \subset X$. Then interior$(A)$, Int$(A)$, or $A^\circ$ denotes the largest open set contained in $A$. Thus

$$A^\circ = \bigcup \{O \subset X : O \text{ is open and } O \subseteq A\}$$

and equivalently

$$A^\circ = \bigcup \{B_\epsilon(x) : x \in A,\ \epsilon > 0,\ B_\epsilon(x) \subset A\},$$

that is, the union of all open balls contained in $A$.

### 2.2.11 Closure in $(X, d)$

The closure of a set $A \subset X$ is the smallest closed set containing $A$. Thus

$$\overline{A} = \bigcap \{F \subset X : F \text{ is closed and } A \subset F\},$$

and, in metric spaces,

$$\overline{A} = \{x \in X : \exists x_n \in A \text{ with } x_n \to x\}.$$

In other words, $\overline{A}$ is the set of limits of convergent sequences drawn from $A$.

**Example 2.78.** Let

$$A = \left\{(n, \frac{1}{n}) : n \in \mathbb{N}\right\}.$$

Then the closure of $A$ in $(\mathbb{R}^2, u)$ is $\overline{A} = A$, and $A^\circ = \varnothing$.

**Proposition** (Closure via open balls)**.** *Let $A \subseteq X$ and let $x \in X$. Then*

$$x \in \overline{A} \quad \iff \quad B_\epsilon(x) \cap A \neq \varnothing \quad \textit{for every } \epsilon > 0.$$

*Proof.* Assume first that $x \in \overline{A}$. If there were $\epsilon_0 > 0$ such that $B_{\epsilon_0}(x) \cap A = \varnothing$, then $A \subseteq X \setminus B_{\epsilon_0}(x)$. Since $X \setminus B_{\epsilon_0}(x)$ is closed and contains $A$, the minimality of $\overline{A}$ would imply

$$\overline{A} \subseteq X \setminus B_{\epsilon_0}(x),$$

which is impossible because $x \in \overline{A}$ but $x \in B_{\epsilon_0}(x)$.

Conversely, assume that every open ball centered at $x$ meets $A$. If $x \notin \overline{A}$, then $x \in X \setminus \overline{A}$, and since $X \setminus \overline{A}$ is open there exists $\epsilon_0 > 0$ such that

$$B_{\epsilon_0}(x) \subseteq X \setminus \overline{A}.$$

Because $A \subseteq \overline{A}$, this gives $B_{\epsilon_0}(x) \cap A = \varnothing$, a contradiction. Hence $x \in \overline{A}$. □

**Proposition** (Sequential characterization of closure)**.** *Let $A \subseteq X$ and let $x \in X$. Then*

$$x \in \overline{A} \quad \iff \quad \textit{there exists a sequence } (x_n) \subseteq A \textit{ with } x_n \to x.$$

*Proof.* Assume that $x \in \overline{A}$. By the previous proposition, $B_{1/n}(x) \cap A \neq \varnothing$ for every $n \in \mathbb{N}$. Choose $x_n \in B_{1/n}(x) \cap A$. Then

$$d(x_n, x) < \frac{1}{n}, \qquad n \in \mathbb{N},$$

so $x_n \to x$.

Conversely, suppose that $(x_n) \subseteq A$ and $x_n \to x$. Let $\epsilon > 0$ be arbitrary. Choose $n_0$ such that $d(x_n, x) < \epsilon$ for all $n \geq n_0$. Then $x_{n_0} \in B_\epsilon(x) \cap A$, so every open ball centered at $x$ meets $A$. The previous proposition now gives $x \in \overline{A}$. □

**Definition 2.79.** A set $A \subset (X, d)$ is said to be *dense* in $X$ if $\overline{A} = X$.

### 2.2.12 Space of Finite Sequences

The space of finite sequences plays a role analogous to that of the space of polynomials:

$$P(x) = a_0 + a_1 x + \ldots + a_n x^n$$

corresponds to

$$(a_0, a_1, \ldots, a_n) \sim (a_0, a_1, \ldots, a_n, 0, 0, 0, \ldots).$$

Let

$$c_{00} = \{x = (x_1, x_2, \ldots, x_n, 0, 0, \ldots) : x_i \in F\}.$$

Then every $x \in c_{00}$ is bounded, and

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i| < \infty$$

defines a norm on $c_{00}$.

As we shall see later, the space of finite sequences $c_{00}$ is dense in $\ell^p$ for every $1 \leq p < \infty$. By contrast, its closure in $\ell^\infty$ is $c_0$, a closed proper subspace of $\ell^\infty$. Indeed, let

$$x_n = \left(1, \frac{1}{2}, \ldots, \frac{1}{n}, 0, 0, \ldots\right) \in c_{00},$$

and

$$x = \left(1, \frac{1}{2}, \ldots, \frac{1}{n}, \frac{1}{n+1}, \ldots\right).$$

Then

$$\|x - x_n\|_\infty = \sup_{k \geq n} \frac{1}{k+1} = \frac{1}{n+1} \to 0,$$

but $x \notin c_{00}$. Hence $c_{00}$ is not closed in $\ell^\infty$. Moreover, $c_{00}$ is not open in $\ell^\infty$. Indeed, let $\epsilon > 0$. Then

$$(\epsilon/2, \epsilon/2, \ldots) \in B_\epsilon(0) \subset \ell^\infty,$$

but $(\epsilon/2, \epsilon/2, \ldots) \notin c_{00}$. Therefore $B_\epsilon(0) \not\subseteq c_{00}$ for any $\epsilon > 0$.

For $1 \le p < \infty$, we have $c_{00} \subsetneq \ell^p$, and $c_{00}$ is neither closed nor open in $\ell^p$. To see this, let

$$x_n = \left(\frac{\epsilon^p}{2^{n+1}}\right)^{1/p},\ 1 \le p < \infty,$$

and consider $x = (x_1, x_2, ...)$. Then $x \in B_\epsilon(0) \subset \ell^p$, but $x \notin c_{00}$. Now write $x_n = (x_1, ..., x_n, 0, ...) \in c_{00}$. Then

$$\|x - x_n\|_p^p = \sum_{k=n+1}^{\infty} \frac{\epsilon^p}{2^{k+1}} \to 0,$$

but $x \notin c_{00}$.

**Example 2.80.** Let $M$ be a non-open subspace of a normed linear space (n.l.s.) $X$. Show that $M = X$.
(*Hint.* $O \in M \implies B_\epsilon(0) \subset M \subset X$. Since $M$ is linear, $\alpha B_\epsilon(0) \subset M \subset X$ for all $\epsilon > 0$ $\implies B_{\epsilon_1}(0) \subset M$ for some $\epsilon_1 > 0$. If $y \in X$, then $y \in B_{\epsilon_1}(0) \subset M \subset X$ for some $\epsilon_1 > 0$.)

Notice that for $x = (x_1, x_2, ..., x_n, ...) \in \ell^p; 1 \le p < \infty$, $x_n = (x_1, ..., x_n, 0, ...) \in c_{00}$. And

$$\|x - x_n\|_p^p = \sum_{k=n+1}^{\infty} |x_k|^p \longrightarrow 0,$$

because $x \in \ell^p$. Hence $x_n \to x$ in $\ell^p$, and therefore $\overline{c_{00}} = \ell^p$.
However, $c_{00}$ is not dense in $\ell^\infty$, but $\overline{c_{00}} = c_0$. For this, let $X = (x_1, x_2, \dots) \in c_0$. Then $\lim_{n\to\infty} x_n = 0$. For $\epsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that $n \ge n_0$ implies $|x_n| < \epsilon/2$. Now, write $X_n = (x_1, \dots, x_n, 0, 0, \dots)$. Then $X_n \in c_{00}$ and for $n \ge n_0$,

$$\|X - X_n\|_\infty = \sup_{i \ge n+1} |x_i| < \epsilon/2.$$

Therefore, $X_n \to X$.

*Remark* 2.81. $\overline{c_{00}} = c_0 \subsetneq \ell^\infty$. That is, $c_{00}$ is not dense in $\ell^\infty$.

**Example 2.82.** Let $f : \mathbb{R} \to \mathbb{R}$ (or $\mathbb{C}$) be a continuous function. Suppose $\lim_{|x|\to\infty} f(x) = 0$. Then for $\epsilon > 0$, there exists $\delta > 0$ such that $|f(x)| < \epsilon$ for $|x| > \frac{1}{\delta}$.

Since $f$ is continuous, it follows that $f$ is bounded. Let $\|f\|_\infty = \sup_{x\in\mathbb{R}} |f(x)| < \infty$. Then

$$C_0 = \left\{ f : \mathbb{R} \to \mathbb{R} \text{ is continuous, } \lim_{|x|\to\infty} |f(x)| = 0 \right\}$$

is a normed linear space.
For any function $f : \mathbb{R} \to \mathbb{R}$, define

$$\operatorname{supp}(f) = \overline{\{x \in \mathbb{R} : f(x) \neq 0\}}$$

called the support of $f$.

Let

$$C_C = \{f\colon \mathbb{R} \to \mathbb{R} \text{ continuous} \quad \text{and } \operatorname{supp}(f) \text{ is compact}\}.$$

Then $f \in C_c$ is a bounded function and

$$\|f\|_\infty = \sup_{x\in\mathbb{R}} |f(x)| = \sup_{x\in\operatorname{supp}(f)} |f(x)| < \infty.$$

Let $K = \operatorname{supp}(f)$ be compact. Then $(C_C, \|\cdot\|_\infty)$ is a dense subspace of $(C_0, \|\cdot\|_\infty)$.
For this, let $f \in C_0$, then for $\epsilon > 0$, there exists $\delta > 0$ such that $|f(x)| < \epsilon$ for $|x| > \frac{1}{\delta}$. Write $K = \{x : |x| \leq \frac{1}{\delta}\}$.
Let $O$ be a bounded open set with $K \subset O$. Define

$$g(x) = \frac{d(x, O^c)}{d(x, O^c) + d(x, K)}$$

Then $g$ is continuous on $\mathbb{R}$, $0 \leq g(x) \leq 1$ and $g(x) = 1$ for $x \in K$ and $g(O^c) = \{0\}$.
Let $h = f \cdot g$. Then $h \in C_C$ and

$$\|f - h\|_\infty = \|f(1-g)\|_\infty = \sup_{x\in\mathbb{R}} |f(x)|(1 - g(x)) \leq \epsilon.$$

Hence, $C_C$ is dense in $C_0$.
Note that $d(x, A) = \inf_{y\in A} |x - y|$.

### 2.2.13 Complete Metric Spaces

We have seen that there are Cauchy sequences whose limits need not necessarily belong to the space.
For example, the sequence $\frac{1}{n} \in ((0,1), u)$ under the usual metric, is a Cauchy sequence but the limit $\frac{1}{n} \to 0 \notin (0,1)$.

It is always possible to enlarge the space so that limits of all Cauchy sequences can be accommodated. This process is known as the completion of metric spaces, we shall see later. However, there are many spaces which do accommodate limits of their Cauchy sequences.

**Definition 2.83.** A metric space $(X, d)$ is called complete if every Cauchy sequence in $X$ has its limit in $X$.

**Example 2.84.** $(\mathbb{R}, u)$ is a complete space.

Let $(x_n)$ be a Cauchy sequence in $\mathbb{R}$. Then it is bounded. And by the Bolzano–Weierstrass theorem, there exists a subsequence $x_{n_k} \to x \in \mathbb{R}$. For any $\epsilon > 0$, there exists a natural number $k_0$ such that

$$|x_{n_k} - x| < \epsilon \quad \text{for all} \quad k \geq k_0 \tag{2.4}$$

But the sequence $(x_n)$ is Cauchy, so for all $\epsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that $|x_n - x_m| < \epsilon$ for all $n, m \geq n_0$. Let $m \geq n_0$ and $m \geq n_{k_0}$. Then

$$|x_n - x_{n_k}| < \epsilon \quad \text{for any } n \geq n_0 \text{ and } k \geq k_0. \tag{2.5}$$

From (2.4) and (2.5), it follows that:

$$|x_n - x| \leq |x_n - x_{n_k}| + |x_{n_k} - x| < 2\epsilon$$

for $n \geq n_0$ and $n_k \geq n_{k_0}$. Thus, for $\epsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that

$$n \geq n_0 \implies |x_n - x| < \epsilon.$$

Notice that the above discussion can be used to prove the following result.

**Proposition** (A Cauchy sequence with a convergent subsequence converges)**.** *Let $(x_n)$ be a Cauchy sequence in a metric space $(X, d)$. If some subsequence $(x_{n_k})$ converges to $x \in X$, then the whole sequence $(x_n)$ converges to $x$.*

*Proof.* Let $\epsilon > 0$. Since $(x_n)$ is Cauchy, there exists $N_1 \in \mathbb{N}$ such that

$$d(x_n, x_m) < \frac{\epsilon}{2} \qquad \text{whenever } n, m \geq N_1.$$

Because $x_{n_k} \to x$, there exists $K \in \mathbb{N}$ such that

$$d(x_{n_k}, x) < \frac{\epsilon}{2} \qquad \text{whenever } k \geq K.$$

Choose $k$ so large that $n_k \geq N_1$ and $k \geq K$, and set $N \coloneqq n_k$. Then for every $n \geq N$ we have

$$d(x_n, x) \leq d(x_n, x_N) + d(x_N, x) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Hence $x_n \to x$. □

**Example 2.85.** $(\mathbb{R}^n, \|\cdot\|_p)$ is complete for every $1 \leq p \leq \infty$.

First assume $1 \leq p < \infty$, and let $(x^k)$ be a Cauchy sequence in $(\mathbb{R}^n, \|\cdot\|_p)$, where

$$x^k = (x_1^k, \ldots, x_n^k).$$

For each fixed $j \in \{1, \ldots, n\}$ we have

$$|x_j^k - x_j^\ell| \leq \|x^k - x^\ell\|_p,$$

so $(x_j^k)_k$ is a Cauchy sequence in $\mathbb{R}$. Since $\mathbb{R}$ is complete, there exists $x_j \in \mathbb{R}$ such that $x_j^k \to x_j$ as $k \to \infty$. Define $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$.

To show that $x^k \to x$ in $\|\cdot\|_p$, fix $\epsilon > 0$. Since $(x^k)$ is Cauchy, there exists $K \in \mathbb{N}$ such that

$$\|x^k - x^\ell\|_p < \epsilon \qquad \text{whenever } k, \ell \geq K.$$

Now fix $k \geq K$ and let $\ell \to \infty$. For each coordinate, $x_j^\ell \to x_j$, so by continuity of the map

$$(t_1, \ldots, t_n) \longmapsto \Big(\sum_{j=1}^n |t_j|^p\Big)^{1/p}$$

on $\mathbb{R}^n$ we obtain

$$\|x^k - x\|_p = \lim_{\ell\to\infty} \|x^k - x^\ell\|_p \leq \epsilon.$$

Thus $x^k \to x$, and $(\mathbb{R}^n, \|\cdot\|_p)$ is complete.

For $p = \infty$, the same coordinatewise argument gives limits $x_j = \lim_k x_j^k$, and for $k, \ell \geq K$ we have

$$\|x^k - x^\ell\|_\infty < \epsilon.$$

Letting $\ell \to \infty$ yields $\|x^k - x\|_\infty \leq \epsilon$ for every $k \geq K$. Hence $(\mathbb{R}^n, \|\cdot\|_\infty)$ is also complete.

**Proposition 2.86** (Completeness of the classical sequence spaces)**.** *For every $1 \leq p \leq \infty$, the space $(\ell^p, \|\cdot\|_p)$ is complete. Thus each $\ell^p$ is a Banach space.*

*Proof.* We first treat the case $1 \leq p < \infty$. Let $x^m = (x_1^m, x_2^m, \ldots)$ be a Cauchy sequence in $\ell^p$. For each fixed $j$, the scalar sequence $(x_j^m)_m$ is Cauchy because

$$|x_j^m - x_j^n| \leq \|x^m - x^n\|_p.$$

Since the scalar field is complete, there is a scalar $x_j$ such that $x_j^m \to x_j$. Put $x = (x_1, x_2, \ldots)$.

Fix $\varepsilon > 0$. Choose $M$ such that

$$\|x^m - x^n\|_p < \varepsilon \qquad (m, n \geq M).$$

For every $N \in \mathbb{N}$ and $m, n \geq M$,

$$\sum_{j=1}^N |x_j^m - x_j^n|^p \leq \|x^m - x^n\|_p^p < \varepsilon^p.$$

Letting $n \to \infty$ gives

$$\sum_{j=1}^N |x_j^m - x_j|^p \leq \varepsilon^p \qquad (m \geq M,\ N \in \mathbb{N}).$$

Now let $N \to \infty$. Since the partial sums are increasing,

$$\|x^m - x\|_p \leq \varepsilon \qquad (m \geq M).$$

In particular $x = x^M + (x - x^M) \in \ell^p$, and $x^m \to x$ in $\ell^p$.

For $p = \infty$, the same coordinatewise argument gives scalars $x_j = \lim_m x_j^m$. If $m, n \geq M$, then $|x_j^m - x_j^n| \leq \varepsilon$ for every $j$. Letting $n \to \infty$ and taking the supremum over $j$ yields

$$\|x^m - x\|_\infty \leq \varepsilon \qquad (m \geq M).$$

Moreover $x \in \ell^\infty$, since $x = x^M + (x - x^M)$. Thus $x^m \to x$ in $\ell^\infty$, completing the proof. □

**Proposition** (Closed subsets of complete spaces)**.** *Every closed subset of a complete metric space is complete.*

*Proof.* Let $F$ be a closed subset of a complete metric space $(X, d)$, and let $(x_n)$ be a Cauchy sequence in $F$. Then $(x_n)$ is also a Cauchy sequence in $X$. Since $X$ is complete, there exists $x \in X$ such that $x_n \to x$. Because $F$ is closed and each $x_n$ belongs to $F$, the limit point $x$ must belong to $F$. Hence every Cauchy sequence in $F$ converges in $F$, so $F$ is complete. □

**Corollary.** *If $(X, d)$ is complete and $F \subseteq X$, then $F$ is complete if and only if $F$ is closed in $X$.*

*Proof.* One implication is the proposition above. Conversely, if $F$ is complete and $(x_n) \subseteq F$ converges in $X$ to some $x \in X$, then $(x_n)$ is Cauchy in $F$. Completeness of $F$ yields a limit $y \in F$. Limits in metric spaces are unique, so $y = x$. Thus $x \in F$, and $F$ is closed. □

**Example 2.87** (The closed subspace $c_0$)**.** The space $c_0$ of scalar sequences converging to 0 is a proper closed subspace of $(\ell^\infty, \|\cdot\|_\infty)$.

It is clearly a linear subspace, and it is proper because the constant sequence $(1, 1, 1, \dots)$ belongs to $\ell^\infty$ but not to $c_0$. To prove closedness, let $x^m \in c_0$ and suppose $x^m \to x$ in $\ell^\infty$. Given $\varepsilon > 0$, choose $m_0$ such that

$$\|x^{m_0} - x\|_\infty < \frac{\varepsilon}{2}.$$

Since $x^{m_0} \in c_0$, there exists $J$ such that

$$|x_j^{m_0}| < \frac{\varepsilon}{2} \qquad (j \geq J).$$

Therefore, for $j \geq J$,

$$|x_j| \leq |x_j - x_j^{m_0}| + |x_j^{m_0}| < \varepsilon.$$

Hence $x_j \to 0$, so $x \in c_0$. Thus $c_0$ is closed in $\ell^\infty$; by Proposition 2.86, it is complete in the inherited norm.

**Example 2.88** (Completeness of $C[a, b]$ in the uniform norm)**.** The space $(C[a, b], \|\cdot\|_\infty)$ is complete.

Indeed, let $(f_n)$ be a Cauchy sequence in $(C[a, b], \|\cdot\|_\infty)$. For each $t \in [a, b]$, the scalar sequence $(f_n(t))$ is Cauchy; hence it converges. Define

$$f(t) := \lim_{n \to \infty} f_n(t).$$

We claim that $f_n \to f$ uniformly. Given $\varepsilon > 0$, choose $N$ such that

$$\|f_n - f_m\|_\infty < \varepsilon \qquad (m, n \geq N).$$

Letting $m \to \infty$ pointwise gives

$$|f_n(t) - f(t)| \leq \varepsilon \qquad (n \geq N,\ t \in [a, b]).$$

Thus $f_n \to f$ uniformly. Since a uniform limit of continuous functions is continuous, $f \in C[a, b]$. Hence $C[a, b]$ is complete in the uniform norm.

**Example 2.89** (Incompleteness of $C[0, 1]$ in the $L^1$-norm)**.** The normed space $(C[0, 1], \|\cdot\|_1)$, where

$$\|f\|_1 = \int_0^1 |f(t)|\, dt,$$

is not complete.

For $n \geq 3$, define $f_n \in C[0, 1]$ by

$$f_n(t) = \begin{cases} 0, & 0 \leq t \leq \frac{1}{2} - \frac{1}{n}, \\ \dfrac{n}{2}\left(t - \frac{1}{2} + \frac{1}{n}\right), & \frac{1}{2} - \frac{1}{n} < t < \frac{1}{2} + \frac{1}{n}, \\ 1, & \frac{1}{2} + \frac{1}{n} \leq t \leq 1. \end{cases}$$

Let

$$H(t) = \begin{cases} 0, & 0 \leq t \leq \frac{1}{2}, \\ 1, & \frac{1}{2} < t \leq 1. \end{cases}$$

Then $|f_n - H| \leq 1$ and the functions differ only on an interval of length $2/n$. Therefore

$$\|f_n - H\|_1 \leq \frac{2}{n} \to 0.$$

It follows that $(f_n)$ is Cauchy in the $L^1$-norm.

If $C[0, 1]$ were complete in $\|\cdot\|_1$, there would exist $g \in C[0, 1]$ such that $\|f_n - g\|_1 \to 0$. Since also $f_n \to H$ in $L^1$, uniqueness of $L^1$-limits gives $g = H$ almost everywhere. This is impossible: a continuous function that is 0 almost everywhere on $[0, \frac{1}{2}]$ must be identically 0 on that interval, while a continuous function that is 1 almost everywhere on $(\frac{1}{2}, 1]$ must be identically 1 there. Continuity at $1/2$ is then contradicted. Hence $(C[0, 1], \|\cdot\|_1)$ is not complete.

### 2.2.14 Sequences of Functions

Notice that in the previous exercises, we have seen that $(C([0, 1]), \|\cdot\|_\infty)$ is complete. That is, if $\|f_n - f_m\|_\infty \to 0$, then there exists $f \in C([0, 1])$ such that $\|f_n - f\|_\infty \to 0$. But then,

$$|f_n(t) - f(t)| < \|f_n - f\|_\infty \to 0, \quad \forall t \in [0, 1],$$

that is, $f_n(t) \to f(t)$ for each $t \in [0,1]$. We say that $f_n \to f$ uniformly if

$$\sup_t |f_n(t) - f(t)| \to 0.$$

But there are sequence of functions which converge pointwise but not uniformly.

**Example 2.90.** Let $f_n(t) = t^n, \ \ t \in [0,1]$. Then,

$$f(t) = \lim_{n\to\infty} f_n(t) = \begin{cases} 0 & 0 \le t < 1 \\ 1 & t = 1 \end{cases}$$

So,

$$\sup_t |f_n(t) - f(t)| = 1 \not\to 0.$$

**Example 2.91.** Let $f_n : \mathbb{R} \to \mathbb{R}$ be given by

$$f_n(t) = e^{-nt^2}, \quad n \in \mathbb{N}$$

Then,

$$f(t) = \lim_{n\to\infty} f_n(t) = \begin{cases} 1 & t = 0 \\ 0 & |t| > 0 \end{cases}$$

Notice that for $t = 0$, $|f_n(0) - f(0)| = |1 - 1| = 0 < \epsilon, \quad \forall n \in \mathbb{N}$ If $|t_0| > 0$, $t_0^2 > 0$. Then for $|f_n(t_0) - 0| < \epsilon$, we get

$$e^{-nt_0^2} < \epsilon \implies n > \frac{\log \frac{1}{\epsilon}}{t_0^2}$$

Let $n_0 = \left\lceil \frac{\log \frac{1}{\epsilon}}{t_0^2} \right\rceil + 1$. Then, $|f_n(t_0) - f(t_0)| < \epsilon$ for $n \ge n_0$

Notice that $n_0 = n_0(\epsilon, t_0)$ and $n_0$ is large for $|t_0|$ close to zero. Thus, $n_0$ cannot be free from $t_0$. Therefore, $f_n \to f$ pointwise but not uniformly. Also,

$$\|f_n - f\|_\infty = \sup_{t\in\mathbb{R}} e^{-nt^2} = 1 \not\to 0$$

If $f_n(t) = e^{-nt}$ for $t \in [1, \infty)$, then

$$\sup_t |f_n(t) - 0| = e^{-n} \to 0 \implies e^{-nt} \xrightarrow[{[1,\infty)}]{\text{unif.}} 0$$

**Exercise 2.92.** Let $f_n, f : A(\subseteq \mathbb{R}) \to \mathbb{R}$ be such that $f_n \to f$ uniformly on $A$. Then for $|f_n(t)| \le M_n$ (that is $f_n$'s are bounded), that implies $f$ is bounded.
(*Hint.* $|f(t)| \le |f_{n_0}(t) - f(t)| + |f_{n_0}(t)| < \epsilon + M_{n_0} < \infty \quad \forall t \in A$)

We shall see later that uniform convergence preserves many important analytic properties.

**Theorem 2.93** (Uniform limits preserve continuity)**.** *Let $A \subseteq \mathbb{R}$, and let $f_n, f : A \to \mathbb{R}$. If $f_n \to f$ uniformly on $A$ and each $f_n$ is continuous on $A$, then $f$ is continuous on $A$.*

*Proof.* Fix $t \in A$ and let $\epsilon > 0$. By uniform convergence, there exists $n_0 \in \mathbb{N}$ such that

$$\sup_{s\in A} |f_{n_0}(s) - f(s)| < \frac{\epsilon}{3}.$$

Since $f_{n_0}$ is continuous at $t$, there exists $\delta > 0$ such that

$$|s - t| < \delta \implies |f_{n_0}(s) - f_{n_0}(t)| < \frac{\epsilon}{3}.$$

Therefore, whenever $s \in A$ and $|s - t| < \delta$,

$$|f(s) - f(t)| \leq |f(s) - f_{n_0}(s)| + |f_{n_0}(s) - f_{n_0}(t)| + |f_{n_0}(t) - f(t)| < \epsilon.$$

Hence $f$ is continuous at $t$. Since $t \in A$ was arbitrary, $f$ is continuous on $A$. □

**Proposition** (Uniform limits and the Riemann integral)**.** *Let $f_n, f \in \mathcal{R}[a, b]$, and suppose that $f_n \to f$ uniformly on $[a, b]$. Then*

$$\int_a^b f_n(x)\,dx \longrightarrow \int_a^b f(x)\,dx.$$

*In particular,*

$$\lim_{n\to\infty} \int_a^b f_n(x)\,dx = \int_a^b \lim_{n\to\infty} f_n(x)\,dx.$$

*Proof.* For each $n$,

$$\left|\int_a^b (f_n - f)\right| \leq \int_a^b |f_n - f| \leq (b - a)\|f_n - f\|_\infty.$$

Since $\|f_n - f\|_\infty \to 0$, the right-hand side tends to 0. Therefore

$$\int_a^b f_n \to \int_a^b f.$$

□

**Corollary** (Term-by-term integration of uniformly convergent series)**.** *Let $f_n \in \mathcal{R}[a, b]$ for all $n$, and suppose that the partial sums*

$$S_N \coloneqq \sum_{n=1}^N f_n$$

*converge uniformly on $[a, b]$ to a function $S$. Then*

$$\int_a^b \sum_{n=1}^\infty f_n(x)\,dx = \sum_{n=1}^\infty \int_a^b f_n(x)\,dx.$$

*Proof.* Each $S_N$ is Riemann integrable, and the preceding proposition gives

$$\int_a^b S_N(x)\,dx \longrightarrow \int_a^b S(x)\,dx.$$

But

$$\int_a^b S_N(x)\,dx = \sum_{n=1}^{N} \int_a^b f_n(x)\,dx$$

for every $N$. Passing to the limit in $N$ yields the desired identity. $\square$

**Theorem 2.94** (Uniform convergence of derivatives)**.** *Let $f_n \in C^1[a,b]$ and suppose that $f_n' \to g$ uniformly on $[a,b]$. Assume also that $f_n(x_0)$ converges for some $x_0 \in [a,b]$. Then there exists $f \in C^1[a,b]$ such that $f_n \to f$ uniformly on $[a,b]$ and $f' = g$.*

*Proof.* Since each $f_n'$ is continuous and $f_n' \to g$ uniformly, the function $g$ is continuous on $[a,b]$. Define

$$f(x) \coloneqq \lim_{n\to\infty} f_n(x_0) + \int_{x_0}^{x} g(t)\,dt.$$

Then $f \in C^1[a,b]$ and $f' = g$.

For $m, n \in \mathbb{N}$ and $x \in [a,b]$, the Fundamental Theorem of Calculus gives

$$f_n(x) - f_m(x) = f_n(x_0) - f_m(x_0) + \int_{x_0}^{x} (f_n'(t) - f_m'(t))\,dt.$$

Hence

$$|f_n(x) - f_m(x)| \le |f_n(x_0) - f_m(x_0)| + (b-a)\|f_n' - f_m'\|_\infty.$$

Taking the supremum over $x \in [a,b]$, we obtain

$$\|f_n - f_m\|_\infty \le |f_n(x_0) - f_m(x_0)| + (b-a)\|f_n' - f_m'\|_\infty.$$

Because $f_n(x_0)$ converges and $f_n'$ converges uniformly, the right-hand side tends to 0 as $m, n \to \infty$. Thus $(f_n)$ is uniformly Cauchy on $[a,b]$, so $f_n \to h$ uniformly for some continuous function $h$.

It remains to identify $h$ with the function $f$ defined above. For every $x \in [a,b]$,

$$f_n(x) - f_n(x_0) = \int_{x_0}^{x} f_n'(t)\,dt.$$

Passing to the limit as $n \to \infty$, using the convergence of $f_n(x_0)$ and the uniform convergence $f_n' \to g$, yields

$$h(x) - h(x_0) = \int_{x_0}^{x} g(t)\,dt.$$

Since $h(x_0) = \lim_{n\to\infty} f_n(x_0) = f(x_0)$, it follows that $h(x) = f(x)$ for every $x \in [a,b]$. Therefore $f_n \to f$ uniformly and $f' = g$. $\square$

*Remark* 2.95. Convergence of $(f_n(x_0))$ is necessary in the above result. Consider

$$f_n(t) = \sqrt{t+n}, \qquad t \in [0,1]$$

Then $f_n$ does not converge at any point of $[0,1]$, but

$$f_n'(t) = \frac{1}{2\sqrt{t+n}} \xrightarrow{\text{unif.}} 0$$

Since

$$\sup_{t\in[0,1]} |f_n'(t) - 0| = \sup_{t\in[0,1]} \frac{1}{2\sqrt{t+n}} = \frac{1}{2\sqrt{n}} \to 0.$$

**Exercise 2.96.** Let $f_n : \mathbb{R} \to \mathbb{R}$. Check for uniform convergence of $f_n$ to some $f$:

1. $f_n(t) = \frac{\sin(nt)}{\sqrt{n}}$.
2. $f_n(t) = n^2 t(1-t^2)^n$.
3. $f_n(t) = te^{-nt}$.

Also, verify for term-by-term integration and differentiation for each of the above.

**Theorem 2.97** (Interchange of uniform and pointwise limits)**.** *Let $E \subseteq \mathbb{R}$, and suppose $f_n \to f$ uniformly on $E$. Let $x$ be a limit point of $E$, and assume that*

$$\lim_{t\to x} f_n(t) = A_n \quad \textit{(finite)} \tag{2.6}$$

*Then $(A_n)$ is convergent and*

$$\lim_{t\to x} f(t) = \lim_{n\to\infty} A_n.$$

*That is,*

$$\lim_{t\to x}\lim_{n\to\infty} f_n(t) = \lim_{n\to\infty}\lim_{t\to x} f_n(t)$$

*Proof.* Since $f_n \to f$ uniformly on $E$, for each $\epsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that

$$|f_n(t) - f_m(t)| < \epsilon, \quad \forall n, m \geq n_0,\ \forall t \in E \tag{2.7}$$

By (2.7), it implies that $|A_n - A_m| < \epsilon, \quad \forall n, m \geq n_0$. So $(A_n)$ is Cauchy, hence convergent $\implies A_n \to A$ (Say). Now,

$$\begin{aligned} |f(t) - A| &= |f(t) - f_n(t) + f_n(t) - A_n + A_n - A| \\ &\leq |f(t) - f_n(t)| + |f_n(t) - A_n| + |A_n - A| \\ &< \epsilon + \epsilon + \epsilon \end{aligned}$$

for $t \in (x-\delta, x+\delta) \setminus \{x\}$ and $n \geq n_0$, with $\delta$ independent of $t$.

$$\lim_{t\to x} f(t) = A = \lim_{n\to\infty} A_n$$

$$\text{Thus,} \quad \lim_{t\to x}\lim_{n\to\infty} f_n(t) = \lim_{n\to\infty}\lim_{t\to x} f_n(t)$$

□

**Theorem 2.98** (Uniform convergence from uniformly convergent derivatives)**.** *Let $f_n : [a,b] \to \mathbb{R}$ be such that $(f_n')$ converges uniformly. If there exists $x_0 \in [a,b]$ such that $(f_n(x_0))$ is convergent, then $(f_n)$ is uniformly convergent, and*

$$\lim_{n\to\infty} f_n'(x) = \Big(\lim_{n\to\infty} f_n(x)\Big)'$$

*(that is limit and derivative commute).*

*Proof.* As in the previous theorem, the Mean Value Theorem gives

$$|f_n(x) - f_m(x)| \le (b-a)\|f_n' - f_m'\|_\infty + |f_n(x_0) - f_m(x_0)|$$

for all $x \in [a,b]$. Since $(f_n')$ converges uniformly and $(f_n(x_0))$ converges, the sequence $(f_n)$ is uniformly Cauchy, hence converges uniformly on $[a,b]$ to some function $f$.

Fix $x \in [a,b]$. For $t \neq x$, define

$$\varphi_n(t) = \frac{f_n(x) - f_n(t)}{x-t}, \qquad \varphi(t) = \frac{f(x) - f(t)}{x-t}.$$

Because $f_n \to f$ uniformly, we have $\varphi_n(t) \to \varphi(t)$ for each $t \neq x$. Moreover, by the Mean Value Theorem,

$$|\varphi_n(t) - \varphi_m(t)| = \left|\frac{(f_n - f_m)(x) - (f_n - f_m)(t)}{x-t}\right| = |(f_n - f_m)'(\xi_{n,m,t})| \le \|f_n' - f_m'\|_\infty$$

for some $\xi_{n,m,t}$ between $x$ and $t$. Hence $(\varphi_n)$ is uniformly Cauchy on $[a,b] \setminus \{x\}$, so $\varphi_n \to \varphi$ uniformly there.

For each $n$, the limit of $\varphi_n(t)$ as $t \to x$ exists and equals $f_n'(x)$. The preceding theorem on interchanging limits therefore yields

$$\lim_{n\to\infty} f_n'(x) = \lim_{n\to\infty}\lim_{t\to x}\varphi_n(t) = \lim_{t\to x}\lim_{n\to\infty}\varphi_n(t) = \lim_{t\to x}\varphi(t) = f'(x).$$

Since $x$ was arbitrary, $f$ is differentiable on $[a,b]$ and

$$\lim_{n\to\infty} f_n'(x) = \Big(\lim_{n\to\infty} f_n(x)\Big)' \qquad (x \in [a,b]).$$

□

### 2.2.15 Term-by-term differentiation

**Corollary 2.99** (Term-by-term differentiation of series)**.** *Let $S_n = f_1 + \cdots + f_n$, where each $f_i : [a,b] \to \mathbb{R}$ is differentiable. Assume that $S_n' \to S$ uniformly on $[a,b]$ and that $S_n(x_0) \to L$ for some $x_0 \in [a,b]$. Then $S_n$ converges uniformly on $[a,b]$ to a differentiable function $F$, and*

$$F' = S.$$

*Equivalently,*

$$\left(\sum_{n=1}^{\infty} f_n\right)' = \sum_{n=1}^{\infty} f_n'$$

*on $[a,b]$.*

*Proof.* Apply Theorem 2.94 to the sequence $(S_n)$. □

This leads naturally to the corresponding relation between differentiation and integration. If $f$ is continuous on $[a,b]$, then the Fundamental Theorem of Calculus gives

$$\frac{d}{dx}\int_a^x f(t)\,dt = f(x). \tag{2.8}$$

On the other hand, if $f$ is differentiable and $f'$ is Riemann integrable, then

$$\int_a^x f'(t)\,dt = f(x) - f(a).$$

The equality in (2.8) concerns the derivative of an integral with variable upper limit; it should not be confused with the identity above for the integral of $f'$.

### 2.2.16 Uniform continuity

**Definition 2.100.** A function $f : A(\subset (X,d)) \to \mathbb{R}$ is said to be *uniformly continuous* on $A$ if for each $\epsilon > 0$, there exists $\delta > 0$ such that for all $x, y \in A$,

$$d(x,y) < \delta \implies |f(x) - f(y)| < \epsilon$$

Notice that $\delta$ is free of choice of locations of points $x, y \in A$; it only depends on their separation.

**Example 2.101.** For $x_0 \in X$, let $f(x) = d(x, x_0)$. Then $f$ is uniformly continuous on $X$. (Hint: $d(x,x_0) \le d(x,y) + d(y,x_0) \implies f(x) - f(y) < d(x,y)$.) Similarly, by replacing $x$ with $y$, it follows.

**Example 2.102.** For $x \in X$, $A \subset X$, define $d(x,A) = \inf\{d(x,a) : a \in A\}$, which is called the *distance of $A$ from $x$*, and is uniformly continuous as a function of $x$. (Hint: $d(x,a) \le d(x,y) + d(y,a)$.) Thus, $d(x,A) \le d(x,y) + d(y,A)$ and so,

$$|f(x) - f(y)| \le d(x,y) \quad (\because x \leftrightarrow y)$$

**Example 2.103.** The function $f : (0,1) \to \mathbb{R}$ given by $f(x) = \frac{1}{x}$ is continuous on $(0,1)$, but not uniformly continuous.

### 2.2.17 Pointwise continuity of $f$

Let $x_0 \in (0,1)$. Then for $\epsilon > 0$, there exists $n \in \mathbb{N}$ such that $(x_0 - \frac{\epsilon}{n},\, x_0 + \frac{\epsilon}{n}) \subset (0,1)$. Suppose $\left|\frac{1}{x_0} - \frac{1}{y}\right| < \epsilon$ for $y \in (x_0 - \frac{\epsilon}{n},\, x_0 + \frac{\epsilon}{n}) =: I_{x_0}$. Then $|x_0 - y| < \epsilon x_0 y$. Let $\delta = \min_{y \in I_{x_0}}\{\epsilon x_0 y\} = \epsilon x_0(x_0 - \epsilon/n) > 0$. If $|x_0 - y| < \delta$. Then

$$\left|\frac{1}{x_0} - \frac{1}{y}\right| = \frac{|x_0 - y|}{x_0 y} < \frac{\delta}{x_0 y} \leq \frac{\epsilon x_0(x_0 - \epsilon/n)}{x_0 y} < \epsilon$$

Hence, $f$ is continuous at each $x_0 \in (0,1)$.
*$f$ is not uniformly continuous:* Let $\epsilon = \frac{1}{2}$, $x = \frac{1}{n}$, $y = \frac{1}{n+1}$, $n \in \mathbb{N}$. Then for any $\delta > 0$, there exists $n_0 \in \mathbb{N}$ such that

$$|x - y| = \left|\frac{1}{n} - \frac{1}{n+1}\right| < \delta$$

but

$$|f(x) - f(y)| = 1 \not< \frac{1}{2}.$$

Hence, $f$ is not uniformly continuous on $(0,1)$. From the above argument, we can prove the following result.

**Theorem 2.104** (Sequential characterization of uniform continuity)**.** *Let $f : A \subseteq X \to \mathbb{R}$, where $(X,d)$ is a metric space. Then $f$ is uniformly continuous on $A$ if and only if for every pair of sequences $(x_n)$ and $(y_n)$ in $A$ satisfying $d(x_n, y_n) \to 0$, one has $|f(x_n) - f(y_n)| \to 0$.*

*Proof.* Suppose first that $f$ is uniformly continuous on $A$. Given $\epsilon > 0$, choose $\delta > 0$ such that

$$d(x,y) < \delta \implies |f(x) - f(y)| < \epsilon \qquad (x, y \in A). \tag{2.9}$$

If $d(x_n, y_n) \to 0$, then there exists $N$ such that $d(x_n, y_n) < \delta$ for all $n \geq N$. Hence $|f(x_n) - f(y_n)| < \epsilon$ for all $n \geq N$, and therefore $|f(x_n) - f(y_n)| \to 0$.

Conversely, suppose that the sequential condition holds, but that $f$ is not uniformly continuous. Then there exists $\epsilon_0 > 0$ such that for every $\delta > 0$ there are $x, y \in A$ with

$$d(x,y) < \delta \qquad \text{and} \qquad |f(x) - f(y)| \geq \epsilon_0.$$

For each $n \in \mathbb{N}$, choose $x_n, y_n \in A$ such that

$$d(x_n, y_n) < \frac{1}{n} \qquad \text{and} \qquad |f(x_n) - f(y_n)| \geq \epsilon_0.$$

Then $d(x_n, y_n) \to 0$, but $|f(x_n) - f(y_n)| \not\to 0$, contradicting the assumed sequential condition. Hence $f$ is uniformly continuous. $\square$

**Exercise 2.105.** Show that a uniformly continuous function on a metric space $(X, d)$ sends Cauchy sequences to Cauchy sequences. (*Hint.* If $f : (X, d) \to \mathbb{R}$ is uniformly continuous, so for $d(x_n, x_m) \to 0 \implies |f(x_n) - f(x_m)| \to 0$.)

**Result:** Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then $f$ is uniformly continuous.

*Proof.* Suppose, on the contrary, that $f$ is not uniformly continuous on $[a, b]$. Then there exists $\epsilon_0 > 0$ such that for every $\delta > 0$, there exist $x, y \in [a, b]$ with $|x - y| < \delta$ but $|f(x) - f(y)| \geq \epsilon_0$. For $\delta = \frac{1}{n}$, there exist $x_n, y_n \in [a, b]$ such that $|x_n - y_n| < \frac{1}{n}$ but $|f(x_n) - f(y_n)| \geq \epsilon_0$. By the Bolzano–Weierstrass theorem, $x_n, y_n$ have convergent subsequences, say $x_{n_k} \to x$ and $y_{n_k} \to y$. Now,

$$|x - y| = \lim_{k\to\infty} |x_{n_k} - y_{n_k}| \leq \lim_{k\to\infty} \frac{1}{n_k} = 0,$$

so $x = y$. Since $f$ is continuous, $f(x_{n_k}) - f(y_{n_k}) \to f(x) - f(y) = 0$, but $|f(x_{n_k}) - f(y_{n_k})| \geq \epsilon_0$, contradiction. □

**Example 2.106.** Let $f : \mathbb{R} \to \mathbb{R}$ be continuous such that $\lim_{|x|\to\infty} f(x) = 0$. Then $f$ is uniformly continuous.

*Proof.* For $\epsilon > 0$, there exists $[-a, a]$ such that $|f(x)| < \epsilon/2$ if $x \in [-a, a]^c$. Hence, if $x, y \in [-a, a]^c$, then

$$|f(x) - f(y)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \qquad (1)$$

Since $f$ is uniformly continuous on $[-a, a]$. For $\epsilon > 0$, there exists $\delta > 0$ such that

$$|x - y| < \delta \implies |f(x) - f(y)| < \epsilon \qquad (2)$$

Since (2.9) holds true for $x, y$ with $|x - y| < \delta$. It follows that for $\epsilon > 0$, we get $\delta > 0$ such that $|x - y| < \delta \implies |f(x) - f(y)| < \epsilon$ (for any $x, y \in \mathbb{R}$). Hence, $f$ is uniformly continuous on $\mathbb{R}$. □

Notice that if $f \in C_0(\mathbb{R})$, that is $f$ is continuous and $\lim_{|x|\to\infty} f(x) = 0$ and hence $f$ is uniformly continuous. But if $f$ is continuous and bounded, then $f$ need not be uniformly continuous on $\mathbb{R}$.

**Example 2.107.** $f(x) = \sin x^2$, which is continuous and bounded but not uniformly continuous on $\mathbb{R}$.

**Example 2.108.** Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded continuous function. If $f$ is monotone, then $f$ is uniformly continuous on $\mathbb{R}$. Since $f$ is bounded, let $\inf f(x) = L, \quad \sup f(x) = M$. For $\epsilon > 0$, there exist $x_0, y_0 \in \mathbb{R}$ such that $f(x_0) < L + \epsilon$ and $f(y_0) > M - \epsilon$.
If $f$ is monotone increasing, then for $x, y \in [x_0, y_0]^c$ and $x, y \geq y_0$

$$f(y) - f(x) \leq M - f(y_0) < M - (M - \varepsilon) = \varepsilon.$$

Similarly, if $x, y \leq x_0$ then

$$f(y) - f(x) \leq L + \varepsilon - f(x_0) < L + \varepsilon - L = \varepsilon.$$

Thus, for $x, y \in [x_0, y_0]^c$, we get $|f(x) - f(y)| < \varepsilon \qquad (1).$
Since $f$ is continuous on $[x_0, y_0]$, $f$ is uniformly continuous on $[x_0, y_0]$. For any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$x, y \in [x_0, y_0],\ |x - y| < \delta \implies |f(x) - f(y)| < \varepsilon \qquad (2)$$

Observe that (1) also holds for $x, y \in [x_0, y_0]^c$ whenever $|x - y| < \delta$. Combining this with (2), we obtain a single $\delta > 0$ such that

$$|x - y| < \delta \implies |f(x) - f(y)| < \varepsilon$$

for all $x, y \in \mathbb{R}$. Thus $f$ is uniformly continuous on $\mathbb{R}$.

**Exercise 2.109.** If $f : \mathbb{R} \to \mathbb{R}$ is bounded, continuous, and monotone, show that

$$\lim_{x \to -\infty} f(x) \in \mathbb{R}, \qquad \lim_{x \to +\infty} f(x) \in \mathbb{R}.$$

(*Hint.* If $f$ is increasing and $x_n \to \infty$, then $(f(x_n))$ is bounded and monotone, hence convergent; identify its limit with the supremum of the tail values.)

**Example 2.110.** Let $f : (a, b] \to \mathbb{R}$ and $f : (b, c) \to \mathbb{R}$ be uniformly continuous. Then $f : (a, c) \to \mathbb{R}$ is uniformly continuous.

*Proof.* Fix $\varepsilon > 0$. Since $f$ is uniformly continuous on $(a, b]$ and on $(b, c)$, there exists $\delta > 0$ such that whenever $x, y$ both belong to $(a, b]$ or both belong to $(b, c)$ and satisfy $|x - y| < \delta$, one has $|f(x) - f(y)| < \varepsilon$.

Now let $x, y \in (a, c)$ with $|x - y| < \delta$. If $x$ and $y$ lie on the same side of $b$, the conclusion is immediate. Otherwise, after relabeling if necessary, we may assume $x \leq b < y$. Then $|x - b| < \delta$ and $|y - b| < \delta$, so

$$|f(x) - f(y)| \leq |f(x) - f(b)| + |f(b) - f(y)| < 2\varepsilon.$$

Thus $f$ is uniformly continuous on $(a, c)$. □

We next show that a uniformly continuous function admits a unique uniformly continuous extension to the closure of its domain.

**Theorem 2.111** (Uniformly continuous extension to the closure)**.** *Let $f : A \subset \mathbb{R} \to \mathbb{R}$ be uniformly continuous on $A$. Then $f$ extends uniquely to a uniformly continuous function on $\overline{A}$.*

*Proof.* For $x \in \overline{A}$, choose a sequence $(x_n) \subset A$ such that $x_n \to x$. Since $(x_n)$ is Cauchy and $f$ is uniformly continuous on $A$, the sequence $(f(x_n))$ is Cauchy in $\mathbb{R}$, hence convergent. Define

$$\widetilde{f}(x) \coloneqq \lim_{n\to\infty} f(x_n).$$

We claim that this is well defined. Indeed, if $(y_n) \subset A$ also satisfies $y_n \to x$, then $|x_n - y_n| \to 0$, so uniform continuity gives $|f(x_n) - f(y_n)| \to 0$. Hence $\lim f(x_n) = \lim f(y_n)$.

Clearly $\widetilde{f} = f$ on $A$. We now show that $\widetilde{f}$ is uniformly continuous on $\overline{A}$. Let $\varepsilon > 0$. By uniform continuity of $f$ on $A$, there exists $\delta > 0$ such that

$$|u - v| < \delta \quad \Longrightarrow \quad |f(u) - f(v)| < \varepsilon$$

for all $u, v \in A$. Let $x, y \in \overline{A}$ with $|x - y| < \delta/3$. Choose sequences $(x_n), (y_n) \subset A$ such that $x_n \to x$ and $y_n \to y$. For sufficiently large $n$ we have

$$|x_n - x| < \delta/3, \qquad |y_n - y| < \delta/3,$$

and therefore

$$|x_n - y_n| \le |x_n - x| + |x - y| + |y - y_n| < \delta.$$

Hence $|f(x_n) - f(y_n)| < \varepsilon$ for all sufficiently large $n$. Passing to the limit yields

$$|\widetilde{f}(x) - \widetilde{f}(y)| \le \varepsilon.$$

Thus $\widetilde{f}$ is uniformly continuous on $\overline{A}$.

Finally, the extension is unique. If $g : \overline{A} \to \mathbb{R}$ is another uniformly continuous extension of $f$, and if $x \in \overline{A}$, choose $(x_n) \subset A$ with $x_n \to x$. Then

$$\widetilde{f}(x) = \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} g(x_n) = g(x).$$

Hence $g = \widetilde{f}$ on $\overline{A}$. □

Next, we shall see that a uniformly continuous function grows at most linearly.

