Analysis II
Lecture Notes

Work in progress, last updated: October 18, 2022

Joris Roos and Andreas Seeger

J.R., Department of Mathematical Sciences, University of Massachusetts Lowell, 265 Riverside St., Lowell, MA 01854, USA

A.S., Department of Mathematics, University of Wisconsin-Madison, 480 Lincoln Dr, Madison, WI-53706, USA

Contents

Note to students

These are lecture notes for a second undergraduate course in analysis, taught as Math 522 at UW Madison. J.R. prepared a full set of lecture notes for the class in the fall semesters of 2018 and 2019; they were preceded by individual notes on some of the topics, written by A.S. for previous classes. The current version is by no means a final one; all chapters are still undergoing revisions and some will be further expanded. We are grateful to the students of several Math 522 classes for useful questions and remarks on previous versions of the notes.

There is already more content in these notes than we can cover in Math 522, and you will receive updates about the precise lecture contents throughout the course.

The notes in the present form are likely to contain typos, errors and imprecisions of all kinds. Possibly lots. Do not ever take anything that you read in a mathematical text for granted. Think hard about what you are reading and try to make sense of it independently. If that fails, then it's time to ask somebody a question and that usually helps. In the spring semester of 2023 the course will be taught by A.S. He will welcome all comments about the contents of these notes - please let him know about any misprints or inaccuracies that you may find.

There are many books on mathematical analysis, each of which will likely have a large intersection with this course. Here are three very good ones:

- W. Rudin, *Principles of mathematical analysis*
- T. Apostol, *Mathematical analysis: A modern approach to advanced calculus*
- T. Körner, *A Companion to Analysis: A Second First and First Second Course in Analysis*

For further self study in analysis we recommend the Princeton Lectures in Analysis I-IV, by Stein and Shakarchi. Throughout the course A.S. will make concrete suggestions for further reading related to the content of these lecture notes.

- E. M. Stein, R. Shakarchi, *Fourier Analysis,* an introduction.
- _____, *Complex Analysis*
- _____, *Real Analysis* : measure theory, integration, and Hilbert spaces
- _____, *Functional Analysis*

Finally we mention two excellent books used in first year analysis graduate courses at UW Madison.

- W. Rudin, *Real and Complex Analysis*
- G. Folland *Real Analysis,* modern techniques and their applications.

CHAPTER 1

Metric spaces

1. Topology

The notion of a metric space serves as a convenient abstract setting that underlies all topics discussed in this course. A metric space can be thought of as a collection of distinct objects that come with a *distance* between them. This provides a structure that makes it meaningful to speak of notions such as *convergence* and *continuity*. It will allow us to use the same terminology for potentially very different kinds of objects.

DEFINITION 1.1 (Metric space). A set X equipped with a map $d : X \times X \to [0, \infty)$ is called a *metric space* if X is not the empty set and for all $x, y, z \in X$,

(1) $d(x, y) = d(y, x)$,
(2) $d(x, z) \leq d(x, y) + d(y, z)$,
(3) $d(x, y) = 0$ if and only if $x = y$.

d is called a *metric* .

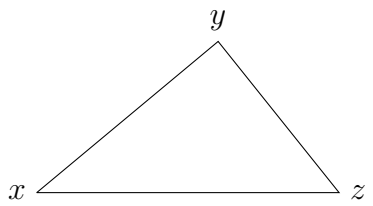

FIGURE 1. Property (2) is called the *triangle inequality*.

One may imagine the d to stand for 'distance'. If multiple metric spaces are relevant at the same time, then we may also write d_X for the metric d on the metric space X.

EXAMPLES 1.2. Some fundamental examples of metric spaces that will be important in this course are

- the real numbers $\mathbb{R}$ with $d(x, y) = |x - y|$,
- closed and open intervals of real numbers (with the same metric),
- the complex numbers $\mathbb{C}$ with $d(z, w) = |z - w|$,
- n-dimensional Euclidean space $\mathbb{R}^n$ consisting of vectors $x = (x_1, \ldots, x_n)$ with the *Euclidean metric*

$$d(x, y) = \Big(\sum_{i=1}^{n} |x_i - y_i|^2 \Big)^{1/2}, \tag{1.1}$$

- the space $C([a, b])$ of continuous functions $[a, b] \to \mathbb{C}$ with

$$d(f, g) = \sup_{x \in [a,b]} |f(x) - g(x)|. \tag{1.2}$$

- the space ℓ^∞ of bounded sequences $(a_n)_{n\in\mathbb{N}}$ of complex numbers with

$$d(a,b) = \sup_{n\in\mathbb{N}} |a_n - b_n| \tag{1.3}$$

- the space c_0 of sequences $(a_n)_{n\in\mathbb{N}}$ of complex numbers with $\lim_{n=0} a_n = 0$, with the same metric as for ℓ^∞.

EXERCISE 1.3. Verify that each of the preceding examples is really a metric space.

In the following let X be a metric space with metric d.

1.1. Open and closed sets. For every $x_0 \in X$ and $r > 0$ define the *open ball*

$$B(x_0, r) = \{x \in X \,:\, d(x, x_0) < r\}, \tag{1.4}$$

and the *closed ball*

$$\overline{B}(x_0, r) = \{x \in X \,:\, d(x, x_0) \le r\}. \tag{1.5}$$

EXAMPLE 1.4. If $X = \mathbb{R}$ (always with the usual metric), then the open balls are open intervals and the closed balls are closed intervals.

Should multiple metric spaces be involved we use subscripts on the metric and balls to indicate which metric space we mean, *i.e.* $B_X(x_0, r)$ is a ball in the metric space X.

DEFINITION 1.5 (Open set). Let X be a metric space and $U \subset X$. A point $x \in U$ is called *interior* in U if there exists $r > 0$ such that $B(x, r) \subset U$.
The set $U \subset X$ is called *open* if every point $x \in U$ is interior.

Clarification of notation: $A \subset B$ means for us that A is a subset of B, not necessarily a proper subset. That is, we also allow $A = B$. We will write $A \subsetneq B$ to refer to proper subsets.

Note that a union of open sets is open. The family $(U)_{U\subset X \text{ open}}$ of open sets is also called the *topology of* X. Note from the definition that the topology of a metric space X is determined by the open balls $(B(x, r))_{x\in X, r>0}$. The notion of open sets can be generalized and leads to the concept of *topological spaces*, which we will not need in this course.

DEFINITION 1.6 (Closed set). A set $A \subset X$ is called closed if its complement $A^{\complement} = X \setminus A$ is open. If $A \subset X$ is an arbitrary set, then $\overline{A}$ denotes the intersection of all closed sets containing A.

Since an intersection of closed sets is closed, $\overline{A}$ is closed by definition and called the *closure* of A. It is the 'smallest' closed set containing A in the sense that if A' is a closed set with $A \subset A'$, then $\overline{A} \subset A'$. As a consequence, a set A is closed if and only if $A = \overline{A}$.

EXERCISE 1.7. Verify that open balls are open and closed balls are closed.

Note that $\overline{B(x_0, r)}$, the closure of the open ball $B(x_0, r)$, often coincides with the closed ball $\overline{B}(x_0, r)$. While this is the case in most of the metric spaces encountered in this course, it is generally only true that

$$B(x_0, r) \subset \overline{B(x_0, r)} \subset \overline{B}(x_0, r). \tag{1.6}$$

EXAMPLE 1.8. Let X be a non-empty set. For $x, y \in X$ let

$$d(x,y) = \begin{cases} 0, & \text{if } x = y, \\ 1, & \text{if } x \neq y. \end{cases} \tag{1.7}$$

This defines a metric on X (called the *trivial metric*). The topology on X is very boring: every set is open (hence also every set is closed). Then, for every $x \in X$, $B(x,1) = \overline{B(x,1)} = \{x\} \subset \overline{B}(x,1) = X$.

DEFINITION 1.9 (Accumulation point). Let $A \subset X$. A point $x \in X$ is called an *accumulation point* of A if for every $r > 0$, there exists $y \in B_r(x) \cap A$ with $y \neq x$.

LEMMA 1.10. *Let X be a metric space and $A \subset X$. Then $\overline{A}$ is equal to the union of A and the set of accumulation points of A.*

PROOF. For one direction we take an arbitrary closed set C containing A and we have to show that every accumulation point x belongs to C. If x were in $X \setminus C$ (an open set not intersecting A) then there would be an $\varepsilon > 0$ and a ball $B(x,\varepsilon)$ such that $B(x,\varepsilon) \subset X \setminus C$ and hence $B(x,\varepsilon) \cap A \subset B(x,\varepsilon) \cap C = \emptyset$, in contradiction to x being an accumulation point. Since C was an arbitrary closed set containing A we find that A and the set of accumulation points are both subsets of $\overline{A}$.

To show the converse let $x \in \overline{A} \setminus A$; we have to show that x is an accumulation point. Again argue by contradiction and suppose that x is not an accumulation point. Then there would exist a ball $B(x,\varepsilon)$ containing no points in A (other than x itself, but that is excluded by assumption). Hence $C = (X \setminus B(x,\varepsilon)) \cap \overline{A}$ would be a closed set containing A, with $C \subsetneq \overline{A}$, a contradiction to the definition of closure of A. □

1.2. Relative topology. If we have a metric space X with metric d and a non-empty subset $A \subset X$, then A can be made a metric space by restricting the metric: we define the metric $d_A : A \times A \to [0,\infty)$ by setting

$$d_A(x,y) = d(x,y) \quad \text{for all } x, y \in A. \tag{1.8}$$

In other words, d_A is the restriction of d to the set $A \times A \subset X \times X$, also denoted by $d_A = d|_{A\times A}$. As a metric space, A comes with its own open sets: unpacking the definition, a set $U \subset A$ is open in A if and only if for every $x \in U$ there exists $r > 0$ such that

$$B_A(x,r) = \{y \in A \,:\, d(x,y) < r\} \subset U. \tag{1.9}$$

Observe that the open balls in A are not necessarily open balls in X. As a consequence, a set $U \subset A$ that is open in A is not necessarily open in X. However, the open sets in A can be characterized by the open sets in X.

LEMMA 1.11. *Let $A \subset X$. A set $U \subset A$ is open in A if and only if there exists an open set $V \subset X$ such that $U = V \cap A$.*

PROOF. Suppose that $U = V \cap A$ with V open. We have to show that U is open in A. Let $x \in U \subset V$ then there is a ball $B(x,r) = \{y \in X : d(x,y) < r_x\}$ contained in V. Then $B_A(x,r_x) \subset U$, so x is an interior point of U (with respect to the metric on A).

Vice versa, let U be open in A. Then for every $x \in U$ there is $r_x > 0$ such that $x \in B_A(x,r_x) \subset U$, and thus $U = \cup_{x\in U} B_A(x,r_x)$. Define $V = \cup_{x\in U} B(x,r_x)$. Then V is open in X and $V \cap A = U$. □

EXAMPLE 1.12. Let $X = \mathbb{R}$, $A = [0,1]$. Then $U = [0,\frac{1}{2}) \subset A \subset X$ is open in A, but not open in $\mathbb{R}$. However, there exists $V \subset \mathbb{R}$ open such that $U = V \cap A$: for example, $V = (-1,\frac{1}{2})$.

1.3. Convergence.

DEFINITION 1.13 (Convergence). Let X be a metric space, $(x_n)_{n\in\mathbb{N}} \subset X$ a sequence and $x \in X$. We say that $(x_n)_{n\in\mathbb{N}}$ *converges* to x if for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n \geq N$ it holds that $d(x_n, x) < \varepsilon$.

If $(x_n)_{n\in\mathbb{N}}$ converges to x we also call x the limit of the sequence and write $x = \lim_{n\to\infty} x_n$; alternatively we may also write that $x_n \to x$ in X.

DEFINITION 1.14 (Cauchy sequence). Let X be a metric space. A sequence $(x_n)_{n\in\mathbb{N}}$ in X is called *Cauchy sequence* if for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n, m \geq N$ we have

$$d(x_n, x_m) < \varepsilon. \tag{1.10}$$

LEMMA 1.15. *Every convergent sequence is a Cauchy sequence.*

PROOF. Let $\varepsilon > 0$. Since $x_n \to x$ there is N so that for all $k \geq N$ we have $d(x_k, x) \leq \varepsilon/2$. If $m \geq N$, $n \geq N$ we get by the triangle inequality

$$d(x_n, x_m) \leq d(x_m, x) + d(x, x_n) < \varepsilon/2 + \varepsilon/2 = \varepsilon \tag{1.11}$$

and since ε was arbitrary the result is proved. □

DEFINITION 1.16 (Completeness). A metric space X is called *complete* if every Cauchy sequence $(x_n)_{n\in\mathbb{N}} \subset X$ converges.

EXAMPLE 1.17. The metric space of rational numbers, $\mathbb{Q}$ (with the usual metric) is not complete: the sequence of rational numbers

$$(10^{-n}\lfloor 10^n\sqrt{2}\rfloor)_{n\in\mathbb{N}} = (1.4, 1.41, 1.414, \dots) \tag{1.12}$$

is a Cauchy sequence, but it does not converge in $\mathbb{Q}$. This is because it converges as a sequence of real numbers to the irrational number $\sqrt{2} \notin \mathbb{Q}$.

The real numbers form an example of a complete metric space (in fact, they are usually defined via *completion* of the rational numbers).

LEMMA 1.18. *If X is complete and $A \subset X$ is closed, then A is a complete metric space.*

PROOF. Let $(x_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in A. Since $d_A = d_X|_{A\times A}$, (x_n) is a Cauchy sequence in X, it has, by assumption, a limit $x \in X$. By Lemma 1.10 $x \in \overline{A}$, but by assumption $\overline{A} = A$. Hence x_n converges to x in A. □

Note that this is not true if X is not complete: for example, every metric space is a closed subset of itself, but not every metric space is complete.

1.4. Continuity.

DEFINITION 1.19 (Continuity). Let X, Y be metric spaces.

(i) A map $f : X \to Y$ is called *continuous at $x \in X$* if for every $\varepsilon > 0$ there exists $\delta > 0$ such that if $d_X(x, y) < \delta$, then $d_Y(f(x), f(y)) < \varepsilon$.

(ii) f is called *continuous* if it is continuous at every $x \in X$. We also write $f \in C(X, Y)$.

LEMMA 1.20. *Let $f : X \to Y$ and $x \in X$. The following are equivalent.*

(i) f continuous at x.