**Theorem 2.112** (Linear growth of uniformly continuous functions on $\mathbb{R}$)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be uniformly continuous. Then there exist constants $A, B \ge 0$ such that $|f(x)| \le A|x| + B$ for all $x \in \mathbb{R}$.*

*Proof.* For any $\epsilon > 0$, there exists $\delta > 0$ such that $|x-y| < \delta$ implies $|f(x)-f(y)| < 1$. We divide the proof into two parts: one near 0 and the other away from 0. Let $a > 0$. Then $|f(x)| \le A < \infty$ for $x \in [-a, a]$. Now consider $f : [a, \infty) \to \mathbb{R}$.Then for $x \in [a, \infty)$, we can find $n \in \mathbb{N}$ such that $x \in [a + n\delta,\, a + (n+1)\delta]$. Then,

$$f(x) - f(a) = f(x) - f(a + n\delta) + f(a + n\delta) - f(a)$$

$$= f(x) - f(a+n\delta) + \sum_{j=1}^{n}[f(a+j\delta) - f(a+(j+1)\delta)]$$

$$\Rightarrow |f(x)| < 1 + n + |f(a)|$$

$$\Rightarrow \left|\frac{f(x)}{x}\right| < \frac{(n+1)+|f(a)|}{a+n\delta} < \frac{(n+1)+|f(a)|}{n\delta} < \left(1+\frac{1}{n}\right)\frac{1}{\delta} + \frac{|f(a)|}{n\delta} \leq B < \infty$$

Notice that $B$ is independent of $n$, hence $B$ is independent of $x$. That is, $|f(x)| \leq B|x|$ if $x > a$. Hence, we can summarize that $|f(x)| \leq B|x| + A \qquad$ for all $x \in \mathbb{R}$. $\square$

**Example 2.113.** Notice that $f(x) = x^2$ is not uniformly continuous on $\mathbb{R}$, as it does not satisfy the conclusion of the above theorem.

**Example 2.114.** Let $f : \mathbb{R} \to \mathbb{R}$ be differentiable and its derivative is bounded. Then $f$ is uniformly continuous on $\mathbb{R}$. For any $x, y \in \mathbb{R}$, by the Mean Value Theorem,

$$|f(x) - f(y)| = |f'(t)(x-y)| \leq M|x-y|$$

where $t$ is between $x$ and $y$, and $M$ is an upper bound for $|f'(t)|$. However, $f(x) = \sqrt{x}$ for $x \in (0, \infty)$ is uniformly continuous, although its derivative $f'(x) = \frac{1}{2\sqrt{x}}$ is not bounded.

## Exercises

**Exercise 2.115.** Verify that $d(x, y) = |x - y|$ defines a metric on $\mathbb{R}$, and that $d(x, y) = \|x - y\|_2$ defines a metric on $\mathbb{R}^n$.

**Exercise 2.116.** Let $X$ be a set and define $d(x, y) = 0$ if $x = y$ and $d(x, y) = 1$ if $x \neq y$. Show that $d$ is a metric (the discrete metric). Describe the open sets in this metric.

**Exercise 2.117.** Let $d$ be a metric on $X$. Prove that every open ball $B_r(x)$ is an open set and every closed ball $\overline{B}_r(x)$ is a closed set.

**Exercise 2.118.** Let $(X, d)$ be a metric space. Show that a sequence $(x_n)$ converges to $x$ if and only if for every $\varepsilon > 0$ there exists $N$ such that $x_n \in B_\varepsilon(x)$ for all $n \geq N$.

**Exercise 2.119.** Let $f : (X, d_X) \to (Y, d_Y)$ be Lipschitz: $d_Y(f(x), f(y)) \leq L\, d_X(x, y)$. Show that $f$ is uniformly continuous.

**Exercise 2.120.** Let $\|\cdot\|$ be a norm on a vector space $V$ and define $d(x, y) = \|x - y\|$. Prove that $d$ is a metric. Conversely, show that if $d$ is translation-invariant and homogeneous in the sense $d(x + z, y + z) = d(x, y)$ and $d(\lambda x, \lambda y) = |\lambda| d(x, y)$, then $d$ comes from a norm.

**Exercise 2.121.** Prove Hölder's inequality and Minkowski's inequality for sequences, and use them to show that $\ell^p$ is a normed linear space for $1 \leq p \leq \infty$.

**Exercise 2.122.** Show that for $0 < p < 1$, $\|x\|_p = \left(\sum |x_n|^p\right)^{1/p}$ fails the triangle inequality, and thus is not a norm.

**Exercise 2.123.** Let $(X, d)$ be a metric space and $A \subset X$. Define $\operatorname{dist}(x, A) = \inf_{a \in A} d(x, a)$. Prove that $x \mapsto \operatorname{dist}(x, A)$ is 1-Lipschitz.

**Exercise 2.124.** Show that two metrics $d_1, d_2$ on the same set $X$ generate the same open sets if there exist constants $c, C > 0$ such that $c\, d_1 \leq d_2 \leq C\, d_1$.

## Chapter 3

# Completeness and Compactness in Metric and Normed Spaces

*This chapter studies completeness and compactness as the central structural principles of analysis in metric and normed spaces. After establishing the basic theory of Cauchy sequences and complete spaces, we prove the Banach contraction principle as a fundamental existence and uniqueness theorem. We then examine total boundedness, sequential compactness, and compactness, clarifying the precise logical relations among these notions and the roles they play in later arguments.*

**Learning objectives.**

- Separate the notions of completeness, total boundedness, and compactness, and understand exactly where they interact.
- Apply the contraction mapping principle as a structural theorem rather than as an isolated trick.
- Use sequential arguments and covering arguments interchangeably when studying compactness in metric spaces.

## 3.1 Completeness and Fixed Points

**Section overview.**

- Completeness is the hypothesis that turns Cauchy control into genuine convergence.
- The contraction mapping principle is the flagship application: it transforms a quantitative estimate into existence and uniqueness.
- Many later local inversion arguments in several variables should be read as sophisticated descendants of this section.

### 3.1.1 Fixed Points

Fixed-point problems provide a unifying framework for many existence and approximation questions. Let $(X, d)$ be a metric space and let $\varphi : X \to X$ be a self-map. A point $x \in X$ is called a *fixed point* of $\varphi$ if $\varphi(x) = x$.

A standard method is *Picard iteration*: starting from $x_0 \in X$, define a sequence by $x_{n+1} := \varphi(x_n)$ for $n \geq 0$. If $(x_n)$ converges to some $x \in X$ and $\varphi$ is continuous, then $\varphi(x) = x$. In a complete metric space it is often enough to show that $(x_n)$ is a Cauchy sequence, because completeness then guarantees convergence. The most important sufficient condition ensuring that $(x_n)$ is Cauchy is that $\varphi$ is a *contraction.*

**Definition 3.1** (Contraction)**.** A map $\varphi : (X, d) \to (X, d)$ is called a *contraction* if there exists a constant $\alpha \in (0, 1)$ such that

$$d(\varphi(x), \varphi(y)) \le \alpha\, d(x, y), \qquad \forall x, y \in X. \tag{3.1}$$

**Theorem 3.2** (Banach fixed point theorem)**.** *Let $(X, d)$ be a complete metric space. If $\varphi : (X, d) \to (X, d)$ is a contraction, then $\varphi$ has a unique fixed point.*

*Proof.* Fix $x_0 \in X$ and define the Picard iterates by $x_n := \varphi^n(x_0)$ for $n \ge 0$. Since $\varphi$ is a contraction with constant $\alpha \in (0, 1)$,

$$d(x_{n+1}, x_n) = d(\varphi(x_n), \varphi(x_{n-1})) \le \alpha\, d(x_n, x_{n-1}) \le \alpha^n d(x_1, x_0)$$

for every $n \ge 1$. Therefore, if $m > n$, then by the triangle inequality,

$$d(x_m, x_n) \le \sum_{k=n}^{m-1} d(x_{k+1}, x_k) \le \sum_{k=n}^{m-1} \alpha^k d(x_1, x_0) \le \frac{\alpha^n}{1-\alpha}\, d(x_1, x_0).$$

As $n \to \infty$, the right-hand side tends to 0, so $(x_n)$ is a Cauchy sequence. Completeness of $X$ now yields a point $x \in X$ such that $x_n \to x$.

Because every contraction is continuous, we may pass to the limit in $x_{n+1} = \varphi(x_n)$ to obtain

$$\varphi(x) = \varphi\Big(\lim_{n\to\infty} x_n\Big) = \lim_{n\to\infty} \varphi(x_n) = \lim_{n\to\infty} x_{n+1} = x.$$

Thus $x$ is a fixed point of $\varphi$.

To prove uniqueness, let $y \in X$ be any other fixed point. Then

$$d(x, y) = d(\varphi(x), \varphi(y)) \le \alpha\, d(x, y).$$

Since $0 < \alpha < 1$, this is possible only if $d(x, y) = 0$, and hence $x = y$. Therefore $\varphi$ has a unique fixed point. □

Observe that completeness of the ambient metric space is a sufficient condition for the existence of fixed points in many important situations. For example, the map

$$\varphi : (0, \infty) \to (0, \infty), \qquad \varphi(x) = \frac{1}{2}\left(x + \frac{a}{x}\right), \qquad a > 0,$$

has the fixed point $\sqrt{a}$.

Likewise, the contraction condition is sufficient for the existence of a fixed point. The map $\varphi$ above is not a contraction on $(0, \infty)$, because

$$|\varphi(x) - \varphi(y)| = \frac{1}{2} \left|1 - \frac{a}{xy}\right| |x - y|,$$

and the factor $\left|1 - a/(xy)\right|$ is unbounded as $xy \to 0^+$.

**Exercise 3.3.** If $(X, d)$ is a complete metric space and $f : X \to X$ is such that $f^k$ is a contraction for some $k \in \mathbb{N}$, show that $f$ has a unique fixed point. (*Hint.* First consider the case $k = 2$. Use the fact that $f^k$ has a unique fixed point. If $f^2(x_0) = x_0$ and $y_0 = f(x_0)$, then $f(y_0) = x_0$; show that $y_0$ is also a fixed point of $f^2$, hence $y_0 = x_0$.)

**Exercise 3.4.** Let $T : C[0, 1] \to C[0, 1]$ be defined by

$$T(f)(x) = \int_0^x f(t)dt$$

Show that $T^2$ is a contraction but $T$ is not a contraction.

The preceding examples also illustrate a useful principle: the tail of an orbit determines its convergence behaviour, so finitely many initial iterates do not affect the limit.

We next use the fixed point theorem to study existence and uniqueness for the initial value problem

$$\begin{cases} y' = f(x, y) \\ y(0) = y_0 \end{cases} \tag{3.2}$$

using the fixed point theorem.

Assume that $f$ is continuous on a rectangle containing $(0, y_0)$ in its interior and is Lipschitz in the second variable; that is,

$$|f(x, y_1) - f(x, y_2)| \leq K|y_1 - y_2|,$$

where $K$ is a fixed constant. Then equation (3.2) has a unique solution in some neighbourhood of $x = 0$. Observe that solving (3.2) is equivalent to solving

$$\int_0^x y'(t)dt = \int_0^x f(t, y(t))dt$$

that is,

$$y(x) = y_0 + \int_0^x f(t, y(t))dt \tag{3.3}$$

That is, we want $y(t)$ such that (3.3) holds. In other words, we want to get fixed point for the map $\varphi \mapsto F(\varphi)$, where

$$F(\varphi)(x) = y_0 + \int_0^x f(t, \varphi(t))dt,$$

with $\varphi \in C[-\delta, \delta]$ for some $\delta > 0$, which we get very soon. Now,

$$|F(\varphi)(x) - F(\psi)(x)| \leq \int_0^x |f(t, \varphi(t)) - f(t, \psi(t))|dt,$$

$$\leq K \int_0^x |\varphi(t) - \psi(t)|dt$$

$$\leq K \cdot 2\delta \cdot \|\varphi - \psi\|_\infty.$$

Thus, $F : C[-\delta, \delta] \to C[-\delta, \delta]$ is a contraction as long as $2K\delta < 1$, that is if $\delta < \frac{1}{2K}$. Hence $F$ has a unique fixed point in $C[-\frac{1}{2K}, \frac{1}{2K}]$. That is, (3.2) has a unique solution in $|x| < \frac{1}{2K}$.

**Example 3.5.** Consider $y' = 2x(1 + y), \qquad y(0) = 0$. Then

$$\varphi(x) = \int_0^x 2t(1 + \varphi(t))dt.$$

With the initial guess $\varphi^0 \equiv 0$, we get

$$\varphi^1(x) = \int_0^x 2t(1 + 0)\, dt = x^2,$$

$$\varphi^2(x) = \int_0^x 2t(1 + t^2)\, dt = x^2 + \frac{x^4}{2},$$

$$\varphi^3(x) = x^2 + \frac{x^4}{2} + \frac{x^6}{6}.$$

Thus, by induction,

$$\varphi^n(x) = \sum_{k=1}^{n} \frac{x^{2k}}{k!} \longrightarrow e^{x^2} - 1, \tag{3.4}$$

and $\varphi(x) = e^{x^2} - 1$ is a solution, which is same as method of separation of variables. Notice that the series (3.4) converges uniformly on every interval $[-a, a]$, or on any interval $[a, b]$. On the other hand, $\varphi'(x) = 2x(1 + \varphi(x))$ has unique solution in neighborhood of any point $x_0$, that is, $[x_0 - \delta, x_0 + \delta]$ with $\delta < \frac{1}{4}$. (*Hint.* Lipschitz constant = 2.)

## 3.2 Total Boundedness and Compactness

**Section overview.**

- In general metric spaces, boundedness alone is too weak to capture compactness.
- Total boundedness provides the finite-scale approximation principle that replaces the compact interval intuition from $\mathbb{R}$.

### 3.2.1 Totally Bounded Set

Suppose $A$ be a bounded set in $\mathbb{R}$ and, without loss of generality, $A \subset (0, 1)$. Then for $\epsilon = \frac{1}{n} > 0$:

$$A \subset \bigcup_{k=1}^{n} \left( \frac{k-1}{n}, \frac{k}{n} \right]$$

That is, $A$ can be covered by finitely many intervals of arbitrarily small length. A similar argument can be produced for a bounded set $A \subset \mathbb{R}^m$ (or in finite dimensional spaces). Notice that if $A$ is bounded in $\mathbb{R}$, then $A \subset [a, b]$ ($a = \inf A, b = \sup A, b - a < \infty$). Hence,

$$A \subseteq \bigcup_{k=1}^{n} \left[ a + \frac{(k-1)(b-a)}{n},\ a + \frac{k(b-a)}{n} \right]$$

Notice that, with small perturbation of the intervals, $A$ can be covered by open intervals of arbitrarily small length $\epsilon > 0$. However, if the dimension of the space $X$ is infinite, then the above property need not be inherited for an arbitrary bounded set. For example, let $X = \ell^1$, $e_n = (0, 0, \ldots, 1, 0, \ldots)$ (with 1 in the $n$th position):

$$\|e_n - e_m\|_1 = 2, \quad \text{if } n \neq m$$

$$\Rightarrow A = \{e_n : n \in \mathbb{N}\} \subset B_1[0] \subset B_2[0]$$

This means $A$ is bounded. Notice that for any $\epsilon$, $0 < \epsilon < 1$, if $A \subset \bigcup_{n=1}^{\infty} B_\epsilon(e_n)$. But $A$ cannot be covered by finitely many balls of arbitrarily small radius, that is if

$$A \subset \bigcup_{i=1}^{n} B_\epsilon(f_i) \quad \forall f_i \in \ell^1$$

Then, for $\epsilon < 1$, each ball $B_\epsilon(f_i)$ can contain exactly one point of $A$ since $\|e_n - e_m\|_1 = 2$. Also, notice that $A$ has no convergent subsequence. Since $\ell^1$ is complete, it is equivalent to say that $A$ has no Cauchy subsequence.

**Definition 3.6.** $A \subseteq (X, d)$ is said to be totally bounded if for every $\epsilon > 0$, there exist $x_1, x_2, \ldots, x_n \in X$ such that

$$A \subseteq \bigcup_{i=1}^{n} B_\epsilon(x_i)$$

We can show that centers of these balls can be taken from some points of $A$, since

$$A \subseteq \bigcup_{i=1}^{n} B_{\epsilon/2}(x_i)$$

Also, we can assume that $A \cap B_{\epsilon/2}(x_i) \neq \varnothing$, for all $i = 1, 2, \ldots, n$. Then there exists $y_i \in A \cap B_{\epsilon/2}(x_i)$. And a direct verification shows that

$$A \subset \bigcup_{i=1}^{n} B_\epsilon(y_i)$$

(*Hint.* $x \in A \implies d(x, x_i) < \epsilon/2$ for some $i$ and $y_i \in A \cap B_{\epsilon/2}(x_i) \implies d(x, y_i) < d(x, x_i) + d(x_i, y_i) < \epsilon$. Moreover, if $A$ is totally bounded, then we can replace balls with sets in $A$ with arbitrarily small diameter.

**Proposition** (Total boundedness via small-diameter coverings)**.** *A subset $A$ of a metric space $(X, d)$ is totally bounded if and only if for every $\epsilon > 0$ there exist sets $A_1, \dots, A_n \subseteq A$ such that*

$$A \subseteq \bigcup_{i=1}^{n} A_i \qquad \textit{and} \qquad \operatorname{diam}(A_i) < \epsilon \quad (1 \le i \le n).$$

*Proof.* Assume first that $A$ is totally bounded. Given $\epsilon > 0$, choose points $x_1, \dots, x_n \in X$ such that

$$A \subseteq \bigcup_{i=1}^{n} B_{\epsilon/2}(x_i).$$

Set $A_i := A \cap B_{\epsilon/2}(x_i)$. Then $A_i \subseteq A$ and $A \subseteq \bigcup_{i=1}^{n} A_i$. Moreover, if $u, v \in A_i$, then

$$d(u, v) \le d(u, x_i) + d(x_i, v) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

so $\operatorname{diam}(A_i) \le \epsilon$.

Conversely, suppose that for every $\epsilon > 0$ there exist sets $A_1, \dots, A_n \subseteq A$ with

$$A \subseteq \bigcup_{i=1}^{n} A_i \qquad \text{and} \qquad \operatorname{diam}(A_i) < \epsilon.$$

Choose $x_i \in A_i$ whenever $A_i \neq \varnothing$. Then every $u \in A_i$ satisfies

$$d(u, x_i) \le \operatorname{diam}(A_i) < \epsilon,$$

so $A_i \subseteq B_\epsilon(x_i)$. Hence

$$A \subseteq \bigcup_{i=1}^{n} B_\epsilon(x_i),$$

which shows that $A$ is totally bounded. $\square$

A direct verification shows that every totally bounded set is bounded. Also, every finite set is totally bounded.

Observe that total boundedness depends only on the metric. For example, in the discrete metric space $(X, d_0)$, a set $A \subset X$ is totally bounded if and only if $A$ is finite. (*Hint.* If $A \subseteq \bigcup_{i=1}^{n} B_\epsilon(x_i)$ with $x_i \in A$, then for $0 < \epsilon < \frac{1}{2}$ each ball $B_\epsilon(x_i)$ reduces to $\{x_i\}$.) By contrast, if $X = \ell^1$, then $\|e_n - e_m\|_1 = 2$ for $n \neq m$, so the set $A = \{e_n : n \in \mathbb{N}\}$ cannot be covered by finitely many balls of radius $< 2$. Indeed, $A = \{e_n : n \in \mathbb{N}\}$, equipped with the induced metric, satisfies

$$d(e_n, e_m) = \begin{cases} 2 & n \neq m \\ 0 & \text{otherwise} \end{cases}$$

(in its own discrete metric) is not totally bounded.

**Exercise 3.7.** Every subset of a totally bounded set is totally bounded.

**Exercise 3.8.** $A \subset \mathbb{R}$ is totally bounded if and only if $A$ is bounded.

**Exercise 3.9.** $A$ is totally bounded if and only if $A$ is covered by finitely many closed sets of arbitrarily small diameters. (*Hint.* $A \subset \bigcup_{i=1}^{n} A_i$; $\delta(A_i) < \epsilon$ but $\delta(\overline{A_i}) = \delta(A_i) < \epsilon$ and $A \subset \bigcup_{i=1}^{n} \overline{A_i}$)

**Exercise 3.10.** $A$ is totally bounded if and only if $\overline{A}$ is totally bounded. If $A$ is totally bounded, then $A \subseteq \bigcup_{i=1}^{n} A_i$, $\delta(A_i) < \epsilon$. So $\overline{A} \subseteq \bigcup_{i=1}^{n} \overline{A_i}$, $\delta(\overline{A_i}) < \epsilon$, so $\overline{A}$ is totally bounded. On the other hand, if $\overline{A}$ is totally bounded, then for $\epsilon > 0$, $\exists x_1, \ldots, x_n \in X$ such that

$$A \subseteq \overline{A} \subseteq \bigcup_{i=1}^{n} B_i, \qquad \delta(B_i) < \epsilon$$

**Exercise 3.11.** If $A \subset \mathbb{R}^n$ is bounded, then $A$ is totally bounded.

**Proposition** (Sequences and total boundedness). *Let $(x_n)$ be a sequence in a metric space $(X, d)$, and let*

$$A := \{x_n : n \in \mathbb{N}\}$$

*be its range.*

*(i) If $(x_n)$ is Cauchy, then $A$ is totally bounded.*

*(ii) If $A$ is totally bounded, then $(x_n)$ has a Cauchy subsequence.*

*Proof.* **(i)** Let $\epsilon > 0$. Since $(x_n)$ is Cauchy, there exists $N \in \mathbb{N}$ such that

$$d(x_n, x_m) < \epsilon \qquad \text{whenever } m, n \geq N.$$

In particular, for every $n \geq N$,

$$d(x_n, x_N) < \epsilon.$$

Hence

$$A \subseteq \bigcup_{i=1}^{N-1} B_\epsilon(x_i) \cup B_\epsilon(x_N),$$

so $A$ is totally bounded.

**(ii)** Assume that $A$ is totally bounded. For each $k \in \mathbb{N}$, cover $A$ by finitely many sets of diameter less than $2^{-k}$. One of those sets contains infinitely many terms of the sequence; call it $E_k$. Choose the sets inductively so that

$$E_1 \supseteq E_2 \supseteq E_3 \supseteq \cdots, \qquad \operatorname{diam}(E_k) < 2^{-k},$$

and each $E_k$ contains infinitely many terms of $(x_n)$. Now choose indices $n_1 < n_2 < \cdots$ with $x_{n_k} \in E_k$ for every $k$. If $p, q \geq k$, then $x_{n_p}, x_{n_q} \in E_k$, and therefore

$$d(x_{n_p}, x_{n_q}) < 2^{-k}.$$

This shows that $(x_{n_k})$ is a Cauchy subsequence. $\square$

**Corollary 3.12** (Bolzano–Weierstrass limit-point theorem)**.** ***(The Bolzano-Weierstrass Theorem)*** *Every bounded infinite subset of* $\mathbb{R}$ *has a limit point in* $\mathbb{R}$.

*Proof.* Let $A$ be an infinite bounded set in $\mathbb{R}$. Then there exists a sequence of distinct points $x_n \in A$. Since $A$ is totally bounded, $(x_n)$ has a Cauchy subsequence, say $(x_{n_k})$. But $\mathbb{R}$ is complete, so $x_{n_k} \to x \in \mathbb{R}$. Thus, $x$ is a limit point of $A$. $\square$

Recall that a metric space $X$ is complete if and only if every Cauchy sequence in $X$ converges to a point of $X$. Moreover, if $X$ is complete, then a subset $A \subseteq X$ is complete if and only if $A$ is closed in $X$.

The next theorem records two useful forms of completeness that are often used in compactness arguments.

**Theorem 3.13** (Equivalent forms of completeness)**.** *Let* $(X, d)$ *be a metric space. The following assertions are equivalent.*

(i) $(X, d)$ *is complete.*

(ii) *Whenever* $F_1 \supseteq F_2 \supseteq \cdots$ *is a decreasing sequence of nonempty closed subsets of* $X$ *with* $\operatorname{diam}(F_n) \to 0$*, the intersection* $\bigcap_{n=1}^{\infty} F_n$ *consists of exactly one point.*

(iii) *Every infinite totally bounded subset of* $X$ *has a limit point in* $X$*.*

*Proof.* *(i)* $\Rightarrow$ *(ii)*. Choose $x_n \in F_n$. If $m \geq n$, then $x_m, x_n \in F_n$; hence

$$d(x_m, x_n) \leq \operatorname{diam}(F_n) \to 0.$$

Thus $(x_n)$ is Cauchy and, by completeness, $x_n \to x \in X$. Since each $F_n$ is closed and contains the tail $\{x_m : m \geq n\}$, we have $x \in F_n$ for every $n$. Hence $x \in \bigcap_n F_n$. If $y$ also belongs to the intersection, then $x, y \in F_n$ for every $n$, so

$$d(x, y) \leq \operatorname{diam}(F_n) \to 0.$$

Therefore $x = y$, and the intersection is a singleton.

*(ii)* $\Rightarrow$ *(iii)*. Let $A \subseteq X$ be infinite and totally bounded. Choose a sequence of distinct points $(a_n)$ in $A$. By total boundedness, $(a_n)$ has a Cauchy subsequence $(a_{n_k})$. Set

$$F_m := \overline{\{a_{n_k} : k \geq m\}} \qquad (m \geq 1).$$

Then $F_1 \supseteq F_2 \supseteq \cdots$ is a decreasing sequence of nonempty closed sets. Since $(a_{n_k})$ is Cauchy,

$$\operatorname{diam}(F_m) = \operatorname{diam}\{a_{n_k} : k \geq m\} \to 0.$$

By (ii), $\bigcap_m F_m = \{x\}$ for some $x \in X$. Every neighbourhood of $x$ meets infinitely many terms of the subsequence, so $x$ is a limit point of $A$.

*(iii)* $\Rightarrow$ *(i)*. Let $(x_n)$ be a Cauchy sequence in $X$. If the set $\{x_n : n \in \mathbb{N}\}$ is finite, then the sequence is eventually constant and hence convergent. Otherwise this set is infinite and totally bounded, because every Cauchy sequence is totally bounded. By (iii), it has a limit point $x \in X$; hence some subsequence $(x_{n_k})$ converges to $x$. A Cauchy sequence with a convergent subsequence converges to the same limit, so $x_n \to x$. Therefore $X$ is complete. □

**Exercise 3.14.** Suppose that every countable, closed subset in $X$ is complete. Show that $X$ is complete.

**Exercise 3.15.** Show that $X$ is complete if and only if every closed ball in $X$ is complete.

*Remark* 3.16. Total boundedness expresses the idea that a set cannot be too widely dispersed: up to finitely many pieces, it can always be covered by sets of arbitrarily small diameter.

### 3.2.2 Compact Metric Spaces

**Definition 3.17.** A metric space $(X, d)$ is said to be compact if $X$ is complete and totally bounded.

**Theorem 3.18** (Compactness and sequential compactness)**.** $(X, d)$ *is compact if and only if every sequence in $X$ has a convergent subsequence.*

*Proof.* Suppose $X$ is compact (complete and totally bounded). Let $x_n \in X$. Then $A = \{x_n : n \in \mathbb{N}\}$ is totally bounded and hence has Cauchy subsequence, say $x_{n_k}$. But $X$ is complete, which implies $x_{n_k} \to x \in X$.

Conversely, if every sequence $(x_n)$ in $X$ has a convergent subsequence $(x_{n_k})$, then every sequence has a Cauchy subsequence; hence $X$ is totally bounded. Also, let $(x_n)$ be a Cauchy sequence in $X$. Then again $A = \{x_n : n \in \mathbb{N}\}$ is totally bounded and has convergent subsequence, say $x_{n_k} \to x \in X$. Thus, $x_n \to x$.

$$\begin{bmatrix} \text{Totally Bounded} \\ + \text{ Complete} \end{bmatrix} \Longleftrightarrow \begin{bmatrix} \text{every sequence has a Cauchy subsequence} \\ + \text{ Cauchy sequence is convergent} \end{bmatrix}$$

□

**Corollary 3.19** (Basic consequences of compactness)**.**

1. *Let $A \subset X$. If $A$ is compact, then $A$ is closed.*
2. *If $X$ is compact and $A$ is closed, then $A$ is compact.*
   *that is compact subsets of a compact metric space are closed sets.*

(Hint. *(i) If $A$ is compact, then for $x_n \in A$ and $x_n \to x \implies x_{n_k} \to y \in A \implies x_n \to x = y$.*

*(ii) If $A$ is closed and $x_n \in A$, then $x_n \in X \implies x_{n_k} \to x \in X \implies x \in A$, since $A$ is closed.)*

**Exercise 3.20.** If $K$ is a compact subset of $(\mathbb{R}, u)$, then $\inf K$ and $\sup K \in K$. By definition of infimum, there exists $x_n \in K$ such that $x_n \to \inf K$. But, since $K$ is compact, there exists $x_{n_k} \to x \in K$, which implies $\inf K = x$ etc.

**Exercise 3.21.** Let $E = \{x \in \mathbb{Q} : 2 < x^2 < 3\}$. Show that $E$ is closed and bounded in $(\mathbb{Q}, u)$, but not compact. (*Hint.* $\mathbb{Q}$ is not complete.)

**Exercise 3.22.** Suppose $f : (X, d) \to (Y, \rho)$ is continuous, then for $K \subset X$ to be compact, $f(K)$ is compact in $Y$.

Let $y_n \in f(K)$, then $y_n = f(x_n)$ for some $x_n \in K$. Therefore, there exists a subsequence $x_{n_k} \to x \in K$ such that $f(x_{n_k}) \to f(x) \in f(K)$.

**Exercise 3.23.** If $A \subset X$ is compact, then show that $\delta(A) < \infty$. If $A \neq \varnothing$, then there exist $x, y \in A$ such that $\delta(A) = d(x, y)$. Note that

$$\delta(A) = \sup\{d(x, y) : (x, y) \in A \times A\} \quad (\text{say} \quad S)$$

And $d : A \times A \to \mathbb{R}$ is (jointly) continuous. As $A \times A$ is compact, the set $S$ is compact in $\mathbb{R}$. Hence, there exist $(x_0, y_0) \in A \times A$ such that $\delta(A) = d(x_0, y_0)$.

**Exercise 3.24.** Show that $S_1[0] = \{x \in \ell^2 : \|x\|_2 \leq 1\}$ is not compact. (*Hint.* The set $\{e_n : n \in \mathbb{N}\}$ is not totally bounded.)

**Exercise 3.25.** Show that $A = \left\{x \in \ell^2 : |x_n| \leq \frac{1}{n},\ n = 1, 2, \ldots\right\}$ is compact. (*Hint.* $A$ is closed, hence complete. $A$ is totally bounded, since for $\varepsilon = 1$, okay. For $\varepsilon < 1$, only finitely many coordinates are left unpatched (uncovered), hence for each $\varepsilon < 1$, $A = A_\varepsilon \cup B_\varepsilon$, $A_\varepsilon \in \mathbb{R}^n$ for some $n$.

**Corollary 3.26** (Extreme value theorem on compact spaces)**.** *Let $(X, d)$ be compact. Suppose $f : X \to \mathbb{R}$ is continuous, then $f$ is bounded. Moreover, $f$ attains its maximum and minimum.*

*Proof.* $f(X)$ is compact in $\mathbb{R}$, which implies $f(X)$ is closed and bounded. Hence,

$$\sup_{x \in X} f(x) \in \mathbb{R}, \qquad \inf_{x \in X} f(x) \in \mathbb{R}.$$

that is there exist $x_0, y_0 \in X$ such that $f(y_0) = \sup_{x \in X} f(x)$, $f(x_0) = \inf_{x \in X} f(x)$. Hence,

$$f(x_0) \leq f(x) \leq f(y_0) \quad \forall x \in X.$$

□

**Corollary 3.27** (Continuous image of a compact interval)**.** *If $f : [a, b] \to \mathbb{R}$ is continuous, then $f([a, b])$ is compact and $f([a, b]) = [c, d]$ for some $c, d \in \mathbb{R}$.*

**Corollary 3.28** (Completeness of $C(X)$ in the uniform norm)**.** *If $(X, d)$ is a compact metric space and*

$$C(X) = \{f : X \to \mathbb{R} \quad or \quad \mathbb{C} \mid f \text{ is continuous}\}.$$

*Define*

$$\|f\|_\infty = \sup_{x \in X} |f(x)| < \infty.$$

*Then $(C(X), \|\cdot\|_\infty)$ is complete normed linear space.*

**Lemma 3.29** (Open-cover and finite-intersection principles)**.** *Let $(X, d)$ be a metric space. Then the following are equivalent:*

*(a) If $\mathcal{G}$ is an arbitrary collection of open sets in $X$ with $\bigcup_{G \in \mathcal{G}} G \supseteq X$, then there exist $G_1, \ldots, G_n$ (finitely many) such that $\bigcup_{i=1}^n G_i \supseteq X$. (In other words, every open cover has a finite subcover.).*

*(b) If $\mathcal{F}$ is a collection of closed sets in $X$ with $\bigcap_{i=1}^n F_i \neq \varnothing$ for every choice of finitely many $F_i$'s in $\mathcal{F}$, then*

$$\bigcap_{F \in \mathcal{F}} F \neq \varnothing.$$

*(This is called the finite intersection property.).*

*Notice that* (a) $\implies$ $X$ is totally bounded, since

$$X \subseteq \bigcup_{x \in X} B_\varepsilon(x) \implies X \subseteq \bigcup_{i=1}^n B_\varepsilon(x_i).$$

(b) $\implies$ $X$ is complete, since every decreasing sequence of closed sets has nonempty intersection.

*Proof.*

(a) $\implies$ (b): Let $\mathcal{F}$ be a collection of closed sets in $X$ such that $\bigcap_{i=1}^n F_i \neq \varnothing$ for every choice of finitely many $F_i$'s in $\mathcal{F}$. On contrary, suppose $\bigcap_{F \in \mathcal{F}} F = \varnothing$. Then $X = \bigcup_{F \in \mathcal{F}} F^c$ is an open cover of $X$. Hence $X = \bigcup_{i=1}^n \{F_i^c \colon F_i \in \mathcal{F}\}$. This implies $\bigcap_{i=1}^n F_i = \varnothing$, a contradiction.

(b) $\implies$ (a): Suppose $X = \bigcup_{G \in \mathcal{G}} G$ but $X \neq \bigcup_{i=1}^n G_i$ for any choice of finitely many $G_i$'s in $\mathcal{G}$. Then $X \setminus \bigcup_{i=1}^n G_i \neq \varnothing$ for every choice of finitely many $G_i$'s in $\mathcal{G}$, which implies $\bigcap_{i=1}^n G_i^c \neq \varnothing$ for every choice of finitely many sets $G_i$'s from $\mathcal{G}$.

$$\bigcap_{G \in \mathcal{G}} G^c \neq \varnothing \implies \bigcup\{G : G \in \mathcal{G}\} \neq X.$$

□

**Theorem 3.30** (Compactness via cover and intersection criteria)**.** *$X$ is compact if and only if either (a) or (b) (hence both) of the previous lemma is satisfied.*

*Proof.* If either condition (a) or condition (b) of the previous lemma holds, then $X$ is both totally bounded and complete; hence $X$ is compact.

Conversely, assume that $X$ is compact, and let $\mathcal{G}$ be an open cover of $X$ with no finite subcover. Since $X$ is totally bounded, it can be covered by finitely many closed sets of diameter at most 1. At least one of these closed sets, call it $A_1$, cannot be covered by finitely many members of $\mathcal{G}$.

Proceed inductively. Once $A_n$ has been chosen, it is compact and therefore totally bounded, so it can be covered by finitely many closed sets of diameter at most $1/(n+1)$. One of these closed sets, call it $A_{n+1}$, still cannot be covered by finitely many members of $\mathcal{G}$. In this way we obtain a decreasing sequence

$$A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots$$

of nonempty closed sets such that $\operatorname{diam}(A_n) \leq 1/n$ for every $n$ and no $A_n$ admits a finite subcover from $\mathcal{G}$.

Choose $x_n \in A_n$. Since $X$ is compact, the sequence $(x_n)$ has a convergent subsequence $x_{n_k} \to x \in X$. Because the sets are nested and closed, for each fixed $m$ all sufficiently large terms of the subsequence lie in $A_m$, so $x \in A_m$. Thus $x \in \bigcap_{n=1}^{\infty} A_n$.

Now choose $G \in \mathcal{G}$ with $x \in G$. Since $G$ is open, there exists $\epsilon > 0$ such that $B_\epsilon(x) \subset G$. Pick $n$ so large that $1/n < \epsilon$. Because $x \in A_n$ and $\operatorname{diam}(A_n) \leq 1/n$, every point of $A_n$ lies within distance $1/n < \epsilon$ of $x$. Hence $A_n \subset B_\epsilon(x) \subset G$, contradicting the choice of $A_n$. Therefore every open cover has a finite subcover, and $X$ is compact. ◻

**Corollary 3.31** (Cantor intersection property)**.** *$X$ is compact if and only if every decreasing sequence of nonempty closed sets has nonempty intersection.*

$$\textit{that is} \quad F_1 \supset F_2 \supset \cdots \supset F_n \supset F_{n+1} \cdots \implies \bigcap_{n=1}^{\infty} F_n \neq \varnothing.$$

*Proof.* The forward implication follows from the previous theorem.

Conversely, assume that every decreasing sequence of nonempty closed subsets of $X$ has nonempty intersection. To prove compactness, it is enough to show that every sequence in $X$ has a convergent subsequence. Let $(x_n)$ be a sequence in $X$, and for each $n$ set

$$A_n \coloneqq \{x_k : k \geq n\}.$$

Then the closed sets $\overline{A_n}$ are nonempty and decreasing, so by assumption there exists

$$x \in \bigcap_{n=1}^{\infty} \overline{A_n}.$$

Since $x \in \overline{A_n}$ for every $n$, each ball centred at $x$ meets $A_n$. Choose $n_1$ such that $B_1(x) \cap A_{n_1} \neq \varnothing$, and then choose $k_1 \geq n_1$ with $x_{k_1} \in B_1(x)$. Proceed inductively: once $k_j$ is chosen, pick $k_{j+1} > k_j$ such that $x_{k_{j+1}} \in B_{1/(j+1)}(x) \cap A_{k_j+1}$. Then

$$d(x_{k_j}, x) < \frac{1}{j}, \qquad j \in \mathbb{N},$$

so $x_{k_j} \to x$. Thus every sequence has a convergent subsequence, and hence $X$ is compact. $\square$

*Remark* 3.32. Note that, as long as compactness is concerned, we do not require the diameter of $F_n$ tends to zero. Hence $\bigcap_{n=1}^{\infty} F_n$ can contain more than one point. This is in sharp contrast with the condition for completeness.

**Corollary 3.33** (Countable open-cover criterion)**.** *$X$ is compact if and only if every countable open cover admits a finite subcover.*

*Proof.* ( $\implies$ :) Compact $\implies$ lemma (a) holds $\implies$ countable cover has finite subcover. ( $\impliedby$ :) Suppose every countable open cover has a finite subcover. This is equivalent to every countable family of closed sets having the finite intersection property (can be proved similar to the previous lemma). Let $(x_n) \subset X$ be a sequence of distinct terms. Write $A_n = \overline{\{x_k : k \geq n\}}$. Then $x \in \bigcap_{n=1}^{\infty} A_n \neq \varnothing$, so there exists $x_{n_k} \in X$ such that $x_{n_k} \to x$. Hence, $X$ is compact. $\square$

### 3.2.3 Separable Metric Spaces

If a space admits a countable dense set, we say that the space is *separable.* Eventually, it helps determine the size of the space, certainly not in terms of cardinality only, rather dimensions, or in a more general sense of size. Evidently, every totally bounded space is separable.

**Definition 3.34.** A metric space $(X, d)$ is called *separable* if it contains a countable dense subset; that is, if there exists a countable set $D \subseteq X$ such that $\overline{D} = X$.

For example, $\mathbb{Q}$ is a countable dense subset of $\mathbb{R}$. Similarly, $\mathbb{Q}^n$ is dense in $\mathbb{R}^n$, and $\mathbb{Q}^n + i\mathbb{Q}^n$ is dense in $\mathbb{C}^n$.

**Proposition 3.35** (Separable and nonseparable sequence spaces)**.** *For $1 \leq p < \infty$, the space $\ell^p$ is separable. The space $\ell^\infty$ is not separable.*

*Proof.* Let $1 \leq p < \infty$. Denote by

$$c_{00}(\mathbb{Q} + i\mathbb{Q})$$

the set of all finitely supported complex sequences whose nonzero coordinates lie in $\mathbb{Q} + i\mathbb{Q}$. This set is countable, since it is a countable union of finite products of countable sets.

We show that $c_{00}(\mathbb{Q} + i\mathbb{Q})$ is dense in $\ell^p$. Let $x = (x_1, x_2, \ldots) \in \ell^p$ and let $\varepsilon > 0$. Choose $N$ so large that

$$\left( \sum_{j>N} |x_j|^p \right)^{1/p} < \frac{\varepsilon}{2}.$$

Using the density of $\mathbb{Q} + i\mathbb{Q}$ in $\mathbb{C}$, choose $q_1, \ldots, q_N \in \mathbb{Q} + i\mathbb{Q}$ such that

$$\left( \sum_{j=1}^{N} |x_j - q_j|^p \right)^{1/p} < \frac{\varepsilon}{2}.$$

Set $q = (q_1, \ldots, q_N, 0, 0, \ldots)$. Then $q \in c_{00}(\mathbb{Q} + i\mathbb{Q})$ and

$$\|x - q\|_p \le \left( \sum_{j=1}^{N} |x_j - q_j|^p \right)^{1/p} + \left( \sum_{j>N} |x_j|^p \right)^{1/p} < \varepsilon.$$

Hence $\ell^p$ is separable.

To prove that $\ell^\infty$ is not separable, consider

$$S := \{0,1\}^{\mathbb{N}} \subset \ell^\infty.$$

The set $S$ is uncountable, and if $x, y \in S$ with $x \neq y$, then $\|x - y\|_\infty = 1$. Therefore the open balls

$$B_{1/3}(x), \qquad x \in S,$$

are pairwise disjoint. Any dense subset of $\ell^\infty$ must meet every one of these balls, and hence must be uncountable. Thus $\ell^\infty$ is not separable. $\square$

**Exercise 3.36.** Show that $\overline{c_{00}} = c_0$ and hence deduce $c_0$ is separable.

**Exercise 3.37.** Let $B([0,1])$ be the space of all bounded functions on $[0,1]$. Show that $(B([0,1]), \|\cdot\|_\infty)$ is not separable. For $t \in (0,1)$, define $f_t = \chi_{[0,t)}$. Then for $s \neq t$, $s, t \in (0,1)$, we get $\|f_s - f_t\|_\infty = 1$. Then $S = \{B_{1/2}(f_t) : t \in (0,1)\}$ is an uncountable collection of disjoint open balls in $B([0,1])$. If $A$ is any countable set, say $A = \{g_1, g_2, \ldots\} \subset B([0,1])$, then there exists $t_0 \in (0,1)$ such that $B_{1/2}(f_{t_0}) \cap A = \varnothing$. That is, except countably many, all the balls in $S$ are left un-intersected by $A$.

**Exercise 3.38.** The space $(C([0,1]), \|\cdot\|_\infty)$ is separable. (*Hint.* proof of this will be done by Weierstrass approximation theorem, which we do later.)

**Exercise 3.39.** Every totally bounded metric space is separable.

Let $(X, d)$ be totally bounded. For $\epsilon = \frac{1}{n}$, there exist $x_{n_1}, \ldots, x_{n_k}$ such that

$$X = \bigcup_{j=1}^{n_k} B_{\frac{1}{n}}(x_{n_j}).$$

Let $D_{n_k} = \{x_{n_1}, \ldots, x_{n_k}\}$. Then $\mathcal{D} = \bigcup D_{n_k}$ is a countable dense set in $X$.

Next, we consider the compact subsets of the space of continuous functions $C(X)$, then $X$ is a compact metric space. Notice that $\dim C(X) < \infty$ if and only if $X$ is a finite set. Hence, closed and bounded subset of $C(X)$ are compact if $X$ is finite. But the question of compact subsets of $C(X)$, $X$ is compact, is same as when a subset of $C(X)$ is totally bounded? In terms of the Bolzano–Weierstrass theorem, we can rephrase, when (uniformly) bounded sequence in $C(X)$ have a uniformly convergent subsequence?

We will see later that this question is related to the earlier question of asking, When does a pointwise convergent sequence imply uniform convergence? That is, pointwise convergence + [something] $\implies$ uniform convergence.

**Example 3.40.** If $f_n \in C(X)$, $X$ compact, $f_n \xrightarrow{\text{unif}} f$, then $\{f\} \cup \{f_n : n \in \mathbb{N}\}$ is compact. (that is, every Cauchy sequence is totally bounded).

**Definition 3.41.** A collection $\mathcal{F} \subset C(X)$ is said to be uniformly bounded if

$$\sup_{f\in\mathcal{F}} \sup_{x\in X} |f(x)| = \sup_{f\in\mathcal{F}} \|f\|_\infty < \infty.$$

**Example 3.42.** Any uniformly convergent sequence $f_n$ in $B(X)$ (or $C(X)$) is uniformly bounded. (*Hint.* $\|f_n\|_\infty \le \|f\|_\infty + 1$ (for $\epsilon = 1$) for all $n \ge N$, $n \in \mathbb{N}$.)

**Definition 3.43.** A collection $\mathcal{F} \subset C(X)$ is said to be pointwise bounded if for each $x \in X$, $\sup_{f\in\mathcal{F}} |f(x)| < \infty$.

**Example 3.44.** If $f_n \to f$ pointwise, then $f_n$ is pointwise bounded.

**Theorem 3.45** (Heine–Cantor theorem in compact metric spaces)**.** *Let $(X, d)$ be a compact metric space and $f : X \to \mathbb{R}$ (or $\mathbb{C}$) be continuous. Then $f$ is uniformly continuous.*

*Proof.* Let $\epsilon > 0$. For each $x \in X$, continuity of $f$ at $x$ gives $\delta_x > 0$ such that

$$d(x, y) < \delta_x \implies |f(x) - f(y)| < \frac{\epsilon}{2}.$$

Then $\{B_{\delta_x/2}(x) : x \in X\}$ is an open cover of $X$. Since $X$ is compact, there exist $x_1, \dots, x_n \in X$ such that

$$X = \bigcup_{i=1}^{n} B_{\delta_{x_i}/2}(x_i).$$

Set

$$\delta = \frac{1}{2} \min_{1\le i\le n} \delta_{x_i} > 0.$$

Now let $x, y \in X$ with $d(x, y) < \delta$. Choose $i$ such that $x \in B_{\delta_{x_i}/2}(x_i)$. Then

$$d(y, x_i) \le d(y, x) + d(x, x_i) < \delta + \frac{\delta_{x_i}}{2} \le \delta_{x_i}.$$

Hence both $x$ and $y$ belong to $B_{\delta_{x_i}}(x_i)$, and therefore

$$|f(x) - f(y)| \le |f(x) - f(x_i)| + |f(x_i) - f(y)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus $f$ is uniformly continuous on $X$. ☐

Next, we shall discuss the missing ingredient of pointwise convergence to the uniform convergence.

## Exercises

**Exercise 3.46.** Prove that a metric space $(X, d)$ is complete if and only if every Cauchy sequence converges.

**Exercise 3.47.** Let $(X, d)$ be complete and $\varphi : X \to X$ a contraction with constant $0 < q < 1$. Prove that $\varphi$ has a unique fixed point and that the Picard iterates $x_{n+1} = \varphi(x_n)$ converge to it.

**Exercise 3.48.** Show that every compact metric space is complete.

**Exercise 3.49.** Show that every totally bounded set in a metric space is bounded.

**Exercise 3.50.** Prove: a metric space is compact if and only if it is complete and totally bounded.

**Exercise 3.51.** Let $(x_n)$ be a sequence in a compact metric space. Prove that $(x_n)$ has a convergent subsequence (sequential compactness).

**Exercise 3.52.** Let $X$ be a normed space. Prove that if $X$ is finite-dimensional, then every closed and bounded subset of $X$ is compact.

**Exercise 3.53.** Show that $\ell^p$ is complete for $1 \leq p \leq \infty$.

**Exercise 3.54.** Let $A \subset X$ and define $\operatorname{diam}(A) = \sup\{d(x, y) : x, y \in A\}$. Prove that $\operatorname{diam}(\overline{A}) = \operatorname{diam}(A)$.

**Exercise 3.55.** Give an example of a complete metric space that is not compact, and of a totally bounded metric space that is not complete.

Chapter 4

# Connectedness in Metric and Normed Linear Spaces

*This chapter develops connectedness as the topological counterpart of the idea of continuity without jumps. We define connected and disconnected subsets of metric spaces, establish basic permanence properties, and compare connectedness with path-connectedness. Special attention is given to the real line, where connected sets are precisely intervals; this identification explains why the Intermediate Value Theorem is one of the most natural analytic manifestations of connectedness.*

**Learning objectives.**

- Detect connectedness and path-connectedness using both separation arguments and image arguments.
- Translate connectedness into analytic consequences such as interval-valued images and intermediate value principles.
- Build and analyze standard examples that distinguish connected, totally disconnected, and non-path-connected sets.

## 4.1 Connected and Disconnected Sets

**Section overview.**

- Connectedness is the inability to decompose a set into two separated nonempty open pieces.
- The central theme is that connected sets cannot support jump behavior under continuous maps.
- The examples matter: intervals, punctured spaces, and classical pathological subsets should be compared throughout.