(ii) For every sequence $(x_n)_{n\in\mathbb{N}} \subset X$ convergent to x, the sequence $(f(x_n))_{n\in\mathbb{N}}$ converges to $f(x)$.

PROOF. (i) $\Longrightarrow$ (ii). Suppose $x_n \to x$ in X. Let $\varepsilon > 0$. Since by assumption f is continuous at x we find $\delta_\varepsilon > 0$ such that $d_Y(f(\tilde{x}), f(x)) < \varepsilon$ whenever $d_X(\tilde{x}, x) < \delta_\varepsilon$. Since $\lim_{n\to\infty} x_n = x$, we find N_ε so that $d_X(x_n, x) < \delta_\varepsilon$ for $n > N_\varepsilon$ and thus $d_Y(f(x_n), f(x)) < \varepsilon$ for $n > N_\varepsilon$. Since $\varepsilon > 0$ was arbitrary this shows that $f(x_n) \to f(x)$ in X.

(ii) $\Longrightarrow$ (i). We argue by contradiction and assume that f is not continuous at x. Then there exists an $\varepsilon > 0$ so that there is no $\delta > 0$ with the property that $d_Y(f(\tilde{x}), f(x)) < \varepsilon$ for all $\tilde{x} \in B(x, \delta)$. Taking δ to be a reciprocal of a positive integer, we see that for every $n \in \mathbb{N}$ there exists an x_n with $d_X(x_n, x) < 1/n$ but $d(f(x_n), f(x)) \geq \varepsilon$. Clearly $x_n \to x$ in X but $(f(x_n))_{n\in\mathbb{N}}$ does not converge to $f(x)$; this is a contradicton.

□

DEFINITION 1.21 (Inverse images). Let $f : X \to Y$ be a function and $A \subset Y$. The inverse image of A under f is defined as

$$f^{-1}(A) = \{x \in X : f(x) \in A\}. \tag{1.13}$$

This definition makes sense for all functions from X to Y. The notation does *not* imply that f is invertible. Find examples to illustrate this. Convince yourself that *if* f is invertible then $f^{-1}(A)$ is the image of A under the inverse map.

LEMMA 1.22. *Let $f : X \to Y$ be a map between metric spaces X and Y. The following are equivalent:*

(i) f continuous.
(ii) $f^{-1}(U) \subset X$ is open for every open set $U \subset Y$.

PROOF. (i) $\Longrightarrow$ (ii). Let $U \subset Y$ be open, and let $x \in f^{-1}(U)$. Since U is open there is an open ball $B(f(x), \varepsilon)$ contained in U. Since f is continuous at x there is an open ball $B(x, \delta)$ in X such that $d_Y(f(\tilde{x}), f(x)) < \varepsilon$ for every $\tilde{x} \in B((x, \delta)$. In particular $f(\tilde{x}) \in U$ for every $\tilde{x} \in B(x, \delta)$. Hence $B(x, \delta) \subset f^{-1}(U)$ so that x is an interior point of $f^{-1}(U)$. Since x was chosen arbitrary in $f^{-1}(U)$ we conclude that $f^{-1}(U)$ is open.

(ii) $\Longrightarrow$ (i). Let $\varepsilon > 0$. Let $x \in X$. By assumption $f^{-1}(B(f(x), \varepsilon))$ is open. The point x belongs to this set and thus is an interior point. Hence there exists a $\delta > 0$ (depending on x and ε) so that

$$B(x, \delta) \subset f^{-1}(B(f(x), \varepsilon)). \tag{1.14}$$

This means that $d_Y(f(\tilde{x}), f(x)) < \varepsilon$ provided that $d_X(\tilde{x}, x) < \delta)$ and since $\varepsilon > 0$ was arbitrary and thus f is continuous at x. Since $x \in X$ was arbitrary f is continuous. □

As a consequence of Lemma 1.20 and Lemma 1.22, continuity of f can also be characterized by saying that f *commutes with limits*. That is,

$$\lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n), \tag{1.15}$$

provided that $(x_n)_{n\in\mathbb{N}}$ is a convergent sequence in X.

In this course we will mostly study real- or complex-valued *functions* on metric spaces, i.e. $f : X \to \mathbb{R}$ or $f : X \to \mathbb{C}$. Whether functions are real- or complex-valued is often of little consequence to the heart of the matter. For definiteness we make the convention that functions are always complex-valued, unless specified otherwise. The space of continuous functions will be denoted by $C(X)$, while the space of bounded continuous functions is denoted $C_b(X)$.

1.5. Uniform convergence.

DEFINITION 1.23. A sequence $(f_n)_{n\in\mathbb{N}}$ of functions on a metric space is called *uniformly convergent* to a function f if for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n \geq N$ and all $x \in X$,

$$|f_n(x) - f(x)| < \varepsilon. \tag{1.16}$$

Compare this to *pointwise convergence* . To see the difference between the two it helps to write down the two definitions using the symbolism of predicate logic:

$$\forall \varepsilon > 0\, \exists N \in \mathbb{N}\, \forall x \in X\, \forall n \geq N \,:\, |f_n(x) - f(x)| < \varepsilon. \tag{1.17}$$

$$\forall \varepsilon > 0\, \forall x \in X\, \exists N \in \mathbb{N}\, \forall n \geq N \,:\, |f_n(x) - f(x)| < \varepsilon. \tag{1.18}$$

Formally, the difference is an interchange in the order of universal and existential quantifiers. The first is uniform convergence, where N needs to be chosen independently of x (*uniformly* in x) and the second is pointwise convergence, where N is allowed to depend on x. One can rephrase uniform convergence as follows:

LEMMA 1.24. *Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of functions on a metric space X. Then $(f_n)_{n\in\mathbb{N}}$ converges uniformly to f if and only if* $\lim_{n\to\infty} \sup_{x\in X} |f_n(x) - f(x)| = 0$.

To illustrate the difference between the notions of pointwise convergence and uniform convergence we consider

$$f_n(x) = \begin{cases} 1 - nx & \text{if } 0 \leq x \leq n^{-1} \\ 0 & \text{if } n^{-1} < x \leq 1 \end{cases} \tag{1.19}$$

as a sequence of functions on the metric space $X = [0, 1]$. For every $x \in [0, 1]$ the numerical sequence $\{f_n(x)\}_{n\in\mathbb{N}}$ converges and we have

$$\lim_{n\to\infty} f_n(x) = \begin{cases} 0 & \text{if } 0 < x \leq 1 \\ 1 & \text{if } x = 0 \end{cases}. \tag{1.20}$$

Thus f_n converges to f pointwise. However, for every $n \in \mathbb{N}$

$$\sup_{x\in[0,1]} |f_n(x) - f(x)| = 1 \tag{1.21}$$

and hence f_n does not converge uniformly.

In the following we collect some important facts surrounding uniform convergence that will be used in this lecture. We give the short proofs, or at least sketches. However, if you are feeling a bit rusty on these concepts, all of these are good exercises to try and prove yourself directly from first principles.

LEMMA 1.25. *A sequence $(f_n)_{n\in\mathbb{N}}$ of functions on a metric space X converges uniformly if and only if it is uniformly Cauchy , i.e. for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n, m \geq N$ and all $x \in X$, $|f_n(x) - f_m(x)| < \varepsilon$.*

PROOF. First assume that f_n converges uniformly to f. Then use

$$|f_n(x) - f_m(x)| = |f_n(x) - f(x) + f(x) - f_m(x)| \leq |f_n(x) - f(x)| + |f(x) - f_m(x)| \tag{1.22}$$

to see that f_n is uniformly Cauchy.

Now assume that (f_n) is uniformly Cauchy. Given $\varepsilon > 0$ there is N such that $|f_n(x) - f_m(x)| < \varepsilon/2$ for $n \geq N$, $x \in X$. In particular for every $x \in X$ the numerical sequence $(f_n(x))_{n\in\mathbb{N}}$ is Cauchy. Since all Cauchy sequences in $\mathbb{R}$ and $\mathbb{C}$ converge this numerical sequence has a limit, call it $f(x)$. Letting $m \to \infty$ we see that $\lim_{m\to\infty} |f_n(x) - f_m(x)| = |f_n(x) - f(x)|$ and it follows that for $n \geq N$ we get $|f_n(x) - f(x)| \leq \varepsilon/2$. Since ε is arbitrary this means that (f_n) converges to f uniformly. □

LEMMA 1.26. *If $(f_n)_{n\in\mathbb{N}}$ converges uniformly to f and each f_n is bounded, then f is bounded.*

(Recall that a function $f : X \to \mathbb{C}$ is called *bounded* if there exists $C > 0$ such that $|f(x)| \leq C$ for all $x \in X$.)

PROOF. By the assumed uniform convergence there is N such that $|f_N(x) - f(x)| < 1$ for all $n \geq N$, $x \in X$. Since f_N is bounded there is $M > 0$ such that $|f_n(x)| \leq M$ for all $x \in X$. Now use

$$|f(x)| = |f(x) - f_N(x) + f_N(x)| \leq |f(x) - f_N(x)| + |f_N(x)| < 1 + M. \tag{1.23}$$

□

LEMMA 1.27. *Let $a \in X$. If $(f_n)_{n\in\mathbb{N}}$ converges uniformly to f and each f_n is continuous at a, then f is continuous at a.*

PROOF. We have to show that given $\varepsilon > 0$ there is $\delta > 0$ such that $|f(x) - f(a)| < \varepsilon$ provided that $d(x, a) < \delta$.

Since f_n converges uniformly to f there is an $N \in \mathbb{N}$ such that $|f_n(x) - f(x)| < \varepsilon/3$ for $n \geq N$, and all $x \in X$. Consider the continuous function f_N. There is $\delta > 0$ such that $|f_N(x) - f_N(a)| < \varepsilon/3$ provided that $d(x, a) < \delta$. For such x we get

$$\begin{aligned} &|f(x) - f(a)| = |f(x) - f_N(x) + f_N(x) - f_N(a) + f_N(a) - f(a)| \\ &\leq |f(x) - f_N(x)| + |f_N(x) - f_N(a)| + |f_N(a) - f(a)| < \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon. \end{aligned}$$

□

LEMMA 1.28. *Let X be a metric space. The space of bounded continuous functions $C_b(X)$ is a complete metric space with the supremum metric*

$$d_\infty(f, g) = \sup_{x\in X} |f(x) - g(x)|. \tag{1.24}$$

(Recall that a metric space is *complete* if every Cauchy sequence converges.)

PROOF. We leave the verification of the metric properties to the reader. Let $(f_n)_{n\in\mathbb{N}}$ be a Cauchy sequence. Then (f_n) is uniformly Cauchy and by Fact 1.25 it is uniformly convergent to a limiting function f. By Facts 1.26 and 1.27 f belongs to $C_b(X)$ and thus indeed (f_n) converges to F with respect to the metric in $C_b(X)$. □

Rephrasing Lemma 1.24 we get

LEMMA 1.29. *Let $(f_n)_{n\in\mathbb{N}} \subset C_b(X)$ be a sequence. Then $(f_n)_{n\in\mathbb{N}}$ converges in $C_b(X)$ (with respect to d_∞) if and only if it converges uniformly to f for some $f \in C_b(X)$.*

EXERCISE 1.30. The concept of uniform convergence can be extended to sequences of functions $f_n : X \to Y$ where (Y, d_Y) is a metric space. Extend the above definitions and prove the relevant theorems.

LEMMA 1.31. *Let $(f_n)_{n\in\mathbb{N}} \subset C_b([a, b])$ be a sequence that converges uniformly to f. Then f is Riemann integrable on $[a, b]$ and $\lim_{n\to\infty} \int_a^b f_n = \int_a^b f$.*

PROOF. By Lemma 1.29 f is continuous and bounded on $[a,b]$ and thus f_n, f, $|f_n - f|$ are Riemann integrable. We have by Theorem A.29

$$\Big|\int_a^b f_n - \int_a^b f\Big| = \Big|\int_a^b (f_n - f)\Big|$$
$$\leq \int_a^b |f_n - f| \leq (b-a)\sup_{a\leq t\leq b}|f_n(t) - f(t)| = (b-a)d_\infty(f_n, f)$$

and the conclusion immediately follows from Lemma 1.29 □

The preceding theorem can be generalized as follows.

LEMMA 1.32. *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of Riemann integrable functions on an interval $[a,b]$ that converges uniformly of f on $[a,b]$. Then f is Riemann integrable on $[a,b]$ and $\lim_{n\to\infty}\int_a^b f_n = \int_a^b f$.*

PROOF. Left as an exercise. This requires a closer review of the Riemann integral, see the Appendix, §4 □

LEMMA 1.33. *Let (f_n) be a sequence of functions differentiable in $[a,b]$ such that f_n' is continuous in $[a,b]$ for every $n \in \mathbb{N}$. Suppose that there is $x_0 \in [a,b]$ such that the numerical sequence $(f_n(x_0))_{n\in\mathbb{N}}$ converges and that the sequence of derivatives converges uniformly on $[a,b]$ to a function g.*

Then $(f_n)_{n\in\mathbb{N}}$ converges uniformly to a function f, and f is differentiable with $f'(x) = g(x)$ for all $x \in [a,b]$.

PROOF OF LEMMA 1.33. By the fundamental theorem of calculus we have $f_n(x) - f_n(x_0) = \int_{x_0}^x f_n'(t)dt$. The right hand side converges uniformly to $\int_a^x g(t)dt$, and if $c = \lim_{n\to\infty} f_n(x_0)$ we see that f_n converges uniformly to $f(x) = c + \int_a^x g(t)dt$. Since g is continuous, the function f is differentiable and $f' = g$. □

Careful: If $f_n \to f$ uniformly and f_n is differentiable on $[a,b]$, then this does *not* imply that f is differentiable.

2. The contraction principle

The contraction principle is a powerful tool in analysis.

DEFINITION 1.34. A map $\varphi : X \to X$ is called a *contraction* (of X) if there exists a constant $c \in (0,1)$ such that

$$d(\varphi(x), \varphi(y)) \leq c \cdot d(x,y) \tag{1.25}$$

holds for all $x, y \in X$.

Observe that contractions are continuous.

EXERCISE 1.35. Let $f : (a,b) \to (a,b)$ be differentiable and suppose that $c \in (0,1)$ is such that $|f'(x)| \leq c$ for all $x \in (a,b)$. Show that f is a contraction of (a,b).

THEOREM 1.36 (Banach fixed point theorem). *Let X be a complete metric space and $\varphi : X \to X$ a contraction. Then there exists a unique $x_* \in X$ such that $\varphi(x_*) = x_*$.*

Remark. A point $x \in X$ such that $\varphi(x) = x$ is called a *fixed point* of φ.