The topology of the real line admits a particularly concrete description, because every open set is a countable disjoint union of open intervals. We have already seen that any open set $O \subset \mathbb{R}$ can be expressed as the disjoint union of countably many open intervals. That is,

$$O = \bigcup_{n=1}^{\infty} I_n, \text{ where } I_n = (a_n, b_n).$$

Hence, for any set $A \subset \mathbb{R}$, we get an open set $O \supset A$, and thus

$$A \subset O \subset \bigcup_{n=1}^{\infty} I_n.$$

Hence, any set can be embedded into countably many open intervals. The "connected set" has its natural meaning, and we can extract its definition from the intervals. Recall that an interval cannot be broken into two relatively open parts. On the contrary, suppose that

$$[a, b] = A \cup B,$$

where $A$ and $B$ are nonempty, disjoint, and relatively open sets in $[a, b]$. This implies that $A$ and $B$ are disjoint closed sets too, as

$$A = [a, b] \setminus B, \qquad B = [a, b] \setminus A.$$

Thus $A$ and $B$ are disjoint, nonempty open and closed sets (called clopen sets). To start with, let $b \in B$. Since $B$ is open, $(b - \varepsilon, b] \subset B$ for some $\varepsilon > 0$. Now, let $c = \sup A$. Then $a < c < b$. (if $a = c$, then $A = \{a\}$ (not open), and if $c = b$, then $A \cap B \neq \varnothing$.) By definition of supremum, $(c - \varepsilon, c) \cap A \neq \varnothing$ and $(c, c + \varepsilon) \cap B \neq \varnothing$ (since $c$ is a boundary point of $A$). That is, $c \in \overline{A} = A$ and $c \in \overline{B} = B$, which is a contradiction that $A \cap B = \varnothing$. Motivated by this observation, we now introduce the general notions of connected and disconnected sets.

**Definition 4.1.** A metric space $X$ is said to be *disconnected* (not connected) if there exist two nonempty open sets $A$ and $B$ such that $X = A \cup B$. The sets $A$ and $B$ are called a *disconnection* of $X$.

We say that $X$ is *connected* if $X$ cannot be expressed as a disjoint union of two nonempty open sets in $X$.

Thus, the interval $[a, b]$ is connected.

Note that, when $X = A \cup B$ where $A$ and $B$ are disjoint, nonempty open sets, it follows that $A$ and $B$ are closed sets too (as $A = B^c$, $B = A^c$). Thus, $A$ and $B$ are disjoint, nonempty clopen sets. Thus, $X$ is connected if and only if $X$ has no nontrivial clopen sets. (*Hint.* if $A$ is clopen, then $X = A \cup A^c$, and $A^c$ is also open.)

**Definition 4.2.** A subset $E$ of a metric space $X$ is called *disconnected in* $E$ if there exist nonempty disjoint open sets $U$ and $V$ in $E$ such that $E = U \cup V$.

Note that there exist open sets $A$ and $B$ in $X$ such that

$$U = A \cap E, \qquad V = B \cap E$$

$$\implies E = (A \cap E) \cup (B \cap E) = (A \cup B) \cap E \implies E \subset A \cup B.$$

The sets $A$ and $B$ need not be disjoint. However, they can be refined to disjoint open sets of $X$ whose intersections with $E$ are still $U$ and $V$, respectively.

**Lemma 4.3** (Separating relatively open sets)**.** *Let $E \subset X$. If $U$ and $V$ are disjoint open sets in $E$, then there exist disjoint open sets $A$ and $B$ in $X$ such that $U = A \cap E \quad \text{and} \quad V = B \cap E$.*

*Proof.* For each $x \in U$, because $U$ is open in the subspace topology on $E$, there exists $\varepsilon_x > 0$ such that

$$E \cap B_{\varepsilon_x}(x) \subset U.$$

Likewise, for each $y \in V$ there exists $\varepsilon_y > 0$ such that

$$E \cap B_{\varepsilon_y}(y) \subset V.$$

Set

$$A := \bigcup_{x \in U} B_{\varepsilon_x}(x), \qquad B := \bigcup_{y \in V} B_{\varepsilon_y}(y).$$

Then $A$ and $B$ are open in $X$, and clearly $U \subseteq A \cap E$ and $V \subseteq B \cap E$. The reverse inclusions also hold: if $z \in A \cap E$, then $z \in B_{\varepsilon_x}(x) \cap E$ for some $x \in U$, hence $z \in U$; similarly, $B \cap E = V$.

Finally, $A \cap B = \varnothing$. Indeed, if $z \in A \cap B$, then for some $x \in U$ and $y \in V$ we would have $z \in B_{\varepsilon_x}(x) \cap B_{\varepsilon_y}(y)$. Since $z \in E$ would then imply $z \in U \cap V$, which is impossible, we may shrink the radii if necessary so that the corresponding balls are disjoint; hence the unions remain disjoint. Therefore $A$ and $B$ are disjoint open subsets of $X$ with $A \cap E = U$ and $B \cap E = V$. □

**Theorem 4.4** (Connected subsets of the real line)**.** *A subset $E$ of $\mathbb{R}$ (containing more than one point) is connected if and only if for every $x, y \in E$ with $x < y$ it follows that $[x, y] \subset E$.*

*Proof.* Suppose first that there exist $x, y \in E$ with $x < y$ and some $z$ satisfying $x < z < y$ but $z \notin E$. Then

$$E \subseteq (-\infty, z) \cup (z, \infty),$$

and both sets on the right meet $E$ because they contain $x$ and $y$, respectively. Hence $E$ is disconnected.

Conversely, assume that $[x, y] \subseteq E$ whenever $x, y \in E$ with $x < y$. We show that $E$ is connected. If not, there exist disjoint nonempty open sets $A, B \subset \mathbb{R}$ such that $E \subseteq A \cup B$, with $A \cap E \neq \varnothing$ and $B \cap E \neq \varnothing$. Choose $a \in A \cap E$ and $b \in B \cap E$ with $a < b$. By hypothesis, $[a, b] \subseteq E \subseteq A \cup B$. Then $A \cap [a, b]$ and $B \cap [a, b]$ are disjoint nonempty open subsets of $[a, b]$ whose union is $[a, b]$, contradicting the connectedness of the interval $[a, b]$.

Thus $E$ is connected exactly when it contains the interval between any two of its points. This property is equivalent to saying that $E$ itself is an interval: indeed, if $\alpha = \inf E$ and $\beta = \sup E$ (allowing the values $\pm\infty$), then every point of $(\alpha, \beta)$ lies between two points of $E$ and hence belongs to $E$. □

**Exercise 4.5.** Show that the connected subsets of Cantor's set are only singletons (that is, Cantor set is totally disconnected).

Now, we simplify our study of connected sets with the help of continuous functions. Notice that a discrete metric space (containing more than one point) is always disconnected. We use this fact to identify disconnected sets through comparison via continuous maps.

**Theorem 4.6** (Disconnectedness and two-valued continuous maps)**.** *A space $X$ is disconnected if and only if there exists a continuous surjective map $f : X \to \{0, 1\}$ (two-point discrete space).*

*Proof.* If $f : X \to \{0, 1\}$ is continuous and surjective, then

$$A = f^{-1}(\{0\}), \quad B = f^{-1}(\{1\}). \tag{4.1}$$

are nonempty, disjoint open sets and $A \cup B = X$. Since $f$ is continuous, $A, B$ are closed. Thus, $X$ has a disconnection.

Conversely, if $X = A \cup B$ where $A$ and $B$ are nonempty, disjoint open sets in $X$. Define

$$f(x) = \begin{cases} 0 & x \in A \\ 1 & x \in B \end{cases} \tag{4.2}$$

which is a continuous and surjective map. □

This result gives a perfect replacement of definition of connected sets. Thus, we conclude that $X$ is connected if and only if every continuous map from $X$ into a discrete space is constant.

**Theorem 4.7** (Continuous images of connected sets)**.** *Let $f : (X, d) \to (Y, \rho)$ be continuous, and let $E \subseteq X$. If $E$ is connected, then $f(E)$ is connected.*

*Proof.* Suppose $f(E)$ is not connected. Then there exists a continuous surjective map $g : f(E) \to \{0, 1\}$. Thus, $g \circ f : E \to \{0, 1\}$ is continuous and surjective, so $E$ is disconnected, a contradiction. □

*Remark* 4.8. A non-constant continuous image of an interval is again an interval. This is nothing but the intermediate value theorem.

**Corollary 4.9** (Intermediate value theorem)**.** *Let $I$ be an interval in $\mathbb{R}$, and $f : I \to \mathbb{R}$ be a non-constant continuous function, then $f(I)$ is an interval. In particular, if $a, b \in I$ and $f(a) \neq f(b)$, then $f$ assumes all values between $f(a)$ and $f(b)$.*

**Example 4.10.** If $A, B$ are connected subsets of a metric space $X$, then $A \times B$ is connected in $(X \times X, d \times d)$, where

$$(d \times d)\{(x_1, y_1), (x_2, y_2))\} = d(x_1, x_2) + d(y_1, y_2). \tag{4.3}$$

Suppose $f : A \times B \to \{0, 1\}$ is continuous. We claim that $f$ is constant. For each fixed $a \in A$, the map $f(a, \cdot)$ is continuous on the connected set $B$, hence constant. Similarly, for each fixed

$b \in B$, the map $f(\cdot, b)$ is constant on $A$. Thus $f$ is constant on every vertical and horizontal fibre, and therefore constant on $A \times B$.

**Exercise 4.11.** Show that $(0,1) \times (0,1)$ cannot be written as disjoint union of countably many open balls. (*Hint:* $(0,1) \times (0,1)$ is connected)

**Exercise 4.12.** Let $D \subset \mathbb{R}$ and $f : D \to \mathbb{R}$ continuous. Show that $D$ is connected if and only if the graph of $f$, $G_f = \{(x, f(x)) : x \in D\}$ is connected in $\mathbb{R}^2$. (*Hint:* The map $g : XoXimesX$, defined by $g(x) = (x, f(x))$, is continuous; hence $G_f = g(X)$ is connected because $X$ is connected.) On the other hand, projection $\rho_1 : G_f \to X \implies \rho_1(x, f(x)) = x$, is continuous $\implies$ $X$ is connected)

**Exercise 4.13.** If $A \subset X$ is connected, then for $A \subseteq B \subseteq \overline{A}$, it implies that $B$ is connected. In particular, $\overline{A}$ is connected.

Suppose $f : B \to \{0,1\}$ is continuous and surjective, then $f|_A : A \to \{0,1\}$ is continuous $\implies$ $f$ is constant on $B$.

**Exercise 4.14.** Let $A \subset B \subset X$. If $A$ and $X$ are connected, does it imply $B$ is connected? ( $(0,1) \subset (0,1) \cup (1,2) \subset \mathbb{R}$)

**Exercise 4.15.** *(Topologist's sine curve)* Let $f : [0,1] \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} \sin\left(\frac{\pi}{x}\right) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Show that $f$ is not continuous, but $G_f$ is connected.
(*Hint.* Consider $g : (0,1] \to [-1,1]$ by $g(x) = \sin\left(\frac{\pi}{x}\right)$. Then $g$ is continuous, and hence $g((0,1])$ is connected $\implies$ $g$ is onto. Also, $G_g$ is connected. Since $G_g \subset G_f \subset \overline{G_g} \implies G_f$ is connected.

**Exercise 4.16.** If $f : X \to Y$ is continuous and onto, $Y$ not connected, then $X$ is not connected. (*Hint.* $Y = C \cup D \implies X = f^{-1}(C) \cup f^{-1}(D)$)

**Exercise 4.17.** $\mathrm{Ln}(\mathbb{R}) = \{$space of all $n \times n$ real matrices$\}$ and $\mathrm{GL}_n(\mathbb{R}) = \{A = (x_{ij}) \in \mathrm{Ln}(\mathbb{R}) : \det A \neq 0\}$. Then $\mathrm{GL}_n(\mathbb{R})$ is disconnected in the usual metric on $\mathrm{Ln}(\mathbb{R})$. (*Hint.* $\det(A) = \sum_{i=1}^{n} x_{ii}$ $\implies$ det is continuous. $\implies$ $\mathrm{GL}_n(\mathbb{R}) = (\det)^{-1}(\mathbb{R} \setminus \{0\})$ is open. Now,

$$\det : \ \mathrm{GL}_n(\mathbb{R}) \xrightarrow{\text{continuous, onto}} \mathbb{R} \setminus \{0\}$$

$\implies$ $\mathrm{GL}_n(\mathbb{R}) = \mathrm{GL}_n^+(\mathbb{R}) \cup \mathrm{GL}_n^-(\mathbb{R})$ is disconnected, where

$$(\det)^{-1}\{(-\infty, 0))\} = \mathrm{GL}_n^-(\mathbb{R}), \quad (\det)^{-1}\{(0, \infty))\} = \mathrm{GL}_n^+(\mathbb{R}).$$

(*Hint.* One convenient metric on $\mathrm{Ln}(\mathbb{R})$ is $d(A, B) = \max_{ij} |a_{ij} - b_{ij}|$.)

## 4.2 Path Connectedness

**Section overview.**

- Path connectedness strengthens connectedness by requiring that points be joined by continuous curves.
- In Euclidean spaces it is often the most geometric way to certify that a set is connected.

### 4.2.1 Path Connectedness

A set $E \subset X$ is said to be *path-connected* if for every $x, y \in E$, there exists a continuous function $\gamma : [0,1] \to E$ such that $\gamma(0) = x$ and $\gamma(1) = y$.

**Example 4.18.** Show that the continuous image of a path-connected set is path-connected.

Let $E \subset X$ be path-connected and let $f : E \to \mathbb{C}$ be continuous. Take any $f(x), f(y) \in f(E)$ with $x, y \in E$. Since $E$ is path-connected, there exists a path $\gamma : [0,1] \to E$ such that $\gamma(0) = x$ and $\gamma(1) = y$. Therefore, $(f \circ \gamma)(0) = f(x)$ and $(f \circ \gamma)(1) = f(y)$, so $f \circ \gamma$ is a path joining $f(x)$ and $f(y)$.

**Example 4.19.** Let $P$ be a polynomial in $\mathbb{C}^n$. Then $\mathbb{C}^n \setminus P^{-1}(0)$ is path-connected.

Let $z, w \in \mathbb{C}^n \setminus P^{-1}(0)$. Define $\gamma : \mathbb{C} \to \mathbb{C}^n$ by $\gamma(t) = (1-t)z + tw$, $t \in \mathbb{C}$. Then $\{t \in \mathbb{C} : \gamma(t) \in P^{-1}(0)\} = (P \circ \gamma)^{-1}(0)$. Since $(P \circ \gamma)$ is a polynomial on $\mathbb{C}$, it implies that $(P\circ\gamma)^{-1}(0)$ is a finite set. Hence, $\mathbb{C}\setminus(P\circ\gamma)^{-1}(0)$ is path-connected in $\mathbb{C}$. Hence, $f(\mathbb{C}\setminus(P\circ\gamma)^{-1}(0))$ is path-connected in $\mathbb{C}^n \setminus P^{-1}(0)$ (since $\gamma(\mathbb{C} \setminus P^{-1}(0))$ is contained in $\mathbb{C}^n \setminus P^{-1}(0)$) containing $z$ and $w$. Hence, $\mathbb{C}^n \setminus P^{-1}(0)$ is path-connected. (Note that $\gamma$ is not onto unless $n = 1$, hence $\gamma(\mathbb{C} \setminus (P \circ \gamma)^{-1}(0)) \subsetneq \mathbb{C}^n \setminus P^{-1}(0)$.)

*Topologist's Sine Curve:* Let $f : [0,1] \to [-1,1]$ by

$$f(x) = \begin{cases} \sin \frac{\pi}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

Then

$$G_f = \{(x, \sin \frac{\pi}{x}) : x \in (0,1]\} \cup \{(0,0)\}$$

is not open. $G_f$ is not path-connected. (The hope comes from the fact that $f$ is not continuous at 0.) Suppose, on the contrary, that there is a continuous path

$$\gamma : [0,1] \to G_f = \{(x, \sin \frac{1}{x}) : x \neq 0\} \cup \{(0,0)\}$$

where $\gamma(0) = (0,0)$ and $\gamma(1) = (1,0)$; write $\gamma = (\gamma_1, \gamma_2)$. Since $\gamma$ is continuous, $\gamma$ becomes uniformly continuous. For $\varepsilon = 1 > 0$, there exists $\delta > 0$ such that

$$|s-t| < \delta \implies |\gamma_2(s) - \gamma_2(t)| < 1$$

Since $0 \in \gamma^{-1}\{(0,0)\}$, let $t^* = \sup\ \gamma^{-1}\{(0,0)\} < 1$ since $\gamma(1) = (1,0)$.

Choose $\delta_1 > 0$ such that $0 \le t^* < t^* + \delta_1 < 1$ and $\delta_1 < \delta$. Note that

$$t^* = \sup\{ t : \gamma(t) = (\gamma_1(t), \gamma_2(t)) = (0,0) \}$$

So, there exists $t_n \to t^*$, with $\gamma_1(t_n) = 0 \implies \gamma_1(t^*) = 0$, but $\gamma_1(t^* + \delta_1) > 0$. Since $0 = \gamma_1(t^*) < \gamma_1(t^* + \delta_1) < 1$, by IVP, for large $N$, there exists $s, t \in (t^*, t^* + \delta_1)$ such that

$$\gamma_1(t) = \frac{2}{N+1}, \quad \gamma_1(s) = \frac{2}{N}$$

Therefore,

$$\gamma_2(t) = \sin\left(\frac{N+1}{2}\right)\pi, \quad \gamma_2(s) = \delta_1 \sin\left(\frac{N\pi}{2}\right)$$

So,

$$|\gamma_2(t) - \gamma_2(s)| = 1$$

This is a contradiction.

**Example 4.20.** $\mathbb{R}^n \setminus \{0\}$, with $n \ge 2$, is connected.

Indeed, it is path-connected. Given $x, y \in \mathbb{R}^n \setminus \{0\}$, one may join them by a polygonal path that avoids the origin; this is possible precisely because $n \ge 2$. Hence $\mathbb{R}^n \setminus \{0\}$ is connected.

**Example 4.21.** Let $S^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$. Then $S^{n-1}$ is connected.

Define $\varphi : \mathbb{R}^n \setminus \{0\} \to S^{n-1}$ by

$$\varphi(x) = \frac{x}{\|x\|}$$

Then $\varphi$ is continuous and onto, hence $S^{n-1}$ is connected. In fact, $S^{n-1}$ is the continuous image of the path-connected set $\mathbb{R}^n \setminus \{0\}$, and is therefore path-connected.

**Example 4.22. Alternative argument.** If $A \subset \mathbb{R}$ is connected, then $A$ is an interval. Indeed, suppose that $x, y \in A$ with $x < z < y$, but $z \notin A$. Define

$$f(s) = \begin{cases} 1, & s < z, \\ -1, & s > z. \end{cases}$$

Since $z \notin A$, the map $f : A \to \{1, -1\}$ is well defined and continuous. It is also onto, because $x$ and $y$ lie on different sides of $z$. This contradicts the connectedness of $A$. Hence every point between $x$ and $y$ belongs to $A$, and therefore $A$ is an interval.

**Example 4.23.** Let $f : \mathbb{R} \to \mathbb{R}$ be such that

$$G_f = \{(x, f(x)) : x \in \mathbb{R}\}$$

is closed and connected in $\mathbb{R}^2$. Then $f$ is continuous.

*Proof.* Fix $x \in \mathbb{R}$, and let $(x_n)$ be any sequence with $x_n \to x$. We show that $f(x_n) \to f(x)$.

First assume that $(f(x_n))$ has a bounded subsequence $(f(x_{n_k}))$. Passing to a further subsequence if necessary, we may suppose that $f(x_{n_k}) \to y$ for some $y \in \mathbb{R}$. Then

$$(x_{n_k}, f(x_{n_k})) \to (x, y).$$

Since $G_f$ is closed, $(x, y) \in G_f$, and hence $y = f(x)$. Thus every bounded subsequence of $(f(x_n))$ converges to $f(x)$.

Suppose now that $f(x_n) \not\to f(x)$. After passing to a subsequence, we may assume that

$$|f(x_n) - f(x)| \geq \varepsilon_0 \qquad (n \geq 1)$$

for some $\varepsilon_0 > 0$. By the previous paragraph, this subsequence cannot contain a bounded subsequence. Therefore, after passing to a further subsequence if necessary, we may assume that $|f(x_n)| \to \infty$.

We claim that there exists $\delta > 0$ such that whenever $|t - x| < \delta$, either $|f(t) - f(x)| < 1$ or $|f(t) - f(x)| > 2$. Indeed, if this were false, we could find a sequence $u_n \to x$ such that

$$1 \leq |f(u_n) - f(x)| \leq 2 \qquad (n \geq 1).$$

Then $(f(u_n))$ is bounded, so it has a convergent subsequence $f(u_{n_k}) \to w$. Since $u_{n_k} \to x$ and $G_f$ is closed, we get $(x, w) \in G_f$, hence $w = f(x)$, which contradicts $1 \leq |f(u_{n_k}) - f(x)| \leq 2$.

Choose such a $\delta$, set $I = [x - \delta, x + \delta]$, and consider

$$A = G_f \cap (I \times \{s : |s - f(x)| < 1\}),$$
$$B = G_f \cap (I \times \{s : |s - f(x)| > 1\}).$$

These are disjoint relatively open subsets of $G_f \cap (I \times \mathbb{R})$, and by construction

$$G_f \cap (I \times \mathbb{R}) = A \cup B.$$

Moreover, $(x, f(x)) \in A$, while $(x_n, f(x_n)) \in B$ for all sufficiently large $n$. Thus the portion of the graph lying over the interval $I$ is disconnected, contradicting the connectedness hypothesis on $G_f$. Therefore $f(x_n) \to f(x)$ for every sequence $x_n \to x$, and hence $f$ is continuous at $x$. Since $x$ was arbitrary, $f$ is continuous on $\mathbb{R}$. □

**Example 4.24.** Let $K = \{\frac{1}{n} : n \geq 1\}$ and $E = ([0,1] \times \{0\}) \cup (K \times [0,1])$. Then $E$ is path-connected (Why?)

Let $C = E \times (\{0\} \times [0,1])$, known as the *comb space*, which is path-connected. The deleted comb space $C_0 = E \cup \{(0,1)\}$ is connected, since $E \subset C_0 \subset \overline{E}$ and $E$ is connected. But $C_0$ is *not* path-connected, because there is no path connecting $(0,1)$ and $(1,0)$. On the contrary, suppose

$$\gamma : [0,1] \to C_0$$

Suppose, for contradiction, that there is a continuous path with $\gamma(0) = (0,1)$ and $\gamma(1) = (1,0)$. Then $\gamma^{-1}((0,1))$ is a closed set, and let $t_0 = \sup \gamma^{-1}((0,1)) = \sup\{t \in [0,1] : \gamma(t) = (0,1)\}$. We claim that there exists $t_1 \in (t_0, 1]$ such that

$$(P_1 \circ \gamma)\{(t_0, t_1)\} \subseteq K,$$

where $P_1 : \mathbb{R}^2 \to \mathbb{R}$ is the projection onto the $x$-axis. Suppose the claim is false. Then $\exists t_n \in (t_0, 1]$ with $t_n \to t_0$. By assumption, $\exists s_n \in (t_0, t_n)$ such that $\gamma(s_n) = (x_n, 0)$ for some $x_n \in [0,1] \setminus K$.

Note that $s_n \to t_0$. By continuity, $(x_n, 0) = \gamma(s_n) \to \gamma(t_0) = (0,1)$, which is absurd. Thus, there exists $t_1 \in (t_0, 1]$ such that $(P_1 \circ \gamma)\{(t_0, t_1)\} \subseteq K \implies 1 \in (P_1 \circ \gamma)(t_0, t_1)$ is a connected subset of $K$. Hence $(P_1 \circ \gamma)(t_0, t_1) = \{1\}$ (by continuity), but $(P_1 \circ \gamma)(t_0) = 0$, an absurd.

**Example 4.25.** Let $U$ be an open set in $\mathbb{R}^n$ (or $\mathbb{C}^n$). Then $U$ is path-connected if and only if $U$ is connected.

Fix a point $p \in U$, and let

$$\mathcal{A} = \{q \in U : \text{there exists a path in } U \text{ joining } p \text{ to } q\}.$$

We show that $\mathcal{A}$ is both open and closed in $U$.

If $q \in \mathcal{A}$, then $q \in U$, so there exists $r > 0$ such that $B_r(q) \subset U$. For any $s \in B_r(q)$, the line segment joining $q$ to $s$ lies in $B_r(q)$, hence in $U$. Since $q$ is already joined to $p$ by a path in $U$, concatenating the two paths shows that $s \in \mathcal{A}$. Therefore $B_r(q) \subset \mathcal{A}$, and $\mathcal{A}$ is open in $U$.

Now let $q \in U \setminus \mathcal{A}$. Again choose $r > 0$ such that $B_r(q) \subset U$. If some $s \in B_r(q)$ belonged to $\mathcal{A}$, then the line segment from $s$ to $q$ would lie in $U$, and concatenating it with a path from $p$ to $s$ would produce a path from $p$ to $q$, contradicting $q \notin \mathcal{A}$. Hence $B_r(q) \subset U \setminus \mathcal{A}$, so $U \setminus \mathcal{A}$ is open in $U$.

Thus $\mathcal{A}$ is nonempty, open, and closed in the connected set $U$. Therefore $\mathcal{A} = U$, and every point of $U$ can be joined to $p$ by a path in $U$. Hence $U$ is path-connected.

## Exercises

**Exercise 4.26.** Prove that the image of a connected set under a continuous map is connected.

**Exercise 4.27.** Prove that an interval in $\mathbb{R}$ is connected.

**Exercise 4.28.** Show that if $E \subset \mathbb{R}$ is connected, then $E$ is an interval (possibly a single point).

**Exercise 4.29.** Let $X$ be a metric space. Prove that path-connectedness implies connectedness.

**Exercise 4.30.** Give an example of a connected set that is not path-connected.

**Exercise 4.31.** Let $A \subset X$. Define the connected component of $x \in A$ to be the union of all connected subsets of $A$ that contain $x$. Prove that components form a partition of $A$ and are closed in $A$.

**Exercise 4.32.** Prove the Intermediate Value Theorem using connectedness: if $f : [a, b] \to \mathbb{R}$ is continuous, then $f([a, b])$ is an interval.

**Exercise 4.33.** Let $f : \mathbb{R} \to \mathbb{R}$ be continuous and periodic. Prove that $f$ is uniformly continuous.

**Exercise 4.34.** Show that $\mathbb{R} \setminus \{0\}$ is disconnected, but $\mathbb{R}^2 \setminus \{0\}$ is connected.

**Exercise 4.35.** Let $U, V$ be disjoint nonempty open sets in $X$ with $X = U \cup V$. Prove that every continuous map $f : X \to \mathbb{R}$ has disconnected range unless it is constant on one of $U$ or $V$.

Chapter 5

# Continuity and Monotone Functions on the Real Line

*This chapter develops the basic theory of limits and continuity for functions on the real line, with special emphasis on the logical form of the $\varepsilon$–$\delta$ definitions and their sequential counterparts. A substantial part of the chapter is devoted to monotone functions, which form one of the most instructive classes of examples in elementary analysis: they admit one-sided limits everywhere, their discontinuities are necessarily jump discontinuities, and those discontinuities form a countable set. We also examine inverse maps of strictly monotone functions and construct monotone functions with prescribed jumps on countable sets.*

**Learning objectives.**

- Compare pointwise continuity, monotonicity, and one-sided behavior on the real line.
- Understand how inverse maps of monotone functions inherit continuity and order properties.
- Use constructive examples to see how prescribed discontinuity patterns can coexist with monotonicity.

## 5.1 Limits and Continuity

**Section overview.**

- Limit processes encode the passage from local estimates to global statements about functions.
- Continuity identifies the class of functions that interact stably with algebraic operations, order, and composition.

### 5.1.1 Limits and one-sided limits

Let $I \subseteq \mathbb{R}$ be an interval and let $a$ be a point such that $I \cap (a-\delta, a+\delta) \setminus \{a\} \neq \varnothing$ for every $\delta > 0$. Let $f : I \setminus \{a\} \to \mathbb{R}$ be a function.

**Definition 5.1** (Limit at a point)**.** We say that $L \in \mathbb{R}$ is the *limit* of $f$ at $a$, and we write $\lim_{x\to a} f(x) = L$, if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$0 < |x-a| < \delta,\ x \in I \implies |f(x) - L| < \varepsilon. \tag{5.1}$$

**Definition 5.2** (One-sided limits)**.** Assume that $I$ contains points to the left of $a$. We say that $L \in \mathbb{R}$ is the *left-hand limit* of $f$ at $a$, and we write $\lim_{x\to a^-} f(x) = L$, if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$a - \delta < x < a \implies |f(x) - L| < \varepsilon. \tag{5.2}$$

Assume that $I$ contains points to the right of $a$. We say that $M \in \mathbb{R}$ is the *right-hand limit* of $f$ at $a$, and we write $\lim_{x\to a^+} f(x) = M$, if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$a < x < a + \delta \implies |f(x) - M| < \varepsilon. \tag{5.3}$$

*Remark* 5.3. If both one-sided limits exist and are equal, then the (two-sided) limit exists and equals their common value. Conversely, if $\lim_{x\to a} f(x)$ exists, then both one-sided limits exist and coincide with it.

**Theorem 5.4** (Sequential characterization of limits). *Let $f : I \setminus \{a\} \to \mathbb{R}$ and $L \in \mathbb{R}$. Then $\lim_{x\to a} f(x) = L$ if and only if for every sequence $(x_n)$ in $I \setminus \{a\}$ with $x_n \to a$ we have $f(x_n) \to L$.*

*Proof.* Assume first that $\lim_{x\to a} f(x) = L$. Let $(x_n)$ be any sequence in $I \setminus \{a\}$ with $x_n \to a$. Given $\varepsilon > 0$, choose $\delta > 0$ as in (5.1). Since $x_n \to a$, there exists $N$ such that $|x_n - a| < \delta$ for all $n \geq N$, and hence $|f(x_n) - L| < \varepsilon$ for all $n \geq N$. Thus $f(x_n) \to L$.

Conversely, assume that the sequential condition holds but the $\varepsilon$–$\delta$ condition fails. Then there exists $\varepsilon_0 > 0$ such that for every $k \in \mathbb{N}$ there is a point $x_k \in I$ with $0 < |x_k - a| < 1/k$ and $|f(x_k) - L| \geq \varepsilon_0$. The sequence $(x_k)$ satisfies $x_k \to a$, but $f(x_k) \not\to L$, which contradicts the assumption. □

### 5.1.2 Continuity

**Definition 5.5** (Continuity at a point). Let $f : I \to \mathbb{R}$ and let $a \in I$. We say that $f$ is *continuous at $a$* if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$|x - a| < \delta \text{ and } x \in I \implies |f(x) - f(a)| < \varepsilon.$$

Equivalently, $f$ is continuous at $a$ if and only if $\lim_{x\to a} f(x) = f(a)$.

*Remark* 5.6. If both one-sided limits $f(a^-)$ and $f(a^+)$ exist and satisfy $f(a^-) = f(a) = f(a^+)$, then $f$ is continuous at $a$. If $f(a^-)$ and $f(a^+)$ exist but are unequal, then $f$ has a *jump discontinuity* at $a$.

## 5.2 Monotone Functions

**Section overview.**

- Monotone functions are among the simplest nonlinear maps, yet they already exhibit one-sided limits, jump discontinuities, and strong regularity properties.
- They form an important bridge between order structure and analytic behavior on the real line.

### 5.2.1 One-sided limits and discontinuities

**Definition 5.7** (Monotonicity). Let $I \subseteq \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$. We say that $f$ is *increasing* on $I$ if $x < y$ implies $f(x) \leq f(y)$. We say that $f$ is *decreasing* on $I$ if $x < y$ implies $f(x) \geq f(y)$. A function is *monotone* if it is either increasing or decreasing.

**Theorem 5.8** (Existence of one-sided limits)**.** *Let $f : (a, b) \to \mathbb{R}$ be monotone and let $c \in (a, b)$. Then both one-sided limits $f(c^-)$ and $f(c^+)$ exist in $\mathbb{R}$.*

*Proof.* Assume that $f$ is increasing. Set

$$L := \sup\{f(x) : a < x < c\}, \qquad M := \inf\{f(x) : c < x < b\}.$$

Then $L \le f(c) \le M$. Fix $\varepsilon > 0$. Choose $x_0 \in (a, c)$ such that $f(x_0) > L - \varepsilon$. If $c - \delta < x < c$ with $\delta := c - x_0$, then $x_0 \le x < c$ and hence

$$L - \varepsilon < f(x_0) \le f(x) \le L,$$

so $|f(x) - L| < \varepsilon$. This shows that $\lim_{x\to c^-} f(x) = L$. The proof for $M = \lim_{x\to c^+} f(x)$ is analogous. □

**Corollary 5.9** (Discontinuities of a monotone function)**.** *Let $f : (a, b) \to \mathbb{R}$ be monotone. Then every discontinuity of $f$ is a jump discontinuity, and the set of discontinuities is at most countable.*

*Proof.* Assume that $f$ is increasing. By the theorem, $f(c^-)$ and $f(c^+)$ exist for every $c \in (a, b)$ and satisfy $f(c^-) \le f(c) \le f(c^+)$. Thus $f$ is discontinuous at $c$ precisely when $f(c^-) < f(c^+)$, in which case the discontinuity is a jump.

For each discontinuity point $c$, choose a rational number $q_c \in \mathbb{Q}$ such that

$$f(c^-) < q_c < f(c^+).$$

If $c \ne d$ are two discontinuity points, then the open intervals $(f(c^-), f(c^+))$ and $(f(d^-), f(d^+))$ are disjoint, because monotonicity implies $c < d$ yields $f(c^+) \le f(d^-)$. Hence the map $c \mapsto q_c$ is injective from the set of discontinuities into $\mathbb{Q}$. Since $\mathbb{Q}$ is countable, the set of discontinuities is at most countable. □

## 5.3 Surjections, Inverses, and Continuity

**Section overview.**

- Once monotonicity and continuity are understood, a natural next question is when inverse maps inherit the same regularity.
- Surjective monotone maps provide a clean setting in which this issue can be analyzed completely.

**Theorem 5.10** (Monotone surjections are continuous)**.** *Let $f : [a, b] \to [c, d]$ be monotone and surjective. Then $f$ is continuous on $[a, b]$.*

*Proof.* Assume that $f$ is increasing. Surjectivity forces $f(a) = c$ and $f(b) = d$. If $f$ were discontinuous at some $x_0 \in (a, b)$, then $f(x_0^-) < f(x_0^+)$ by the previous corollary. Choose $y$ with

$f(x_0^-) < y < f(x_0^+)$. For $x < x_0$ we have $f(x) \le f(x_0^-)$, and for $x > x_0$ we have $f(x) \ge f(x_0^+)$. Thus $f(x) \ne y$ for every $x \in [a, b]$, contradicting surjectivity. Therefore $f$ is continuous. □

**Proposition 5.11** (Monotone surjections onto open intervals)**.** *Let $f : (a, b) \to (c, d)$ be monotone and surjective. Then $f$ is continuous on $(a, b)$.*

*Proof.* Assume that $f$ is increasing. If $f$ were discontinuous at some $x_0 \in (a, b)$, then $f(x_0^-) < f(x_0^+)$. Choose $y$ with $f(x_0^-) < y < f(x_0^+)$. For $x < x_0$ we have $f(x) \le f(x_0^-)$, and for $x > x_0$ we have $f(x) \ge f(x_0^+)$. Thus $f(x) \ne y$ for every $x \in (a, b)$, contradicting surjectivity. Therefore $f$ is continuous. □

**Proposition 5.12** (Inverse of a strictly monotone map)**.** *Let $f : (a, b) \to (c, d)$ be strictly monotone and surjective. Then $f$ is bijective, and its inverse $f^{-1} : (c, d) \to (a, b)$ is strictly monotone. If, in addition, $f$ is continuous, then $f^{-1}$ is continuous.*

*Proof.* Strict monotonicity implies injectivity, and surjectivity gives bijectivity. If $f$ is strictly increasing, then $y_1 < y_2$ implies $f^{-1}(y_1) < f^{-1}(y_2)$, so $f^{-1}$ is strictly increasing, and similarly for the decreasing case.

Assume that $f$ is strictly increasing and continuous. Fix $y_0 \in (c, d)$ and set $x_0 = f^{-1}(y_0)$. Given $\varepsilon > 0$, continuity of $f$ at $x_0$ yields $\delta > 0$ such that $|x - x_0| < \delta$ implies $|f(x) - y_0| < \varepsilon$. Since $f$ is increasing, this implies that $y_0 - \varepsilon < f(x_0 - \delta) \le y \le f(x_0 + \delta) < y_0 + \varepsilon$ forces $|f^{-1}(y) - x_0| < \delta$. Hence $f^{-1}$ is continuous at $y_0$. □

*Remark* 5.13. A monotone map from a compact interval onto an open interval cannot be continuous. Indeed, if $f : [a, b] \to (c, d)$ were continuous and onto, then the image $f([a, b])$ would be compact, which is impossible because $(c, d)$ is not compact.

**Example 5.14** (Cantor function)**.** The Cantor (or "devil's staircase") function $\tilde{f} : [0, 1] \to [0, 1]$ is continuous, increasing, and surjective, but it is not injective.

### 5.3.1 Preimages of intervals under monotone maps

**Proposition 5.15** (Level sets of a monotone function)**.** *Let $I \subseteq \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$ be increasing. For every $\alpha \in \mathbb{R}$, the set*

$$E_\alpha := \{x \in I : f(x) > \alpha\} = f^{-1}((\alpha, \infty))$$

*is either empty or an interval.*

*Proof.* Assume that $E_\alpha \ne \varnothing$. If $x' \in E_\alpha$ and $x \in I$ satisfies $x > x'$, then monotonicity gives

$$f(x) \ge f(x') > \alpha,$$

so $x \in E_\alpha$. Thus $E_\alpha$ is an upper interval in $I$, hence an interval. □

*Remark* 5.16. The analogous sets $\{x \in I : f(x) < \alpha\}$ and $\{x \in I : f(x) \geq \alpha\}$ are likewise intervals (possibly empty). This simple observation is often useful when studying the interaction between order and continuity.

## 5.4 Constructing Monotone Functions with Prescribed Jumps

**Section overview.**

- This construction makes concrete the principle that discontinuities of monotone functions are highly structured.
- It shows that prescribed jump data on a countable set can be encoded into an increasing function.

Let $D$ be a countable subset of $\mathbb{R}$. We now construct an increasing function whose discontinuities are contained in $D$, and whose jump at each point of $D$ can be prescribed.

**Proposition 5.17** (Construction)**.** *Write $D = \{x_1, x_2, \dots\}$, and choose numbers $\varepsilon_n \in (0,1)$ such that $\sum_{n=1}^{\infty} \varepsilon_n < \infty$. Define*

$$f(x) := \sum_{\substack{n \in \mathbb{N} \\ x_n \leq x}} \varepsilon_n,$$

*with the convention that $f(x) = 0$ if the index set $\{n : x_n \leq x\}$ is empty. Then $f$ is increasing on $\mathbb{R}$. Moreover, for each $k \in \mathbb{N}$,*

$$f(x_k^-) = f(x_k) - \varepsilon_k, \qquad f(x_k^+) = f(x_k),$$

*and hence the jump size at $x_k$ equals $\varepsilon_k$. Finally, $f$ is continuous at every point of $\mathbb{R} \setminus D$.*

*Proof.* If $x < y$, then the index set $\{n : x_n \leq x\}$ is contained in $\{n : x_n \leq y\}$, so $f(y) \geq f(x)$. Fix $k$ and let $y \downarrow x_k$. Then the set $\{n : x_k < x_n \leq y\}$ eventually becomes empty, so $\sum_{x_k < x_n \leq y} \varepsilon_n \to 0$, which gives $f(x_k^+) = f(x_k)$. Similarly, if $x \uparrow x_k$, then the only term that can "disappear" at the limit is $\varepsilon_k$, which yields $f(x_k^-) = f(x_k) - \varepsilon_k$.

If $x \notin D$, then for $y \to x$ the index set $\{n : \min\{x, y\} < x_n \leq \max\{x, y\}\}$ eventually becomes empty, so the corresponding sum tends to 0 and $f$ is continuous at $x$. □

**Example 5.18.** If $D = \mathbb{Z}$ and $\sum_{n \in \mathbb{Z}} \varepsilon_n < \infty$ (for instance by choosing $\varepsilon_n = 2^{-|n|-1}$), then the above construction yields a monotone function that is constant on each open interval $(n, n+1)$.

**Example 5.19.** Let $D$ be the set of endpoints of the intervals removed in the construction of the Cantor set. By choosing a suitable summable sequence $(\varepsilon_n)$ and applying the above construction to an enumeration of $D$, one obtains a monotone function whose discontinuities occur only at points of $D$. This viewpoint provides one route to the Cantor function.

**Example 5.20.** Define $f : [0,1] \to \mathbb{R}$ by

$$f(x) = x + \sum_{n=0}^{n_x} 2^{-n}, \qquad n_x := \left\lfloor \frac{1}{1-x} \right\rfloor \text{ for } x < 1,$$

and set $f(1) = 3$. Show that $f$ is strictly increasing and discontinuous at the countable set $\left\{1 - \frac{1}{k} : k \in \mathbb{N}\right\}$.

## Exercises

**Exercise 5.21.** Give the $\varepsilon$–$\delta$ proof that $\lim_{x\to a}(x^2) = a^2$.

**Exercise 5.22.** Show that $f$ is continuous at $a$ if and only if $x_n \to a$ implies $f(x_n) \to f(a)$.

**Exercise 5.23.** Prove that sums, products, and compositions of continuous functions are continuous (with full $\varepsilon$–$\delta$ arguments).

**Exercise 5.24.** Show that a continuous function on a compact set is bounded and attains its maximum and minimum.

**Exercise 5.25.** Prove that a continuous function on $[a, b]$ is uniformly continuous.

**Exercise 5.26.** Let $f$ be monotone on an interval. Prove that the one-sided limits $f(x^-)$ and $f(x^+)$ exist at every interior point.

**Exercise 5.27.** Show that a monotone function has at most countably many discontinuities.

**Exercise 5.28.** Let $f$ be strictly increasing and continuous on an interval $I$. Prove that $f(I)$ is an interval and that $f^{-1}$ is continuous on $f(I)$.

**Exercise 5.29.** Construct a monotone function with jump discontinuities exactly at the points of a given countable subset of $(0, 1)$.

**Exercise 5.30.** Prove that if $f$ is continuous and injective on an interval, then $f$ is strictly monotone.

Chapter 6

# Function Spaces, Approximation, and Power Series

*This chapter studies compactness and approximation in spaces of functions. The first part develops equicontinuity and the Arzelà–Ascoli theorem, which identify compactness in function spaces through uniform control rather than finite-dimensional geometry. The second part treats semicontinuity and approximation results, including the Weierstrass approximation theorem, and the chapter concludes with a systematic treatment of power series as analytic objects governed by their radius of convergence.*

**Learning objectives.**

- Control families of functions by uniform boundedness, equicontinuity, and compactness in function spaces.
- Distinguish pointwise from uniform phenomena and understand which analytic properties pass to the limit.
- Use Arzelà–Ascoli, approximation arguments, and power-series methods as a coherent package.

## 6.1 Compactness in Spaces of Functions

**Section overview.**

- Compactness in $C(X)$ is subtler than compactness in Euclidean space because one must control entire families of functions at once.
- Uniform boundedness and equicontinuity are the correct replacements for ordinary boundedness.
- Arzelà–Ascoli should be read as a compactness criterion and as a blueprint for extracting convergent subsequences.

### 6.1.1 Equicontinuity

A collection $\mathcal{F} \subset C(X)$ is said to be (uniformly) equicontinuous if for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$d(x, y) < \delta \implies |f(x) - f(y)| < \varepsilon, \quad \forall f \in \mathcal{F}.$$

**Example 6.1.**

(i) Finite subset of $C(X)$ is (uniformly) equicontinuous and every sub-collection of a (uniformly) equicontinuous collection is equicontinuous.

(ii) Let $0 < \alpha \le 1$ and $k > 0$. Define

$$\text{Lip}_K^\alpha = \{f \in C([0,1]) : |f(x) - f(y)| \le k|x-y|^\alpha\}$$

This collection is equicontinuous, but not totally bounded, since all constant functions are satisfying this condition.

**Lemma 6.2** (Total boundedness implies uniform boundedness and equicontinuity)**.** *If $\mathcal{F} \subset C(X)$ is totally bounded, then $\mathcal{F}$ is uniformly bounded and (uniformly) equicontinuous.*

*Proof.* Since a totally bounded set is (uniformly) bounded, we only need to show that $\mathcal{F}$ is equicontinuous. Since $\mathcal{F}$ is totally bounded, for $\varepsilon > 0$, there exists $f_1, \ldots, f_n \in \mathcal{F}$ such that for $f \in \mathcal{F}$, there exists $f_i$ with $\|f - f_i\|_\infty < \varepsilon$. But $\{f_1, \cdots, f_n\}$ is equicontinuous, so for $\varepsilon > 0$, there exists $\delta > 0$ such that

$$d(x,y) < \delta \implies |f_i(x) - f_i(y)| < \varepsilon, \quad \forall i = 1, \ldots, n.$$

Now, for any $f \in \mathcal{F}$,

$$|f(x) - f(y)| \le |f(x) - f_i(x)| + |f_i(x) - f_i(y)| + |f_i(y) - f(y)| \le \varepsilon + \varepsilon + \varepsilon = 3\varepsilon.$$

$\square$

**Corollary 6.3** (Uniformly convergent sequences in $C(X)$ are equicontinuous)**.** *If $f_n \xrightarrow{unif} f$ in $C(X)$, then $\{f_n\}$ is uniformly bounded and (uniformly) equicontinuous.*

*Proof.* Since $f_n \to f$ uniformly, the set $\{f\} \cup \{f_n : n \in \mathbb{N}\}$ is compact in $C(X)$. Hence it is totally bounded. By the preceding lemma, every totally bounded subset of $C(X)$ is uniformly bounded and uniformly equicontinuous. In particular, $\{f_n : n \in \mathbb{N}\}$ has both properties. $\square$

### 6.1.2 Arzelà-Ascoli Theorem

**Theorem 6.4** (Arzelà–Ascoli)**.** *Let $X$ be a compact metric space, and let $\mathcal{F} \subset C(X)$. Then $\mathcal{F}$ is compact in $(C(X), \|\cdot\|_\infty)$ if and only if $\mathcal{F}$ is closed, uniformly bounded, and uniformly equicontinuous.*

**Proof strategy.** The forward direction packages compactness into total boundedness and hence into uniform boundedness and equicontinuity. For the converse, one proves that every sequence in $\mathcal{F}$ admits a uniformly Cauchy subsequence by combining finitely many point evaluations with equicontinuity on a finite $\delta$-net of the compact space.

*Proof.* The forward implication follows from the previous lemma.

Conversely, let $(f_n) \subset \mathcal{F}$ be a sequence. We show that it has a uniformly convergent subsequence. Fix $\varepsilon > 0$. Since $\mathcal{F}$ is uniformly equicontinuous, there exists $\delta > 0$ such that

$$d(x,y) < \delta \implies |f_n(x) - f_n(y)| < \varepsilon \qquad \text{for all } n \in \mathbb{N}.$$

Because $X$ is compact, it is totally bounded, so there exist points $x_1, \dots, x_k \in X$ such that

$$X = \bigcup_{i=1}^{k} B_\delta(x_i).$$

For each fixed $i$, the sequence $(f_n(x_i))$ is bounded, since $\mathcal{F}$ is uniformly bounded. By a diagonal argument, there exists a subsequence $(f_{n_j})$ such that, for every $i = 1, \dots, k$, the numerical sequence $(f_{n_j}(x_i))_j$ converges. In particular, there exists $J \in \mathbb{N}$ such that

$$|f_{n_p}(x_i) - f_{n_q}(x_i)| < \varepsilon$$

for all $p, q \geq J$ and all $i = 1, \dots, k$.

Now let $x \in X$. Choose $x_i$ with $d(x, x_i) < \delta$. Then for $p, q \geq J$,

$$\begin{aligned} |f_{n_p}(x) - f_{n_q}(x)| &\leq |f_{n_p}(x) - f_{n_p}(x_i)| + |f_{n_p}(x_i) - f_{n_q}(x_i)| + |f_{n_q}(x_i) - f_{n_q}(x)| \\ &< 3\varepsilon. \end{aligned}$$

Taking the supremum over $x \in X$, we obtain

$$\|f_{n_p} - f_{n_q}\|_\infty < 3\varepsilon \qquad (p, q \geq J).$$

Thus $(f_{n_j})$ is Cauchy in $(C(X), \|\cdot\|_\infty)$, and since $C(X)$ is complete, it converges uniformly to some $f \in C(X)$. Because $\mathcal{F}$ is closed, we have $f \in \mathcal{F}$. Therefore every sequence in $\mathcal{F}$ has a convergent subsequence, so $\mathcal{F}$ is compact. □

**Takeaway.** Arzelà–Ascoli is the function-space analogue of Bolzano–Weierstrass: compactness is recovered not from pointwise boundedness alone, but from a simultaneous control of size and oscillation.

**Corollary 6.5** (Sequential compactness form of Arzelà–Ascoli)**.** *Let $X$ be compact. If $(f_n)$ is uniformly bounded and (uniformly) equicontinuous in $C(X)$, then $(f_n)$ has a uniformly convergent subsequence. (Hint: $A = \overline{\{f_n : n \in \mathbb{N}\}}$ is closed.)*

**Example 6.6.** Let $X = (0, 1)$ and define

$$f_n(t) = \begin{cases} 1 - nt & \text{if } t < \frac{1}{n} \\ 0 & \text{if } t \geq \frac{1}{n} \end{cases}$$

Show that $(f_n)_{n=1}^{\infty}$ is pointwise equicontinuous but not uniformly equicontinuous on $(0, 1)$.

Fix $t \in (0, 1)$. Choose $n_0 \in \mathbb{N}$ so large that $1/n_0 < t/2$. Then for every $n \geq n_0$, the support of $f_n$ is contained in $(0, 1/n)$ and therefore does not meet a sufficiently small neighborhood of $t$. Hence $f_n = 0$ on that neighborhood for all $n \geq n_0$. Since the finite family $\{f_1, \dots, f_{n_0-1}\}$ is equicontinuous at $t$, it follows that $(f_n)$ is pointwise equicontinuous at $t$. As $t$ was arbitrary, $(f_n)$ is pointwise equicontinuous on $(0, 1)$.

However,
$$\left|f_n\left(\frac{1}{2n}\right) - f_n\left(\frac{1}{n}\right)\right| = \left|\frac{1}{2} - 0\right| = \frac{1}{2},$$
while
$$\left|\frac{1}{2n} - \frac{1}{n}\right| = \frac{1}{2n} \to 0.$$
Therefore $(f_n)$ is not uniformly equicontinuous on $(0,1)$.

**Example 6.7.** For $X = [0,1]$, define
$$f_n(t) = \max\left\{1 - 2(n+1)^2 \left|t - \frac{1}{n}\right|, 0\right\}.$$
Then $(f_n)_{n\geq 1}$ is equicontinuous at each point $t > 0$, but not at $t = 0$.

Fix $t_0 > 0$. Choose $n_0 \in \mathbb{N}$ so large that
$$\frac{1}{n} + \frac{1}{2(n+1)^2} < \frac{t_0}{2} \qquad (n \geq n_0).$$
Then for every $n \geq n_0$, the support of $f_n$ lies in a neighborhood of $1/n$ contained in $(0, t_0/2)$, and therefore misses some neighborhood of $t_0$. Consequently $f_n = 0$ near $t_0$ for all $n \geq n_0$. Since the finite family $\{f_1, \ldots, f_{n_0-1}\}$ is equicontinuous at $t_0$, the whole sequence $(f_n)$ is equicontinuous at $t_0$. Thus $(f_n)$ is pointwise equicontinuous at every point $t_0 > 0$.

At $t = 0$, however, we have $f_n(0) = 0$ and $f_n(1/n) = 1$, while $1/n \to 0$. Hence
$$|f_n(0) - f_n(1/n)| = 1 \qquad (n \geq 1),$$
so $(f_n)$ is not pointwise equicontinuous at 0.

*Remark* 6.8. We end this section with a remark on a structural property of subsets of the real line. Any open set can be written as a countable union of pairwise disjoint open intervals; moreover, a bounded (equivalently, totally bounded) set can be covered by finitely many intervals of arbitrarily small length.

*Remark* 6.9. A closer look at totally bounded sets reveals that many properties verified on finitely many centers can be propagated to the whole space, since every point lies in a ball of arbitrarily small radius.

### 6.1.3 Dini's Theorem

**Theorem 6.10** (Dini's theorem for decreasing sequences)**.** *Let $X$ be a compact metric space, and $f, f_n \in C(X)$ such that $f_n \downarrow f$ pointwise on $X$. Then $f_n \downarrow f$ uniformly on $X$.*

*Proof.* Let $g_n = f_n - f$. Then $g_n \downarrow 0$ pointwise on $X$. Notice that for each $\epsilon > 0$, $|g_n(x)| < \epsilon$ for large $n$ (depending upon $x$). Let
$$E_n = \{x \in X : g_n(x) < \epsilon\}.$$

Then $E_n = g_n^{-1}(-\infty, \epsilon)$, hence open. Also, $E_n \subset E_{n+1} \subset \cdots$. Since $g_n \downarrow 0$ at each point, it follows that $X = \bigcup_{n=1}^{\infty} E_n$. (If $x \in X$ and $x \notin E_n$ for all $n \in \mathbb{N}$, then $g_n(x) \geq \epsilon$ for all $n \in \mathbb{N}$, which is a contradiction.)

But $X$ is compact, hence there exists $N \in \mathbb{N}$ such that $X = \bigcup_{n=1}^{N} E_n = E_N$. Thus, for $x \in X$ and $n \geq N$, we have $g_n(x) \leq g_N(x) < \epsilon$, that is, $|g_n(x)| < \epsilon$ for all $n \geq N$ and all $x \in X$. Hence $g_n \downarrow 0$ uniformly on $X$. □

**Corollary 6.11** (Dini's theorem for increasing sequences)**.** *Suppose $f_n, f \in C(X)$ and $f_n \uparrow f$ pointwise, then $f_n \uparrow f$ uniformly. (*Hint. *$g_n = f - f_n \downarrow 0$ pointwise, so use the above argument.)*

1. Notice that the limit function $f$ must be continuous, else $f_n(x) = x^n$ will contradict the above theorem.
2. If $X$ is not compact, then the conclusion of the theorem might not be true.
   For $X = \mathbb{R}$,
   $$f_n(t) = \begin{cases} 0 & \text{if } -\infty < t \leq n \\ \frac{t}{n} - 1 & \text{if } n < t \leq 2n \\ 1 & \text{if } t > 2n \end{cases}$$
   $f_n \downarrow 0$ pointwise, but $\|f_n\|_\infty = 1$..

*Remark* 6.12. However, a pointwise convergent sequence can differ with uniform convergence on an arbitrarily small set (*Egoroff's Theorem*).

## 6.2 Semicontinuity and Approximation

**Section overview.**

- Semicontinuity captures one-sided stability properties that arise naturally in optimization and variational arguments.
- Approximation theorems show that complicated continuous functions can often be modeled uniformly by much simpler ones.

### 6.2.1 Upper Semicontinuity

Let $f : (X, d) \to \mathbb{R}$. Then $f$ is said to be *upper semicontinuous* on $X$ if for each $\alpha \in \mathbb{R}$, the set $\{x \in X : f(x) < \alpha\}$ is open.

**Proposition** (Sequential characterization of upper semicontinuity)**.** *Let $f : X \to \mathbb{R}$. Then $f$ is upper semicontinuous on $X$ if and only if for every $x \in X$ and every sequence $(x_n)$ with $x_n \to x$, one has*
$$\limsup_{n\to\infty} f(x_n) \leq f(x).$$

*Proof.* Assume first that $f$ is upper semicontinuous, and let $x_n \to x$. Fix $\epsilon > 0$. The set
$$U_\epsilon := \{y \in X : f(y) < f(x) + \epsilon\}$$

is open and contains $x$. Hence there exists $N \in \mathbb{N}$ such that $x_n \in U_\epsilon$ for all $n \geq N$. Therefore

$$f(x_n) < f(x) + \epsilon \qquad \text{for all } n \geq N,$$

which implies

$$\limsup_{n\to\infty} f(x_n) \leq f(x) + \epsilon.$$

Since $\epsilon > 0$ is arbitrary, we conclude that $\limsup_{n\to\infty} f(x_n) \leq f(x)$.

Conversely, suppose the sequential condition holds. Fix $\alpha \in \mathbb{R}$ and set

$$U_\alpha := \{x \in X : f(x) < \alpha\}.$$

To show that $f$ is upper semicontinuous, it is enough to prove that $U_\alpha$ is open. Let $x \in U_\alpha$. If $U_\alpha$ were not open at $x$, then for each $n \in \mathbb{N}$ there would exist $x_n \in B_{1/n}(x) \setminus U_\alpha$. Then $x_n \to x$, while

$$f(x_n) \geq \alpha > f(x),$$

so

$$\limsup_{n\to\infty} f(x_n) \geq \alpha > f(x),$$

contrary to the assumed sequential condition. Thus $U_\alpha$ is open, and $f$ is upper semicontinuous. □

**Theorem 6.13** (Maximum principle for upper semicontinuous functions)**.** *If $X$ is compact and $f : X \to \mathbb{R}$ is upper semicontinuous, then $f$ attains its maximum on $X$.*

*Proof.* For each $\alpha \in \mathbb{R}$, let $U_\alpha = \{x \in X : f(x) < \alpha\}$. Since $f$ is upper semicontinuous, each $U_\alpha$ is open, and clearly

$$X = \bigcup_{\alpha\in\mathbb{R}} U_\alpha.$$

By compactness, there exist $\alpha_1, \ldots, \alpha_k$ such that

$$X = \bigcup_{i=1}^{k} U_{\alpha_i}.$$

Hence $f(x) < \max\{\alpha_1, \ldots, \alpha_k\}$ for every $x \in X$, so $f$ is bounded above.

Let $M = \sup_X f$. Suppose, for contradiction, that $f(x) < M$ for every $x \in X$. Choose $x_n \in X$ such that

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

Since $X$ is compact, some subsequence $x_{n_k}$ converges to a point $x \in X$. Upper semicontinuity yields

$$M \leq \limsup_{k\to\infty} f(x_{n_k}) \leq f(x) \leq M.$$

Thus $f(x) = M$, contradicting the assumption that $f(x) < M$ for all $x \in X$. Therefore $f$ attains its maximum. □

Similarly, one defines *lower semicontinuity*: a function $f\colon X \to \mathbb{R}$ is lower semicontinuous if $\{x \in X : f(x) > \alpha\}$ is open for every $\alpha \in \mathbb{R}$. Equivalently,

$$f(x) \le \liminf_{n\to\infty} f(x_n)$$

whenever $x_n \to x$. Thus, $f$ is continuous if and only if it is both lower semicontinuous and upper semicontinuous.

*Remark* 6.14. If $f : X \to \mathbb{R}$ is upper semicontinuous, then $f^{-1}((-\infty, \alpha))$ is open for every $\alpha \in \mathbb{R}$. Equivalently, $f^{-1}([\alpha, \infty))$ is closed for every $\alpha \in \mathbb{R}$. One should *not* expect preimages of bounded intervals such as $(\beta, \alpha)$ or $[\beta, \alpha)$ to be open in general.

If $f$ is both lower semicontinuous and upper semicontinuous, then for every $\alpha < \beta$,

$$f^{-1}((\alpha, \beta)) = f^{-1}((-\infty, \beta)) \cap f^{-1}((\alpha, \infty)).$$

The right-hand side is an intersection of two open sets, so $f^{-1}((\alpha, \beta))$ is open. Hence $f$ is continuous.

*Remark* 6.15. There is, in general, no direct relation between lower or upper semicontinuity and the existence of one-sided limits.

**Example 6.16.**

$$f(x) = \begin{cases} \sin\frac{1}{x} & x \neq 0 \\ 1 & x = 0 \end{cases}$$

is upper semicontinuous, but none of left limit and right limit exists at $x = 0$.

**Exercise 6.17.** Check for lower semicontinuity and upper semicontinuity for $f(x) = \lfloor x \rfloor$, the greatest integer function.

### 6.2.2 Weierstrass Approximation Theorem

We shall see that polynomials are dense in $(C[a, b], \|\cdot\|_\infty)$ if $b - a < \infty$. As a consequence, $C[a, b]$ is a separable space. The question of density of polynomials in $C[a, b]$ can be transferred to $C[0, 1]$ with the help of the map:

$$f(t) = \frac{t-a}{b-a}$$

For $f \in C[0, 1]$ and $n = 0, 1, 2, \ldots$, define (Bernstein polynomial):

$$B_n(f)(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k}$$

Then $B_n(f)$ is a polynomial of degree at most $n$. Here, $B_n(f)$ is known as *Bernstein polynomial.* In fact, we have

$$B_n(f)(0) = f(0), \qquad B_n(f)(1) = f(1)$$

Let us write $f_n(x) = x^n$ for $n = 0, 1, 2, \ldots$. The following combinatorial lemma is the key step in proving the density of $B_n(f)$ in $C[0,1]$.

**Lemma 6.18** (Basic Bernstein polynomial identities)**.**

*(i)* $B_n(f_0) = f_0$ *and* $B_n(f_1) = f_1$.

*(ii)* $B_n(f_2) = \left(1 - \frac{1}{n}\right) f_2 + \frac{1}{n} f_1$, *hence* $B_n(f_2) \to f_2$ *uniformly.*

*(iii)* $\sum_{k=0}^{n} \left(\frac{k}{n} - x\right)^2 \binom{n}{k} x^k (1-x)^{n-k} = \frac{x(1-x)}{n} \le \frac{1}{4n}$.

*(iv)* *Given* $\delta > 0$, $0 \le x \le 1$, *let* $F$ *denote the set of* $F = \left\{k \in \{0, 1, \ldots, n\} : \left|\frac{k}{n} - x\right| \ge \delta\right\}$. *Then*

$$\sum_{k \in F} \binom{n}{k} x^k (1-x)^{n-k} \le \frac{1}{4n\delta^2}$$

*Proof.* (i) is trivial, as it follows from simple binomial expansions. *Hint.*

$$\sum_{k=0}^{n} \frac{k}{n} \binom{n}{k} x^k (1-x)^{n-k} = x \sum_{k=1}^{n} \binom{n-1}{k-1} x^{k-1} (1-x)^{n-k}$$

$$= x \sum_{j=0}^{n-1} \binom{n-1}{j} x^j (1-x)^{(n-1)-j} = x[x + (1-x)]^{n-1} = x$$

So $B_n(f_1) = f_1$.

(ii) To compute $B_n(f_2)$, we break the sum into two parts:

$$\left(\frac{k}{n}\right)^2 \binom{n}{k} = \frac{k}{n} \binom{n-1}{k-1} = (1 - \frac{1}{n}) \binom{n-2}{k-2} + \frac{1}{n} \binom{n-1}{k-1} \quad \text{for } k \ge 2$$

Thus,

$$B_n(f_2) = \left(1 - \frac{1}{n}\right) \sum_{k=2}^{n} \binom{n-2}{k-2} x^k (1-x)^{n-k} + \frac{1}{n} \sum_{k=1}^{n} \binom{n-1}{k-1} x^k (1-x)^{n-k}$$

$$= \left(1 - \frac{1}{n}\right) x^2 + \frac{1}{n} x \to f_2 \text{ uniformly}$$

(iii) Note that

$$\left(\frac{k}{n} - x\right)^2 = \left(\frac{k}{n}\right)^2 - 2x\frac{k}{n} + x^2$$

Hence,

$$\sum_{k=0}^{n} \left(\frac{k}{n} - x\right)^2 \binom{n}{k} x^k (1-x)^{n-k} = (1 - \frac{1}{n}) x^2 + \frac{1}{n} x - 2x^2 + x^2 = \frac{x(1-x)}{n} \le \frac{1}{4n} \quad (\text{ by (ii)})$$

(iv) For $k \in F$, $1 \leq \frac{(\frac{k}{n}-x)^2}{\delta^2}$. Hence,

$$\sum_{k\in F}\binom{n}{k}x^k(1-x)^{n-k} \leq \frac{1}{\delta^2}\sum_{k\in F}\left(\frac{k}{n}-x\right)^2\binom{n}{k}x^k(1-x)^{n-k}$$
$$\leq \frac{1}{\delta^2}\sum_{k=0}^{n}\left(\frac{k}{n}-x\right)^2\binom{n}{k}x^k(1-x)^{n-k} \leq \frac{1}{4n\delta^2}.$$

□

**Theorem 6.19** (Bernstein)**.** *Let $f \in C[0,1]$. Then $B_n(f) \to f$ uniformly on $[0,1]$.*

*Proof.* Since $f$ is uniformly continuous, for $\varepsilon > 0$ there exists $\delta > 0$ such that $|x-y| < \delta \implies |f(x)-f(y)| < \frac{\varepsilon}{2}$. Now,

$$\begin{aligned}|f(x) - B_n(f)(x)| &= \left|\sum_{k=0}^{n}\left(f(x) - f\left(\frac{k}{n}\right)\right)\binom{n}{k}x^k(1-x)^{n-k}\right| \\ &\leq \sum_{k=0}^{n}\left|f(x) - f\left(\frac{k}{n}\right)\right|\binom{n}{k}x^k(1-x)^{n-k},\end{aligned}$$

Fix an integer $n$ (to be specified shortly). Let $F$ denote the set of $k \in \{0,1,\ldots,n\}$ such that $\left|\frac{k}{n}-x\right| \geq \delta$. Then

$$|f(x) - f\left(\frac{k}{n}\right)| < \frac{\varepsilon}{2} \quad \text{for } k \notin F,$$

and

$$|f(x) - f\left(\frac{k}{n}\right)| \leq 2\|f\|_\infty \quad \text{for } k \in F.$$

Thus,

$$\begin{aligned}|f(x) - B_n(f)(x)| &\leq \frac{\varepsilon}{2}\sum_{k\notin F}\binom{n}{k}x^k(1-x)^{n-k} + 2\|f\|_\infty\sum_{k\in F}\binom{n}{k}x^k(1-x)^{n-k} \\ &\leq \frac{\varepsilon}{2}\cdot 1 + 2\|f\|_\infty\left(\frac{1}{4n\delta^2}\right) \\ &< \varepsilon\end{aligned}$$

provided $n > \frac{\|f\|_\infty}{\varepsilon\delta^2}$. Therefore,

$$\|B_n(f) - f\|_\infty < \varepsilon \quad \text{whenever} \quad n > \frac{\|f\|_\infty}{\varepsilon\delta^2}.$$

□

**Exercise 6.20.** If $f \in C[0,1]$ and $\int_0^1 x^n f(x)\,dx = 0$ for all $n \geq 0$, then $f = 0$.

## 6.3 Power Series

**Section overview.**

- Power series provide local analytic models and a flexible source of uniformly convergent series of functions.
- The radius of convergence separates formal algebraic manipulations from statements that are analytically valid.

Power series provide a canonical source of uniformly convergent series of functions and connect real analysis with elementary complex-analytic ideas. We record the basic convergence theory and the standard termwise operations.

**Definition 6.21.** A *power series* centered at $x_0 \in \mathbb{R}$ is a series of the form

$$\sum_{n=0}^{\infty} a_n (x - x_0)^n, \tag{6.1}$$

where $(a_n)_{n \geq 0} \subset \mathbb{R}$ (or $\mathbb{C}$).

**Theorem 6.22** (Radius of convergence)**.** *For the power series* (6.1) *there exists* $R \in [0, \infty]$ *such that:*

*(i) the series converges absolutely for all* $x$ *with* $|x - x_0| < R$*;*

*(ii) the series diverges for all* $x$ *with* $|x - x_0| > R$*.*

*The number* $R$ *is called the* radius of convergence*. Moreover,*

$$\frac{1}{R} = \limsup_{n \to \infty} |a_n|^{1/n} \tag{6.2}$$

*(with the convention* $1/\infty = 0$ *and* $1/0 = \infty$*).*

*Proof.* Set $\rho := \limsup_{n \to \infty} |a_n|^{1/n} \in [0, \infty]$ and define $R$ by $1/R = \rho$. Fix $x$ with $|x - x_0| < R$. Choose $r$ with $|x - x_0| < r < R$. Then $\rho < 1/r$, hence there exists $N$ such that $|a_n|^{1/n} \leq 1/r$ for all $n \geq N$. Thus $|a_n(x - x_0)^n| \leq (|x - x_0|/r)^n$ for $n \geq N$, and the tail is dominated by a geometric series; hence convergence is absolute.

If $|x - x_0| > R$, choose $r$ with $R < r < |x - x_0|$. Then $\rho > 1/r$, hence for infinitely many $n$ we have $|a_n|^{1/n} \geq 1/r$, that is $|a_n(x - x_0)^n| \geq (|x - x_0|/r)^n$, which does not tend to 0. Therefore the series diverges. □

**Theorem 6.23** (Uniform convergence on compact subintervals)**.** *Let* $\sum_{n=0}^{\infty} a_n (x - x_0)^n$ *have radius of convergence* $R > 0$*. Then for every* $0 < r < R$ *the series converges uniformly on the closed interval*

$$[x_0 - r,\ x_0 + r].$$

*Proof.* Fix $r \in (0, R)$. By the definition of $R$ there exists $M > 0$ such that $|a_n| r^n \leq M$ for all $n$ large enough. For $x \in [x_0 - r, x_0 + r]$ we have $|x - x_0| \leq r$, hence $|a_n(x - x_0)^n| \leq |a_n| r^n$.

Since $\sum |a_n| r^n$ converges (as $r < R$), the Weierstrass M-test implies uniform convergence on $[x_0 - r, x_0 + r]$. $\square$

**Theorem 6.24** (Termwise differentiation and integration). *Let $\sum_{n=0}^{\infty} a_n(x-x_0)^n$ have radius of convergence $R > 0$, and define*

$$f(x) := \sum_{n=0}^{\infty} a_n(x-x_0)^n \qquad (|x-x_0| < R).$$

*Then $f$ is differentiable on $(x_0 - R, x_0 + R)$ and*

$$f'(x) = \sum_{n=1}^{\infty} n a_n (x-x_0)^{n-1}, \qquad |x-x_0| < R. \tag{6.3}$$

*Moreover, for every $x$ with $|x - x_0| < R$ we have*

$$\int_{x_0}^{x} f(t)\,dt = \sum_{n=0}^{\infty} a_n \frac{(x-x_0)^{n+1}}{n+1}. \tag{6.4}$$

*Both derived series have the same radius of convergence $R$.*

*Proof.* Fix $r \in (0, R)$ and work on the closed interval $I = [x_0 - r, x_0 + r]$. By theorem 6.23, the original series converges uniformly on $I$. The differentiated series in (6.3) also converges uniformly on $I$: indeed, since $\sum |a_n| r^n$ converges, the series $\sum n|a_n|r^{n-1}$ converges by a standard comparison (e.g. ratio test), and hence the M-test applies to $\sum n a_n (x-x_0)^{n-1}$ on $I$. Let $S_N(x) = \sum_{n=0}^{N} a_n(x-x_0)^n$. Then $S_N$ is differentiable and $S_N'(x) = \sum_{n=1}^{N} n a_n (x-x_0)^{n-1}$. Uniform convergence of $S_N'$ on $I$ together with pointwise convergence of $S_N(x_0)$ implies that $(S_N)$ converges uniformly to a differentiable limit $f$ on $I$ and that $f' = \lim S_N'$ on $I$. Since $r < R$ is arbitrary, the formula holds for all $|x - x_0| < R$. The integral formula (6.4) follows by integrating the partial sums and passing to the limit using uniform convergence. $\square$

**Example 6.25.** The exponential series $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ has infinite radius of convergence and defines a $C^\infty$ function on $\mathbb{R}$. Moreover, it satisfies $f' = f$ and $f(0) = 1$, hence it coincides with the exponential function $e^x$.

## Exercises

**Exercise 6.26.** Let $(f_n)$ be a sequence of continuous functions on $[a, b]$ that converges uniformly to $f$. Prove that $f$ is continuous.

**Exercise 6.27.** Give an example of pointwise convergence $f_n \to f$ on $[0, 1]$ where $f$ is discontinuous.

**Exercise 6.28.** Prove that if $f_n \to f$ uniformly and each $f_n$ is bounded, then $\sup |f_n - f| \to 0$.

**Exercise 6.29.** State and prove the Weierstrass M-test, and use it to prove uniform convergence of a power series inside its radius of convergence.

**Exercise 6.30.** Prove the Arzelà–Ascoli theorem for a uniformly bounded, equicontinuous family in $C([a, b])$.

**Exercise 6.31.** Let $f_n$ be differentiable on $[a, b]$, assume $f_n(x_0)$ converges for some $x_0$ and $f'_n$ converges uniformly. Prove that $f_n$ converges uniformly and that $(\lim f_n)' = \lim f'_n$.

**Exercise 6.32.** Prove that every continuous function on $[0, 1]$ can be uniformly approximated by polynomials (Weierstrass approximation theorem).

**Exercise 6.33.** For a power series $\sum a_n(x - a)^n$ with radius $R > 0$, prove that it converges uniformly on $[a - r, a + r]$ for each $0 < r < R$.

**Exercise 6.34.** Compute the radius of convergence of $\sum_{n \geq 1} \frac{n!}{n^n}(x - 2)^n$.

**Exercise 6.35.** Let $\sum a_n x^n$ and $\sum b_n x^n$ have radii $R_a$ and $R_b$. Prove that the Cauchy product has radius at least $\min(R_a, R_b)$.

## Chapter 7

# Differential Calculus in One Variable

*This chapter develops the differential calculus of one real variable from the viewpoint of first-order approximation. After reviewing differentiability and the standard algebraic rules, we prove Rolle's theorem and the mean value theorems and derive their principal consequences, including monotonicity criteria, comparison estimates, and convexity tests. The chapter culminates in Taylor's theorem with remainder, which formalises local polynomial approximation and supplies one of the principal computational and conceptual tools of classical analysis.*

**Learning objectives.**

- Treat differentiability as first-order approximation and use it to derive qualitative global information.
- Apply Rolle's theorem, the mean value theorems, and Taylor's theorem with full control of their hypotheses.
- Convert derivative information into monotonicity, Lipschitz estimates, and approximation results.

## 7.1 Differentiability

**Section overview.**

- Differentiability is stronger than continuity because it asserts the existence of a linear model at small scales.
- The goal of this section is to identify the exact approximation statement hidden behind the derivative.
- Later results in the chapter should be read as consequences of this linear approximation viewpoint.

**Definition 7.1.** Let $I \subseteq \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$. We say that $f$ is *differentiable at* $x_0 \in I$ if the limit

$$f'(x_0) := \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} \tag{7.1}$$

exists (finite). If $f$ is differentiable at every $x \in I$, then $f$ is *differentiable on* $I$.

**Proposition 7.2** (Differentiability implies continuity)**.** *If $f$ is differentiable at $x_0$, then $f$ is continuous at $x_0$.*

*Proof.* Assume (7.1) holds. Write

$$f(x_0 + h) - f(x_0) = h \cdot \frac{f(x_0 + h) - f(x_0)}{h}.$$

As $h \to 0$, the quotient tends to $f'(x_0)$, hence is bounded near 0. Therefore the product tends to 0, that is $f(x_0 + h) \to f(x_0)$. □

*Remark* 7.3. Differentiability is precisely the existence of a first-order approximation: $f$ is differentiable at $x_0$ if and only if there exists $m \in \mathbb{R}$ and a function $\eta(h)$ with $\eta(h) \to 0$ as $h \to 0$ such that

$$f(x_0 + h) = f(x_0) + mh + h\eta(h). \tag{7.2}$$

In that case $m = f'(x_0)$.

## 7.2 Mean Value Theorems

**Section overview.**

- The mean value theorems translate local derivative information into global conclusions on an interval.
- They are the fundamental mechanism behind monotonicity tests, error estimates, and Taylor expansions.

**Theorem 7.4** (Rolle)**.** *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on $(a, b)$. If $f(a) = f(b)$, then there exists $c \in (a, b)$ such that $f'(c) = 0$.*

*Proof.* By the extreme value theorem, $f$ attains a maximum and a minimum on $[a, b]$. If both extrema occur at the endpoints, then $f$ is constant and $f' \equiv 0$ on $(a, b)$. Otherwise, at least one of the extrema occurs at some interior point $c \in (a, b)$. At an interior maximum or minimum, the difference quotients from the left and right have opposite signs, which forces $f'(c) = 0$. □

**Theorem 7.5** (Mean Value Theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on $(a, b)$. Then there exists $c \in (a, b)$ such that*

$$f'(c) = \frac{f(b) - f(a)}{b - a}. \tag{7.3}$$

**Idea.** Subtract the secant line joining $(a, f(a))$ and $(b, f(b))$. The resulting function has equal endpoint values, so Rolle's theorem turns the mean slope statement into a zero-derivative statement.

*Proof.* Consider the auxiliary function

$$g(x) := f(x) - \ell(x), \qquad \ell(x) := f(a) + \frac{f(b) - f(a)}{b - a}(x - a).$$

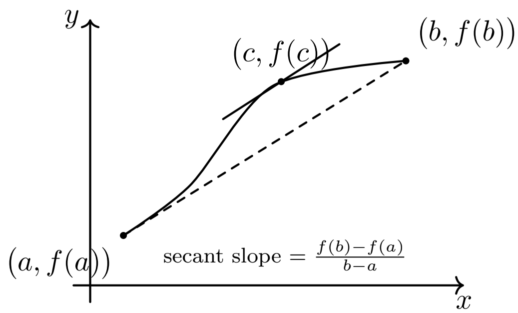



**Figure 7.1.** Geometric content of the Mean Value Theorem: some tangent slope equals the secant slope.

Then $g$ is continuous on $[a, b]$, differentiable on $(a, b)$, and $g(a) = g(b) = 0$. By Rolle's theorem theorem 7.4, there exists $c \in (a, b)$ with $g'(c) = 0$, that is $f'(c) = \ell'(c)$, which is exactly (7.3). $\square$

**Theorem 7.6** (Cauchy Mean Value Theorem)**.** *Let $f, g : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on $(a, b)$. Then there exists $c \in (a, b)$ such that*

$$(f(b) - f(a))g'(c) = (g(b) - g(a))f'(c). \tag{7.4}$$

*If moreover $g'(x) \neq 0$ on $(a, b)$, then*

$$\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{g(b) - g(a)}. \tag{7.5}$$

*Proof.* Define

$$h(x) := (f(b) - f(a))\, g(x) - (g(b) - g(a))\, f(x).$$

Then $h$ is continuous on $[a, b]$, differentiable on $(a, b)$, and $h(a) = h(b)$. By Rolle's theorem theorem 7.4, there exists $c \in (a, b)$ such that $h'(c) = 0$, which is precisely (7.4). $\square$

## 7.3 Consequences of the MVT

**Section overview.**

- Once the mean value theorem is available, many structural consequences follow with very little additional effort.
- The emphasis here is on learning how derivative inequalities become geometric information about the function itself.

**Proposition 7.7** (Monotonicity test)**.** *Let $f : (a, b) \to \mathbb{R}$ be differentiable.*

*(i) If $f'(x) \geq 0$ for all $x \in (a, b)$, then $f$ is nondecreasing on $(a, b)$.*

*(ii) If $f'(x) > 0$ for all $x \in (a, b)$, then $f$ is strictly increasing on $(a, b)$.*

*Proof.* Fix $x < y$ in $(a, b)$. Apply the Mean Value Theorem theorem 7.5 to $f$ on $[x, y]$ to find $c \in (x, y)$ such that

$$f(y) - f(x) = f'(c)(y - x).$$

If $f' \geq 0$, then the right-hand side is $\geq 0$, hence $f(y) \geq f(x)$. If $f' > 0$, then $f(y) > f(x)$. □

**Proposition 7.8** (Lipschitz estimate)**.** *Let $f : [a, b] \to \mathbb{R}$ be differentiable and suppose $|f'(x)| \leq M$ on $(a, b)$. Then for all $x, y \in [a, b]$,*

$$|f(x) - f(y)| \leq M|x - y|. \tag{7.6}$$

*Proof.* Assume $x < y$. By theorem 7.5, there exists $c \in (x, y)$ with $f(y) - f(x) = f'(c)(y - x)$, hence $|f(y) - f(x)| \leq M|y - x|$. □

## 7.4 Taylor's theorem

**Section overview.**

- Taylor's theorem formalizes the idea that a smooth function is approximated by a polynomial at small scales.
- The remainder term measures precisely how much information is lost when the expansion is truncated.

**Theorem 7.9** (Taylor with Lagrange remainder)**.** *Let $n \in \mathbb{N}$ and let $f : [a, b] \to \mathbb{R}$ be $(n + 1)$ times continuously differentiable on $[a, b]$. Fix $x_0 \in [a, b]$. Then for every $x \in [a, b]$ there exists $\xi$ between $x$ and $x_0$ such that*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + \frac{f^{(n+1)}(\xi)}{(n + 1)!}(x - x_0)^{n+1}. \tag{7.7}$$

*Proof.* Fix $x \neq x_0$ and define the degree-$n$ Taylor polynomial at $x_0$:

$$P_n(t) := \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(t - x_0)^k.$$

Consider the function

$$\Phi(t) := f(t) - P_n(t) - \lambda(t - x_0)^{n+1},$$

where $\lambda$ is chosen so that $\Phi(x) = 0$. Since $\Phi(x_0) = 0$ as well, repeated application of Rolle's theorem yields a point $\xi$ between $x$ and $x_0$ with $\Phi^{(n+1)}(\xi) = 0$. But $P_n^{(n+1)} \equiv 0$ and $\frac{d^{n+1}}{dt^{n+1}}(t - x_0)^{n+1} = (n + 1)!$, hence

$$0 = \Phi^{(n+1)}(\xi) = f^{(n+1)}(\xi) - \lambda(n + 1)!,$$

so $\lambda = \frac{f^{(n+1)}(\xi)}{(n+1)!}$. Substituting the definition of $\lambda$ (from $\Phi(x) = 0$) gives (7.7). □

**Corollary 7.10** (Uniform Taylor error bound)**.** *Under the assumptions of theorem 7.9, if* $|f^{(n+1)}(t)| \le M$ *on* $[a, b]$*, then*

$$\left| f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k \right| \le \frac{M}{(n+1)!}|x - x_0|^{n+1}.$$

## Exercises

**Exercise 7.11.** Prove Rolle's theorem and deduce the Mean Value Theorem.

**Exercise 7.12.** Let $f$ be differentiable on $(a, b)$ and continuous on $[a, b]$. Show that if $f' = 0$ on $(a, b)$ then $f$ is constant on $[a, b]$.

**Exercise 7.13.** Show that if $f'$ exists and is bounded on $(a, b)$, then $f$ is Lipschitz on $[a, b]$.

**Exercise 7.14.** Prove that if $f' > 0$ on an interval then $f$ is strictly increasing there.

**Exercise 7.15.** Let $f$ be twice differentiable on $(a, b)$. Prove that $f'' \ge 0$ implies $f$ is convex.

**Exercise 7.16.** Prove Taylor's theorem with Lagrange remainder and use it to derive $\sin x < x$ for $x > 0$.

**Exercise 7.17.** Show that if $\sum a_n$ converges absolutely, then the series $\sum a_n x^n$ defines a differentiable function on $(-1, 1)$ with termwise differentiation.

**Exercise 7.18.** Give an example of a function differentiable everywhere whose derivative is unbounded on every interval.

**Exercise 7.19.** Prove that if $f$ is differentiable at $a$, then $f$ is continuous at $a$.

**Exercise 7.20.** Suppose $f$ is differentiable on $(a, b)$ and $f'$ is increasing. Prove that $f$ is convex.

Chapter 8

# Differential Calculus in Several Variables

*This chapter develops the analytic framework for functions of several variables in Euclidean space. We begin with limits and continuity in $\mathbb{R}^n$, then introduce partial derivatives, directional derivatives, and, crucially, differentiability in the Fréchet sense as the existence of a best linear approximation. The Jacobian and Hessian are developed as tools for local analysis, leading naturally to multivariable Taylor expansion and the study of local extrema. The chapter concludes with the inverse and implicit function theorems, which describe the local structure of smooth maps and level sets.*

**Learning objectives.**

- Understand differentiability in several variables as a linear approximation, not merely as the existence of partial derivatives.
- Use the Jacobian, Hessian, and Taylor expansion to analyze local geometry and local extrema.
- Master the conceptual content of the inverse and implicit function theorems as local structure theorems.

## 8.1 Orientation and Chapter Roadmap

This chapter develops the theory of functions of several variables in a logically ordered and essentially self-contained manner. The central viewpoint is that differentiability is *linear approximation with a remainder that is small relative to* $\|h\|$. The chapter develops this idea in a sequence that mirrors the syllabus:

(1) **Limits and continuity in $\mathbb{R}^n$:** the $\varepsilon$–$\delta$ definition and its sequential characterization, together with basic permanence properties.

(2) **Differentiability and the Fréchet derivative:** partial and directional derivatives, differentiability as the existence of a best linear approximation, and the chain rule.

(3) **Taylor's theorem and quantitative estimates:** higher-order expansions and remainder bounds that control approximation errors.

(4) **Local structure of smooth maps:** the inverse function theorem and implicit function theorem, with applications to local parametrizations and constraint manifolds.

## 8.2 Limits and Continuity

**Section overview.**

- In several variables, local behavior must be controlled simultaneously in all directions.
- The goal is to replace one-dimensional intuition with genuinely Euclidean notions of limit and continuity.

### 8.2.1 Notation and basic definitions in Euclidean space

Throughout this chapter we work in $\mathbb{R}^n$ with the standard inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle := \sum_{j=1}^{n} x_j y_j, \qquad \mathbf{x} = (x_1, \ldots, x_n),\ \mathbf{y} = (y_1, \ldots, y_n),$$

and the associated Euclidean norm

$$\|\mathbf{x}\| := \langle \mathbf{x}, \mathbf{x} \rangle^{1/2} = \left(x_1^2 + \cdots + x_n^2\right)^{1/2}. \tag{8.1}$$

We will use (8.1) as the default norm on $\mathbb{R}^n$ (and on $\mathbb{R}^m$ when needed). The Euclidean distance is

$$d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\|.$$

For $\mathbf{x} \in \mathbb{R}^n$ and $r > 0$, the open ball of radius $r$ centered at $\mathbf{x}$ is

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

**Conventions.** Vectors are written in boldface, and $\mathbf{x}_k \to \mathbf{x}$ always means $\|\mathbf{x}_k - \mathbf{x}\| \to 0$. When the meaning is clear from context we may suppress boldface.

We record two inequalities that will be used repeatedly.

**Theorem 8.1** (Cauchy–Schwarz inequality)**.** *For all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$*,*

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \le \|\mathbf{x}\| \, \|\mathbf{y}\|.$$

*Moreover, equality holds if and only if* $\mathbf{x}$ *and* $\mathbf{y}$ *are linearly dependent.*

*Proof.* If $\mathbf{y} = \mathbf{0}$ the claim is trivial. Otherwise, consider the quadratic polynomial

$$\begin{aligned} p(t) :&= \|\mathbf{x} - t\mathbf{y}\|^2 \\ &= \langle \mathbf{x} - t\mathbf{y}, \mathbf{x} - t\mathbf{y} \rangle \\ &= \|\mathbf{x}\|^2 - 2t\langle \mathbf{x}, \mathbf{y} \rangle + t^2\|\mathbf{y}\|^2. \end{aligned}$$

Since $p(t) \ge 0$ for all $t \in \mathbb{R}$, its discriminant is non-positive:

$$4\langle \mathbf{x}, \mathbf{y} \rangle^2 - 4\|\mathbf{x}\|^2\|\mathbf{y}\|^2 \le 0.$$

This is exactly $|\langle \mathbf{x}, \mathbf{y} \rangle| \le \|\mathbf{x}\| \, \|\mathbf{y}\|$. Equality holds if and only if $p(t)$ has a real root, i.e. $\mathbf{x} - t\mathbf{y} = \mathbf{0}$ for some $t$. $\square$

**Proposition 8.2** (Triangle inequality)**.** *For all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$*,*

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|.$$

*Proof.* By expanding the square and applying Cauchy–Schwarz,

$$\begin{aligned}\|\mathbf{x} + \mathbf{y}\|^2 &= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\langle \mathbf{x}, \mathbf{y} \rangle \\ &\leq (\|\mathbf{x}\| + \|\mathbf{y}\|)^2.\end{aligned}$$

Taking square roots gives the result. □

**Theorem 8.3** (Bolzano–Weierstrass in $\mathbb{R}^n$)**.** *Every bounded sequence in* $\mathbb{R}^n$ *has a convergent subsequence.*

*Proof.* Let $\mathbf{x}_k = (x_1^k, \ldots, x_n^k)$ be bounded. Then each coordinate sequence $(x_j^k)_{k \geq 1}$ is bounded in $\mathbb{R}$. Apply the one-dimensional Bolzano–Weierstrass theorem to $(x_1^k)$ to obtain a subsequence along which the first coordinate converges. Restrict to that subsequence and repeat for the second coordinate, and so on. After $n$ steps we obtain a subsequence $(\mathbf{x}_{k_\ell})$ such that each coordinate converges; hence $\mathbf{x}_{k_\ell} \to \mathbf{x} \in \mathbb{R}^n$. □

### 8.2.2 Limits in Euclidean space

The definition of a limit in $\mathbb{R}^n$ uses the Euclidean distance.

**Definition.** Let $D \subseteq \mathbb{R}^n$ and let $f : D \to \mathbb{R}^m$. Given $x_0 \in \mathbb{R}^n$ that is a limit point of $D$, we write

$$\lim_{x \to x_0} f(x) = L$$

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

$$0 < \|x - x_0\| < \delta \ \text{ and } \ x \in D \quad \Longrightarrow \quad \|f(x) - L\| < \varepsilon.$$

**Sequential criterion.** The limit equals $L$ if and only if for every sequence $(x_k) \subset D$ with $x_k \to x_0$ and $x_k \neq x_0$, one has $f(x_k) \to L$. We will freely use either formulation, depending on convenience.

**Paths and polar coordinates in $\mathbb{R}^2$.** When $n = 2$ and $x_0 = (0, 0)$, a common way to *disprove* the existence of a limit is to find two paths approaching $(0, 0)$ along which $f$ approaches different values. A common way to *compute* limits is to use polar coordinates $x = r\cos\theta$, $y = r\sin\theta$; then $(x, y) \to (0, 0)$ is equivalent to $r \to 0^+$. In many examples, showing that $\lim_{r \to 0^+} f(r\cos\theta, r\sin\theta)$ exists and is independent of $\theta$ suffices.

**A caution about the path test.** Agreement along a few special curves never *proves* the existence of a multivariable limit; it only rules out some obvious obstructions. To establish existence, one needs an estimate that works for *all* nearby points—for example, a bound in terms of $\sqrt{x^2+y^2}$, or a representation in polar coordinates whose absolute value tends to 0 uniformly in $\theta$.

### 8.2.3 Continuity in Euclidean space

Let $D \subseteq \mathbb{R}^n$ and $f : D \to \mathbb{R}^m$.

**Definition 8.4.** The function $f$ is *continuous at* $x_0 \in D$ if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\|x - x_0\| < \delta \ \text{ and } \ x \in D \quad \Longrightarrow \quad \|f(x) - f(x_0)\| < \varepsilon.$$

Equivalently, $f$ is continuous at $x_0$ if and only if $\lim_{x \to x_0} f(x) = f(x_0)$.

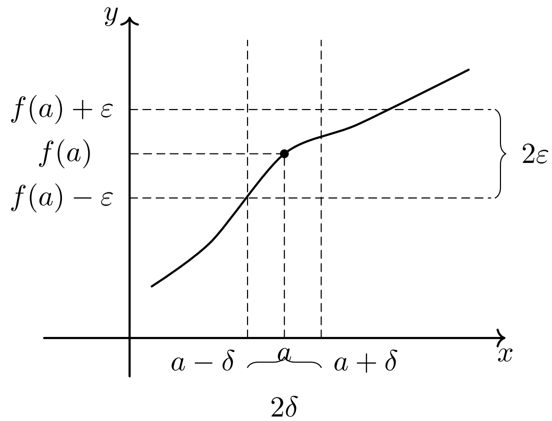



**Figure 8.1.** Geometric interpretation of the $\varepsilon$–$\delta$ definition of continuity at $a$.

**Sequential criterion.** The function $f$ is continuous at $x_0$ if and only if for every sequence $(x_k) \subset D$ with $x_k \to x_0$ one has $f(x_k) \to f(x_0)$.

**Componentwise continuity.** If $f = (f_1, \dots, f_m) : D \to \mathbb{R}^m$, then $f$ is continuous at $x_0$ if and only if each component $f_i$ is continuous at $x_0$. Indeed, $\|f(x) - f(x_0)\| < \varepsilon$ implies $|f_i(x) - f_i(x_0)| \le \|f(x) - f(x_0)\| < \varepsilon$, while conversely one can estimate

$$\|f(x) - f(x_0)\|^2 = \sum_{i=1}^{m} |f_i(x) - f_i(x_0)|^2.$$

**Negation of continuity.** The function $f$ fails to be continuous at $x_0$ if and only if there exists $\varepsilon_0 > 0$ such that for every $\delta > 0$ there exists $x \in D$ with $\|x - x_0\| < \delta$ but $\|f(x) - f(x_0)\| \ge \varepsilon_0$.

**Example 8.5.** Define

$$f(x,y) = \begin{cases} 1, & xy \neq 0, \\ 0, & xy = 0. \end{cases}$$

Then $\lim_{(x,y)\to(0,0)} f(x,y)$ does not exist, because along the path $y=0$ the function is identically 0, whereas along the path $y=x$ it is identically 1 for $x\neq 0$. By contrast, if one defines

$$g(x,y)=\begin{cases}1, & xy=1,\\ 0, & xy\neq 1,\end{cases}$$

then $\lim_{(x,y)\to(0,0)} g(x,y)=0$, since the condition $xy=1$ cannot hold near $(0,0)$.

**Exercise 8.6.** Let $f:\mathbb{R}^2\to\mathbb{R}$, check the continuity of $f$ at $(0,0)$.

1. $f(x,y)=\begin{cases}\dfrac{xy}{\sqrt{x^2+y^2}} & \text{if } x^2+y^2\neq 0\\ 0 & \text{otherwise}\end{cases}$
2. $f(x,y)=\dfrac{\sin^2(x-y)}{\sqrt{x^2+y^2}}, \qquad f(0,0)=0.$
3. $f(x,y)=\begin{cases}\dfrac{x^2y}{x^2+y} & \text{if } x^2+y\neq 0\\ 0 & \text{otherwise}\end{cases}$
4. $f(x,y)=\begin{cases}\dfrac{x^2y}{x^4+y^2} & \text{if } x^4+y^2\neq 0\\ 0 & \text{otherwise}\end{cases}$
5. $f(x,y)=\begin{cases}\dfrac{\sin xy}{xy} & \text{if } xy\neq 0\\ 0 & \text{otherwise}\end{cases}$

*Remark* 8.7 (A typical $\varepsilon$–$\delta$ negation). Consider the function

$$f(x,y):=\begin{cases}\dfrac{xy}{x^2+y^2}, & (x,y)\neq(0,0),\\ 0, & (x,y)=(0,0).\end{cases}$$

Along the diagonal $y=x$ we have $f(x,x)=\frac{1}{2}$ for all $x\neq 0$. Hence, for $\varepsilon_0:=\frac{1}{4}$ and for every

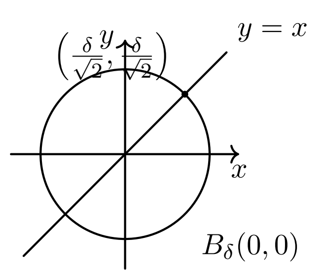



**Figure 8.2.** The diagonal path $y=x$ intersects every ball $B_\delta(0,0)$, which forces the negation estimate in Remark 8.7.

$\delta>0$, the point

$$(x_\delta,y_\delta):=\left(\frac{\delta}{\sqrt{2}},\frac{\delta}{\sqrt{2}}\right)$$

satisfies $\sqrt{x_\delta^2 + y_\delta^2} = \delta$ but

$$|f(x_\delta, y_\delta) - f(0,0)| = \frac{1}{2} \geq \varepsilon_0.$$

Therefore $f$ is not continuous at $(0,0)$, and consequently $\lim_{(x,y)\to(0,0)} f(x,y)$ does not exist.

**Proposition 8.8** (Composition preserves continuity)**.** *Let $D \subseteq \mathbb{R}^n$ and let $f\colon D \to \mathbb{R}^m$ be continuous at $x_0 \in D$. Let $I \subseteq \mathbb{R}^m$ and let $g\colon I \to \mathbb{R}^k$ be continuous at $f(x_0)$, and assume that $f(D) \subseteq I$. Then the composition $g \circ f\colon D \to \mathbb{R}^k$ is continuous at $x_0$.*

*Proof.* Fix $\varepsilon > 0$. Since $g$ is continuous at $f(x_0)$, there exists $\eta > 0$ such that

$$\|u - f(x_0)\| < \eta \implies \|g(u) - g(f(x_0))\| < \varepsilon, \qquad u \in I. \tag{8.2}$$

Since $f$ is continuous at $x_0$, there exists $\delta > 0$ such that

$$\|x - x_0\| < \delta \implies \|f(x) - f(x_0)\| < \eta, \qquad x \in D. \tag{8.3}$$

Combining (8.3) with (8.2) yields

$$\|x - x_0\| < \delta \implies \|g(f(x)) - g(f(x_0))\| < \varepsilon.$$

Hence $g \circ f$ is continuous at $x_0$. □

*Remark* 8.9 (Sequential proof). Alternatively, if $x_n \to x_0$ in $D$, then $f(x_n) \to f(x_0)$ by continuity of $f$, and hence $g(f(x_n)) \to g(f(x_0))$ by continuity of $g$.

**Example 8.10.**

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

$$f(x,y) = p \circ g(x,y), \quad \text{where } p(t) = \begin{cases} \dfrac{\sin t}{t} & t \neq 0 \\ 1 & t = 0 \end{cases}$$

## 8.3 Differentiation in $\mathbb{R}^n$

**Section overview.**

- Partial derivatives are useful coordinates, but differentiability is fundamentally the existence of a single linear map approximating the function.
- This section emphasizes how the Jacobian packages the first-order behavior of a multivariable map.
- Watch carefully which statements require only directional information and which require a genuinely uniform linear approximation.

### 8.3.1 Partial derivatives

Let $D \subseteq \mathbb{R}^2$ be open and $f : D \to \mathbb{R}$. Fix $(x_0, y_0) \in D$.

$$\frac{\partial f}{\partial x}(x_0, y_0) := \lim_{h \to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}, \tag{8.4}$$

provided the limit exists. Equation (8.4) is the defining limit for the partial derivative in the $x$-direction. Similarly,

$$\frac{\partial f}{\partial y}(x_0, y_0) := \lim_{k \to 0} \frac{f(x_0, y_0 + k) - f(x_0, y_0)}{k},$$

when the limit exists. We also use the notation $f_x(x_0, y_0)$ and $f_y(x_0, y_0)$.

**First-order expansion along coordinate lines.** If $f_x(x_0, y_0)$ exists, then

$$f(x_0 + h, y_0) = f(x_0, y_0) + h\, f_x(x_0, y_0) + o(h) \qquad (h \to 0),$$

and analogously for $f_y$. The important point is that the existence of $\partial f / \partial x$ depends only on the values of $f$ along the line segment $\{(x_0 + h, y_0) : |h| \text{ small}\}$, not on the behavior of $f$ in two dimensions.

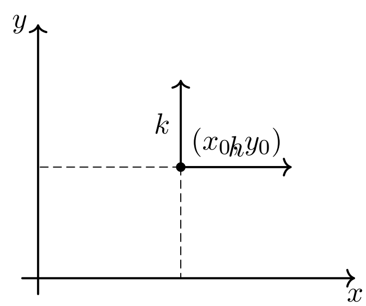



**Figure 8.3.** Partial derivatives measure change along coordinate directions: $f_x$ varies $x$ while holding $y = y_0$ fixed (horizontal arrow), and $f_y$ varies $y$ while holding $x = x_0$ fixed (vertical arrow).

**Warning.** The existence of partial derivatives at a point does *not* imply continuity or differentiability at that point.

**Example 8.11.** $f(x, y) = \dfrac{xy}{x^2 + y^2}$, $f(0, 0) = 0$. Then $f_x(0, 0) = 0 = f_y(0, 0)$ but $f$ is not continuous at $(0, 0)$.

### 8.3.2 Directional derivatives

The directional derivative measures the instantaneous rate of change of $f$ along a prescribed direction.