PROOF. Uniqueness: Suppose $x_0, x_1 \in X$ are fixed points of φ. Then

$$0 \leq d(x_0, x_1) = d(\varphi(x_0), \varphi(x_1)) \leq c \cdot d(x_0, x_1), \tag{1.26}$$

which implies $d(x_0, x_1) = 0$, since $c \in (0, 1)$. Thus $x_0 = x_1$.
Existence: Pick $x_0 \in X$ arbitrarily and define a sequence $(x_n)_{n\geq 0}$ recursively by

$$x_{n+1} = \varphi(x_n). \tag{1.27}$$

We claim that $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. Indeed, by induction we see that

$$d(x_{n+1}, x_n) \leq cd(x_n, x_{n-1}) \leq c^2 d(x_{n-1}, x_{n-2}) \leq \cdots \leq c^n d(x_1, x_0). \tag{1.28}$$

Thus, for $n < m$ we can use the triangle inequality to obtain

$$d(x_m, x_n) \leq d(x_m, x_{m-1}) + d(x_{m-1}, x_{m-2}) + \cdots + d(x_{n+1}, x_n) \tag{1.29}$$

$$= \sum_{i=n}^{m-1} d(x_{i+1}, x_i) \leq \sum_{i=n}^{m-1} c^i d(x_1, x_0) \leq d(x_1, x_0) \sum_{i=n}^{\infty} c^i = c^n \frac{d(x_1, x_0)}{1-c}. \tag{1.30}$$

Thus, $d(x_m, x_n)$ converges to 0 as $m > n \to \infty$. This shows that $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. By completeness of X, it must converge to a limit which we call $x_* \in X$. By continuity of φ,

$$\varphi(x_*) = \varphi(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} \varphi(x_n) = \lim_{n\to\infty} x_{n+1} = x_*. \tag{1.31}$$

□

Remarks. 1. The proof not only demonstrates the existence of the fixed point x_*, but also gives an algorithm to compute it via successive applications of the map φ. We can say something about how quickly the algorithm converges: the sequence $(x_n)_{n\in\mathbb{N}}$ defined in the proof satisfies the inequality

$$d(x_n, x_*) \leq \frac{c^n}{1-c} d(x_0, x_1), \tag{1.32}$$

so speed of convergence depends only on the parameter $c \in (0, 1)$ and the quality of the initial guess $x_0 \in X$.
2. The contraction principle can be used to solve equations. For example, say we want to solve $F(x) = 0$ (F is some function). Then we can set $G(x) = F(x) + x$. Then $F(x) = 0$ if and only if x is a fixed point of G.
3. The conclusion does not necessarily hold if we drop the assumption that X is complete: the map $f : (0, 1) \to (0, 1)$ defined by $f(x) = x/2$ is a contraction (in which metric space?) but has no fixed point.
4. If we replace the contraction assumption (1.25) by the weaker condition

$$|\varphi(x) - \varphi(y)| < |x - y| \tag{1.33}$$

for all x, y with $x \neq y$ then φ may not have a fixed point in X. Consider $\varphi(x) = x + e^{-x}$ on the complete metric space $X = [0, \infty)$. One verifies that $\varphi'(x) = 1 - e^{-x} \in (0, 1)$ for $x \geq 0$ thus (1.33) is satisfied if $x \neq y$. But clearly $\varphi(x) - x > 0$ for $x \geq 0$ so φ does not have a fixed point.

EXERCISE 1.37. We are given $h \in C([0, 1])$ and $K \in C([0, 1]^2)$ such that $|K(x, t)| \leq 3/4$ for $(x, t) \in [0, 1]^2$. Consider the integral equation

$$f(x) = \int_0^1 K(x, t) f(t) dt + h(x), \quad x \in [0, 1] \tag{1.34}$$

Show that there exists a unique function continuous in $[0,1]$ such that (1.34) holds. Follow the following steps.

(i) Define for $f \in C([0,1])$

$$T[f](x) = \int_0^1 K(x,t)f(t)dt + h(x) \tag{1.35}$$

.

(ii) Show that T maps $C([0,1])$ to $C([0,1])$.

(iii) Show that $\sup_{x\in[0,1]} |T[f](x) - T[g](x)| \leq \frac{3}{4} \sup_{t\in[0,1]} |f(t) - g(t)|$ and conclude.

Exercise 1.38. Let $A > 0$. We are given $h \in C([0,A])$ and $K \in C([0,A]^2)$ such that $|K(x,t)| \leq B$ for all $(x,t) \in [0,A]^2$.

Consider the *Volterra integral equation*

$$f(x) = \int_0^x K(x,t)f(t)dt + h(x), \quad x \in [0,A] \tag{1.36}$$

Show that there exists a unique function continuous in $[0,A]$ such that (1.36) holds. Fill in the details for the following steps.

(i) Define for $f \in C([0,A])$

$$V[f](x) = \int_0^x K(x,t)f(t)dt + h(x) \tag{1.37}$$

.

(ii) Show that V maps $C([0,A])$ to $C([0,A])$.

(iii) Given a positive number M define a metric on $C([0,A])$ by

$$d_M(f,g) = \sup_{x\in[0,A]} |f(x) - g(x)|e^{-Mx}. \tag{1.38}$$

Show that $C([0,A])$ with this metric is a complete metric space. Show that a sequence $(f_n)_{n\in\mathbb{N}}$ of functions in $C([0,A])$ converges uniformly if and only if it converges with respect to d_M.

(iv) Show

$$d(V[f],V[g]) \leq \frac{B}{M} d(f,g) \tag{1.39}$$

so that with the choice $M > B$ the map V becomes a contraction on $(C([0,A]), d_M)$.

Remark. The preceding example shows that a smart choice of the metric or metric space can be crucial in solving such equations. This can often present highly nontrivial problems in applications.

Exercise 1.39. Show that the system of equations

$$x_1 + \tfrac{1}{10}\cos(\sin(2x_2 + x_1)) = 6$$
$$x_2 + \tfrac{1}{12}e^{-x_1^2} + \tfrac{1}{10}\cos(x_1 + x_2) = 7$$

has a unique solution $(x_1, x_2) \in \mathbb{R}^2$.

Exercise 1.40. (i) Show there is exactly one $u \in C([-1,1])$ that satisfies the integral equation

$$u(x) \;=\; x \int_0^x t^2 \cos\big(u(t)\big)\, dt, \qquad x \in [-1,1].$$

Hint: Use the contraction principle in the space $C([-1,1])$.

(ii) Show that u is differentiable. Is u' differentiable?

EXERCISE 1.41. Let $f : \mathbb{R} \to \mathbb{R}$ be a C^1-function, such that $|f'(x)| \leq a < 1$ for all $x \in \mathbb{R}$. Define a C^1-function $g : \mathbb{R}^2 \to \mathbb{R}^2$ by $g(x, y) = (x + f(y), y + f(x))$. Show that the range of g is all of $\mathbb{R}^2$.

EXERCISE 1.42. Show that there exists a unique $(x, y) \in \mathbb{R}^2$ such that $\cos(\sin(x)) = y$ and $\sin(\cos(y)) = x$.

3. Compactness

The goal in this section is to study the general theory of compactness in metric spaces. From Analysis I, you might already be familiar with compactness in $\mathbb{R}$. By the Heine-Borel theorem, a subset of $\mathbb{R}^n$ is compact if and only if it is bounded and closed. We will see that this no longer holds in general metric spaces. We will also study in detail compact subsets of the space of continuous functions $C(K)$ where K is a compact metric space (Arzelà-Ascoli theorem). Let (X, d) be a metric space. We first review some basic definitions.

DEFINITION 1.43. A collection $(G_i)_{i\in I}$ (I is an arbitrary index set) of open sets $G_i \subset X$ is called an *open cover of* X if $X \subset \bigcup_{i\in I} G_i$.

DEFINITION 1.44. X is *compact* if every open cover of X contains a finite subcover. That is, if for every open cover $(G_i)_{i\in I}$ there exists $m \in \mathbb{N}$ and $i_1, \ldots, i_m \in I$ such that $X \subset \bigcup_{j=1}^m G_{i_j}$. This is also called the *Heine-Borel property* .

DEFINITION 1.45. A subset $A \subset X$ is called *compact* if $(A, d|_{A\times A})$ is a compact metric space.

THEOREM 1.46 (Heine-Borel). *A subset $A \subset \mathbb{R}$ is compact if and only if A is closed and bounded.*

This theorem also holds for subsets of $\mathbb{R}^n$ but not for subsets of general metric spaces. We will later identify this as a special case of a more general theorem.

DEFINITION 1.47. A subset $A \subset X$ is called *relatively compact* or *precompact* if the closure $\overline{A} \subset X$ is compact.

EXAMPLES 1.48.
- If X is finite, then it is compact.
- $[a, b] \subset \mathbb{R}$ is compact. $[a, b), (a, b) \subset \mathbb{R}$ are relatively compact.
- $\{x \in \mathbb{R}^n : \sum_{i=1}^n |x_i|^2 = 1\} \subset \mathbb{R}^n$ is compact.
- The set of orthogonal $n \times n$ matrices with real entries $O(n, \mathbb{R})$ is compact as a subset of $\mathbb{R}^{n^2}$.
- For general X, the closed ball

$$\overline{B}(x_0, r) = \{x \in X : d(x, x_0) \leq r\} \subset X \tag{1.40}$$

is *not* necessarily compact (examples later).

As a warm-up in dealing with the definition of compactness let us prove the following.

LEMMA 1.49. *A closed subset of a compact metric space is compact.*

PROOF. Let $(G_i)_{i\in I}$ be an open cover of a closed subset $A \subset X$. That is, $G_i \subset A$ is open with respect to A. Then $G_i = U_i \cap A$ for some open $U_i \subset X$ (see Theorem 2.30 in Rudin's book). Note that $X\backslash A$ is open. Thus,

$$\{U_i : i \in I\} \cup \{X\backslash A\} \tag{1.41}$$

is an open cover of X, which by compactness has a finite subcover $\{U_{i_k} : k = 1, \dots, M\} \cup \{X \backslash A\}$. Then $\{G_{i_k} : k = 1, \dots, M\}$ is an open cover of A. □

EXERCISE 1.50. A collection $\{F_\alpha : \alpha \in A\}$ of closed sets has the *finite intersection property* if for every finite subset A_o of A the intersection $\cap_{\alpha \in A_o} F_\alpha$ is not empty.

Prove that the following statements (i), (ii) are equivalent.

(i) A metric space X, with metric d, is compact.

(ii) For every collection $\{F_\alpha\}_{\alpha \in A}$ of closed sets with the finite intersection property it follows that

$$\bigcap_{\alpha \in A} F_\alpha \neq \emptyset.$$

EXERCISE 1.51. Let X be a compact metric space. Prove that there exists a countable, dense set $E \subset X$ (recall that $E \subset X$ is called *dense* if $\overline{E} = X$).

EXERCISE 1.52. Construct a compact subset of real numbers whose accumulation points form a countable set.

3.1. Compactness and continuity. We will now prove three key theorems that relate compactness to continuity. In Analysis I you might have seen versions of these on $\mathbb{R}$ or $\mathbb{R}^n$. The proofs are not very interesting, but can serve as instructive examples of how to prove statements involving the Heine-Borel property.

THEOREM 1.53. *Let X, Y be metric spaces and assume that X is compact. If a map $f : X \to Y$ is continuous, then it is uniformly continuous.*

PROOF. Let $\varepsilon > 0$. We need to demonstrate the existence of a number $\delta > 0$ such that for all $x, y \in X$ we have that $d_X(x, y) \leq \delta$ implies $d_Y(f(x), f(y)) \leq \varepsilon$. By continuity, for every $x \in X$ there exists a number $\delta_x > 0$ such that for all $y \in X$, $d_X(x, y) \leq \delta_x$ implies $d_Y(f(x), f(y)) \leq \varepsilon/2$. Let

$$B_x = B(x, \delta_x/2) = \{y \in X : d_X(x, y) < \delta_x/2\}. \tag{1.42}$$

Then $(B_x)_{x \in X}$ is an open cover of X. By compactness, there exists a finite subcover by $B_{x_1}, \dots, B_{x_m}$. Now we set

$$\delta = \tfrac{1}{2} \min(\delta_{x_1}, \dots, \delta_{x_m}). \tag{1.43}$$

We claim that this δ does the job. Indeed, let $x, y \in X$ satisfy $d_X(x, y) \leq \delta$. There exists $i \in \{1, \dots, m\}$ such that $x \in B_{x_i}$. Then

$$d_X(x_i, y) \leq d_X(x_i, x) + d_X(x, y) \leq \tfrac{1}{2}\delta_{x_i} + \delta \leq \delta_{x_i}. \tag{1.44}$$

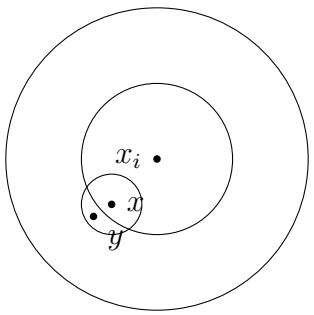

FIGURE 2. The balls B_{x_i}, $B(x_i, \delta_{x_i})$, $B(x, \delta)$.

Thus, by definition of δ_{x_i},

$$d_Y(f(x), f(y)) \leq d_Y(f(x), f(x_i)) + d_Y(f(x_i), f(y)) \leq \varepsilon/2 + \varepsilon/2 = \varepsilon. \tag{1.45}$$

□

THEOREM 1.54. *Let X, Y be metric spaces and assume that X is compact. If a map $f : X \to Y$ is continuous, then $f(X) \subset Y$ is compact.*

Note that for $A \subset X$ we have $A \subset f^{-1}(f(A))$ and for $B \subset Y$ we have $f(f^{-1}(B)) \subset B$, but equality need not hold in either case.