Let $D \subseteq \mathbb{R}^n$ be open and $f : D \to \mathbb{R}$. Fix $x_0 \in D$ and a unit vector $\mathbf{v} \in \mathbb{R}^n$ (i.e. $\|\mathbf{v}\| = 1$). The *directional derivative* of $f$ at $x_0$ in the direction $\mathbf{v}$ is

$$D_{\mathbf{v}} f(x_0) := \lim_{t \to 0} \frac{f(x_0 + t\mathbf{v}) - f(x_0)}{t}, \tag{8.5}$$

provided the limit exists. As in (8.5), only the values of $f$ along the line segment $\{x_0 + t\mathbf{v} : |t| \text{ small}\}$ matter.

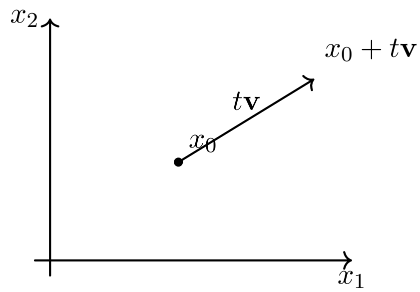



**Figure 8.4.** Directional derivatives probe $f$ along the line $t \mapsto x_0 + t\mathbf{v}$. Only the behavior on this one-dimensional slice is relevant.

**Example 8.12.** Define

$$f(x, y) = \begin{cases} \dfrac{x^2 y}{x^4 + y^2} & \text{if } (x, y) \neq (0, 0), \\ 0 & \text{if } (x, y) = (0, 0). \end{cases}$$

Then every directional derivative $D_{\mathbf{v}} f(0, 0)$ exists at $(0, 0)$. Indeed, for $\mathbf{v} = (v_1, v_2)$ with $\|\mathbf{v}\| = 1$,

$$\frac{f(t\mathbf{v}) - f(0, 0)}{t} = \frac{v_1^2 v_2}{t^2 v_1^4 + v_2^2} \xrightarrow[t \to 0]{} \begin{cases} 0, & v_2 = 0, \\ \dfrac{v_1^2}{v_2}, & v_2 \neq 0. \end{cases}$$

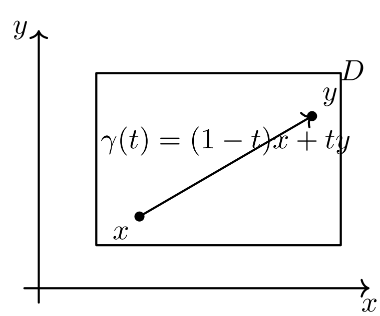



**Figure 8.5.** On a convex set $D$, the straight-line path from $x$ to $y$ stays inside $D$, so the one-variable mean value theorem applied to $t \mapsto f(\gamma(t))$ yields constancy when $\nabla f \equiv 0$.

### 8.3.3 Differentiability

Partial and directional derivatives capture *one-dimensional* behavior of $f$ along lines. The correct multivariable notion is differentiability as *linear approximation.*

Let $D \subseteq \mathbb{R}^n$ be open and let $f : D \to \mathbb{R}^m$. Fix $x_0 \in D$ and write $H \in \mathbb{R}^n$ for an increment.

**Definition 8.13** (Fréchet differentiability)**.** We say that $f$ is *differentiable at* $x_0$ if there exists a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\frac{\|f(x_0 + H) - f(x_0) - AH\|}{\|H\|} \longrightarrow 0 \qquad \text{as } \|H\| \to 0. \tag{8.6}$$

The linear map $A$ is unique; it is called the *derivative* of $f$ at $x_0$ and is denoted by $Df(x_0)$ or $f'(x_0)$.

*Uniqueness of the derivative.* Suppose $A$ and $B$ both satisfy (8.6). Then

$$\frac{\|(A - B)H\|}{\|H\|} \leq \frac{\|f(x_0 + H) - f(x_0) - AH\|}{\|H\|} + \frac{\|f(x_0 + H) - f(x_0) - BH\|}{\|H\|} \longrightarrow 0.$$

Taking $H = t\mathbf{v}$ with $\mathbf{v} \neq \mathbf{0}$ and letting $t \to 0$ shows $\|(A-B)\mathbf{v}\| = 0$ for every $\mathbf{v}$, hence $A = B$. □

**Consequences.** If $f$ is differentiable at $x_0$, then $f$ is continuous at $x_0$, and every directional derivative exists with

$$D_{\mathbf{v}} f(x_0) = f'(x_0)\mathbf{v} \qquad (\mathbf{v} \in \mathbb{R}^n,\ \|\mathbf{v}\| = 1).$$

In particular, for scalar-valued $f : D \to \mathbb{R}$ the derivative $f'(x_0)$ can be identified with the gradient row vector $\nabla f(x_0)$, and $f'(x_0)\mathbf{v} = \nabla f(x_0) \cdot \mathbf{v}$.

**Example 8.14** (Bounded partial derivatives imply continuity)**.** Let $D \subseteq \mathbb{R}^2$ be open and let $f : D \to \mathbb{R}$ have bounded partial derivatives $f_x$ and $f_y$ on $D$. Then $f$ is (locally) Lipschitz on $D$, hence continuous on $D$.

*Proof.* Fix $(x_0, y_0) \in D$ and consider $(x_0 + h, y_0 + k)$ close enough to $(x_0, y_0)$ so that the rectangle with corners $(x_0, y_0)$ and $(x_0 + h, y_0 + k)$ lies in $D$. Using the one-variable mean value theorem along the two coordinate segments,

$$f(x_0 + h, y_0 + k) - f(x_0, y_0) = \big(f(x_0 + h, y_0 + k) - f(x_0, y_0 + k)\big) + \big(f(x_0, y_0 + k) - f(x_0, y_0)\big),$$

and hence

$$\begin{aligned} |f(x_0 + h, y_0 + k) - f(x_0, y_0)| &\leq |h| \sup_D |f_x| + |k| \sup_D |f_y| \\ &\leq \sqrt{h^2 + k^2}\, \sqrt{\big(\sup_D |f_x|\big)^2 + \big(\sup_D |f_y|\big)^2}. \end{aligned}$$

Letting $(h, k) \to (0, 0)$ gives continuity at $(x_0, y_0)$. □

**Exercise 8.15.** Even if $f_x(x_0)$ and $f_y(x_0)$ exist, $f$ need not be differentiable at $x_0$. Find an explicit example (or verify one of the examples below).

**Example 8.16** (Directional derivatives may all exist without differentiability)**.** Define

$$f(x,y) = \begin{cases} \dfrac{y}{|y|}\sqrt{x^2+y^2}, & y \neq 0, \\ 0, & y = 0. \end{cases}$$

Then $f$ is continuous at $(0,0)$ and every directional derivative $D_{\mathbf{v}}f(0,0)$ exists. However, $f$ is not differentiable at $(0,0)$ since the map $\mathbf{v} \mapsto D_{\mathbf{v}}f(0,0)$ is not linear.

*Proof.* For $\mathbf{v} = (v_1, v_2)$ with $\|\mathbf{v}\| = 1$,

$$\frac{f(t\mathbf{v}) - f(0,0)}{t} = \begin{cases} 1, & v_2 > 0, \\ -1, & v_2 < 0, \\ 0, & v_2 = 0, \end{cases} \qquad (t \downarrow 0).$$

Thus $D_{\mathbf{v}}f(0,0)$ exists for every $\mathbf{v}$. If $f$ were differentiable at $(0,0)$, then $D_{\mathbf{v}}f(0,0) = Df(0,0)\cdot\mathbf{v}$ would depend linearly on $\mathbf{v}$, which is impossible. □

**Theorem 8.17** (A sufficient condition for differentiability)**.** *Let $D \subseteq \mathbb{R}^2$ be an open set, and let $(x_0,y_0) \in D$. Suppose that the partial derivatives $f_x$ and $f_y$ exist in a neighborhood of $(x_0,y_0)$ and are continuous at $(x_0,y_0)$. Then $f$ is differentiable at $(x_0,y_0)$. Moreover,*

$$f'(x_0,y_0) = \big(f_x(x_0,y_0),\ f_y(x_0,y_0)\big).$$

**Proof strategy.** Decompose the increment of $f$ into a horizontal step and a vertical step, then apply the one-variable mean value theorem to each leg. Continuity of the partial derivatives lets the coefficients in this decomposition converge to the values of the Jacobian at the base point.

*Proof.* Write $X_0 = (x_0,y_0)$ and let $H = (h,k)$. Since $D$ is open, there exists $r > 0$ such that the closed ball $\overline{B(X_0,r)}$ is contained in $D$. We may therefore assume that $\|H\| < r$, so that all points used below lie in $D$.

Decompose the increment as

$$\begin{aligned} f(x_0+h, y_0+k) - f(x_0,y_0) = {}& \big[f(x_0+h,y_0+k) - f(x_0,y_0+k)\big] \\ &+ \big[f(x_0,y_0+k) - f(x_0,y_0)\big]. \end{aligned}$$

By the one-variable mean value theorem, there exist $\theta_i \in (0,1)$; $i = 1,2$, such that

$$f(x_0+h,y_0+k) - f(x_0,y_0) = f_x(x_0+\theta_1 h, y_0+k)\,h + f_y(x_0, y_0+\theta_2 k)\,k.$$

Subtracting $f_x(x_0, y_0)h + f_y(x_0, y_0)k$, we get

$$\begin{aligned}&f(x_0 + h, y_0 + k) - f(x_0, y_0) - f_x(x_0, y_0)h - f_y(x_0, y_0)k\\&= \big[f_x(x_0 + \theta_1 h, y_0 + k) - f_x(x_0, y_0)\big]h\\&\qquad + \big[f_y(x_0, y_0 + \theta_2 k) - f_y(x_0, y_0)\big]k.\end{aligned}$$

Now let $\epsilon > 0$. Since $f_x$ and $f_y$ are continuous at $(x_0, y_0)$, there exists $\delta > 0$ such that

$$\|(u, v) - (x_0, y_0)\| < \delta \quad \Longrightarrow \quad |f_x(u, v) - f_x(x_0, y_0)| < \epsilon, \quad |f_y(u, v) - f_y(x_0, y_0)| < \epsilon.$$

If $\|H\| < \delta$, then both points $(x_0 + \theta_1 h, y_0 + k)$ and $(x_0, y_0 + \theta_2 k)$ lie within distance at most $\|H\|$ of $(x_0, y_0)$. Therefore

$$\begin{aligned}&\big|f(x_0 + h, y_0 + k) - f(x_0, y_0) - f_x(x_0, y_0)h - f_y(x_0, y_0)k\big|\\&\le \epsilon|h| + \epsilon|k| \le \sqrt{2}\,\epsilon\,\|(h, k)\|.\end{aligned}$$

Thus

$$\frac{\big|f(X_0 + H) - f(X_0) - \nabla f(X_0) \cdot H\big|}{\|H\|} \le \sqrt{2}\,\epsilon$$

whenever $0 < \|H\| < \delta$. □

**Exercise 8.18.** Prove that

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

is differentiable at $(0, 0)$ and $f'(0, 0) = \mathbf{0}$, but neither $f_x$ nor $f_y$ is continuous at $(0, 0)$.

**Geometric interpretation (scalar case).** If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $X_0$, then the graph of $f$ has a tangent hyperplane at $(X_0, f(X_0))$ given by

$$z = f(X_0) + f'(X_0)\,(X - X_0) = f(X_0) + \nabla f(X_0) \cdot (X - X_0).$$

For $n = 1$ this is the familiar tangent line; for $n = 2$ it is a tangent plane. In general, $f'(X_0)$ is the best linear approximation to $f$ near $X_0$.

**Proposition 8.19** (Tangent line, tangent plane, and tangent hyperplane)**.** *If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $X_0$, then*

$$f(X_0 + H) = f(X_0) + f'(X_0)\,H + o(\|H\|),$$

*so the graph of $f$ near $(X_0, f(X_0))$ is approximated by the affine hyperplane*

$$z = f(X_0) + f'(X_0)\,(X - X_0).$$

*For $n = 1$ this reduces to the tangent line*

$$z = f(x_0) + f'(x_0)(x - x_0),$$

*and for $n = 2$ it becomes the tangent plane*

$$z = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

*Proof.* The differentiability relation

$$f(X_0 + H) - f(X_0) - f'(X_0)\, H = o(\|H\|)$$

says precisely that the affine function

$$L(X) := f(X_0) + f'(X_0)\,(X - X_0)$$

approximates $f(X)$ with an error negligible compared to $\|X - X_0\|$. Thus $L$ is the tangent hyperplane to the graph at $(X_0, f(X_0))$. The formulas for $n = 1$ and $n = 2$ are the corresponding coordinate expressions. $\square$

### 8.3.4 Chain rule

**One-variable chain rule.** Let $I, J \subseteq \mathbb{R}$ be intervals, $g : I \to J$, and $f : J \to \mathbb{R}$. If $g$ is differentiable at $x \in I$ and $f$ is differentiable at $g(x) \in J$, then $f \circ g$ is differentiable at $x$ and

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

$$\begin{tikzcd} I \subseteq \mathbb{R} \arrow[r, "g"] \arrow[rr, bend right, "f \circ g"'] & J \subseteq \mathbb{R} \arrow[r, "f"] & \mathbb{R} \end{tikzcd}$$

**Figure 8.6.** The chain rule differentiates the composition $f \circ g$ by combining the derivatives of $g$ and $f$ at the appropriate points.

*Proof.* Set $y = g(x)$ and $k = g(x+h) - g(x)$. Since $f$ is differentiable at $y$, there exists a function $\eta$ with $\eta(k) \to 0$ as $k \to 0$ such that

$$f(y + k) - f(y) - f'(y)k = k\,\eta(k).$$

Since $g$ is differentiable at $x$, we may write $k = hg'(x) + h\mu(h)$ with $\mu(h) \to 0$ as $h \to 0$. Consider the error term

$$\varepsilon(h) := \frac{f(g(x+h)) - f(g(x)) - f'(g(x))\, g'(x)\, h}{h}. \tag{8.7}$$

We will show that $\varepsilon(h) \to 0$ as $h \to 0$; this is exactly the statement that the difference quotient in (8.7) tends to $f'(g(x))g'(x)$. Substituting the expansions above and simplifying gives

$$\varepsilon(h) = \eta(k)\big(g'(x) + \mu(h)\big) + f'(y)\mu(h).$$

As $h \to 0$ we have $k \to 0$, hence $\eta(k) \to 0$; also $\mu(h) \to 0$. Therefore $\varepsilon(h) \to 0$, which means that the difference quotient tends to $f'(g(x))g'(x)$. $\square$

*Notation* 8.20 (Matrix spaces and the operator norm). Let $L_n(\mathbb{R})$ denote the space of linear maps $\mathbb{R}^n \to \mathbb{R}^n$ (equivalently, real $n \times n$ matrices), and let $GL_n(\mathbb{R}) \subset L_n(\mathbb{R})$ be the group of invertible matrices.

For a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ we use the (subordinate) *operator norm*

$$\|A\| := \sup_{\|x\|=1} \|Ax\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}.$$

It satisfies the basic estimates

$$\|Ax\| \leq \|A\|\,\|x\|, \qquad \|AB\| \leq \|A\|\,\|B\|.$$

**Examples.**

(1) If $A : \mathbb{R}^2 \to \mathbb{R}$ is given by $A(x,y) = 4x + 3y$, then by Cauchy–Schwarz, $|A(x,y)| \leq \sqrt{4^2+3^2}\,\sqrt{x^2+y^2} = 5\|(x,y)\|$, and equality is attained at $(x,y) = (4,3)/5$. Hence $\|A\| = 5$.

(2) If $A : \mathbb{R}^2 \to \mathbb{R}^2$ is given by $A(x,y) = (3x, 4y)$, then $\|A(x,y)\|^2 = 9x^2 + 16y^2 \leq 16(x^2+y^2)$, with equality when $x = 0$. Hence $\|A\| = 4$.

Since $L_n(\mathbb{R})$ is finite-dimensional, all norms on $L_n(\mathbb{R})$ are equivalent, and the choice affects only constants, not topological statements.

**Proposition 8.21** (Existence of the operator norm)**.** *Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be linear. Then the quantity*

$$\|A\| = \sup_{\|x\|=1} \|Ax\|$$

*is finite, and in fact the supremum is attained:*

$$\|A\| = \max_{\|x\|=1} \|Ax\|.$$

*Equivalently,*

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|},$$

*and therefore*

$$\|Ax\| \leq \|A\|\,\|x\| \qquad \textit{for all } x \in \mathbb{R}^n.$$

*Proof.* Write $A = (R_1, \ldots, R_m)^t$, where $R_i \in \mathbb{R}^n$ is the $i$th row of the matrix of $A$. For any $x \in \mathbb{R}^n$, Cauchy–Schwarz gives

$$|(Ax)_i| = |R_i \cdot x| \leq \|R_i\| \, \|x\|, \qquad i = 1, \ldots, m.$$

Hence

$$\|Ax\|^2 = \sum_{i=1}^{m} |(Ax)_i|^2 \leq \left( \sum_{i=1}^{m} \|R_i\|^2 \right) \|x\|^2,$$

so $\|Ax\|$ is bounded on the unit sphere $S^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$. The map $x \mapsto \|Ax\|$ is continuous, and $S^{n-1}$ is compact; therefore a maximum is attained on $S^{n-1}$.

If $x \neq 0$, write $x = \|x\|u$ with $\|u\| = 1$. Then

$$\frac{\|Ax\|}{\|x\|} = \|Au\| \leq \max_{\|v\|=1} \|Av\| = \|A\|.$$

Taking the supremum over $x \neq 0$ gives

$$\sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} \leq \|A\|.$$

The reverse inequality follows by restricting the supremum to vectors with $\|x\| = 1$, so the two formulas agree. The estimate $\|Ax\| \leq \|A\| \, \|x\|$ is immediate. □

**Chain rule in $\mathbb{R}^n$.** Let $U \subseteq \mathbb{R}^n$, $V \subseteq \mathbb{R}^m$ be open. If $g : U \to V$ is differentiable at $x \in U$ and $f : V \to \mathbb{R}^k$ is differentiable at $g(x)$, then $f \circ g$ is differentiable at $x$ and

$$(f \circ g)'(x) = f'(g(x)) \, g'(x),$$

where the right-hand side is ordinary matrix multiplication of the Jacobians.

*Proof.* Set $y = g(x)$. Since $f$ is differentiable at $y$, there exists a function $\eta : \mathbb{R}^m \to \mathbb{R}^k$ such that

$$f(y + k) - f(y) - f'(y)k = \|k\| \, \eta(k), \qquad \eta(k) \to 0 \quad (k \to 0).$$

Since $g$ is differentiable at $x$, there exists a function $\varepsilon : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$g(x + h) - g(x) = g'(x)h + \|h\| \, \varepsilon(h), \qquad \varepsilon(h) \to 0 \quad (h \to 0).$$

Now set

$$k := g(x + h) - g(x).$$

Because differentiability implies continuity, $g$ is continuous at $x$, hence $\|h\| \to 0$ implies $\|k\| \to 0$. Moreover,

$$\|k\| = \|g'(x)h + \|h\| \, \varepsilon(h)\| \leq \|g'(x)\| \, \|h\| + \|h\| \, \|\varepsilon(h)\|.$$

Consider

$$\mu(h) := \frac{(f \circ g)(x+h) - (f \circ g)(x) - f'(g(x))\, g'(x)\, h}{\|h\|}.$$

Using $y = g(x)$ and $k - g'(x)h = \|h\|\, \varepsilon(h)$, we obtain

$$\begin{aligned}
\mu(h) &= \frac{f(y+k) - f(y) - f'(y)\, g'(x)h}{\|h\|} \\
&= \frac{f(y+k) - f(y) - f'(y)k + f'(y)\big(k - g'(x)h\big)}{\|h\|} \\
&= \frac{\|k\|\, \eta(k) + \|h\|\, f'(y)\varepsilon(h)}{\|h\|}.
\end{aligned}$$

Hence

$$\|\mu(h)\| \le \frac{\|k\|}{\|h\|}\, \|\eta(k)\| + \|f'(y)\|\, \|\varepsilon(h)\| \le \|\eta(k)\| \big(\|g'(x)\| + \|\varepsilon(h)\|\big) + \|f'(y)\|\, \|\varepsilon(h)\|.$$

As $h \to 0$, we have $k \to 0$, so $\eta(k) \to 0$, and also $\varepsilon(h) \to 0$. Therefore $\mu(h) \to 0$. This proves that $f \circ g$ is differentiable at $x$ and

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

□

**Proposition 8.22** (Vanishing gradient implies constancy on convex sets)**.** *Let $D \subseteq \mathbb{R}^n$ be open and convex, and let $f \in C^1(D)$ satisfy $\nabla f \equiv \mathbf{0}$ on $D$. Then $f$ is constant on $D$.*

*Proof.* Fix $x, y \in D$. Since $D$ is convex, the line segment $\gamma(t) = (1-t)x + ty$ lies in $D$ for $t \in [0, 1]$. Define $\varphi(t) = f(\gamma(t))$. By the chain rule,

$$\varphi'(t) = \nabla f(\gamma(t)) \cdot (y - x) = 0 \quad \text{for all } t \in [0, 1].$$

Hence $\varphi$ is constant, so $f(y) = \varphi(1) = \varphi(0) = f(x)$. □

**Worked examples and a useful identity.**

**Example 8.23** (Radial compositions)**.** Let $\phi : \mathbb{R} \to \mathbb{R}$ be differentiable and define $F : \mathbb{R}^n \to \mathbb{R}$ by $F(x) = \phi(\|x\|^2)$. Then $F$ is differentiable and

$$DF(x) = \phi'(\|x\|^2)\, D(\|x\|^2) = 2\phi'(\|x\|^2)\, x^t, \qquad \text{so} \qquad \nabla F(x) = 2\phi'(\|x\|^2)\, x.$$

More generally, for an integer $k \ge 1$ and $G(x) = \phi(\|x\|^{2k})$ one has (for $x \ne 0$)

$$\nabla G(x) = 2k\, \|x\|^{2k-2}\, \phi'(\|x\|^{2k})\, x.$$

**Theorem 8.24** (Euler's homogeneous function theorem)**.** *Let $U \subseteq \mathbb{R}^n \setminus \{0\}$ be open and let $f : U \to \mathbb{R}$ be differentiable. Assume that $f$ is* homogeneous of degree $\alpha \in \mathbb{R}$*, meaning that*

$$f(tx) = t^{\alpha} f(x) \qquad \text{for all } x \in U \text{ and all } t > 0 \text{ with } tx \in U.$$

*Then for every $x \in U$,*

$$x \cdot \nabla f(x) = \alpha f(x).$$

*Proof.* Fix $x \in U$ and consider the one-variable function $\psi(t) = f(tx)$ for $t$ near 1. By the chain rule, $\psi'(t) = f'(tx)\,x = \nabla f(tx) \cdot x$. The homogeneity hypothesis gives $\psi(t) = t^{\alpha} f(x)$, so $\psi'(t) = \alpha t^{\alpha-1} f(x)$. Evaluating at $t = 1$ yields $\nabla f(x) \cdot x = \alpha f(x)$. □

**Corollary 8.25** (Homogeneous maps near the origin)**.** *Suppose $f : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ is homogeneous of degree $\alpha > 1$ and extends continuously to 0 by setting $f(0) := 0$. Then $f$ is differentiable at 0 and $f'(0) = 0$.*

*Proof.* Write $h = \|h\| v$ with $\|v\| = 1$. By homogeneity, $|f(h)| = \|h\|^{\alpha} |f(v)|$. Continuity on the compact unit sphere implies $M := \sup_{\|v\|=1} |f(v)| < \infty$, hence

$$\frac{|f(h) - f(0)|}{\|h\|} = \frac{|f(h)|}{\|h\|} \le M\, \|h\|^{\alpha-1} \xrightarrow[h \to 0]{} 0.$$

This is exactly differentiability at 0 with derivative 0. □

**Mean value theorem on convex domains.** Let $\mathcal{D} \subseteq \mathbb{R}^n$ be open and convex, and let $f : \mathcal{D} \to \mathbb{R}$ be differentiable. Then for any $x, y \in \mathcal{D}$ there exists a point $c$ on the open line segment $(x, y)$ such that

$$f(y) - f(x) = f'(c)\,(y - x) = \nabla f(c) \cdot (y - x).$$

*Proof.* Define $\varphi(t) = f\big((1-t)x + ty\big)$ for $t \in [0,1]$. By the chain rule, $\varphi$ is differentiable on $(0,1)$ and

$$\varphi'(t) = f'\big((1-t)x + ty\big)\,(y - x).$$

Applying the one-variable mean value theorem to $\varphi$ yields $\lambda \in (0,1)$ such that

$$\begin{aligned} f(y) - f(x) &= \varphi(1) - \varphi(0) \\ &= \varphi'(\lambda) \\ &= f'\big((1-\lambda)x + \lambda y\big)\,(y - x). \end{aligned}$$

Set $c = (1-\lambda)x + \lambda y$. □

**Exercise 8.26.** Let $U$ be an open and connected set in $\mathbb{R}^n$ and $f : U \to \mathbb{R}$ be such that $f'(x) = 0$ for all $x \in U$. Show that $f \equiv 0$.

**Jacobian matrix for vector-valued maps.** Let $U \subseteq \mathbb{R}^n$ be open and $f = (f_1, \dots, f_m) : U \to \mathbb{R}^m$ be differentiable at $x_0 \in U$. Then each partial derivative $\partial f_i / \partial x_j (x_0)$ exists and

$$f'(x_0) = \left( \frac{\partial f_i}{\partial x_j}(x_0) \right)_{1 \le i \le m,\ 1 \le j \le n}.$$

*Proof.* Let $e_j$ be the $j$th standard basis vector of $\mathbb{R}^n$. Differentiability gives

$$\frac{\| f(x_0 + he_j) - f(x_0) - f'(x_0)(he_j) \|}{|h|} \longrightarrow 0 \qquad (h \to 0).$$

Since $f'(x_0)(he_j) = h\, f'(x_0)e_j$, dividing by $h$ shows that

$$\lim_{h \to 0} \frac{f(x_0 + he_j) - f(x_0)}{h} = f'(x_0)e_j \in \mathbb{R}^m,$$

so every component limit exists; equivalently, each $\partial f_i / \partial x_j (x_0)$ exists. Moreover, the vector $f'(x_0)e_j$ is exactly the $j$th column of the Jacobian matrix, hence $f'(x_0)$ is the matrix of partial derivatives. □

**Example 8.27** (Jacobian at a point does not guarantee differentiability)**.** Define $F : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$F(x, y) := \begin{cases} \left( \dfrac{x^2 y}{x^2 + y^2}, \dfrac{xy^2}{x^2 + y^2} \right), & (x, y) \neq (0, 0), \\ (0, 0), & (x, y) = (0, 0). \end{cases}$$

Then all first partial derivatives of the component functions exist at $(0, 0)$ and the Jacobian matrix $F'(0, 0)$ is the zero matrix, but $F$ is *not* differentiable at $(0, 0)$.

*Proof.* Along the coordinate axes we have $F(h, 0) = (0, 0)$ and $F(0, k) = (0, 0)$, so each partial derivative at $(0, 0)$ exists and equals 0, hence $F'(0, 0) = 0$. If $F$ were differentiable at $(0, 0)$ with derivative 0, we would have $\|F(h, k)\| / \sqrt{h^2 + k^2} \to 0$ as $(h, k) \to (0, 0)$. However, along the diagonal $y = x$,

$$F(t, t) = \left( \frac{t}{2}, \frac{t}{2} \right), \qquad t \neq 0.$$

Thus

$$\frac{\|F(t, t)\|}{\|(t, t)\|} = \frac{\sqrt{(t/2)^2 + (t/2)^2}}{\sqrt{t^2 + t^2}} = \frac{1}{2}$$

for all $t \neq 0$, which does not tend to 0 as $t \to 0$. Therefore $F$ is not differentiable at $(0, 0)$. □

**Higher derivatives and the Hessian (scalar case).** If $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ is differentiable on $U$, then each component of $\nabla f$ may itself be differentiated. The matrix of second partial derivatives, when it exists, is called the *Hessian*:

$$f''(x) = \left( \frac{\partial^2 f}{\partial x_j\, \partial x_k}(x) \right)_{1 \le j, k \le n}.$$

The next theorem justifies that under mild regularity, mixed second partials agree.

**Example 8.28** (Mixed partials can disagree without continuity)**.** Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x,y) := \begin{cases} \dfrac{xy(x^2-y^2)}{x^2+y^2}, & (x,y) \neq (0,0), \\ 0, & (x,y) = (0,0). \end{cases}$$

Then the mixed partials at $(0,0)$ exist but are different: $f_{yx}(0,0) = 1$ whereas $f_{xy}(0,0) = -1$. Indeed, for $h \neq 0$ one computes

$$f_y(h,0) = \lim_{k\to 0} \frac{f(h,k) - f(h,0)}{k} = \lim_{k\to 0} h\,\frac{h^2-k^2}{h^2+k^2} = h,$$

so $f_{yx}(0,0) = \lim_{h\to 0} \frac{f_y(h,0)-f_y(0,0)}{h} = 1$. Similarly, for $k \neq 0$,

$$f_x(0,k) = \lim_{h\to 0} \frac{f(h,k) - f(0,k)}{h} = \lim_{h\to 0} k\,\frac{h^2-k^2}{h^2+k^2} = -k,$$

so $f_{xy}(0,0) = \lim_{k\to 0} \frac{f_x(0,k)-f_x(0,0)}{k} = -1$.

*Notation* 8.29 (Smoothness classes). Let $\mathcal{D} \subset \mathbb{R}^n$ be open. For $k \in \mathbb{N}$, we write $C^k(\mathcal{D})$ for the space of functions $f : \mathcal{D} \to \mathbb{R}$ whose partial derivatives of order $\leq k$ exist and are continuous on $\mathcal{D}$. In particular, $C^1(\mathcal{D})$ consists of continuously differentiable functions and $C^2(\mathcal{D})$ consists of twice continuously differentiable functions.

**Theorem 8.30** (Clairaut–Schwarz)**.** *Let $\mathcal{D} \subset \mathbb{R}^2$ be open and let $f \in C^2(\mathcal{D})$. Then for every $(x_0, y_0) \in \mathcal{D}$,*

$$\frac{\partial^2 f}{\partial x\,\partial y}(x_0,y_0) = \frac{\partial^2 f}{\partial y\,\partial x}(x_0,y_0).$$

*More generally, if $\mathcal{D} \subset \mathbb{R}^n$ is open and $f \in C^2(\mathcal{D})$, then*

$$\frac{\partial^2 f}{\partial x_j\,\partial x_k} = \frac{\partial^2 f}{\partial x_k\,\partial x_j}, \qquad 1 \leq j,k \leq n.$$

*Proof.* Fix $(x_0,y_0) \in \mathcal{D}$. Since $\mathcal{D}$ is open, there exists $r > 0$ such that the rectangle $[x_0 - r, x_0 + r] \times [y_0 - r, y_0 + r]$ is contained in $\mathcal{D}$. For $(x,y)$ in this rectangle define

$$F(x,y) := f(x,y) - f(x_0,y) - f(x,y_0) + f(x_0,y_0).$$

Note that $F(x_0,y) = 0$ and $F(x,y_0) = 0$. Fix $x \neq x_0$ and apply the one-variable mean value theorem to the map $y \mapsto f(x,y) - f(x_0,y)$ to obtain a point $\eta$ between $y$ and $y_0$ such that

$$F(x,y) = (y-y_0)\big(f_y(x,\eta) - f_y(x_0,\eta)\big).$$

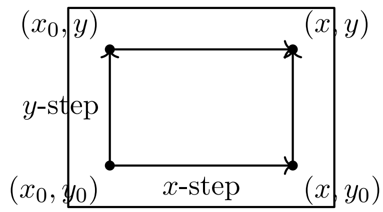



**Figure 8.7.** The rectangle used in the proof of Clairaut–Schwarz: compare increments by moving in the $x$-direction then $y$-direction, versus $y$ then $x$.

Now fix $y \neq y_0$ and apply the mean value theorem to the map $x \mapsto f_y(x, \eta)$ to obtain a point $\xi$ between $x$ and $x_0$ such that

$$f_y(x, \eta) - f_y(x_0, \eta) = (x - x_0) f_{xy}(\xi, \eta).$$

Consequently,

$$\frac{F(x, y)}{(x - x_0)(y - y_0)} = f_{xy}(\xi, \eta).$$

Letting $(x, y) \to (x_0, y_0)$ forces $(\xi, \eta) \to (x_0, y_0)$, hence by continuity of $f_{xy}$ we obtain

$$\lim_{(x,y)\to(x_0,y_0)} \frac{F(x, y)}{(x - x_0)(y - y_0)} = f_{xy}(x_0, y_0).$$

Repeating the argument with the roles of $x$ and $y$ interchanged yields the same limit equal to $f_{yx}(x_0, y_0)$. Therefore $f_{xy}(x_0, y_0) = f_{yx}(x_0, y_0)$. □

### 8.3.5 Taylor's theorem

**Theorem 8.31** (Second-order Taylor formula in $\mathbb{R}^n$)**.** *Let $\mathcal{D} \subseteq \mathbb{R}^n$ be open and let $f \in C^2(\mathcal{D})$. Fix $X \in \mathcal{D}$. Then there exists $\delta > 0$ such that for every $H \in \mathbb{R}^n$ with $\|H\| < \delta$ there exists $\lambda \in (0, 1)$ for which*

$$f(X + H) = f(X) + f'(X)\, H + \frac{1}{2} H^t f''(C)\, H, \qquad C = X + \lambda H,$$

*where $f'(X)$ is the derivative (gradient row vector) and $f''(C)$ is the Hessian matrix at $C$.*

**How to read the formula.** For scalar-valued $f$, the derivative $f'(X)$ can be identified with the gradient row vector $\nabla f(X)$, so the first-order term is $f'(X)\, H = \nabla f(X) \cdot H$. The matrix $f''(C)$ is the Hessian at $C$, and $H^t f''(C) H$ is the associated quadratic form.

*Proof.* Fix $X \in \mathcal{D}$ and choose $\delta > 0$ so that $X + tH \in \mathcal{D}$ for all $t \in [0, 1]$ whenever $\|H\| < \delta$. For such $H$, define $g : [0, 1] \to \mathbb{R}$ by $g(t) = f(X + tH)$. Then $g \in C^2([0, 1])$, and by the chain rule,

$$g'(t) = f'(X + tH)\, H, \qquad g''(t) = H^t f''(X + tH)\, H.$$

Apply the one-variable Taylor theorem (with Lagrange remainder) at $t = 0$ to obtain $\lambda \in (0, 1)$ such that

$$g(1) = g(0) + g'(0) + \frac{1}{2}\, g''(\lambda).$$

Substituting the expressions for $g, g', g''$ and writing $C = X + \lambda H$ gives the stated formula. $\square$

### 8.3.6 Multi-index notation and the general Taylor formula

For higher-order expansions in $\mathbb{R}^n$, it is convenient to use *multi-indices*. A multi-index is a vector $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}_0^n$. We write $|\alpha| := \alpha_1 + \cdots + \alpha_n$, $\alpha! := \alpha_1! \cdots \alpha_n!$, and for $h = (h_1, \ldots, h_n) \in \mathbb{R}^n$,

$$h^\alpha := h_1^{\alpha_1} \cdots h_n^{\alpha_n}.$$

For a sufficiently differentiable function $f$, define

$$D^\alpha f := \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}.$$

**Theorem 8.32** (Taylor's theorem in $\mathbb{R}^n$ (multi-index form)). *Let $U \subset \mathbb{R}^n$ be open, let $k \in \mathbb{N}$, and let $f \in C^{k+1}(U)$. Fix $a \in U$. Then there exists a neighborhood $V$ of $0$ such that $a + h \in U$ for all $h \in V$, and for each such $h$,*

$$f(a + h) = \sum_{|\alpha| \le k} \frac{D^\alpha f(a)}{\alpha!}\, h^\alpha + R_k(a, h),$$

*where the remainder satisfies*

$$\lim_{h \to 0} \frac{R_k(a, h)}{\|h\|_2^k} = 0.$$

*In particular, the polynomial part gives the best $k$-th order approximation of $f$ near $a$ in the sense of little-o.*

*Remark* 8.33. There are several equivalent remainder formulations (integral form, mean-value form) under additional smoothness. For many applications in analysis, the little-$o$ remainder above is the most robust.

**A mean value inequality for vector-valued functions.**

**Theorem 8.34** (Vector-valued mean-value estimate). *Let $f : [a, b] \to \mathbb{R}^n$ be differentiable on $(a, b)$ and continuous on $[a, b]$. Then there exists $\lambda \in (a, b)$ such that*

$$\|f(b) - f(a)\| \le \|f'(\lambda)\|\, (b - a).$$

*Proof.* Consider the scalar function $g(t) = (f(b) - f(a)) \cdot f(t)$ on $[a, b]$. Then $g$ is continuous on $[a, b]$ and differentiable on $(a, b)$, with

$$g'(t) = (f(b) - f(a)) \cdot f'(t).$$

By the one-variable mean value theorem, there exists $\lambda \in (a, b)$ such that

$$g(b) - g(a) = g'(\lambda)\,(b - a).$$

Since $g(b) - g(a) = (f(b) - f(a)) \cdot (f(b) - f(a)) = \|f(b) - f(a)\|^2$, we obtain

$$\begin{aligned}\|f(b) - f(a)\|^2 &= (b - a)\,(f(b) - f(a)) \cdot f'(\lambda)\\ &\le (b - a)\,\|f(b) - f(a)\|\,\|f'(\lambda)\|,\end{aligned}$$

and the conclusion follows (trivially if $f(b) = f(a)$). □

**Theorem 8.35** (Mean value inequality along line segments)**.** *Let $\mathcal{D} \subseteq \mathbb{R}^n$ be open and let $f : \mathcal{D} \to \mathbb{R}^m$ be differentiable. Fix $X \in \mathcal{D}$. Then there exists $\varepsilon > 0$ such that whenever $\|H\| < \varepsilon$ we can find $\lambda \in (0, 1)$ with*

$$\|f(X + H) - f(X)\| \le \|f'(X + \lambda H)\|\,\|H\|.$$

*Proof.* Choose $\varepsilon > 0$ so that $X + tH \in \mathcal{D}$ for all $t \in [0, 1]$ whenever $\|H\| < \varepsilon$. Fix such $H$ and define $g(t) = f(X + tH)$ for $t \in [0, 1]$. Then $g : [0, 1] \to \mathbb{R}^m$ is differentiable and $g'(t) = f'(X + tH)\,H$. Apply Theorem 8.34 to $g$ on $[0, 1]$ to obtain $\lambda \in (0, 1)$ such that

$$\|f(X + H) - f(X)\| = \|g(1) - g(0)\| \le \|g'(\lambda)\| \le \|f'(X + \lambda H)\|\,\|H\|.$$

□

*Remark* 8.36 (Equality need not hold in the vector mean value inequality)*.* The scalar mean value theorem produces an exact identity, but for vector-valued functions the correct statement is generally only an inequality. For example, define $g : [-1, 1] \to \mathbb{R}^2$ by

$$g(t) = (t^3, 1 - t^2).$$

Then

$$g(1) - g(-1) = (2, 0), \qquad \|g(1) - g(-1)\| = 2.$$

Also

$$g'(t) = (3t^2, -2t), \qquad \|g'(t)\| = \sqrt{9t^4 + 4t^2} = |t|\sqrt{9t^2 + 4}.$$

If one had an exact vector-valued mean value formula

$$\|g(1) - g(-1)\| = \|g'(c)\|\,(1 - (-1))$$

for some $c \in (-1,1)$, then $\|g'(c)\| = 1$, hence $9c^4 + 4c^2 - 1 = 0$, so $c^2 = \frac{1}{3}$. At such a point,

$$g'(c) = (1, -2c)\,,$$

which is not parallel to $g(1) - g(-1) = (2, 0)$.

*Geometric picture.* The image of $g$ is the planar curve

$$g(t) = (t^3, 1 - t^2), \qquad -1 \le t \le 1,$$

which joins $g(-1) = (-1, 0)$ to $g(1) = (1, 0)$. The secant vector $g(1) - g(-1)$ is horizontal, whereas at the distinguished point $c = \frac{1}{\sqrt{3}}$ the tangent direction $g'(c) = \big(1, -\frac{2}{\sqrt{3}}\big)$ is not horizontal. The figure below makes this mismatch visible.

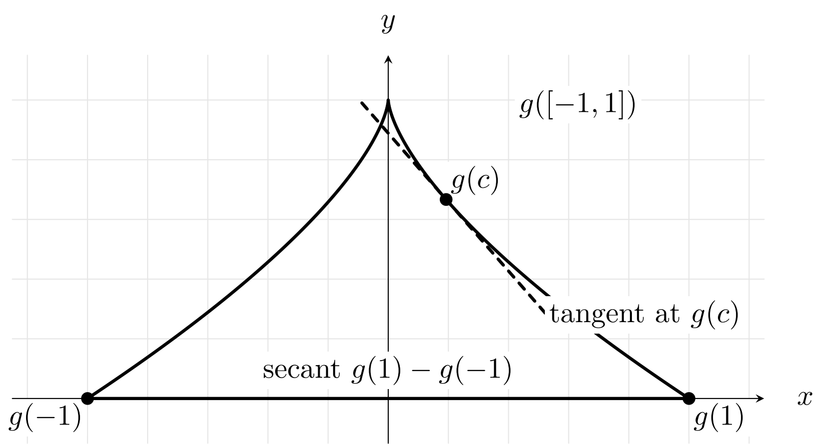



Figure for Remark 8.36: the secant direction determined by the endpoints is horizontal, but the tangent direction at the candidate mean-value point is not.

Thus the one-dimensional equality has no direct vector-valued analogue in general; the norm inequality is the correct replacement.

**Example 8.37** (Derivative of the inverse in one dimension)**.** Let $f : \mathbb{R} \to \mathbb{R}$ be bijective and continuously differentiable on a neighborhood of $x_0$, with $f'(x_0) \neq 0$. Then $f^{-1}$ is differentiable at $y_0 = f(x_0)$ and

$$(f^{-1})'(y_0) = \frac{1}{f'(x_0)}.$$

*Proof.* Let $k \to 0$ and set $h = f^{-1}(y_0 + k) - f^{-1}(y_0)$, so that $y_0 + k = f(x_0 + h)$ and $k = f(x_0 + h) - f(x_0)$. By the mean value theorem applied to $f$ on the interval between $x_0$ and $x_0 + h$, there exists $\theta \in (0, 1)$ such that

$$k = h\, f'(x_0 + \theta h).$$

Since $f'(x_0) \neq 0$ and $f'$ is continuous, for $h$ small we have $|f'(x_0 + \theta h)| \ge m > 0$, hence $k \to 0$ implies $h \to 0$. Now compute

$$\frac{f^{-1}(y_0+k)-f^{-1}(y_0)-\dfrac{k}{f'(x_0)}}{|k|}=\frac{h-\dfrac{f(x_0+h)-f(x_0)}{f'(x_0)}}{|f(x_0+h)-f(x_0)|}$$
$$=\frac{|f'(x_0)-f'(x_0+\theta h)|}{|f'(x_0+\theta h)|\,|f'(x_0)|}\xrightarrow[h\to 0]{}0.$$

Since $f^{-1}$ is differentiable at $y_0=f(x_0)$, by differentiating $f^{-1}(f(x))=x$ at $x=x_0$ yields $(f^{-1})'(f(x_0))\,f'(x_0)=1$. ☐

**Example 8.38** (Invertible Jacobian does not imply global one-to-one behavior)**.** Define $F:\mathbb{R}^2\to\mathbb{R}^2$ by

$$F(x,y)=\big(e^x\cos y,\ e^x\sin y\big).$$

Then

$$DF(x,y)=\begin{pmatrix}e^x\cos y & -e^x\sin y\\ e^x\sin y & e^x\cos y\end{pmatrix},\qquad \det DF(x,y)=e^{2x}>0$$

for every $(x,y)\in\mathbb{R}^2$. Thus the Jacobian matrix is invertible at every point. Nevertheless $F$ is not injective on $\mathbb{R}^2$, because

$$F(x,y+2\pi)=F(x,y)\qquad\text{for all }(x,y)\in\mathbb{R}^2.$$

This example is a useful warning before the inverse function theorem: invertibility of the derivative guarantees *local* invertibility, not a global inverse on the whole domain.

## 8.4 Inverse and implicit function theorems

**Section overview.**

- These theorems explain when nonlinear equations can be solved locally and smoothly.
- They are the nonlinear analogues of solving linear systems with an invertible matrix.

We now turn to two fundamental results describing the local structure of $C^1$ maps: when a map is locally invertible (inverse function theorem) and when a level set can be solved locally as a graph (implicit function theorem).

**Proposition 8.39** (Stability of invertibility and continuity of inversion)**.** *Let $A\in GL_n(\mathbb{R})$ and $B\in L_n(\mathbb{R})$ satisfy*

$$\|B-A\|<\frac{1}{\|A^{-1}\|}.$$

*Then $B\in GL_n(\mathbb{R})$. Moreover, the map $A\mapsto A^{-1}$ is continuous on $GL_n(\mathbb{R})$.*

*Proof.* Set $\alpha=\frac{1}{\|A^{-1}\|}$ and $\beta=\|B-A\|$, so $\beta<\alpha$. For any $x\in\mathbb{R}^n$,

$$\begin{aligned}\alpha\|x\| &= \alpha\|A^{-1}Ax\| \\ &\le \alpha\|A^{-1}\|\,\|Ax\| \\ &= \|Ax\| \\ &\le \|(A-B)x\| + \|Bx\| \\ &\le \beta\|x\| + \|Bx\|.\end{aligned}$$

Hence $(\alpha-\beta)\|x\| \le \|Bx\|$ for all $x$. If $Bx = 0$, this forces $x = 0$, so $B$ is injective and therefore invertible (finite dimensions). For continuity, note that for $y \neq 0$ and $x = B^{-1}y$,

$$\begin{aligned}(\alpha-\beta)\|B^{-1}y\| \le \|y\| \\ &\Rightarrow \quad \|B^{-1}\| \\ &\le \frac{1}{\alpha-\beta}.\end{aligned}$$

Then

$$\begin{aligned}\|B^{-1} - A^{-1}\| &= \|B^{-1}(A-B)A^{-1}\| \\ &\le \|B^{-1}\|\,\|A-B\|\,\|A^{-1}\| \\ &\le \frac{\|A-B\|}{\alpha(\alpha-\beta)} \xrightarrow[B\to A]{} 0.\end{aligned}$$

Thus inversion is continuous on $GL_n(\mathbb{R})$. □

### 8.4.1 Inverse function theorem

**Theorem 8.40** (Inverse Function Theorem)**.** *Let $\Omega$ be an open set in $\mathbb{R}^n$. Suppose $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$ be a $C^1$ map such that* $\det f'(x_0) \neq 0$. *Then*

*(i) There exist open sets $U$ and $V \subset \mathbb{R}^n$ such that $f : U \to V$ $(= f(U))$ is bijective.*

*(ii) $f^{-1}$ is a $C^1$ map on $V$, and*

$$(f^{-1})'(f(x_0)) = (f'(x_0))^{-1}.$$

**Idea.** Near $x_0$, the map $f$ is well-approximated by its linearization $A = f'(x_0)$. The hypothesis $\det A \neq 0$ means this linear approximation is invertible. The proof turns the equation $f(x) = y$ into a fixed-point problem whose solution follows from the contraction mapping principle, uniformly for $y$ in a small neighborhood of $f(x_0)$.