PROOF. Let $(V_i)_{i\in I}$ be an open cover of $f(X)$. Since f is continuous, the sets $U_i = f^{-1}(V_i) \subset X$ are open. We have $f(X) \subset \bigcup_{i\in I} V_i$. So,

$$X \subset f^{-1}(f(X)) \subset \bigcup_{i\in I} f^{-1}(V_i) = \bigcup_{i\in I} U_i. \tag{1.46}$$

Thus $(U_i)_{i\in I}$ is an open cover of X and by compactness there exists a finite subcover $\{U_{i_1}, \ldots, U_{i_m}\}$. That is,

$$X \subset \bigcup_{k=1}^{m} U_{i_k} \tag{1.47}$$

Consequently,

$$f(X) \subset \bigcup_{k=1}^{m} f(U_{i_k}) \subset \bigcup_{k=1}^{m} V_{i_k}. \tag{1.48}$$

Thus $\{V_{i_1}, \ldots, V_{i_m}\}$ is an open cover of $f(X)$. □

THEOREM 1.55. *Let X be a compact metric space and $f : X \to \mathbb{R}$ a continuous function. Then there exists $x_0 \in X$ such that $f(x_0) = \sup_{x\in X} f(x)$.*

By passing from f to $-f$ we see that the theorem also holds with sup replaced by inf.

PROOF. By Theorem 1.54, $f(X) \subset \mathbb{R}$ is compact. By the Heine-Borel Theorem 1.46, it is therefore closed and bounded. By completeness of the real numbers, $f(X)$ has a finite supremum $\sup f(X)$ and since $f(X)$ is closed we have $\sup f(X) \in f(X)$, so there exists $x_0 \in X$ such that $f(x_0) = \sup f(X) = \sup_{x\in X} f(x)$. □

COROLLARY 1.56. *Let X be a compact metric space. Then every continuous function on X is bounded: $C(X) = C_b(X)$.*

For a converse of this statement, see Exercise 1.104 below.

PROOF. Let $f \in C(X)$. Then $|f| : X \to [0, \infty)$ is also continuous. By Theorem 1.55 there exists $x_0 \in X$ such that $|f(x_0)| = \sup_{x\in X} |f(x)|$. Set $C = |f(x_0)|$. Then $|f(x)| \leq C$ for all $x \in X$, so f is bounded. □

3.2. Sequential compactness and total boundedness.

DEFINITION 1.57. A metric space X is *sequentially compact* if every sequence in X has a convergent subsequence. This is also called the *Bolzano-Weierstrass property* .

Let us recall the Bolzano-Weierstrass theorem which you should have seen in Analysis I.

THEOREM 1.58 (Bolzano-Weierstrass). *Every bounded sequence in $\mathbb{R}$ has a convergent subsequence.*

DEFINITION 1.59. A metric space X is *bounded* if it is contained in a single fixed ball, i.e. if there exist $x_0 \in X$ and $r > 0$ such that $X \subset B(x_0, r)$.

DEFINITION 1.60. A metric space X is *totally bounded* if for every $\varepsilon > 0$ there exist finitely many balls of radius ε that cover X.

Similarly, we define these terms for subsets $A \subset X$ by considering $(A, d|_{A\times A})$ as its own metric space.

Note that

$$X \text{ totally bounded} \implies X \text{ bounded}. \tag{1.49}$$

The converse is generally false. However, for $A \subset \mathbb{R}^n$ we have that A is totally bounded if and only if A is bounded.

THEOREM 1.61. *Let X be a metric space. The following are equivalent:*

(1) X is compact
(2) X is sequentially compact
(3) X is totally bounded and complete

COROLLARY 1.62.

(1) (Heine-Borel Theorem) A subset $A \subset \mathbb{R}^n$ is compact if and only if it is bounded and closed.
(2) (Bolzano-Weierstrass Theorem) A subset $A \subset \mathbb{R}^n$ is sequentially compact if and only if it is bounded and closed.

PROOF OF COROLLARY 1.62. A subset $A \subset \mathbb{R}^n$ is closed if and only if A is complete as a metric space (this is because $\mathbb{R}^n$ is complete). Also, $A \subset \mathbb{R}^n$ is bounded if and only if it is totally bounded. Therefore, both claims follow from Theorem 1.61. □

EXAMPLE 1.63. Let ℓ^∞ be the space of bounded sequences $(a_n)_{n\in\mathbb{N}} \subset \mathbb{C}$ with $d(a, b) = \sup_{n\in\mathbb{N}} |a_n - b_n|$ (that is, $\ell^\infty = C_b(\mathbb{N})$). We claim that the closed unit ball around $0 = (0, 0, \dots)$,

$$\overline{B}(0, 1) = \{a \in \ell^\infty : |a_n| \leq 1 \, \forall n \in \mathbb{N}\} \tag{1.50}$$

is bounded and closed, but *not* compact. Indeed, let $e^{(k)} \in \ell^\infty$ be the sequence with

$$e^{(k)}_n = \begin{cases} 0, & k \neq n, \\ 1, & k = n. \end{cases} \tag{1.51}$$

Then, $e^{(k)} \in \overline{B}(0, 1)$ for all $k = 1, 2, \dots$ but $(e^{(k)})_k \subset \overline{B}(0, 1)$ does not have a convergent subsequence, because $d(e^{(k)}, e^{(j)}) = 1$ for all $k \neq j$ and therefore no subsequence can be Cauchy. Thus $\overline{B}(0, 1)$ is not sequentially compact and by Theorem 1.61 it is not compact.

EXAMPLE 1.64. Let ℓ^1 be the space of complex sequences $(a_n)_{n\in\mathbb{N}} \subset \mathbb{C}$ such that $\sum_n |a_n| < \infty$. We define a metric on ℓ^1 by

$$d(a, b) = \sum_n |a_n - b_n| \tag{1.52}$$

EXERCISE 1.65. Show that the closed and bounded set $\overline{B}(0, 1) \in \ell^1$ is not compact.

PROOF OF THEOREM 1.61. X compact $\Rightarrow$ X sequentially compact: Suppose that X is compact, but not sequentially compact. Then there exists a sequence $(x_n)_{n\in\mathbb{N}} \subset X$ without a convergent subsequence. Let $A = \{x_n : n \in \mathbb{N}\} \subset X$. Note that A must be an infinite set (otherwise $(x_n)_{n\in\mathbb{N}}$ has a constant subsequence). Since A has no accumulation points, we have that for every x_n there is an open ball B_n such that $B_n \cap A = \{x_n\}$. Also, A is a closed set, so $X\backslash A$ is open. Thus, $\{B_n : n \in \mathbb{N}\} \cup \{X\backslash A\}$ is an open cover of X. By compactness of X, it has a finite subcover, but that is a contradiction since A is an infinite set.

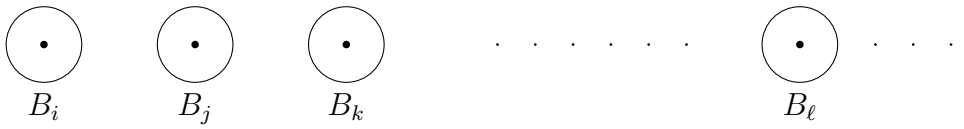

FIGURE 3.

X sequentially compact $\Rightarrow$ X totally bounded and complete: Suppose X is sequentially compact. Then it is complete, because every Cauchy sequence that has a convergent subsequence must converge (prove this!). Suppose that X is not totally bounded. Then there exists $\varepsilon > 0$ such that X cannot be covered by finitely many ε-balls.
Claim: There exists a sequence $p_1, p_2, \ldots$ in X such that $d(p_i, p_j) \geq \varepsilon$ for all $i \neq j$.
Proof of claim. Pick p_1 arbitrarily and then proceed inductively: say that we have constructed $p_1, \ldots, p_n$ already. Then there exists p_{n+1} such that $d(p_i, p_{n+1}) \geq \varepsilon$ for all $i = 1, \ldots, n$ since otherwise we would have $\bigcup_{i=1}^{n} B(p_i, \varepsilon) \supset X$. $\square$
Now it remains to observe that the sequence $(p_n)_{n\in\mathbb{N}}$ has no convergent subsequence (no subsequence can be Cauchy). Contradiction! Thus, X is totally bounded.