*Proof.* Let $A = f'(x_0)$. Since $\det A \neq 0$, the matrix $A$ is invertible. **Step 0 (fixed-point reformulation).** For $y \in \mathbb{R}^n$, define

$$\varphi_y(x) := x + A^{-1}(y - f(x)), \qquad x \in \Omega. \tag{8.8}$$

Then $\varphi_y(x) = x$ if and only if $f(x) = y$. In other words, solving $f(x) = y$ is equivalent to finding a fixed point of $\varphi_y$. **Step 1 (a uniform contraction estimate).** Because $f'$ is continuous at $x_0$, for

$$\varepsilon = \frac{1}{2\|A^{-1}\|} > 0$$

there exists $\delta > 0$ such that

$$\|x - x_0\| < \delta \implies \|f'(x) - A\| < \frac{1}{2\|A^{-1}\|}.$$

Set $U := B_\delta(x_0) \subset \Omega$ and $V := f(U)$. For $x \in U$ we have

$$\begin{aligned}\|\varphi_y'(x)\| &= \|I - A^{-1}f'(x)\| \\ &= \|A^{-1}(A - f'(x))\| \\ &\le \|A^{-1}\|\,\|A - f'(x)\| < \frac{1}{2}.\end{aligned}$$

By the mean value estimate for $C^1$ maps on convex sets (applied to $\varphi_y$ on $U$), it follows that

$$\|\varphi_y(x_1) - \varphi_y(x_2)\| \le \frac{1}{2}\,\|x_1 - x_2\| \qquad \text{for all } x_1, x_2 \in U. \tag{8.9}$$

Thus $\varphi_y$ is a contraction on $U$ (with contraction constant $1/2$), uniformly in $y$. **(i) $f$ is one-to-one on $U$.** If $f(x_1) = f(x_2) = y$ with $x_1, x_2 \in U$, then both $x_1$ and $x_2$ are fixed points of $\varphi_y$. By (8.9), a contraction has at most one fixed point, hence $x_1 = x_2$. **(ii) $V$ is open (and $f : U \to V$ is bijective).** Fix $y^* \in V$. Choose $x^* \in U$ with $y^* = f(x^*)$ and pick $r > 0$ such that the closed ball $\overline{B}_r(x^*) \subset U$. We claim that

$$\|y - y^*\| < \frac{r}{2\|A^{-1}\|} \quad \implies \quad y \in V. \tag{8.10}$$

Indeed, for such $y$ we have

$$\|\varphi_y(x^*) - x^*\| = \|A^{-1}(y - y^*)\| \le \|A^{-1}\|\,\|y - y^*\| < \frac{r}{2}.$$

If $x \in \overline{B}_r(x^*)$, then using (8.9) we obtain

$$\begin{aligned}\|\varphi_y(x) - x^*\| &\le \|\varphi_y(x) - \varphi_y(x^*)\| + \|\varphi_y(x^*) - x^*\| < \frac{1}{2}\|x - x^*\| + \frac{r}{2} \\ &\le r.\end{aligned}$$

Hence $\varphi_y(\overline{B}_r(x^*)) \subset \overline{B}_r(x^*)$. Since $\overline{B}_r(x^*)$ is complete and $\varphi_y$ is a contraction there, the contraction mapping principle gives a unique fixed point $x \in \overline{B}_r(x^*)$. This fixed point satisfies $f(x) = y$, so $y \in f(U) = V$, proving (8.10). Thus every point of $V$ is an interior point, so $V$ is open. Since $V = f(U)$ and $f$ is injective on $U$, the restriction $f : U \to V$ is bijective. **(iii)**

**Differentiability of $f^{-1}$ at $f(x_0)$.** Let $y_0 = f(x_0)$ and let $k$ be small enough that $y_0 + k \in V$. Set

$$h := f^{-1}(y_0 + k) - x_0, \qquad \text{so that} \qquad k = f(x_0 + h) - f(x_0).$$

Using (8.8) with $y = y_0 + k$, the fixed-point property $\varphi_{y_0+k}(x_0 + h) = x_0 + h$ and $\varphi_{y_0+k}(x_0) = x_0 + A^{-1}k$ yield

$$h - A^{-1}k = \varphi_{y_0+k}(x_0 + h) - \varphi_{y_0+k}(x_0).$$

By the contraction estimate (8.9),

$$\|h - A^{-1}k\| \le \frac{1}{2}\,\|h\|.$$

Consequently,

$$\frac{1}{2}\,\|h\| \le \|A^{-1}k\| \le \|A^{-1}\|\,\|k\|, \qquad \text{so} \qquad \|h\| \le 2\|A^{-1}\|\,\|k\|. \tag{8.11}$$

Since $f$ is differentiable at $x_0$, we can write

$$f(x_0 + h) - f(x_0) - Ah = \|h\|\,\rho(h), \qquad \rho(h) \to 0 \text{ as } \|h\| \to 0.$$

But $k = f(x_0 + h) - f(x_0)$, hence

$$\begin{aligned} k - Ah = \|h\|\,\rho(h), \qquad \text{and therefore} \qquad h - A^{-1}k \\ = -A^{-1}(k - Ah) \\ = -\|h\|\,A^{-1}\rho(h). \end{aligned}$$

Divide by $\|k\|$ and use (8.11):

$$\begin{aligned} \frac{\|h - A^{-1}k\|}{\|k\|} &\le \frac{\|h\|\,\|A^{-1}\|\,\|\rho(h)\|}{\|k\|} \\ &\le 2\|A^{-1}\|^2\,\|\rho(h)\| \longrightarrow 0. \end{aligned}$$

Thus $f^{-1}$ is differentiable at $y_0$ with

$$(f^{-1})'(y_0) = A^{-1} = (f'(x_0))^{-1}.$$

**(iv) $f^{-1}$ is $C^1$ on $V$.** The above argument applies at any $y \in V$ with $x = f^{-1}(y)$ and $A = f'(x)$ (which is invertible on $U$ by continuity of $\det f'$). Hence

$$(f^{-1})'(y) = \big(f'(f^{-1}(y))\big)^{-1}, \qquad y \in V.$$

The map $y \mapsto f^{-1}(y)$ is continuous, the derivative $f'$ is continuous on $U$, and the inversion map $A \mapsto A^{-1}$ is continuous on $GL_n(\mathbb{R})$. Therefore $y \mapsto (f^{-1})'(y)$ is continuous on $V$, i.e., $f^{-1} \in C^1(V)$. □

**Example 8.41** (A concrete Jacobian computation)**.** Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$f(x, y) = (x - e^{-y},\, y - e^x).$$

Then

$$f'(0,0) = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \qquad \det f'(0,0) = 2 \neq 0. \tag{8.12}$$

The non-vanishing determinant in (8.12) is precisely the inverse-function hypothesis. Therefore $f$ is locally invertible near $(0, 0)$, and

$$(f^{-1})'(f(0,0)) = (f'(0,0))^{-1} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

This example is best read as a computational template: first verify invertibility of the Jacobian, then obtain the derivative of the local inverse for free.

**Example 8.42** (Local vs. global invertibility)**.** Define $F : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$F(x, y) = (e^x \cos y,\, e^x \sin y).$$

Then $F$ is $C^1$ and

$$\det F'(x, y) = e^{2x} \neq 0 \qquad \text{for all } (x, y) \in \mathbb{R}^2.$$

Hence, by the inverse function theorem, $F$ is locally one-to-one (and locally onto its image) at every point. Nevertheless, $F$ is not injective on all of $\mathbb{R}^2$, since

$$F(x, y + 2\pi) = F(x, y) \qquad \text{for all } (x, y) \in \mathbb{R}^2.$$

This illustrates that the inverse function theorem is inherently a *local* result.

### 8.4.2 Implicit function theorem

**A motivating example.** Consider $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x, y) = x^2 + y^2 - 1.$$

The level set $\{(x, y) : f(x, y) = 0\}$ is the unit circle.

We have $f'(x, y) = (2x, 2y)$, so at $(1, 0)$

$$\left.\frac{\partial f}{\partial x}\right|_{(1,0)} = 2 \neq 0, \qquad \left.\frac{\partial f}{\partial y}\right|_{(1,0)} = 0.$$

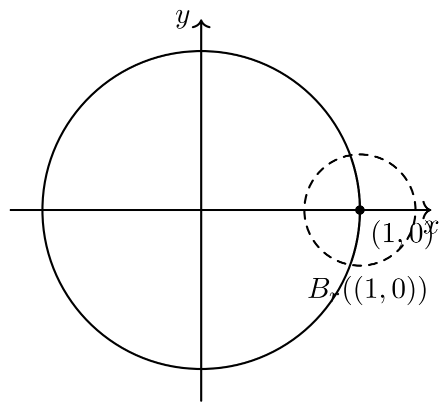



**Figure 8.8.** Near $(1,0)$, the circle $x^2+y^2=1$ is the graph $x=\sqrt{1-y^2}$, but it cannot be written as a single-valued graph $y=\psi(x)$ near $x=1$.

Because $\partial f/\partial x \neq 0$ at $(1,0)$, we can solve $f(x,y)=0$ locally for $x$ as a $C^1$ function of $y$:

$$x=\varphi(y)=\sqrt{1-y^2}, \qquad |y|<r,$$

for some small $r>0$. On the other hand, we *cannot* solve locally for $y$ as a single-valued function of $x$ near $x=1$, since for $x$ slightly less than 1 the equation $x^2+y^2=1$ has two solutions $y=\pm\sqrt{1-x^2}$. This illustrates that which variable can be solved for depends on a non-degeneracy condition on the relevant partial derivative.

**A linear-algebra model.** Let

$$A:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^n$$

be a linear map. Writing $(h,k)=(h,0)+(0,k)$, we may decompose

$$A(h,k)=A(h,0)+A(0,k)=A_xh+A_yk,$$

where $A_x:\mathbb{R}^n\to\mathbb{R}^n$ and $A_y:\mathbb{R}^m\to\mathbb{R}^n$ are linear.

**Lemma 8.43** (Solving the block linear equation)**.** *If $A_x$ is invertible, then for each $k\in\mathbb{R}^m$ there exists a unique $h\in\mathbb{R}^n$ such that $A(h,k)=0$, namely*

$$h=-A_x^{-1}A_yk.$$

*Proof.* The equation $A(h,k)=0$ is equivalent to $A_xh+A_yk=0$. If $A_x$ is invertible, we can solve uniquely for $h$: $h=-A_x^{-1}A_yk$. □

**Block form of the derivative.** If $\Omega\subset\mathbb{R}^n\times\mathbb{R}^m$ is open and $f:\Omega\to\mathbb{R}^n$ is differentiable, then for $(x,y)\in\Omega$ the derivative $f'(x,y)$ is an $n\times(n+m)$ matrix. It is convenient to write it in block form

$$f'(x,y)=(A_x \;\; A_y),$$

where

$$A_x = \left(\frac{\partial f_i}{\partial x_j}(x,y)\right)_{1\le i\le n,\ 1\le j\le n}, \qquad A_y = \left(\frac{\partial f_i}{\partial y_k}(x,y)\right)_{1\le i\le n,\ 1\le k\le m}.$$

**Theorem 8.44** (Implicit Function Theorem). *Let $\Omega$ be an open subset of $\mathbb{R}^n \times \mathbb{R}^m$ and let $f : \Omega \to \mathbb{R}^n$ be a $C^1$ map. Assume that $f(x_0, y_0) = 0$ for some $(x_0, y_0) \in \Omega$ and that the $n \times n$ block*

$$A_x := [f'(x_0, y_0)]_x$$

*is invertible (equivalently, $\det A_x \neq 0$). Then:*

*(i) There exist open sets $U \subset \mathbb{R}^n \times \mathbb{R}^m$ with $(x_0, y_0) \in U$ and $W \subset \mathbb{R}^m$ with $y_0 \in W$ such that for every $y \in W$ there exists a unique $x \in \mathbb{R}^n$ with $(x, y) \in U$ and $f(x, y) = 0$.*

*(ii) Writing this unique solution as $x = g(y)$, we obtain a $C^1$ map $g : W \to \mathbb{R}^n$ with $g(y_0) = x_0$ and $f(g(y), y) = 0$ for all $y \in W$. Moreover,*

$$g'(y_0) = -A_x^{-1} A_y, \qquad \textit{where} \qquad A_y := [f'(x_0, y_0)]_y .$$

*In particular, near $(x_0, y_0)$ the zero set $\{(x, y) \in \Omega : f(x, y) = 0\}$ is the graph of the function $x = g(y)$.*

**Geometric meaning.** The equation $f(x, y) = 0$ describes a level set in $\mathbb{R}^{n+m}$. The invertibility of the $x$-block of the Jacobian says that, near $(x_0, y_0)$, the equation can be solved uniquely for $x$ as a $C^1$ function of $y$. In other words, the level set is locally the graph of a function $x = \varphi(y)$.

*Proof.* (i) Define $F : \Omega \to \mathbb{R}^n \times \mathbb{R}^m$ by

$$F(x, y) := (f(x, y), y).$$

Then $F$ is a $C^1$ map and its derivative at $(x_0, y_0)$ has block form

$$F'(x_0, y_0) = \begin{bmatrix} A_x & A_y \\ 0 & I \end{bmatrix}.$$

Hence $\det F'(x_0, y_0) = \det(A_x) \neq 0$. By the inverse function theorem, there exist open sets $U \subset \mathbb{R}^n \times \mathbb{R}^m$ and $V \subset \mathbb{R}^n \times \mathbb{R}^m$ such that

$$F : U \to V$$

is a $C^1$ bijection and $F^{-1} : V \to U$ is $C^1$. Let

$$W := \{\, y \in \mathbb{R}^m : \ (0, y) \in V \,\}.$$

Since $V$ is open, $W$ is open in $\mathbb{R}^m$. For any $y \in W$, we have $(0, y) \in V$, and since $F$ is onto $V$ there exists $(x, y) \in U$ such that

$$(0, y) = F(x, y) = (f(x, y), y).$$

Thus $f(x, y) = 0$. If also $(x', y) \in U$ satisfies $f(x', y) = 0$, then $F(x', y) = (0, y) = F(x, y)$, and injectivity of $F$ on $U$ forces $x' = x$. This proves existence and uniqueness of $x$ for each $y \in W$.

(ii) Define $g : W \to \mathbb{R}^n$ by declaring $g(y)$ to be the unique $x$ with $(x, y) \in U$ and $f(x, y) = 0$. Then $(g(y), y) \in U$ and

$$F(g(y), y) = (0, y), \qquad y \in W.$$

Equivalently, $F^{-1}(0, y) = (g(y), y)$ for $y \in W$. Since $F^{-1}$ is $C^1$, it follows that $g$ is $C^1$. To compute $g'(y_0)$, differentiate the identity $f(g(y), y) = 0$ at $y = y_0$ and use the chain rule:

$$0 = \frac{d}{dy} f(g(y), y)\Big|_{y=y_0} = A_x\, g'(y_0) + A_y.$$

Since $A_x$ is invertible, this gives $g'(y_0) = -A_x^{-1} A_y$.

□

**Exercise 8.45.** Prove that $x^2 + ye^x - \sin(xy) = 0$ can be solved for $y$ in a neighborhood of $(0, 0)$, but cannot be solved for $x$ in any neighborhood of $(0, 0)$.

$$F(x, y) = x^2 + ye^x - \sin(xy)$$

(i) $F(0, 0) = 0$, $\frac{\partial F}{\partial y}|_{(0,0)} = 1 \neq 0$. By the implicit function theorem, there exists a neighborhood of $(0, 0)$ and a function $g$ such that

$$F(x, g(x)) = 0$$

for $|x| < r$, equivalently $y = g(x)$ near 0.

(ii) $\frac{\partial F}{\partial x}|_{(0,0)} = 0$. Hence, the implicit function theorem *cannot* be applied.
On the contrary, suppose that $x = \phi(y)$ near $y = 0$. Then $\phi(0) = 0$, and

$$(\phi(y))^2 + ye^{\phi(y)} - \sin(\phi(y)y) = 0$$

for $|y| < r$ for some $r > 0$. Then

$$2\phi(0)\phi'(0) + 1 \cdot e^{\phi(0)} + 0 \cdot e^{\phi(0)}\phi'(0) - \cos(\phi(0)0)\,(\phi'(0)0 + \phi(0) \cdot 1) = 0$$

$$\implies 1 = 0 \quad \text{(contradiction)}$$

**Example 8.46.** Let $f : \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2$

$$f(x, y, z) = (xe^y + ye^z,\ xe^z + ze^y)$$

Then $f$ is a $C^1$-map.

$$f'(x,y,z) = \begin{pmatrix} e^y & xe^y + e^z & ye^z \\ e^z & ze^y & xe^z + e^y \end{pmatrix}$$

$$f(-1,1,1) = (0,0)$$

Let $f = (f_1, f_2)$. Then

$$\begin{pmatrix} \dfrac{\partial f_1}{\partial y} & \dfrac{\partial f_1}{\partial z} \\ \dfrac{\partial f_2}{\partial y} & \dfrac{\partial f_2}{\partial z} \end{pmatrix}(-1,1,1) = \begin{pmatrix} 0 & e \\ e & 0 \end{pmatrix}$$

By the implicit function theorem, there exists an open ball $U$ in $\mathbb{R}^3$ and open ball $V$ in $\mathbb{R}^2$, such that

$$(y,z) = (\phi(x), \psi(x)), \quad |x| < r \quad \text{for some } r > 0.$$

**Exercise 8.47.** Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a $C^1$-map such that $f(0,0) = 0, f_x(0,0) = 1$. Let $F(x,y) = (f(x,y), y)$. Prove that $F$ is injective in some neighborhood of $(0,0)$. Does $F$ remain injective in any neighborhood of $(0,0)$?

*Remark:* Condition in implicit function theorem or inverse mapping theorem on derivatives are sufficient.

**Example 8.48.** $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = x^2 - y^3$.

$$f(0,0) = 0,$$
$$\frac{\partial f}{\partial y}(0,0) = 0,$$

but $y = x^{2/3}$ is a solution of $f(x,y) = 0$ near $(0,0)$.

**Example 8.49.** Let $f : \mathbb{R}^2 \to \mathbb{R}^2$, $f(x,y) = (x^3, y^3)$. Then $\det f'(0,0) = 0$ but $f$ is one-to-one, onto.

## 8.5 Extrema, the Hessian test, and Lagrange multipliers

**Section overview.**

- Optimization in several variables combines first-order geometry with second-order tests.
- Gradients, Hessians, and constraints provide the natural language for locating and classifying extrema.

Many applications of multivariable calculus involve optimizing a function under no constraints (unconstrained extrema) or under one or more constraints (constrained extrema). The correct language is geometric: level sets and gradients.

### 8.5.1 Unconstrained extrema and critical points

Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $a \in \mathbb{R}^n$. We say that $a$ is a *local maximizer* of $f$ if there exists $r > 0$ such that $f(a) \geq f(x)$ for all $x \in B_r(a)$. Local minimizers are defined analogously.

**Theorem 8.50** (Fermat's theorem for $C^1$ functions)**.** *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $a \in \mathbb{R}^n$. If $a$ is a local maximizer or local minimizer of $f$, then $\nabla f(a) = 0$.*

*Proof.* Fix any unit vector $v \in \mathbb{R}^n$ and consider the one-variable function $\varphi(t) = f(a + tv)$ for $t$ near 0. If $a$ is a local extremum of $f$, then $t = 0$ is a local extremum of $\varphi$. Since $\varphi$ is differentiable at 0, we have $\varphi'(0) = 0$. By the chain rule, $\varphi'(0) = \nabla f(a) \cdot v$. Because $v$ is arbitrary, it follows that $\nabla f(a) = 0$. □

Points $a$ with $\nabla f(a) = 0$ are called *critical points*. Fermat's theorem says that local extrema can occur only at critical points (or on the boundary of the domain).

### 8.5.2 The Hessian and the second derivative test

Assume $f \in C^2$ in a neighborhood of $a \in \mathbb{R}^n$. The *Hessian matrix* at $a$ is

$$H_f(a) := \left[ \frac{\partial^2 f}{\partial x_i \partial x_j}(a) \right]_{i,j=1}^{n}.$$

It is symmetric when $f \in C^2$, and thus determines a quadratic form $q(h) = \frac{1}{2} h^\mathsf{T} H_f(a) h$.

**Theorem 8.51** (Second derivative test)**.** *Let $f \in C^2$ in a neighborhood of $a$ and assume $\nabla f(a) = 0$.*

*(i) If $H_f(a)$ is positive definite, then $a$ is a strict local minimizer.*

*(ii) If $H_f(a)$ is negative definite, then $a$ is a strict local maximizer.*

*(iii) If $H_f(a)$ is indefinite, then $a$ is a saddle point.*

*Proof.* By Taylor's theorem with remainder (in $\mathbb{R}^n$),

$$f(a + h) = f(a) + \nabla f(a) \cdot h + \tfrac{1}{2} h^\mathsf{T} H_f(a) h + o(\|h\|^2) \qquad (h \to 0).$$

Since $\nabla f(a) = 0$, the sign of $f(a + h) - f(a)$ for sufficiently small $h$ is governed by the quadratic form $h^\mathsf{T} H_f(a) h$, yielding the three cases. □

### 8.5.3 Constrained extrema and the Lagrange multiplier condition

We now optimize $f$ subject to a constraint $g(x) = 0$. Geometrically, the constraint set $M = \{x : g(x) = 0\}$ is (under mild hypotheses) a smooth hypersurface, and the gradient $\nabla g$ is normal to its level sets. At an extremum of $f$ on $M$, the level set of $f$ must be tangent to $M$, forcing the gradients to be parallel.

**Theorem 8.52** (Lagrange multiplier condition)**.** *Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be $C^1$ in a neighborhood of $a \in \mathbb{R}^n$. Assume $g(a) = 0$ and $\nabla g(a) \neq 0$. If $a$ is a local extremum of $f$ restricted to the constraint set $M = \{x : g(x) = 0\}$, then there exists $\lambda \in \mathbb{R}$ such that*

$$\nabla f(a) = \lambda \nabla g(a).$$

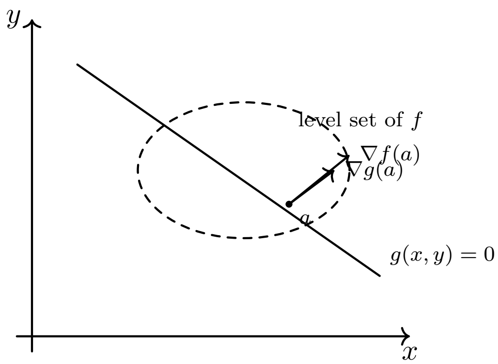



**Figure 8.9.** At a constrained extremum, the level set of $f$ is tangent to the constraint set $g = 0$, so $\nabla f$ is parallel to $\nabla g$.

**Proof strategy.** Use the implicit function theorem to write the constraint set locally as a graph. The constrained extremum then becomes an ordinary local extremum in $\mathbb{R}^{n-1}$, to which Fermat's theorem applies. The chain rule then converts the resulting equations into the proportionality of gradients.

*Proof.* Because $\nabla g(a) \neq 0$, at least one partial derivative of $g$ is nonzero at $a$. After a permutation of coordinates, we may assume that

$$\frac{\partial g}{\partial x_n}(a) \neq 0.$$

Write $a = (a', a_n)$ with $a' \in \mathbb{R}^{n-1}$. By the implicit function theorem, there exist an open neighborhood $U \subset \mathbb{R}^{n-1}$ of $a'$ and a $C^1$ function $h : U \to \mathbb{R}$ such that, in a neighborhood of $a$,

$$M = \{x \in \mathbb{R}^n : g(x) = 0\} = \{(x', h(x')) : x' \in U\}.$$

Define

$$\Phi(x') := (x', h(x')), \qquad F(x') := f(\Phi(x')) = f(x', h(x')).$$

Since $a$ is a local extremum of $f$ subject to the constraint $g = 0$, the point $a'$ is a local extremum of $F$ on $U$. By Fermat's theorem,

$$\frac{\partial F}{\partial x_i}(a') = 0, \qquad i = 1, \dots, n-1.$$

Applying the chain rule gives

$$0 = \frac{\partial F}{\partial x_i}(a') = \frac{\partial f}{\partial x_i}(a) + \frac{\partial f}{\partial x_n}(a)\,\frac{\partial h}{\partial x_i}(a'). \tag{1}$$

On the other hand, since $g(x', h(x')) \equiv 0$ on $U$, differentiation again yields

$$0 = \frac{\partial g}{\partial x_i}(a) + \frac{\partial g}{\partial x_n}(a)\,\frac{\partial h}{\partial x_i}(a'). \tag{2}$$

Because $\partial g/\partial x_n(a) \neq 0$, equation (2) implies

$$\frac{\partial h}{\partial x_i}(a') = -\frac{\partial g/\partial x_i(a)}{\partial g/\partial x_n(a)}.$$

Substituting this into (1), we obtain

$$\frac{\partial f}{\partial x_i}(a) = \frac{\partial f/\partial x_n(a)}{\partial g/\partial x_n(a)}\,\frac{\partial g}{\partial x_i}(a), \qquad i = 1, \dots, n-1.$$

If we now set

$$\lambda := \frac{\partial f/\partial x_n(a)}{\partial g/\partial x_n(a)},$$

then the same identity is trivially true for $i = n$ as well. Therefore

$$\nabla f(a) = \lambda\, \nabla g(a),$$

as required. □

*Remark* 8.53. The condition $\nabla g(a) \neq 0$ is essential: without it, the constraint set may have a corner or self-intersection at $a$, and the geometric tangency argument can fail.

**Example 8.54** (A standard minimization problem)**.** Minimize $f(x,y) = x^2 + y^2$ subject to $g(x,y) = x + y - 1 = 0$. We have $\nabla f = (2x, 2y)$ and $\nabla g = (1,1)$. The Lagrange condition gives $(2x, 2y) = \lambda(1,1)$, hence $x = y$. Imposing $x + y = 1$ yields $x = y = \frac{1}{2}$. Thus the minimum of $x^2 + y^2$ on the line $x + y = 1$ occurs at $\left(\frac{1}{2}, \frac{1}{2}\right)$.

## Exercises

**Exercise 8.55.** Prove that if $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $a$, then $f$ is continuous at $a$.

**Exercise 8.56.** Give an example of a function $f : \mathbb{R}^2 \to \mathbb{R}$ whose partial derivatives exist everywhere but which is not differentiable at $(0,0)$.

**Exercise 8.57.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Prove that the directional derivative in direction $v$ equals $Df(a)\,v$.

**Exercise 8.58.** Prove the multivariable chain rule using linear approximation.

**Exercise 8.59.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be $C^2$. Prove that if the Hessian at $a$ is positive definite then $a$ is a strict local minimum.

**Exercise 8.60.** Compute the Jacobian determinant of the polar coordinate map $(r,\theta) \mapsto (r\cos\theta, r\sin\theta)$ and deduce the change-of-variables formula for integrals over planar regions.

**Exercise 8.61.** State and prove the Implicit Function Theorem in $\mathbb{R}^2$ (one equation, two variables) and apply it to show that the unit circle is locally the graph of a smooth function away from the points where $y = 0$.

**Exercise 8.62.** Use the Inverse Function Theorem to prove that a $C^1$ map with everywhere invertible derivative is locally open.

**Exercise 8.63.** Apply Lagrange multipliers to find the extrema of $f(x, y) = xy$ subject to $x^2 + y^2 = 1$.

**Exercise 8.64.** Let $g : \mathbb{R}^n \to \mathbb{R}$ be $C^1$ with $\nabla g(a) \neq 0$. Show that $\{x : g(x) = g(a)\}$ is locally a smooth hypersurface through $a$.

Chapter 9

# Multiple Integration and Vector Calculus

*This chapter develops the passage from one-variable integration to multiple integration and vector analysis. We begin with the Riemann theory of double and triple integrals, repeated integration, and Fubini-type principles, and we explain how geometric parametrisation leads to change-of-variables formulas. We then introduce line and surface integrals and conclude with Green's theorem, the divergence theorem, and Stokes' theorem, which unify local differential information with global integral identities.*

**Learning objectives.**

- Pass from one-variable integration to multiple integration with full control over iterated integrals and geometry of domains.
- Compute line, surface, and volume integrals while keeping track of orientation and parametrization.
- Interpret divergence, curl, and the integral theorems of vector calculus as global manifestations of local differential data.

## 9.1 A brief review of the Riemann integral in one variable

**Section overview.**

- Multiple integration grows from the one-variable Riemann theory by combining approximation with geometry.
- This short review isolates the core ideas that pass unchanged to higher dimensions.

In one variable, the Riemann integral is constructed from finite interval partitions together with upper and lower estimates on each subinterval. Since the higher-dimensional theory is obtained by refining this same idea over rectangles and boxes, we briefly recall the one-dimensional framework and the criterion based on upper and lower sums.

Let $f : [a, b] \to \mathbb{R}$ be a bounded function and let $P = \{x_0, x_1, \dots, x_n\}$ be a partition of $[a, b]$, where $\{a = x_0 < x_1 < \cdots < x_n = b\}$.

Let $\Delta x_i = x_i - x_{i-1}$, and define

$$m_i \coloneqq \inf\{f(x) : x_{i-1} \le x \le x_i\}, \qquad M_i \coloneqq \sup\{f(x) : x_{i-1} \le x \le x_i\}.$$

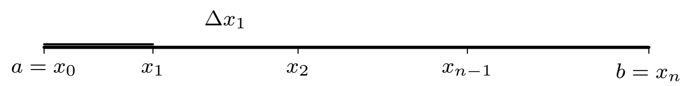



**Figure 9.1.** A partition $P$ of $[a, b]$. The mesh $\|P\| = \max_i(x_i - x_{i-1})$ measures the fineness of the partition.

The associated lower and upper sums are

$$L(P, f) := \sum_{i=1}^{n} m_i \, \Delta x_i, \qquad U(P, f) := \sum_{i=1}^{n} M_i \, \Delta x_i.$$

Since $f$ is bounded, there exist real numbers $m \le M$ such that $m \le f(x) \le M$ for all $x \in [a, b]$. Consequently,

$$m(b - a) \le L(P, f) \le U(P, f) \le M(b - a).$$

If $P_1 \subseteq P_2$, then $P_2$ is a refinement of $P_1$, and a direct verification shows that

$$U(P_1, f) \ge U(P_2, f), \qquad L(P_1, f) \le L(P_2, f).$$

Thus lower sums increase and upper sums decrease under refinement.

**Definition 9.1.** A bounded function $f : [a, b] \to \mathbb{R}$ is called *Riemann integrable* (written $f \in \mathcal{R}[a, b]$) if

$$\inf_P U(P, f) = \sup_P L(P, f).$$

Define the oscillation sum of $f$ relative to $P$ by

$$\omega(P, f) := U(P, f) - L(P, f).$$

Then $f \in \mathcal{R}[a, b]$ if and only if

$$\inf_P \omega(P, f) = 0. \tag{9.1}$$

Indeed, if $f$ is Riemann integrable, then for every $\epsilon > 0$ there exists a partition $P$ such that $\omega(P, f) < \epsilon$. Conversely, if (9.1) holds, then the lower and upper integrals coincide. In particular, in the integrable case one obtains a sequence of partitions $(P_n)$ with

$$\omega(P_n, f) < \frac{1}{n},$$

and hence $\omega(P_n, f) \to 0$.

**Theorem 9.2** (Oscillatory criterion for Riemann integrability)**.** *Let $f : [a, b] \to \mathbb{R}$ be bounded. Then $f \in \mathcal{R}[a, b]$ if and only if there exists a sequence of partitions $(P_n)$ of $[a, b]$ such that*

$$\omega(P_n, f) \longrightarrow 0.$$

*Proof.* If $f \in \mathcal{R}[a,b]$, the preceding discussion already shows that one can choose partitions $(P_n)$ with $\omega(P_n, f) < 1/n$, and hence $\omega(P_n, f) \to 0$.

Conversely, suppose that there exists a sequence of partitions $(P_n)$ with $\omega(P_n, f) \to 0$. Given $\epsilon > 0$, choose $n_0$ such that $\omega(P_{n_0}, f) < \epsilon$. Then

$$\inf_P \omega(P, f) \le \omega(P_{n_0}, f) < \epsilon.$$

Since $\epsilon > 0$ is arbitrary, we obtain $\inf_P \omega(P, f) = 0$. By (9.1), this is equivalent to Riemann integrability of $f$. $\square$

**Example 9.3.** Let $f : [0,1] \to \mathbb{R}$ be given by

$$f(x) = \begin{cases} 1, & x = \frac{1}{2}, \\ 0, & \text{otherwise.} \end{cases}$$

For the uniform partition

$$P_n = \left\{ \frac{i}{n} : 0 \le i \le n \right\},$$

the point 1/2 lies in at most two adjacent subintervals. Hence

$$\omega(P_n, f) = \sum_{i=1}^{n} (M_i - m_i)\,\Delta x_i \le \frac{2}{n} \longrightarrow 0.$$

Therefore $f \in \mathcal{R}[0,1]$.

Recall that if $P_1 \subseteq P_2$, then $U(P_1, f) \ge U(P_2, f)$ and $L(P_1, f) \le L(P_2, f)$, so $\omega(P_1, f) \ge \omega(P_2, f)$. Thus, in testing integrability via oscillation sums, it is enough to work with increasing sequences of partitions.

**Theorem 9.4** (Increasing-partition criterion for Riemann integrability)**.** *Let $f : [a,b] \to \mathbb{R}$ be bounded. Then $f \in \mathcal{R}[a,b]$ if and only if there exists an increasing sequence of partitions $(P_n)$ of $[a,b]$ such that*

$$\omega(P_n, f) \longrightarrow 0.$$

*Proof.* Assume first that $f \in \mathcal{R}[a,b]$. By Theorem 9.2, there exists a sequence of partitions $(P_n)$ such that $\omega(P_n, f) \to 0$. Define $Q_1 = P_1$ and, for $n \ge 2$,

$$Q_n \coloneqq P_1 \cup P_2 \cup \cdots \cup P_n.$$

Then $(Q_n)$ is increasing, and each $Q_n$ refines $P_n$. Therefore

$$\omega(Q_n, f) \le \omega(P_n, f) \longrightarrow 0.$$

Conversely, if there exists an increasing sequence of partitions with oscillation sums tending to 0, then Theorem 9.2 immediately yields $f \in \mathcal{R}[a,b]$. $\square$

*Remark* 9.5. Theorem 9.4 also yields

$$\lim_{n\to\infty} U(P_n, f) = \lim_{n\to\infty} L(P_n, f) = \int_a^b f(x)\,dx,$$

for every increasing sequence of partitions whose oscillatory sums tend to 0.

**Theorem 9.6** (Continuity implies Riemann integrability)**.** *If $f : [a,b] \to \mathbb{R}$ is continuous, then $f \in \mathcal{R}[a,b]$.*

*Proof.* Since $f$ is continuous on the compact interval $[a,b]$, it is bounded and uniformly continuous. Let $\epsilon > 0$. Choose $\delta > 0$ such that

$$|x-y| < \delta \quad \Longrightarrow \quad |f(x) - f(y)| < \frac{\epsilon}{b-a}.$$

Now choose a partition $P$ of $[a,b]$ with $\Delta x_i < \delta$ for every $i$. Because $f$ attains its infimum and supremum on each subinterval, we have

$$M_i - m_i \le \frac{\epsilon}{b-a} \qquad \text{for every } i.$$

Therefore

$$\omega(P,f) = \sum_{i=1}^{n} (M_i - m_i)\Delta x_i \le \sum_{i=1}^{n} \frac{\epsilon}{b-a}\,\Delta x_i = \epsilon.$$

Hence $f \in \mathcal{R}[a,b]$ by Theorem 9.2. □

**Example 9.7.** Every monotone function $f$ on $[a,b]$ is Riemann integrable. Assume $f$ is monotone increasing. Let $P_n = \left\{ x_i = a + \frac{(b-a)i}{n} \;:\; i = 0, 1, \ldots, n \right\}$. Then the oscillatory sum

$$\omega(P_n, f) = \sum_{i=1}^{n} (M_i - m_i)\Delta x_i = \sum_{i=1}^{n} \{f(x_i) - f(x_{i-1})\}\,\frac{b-a}{n} = \{f(b) - f(a)\}\,\frac{b-a}{n} \to 0.$$

Hence by Theorem 9.4 we conclude that $f \in \mathcal{R}[a,b]$.

**Continuity-type condition for Riemann integrability on** $[a,b]$**.**

Recall that the oscillatory sum $\omega(P,f)$ decreases under refinement. Moreover, a bounded function is Riemann integrable if and only if there exists a sequence of partitions $\{P_n\}$ for which $\omega(P_n, f) \to 0$. This criterion leads to a mesh-size characterization of integrability that resembles uniform continuity. For a partition $P = \{x_0, x_1, \ldots, x_n\}$ of $[a,b]$, we write

$$|P| := \max_{1\le i\le n} \Delta x_i, \qquad \Delta x_i = x_i - x_{i-1}.$$

**Theorem 9.8** (Mesh criterion for Riemann integrability)**.** *Let $f : [a,b] \to \mathbb{R}$ be bounded. Then $f \in \mathcal{R}[a,b]$ if and only if for every $\epsilon > 0$ there exists $\delta > 0$ such that*

$$|P| < \delta \quad \Longrightarrow \quad \omega(P,f) < \epsilon$$

*for every partition $P$ of $[a, b]$.*

*Proof.* Assume first that $f \in \mathcal{R}[a, b]$, and fix $\epsilon > 0$. Choose a partition

$$Q: \; a = t_0 < t_1 < \cdots < t_m = b$$

such that

$$\omega(Q, f) < \frac{\epsilon}{2}.$$

Let

$$C := \sup_{x \in [a,b]} f(x) - \inf_{x \in [a,b]} f(x),$$

which is finite because $f$ is bounded. Choose $\delta > 0$ so that

$$\delta < \min_{1 \le k \le m} (t_k - t_{k-1}) \qquad \text{and} \qquad C(m-1)\delta < \frac{\epsilon}{2}.$$

Now let $P$ be any partition of $[a, b]$ with $|P| < \delta$.

Split the subintervals of $P$ into two classes. An interval of $P$ is *good* if it is contained in one of the intervals $[t_{k-1}, t_k]$ of $Q$, and *bad* otherwise. Because $|P| < \delta < \min_k(t_k - t_{k-1})$, each bad interval contains exactly one interior partition point $t_k$ of $Q$. Hence there are at most $m - 1$ bad intervals, and their total length is at most $(m - 1)\delta$.

For a good interval $I \subseteq [t_{k-1}, t_k]$, the oscillation of $f$ on $I$ is at most the oscillation of $f$ on $[t_{k-1}, t_k]$. Therefore the total contribution of all good intervals to $\omega(P, f)$ is at most $\omega(Q, f)$. For each bad interval $I$, the oscillation of $f$ on $I$ is at most $C$, so the total contribution of all bad intervals is at most

$$C(m-1)\delta < \frac{\epsilon}{2}.$$

Consequently,

$$\omega(P, f) \le \omega(Q, f) + C(m-1)\delta < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Conversely, suppose that for every $\epsilon > 0$ there exists $\delta > 0$ such that $|P| < \delta$ implies $\omega(P, f) < \epsilon$. Choose any partition $P$ with mesh smaller than this $\delta$. Then $\omega(P, f) < \epsilon$, and since $\epsilon > 0$ is arbitrary, Theorem 9.2 shows that $f$ is Riemann integrable. □

**Corollary 9.9** (Sequential mesh criterion for Riemann integrability)**.** *Let $f : [a, b] \to \mathbb{R}$ be bounded. Then $f \in \mathcal{R}[a, b]$ if and only if for every sequence of partitions $(P_n)$ with $|P_n| \to 0$, one has*

$$\omega(P_n, f) \to 0.$$

**Question.** How far can a Riemann integrable function be from being continuous?

## 9.2 Double integrals

**Section overview.**

- The passage from one variable to two variables introduces geometry of domains as an essential part of integration.
- Iterated integrals are not merely computational tools; they encode the structure of the domain through slicing.
- Sketching the region before computing is part of the proof-oriented method, not an optional extra.

Recall that, in one variable, the Riemann integral of a non-negative function on a finite interval represents the area under its graph. In two variables, the analogous quantity is volume: if $f(x,y) \geq 0$ on a planar region $D$, then the double integral measures the volume of the solid lying under the surface $z = f(x,y)$ and above $D$.

We begin with rectangular domains, where the definition is closest to the one-variable theory, and only afterwards pass to more general regions with curved boundaries.

Let $D = [a,b] \times [c,d]$, and let $f : D \to \mathbb{R}$ be bounded. Let $P_1 = \{x_0, x_1, \ldots, x_n\}$ be a partition of $[a,b]$ and $P_2 = \{y_0, y_1, \ldots, y_m\}$ a partition of $[c,d]$. Their product partition $P = P_1 \times P_2$ decomposes $D$ into the rectangles

$$D_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j], \qquad 1 \leq i \leq n,\ 1 \leq j \leq m.$$

For each cell $D_{ij}$, set

$$m_{ij} := \inf\{f(x,y) : (x,y) \in D_{ij}\}, \qquad M_{ij} := \sup\{f(x,y) : (x,y) \in D_{ij}\}.$$

The lower and upper sums associated with $P$ are then

$$L(P,f) = \sum_{i=1}^{n} \sum_{j=1}^{m} m_{ij}\, \Delta x_i\, \Delta y_j, \qquad U(P,f) = \sum_{i=1}^{n} \sum_{j=1}^{m} M_{ij}\, \Delta x_i\, \Delta y_j.$$

The lower integral is defined by $\sup_P L(P,f)$ and the upper integral by $\inf_P U(P,f)$. Both quantities exist because $f$ is bounded. We say that $f$ is Riemann integrable on $D$ (written $f \in \mathcal{R}(D)$) if these two numbers coincide. Their common value is denoted by

$$\iint\limits_D f(x,y)\, dx\, dy \qquad \text{or} \qquad \iint\limits_D f(x,y)\, dA.$$

**Example 9.10.** Let $f : D = [0,1] \times [0,1] \to \mathbb{R}$ be given by

$$f(x,y) = \begin{cases} 1 & \text{if } x, y \in \mathbb{Q} \cap [0,1], \\ 0 & \text{otherwise.} \end{cases}$$

Then $f$ is not integrable on $D$, because for any partition $P$ of $D$ defined as above, we get $U(P,f) = 1 \neq 0 = L(P,f)$.

**Theorem 9.11** (Oscillatory criterion for double integrability)**.** *Let $f : D = [a, b] \times [c, d] \to \mathbb{R}$ be bounded. Then $f \in \mathcal{R}(D)$ if and only if for each $\epsilon > 0$ there exists a partition $P$ of $D$ such that $\omega(P, f) = U(P, f) - L(P, f) < \epsilon$.*

**Theorem 9.12** (Increasing-partition criterion for double integrability)**.** *Let $f : D = [a, b] \times [c, d] \to \mathbb{R}$ be bounded. Then $f \in \mathcal{R}(D)$ if and only if there exists an increasing sequence of partitions $\{P_n\}$ of $D$ such that*

$$\omega(P_n, f) \longrightarrow 0.$$

*Proof.* Assume first that $f \in \mathcal{R}(D)$. By Theorem 9.11, for each $n \in \mathbb{N}$ there exists a partition $Q_n$ of $D$ such that

$$\omega(Q_n, f) < \frac{1}{n}.$$

Define

$$P_n := Q_1 \cup Q_2 \cup \cdots \cup Q_n.$$

Then $(P_n)$ is an increasing sequence of partitions, and each $P_n$ refines $Q_n$. Since oscillation sums decrease under refinement, we obtain

$$\omega(P_n, f) \leq \omega(Q_n, f) < \frac{1}{n}.$$

Hence $\omega(P_n, f) \to 0$.

Conversely, suppose there exists an increasing sequence of partitions $(P_n)$ such that $\omega(P_n, f) \to 0$. Given $\epsilon > 0$, choose $n$ large enough that $\omega(P_n, f) < \epsilon$. By Theorem 9.11, this implies that $f \in \mathcal{R}(D)$. □

**Example 9.13.** Let $f : D = [0, 1] \times [0, 1] \to \mathbb{R}$ be given by

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

Then $\iint\limits_D f(x, y)dxdy = 0$. Let $P_n = \{\frac{i}{n} : \ i = 0, 1, \ldots, n\} \times \{\frac{i}{n} : \ i = 0, 1, \ldots, n\}$. In this case, $\Delta x_i = \Delta y_j = \frac{1}{n}$. The oscillatory sum of the function $f$ on $D$ satisfies

$$\omega(P_n, f) = \sum_{i=1}^{n}\sum_{j=1}^{n}(M_{ij} - m_{ij})\Delta x_i \Delta y_j = \sum_{i=1}^{n}\sum_{j=1}^{n}(M_{ij} - 0)\frac{1}{n^2} = \sum_{i=j,\ j=1}^{n} 1.\frac{1}{n^2} = \frac{1}{n} \to 0.$$

**Theorem 9.14** (Continuity implies integrability on rectangles)**.** *Let $D = [a, b] \times [c, d]$. If $f : D \to \mathbb{R}$ is continuous, then $f$ is integrable on $D$.*

*Proof.* Since $f$ is continuous on the compact rectangle $D$, it is bounded and uniformly continuous. Let $A$ denote the area of $D$. Given $\epsilon > 0$, choose $\delta > 0$ such that

$$\sqrt{(x - x')^2 + (y - y')^2} < \delta \quad \Longrightarrow \quad |f(x, y) - f(x', y')| < \frac{\epsilon}{A}$$

for all $(x,y),(x',y') \in D$. Now take a partition $P = \{D_{ij} : 1 \le i \le n,\ 1 \le j \le m\}$ with

$$D_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j], \qquad \operatorname{diam}(D_{ij}) < \delta$$

for every $i, j$. Since $f$ attains its infimum and supremum on each closed cell $D_{ij}$, we obtain

$$M_{ij} - m_{ij} \le \frac{\epsilon}{A}.$$

Therefore

$$\omega(P,f) = \sum_{i=1}^{n}\sum_{j=1}^{m}(M_{ij} - m_{ij})\,\Delta x_i\,\Delta y_j \le \frac{\epsilon}{A}\sum_{i=1}^{n}\sum_{j=1}^{m}\Delta x_i\,\Delta y_j = \epsilon.$$

By Theorem 9.12, it follows that $f \in \mathcal{R}(D)$. □

**A mesh-size criterion for integrability.**

As in one variable, there is a useful characterization of Riemann integrability in terms of sufficiently fine partitions. If

$$P = \{D_{ij} : 1 \le i \le n,\ 1 \le j \le m\}$$

is a partition of $D = [a,b] \times [c,d]$, where $D_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j]$, we define

$$\operatorname{diam}(D_{ij}) = \sqrt{(x_i - x_{i-1})^2 + (y_j - y_{j-1})^2}$$

and set

$$|P| := \max\{\operatorname{diam}(D_{ij}) : 1 \le i \le n,\ 1 \le j \le m\}.$$

**Theorem 9.15** (Mesh criterion for double integrability)**.** *Let $f : D = [a,b] \times [c,d] \to \mathbb{R}$ be a bounded function. Then $f \in \mathcal{R}(D)$ if and only if for each $\epsilon > 0$ there exists $\delta > 0$ such that for each partition $P$ of $D$ with $|P| < \delta$ implies $\omega(P,f) < \epsilon$.*

*Proof.* Assume first that $f \in \mathcal{R}(D)$, and fix $\epsilon > 0$. Choose a partition

$$Q = \{R_{ij} : 1 \le i \le m,\ 1 \le j \le n\}$$

of $D$ such that

$$\omega(Q,f) < \frac{\epsilon}{2}.$$

Let

$$C := \sup_D f - \inf_D f,$$

which is finite because $f$ is bounded. Write

$$R_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j],$$

and set

$$\eta_x := \min_{1 \le i \le m}(x_i - x_{i-1}), \qquad \eta_y := \min_{1 \le j \le n}(y_j - y_{j-1}).$$

Choose $\delta > 0$ such that

$$\delta < \min\{\eta_x, \eta_y\} \qquad \text{and} \qquad C(m+n)\,\mathrm{area}(D)\,\delta < \frac{\epsilon}{2}.$$

Now let $P$ be any partition of $D$ with $|P| < \delta$.

Call a rectangle of $P$ *good* if it is contained in one of the rectangles $R_{ij}$ of $Q$, and *bad* otherwise. Because $|P| < \delta < \min\{\eta_x, \eta_y\}$, a bad rectangle of $P$ can meet at most one vertical grid line $x = x_i$ from $Q$ and at most one horizontal grid line $y = y_j$ from $Q$ in its interior. Hence every bad rectangle is contained in the union of

- a vertical strip of width at most $\delta$ around one of the $m-1$ interior vertical grid lines of $Q$, and
- a horizontal strip of height at most $\delta$ around one of the $n-1$ interior horizontal grid lines of $Q$.

Therefore the union of all bad rectangles has area at most

$$(m-1)(d-c)\,\delta + (n-1)(b-a)\,\delta \le (m+n)\,\mathrm{area}(D)\,\delta.$$

For good rectangles, the oscillation of $f$ is dominated by the oscillation of $f$ on the containing rectangle of $Q$, so the total contribution of all good rectangles to $\omega(P, f)$ is at most $\omega(Q, f)$. For bad rectangles, the oscillation of $f$ is at most $C$, hence their total contribution is at most

$$C(m+n)\,\mathrm{area}(D)\,\delta < \frac{\epsilon}{2}.$$

Consequently,

$$\omega(P, f) \le \omega(Q, f) + C(m+n)\,\mathrm{area}(D)\,\delta < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Conversely, suppose that for every $\epsilon > 0$ there exists $\delta > 0$ such that $|P| < \delta$ implies $\omega(P, f) < \epsilon$. Choose any partition $P$ of $D$ with $|P| < \delta$. Then $\omega(P, f) < \epsilon$, and since $\epsilon > 0$ is arbitrary, Theorem 9.11 implies that $f \in \mathcal{R}(D)$. ☐

**Corollary 9.16** (Sequential mesh criterion for double integrability)**.** *Let $f : D = [a, b] \times [c, d] \to \mathbb{R}$ be a bounded function. Then $f \in \mathcal{R}(D)$ if and only if for each sequence of partitions $\{P_n\}$ of $D$ with $|P_n| \to 0$ implies $\omega(P_n, f) \to 0$.*

**Note that** in order to show $f \notin \mathcal{R}(D)$, it is enough to show that there exists a sequence of partitions $\{P_n\}$ with $|P_n| \to 0$ but $\omega(P_n, f) \not\to 0$.

**Geometric interpretation.** If $f : D = [a, b] \times [c, d] \to [0, \infty)$ is integrable, then

$$\iint_D f(x, y)\, dx\, dy$$

represents the volume of the solid bounded by the planes $x = a$, $x = b$, $y = c$, $y = d$, and the surface $z = f(x, y)$.

### 9.2.1 Repeated Integrals

The next result shows that certain double integrals can be evaluated by repeated integration. This is the content of Fubini's theorem. Before stating the theorem, we examine two instructive examples.

**Example 9.17.** Consider the function $f : D = [0,1] \times [0,1] \to \mathbb{R}$ defined by

$$f(x,y) = \begin{cases} 1, & \text{if } x \in \mathbb{Q} \cap [0,1], \\ 2y, & \text{if } x \in \mathbb{Q}^c \cap [0,1]. \end{cases}$$

Then

$$\int_0^1 \left( \int_0^1 f(x,y)\, dy \right) dx = 1.$$

However, $f$ is not integrable on $D$. (**Hint:** Use Corollary 9.16 to deduce that $f \notin \mathcal{R}(D)$.)

**Example 9.18.** Consider the function $f : D = [0,1] \times [0,1] \to \mathbb{R}$ defined by

$$f(x,y) = \begin{cases} 0, & \text{if } x \neq \frac{1}{2}, \\ 1, & \text{if } x = \frac{1}{2} \text{ and } y \in \mathbb{Q} \cap [0,1], \\ -1, & \text{if } x = \frac{1}{2} \text{ and } y \in \mathbb{Q}^c \cap [0,1]. \end{cases}$$

For $x = \frac{1}{2}$, the inner integral $\int_0^1 f(x,y)\, dy$ does not exist. Nevertheless, the double integral $\iint\limits_D f(x,y)\, dx\, dy$ does exist.

**Theorem 9.19** (Fubini's Theorem)**.** *Let $f : D = [a,b] \times [c,d] \to [0,\infty)$ be integrable. If for each $y \in [c,d]$, the function $f(\cdot, y) \in \mathcal{R}[a,b]$, then the function $F$ defined by $F(y) = \int\limits_a^b f(x,y)dx$ is integrable on $[c,d]$ and*

$$\iint\limits_D f(x,y)dxdy = \int_c^d \left( \int_a^b f(x,y)dx \right) dy.$$

*Proof.* Since $f \in \mathcal{R}(D)$, for each $\epsilon > 0$ there exists a partition

$$P = P_1 \times P_2 = \{D_{ij} :\ i = 1,2,\ldots,n \text{ and } j = 1,2,\ldots,m\}$$

of $D$ such that $U(P,f) - L(P,f) < \epsilon$. Recall that $m_{ij} = \inf\{f(x,y) :\ (x,y) \in D_{ij}\}$ and $M_{ij} = \sup\{f(x,y) :\ (x,y) \in D_{ij}\}$. Let us define $k_j = \inf\{F(y) :\ y_{j-1} \le y \le y_j\}$ and $K_j = \sup\{F(y) :\ y_{j-1} \le y \le y_j\}$. Since $m_{ij} \le f(x,y) \le M_{ij}$ for each $(x,y) \in D_{ij}$, it follows that

$$\sum_{i=1}^n m_{ij} \Delta x_i \le L(P_1, f(\cdot,y)) \le \int_a^b f(x,y)dx = F(y) \le \sum_{i=1}^n M_{ij} \Delta x_i \tag{9.2}$$

for each $y \in [y_{j-1}, y_j]$. Note the **first inequality** in (9.2) follows due to the fact that infimum $m_{ij}$ of $f$ on $D_{ij}$ is smaller than the infimum of $f$ over $[x_{i-1}, x_i] \times \{y\}$.

From the above it follows that

$$L(P, f) = \sum_{i=1}^{n}\sum_{j=1}^{m} m_{ij}\Delta x_i \Delta y_j \leq \sum_{j=1}^{n} k_j \Delta y_j = L(P_2, F) \leq U(P_2, F)$$

and

$$U(P_2, F) = \sum_{j=1}^{n} K_j \Delta y_j \leq \sum_{i=1}^{n}\sum_{j=1}^{m} M_{ij}\Delta x_i \Delta y_j = U(P, f).$$

Hence

$$L(P, f) \leq L(P_2, F) \leq U(P_2, F) \leq U(P, f). \tag{9.3}$$

Since $U(P, f) - L(P, f) < \epsilon$, from (9.3) we get $U(P_2, F) - L(P_2, F) < \epsilon$. That is, $F \in \mathcal{R}[c, d]$, and hence once again from (9.3) we infer that

$$L(P, f) \leq \int_c^d F(y)dy \leq U(P, f) \text{ and } L(P, f) \leq \iint\limits_D f(x, y)dxdy \leq U(P, f).$$

Thus,

$$-\epsilon < \iint\limits_D f(x, y)dxdy - \int_c^d F(y)dy \leq \epsilon$$

for each $\epsilon > 0$. Hence

$$\iint\limits_D f(x, y)dxdy = \int_c^d F(y)dy.$$

This completes the proof. $\square$

If we instead define $G(x) = \int\limits_c^d f(x, y)\, dy$, then an analogous statement also holds.

**Corollary 9.20** (Fubini's Theorem)**.** *Let $f : D = [a, b] \times [c, d] \to \mathbb{R}$ be a continuous function. Then*

$$\iint\limits_D f(x, y)dxdy = \int_c^d \left(\int_a^b f(x, y)dx\right) dy = \int_a^b \left(\int_c^d f(x, y)dy\right) dx.$$

**Example 9.21.** Let $f(x, y) = xe^{xy}$ for $(x, y) \in D = [0, 2] \times [0, 1]$. Then $f$ is continuous and hence by Fubini's theorem

$$\iint\limits_D f(x, y)dxdy = \int_0^2 \left(\int_0^1 xe^{xy}dy\right) dx = \int_0^2 [e^{xy}]_0^1\, dx = \int_0^2 (e^x - 1)dx = e^2 - 3.$$

### 9.2.2 Bounded functions with discontinuities

We know from Theorem 9.14 that if $f$ is continuous on $D$ then $f$ is integrable. In this section, we discuss that the integral of a function $f$ also exists if the set of discontinuities of $f$ is not too large. In order to measure discontinuities, we introduce the following concept.

**Definition 9.22.** Let $A$ be a bounded subset of $\mathbb{R}^2$. We say that $A$ has *content zero* if, for every $\epsilon > 0$, there exist finitely many rectangles $\{R_i\}_{i=1}^n$ such that

$$A \subseteq \bigcup_{i=1}^n R_i \qquad \text{and} \qquad \text{Area}\left(\bigcup_{i=1}^n R_i\right) < \epsilon.$$

**Example 9.23.** (i) Any finite set of points in $\mathbb{R}^2$ has content zero.

(ii) Every subset of a set of content zero has content zero.

(iii) The union of finitely many bounded sets of content zero is also of content zero.

(iv) Every line segment has content zero.

**Exercise 9.24.** Any bounded subset of $\mathbb{R}^2$ having nonempty interior cannot have content zero.

**Theorem 9.25** (Discontinuities of content zero imply integrability)**.** *Let $f : D = [a,b]\times[c,d] \to \mathbb{R}$ be a bounded function. If the set of discontinuities of $f$ in $D$ is a set of content zero, then $f$ is integrable.*

*Proof.* Let $M > 0$ be such that $|f(x,y)| \leq M$ for all $(x,y) \in D$. Suppose $E$ is the set of discontinuities of $f$ in $D$. Let

$$P = \{D_i : D_i \text{ is a subrectangle of } D\}$$

be a partition of $D$, and write

$$m_i = \inf_{D_i} f, \qquad M_i = \sup_{D_i} f, \qquad A(D_i) = \text{Area}(D_i).$$

Since $E$ has content zero, we can choose finitely many subrectangles $D_1, \ldots, D_m$ such that

$$E \subset \bigcup_{i=1}^m D_i \qquad \text{and} \qquad \sum_{i=1}^m A(D_i) < \frac{\epsilon}{4M}.$$

On each of the remaining closed subrectangles $D_i$ $(i = m+1, \ldots, n)$, the function $f$ is continuous and hence uniformly continuous. Therefore, by refining the partition if necessary, we may arrange that

$$M_i - m_i \leq \frac{\epsilon}{2A(D)} \qquad (i = m+1, \ldots, n).$$

Hence

$$\begin{aligned}
\omega(P,f) &= \sum_{i=1}^{n}(M_i-m_i)A(D_i)\\
&= \sum_{i=1}^{m}(M_i-m_i)A(D_i)+\sum_{i=m+1}^{n}(M_i-m_i)A(D_i)\\
&\le \sum_{i=1}^{m}2MA(D_i)+\sum_{i=m+1}^{n}\frac{\epsilon}{2A(D)}A(D_i)\\
&< 2M\frac{\epsilon}{4M}+\frac{\epsilon}{2}\frac{A(D)}{A(D)}=\epsilon.
\end{aligned}$$

Thus, for each $\epsilon > 0$ we have constructed a partition $P$ of $D$ such that $\omega(P,f) < \epsilon$. This implies $f \in \mathcal{R}(D)$. □

### 9.2.3 Double integral over general bounded regions

Let $D$ be a bounded region in $\mathbb{R}^2$, and let $f : D \to \mathbb{R}$ be bounded. Choose a rectangle $Q$ with $D \subseteq Q$, and extend $f$ to a function $\tilde{f} : Q \to \mathbb{R}$ by

$$\tilde{f}(x,y)=\begin{cases} f(x,y), & \text{if } (x,y)\in D,\\ 0, & \text{if } (x,y)\in Q\setminus D.\end{cases}$$

If $\tilde{f}$ is integrable over $Q$, then we say that $f$ is integrable over $D$ and define

$$\iint\limits_D f(x,y)\,dx\,dy=\iint\limits_Q \tilde{f}(x,y)\,dx\,dy.$$

**Theorem 9.26** (Fubini's theorem)**.** *Let $f$ be a bounded continuous function over a bounded region $D$ in $\mathbb{R}^2$.*

*(i) If $D = \{(x,y) : a \le x \le b$ and $f_1(x) \le y \le f_2(x)\}$ for some continuous functions $f_1, f_2 : [a,b] \to \mathbb{R}$, then*

$$\iint\limits_D f(x,y)dxdy=\int_a^b\left(\int_{f_1(x)}^{f_2(x)} f(x,y)dy\right)dx.$$

*(ii) If $D = \{(x,y) : c \le y \le d$ and $g_1(y) \le x \le g_2(y)\}$ for some continuous functions $g_1, g_2 : [c,d] \to \mathbb{R}$, then*

$$\iint\limits_D f(x,y)dxdy=\int_c^d\left(\int_{g_1(y)}^{g_2(y)} f(x,y)dx\right)dy.$$

For a proof of Theorem 9.26, we refer to Chapter 11 of Apostol's *Calculus, Vol. II.*

**Example 9.27.** $(i)$ Let $D$ be the region bounded by the lines joining the points $(0,0),(0,1)$ and $(2,2)$. Evaluate the integral $\iint\limits_D (x+y)^2dxdy$.

$(ii)$ Evaluate the integral $\int\limits_0^2 \left( \int\limits_{\frac{y}{2}}^1 e^{x^2} dx \right) dy.$

Riemann integrable functions on $D$ satisfy the following basic algebraic relations.

**Theorem 9.28** (Algebraic properties of double integrals)**.** *Let $f$ and $g$ be Riemann integrable functions on the region $D$ in the plane and $c \in \mathbb{R}$. Then*

(i) $cf + g \in R(D),\ \iint\limits_D \{cf(x,y) + g(x,y)\} dxdy = c \iint\limits_D f(x,y) dxdy + \iint\limits_D g(x,y) dxdy.$

(ii) *If $f(x,y) \leq g(x,y)$ for all $(x,y) \in D$, then* $\iint\limits_D f(x,y) dxdy \leq \iint\limits_D g(x,y) dxdy.$

(iii) $|f| \in \mathcal{R}(D)$ *and* $\left| \iint\limits_D f(x,y) dxdy \right| \leq \iint\limits_D |f(x,y)| dxdy.$

### 9.2.4 Change of variable

The change-of-variables formula is one of the central tools of multivariable calculus. Many regions and integrands become substantially simpler when expressed in coordinates adapted to the geometry of the problem. The purpose of the formula is to make this change precise while keeping track of the distortion of area or volume through the Jacobian.

### 9.2.5 Change of variable in single integral

Let $f : [a,b] \to \mathbb{R}$ be an integrable function. Suppose $g : [c,d] \to [a,b]$ is continuously differentiable function (that is, a $C^1$ function) such that $g'(t) \neq 0$ for all $t \in (c,d)$. Then $g$ is one-to-one (by the Mean Value Theorem) and hence monotone. We also assume that $g$ is surjective. Put $x = g(t)$. Then $dx = g'(t)dt$. If $g$ is monotone increasing, then

$$\int_a^b f(x)dx = \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(t))g'(t)dt = \int_c^d f(g(t))g'(t)dt.$$

In case, if $g$ is monotone decreasing, then $[c,d] = [g^{-1}(b), g^{-1}(a)]$. Thus, we have the formula

$$\int_a^b f(x)dx = \int_c^d f(g(t))|g'(t)|dt.$$

### 9.2.6 Change of variable in double integral

We now extend the one-variable change-of-variables principle to the setting of double integrals. A complete proof in higher dimensions requires additional material from advanced analysis and differential calculus, so we state the result here and focus on its interpretation and use.

Suppose $S$ is a bounded region in the $uv$-plane and that it is mapped onto a bounded region $D$ in the $xy$-plane by $x = \varphi(u,v)$ and $y = \psi(u,v)$. Consider the associated map $T : D \to S$, assumed to be bijective and continuously differentiable, whose inverse is given by $T^{-1}(u,v) = (\varphi(u,v), \psi(u,v))$. See fig. 9.2.

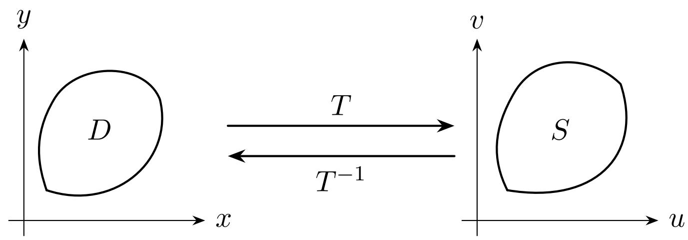



**Figure 9.2.** A bijective $C^1$ change of variables $T : D \to S$ and its inverse.

Assume further that $T^{-1}$ is also continuously differentiable and that its derivative $(T^{-1})'$ is invertible (that is, nonsingular) on the interior of $S$. Then

$$\det(T^{-1})' = \begin{vmatrix} \frac{\partial\varphi}{\partial u} & \frac{\partial\varphi}{\partial v} \\ \frac{\partial\psi}{\partial u} & \frac{\partial\psi}{\partial v} \end{vmatrix} = J(u,v) \neq 0 \ ,$$

where the function $J$ is the Jacobian determinant of the transformation. In this way, a function $f(x, y)$ defined on $D$ may be regarded as the composite function $f(\varphi(u, v), \psi(u, v))$ on $S$.

**Theorem 9.29** (Change of variables for double integrals)**.** *Let $f : D \to \mathbb{R}$ be a continuous function and let $J$ be as above. Then the integrals of $f$ over $D$ and of $f(\varphi(u, v), \psi(u, v))$ over $S$ are related by*

$$\iint\limits_D f(x,y)\,dx\,dy = \iint\limits_S f[\varphi(u,v),\psi(u,v)]\,|J(u,v)|\,du\,dv.$$

**Example 9.30.** Find the area of the region $D$ bounded by the hyperbolas $xy = 1$ and $xy = 2$ and the curves $xy^2 = 3$ and $xy^2 = 4$.
Note that the area of the region is given by $\iint\limits_D dxdy$. Let $u = xy$ and $v = xy^2$. Then $x = \frac{u^2}{v}$ and $y = \frac{v}{u}$. Also $J(u, v) = \frac{1}{v}$. Thus, $\iint\limits_D dxdy = \int\limits_{u=1}^{2} \int\limits_{v=3}^{4} \frac{1}{v} dvdu = \log(\frac{4}{3})$.

**Example 9.31.** Evaluate the double integral $\iint\limits_D \frac{(x-y)}{(x+y+2)^2} dxdy$ over the region $D$ bounded by the lines $x + y = \pm 1$ and $x - y = \pm 1$.
Let $u = x + y$ and $v = x - y$. Then $x = \frac{u+v}{2}$ and $y = \frac{u-v}{2}$. Here $J(u, v) = -\frac{1}{2}$. Also, $u = \pm 1$ and $v = \pm 1$. Hence we have

$$\iint\limits_D \frac{(x-y)}{(x+y+2)^2} dxdy = \frac{1}{2} \int_{u=-1}^{1} \int_{v=-1}^{1} \frac{v}{(u+2)^2} dudv.$$

**Polar coordinates:** In this case the variables $x$ and $y$ are changed to $r$ and $\theta$ by the following two equations

$$x = r\cos\theta \text{ and } y = r\sin\theta.$$

We assume that $r > 0$ and $\theta$ lies in $[0, 2\pi)$ so that the mapping $T^{-1}(r,\theta) = (r\cos\theta, r\sin\theta)$ is bijective. Then

$$J(r,\theta) = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r.$$

Hence the change of variable formula in this case is

$$\iint_D f(x,y)dxdy = \iint_S f(r\cos\theta, r\sin\theta)\, r\, drd\theta.$$

**Example 9.32.** Let $D = \{(x,y) : x^2 + y^2 \leq a^2\}$ and $f : D \to \mathbb{R}$ is given by $f(x,y) = 2\sqrt{a^2 - x^2 - y^2}$. Then

$$\begin{aligned} \iint_D f(x,y)dxdy &= 2\int_0^a \int_0^{2\pi} \sqrt{a^2 - r^2} r d\theta dr \\ &= 4\pi \int_0^a r\sqrt{a^2 - r^2} dr \\ &= 4\pi \left. \frac{(a^2 - r^2)^{\frac{3}{2}}}{-3} \right|_0^a = \frac{4\pi a^3}{3}. \end{aligned}$$

## 9.3 Triple integrals

**Section overview.**

- Triple integrals extend planar integration to volume computations and three-dimensional accumulation problems.
- The geometry of the domain becomes even more important, especially when one chooses coordinates adapted to symmetry.

The definition of the double integral extends in a natural way to bounded functions on a rectangular box $D = [a,b] \times [c,d] \times [e,f] \subset \mathbb{R}^3$. One now uses product partitions of the form $P = P_1 \times P_2 \times P_3$, where $P_1$, $P_2$, and $P_3$ are partitions of $[a,b]$, $[c,d]$, and $[e,f]$, respectively. For a bounded function $f$ on $D$, one defines lower and upper sums exactly as before; if the corresponding lower and upper integrals agree, then $f$ is said to be integrable on $D$, and the common value is called the *triple integral* of $f$ over $D$. We denote it by

$$\iiint_D f(x,y,z)\, dx\, dy\, dz \quad \text{or} \quad \iiint_D f(x,y,z)\, dV.$$

*Remark* 9.33. Most of the basic integrability criteria developed for functions of two variables extend, with only notational changes, to the three-variable setting. We therefore emphasize here the new geometric features rather than repeating every earlier argument in detail.

**Theorem 9.34** (Fubini's theorem)**.** *Let $R$ be a bounded region in $\mathbb{R}^2$ and let $D$ be a bounded domain in $\mathbb{R}^3$ given by $D = \{(x,y,z) : (x,y) \in R \quad \text{and} \quad f_1(x,y) \le z \le f_2(x,y)\}$, where $f_1, f_2$ are continuous functions on $R$. If $f$ is continuous on $D$, then*

$$\iiint\limits_D f(x,y,z)dV = \iint\limits_R \left( \int_{f_1(x,y)}^{f_2(x,y)} f(x,y,z)dz \right) dA.$$

### 9.3.1 Change of variable in a triple integral

The change-of-variables formula for double integrals extends verbatim to the three-dimensional setting.

$$\iiint\limits_S f(x,y,z)dxdydz = \iiint\limits_T f[\varphi(u,v,w), \psi(u,v,w), \eta(u,v,w)]|J(u,v,w)|dudvdw,$$

where

$$J(u,v,w) = \begin{vmatrix} \frac{\partial \varphi}{\partial u} & \frac{\partial \varphi}{\partial v} & \frac{\partial \varphi}{\partial w} \\ \frac{\partial \psi}{\partial u} & \frac{\partial \psi}{\partial v} & \frac{\partial \psi}{\partial w} \\ \frac{\partial \eta}{\partial u} & \frac{\partial \eta}{\partial v} & \frac{\partial \eta}{\partial w} \end{vmatrix}.$$

**Example 9.35.** Consider the integral $\iiint\limits_D x\,dxdydz$, where $D$ is the region in $\mathbb{R}^3$ bounded by $x = 0, y = 0, z = 2$ and the surface $z = x^2 + y^2$. Here $D = \{(x,y,z) : (x,y) \in R, x^2 + y^2 \le z \le 2\}$ and $R = \left\{(x,y) : 0 \le x \le \sqrt{2}, 0 \le y \le \sqrt{2 - x^2}\right\}$. Therefore

$$\begin{aligned} \iiint\limits_D xdxdydz &= \iint\limits_R \left( \int_{x^2+y^2}^{2} x\,dz \right) dA \\ &= \int_0^{\sqrt{2}} \int_0^{\sqrt{2-x^2}} \int_{x^2+y^2}^{2} x\,dzdydx \\ &= \frac{8\sqrt{2}}{15}. \end{aligned}$$

**Cylindrical co-ordinates:** In this case the variables $x, y$ and $z$ are changed to $r, \theta$ and $z$ by the following three equations

$$x = r\cos\theta, y = r\sin\theta \quad \text{and} \quad z = z,$$

where $r > 0$ and $\theta \in [0, 2\pi)$. The Jacobian is

$$J(r,\theta,z) = \begin{vmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{vmatrix} = r.$$

Therefore the change of variable formula is

$$\iiint\limits_S f(x,y,z)dxdydz = \iiint\limits_T f(r\cos\theta, r\sin\theta, z)\, r\, drd\theta dz.$$

**Example 9.36.** Consider $\iiint\limits_D \left(z^2x^2 + z^2y^2\right) dxdydz$, where $D$ is the region determined by $x^2 + y^2 \le 1$ and $-1 \le z \le 1$. We can describe $D$ in cylindrical coordinates by $0 \le r \le 1$, $0 \le \theta \le 2\pi$ and $-1 \le z \le 1$. Therefore,

$$\begin{aligned}\iiint\limits_D \left(z^2x^2 + z^2y^2\right) dxdydz &= \int_{-1}^{1}\int_0^{2\pi}\int_0^1 \left(z^2r^2\right) r\, drd\theta dz \\ &= \int_{-1}^{1}\int_0^{2\pi} z^2\frac{r^4}{4}\bigg|_{r=0}^{1} d\theta dz \\ &= \int_{-1}^{1}\frac{2\pi}{4}z^2dz = \frac{\pi}{3}.\end{aligned}$$

**Spherical co-ordinates:** In this case the variables $x, y$ and $z$ are changed to $r, \theta$ and $\phi$ by the following three equations

$$x = r\sin\phi\cos\theta, \quad y = r\sin\phi\sin\theta, \quad z = r\cos\phi.$$

We assume $r > 0, 0 \le \theta < 2\pi$ and $0 \le \phi < \pi$ to get mapping of transformation one-to-one. The Jacobian is

$$J(r,\theta,\phi) = -r^2\sin\phi.$$

Hence the change of variable formula is

$$\iiint\limits_S f(x,y,z)dxdydz = \iiint\limits_T f(r\sin\phi\cos\theta, r\sin\phi\sin\theta, r\cos\phi)\, r^2\sin\phi\, drd\theta d\phi.$$

## 9.4 Surface area and surface integrals

**Section overview.**

- Curved domains are naturally described by parametrizations rather than by rectangular coordinates.
- Surface integrals combine the geometry of the surface with scalar or vector quantities distributed on it.

A surface is a two-dimensional geometric object in $\mathbb{R}^3$. From the analytic point of view, the most effective language for studying surfaces is that of parametrizations, because a parametrization describes geometry and integration simultaneously: tangent vectors, area elements, and normal directions all arise from the same map.

**Parametric surface.** A parametric surface is the image of a map of two variables with values in $\mathbb{R}^3$. More precisely, if

$$r : T \subseteq \mathbb{R}^2 \to \mathbb{R}^3, \qquad r(u,v) = f(u,v)i + g(u,v)j + h(u,v)k,$$

then the set $r(T) = \{r(u,v) : (u,v) \in T\}$ is called the *parametric surface* determined by $r$. We assume that $r$ is one-to-one on the interior of $T$, so that the surface has no self-intersections there. The equations

$$x = f(u,v), \qquad y = g(u,v), \qquad z = h(u,v), \qquad (u,v) \in T,$$

are the parametric equations of the surface $r(T)$.

**Example 9.37.** $(i)$. Let $T = \{(x,y) : x^2 + y^2 \leq 1\}$. Consider the function $r : T \to \mathbb{R}^3$ given by $r(x,y) = xi + yj + \sqrt{x^2+y^2}k$. This represents a cone of height 1.

$(ii)$. For fixed $a > 0$ with $0 \leq \theta < 2\pi$ and $0 \leq \phi < \pi$, the equations

$$x = a\sin\phi\cos\theta, \quad y = a\sin\phi\sin\theta, \quad z = a\cos\phi$$

represent a sphere. Here the parameters are $\theta$ and $\phi$.

A parametrized surface may degenerate to a lower-dimensional object. For example, if $x = f(u,v)$, $y = g(u,v)$, and $z = h(u,v)$ are all constant, then $r(T)$ reduces to a single point. Likewise, if $x = u+v$, $y = (u+v)^2$, and $z = (u+v)^3$, then after writing $t = u+v$ one sees that $r(T)$ is in fact a curve in $\mathbb{R}^3$.

If, at $(u,v) \in T$, the partial derivatives $\frac{\partial r}{\partial u}$ and $\frac{\partial r}{\partial v}$ are continuous and satisfy $\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} \neq 0$, then $r(u,v)$ is called a **regular point** of the surface $r(T)$. If one of these derivatives fails to be continuous or if $\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} = 0$ at $(u,v)$, then $(u,v)$ is called a **singular point**. A surface $r(T)$ is called **smooth** if every point of the surface is regular.

### 9.4.1 Area of a parametric surface

Let $S = r(T)$ be a smooth parametric surface defined on a domain $T$. Thus $\frac{\partial r}{\partial u}$ and $\frac{\partial r}{\partial v}$ are continuous on $T$, and $\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v}$ never vanishes there.

If we fix $v$ and allow $u$ to run, then the image of $r$ reduces to a curve in $\mathbb{R}^3$. Hence the distance travel along the curve $r(\cdot, v)$ in a small time interval $\Delta u$ is $\|\frac{\partial r}{\partial u}\|\Delta u$. Similarly, if we fix $u$, then the graph of $r$ is a curve in $\mathbb{R}^3$, and the distance traveled along this curve in a small time interval $\Delta v$ is $\|\frac{\partial r}{\partial v}\|\Delta v$. See fig. 9.3.

Thus, we see that a small rectangle in $T$ of area $\Delta u \Delta v$ in the $uv$-plane is transferred to a parallelogram on the surface $r(T)$ with area $\left\|\frac{\partial r}{\partial u}\Delta u \times \frac{\partial r}{\partial v}\Delta v\right\| = \left\|\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v}\right\| \Delta u \Delta v$.

Note that the point, where $\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} = 0$, the parallelogram on $r(T)$ will collapse to a curve or a point. Now at each regular point of $r(T)$, the vectors $\frac{\partial r}{\partial u}$ and $\frac{\partial r}{\partial v}$ determine a plane having $\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v}$ as the normal vector to the surface at the point $(u,v)$. Recall that $\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} = \left\|\frac{\partial r}{\partial u}\right\| \left\|\frac{\partial r}{\partial v}\right\| \sin\theta \ \hat{n}$, where $\hat{n}$ is the unit normal to the surface $r(T)$ at $(u,v) \in T$. Hence, the plane determined by $\frac{\partial r}{\partial u}$

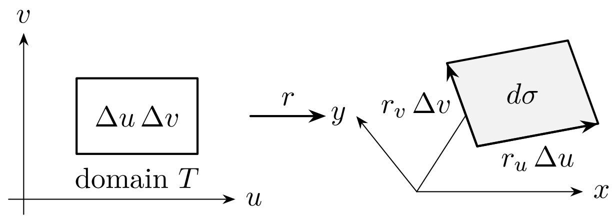



**Figure 9.3.** A small rectangle in the parameter domain maps to a parallelogram on the surface, with area element $d\sigma = \|r_u \times r_v\|\,du\,dv$.

and $\frac{\partial r}{\partial v}$ is called tangent plane of the surface. Note that the continuity of $\frac{\partial r}{\partial u}$ and $\frac{\partial r}{\partial v}$ implies the continuity of $\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v}$. Hence the tangent plane varies continuously on the smooth surface. Thus, the continuity of $\frac{\partial r}{\partial u}$ and $\frac{\partial r}{\partial v}$ prevent the occurrence of sharp edges or corners on the surface.

Let us denote the area of small parallelogram obtained by transferring the small rectangle of areas $\Delta u \Delta v$ in the domain $T$ by $d\sigma$. Let $r_u = \frac{\partial r}{\partial u}$ and $r_v = \frac{\partial r}{\partial v}$. Then $d\sigma = \|r_u \times r_v\|\, \Delta u \Delta v$. Hence the surface area of $r(T)$ denoted by $a(S)$ is given by

$$a(S) = \iint_T \|r_u \times r_v\|\, dudv.$$

**Area of a surface defined by a graph:** Suppose a surface $S$ is given by

$$z = f(x, y), \quad \text{for } (x, y) \in T.$$

That is, $S$ is the graph of the function $f(x, y)$. Then $S$ can be considered as a parametric surface defined by:

$$r(x, y) = xi + yj + f(x, y)k, \quad (x, y) \in T.$$

In this case, $r_x = i + f_x k, \quad r_y = j + f_y k$. Further, $r_x \times r_y = \begin{vmatrix} i & j & k \\ 1 & 0 & f_x \\ 0 & 1 & f_y \end{vmatrix} = -f_x i - f_y j + k.$

Hence the surface area becomes

$$a(S) = \iint_T \sqrt{1 + f_x^2 + f_y^2}\, dxdy.$$

**Example 9.38.** Let us find the area of the surface of the portion of the sphere $x^2 + y^2 + z^2 = 4a^2$ that lies inside the cylinder $x^2 + y^2 = 2ax$. Consider $f(x, y) = \sqrt{4a^2 - x^2 - y^2}$, then

$$f_x = \frac{-x}{\sqrt{4a^2 - x^2 - y^2}}, \quad f_y = \frac{-y}{\sqrt{4a^2 - x^2 - y^2}}$$

$$\text{and} \qquad \sqrt{1 + f_x^2 + f_y^2} = \sqrt{\frac{4a^2}{4a^2 - x^2 - y^2}}.$$

Then

$$\begin{aligned} a(S) =& 2 \iint\limits_{T} \sqrt{\frac{4a^2}{4a^2 - x^2 - y^2}} dxdy \\ =& 2 \times 2 \int_0^{\frac{\pi}{2}} \int_0^{2a\cos\theta} \frac{2ar}{\sqrt{4a^2 - r^2}} \, drd\theta. \end{aligned}$$

*Remark* 9.39. Note that

$$\begin{aligned} \|r_u \times r_v\|^2 &= \|r_u\|^2 \|r_v\|^2 \sin^2\theta \\ &= \|r_u\|^2 \|r_v\|^2 \left(1 - \cos^2\theta\right) \\ &= \|r_u\|^2 \|r_v\|^2 - (r_u \cdot r_v)^2 . \end{aligned}$$

Let $E = r_u \cdot r_u,\ G = r_v \cdot r_v$ and $F = r_u \cdot r_v$, then

$$a(S) = \iint\limits_{T} \sqrt{EG - F^2} dudv.$$

### 9.4.2 Surface integrals

Let $S$ be a parametric surface given by $r(u, v)$ on a parameter domain $T$, and assume that $r_u$ and $r_v$ are continuous. If $g : S \to \mathbb{R}$ is a bounded scalar field for which the right-hand side exists, then the surface integral of $g$ over $S$, denoted by $\iint\limits_{S} g \, d\sigma$, is defined by

$$\iint\limits_{S} g d\sigma = \iint\limits_{T} g(r(u,v)) \|r_u \times r_v\| \, dudv = \iint\limits_{T} g(r(u,v)) \sqrt{EG - F^2} dudv$$

provided double integral in the RHS exists.

*Remark* 9.40. $(i)$.

$$\iint\limits_{S} g d\sigma = \iint\limits_{T} g(r(u,v)) \sqrt{EG - F^2} dudv.$$

$(ii)$. If $S$ is defined by $z = f(x, y)$, then

$$\iint\limits_{S} g d\sigma = \iint\limits_{T} g[x, y, f(x, y)] \sqrt{1 + f_x^2 + f_y^2} \, dxdy,$$

where $T$ is the projection of the surface $S$ over the $xy$-plane.

**Example 9.41.** Let $S$ be the hemispherical surface $z = \left(a^2 - x^2 - y^2\right)^{1/2}$. Evaluate

$$\iint\limits_{S} \frac{d\sigma}{[x^2 + y^2 + (z + a)^2]^{1/2}}.$$

Consider

$$S := r(\theta, \phi) = (a \sin\phi \cos\theta,\ a \sin\phi \sin\theta,\ a\cos\phi),$$

where $0 \le \theta \le 2\pi$, $0 \le \phi \le \frac{\pi}{2}$. Note that

$$\sqrt{EG - F^2} = a^2 \sin\phi \text{ and } \left[x^2 + y^2 + (z+a)^2\right]^{1/2} = 2a\cos\frac{\phi}{2}.$$

Hence

$$\iint\limits_S \frac{d\sigma}{[x^2 + y^2 + (z+a)^2]^{1/2}} = \int_0^{2\pi} \int_0^{\pi/2} \frac{a^2 \sin\phi}{2a\cos\frac{\phi}{2}} d\phi d\theta.$$

## 9.5 Line Integrals

**Section overview.**

- Line integrals measure accumulation along curves and provide the language for work, circulation, and path independence.
- The central conceptual issues are parametrization, orientation, and dependence on endpoints versus dependence on the path.

Let $R : [a, b] \to \mathbb{R}^3$ be a differentiable parametrization of a curve $C$, and let $f : C \to \mathbb{R}^3$ be a bounded vector field such that the integrand below is Riemann integrable. The line integral of $f$ along $C$ is defined by

$$\int\limits_C f \cdot dR = \int_a^b f(R(t)) \cdot R'(t)\, dt.$$

This definition is independent of the particular smooth parametrization, provided orientation is respected. See fig. 9.4.

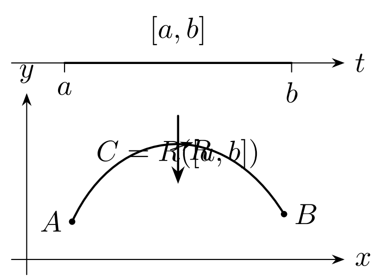



**Figure 9.4.** A curve $C$ parameterized by $R : [a, b] \to \mathbb{R}^3$; the line integral $\int_C F \cdot dR$ is defined as $\int_a^b F(R(t)) \cdot R'(t)\, dt$.

*Remark* 9.42. Suppose $f = (f_1, f_2, f_3)$ and $R(t) = (x(t), y(t), z(t))$. Then the line integral $\int_C f \cdot dR$ is also written as

$$\int_C f_1 dx + f_2 dy + f_3 dz \quad \text{or} \quad \int_C f_1(x, y, z)dx + f_2(x, y, z)dy + f_3(x, y, z)dz.$$

**Example 9.43.** Let $f = x^2 i + yj + (xz - y)k$. Compute the line integral $\int_C f \cdot dR$ along a curve $C$ joining $(0,0,0)$ to $(1,2,4)$ in each of the following cases:
(i) $C$ is the straight-line segment joining these points;
(ii) $C$ is the curve given by $R(t) = (t^2, 2t, 4t^2)$.

**The second fundamental theorem for line integrals:** In one variable, the fundamental theorem of calculus shows that the integral of a derivative depends only on the values of an antiderivative at the endpoints. The line-integral analogue expresses the same principle for gradient fields: if a vector field is the gradient of a scalar potential, then its line integral depends only on the endpoints of the curve and is therefore path independent.

We now formulate this higher-dimensional version precisely.

**Theorem 9.44** (Fundamental theorem for line integrals)**.** *Let $D \subseteq \mathbb{R}^3$ be a domain, let $f \in C^1(D)$, and let $C = \{R(t) : t \in [a,b]\}$ be a continuously differentiable curve in $D$ joining the points $A = R(a)$ and $B = R(b)$. Then*

$$\int_C \nabla f \cdot dR = f(B) - f(A).$$

*Proof.* Define $h(t) := f(R(t))$ for $t \in [a,b]$. By the chain rule,

$$h'(t) = (f \circ R)'(t) = \nabla f(R(t)) \cdot R'(t).$$

Therefore

$$\int_C \nabla f \cdot dR = \int_a^b \nabla f(R(t)) \cdot R'(t)\,dt = \int_a^b h'(t)\,dt = h(b) - h(a) = f(B) - f(A).$$

□

*Remark* 9.45. Theorem 9.44 shows that the line integral of a gradient field depends only on the endpoints of the curve. In particular, it is independent of the choice of path joining $A$ and $B$ inside $D$.

**Definition 9.46.** Let $R : [a,b] \to \mathbb{R}^3$ be a continuous parametrization of a curve $C$. The curve $C$ is called

(i) **simple** if $R$ is one-to-one on $(a,b)$;
(ii) **closed** if $R(a) = R(b)$;
(iii) **smooth** if $R'$ exists and is continuous;
(iv) **piecewise smooth** if $[a,b]$ can be partitioned into finitely many subintervals on each of which $R$ is smooth.

**Theorem 9.47** (Green's theorem)**.** *Let $C$ be a positively oriented piecewise smooth simple closed curve in the xy-plane, and let $D$ denote the bounded region enclosed by $C$. Assume that $M, N \in C^1(U)$ on an open set $U$ containing $D$. Then*

$$\iint_D \left(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y}\right) dx\, dy = \oint_C M\, dx + N\, dy. \tag{9.4}$$

By linearity, it suffices to establish separately the contributions coming from $\partial N/\partial x$ and $-\partial M/\partial y$. We first treat the standard Type I and Type II regions shown in fig. 9.5; the general case then follows by decomposing the domain into finitely many such pieces.

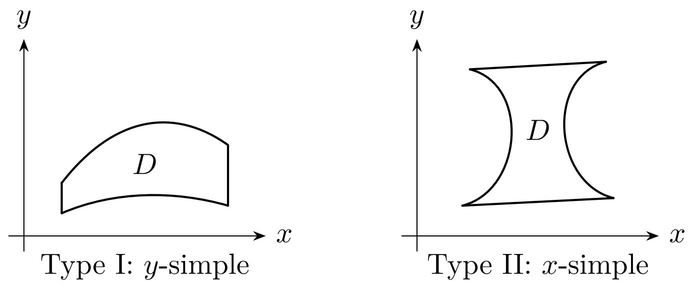



**Figure 9.5.** Two standard classes of planar domains used in the proof of Green's theorem.

*Proof.* $(i)$ Let $D = \{(x, y) : \ a \le x \le b \text{ and } f(x) \le y \le g(x)\}$, where $f$ and $g$ are continuous functions on $[a, b]$. Since $\frac{\partial \mathrm{M}}{\partial y}$ is continuous, by Fubini's Theorem, the double integral

$$-\iint_D \frac{\partial \mathrm{M}}{\partial y} dxdy = \int_a^b \left[\int_{f(x)}^{g(x)} \frac{\partial \mathrm{M}}{\partial y} dy\right] dx = \int_a^b M[x, f(x)]dx - \int_a^b M[x, g(x)]dx. \tag{9.5}$$

On the other hand, we can write

$$\int_C M dx = \int_{C_1} M dx + \int_{C_2} M dx, \tag{9.6}$$

since the line integral along each of vertical segment is zero. Note that $C_1$ and $C_2$ can be represented by $r_1(t) = ti + f(t)j$ and $r_2(t) = ti + g(t)j$ respectively. Hence

$$\int_{C_1} M dx = \int_a^b M[t, f(t)]dt \text{ and } \int_{C_2} M dx = -\int_a^b M[t, g(t)]dt. \tag{9.7}$$

Negative sign appeared in the second equation since the curve $C_2$ traverses in the reverse direction. Thus, from (9.5-9.7) we conclude that the identity (9.4) holds for the type I region. Similarly, we can obtain the result for the type II region. Further, we can obtain the result for any region which can be decomposed into finitely many regions of the above two types. $\square$

**Area expressed as a line integral:** Let $C$ be a simple (piecewise smooth) closed curve and $D$ be the region enclosed by $C$. Let $M(x,y) = -\frac{y}{2}$ and $N(x,y) = \frac{x}{2}$. Then by Green's Theorem the area of $D$ is

$$a(D) = \iint\limits_D dxdy = \iint\limits_D (N_x - M_y)\,dxdy = \int_a^b Mdx + Ndy = \frac{1}{2}\int_C -ydx + xdy.$$

**Example 9.48.** Note that the integral

$$\int_C xy^2dx + \left(x^2y + x\right)dy = \iint\limits_D dxdy = \text{Area(D)}$$

(By Green's Theorem), where $D$ is the region enclosed by $C$. Hence the integral is depending only on the region enclosed by $C$ but not its location.

**Example 9.49.** Find the area bounded by the ellipse $C = \{\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1\}$.
Consider the parametric form of $C = \{(a\cos t, b\sin t) :\ 0 \le t < 2\pi.\}$ Then the area is

$$\frac{1}{2}\int_C -ydx + xdy = \frac{1}{2}\int_0^{2\pi} -(b\sin t)(-a\sin t)dt + (a\cos t)(b\cos t)dt = \frac{1}{2}\int_0^{2\pi} abdt = ab\pi.$$

**Example 9.50.** Let $C_1$ and $C_2$ be two simple (piecewise smooth) closed curves as shown in fig. 9.6.

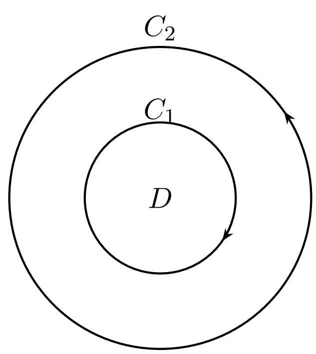



**Figure 9.6.** A multiply connected domain $D$ bounded by an outer curve $C_2$ and an inner curve $C_1$ (annular-type region).

Consider the region $D$ bounded by the curves $C_1$ and $C_2$. Note that $D = D_1 \cup D_2$ and $D_1$ is enclosed by the curves $\gamma_i;\ i = 1,2,3,4$ and $D_2$ is enclosed by curves $\gamma_j;\ i = 1,3,5,6$.

$$\iint\limits_D \left(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y}\right)dxdy = \iint\limits_{D_1} \left(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y}\right)dxdy + \iint\limits_{D_2} \left(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y}\right)dxdy$$
$$= \left(\int_{\gamma_1}\alpha + \int_{\gamma_2}\alpha + \int_{\gamma_3}\alpha + \int_{\gamma_4}\alpha\right) + \left(\int_{\gamma_6}\alpha - \int_{\gamma_3}\alpha + \int_{\gamma_5}\alpha - \int_{\gamma_1}\alpha\right) = \oint_{C_2}\alpha - \oint_{C_1}\alpha,$$

where $\alpha = Mdx + Ndy$.

**Example 9.51.** Let $C_1$ be unit circle and $C_2$ be any simple closed curve as shown in fig. 9.7.

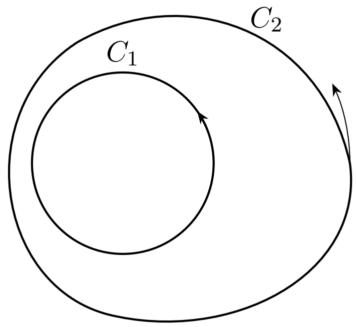



**Figure 9.7.** An inner unit circle $C_1$ and an outer simple closed curve $C_2$ bounding a multiply connected domain.

Find $\int_{C_2} \frac{x\,dy - y\,dx}{x^2+y^2}$. Let $D$ be the domain lying between $C_1$ and $C_2$. A direct calculation shows that $N_x - M_y = 0$ on $D$. Applying Green's theorem to the multiply connected domain $D$, we get

$$\oint_{C_2} (M\,dx + N\,dy) - \oint_{C_1} (M\,dx + N\,dy) = \iint_D (N_x - M_y)\,dx\,dy = 0.$$

Since $C_1 = \{(\cos t, \sin t) : 0 \le t \le 2\pi\}$, we obtain

$$\int_{C_1} \frac{-y\,dx + x\,dy}{x^2+y^2} = \int_0^{2\pi} \frac{\sin^2 t + \cos^2 t}{\sin^2 t + \cos^2 t}\,dt = 2\pi.$$

Hence

$$\oint_{C_2} (M\,dx + N\,dy) = 2\pi.$$

### 9.5.1 Exactness of the line integral

Let $Q$ be a cube in $\mathbb{R}^3$, and let $C$ be a $C^1$ curve in $Q$ with parametrization $R(t) = (x(t), y(t), z(t))$, where $R : [a,b] \to \mathbb{R}^3$. A basic question is whether there exists a scalar potential $F : Q \to \mathbb{R}$ such that

$$\int_C f \cdot dR = \int_C dF$$

for every curve $C$ in $Q$.

Assume that such an $F$ exists. Then, by Theorem 9.44 (the second fundamental theorem for line integrals),

$$\int_C f \cdot dR = F(R(b)) - F(R(a)) = F(B) - F(A) = \int_C \nabla F \cdot dR.$$

Therefore

$$\int_C (f - \nabla F) \cdot dR = 0$$

for every curve $C$ in $Q$. A standard localization argument then shows that $f = \nabla F$ on $Q$.

*Remark* 9.52. On a general domain $D$, the implication from vanishing line integrals to the existence of a potential requires additional topological hypotheses. The following exercise treats a particularly simple case.

**Exercise 9.53.** Let $D = \{(x, y) : x^2 + y^2 < 1\}$. If $f : D \to \mathbb{R}^2$ is continuously differentiable and

$$\int_\Gamma f \cdot dR = 0$$

for every piecewise smooth curve $\Gamma$ in $D$, show that $f \equiv 0$ on $D$.

**Example 9.54.** Show that the line integral

$$\int_C 2x \sin y \; dx + (x^2 \cos y - 3y^2) dy$$

is path independent joining the points $(-1, 0)$ and $(5, 1)$.

## 9.6 Curl and Divergence

**Section overview.**

- Curl and divergence encode two complementary local features of a vector field: rotation and net outward flux.
- The integral theorems show how these local differential quantities govern global circulation and flux.

Let $F : \mathbb{R}^3 \to \mathbb{R}^3$ be a vector field of the form $F(x, y, z) = P(x, y, z)i + Q(x, y, z)j + R(x, y, z)k$.

**Definition 9.55** (Curl of $F$)**.** The *curl* of $F$ is the vector field denoted by $\operatorname{curl} F$ and defined formally by

$$\operatorname{curl} F = \begin{vmatrix} i & j & k \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ P & Q & R \end{vmatrix} = \nabla \times F,$$

where $\nabla = \dfrac{\partial}{\partial x} i + \dfrac{\partial}{\partial y} j + \dfrac{\partial}{\partial z} k$.

**Definition 9.56** (Divergence of $F$)**.** The *divergence* of $F$ is the scalar field denoted by $\operatorname{div} F$ and defined by

$$\operatorname{div} F = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} = \nabla \cdot F.$$

To motivate Stokes' theorem, recall first Green's theorem in the plane. Let $C$ be a piecewise smooth closed curve enclosing a domain $D \subset \mathbb{R}^2$, and let $F(x, y) = M(x, y)i + N(x, y)j + 0k$. Then Green's theorem gives

$$\iint_D (N_x - M_y) \; dx \, dy = \oint_C M \, dx + N \, dy,$$

where $C = \{R(t) : t \in [a,b]\}$. This can be rewritten as

$$\iint\limits_D \operatorname{curl} F \cdot k \, dx \, dy = \oint_C F \cdot dR, \tag{9.8}$$

where $\operatorname{curl} F = \left(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y}\right) k$. Stokes' theorem is the three-dimensional extension of this identity. Before stating it formally, we review the geometry of normal vectors on some basic classes of surfaces.

(i) Suppose the surface $S$ is given by $f(x,y,z) = c$, where $f$ is differentiable function on some domain $D$ in $\mathbb{R}^3$. See fig. 9.8.

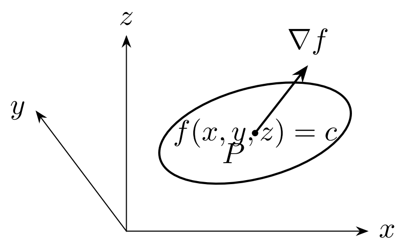



**Figure 9.8.** A level surface $f(x,y,z) = c$ with normal direction given by $\nabla f(P)$.

Consider a smooth curve $C$ with parametrization $R : [a,b] \to \mathbb{R}^3$ lying on $S$ and passing through a point $P \in S$. Since $f(R(t)) = c$ along the curve, the chain rule yields

$$\nabla f(R(t)) \cdot R'(t) = 0.$$

Thus $\nabla f(R(t))$ is orthogonal to every tangent vector $R'(t)$ and therefore provides a normal direction to the surface at $P$. Consequently, whenever $\nabla f \neq 0$, a unit normal vector is given by

$$\hat{n} = \frac{\nabla f}{\|\nabla f\|}.$$

If this unit normal can be chosen continuously on the surface and never vanishes, then the surface $S$ is called **orientable**.

(ii) Let $D$ be a domain in $\mathbb{R}^2$. Let $F : D \to \mathbb{R}^3$ is given by $F(s,t) = x(s,t)i + y(s,t)j + z(s,t)k$ and is a parametrization of the surface $S$, where $F$ is smooth (continuously differentiable). Let $P = F(s_o, t_o)$ be a point on the surface $S$. Then $F(s, t_o)$ and $F(s_o, t)$ are curves on $S$ passing through $P$ as shown in fig. 9.9.

Recall that the cross product $F_s \times F_t$ is orthogonal to the tangent plane of $S$ at $P$. Hence, whenever $F_s \times F_t \neq 0$, a unit normal vector is given by

$$\hat{n} = \frac{F_s \times F_t}{\|F_s \times F_t\|}.$$

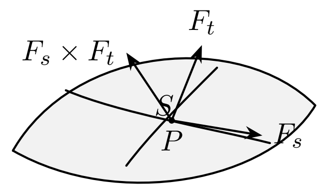



**Figure 9.9.** Coordinate curves on a parametrized surface $F(s,t)$; the tangent vectors $F_s$ and $F_t$ span the tangent plane, and $F_s \times F_t$ is normal.

(iii) If the surface $S$ is given by the graph of a smooth function $f : D \subset \mathbb{R}^2 \to \mathbb{R}$, so that

$$F(x,y) = (x, y, f(x,y)) = x\,i + y\,j + f(x,y)\,k,$$

then a unit normal vector is given by

$$\hat{n} = \frac{F_x \times F_y}{\|F_x \times F_y\|} = \frac{-f_x i - f_y j + k}{\sqrt{1 + f_x^2 + f_y^2}}.$$

**Definition 9.57.** A surface $S$ is called *orientable* if it admits a continuous unit normal field on all of $S$.

Thus an orientable surface is, in essence, a two-sided surface. The Möbius strip provides the standard example of a non-orientable surface.

**Theorem 9.58** (Stokes' Theorem)**.** *Let $S$ be a piecewise smooth orientable surface and $C$ be the piecewise smooth boundary of $S$. Let $F(x,y,z) = P(x,y,z)i + Q(x,y,z)j + R(x,y,z)k$ be a vector field such that $P, Q$ and $R$ are continuously differentiable on an open set containing $S$. If $\hat{n}$ is a unit normal vector to $S$, then*

$$\iint\limits_S \operatorname{curl} F \cdot \hat{n}\, d\sigma = \oint_C F \cdot dR, \tag{9.9}$$

*where the line integral is evaluated around $C$ in the direction of the orientation of $C$ with respect to $\hat{n}$.*

See fig. 9.10.

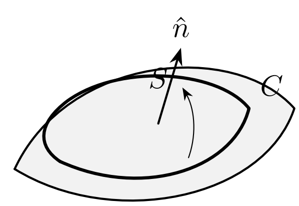



**Figure 9.10.** Stokes' theorem relates the circulation along the oriented boundary $C = \partial S$ to the flux of $\operatorname{curl} F$ through $S$.

(i) The value of the surface integral in (9.9) depends only on the boundary $C$ and not on the particular shape of the surface $S$.

(ii) If $S$ is a plane surface, then identity (9.9) reduces to identity (9.8). Thus, Stokes' theorem may be viewed as a direct extension of Green's theorem.

(iii) If $S$ is a closed smooth surface, such as a sphere or a torus, then $\partial S = \varnothing$, and therefore

$$\iint_S \operatorname{curl} F \cdot \hat{n}\, d\sigma = 0.$$

(iv) Stokes' theorem extends to smooth surfaces whose boundary consists of more than one simple smooth closed curve.

*Remark* 9.59. If a surface $S$ is given by the graph of a smooth function $f$ defined on a domain $D \subset \mathbb{R}^2$, then

$$\oint_C F \cdot dR = \iint_D (-f_x i - f_y j + k) \cdot \operatorname{curl} F \, dx\, dy.$$

**Example 9.60.** Let $S$ be the portion of the surface $z = 1 - x^2$ with $0 \le x \le 1$ and $-2 \le y \le 2$. Let $C = \partial S$, and let $F(x, y, z) = yi + yj + k$. Use Stokes' theorem to compute the line integral $\int_C F \cdot dR$.

Here $\operatorname{curl} F = -\vec{k}$. Write $z = f(x, y) = 1 - x^2$. Then the upward unit normal to $S$ is

$$\hat{n} = \frac{-f_x\vec{i} - f_y\vec{j} + \vec{k}}{\sqrt{1 + f_x^2 + f_y^2}} = \frac{2x\vec{i} + \vec{k}}{\sqrt{1 + 4x^2}},$$

and the surface element is

$$d\sigma(x, y) = \sqrt{1 + f_x^2 + f_y^2}\, dx\, dy.$$

See fig. 9.11.

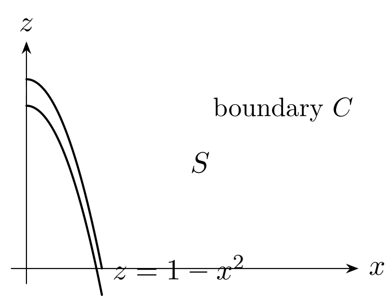



**Figure 9.11.** A schematic portion of the cylinder-like surface $z = 1 - x^2$ over a rectangle in the $(x, y)$-plane; its boundary curve $C$ is oriented consistently with $\hat{n}$.

By Stokes' theorem and Remark 9.59,

$$\oint_C F \cdot dR = \iint_D (-f_x i - f_y j + k) \cdot \operatorname{curl} F \, dx\, dy = \int_{-2}^{2} \int_0^1 (-1)\, dx\, dy = -4.$$

Let us return to Green's theorem in the plane and derive its flux form. Let

$$F(x,y) = M(x,y)\, i + N(x,y)\, j$$

be a $C^1$ vector field on a domain $D \subset \mathbb{R}^2$ whose positively oriented boundary is a simple smooth curve

$$C = \{R(t) : t \in [a,b]\}, \qquad R(t) = (x(t), y(t)).$$

Then

$$T(t) = \frac{R'(t)}{\|R'(t)\|}$$

is the unit tangent vector, and the outward unit normal is

$$\nu(t) = \frac{(y'(t), -x'(t))}{\|R'(t)\|}.$$

Since $ds = \|R'(t)\|\, dt$, we have

$$F(R(t)) \cdot \nu(t)\, ds = M(R(t))\, dy - N(R(t))\, dx.$$

Applying Green's theorem with the choice $M_1 = -N$ and $N_1 = M$, we obtain

$$\begin{aligned} \oint_C F \cdot \nu\, ds &= \oint_C M\, dy - N\, dx \\ &= \iint_D \left( \frac{\partial M}{\partial x} + \frac{\partial N}{\partial y} \right) dx\, dy = \iint_D \operatorname{div} F\, dx\, dy. \end{aligned}$$

Hence

$$\iint_D \operatorname{div} F\, dx\, dy = \oint_C F \cdot \nu\, ds. \tag{9.10}$$

The generalization of (9.10) to surfaces in $\mathbb{R}^3$ is the divergence theorem.

**Theorem 9.61** (Divergence Theorem)**.** *Let $D$ be a solid region in $\mathbb{R}^3$ bounded by a piecewise smooth orientable surface $S$. Let $F(x,y,z) = P(x,y,z)i + Q(x,y,z)j + R(x,y,z)k$ be a vector field that is continuously differentiable on an open set containing $D$. If $\hat{n}$ denotes the outward unit normal to $S$, then*

$$\iiint_D \operatorname{div} F\, dV = \iint_S F \cdot \hat{n}\, d\sigma.$$

See fig. 9.12.

**Example 9.62.** Let $F(x,y,z) = (x+y)i + z^2 j + x^2 k$. Let $\hat{n}$ be the outward unit normal to the hemisphere

$$S = \{(x,y,z) : x^2 + y^2 + z^2 = 1,\ z > 0\}.$$

Compute the surface integral $\iint_S F \cdot \hat{n}\, d\sigma$ using the divergence theorem.

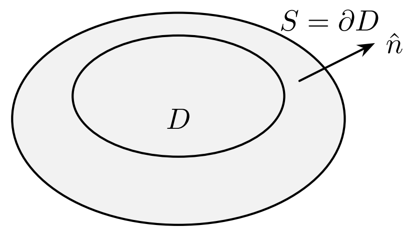



**Figure 9.12.** A bounded domain $D$ with boundary surface $S = \partial D$ and outward unit normal $\hat{n}$ (Gauss–divergence theorem).

Write $F(x, y, z) = (x + y, z^2, x^2)$. Then $\operatorname{div} F = 1$. Since $S$ is not closed, let

$$S_1 = \{(x, y, 0) : x^2 + y^2 \le 1\}.$$

Then $S \cup S_1$ is a closed surface, so the divergence theorem applies; see fig. 9.13.

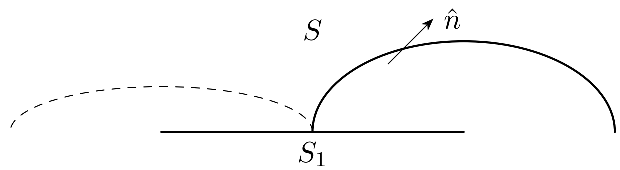



**Figure 9.13.** A hemisphere $S$ closed by the base disk $S_1$ to form a closed surface, enabling the divergence theorem.

By the divergence theorem,

$$\iint\limits_S F \cdot \hat{n}\, d\sigma + \iint\limits_{S_1} F \cdot \hat{n}_1\, d\sigma_1 = \iiint\limits_D \operatorname{div} F\, dV = \frac{2\pi}{3}.$$

Here

$$\iint\limits_{S_1} F \cdot \hat{n}_1\, d\sigma_1 = \iint\limits_{x^2+y^2\le 1} F(x, y, 0) \cdot (-k)\, dx\, dy = -\iint\limits_{x^2+y^2\le 1} x^2\, dx\, dy = -\frac{\pi}{4}.$$

Therefore

$$\iint\limits_S F \cdot \hat{n}\, d\sigma = \frac{2\pi}{3} + \frac{\pi}{4} = \frac{11\pi}{12}.$$

# Problem Sets

These problem sets are designed in three layers: a first pass of diagnostic questions that test command of definitions, a second pass of proof-oriented exercises that consolidate the main theorems, and a final pass of synthesis problems that ask you to combine ideas across sections. A productive workflow is to attempt each set in that order, returning to the later questions only after the structural lemmas and standard examples have become familiar.

## Problem Set 1

**Problem-set architecture.**

- Begin with Item 1 and Item 3 to audit your command of the definitions and to build a bank of counterexamples.
- Use the middle part of the set to compare genuine metric phenomena with artifacts of particular norms or coordinates.
- Reserve the later questions for synthesis: these require you to combine completeness, continuity, and structure of function spaces rather than treating them separately.

1. State whether each of the following statements is true or false, and justify your answer.
   (a) There does not exist a monotone function $f : \mathbb{R} \to \mathbb{Q}$ which is onto.
   (b) There exists a monotone function $f : (0, \infty) \to \mathbb{R}$ such that each $c \in (0, \infty)$ satisfies $|f(c+) - f(c-)| = \frac{1}{c}$.
   (c) There exists a sequence of differentiable functions $f_n$ on $(0, \infty)$ such that $f_n'$ is uniformly convergent on $(0, \infty)$ but $f_n$ is nowhere pointwise convergent.
   (d) There exists a metric space having exactly 36 open sets.
   (e) It is impossible to define a metric $d$ on $\mathbb{R}$ such that only finitely many subsets of $\mathbb{R}$ are open in $(\mathbb{R}, d)$.
   (f) If $A$ and $B$ are open (closed) subsets of a normed vector space $X$, then $A + B = \{a + b : a \in A, b \in B\}$ is open (closed) in $X$.
   (g) If $A$ and $B$ are closed subsets of $[0, \infty)$ (with the usual metric), then $A + B$ is closed in $[0, \infty)$.
   (h) It is possible to define a metric $d$ on $\mathbb{R}$ such that the sequence $(1, 0, 1, 0, \ldots)$ converges in $(\mathbb{R}, d)$.
   (i) It is possible to define a metric $d$ on $\mathbb{R}^2$ such that $((\frac{1}{n}, \frac{n}{n+1}))$ is not a Cauchy sequence in $(\mathbb{R}^2, d)$.
   (j) It is possible to define a metric $d$ on $\mathbb{R}^2$ such that in $(\mathbb{R}^2, d)$, the sequence $((\frac{1}{n}, 0))$ converges but the sequence $((\frac{1}{n}, \frac{1}{n}))$ does not converge.

(k) Let $A \subset (1, \infty)$ be a closed set. Then $A^2 := \{a^2 : a \in A\}$ is a closed set.

(l) Let $A_n = \{(x, y) \in \mathbb{R}^2 : 0 < \frac{1}{x} < y < \frac{1}{n}\}$. Determine whether the set $\bigcap\limits_{n=1}^{\infty} A_n$ is open or closed.

(m) There exists a set $A \subset (\mathbb{R}, u)$ such that $\delta\big(A^\circ \cup \{0\}\big) = 0$ but $\delta\big((\overline{A})^\circ\big) = 1$, where $\delta$ denotes the diameter.

(n) If $(x_n)$ is a sequence in a complete normed vector space $X$ such that $\|x_{n+1} - x_n\| \to 0$ as $n \to \infty$, then $(x_n)$ must converge in $X$.

(o) If $(f_n)$ is a sequence in $C[0, 1]$ such that $|f_{n+1}(x) - f_n(x)| \leq \frac{1}{n^2}$ for all $n \in \mathbb{N}$ and for all $x \in [0, 1]$, then there must exist $f \in C[0, 1]$ such that $\int\limits_0^1 |f_n(x) - f(x)|\, dx \to 0$ as $n \to \infty$.

(p) If $(x_n)$ is a Cauchy sequence in a normed vector space, then $\lim\limits_{n\to\infty} \|x_n\|$ must exist.

(q) $\{f \in C[0, 1] : \|f\|_1 \leq 1\}$ is a bounded subset of the normed vector space $(C[0, 1], \|\cdot\|_\infty)$.

(r) For $x, y \in \ell^\infty$, let $d(x, y) = \min\{1, \limsup_{n\to\infty} |x_n - y_n|\}$. Determine whether this defines a metric on $\ell^\infty$.

(s) The sequence $f_n(t) = e^{-n^2 \sin \pi t}$ converges uniformly to 0 on $(0, 1)$.

(t) If the sequence $(x_n)$ in $\mathbb{R}$ satisfies

$$0 \leq \inf x_n \leq \sup x_n < \infty,$$

determine whether $(x_n)$ must have a convergent subsequence.

(u) If $f : \mathbb{R} \to \mathbb{R}$ is continuous, bounded, and monotone, must the limits $\lim\limits_{x\to\pm\infty} f(x)$ be finite?

2. What is the cardinality of the set $\{f : \mathbb{R} \to \mathbb{R},\ f \text{ is nowhere continuous}\}$?

3. For a monotone increasing function $f : [a, b] \to \mathbb{R}$, define $g(x) = \sup\{f(y) : y < x\}$. If $f$ has limit at $c$, then show that $f(c) = g(c)$.

4. Examine whether $(X, d)$ is a metric space, where

(a) $X = \mathbb{R}$ and $d(x, y) = \frac{|x-y|}{1+|xy|}$ for all $x, y \in \mathbb{R}$.

(b) $X = \mathbb{R}$ and $d(x, y) = |x - y|^p$ for all $x, y \in \mathbb{R}$ $(0 < p < 1)$.

(c) $X = \mathbb{R}$ and $d(x, y) = \min\{\sqrt{|x - y|}, |x - y|^2\}$ for all $x, y \in \mathbb{R}$.

(d) $X = \mathbb{R}$ and for all $x, y \in \mathbb{R}$, $d(x, y) = \begin{cases} 1 + |x - y| & \text{if exactly one of } x \text{ and } y \text{ is positive,} \\ |x - y| & \text{otherwise.} \end{cases}$

(e) $X = \mathbb{R}^2$ and $d(x, y) = (|x_1 - y_1| + |x_2 - y_2|^{\frac{1}{2}})^{\frac{1}{2}}$ for all $x = (x_1, x_2), y = (y_1, y_2) \in \mathbb{R}^2$.

(f) $X = \mathbb{R}^n$ and $d(x, y) = [(x_1 - y_1)^2 + \frac{1}{2}(x_2 - y_2)^2 + \cdots + \frac{1}{n}(x_n - y_n)^2]^{\frac{1}{2}}$ for all $x = (x_1, ..., x_n)$, $y = (y_1, ..., y_n) \in \mathbb{R}^n$.

(g) $X = \mathbb{C}$ and for all $z, w \in \mathbb{C}$, $d(z, w) = \begin{cases} \min\{|z| + |w|, |z - 1| + |w - 1|\} & \text{if } z \neq w, \\ 0 & \text{if } z = w. \end{cases}$

(h) $X = \mathbb{C}$ and for all $z, w \in \mathbb{C}$, $d(z, w) = \begin{cases} |z - w| & \text{if } \frac{z}{|z|} = \frac{w}{|w|}, \\ |z| + |w| & \text{otherwise.} \end{cases}$

(i) $X = \mathbb{C}$ and $d(z,w) = \frac{2|z-w|}{\sqrt{1+|z|^2}\sqrt{1+|w|^2}}$ for all $z, w \in \mathbb{C}$.

(j) $X =$ The class of all finite subsets of a nonempty set and $d(A,B) =$ The number of elements of the set $A \triangle B$ (the symmetric difference of $A$ and $B$).

5. Let $1 \le p \le \infty$ and let $d_i$, $i = 1, 2$, be two metrics on a nonempty set $X$. Show that $d_p = (d_1^p + d_2^p)^{1/p}$ is a metric on $X$ for $1 \le p < \infty$. Determine whether $d_\infty = \max\{d_1, d_2\}$ is a metric on $X$.

6. Examine whether $\|\cdot\|$ is a norm on $\mathbb{R}^2$, where for each $(x,y) \in \mathbb{R}^2$,

   (a) $\|(x,y)\| = (|x|^p + |y|^p)^{\frac{1}{p}}$, where $0 < p < 1$.

   (b) $\|(x,y)\| = \sqrt{\frac{x^2}{9} + \frac{y^2}{4}}$.

   (c) $\|(x,y)\| = \begin{cases} \sqrt{x^2+y^2} & \text{if } xy \ge 0, \\ \max\{|x|, |y|\} & \text{if } xy < 0. \end{cases}$

7. Let $\|f\| = \min\{\|f\|_\infty, 2\|f\|_1\}$ for all $f \in C[0,1]$. Prove that $\|\cdot\|$ is not a norm on $C[0,1]$.

8. Let $X$ be a normed linear space. Prove that norm of any $x \in X$, can be expressed as $\|x\| = \inf\{|\alpha| : \ \alpha \in \mathbb{C} \setminus \{0\} \text{ with } \|x\| \le |\alpha|\}$.

9. Let $(X, \|\cdot\|)$ be a normed linear space. Show that $\|x\| = \sup\{|\alpha| : \ |\alpha| < \|x\|\}$.

10. Let $(X, \|\cdot\|)$ be a normed linear space and let $p$ be a seminorm on $X$. Show that $p : (X, \|\cdot\|) \to \mathbb{R}$ is continuous if and only if there exists $\alpha > 0$ such that $p(x) \le \alpha\|x\|$ for all $x \in X$.

11. If $1 \le p < q \le \infty$, then show that $\|x\|_q \le \|x\|_p$ for all $x \in \ell^p$.

12. If $\mathbf{x} \in \mathbb{R}^n$, then show that $\lim\limits_{p\to\infty} \|\mathbf{x}\|_p = \|\mathbf{x}\|_\infty$. And if $x \in \ell^p$, then show that $\liminf\limits_{p\to\infty} \|\mathbf{x}\|_p \ge \|\mathbf{x}\|_\infty$.

13. Let $d$ be a metric on a real vector space $X$ satisfying the following two conditions:

    (i) $d(x+z, y+z) = d(x,y)$ for all $x, y, z \in X$,

    (ii) $d(\alpha x, \alpha y) = |\alpha| d(x,y)$ for all $x, y \in X$ and for all $\alpha \in \mathbb{R}$.

    Show that there exists a norm $\|\cdot\|$ on $X$ such that $d(x,y) = \|x-y\|$ for all $x, y \in X$.

14. Let $f$ be a non-negative function on a linear space $X$ such that $f(\alpha x) = |\alpha| f(x)$ for all $\alpha \in \mathbb{C}$. Show that $f$ is norm on $X$ if and only if $f$ is a convex map which can vanish at most at one point.

15. Let $f : (X,d) \to [0,1]$ be continuous map. Show that $f^{-1}(0)$ is a closed $G_\delta$ set.

16. Let $(x_n)$ be a sequence in a normed linear space $X$ which converges to a non-zero vector $x \in X$. Show that $\frac{x_1+\cdots+x_n}{n^\alpha} \to x$ if and only if $\alpha = 1$. What are admissible values of $\alpha$ if $x_n \to 0$?

17. For $x = (x_n) \in l^2$, write $\|x\| = (\sum\limits_{n=1}^{\infty} a_n |x_n|^2)^{1/2}$. Find all possible sequences $(a_n)$ such that $\|\cdot\|$ is a norm on $l^2$.

18. Let $\mathbb{R}^\infty$ be the real vector space of all sequences in $\mathbb{R}$, where addition and scalar multiplication are defined componentwise. Let $d((x_n),(y_n)) = \sum\limits_{n=1}^{\infty} \frac{1}{2^n} \cdot \frac{|x_n-y_n|}{1+|x_n-y_n|}$ for all $(x_n),(y_n) \in \mathbb{R}^\infty$. Show that $d$ is a metric on $\mathbb{R}^\infty$ but that no norm on $\mathbb{R}^\infty$ induces $d$.

19. Let $(X, \|\cdot\|)$ be a nonzero normed vector space. Consider the metrics $d_1, d_2$ and $d_3$ on $X$:

$$d_1(x,y) := \min\{1, \|x-y\|\},$$
$$d_2(x,y) := \frac{\|x-y\|}{1+\|x-y\|},$$
$$d_3(x,y) := \begin{cases} 1+\|x-y\| & \text{if } x \neq y, \\ 0 & \text{if } x = y, \end{cases}$$

for all $x, y \in X$. Prove that none of $d_1, d_2$ and $d_3$ is induced by any norm on $X$.

20. Let $X$ be a normed vector space containing more than one point, let $x, y \in X$ and let $\varepsilon, \delta > 0$. If $B_\varepsilon[x] = B_\delta[y]$, show that $x = y$ and $\varepsilon = \delta$. Does the result remain true if $X$ is assumed to be a metric space? Justify.

21. Let $A = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 < 1\}$ and $B = \{(x,y,z) \in \mathbb{R}^3 : z = 0\}$. Examine whether $A \cap B$ is a closed/an open subset of $\mathbb{R}^3$ with respect to the usual metric on $\mathbb{R}^3$.

22. Let $F_n$ be a sequence of closed sets in $\mathbb{R}$ such that $F_n \subset (n, n+1]$ and $F_n \cap F_m = \varnothing$, whenever $m \neq n$. Show that $F = \bigcup\limits_{n=1}^{\infty} F_n$ is a closed set in $\mathbb{R}$.

23. For all $x, y \in \mathbb{R}$, let $d_1(x,y) = |x-y|$, $d_2(x,y) = \min\{1, |x-y|\}$ and $d_3(x,y) = \frac{|x-y|}{1+|x-y|}$. If $G$ is an open set in any one of the three metric spaces $(\mathbb{R}, d_i)$ $(i = 1,2,3)$, then show that $G$ is also open in the other two metric spaces.

24. Let $X$ be a normed vector space and let $Y (\neq X)$ be a subspace of $X$. Show that $Y$ is not open in $X$.

25. Let $(x_n)$ and $(y_n)$ be Cauchy sequences in a metric space $(X, d)$. Show that the sequence $(d(x_n, y_n))$ is convergent.

26. Let $d_o$ be the discrete metric on nonempty set $X$. Show that $(X, d_o)$ is complete.

27. Let $(x_n)$ be a sequence in a complete metric space $(X, d)$ such that $\sum\limits_{n=1}^{\infty} d(x_n, x_{n+1}) < \infty$. Show that $(x_n)$ converges in $(X, d)$.

28. Let $(x_n)$ be a sequence in a metric space $X$ such that each of the subsequences $(x_{2n})$, $(x_{2n-1})$ and $(x_{3n})$ converges in $X$. Show that $(x_n)$ converges in $X$.

29. Show that the following are incomplete metric spaces.
    (a) $(\mathbb{N}, d)$, where $d(m,n) = |\frac{1}{m} - \frac{1}{n}|$ for all $m, n \in \mathbb{N}$
    (b) $((0,\infty), d)$, where $d(x,y) = |\frac{1}{x} - \frac{1}{y}|$ for all $x, y \in (0, \infty)$
    (c) $(\mathbb{R}, d)$, where $d(x,y) = |\frac{x}{1+|x|} - \frac{y}{1+|y|}|$ for all $x, y \in \mathbb{R}$
    (d) $(\mathbb{R}, d)$, where $d(x,y) = |e^x - e^y|$ for all $x, y \in \mathbb{R}$

30. Examine whether the following metric spaces are complete.
    (a) $([0,1), d)$, where $d(x,y) = |\frac{x}{1-x} - \frac{y}{1-y}|$ for all $x, y \in [0,1)$
    (b) $((-1,1), d)$, where $d(x,y) = |\tan\frac{\pi x}{2} - \tan\frac{\pi y}{2}|$ for all $x, y \in (-1,1)$
    (c) $((0,2], d)$, where $d(x,y) = |\frac{1}{x} - \frac{1}{y}|$

31. For $X(\neq \varnothing) \subset \mathbb{R}$, let $d(x,y) = \frac{|x-y|}{1+|x-y|}$ for all $x, y \in X$. Examine the completeness of the metric space $(X, d)$, where $X$ is
    (a) $[0,1] \cap \mathbb{Q}$.
    (b) $[-1,0] \cup [1,\infty)$.
    (c) $\{n^2 : n \in \mathbb{N}\}$.

32. Let $X = C[0,1]$ be the space all the continuous functions on interval $[0,1]$. Prove that norms $\|\ .\ \|_\infty$ and $\|\ .\ \|_1$ on $X$ are not equivalent.
33. Let $C^1[0,1]$ denote the space of all continuously differentiable functions on $[0,1]$. For $f \in C^1[0,1]$, define $\|f\| = \|f\|_\infty + \|f^{'}\|_\infty$. Show that space $\left(C^1[0,1], \|\ .\ \|\right)$ is a Banach space.
34. The space $\left(C^1[0,1], \|\ .\ \|\right)$, where $\|f\| = \left(\|f\|_2^2 + \|f'\|_2^2\right)^{\frac{1}{2}}$ is complete.
35. Let $f \in C^1[0,1]$ and $\|f\| = \|f'\|_2 + \|f\|_\infty$. Then verify if $\left(C^1[0,1], \|\ .\ \|\right)$ is complete.
36. Let $f \in C^1[0,1]$. Then verify if $\|f\| = \min\left(\|f'\|_2,\ \|f\|_\infty\right)$ defines a norm on $C^1[0,1]$.
37. Let $X = \{f \in C^1[0,1] : f(0) = 0\}$. Then $\|f\| = \|f'\|_2$ is a norm on $C^1[0,1]$ but not complete.
38. Let $D = \{z \in \mathbb{C} : |z| < 1\}$. Let $X$ be the class of all functions $f$ which are analytic on $D$ and continuous on $\bar{D}$. Define $\|f\| = \sup\{|f(e^{it})| : 0 \leq t \leq 2\pi\}$. Show that $(X, \|\ .\ \|)$ is complete.

39. Examine whether the sequence $(f_n)$ is convergent in $(C[0,1], d_\infty)$, where for all $n \in \mathbb{N}$ and for all $t \in [0,1]$,
    (a) $f_n(t) = \frac{nt^2}{1+nt}$.
    (b) $f_n(t) = 1 + t + \frac{t^2}{2!} + \cdots + \frac{t^n}{n!}$.
    (c) $f_n(t) = \begin{cases} nt & \text{if } 0 \leq t \leq \frac{1}{n}, \\ \frac{1}{nt} & \text{if } \frac{1}{n} < t \leq 1. \end{cases}$
    (d) $f_n(t) = \begin{cases} nt & \text{if } 0 \leq t \leq \frac{1}{n}, \\ \frac{n}{n-1}(1-t) & \text{if } \frac{1}{n} < t \leq 1. \end{cases}$
40. Find the pointwise limit of the sequence $f_n(t) = e^{-nt^2} \sin nt$. Examine for uniform convergence of $f_n$ on $\mathbb{R}$.
41. Let $f_n, f : \mathbb{R} \to (0,\infty)$ be such that $f_n \to f$ uniformly on $\mathbb{R}$. Examine for $e^{f_n} \to e^f$ uniformly on $\mathbb{R}$.
42. Let $f_n(t) = \sqrt{t^2 + n}$. Examine for the uniform convergence of $f_n'$ on $\mathbb{R}$.
43. Let $X$ be the class of all continuous functions $f : \mathbb{R} \to \mathbb{C}$ such that for each $\epsilon > 0$, there exists a compact set $K \subset \mathbb{R}$ such that $|f(x)| < \epsilon$, for all $x \in \mathbb{R} \setminus K$. Show that $(X,\ \|\ .\ \|_\infty)$ is complete.
44. Let $1 \leq p < \infty$. Let $X_p$ be a class of all the Riemann integrable functions on $[0,1]$. Prove that $\|f\|_p = \left(\int_0^1 |f|^p\right)^{\frac{1}{p}} < \infty$. Prove that $(X_p,\ \|\ .\ \|_p)$ is a normed linear space but not complete.
45. Show that $\{(x_n) \in l^2 : |x_n| < \frac{1}{n} \text{ for all } n \in \mathbb{N}\}$ is a convex set with empty interior.

46. Suppose that $x \in \ell^p$ for some $p \geq 1$. Show that $\liminf\limits_{p\to\infty} \|x\|_p \geq \|x\|_\infty$. Prove/disprove that $\lim\limits_{p\to\infty} \|x\|_p = \|x\|_\infty$.
47. Let $M$ be a subspace of a normed linear space $X$. Then show that $M$ is closed if and only if $\{y \in M : \|y\| \leq 1\}$ is closed in $X$.
48. Let $T : (C[0, \frac{\pi}{2}], \|.\|_\infty) \to (C[0, \frac{\pi}{2}], \|.\|_\infty)$ be defined by $(Tf)(x) = \int_{s=0}^{x} f(s)\sin s ds$. Show that $T$ is not a contraction but $T^2$ is a contraction.
49. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be continuous and let there exist $\alpha > 0$ such that $\|f(\mathbf{x}) - f(\mathbf{y})\| \geq \alpha\|\mathbf{x} - \mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Show that $f(\mathbb{R}^n)$ is complete.
50. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a contraction and let $g(\mathbf{x}) = \mathbf{x} - f(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$. Show that $g : \mathbb{R}^n \to \mathbb{R}^n$ is one-to-one and onto. Also, show that both $g$ and $g^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ are continuous.
51. Find a neighborhood of $x = 0$ in which initial value problem $y' = \frac{x}{1+y^2}$ with $y(0) = 0$ has a unique solution.

# Problem Set 2

**Problem-set architecture.**

- Start with the true/false diagnostics to separate completeness, compactness, and total boundedness cleanly in your mind.
- The next tier asks for proofs that should explicitly invoke the correct compactness criterion rather than an informal heuristic.
- The final questions are best approached after revisiting Arzelà–Ascoli, uniform convergence, and the basic examples of noncompact bounded families.

1. State whether each of the following statements is true or false, and justify your answer.
   (a) If $X$ is a finite metric space, then $C(X)$, the space of continuous functions on $X$, is a finite-dimensional normed linear space.
   (b) If every countable closed subset of a metric space $(X, d)$ is complete, then $X$ is complete.
   (c) Total boundedness is preserved under homeomorphisms.
   (d) Let $f_n : [0, 1] \to \mathbb{R}$ be defined by $f_n = \chi_{[0,1/n]}$, and suppose that $f_n$ converges pointwise to $f$. Then the set $\{f, f_n : n \in \mathbb{N}\}$ is compact in $B[0, 1]$.
   (e) Let $f_n \in C^1[0, 1]$. Then the set $\{f_n : n \in \mathbb{N}\}$ is compact in $C[0, 1]$.
   (f) Determine whether the set $\{x = (x_1, x_2, \ldots) \in \ell^2 : |x_n| \leq 1/n\}$ is totally bounded in $\ell^2$.
2. Prove that if every countable closed subset of a metric space $X$ is complete, then $X$ is complete.
3. Show that a subset $A$ of a metric space $X$ is closed if and only if $A \cap K$ is compact for every compact subset $K$ of $X$.
4. Find a subset of $\ell^\infty$ that is closed and bounded but not totally bounded.

5. Show that a subset $A$ of a metric space $(X, d)$ is totally bounded if and only if every sequence $(x_n)$ in $A$ has a subsequence $(x_{n_k})$ satisfying

$$d(x_{n_k}, x_{n_{k+1}}) \le 2^{-k} \qquad (k \in \mathbb{N}).$$

6. Let $K$ and $F$ be nonempty subsets of a metric space $(X, d)$. If $K$ is compact and $F$ is closed, show that $\operatorname{dist}(K, F) > 0$ whenever $K \cap F = \varnothing$. Does the same conclusion remain valid if $K$ is closed but not compact?

7. A function $f : (X, d) \to \mathbb{R}$ is called lower semicontinuous if, for each $\alpha \in \mathbb{R}$, the set $\{x \in X : f(x) > \alpha\}$ is open in $X$.
   (a) Show that $f$ is lower semicontinuous if and only if

   $$f(x) \le \liminf_{n\to\infty} f(x_n)$$

   whenever $x_n \to x$.

   (b) If $X$ is a compact metric space, prove that every lower semicontinuous function on $X$ is bounded below and attains its minimum.

8. Let $f : (X, d) \to \mathbb{R}$ be lower semicontinuous. Show directly that for every $x \in X$ and every sequence $x_n \to x$,

$$f(x) \le \liminf_{n\to\infty} f(x_n).$$

9. Let $X$ be a compact metric space, and let $f : X \to X$ satisfy $d(f(x), f(y)) = d(x, y)$ for all $x, y \in X$. Show that $f$ is onto. Is compactness of $X$ necessary?

10. Let $X$ be a compact metric space, and let $f : X \to X$ satisfy $d(f(x), f(y)) \ge d(x, y)$ for all $x, y \in X$. Show that $f$ is an onto isometry.

11. Let $X$ be a compact metric space, and let $f : X \to X$ be bijective with

$$d(f(x), f(y)) \le d(x, y) \qquad (x, y \in X).$$

   Show that $f$ is an isometry.

12. Let $X$ be a compact metric space, and let $\mathcal{F} \subset C(X)$.
    (a) Prove that an equicontinuous family $\mathcal{F}$ is pointwise bounded if and only if it is uniformly bounded.
    (b) Prove that $\mathcal{F}$ is pointwise equicontinuous if and only if it is uniformly equicontinuous.

13. Let $X$ be a compact metric space, and let $(f_n)$ be a sequence in $C(X)$.
    (a) Suppose that $(f_n)$ is equicontinuous and converges pointwise. Show that $(f_n)$ converges uniformly.
    (b) If $(f_n)$ decreases pointwise to 0, show that $(f_n)$ is equicontinuous.

(c) If $(f_n)$ is equicontinuous, show that the set

$$\{x \in X : (f_n(x)) \text{ converges}\}$$

is closed in $X$.

14. For fixed $k > 0$ and $0 < \alpha \leq 1$, define

$$\mathrm{Lip}_{k,\alpha} = \{f \in C[0,1] : |f(x) - f(y)| \leq k|x-y|^\alpha\}.$$

Show that $\{f \in \mathrm{Lip}_{k,\alpha} : f(0) = 0\}$ is a compact subset of $C[0,1]$. Determine whether the set $\{f \in \mathrm{Lip}_{k,\alpha} : \int_0^1 f(t)\,dt = 1\}$ is compact.

15. Let $K(x,t)$ be continuous on the square $[0,1] \times [0,1]$. For $f \in C[0,1]$, define

$$Tf(x) = \int_0^1 f(t)K(x,t)\,dt.$$

Show that $T$ maps bounded sets into equicontinuous sets.

16. Let $f_n \in C[0,1]$ satisfy $\|f_n\|_\infty \leq 1$. Define

$$F_n(x) = \int_0^x f_n(t)\,dt.$$

Show that $(F_n)$ has a convergent subsequence.

17. If $f \in B[0,1]$, show that $B_n(f)(x) \to f(x)$ at each point of continuity of $f$.

18. Give an example of a sequence of functions $f_n \in C[0,1]$ that decreases pointwise to a limit function $f$, but not uniformly.

19. For a given polynomial $p$ and $\varepsilon > 0$, show that there exists a polynomial $q$ with rational coefficients such that $\|p - q\|_\infty < \varepsilon$ on $[0,1]$.

20. Let $(x_i)$ be a sequence in $(0,1)$ such that

$$\frac{1}{n}\sum_{i=1}^n x_i^k$$

converges for each $k = 0, 1, 2, \ldots$. Show that

$$\frac{1}{n}\sum_{i=1}^n f(x_i)$$

converges for every $f \in C[0,1]$.

21. For $f \in C^1[0,1]$ and $\varepsilon > 0$, show that there exists a polynomial $p$ such that

$$\|f - p\|_\infty < \varepsilon \qquad \text{and} \qquad \|f' - p'\|_\infty < \varepsilon.$$

22. Let $f : [1, \infty) \to \mathbb{R}$ be continuous, and suppose that $\lim_{x\to\infty} f(x)$ exists. Given $\varepsilon > 0$, show that there exists a polynomial $p$ such that

$$|f(x) - p(1/x)| < \varepsilon \qquad (x \geq 1).$$

## Problem Set 3

**Problem-set architecture.**

- Read this set in two tracks: first the separation/connectedness questions, then the continuity and image-of-an-interval consequences.
- Many problems can be solved elegantly by choosing the right continuous map and pushing connectedness through the image.
- The later questions are deliberately synthesis-heavy and reward a clear stock of standard examples, especially subsets of $\mathbb{R}^2$.

1. State whether each of the following statements is true or false, and justify your answer.
   (a) It is possible to write $\mathbb{R}^2$ as a countable union of path-connected sets.
   (b) The set of all real polynomials whose zero-set complement is connected is countable.
   (c) There exists a nonempty open connected set $A \subset \mathbb{R}^n$ such that every real-valued function on $A$ is continuous.
   (d) If a metric space $X$ is path-connected, then there exists a continuous map $f : [0, 1] \to X$ that is onto.
   (e) Let $f : (X, d) \to \mathbb{R}$ be such that the graph $G_f = \{(x, f(x)) : x \in X\}$ is connected. Does this imply that $X$ is connected?
   (f) There exists a discontinuous function $f : \mathbb{R} \to \mathbb{R}$ such that the graph $G_f$ is connected in $\mathbb{R}^2$, while $\operatorname{int}(\overline{G_f}) \neq \varnothing$.
2. Let $A$ be a connected subset of a metric space $X$, and let $B$ be a clopen subset of $X$ such that $A \cap B \neq \varnothing$. Show that $A \subseteq B$.
3. If $E$ is a connected subset of a metric space $X$ and $E \subseteq A \cup B$, where $A$ and $B$ are disjoint open subsets of $X$, show that either $E \subseteq A$ or $E \subseteq B$.
4. Prove that $E \subseteq X$ is disconnected if and only if there exist nonempty open sets $A$ and $B$ such that $E = A \cup B$, $A \cap \overline{B} = \varnothing$, and $\overline{A} \cap B = \varnothing$.
5. If every pair of points in $X$ is contained in some connected subset of $X$, show that $X$ itself is connected.
6. If $E$ and $F$ are connected subsets of $X$ with $E \cap F \neq \varnothing$, show that $E \cup F$ is connected.
7. If $E$ and $F$ are nonempty subsets of $X$ such that $E \cup F$ is connected, show that $\overline{E} \cap \overline{F} \neq \varnothing$.
8. Show that the complement of any countable subset of $\mathbb{R}^2$ is path-connected.

9. Prove that $X$ is disconnected if and only if there exists a continuous function $f : X \to \mathbb{R}$ such that $f^{-1}(\{0\}) = \varnothing$, while both $f^{-1}((-\infty, 0))$ and $f^{-1}((0, \infty))$ are nonempty.
10. If $X$ is connected and has at least two points, show that $X$ is uncountable.
11. Let $f : [a, b] \to \mathbb{R}$ have the intermediate value property, and assume that $f^{-1}(\{y\})$ is closed for every $y \in \mathbb{R}$. Show that $f$ is continuous.
12. If $f : [a, b] \to [a, b]$ is continuous, show that $f$ has a fixed point.
13. Let $f : [0, 2] \to \mathbb{R}$ be continuous with $f(0) = f(2)$. Show that there exists $x \in [0, 1]$ such that $f(x) = f(x + 1)$.
14. If $f : \mathbb{R} \to \mathbb{R}$ is continuous and open, show that $f$ is strictly monotone.
15. If $f : \mathbb{R} \to \mathbb{R}$ is continuous and one-to-one, show that $f$ is strictly monotone.
16. Prove that there does not exist a continuous function $f : \mathbb{R} \to \mathbb{R}$ such that $f(\mathbb{Q}) \subseteq \mathbb{R} \setminus \mathbb{Q}$ and $f(\mathbb{R} \setminus \mathbb{Q}) \subseteq \mathbb{Q}$.
17. If $A$ and $B$ are closed subsets of $X$ such that both $A \cup B$ and $A \cap B$ are connected, show that $A$ and $B$ are connected.
18. Let $I = (\mathbb{R} \setminus \mathbb{Q}) \cap [0, 1]$ and $Q = \mathbb{Q} \cap [0, 1]$. Show that there exists a continuous map from $I$ onto $Q$, but that there does not exist a continuous map from $[0, 1]$ onto $Q$.
19. Suppose that $f : \mathbb{R} \to \mathbb{R}$ has the intermediate value property. If its graph $G_f$ is closed, show that $f$ is continuous.
20. If $f : \mathbb{R} \to \mathbb{R}$ is differentiable, show that $f'$ has the intermediate value property.
21. Show that the set
$$\{(x, y) \in \mathbb{R}^2 : x^2 + y^3 \in \mathbb{R} \setminus \mathbb{Q}\}$$
is disconnected in the usual topology of $\mathbb{R}^2$.
22. Show that $GL_n(\mathbb{C})$ is path-connected, using the fact that every polynomial over $\mathbb{C}$ has only finitely many zeros. Is $GL_n(\mathbb{C})$ open in $M_n(\mathbb{C})$?

## Problem Set 4

**Problem-set architecture.**

- Begin with the differentiability diagnostics and small Jacobian computations before attempting the inverse/implicit function problems.
- Several questions are most transparent if you first identify the correct linearization and only then compute coordinates.
- The final tier emphasizes local invertibility, homogeneous maps, and geometric applications; these are intended as capstone problems for the chapter.

1. State whether each of the following statements is true or false, and justify your answer.
   (a) There exists a one-to-one continuous function from $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$ onto $\mathbb{R}^2$.

(b) There exists a function $f : \mathbb{R}^2 \to \mathbb{R}^2$ which is differentiable only at $(1,0)$.

(c) Let $f : \mathbb{R}^2 \to \mathbb{R}$ be such that $f_x(0,0) = 0$. Then there exists some $\delta > 0$ such that $f(x,0)$ is continuous on $(-\delta, \delta)$.

(d) If $f : \mathbb{R}^2 \to \mathbb{R}^2$ is differentiable with $f(0,0) = (1,1)$ and $[f'(0,0)] = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, then there cannot exist a differentiable function $g : \mathbb{R}^2 \to \mathbb{R}^2$ with $g(1,1) = (0,0)$ and $(f \circ g)(x,y) = (y,x)$ for all $(x,y) \in \mathbb{R}^2$.

(e) A continuously differentiable function $f : \mathbb{R}^2 \to \mathbb{R}^2$ cannot be one-to-one and onto if $\det[f'(x,y)] = 0$ for some $(x,y) \in \mathbb{R}^2$.

(f) The equation $\sin(xyz) = z$ defines $x$ implicitly as a differentiable function of $y$ and $z$ locally around the point $(x,y,z) = (\frac{\pi}{2}, 1, 1)$.

2. Let $\Omega$ be an open subset of $\mathbb{R}^n$ and let $f : \Omega \to \mathbb{R}^m$ and $g : \Omega \to \mathbb{R}^m$ be continuous at $\mathbf{x}_0 \in \Omega$. If for each $\varepsilon > 0$, there exist $\mathbf{x}, \mathbf{y} \in B_\varepsilon(\mathbf{x}_0)$ such that $f(\mathbf{x}) = g(\mathbf{y})$, then show that $f(\mathbf{x}_0) = g(\mathbf{x}_0)$.

3. Let $A (\neq \varnothing) \subset \mathbb{R}^n$ be such that every continuous function $f : A \to \mathbb{R}$ is bounded. Show that $A$ is a closed and bounded subset of $\mathbb{R}^n$.

4. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be linear and let $f(\mathbf{x}) = T(\mathbf{x}) \cdot \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$. Find $f'(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^n$.

5. Examine the differentiability of $f$ at $\mathbf{0}$, where

   (a) $f : \mathbb{R}^n \to \mathbb{R}$ satisfies $|f(\mathbf{x})| \leq \|\mathbf{x}\|_2^2$ for all $\mathbf{x} \in \mathbb{R}^n$.

   (b) $f : \mathbb{R}^n \to \mathbb{R}$ is defined by $f(\mathbf{x}) = \|\mathbf{x}\|_2$ for all $\mathbf{x} \in \mathbb{R}^n$.

   (c) $f : \mathbb{R}^n \to \mathbb{R}^n$ is defined by $f(\mathbf{x}) = \|\mathbf{x}\|_2 \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$.

6. Let $\Omega$ be a nonempty open subset of $\mathbb{R}^n$. Let $f : \Omega \to \mathbb{R}$ be differentiable at $\mathbf{x}_0 \in \Omega$, let $f(\mathbf{x}_0) = 0$ and let $g : \Omega \to \mathbb{R}$ be continuous at $\mathbf{x}_0$. Prove that $fg : \Omega \to \mathbb{R}$, defined by $(fg)(\mathbf{x}) = f(\mathbf{x})g(\mathbf{x})$ for all $\mathbf{x} \in \Omega$, is differentiable at $\mathbf{x}_0$.

7. Let $\Omega$ be a nonempty open subset of $\mathbb{R}^n$ and let $g : \Omega \to \mathbb{R}^n$ be continuous at $\mathbf{x}_0 \in \Omega$. If $f : \Omega \to \mathbb{R}$ is such that $f(\mathbf{x}) - f(\mathbf{x}_0) = g(\mathbf{x}) \cdot (\mathbf{x} - \mathbf{x}_0)$ for all $\mathbf{x} \in \Omega$, then show that $f$ is differentiable at $\mathbf{x}_0$.

8. The directional derivatives of a differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$ at $(0,0)$ in the directions of $(1,2)$ and $(2,1)$ are 1 and 2 respectively. Find $f_x(0,0)$ and $f_y(0,0)$.

9. Let $A \in GL(\mathbb{R}^n)$ and $\alpha \geq 2$. If $f : \mathbb{R}^n \to \mathbb{R}^n$ satisfies $\|f(x)\| \leq k\|x\|^\alpha$, for some $k > 0$. Determine whether the map $g = f + A$ is continuously differentiable at $\mathbf{0}$ and $g$ is invertible in the neighborhood of $\mathbf{0}$.

10. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be differentiable such that $f(1,1) = 1$, $f_x(1,1) = 2$ and $f_y(1,1) = 5$. If $g(x) = f(x, f(x,x))$ for all $x \in \mathbb{R}$, determine $g'(1)$.

11. Let $A \in GL_n(\mathbb{C})$. Show that the set $E = \{B \in L_n(\mathbb{C}) : \|B - A\| < \frac{1}{2\|A^{-1}\|}\}$ is open in $GL_n(\mathbb{C})$. Hence deduce that $E$ is path-connected in $L_n(\mathbb{C})$.

12. Prove that a differentiable function $f : \mathbb{R}^n \setminus \{\mathbf{0}\} \to \mathbb{R}^m$ is homogeneous of degree $\alpha \in \mathbb{R}$ (*that is* $f(t\mathbf{x}) = t^\alpha f(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ and for all $t > 0$) if and only if $f'(\mathbf{x})(\mathbf{x}) = \alpha f(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$.
13. Suppose $f : \mathbb{R}^n \to \mathbb{R}^n$ is satisfying $f(rx) = r^{\frac{3}{2}} f(x)$ for all $(x, r) \in \mathbb{R}^n \times (0, \infty)$. Determine whether $f$ is differentiable at $\mathbf{0}$.
14. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be continuously differentiable such that $f_x(a, b) = f_y(a, b)$ for all $(a, b) \in \mathbb{R}^2$ and $f(a, 0) > 0$ for all $a \in \mathbb{R}$. Show that $f(a, b) > 0$ for all $(a, b) \in \mathbb{R}^2$.
15. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be such that $f(tx) = t^2 f(x)$ for every $t > 0$ and $x \in \mathbb{R}^n$. Does it imply that $f$ is differentiable at 0?
16. Let $\Omega$ be an open subset of $\mathbb{R}^n$ such that $\mathbf{a}, \mathbf{b} \in \Omega$ and $S = \{(1-t)\mathbf{a} + t\mathbf{b} : t \in [0, 1]\} \subset \Omega$. If $f : \Omega \to \mathbb{R}^m$ is differentiable at each point of $S$, then show that there exists a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ such that $f(\mathbf{b}) - f(\mathbf{a}) = L(\mathbf{b} - \mathbf{a})$.
17. Let $f(x, y) = (2ye^{2x}, xe^y)$ for all $(x, y) \in \mathbb{R}^2$. Show that there exist open sets $U$ and $V$ in $\mathbb{R}^2$ containing $(0, 1)$ and $(2, 0)$ respectively such that $f : U \to V$ is one-to-one and onto.
18. Let $f(x, y) = (3x - y^2, 2x + y, xy + y^3)$ and $g(x, y) = (2ye^{2x}, xe^y)$ for all $(x, y) \in \mathbb{R}^2$. Examine whether $(f \circ g^{-1})'(2, 0)$ exists (with a meaningful interpretation of $g^{-1}$) and find $(f \circ g^{-1})'(2, 0)$ if it exists.
19. For $n \geq 2$, let $B = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 < 1\}$ and let $f(\mathbf{x}) = \|\mathbf{x}\|_2^2 \mathbf{x}$ for all $\mathbf{x} \in B$. Show that $f : B \to B$ is differentiable and invertible but that $f^{-1} : B \to B$ is not differentiable at $\mathbf{0}$.
20. Let $f : \mathbb{R} \to \mathbb{R}^n$ be a differentiable function with $\|f'(x)\| \leq 1$. Show that $f$ satisfies $\|f(x) - f(y)\| \leq |x - y|$ for every $x, y \in \mathbb{R}$. (*Hint.* Use the one-dimensional mean value theorem.)
21. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be continuously differentiable. Find a suitable condition under which the equation $f(x, y) = 0$ can be solved for $x$ in a neighborhood of $(0, 0)$.
22. Let $f : \mathbb{R} \to \mathbb{R}$ be continuously differentiable and $f'(0) \neq 0$. Show that $F(x, y) = (x - yf(y), f(y))$ is locally invertible in some neighborhood of $(0, 0)$. Does there exist some $f$ for which $F$ is globally invertible?
23. Using implicit function theorem, show that the system of equations
$$\begin{aligned} x^3(y^3 + z^3) &= 0, \\ (x - y)^3 - z^2 &= 7, \end{aligned}$$
can be solved locally near the point $(1, -1, 1)$ for $y$ and $z$ as a differentiable function of $x$.
24. Using implicit function theorem, show that in a neighbourhood of any point $(x_0, y_0, u_0, v_0) \in \mathbb{R}^4$ which satisfies the equations
$$\begin{aligned} x - e^u \cos v &= 0, \\ v - e^y \sin x &= 0, \end{aligned}$$

there exists a unique solution $(u, v) = \varphi(x, y)$ satisfying $\det[\varphi'(x, y)] = v/x$.

25. Show that around the point $(0, 1, 1)$, the equation $xy - z \log y + e^{xz} = 1$ can be solved locally as $y = f(x, z)$ but cannot be solved locally as $z = g(x, y)$.
26. Find the 3rd order Taylor polynomial of $f(x, y, z) = x^2 y + z$ about the point $(1, 2, 1)$.
27. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be continuously differentiable. Show that $f$ is not one-to-one.

# References

[1] J. E. Marsden and M. J. Hoffman, *Elementary Classical Analysis*, 2nd ed., W. H. Freeman, 1993.

[2] P. M. Fitzpatrick, *Advanced Calculus*, 2nd ed., American Mathematical Society, Indian Edition, 2010.

[3] N. L. Carothers, *Real Analysis*, Cambridge University Press, Indian Edition, 2009.

[4] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed., McGraw–Hill, 1976.

[5] E. M. Stein and R. Shakarchi, *Real Analysis: Measure Theory, Integration, and Hilbert Spaces* (Princeton Lectures in Analysis), Princeton University Press, 2005.