X totally bounded and complete $\Rightarrow$ X sequentially compact: Assume that X is totally bounded and complete. Let $(x_n)_{n\in\mathbb{N}} \subset X$ be a sequence. We will construct a convergent subsequence. First we cover X by finitely many 1-balls. At least one of them, call it B_0, must contain infinitely many of the x_n (that is, $x_n \in B_0$ for infinitely many n), so there is a subsequence $(x_n^{(0)})_n \subset B_0$. Next, cover X by finitely many $\frac{1}{2}$-balls. There is at least one, B_1, that contains infinitely many of the $x_n^{(0)}$. Thus there is a subsequence $(x_n^{(1)})_n \subset B_1$. Inductively, we obtain subsequences $(x_n^{(0)})_n \supset (x_n^{(1)})_n \supset \ldots$ of $(x_n)_{n\in\mathbb{N}}$ such that $(x_n^{(k)})_n$ is contained in a ball of radius 2^{-k}. Now let $a_n = x_n^{(n)}$. Then $(a_n)_{n\in\mathbb{N}}$ is a subsequence of $(x_n)_{n\in\mathbb{N}}$.
Claim: $(a_n)_{n\in\mathbb{N}}$ is a Cauchy sequence.
Proof of claim. Let $\varepsilon > 0$ and N large enough so that $2^{-N+1} < \varepsilon$. Then for $m > n \geq N$ we have

$$d(a_m, a_n) \leq 2 \cdot 2^{-n} \leq 2^{-N+1} < \varepsilon, \tag{1.53}$$

because $a_n, a_m \in B_n$ and B_n is a ball of radius 2^{-n}. $\square$
Since X is complete, the Cauchy sequence $(a_n)_{n\in\mathbb{N}}$ converges.

X sequentially compact $\Rightarrow$ X compact: Assume that X is sequentially compact. Let $(G_i)_{i\in I}$ be an open cover of X.
Claim: There exists $\varepsilon > 0$ such that every ball of radius ε is contained in one of the G_i.
Proof of claim. Suppose not. Then for every $n \in \mathbb{N}$ there is a ball B_n of radius

$\frac{1}{n}$ that is not contained in any of the G_i. Let p_n be the center of B_n. By sequential compactness, the sequence $(p_n)_{n\in\mathbb{N}}$ has a convergent subsequence $(p_{n_k})_{k\in\mathbb{N}}$ with some limit $p \in X$. Let $i_0 \in I$ be such that $p \in G_{i_0}$. Since G_{i_0} is open there exists $\delta > 0$ such that $B(p,\delta) \subset G_{i_0}$. Let k be large enough such that $d(p_{n_k}, p) < \delta/2$ and $\frac{1}{n_k} < \delta/2$. Then $B_{n_k} \subset B(p,\delta)$ because if $x \in B_{n_k}$, then

$$d(p,x) \le d(p,p_{n_k}) + d(p_{n_k},x) < \delta/2 + \delta/2 = \delta. \tag{1.54}$$

Thus, $B_{n_k} \subset B(p,\delta) \subset G_{i_0}$.

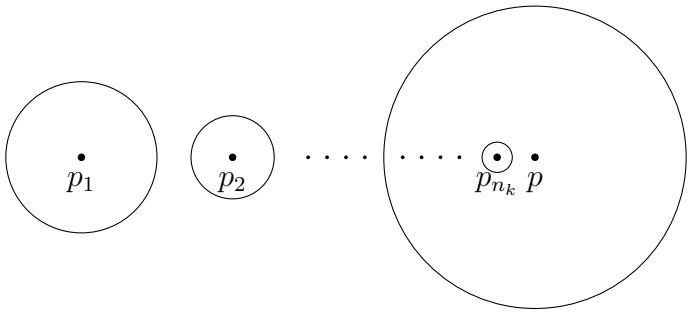

FIGURE 4.

This is a contradiction, because we assumed that the B_n are not contained in any of the G_i. □

Now let $\varepsilon > 0$ be such that every ε-ball is contained in one of the G_i. We have already proven earlier that X is totally bounded if it is sequentially compact. Thus there exist $p_1, \dots, p_M$ such that the balls $B(p_j,\varepsilon)$ cover X. But each $B(p_j,\varepsilon)$ is contained in a G_i, say in G_{i_j}, so we have found a finite subcover:

$$X \subset \bigcup_{j=1}^{M} B(p_j,\varepsilon) \subset \bigcup_{j=1}^{M} G_{i_j}. \tag{1.55}$$

□

COROLLARY 1.66. *Compact subsets of metric spaces are bounded and closed.*

COROLLARY 1.67. *Let X be a complete metric space and $A \subset X$. Then A is totally bounded if and only if it is relatively compact.*

EXERCISE 1.68. Prove this.

3.3. Equicontinuity and the Arzelà-Ascoli theorem. Let (K,d) be a compact metric space. By Corollary 1.56, continuous functions on K are automatically bounded. Thus, $C(K) = C_b(K)$ is a complete metric space with the supremum metric

$$d_\infty(f,g) = \sup_{x\in K} |f(x) - g(x)| \tag{1.56}$$

(see Fact 1.28). Convergence with respect to d_∞ is uniform convergence (see Fact 1.29). In this section we ask ourselves when a subset $\mathcal{F} \subset C(K)$ is compact.

EXAMPLE 1.69. Let $\mathcal{F} = \{f_n : n \in \mathbb{N}\} \subset C([0,1])$, where

$$f_n(x) = x^n, \quad x \in [0,1]. \tag{1.57}$$

$\mathcal{F}$ is not compact, because no subsequence of $(f_n)_{n\in\mathbb{N}}$ converges. This is because the pointwise limit

$$f(x) = \begin{cases} 0, & x \in [0,1), \\ 1, & x = 1. \end{cases} \tag{1.58}$$

is not continuous, i.e. not in $C([0,1])$.

The key concept that characterizes compactness in $C(K)$ is equicontinuity.

DEFINITION 1.70 (Equicontinuity). A subset $\mathcal{F} \subset C(K)$ is called *equicontinuous* if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ for all $f \in \mathcal{F}$, $x, y \in K$ with $d(x,y) < \delta$.

DEFINITION 1.71. $\mathcal{F} \subset C(K)$ is called *uniformly bounded* if there exists $C > 0$ such that $|f(x)| \leq C$ for all $x \in K$ and $f \in \mathcal{F}$.
$\mathcal{F} \subset C(K)$ is called *pointwise bounded* if for all $x \in K$ there exists $C = C(x) > 0$ such that $|f(x)| \leq C$ for all $f \in \mathcal{F}$.

Note that $\mathcal{F} \subset C(K)$ is uniformly bounded if and only if it is bounded (as a metric space, see Definition 1.59). We have

$$\mathcal{F} \text{ uniformly bounded } \Rightarrow \mathcal{F} \text{ pointwise bounded.} \tag{1.59}$$

The converse is false in general.

LEMMA 1.72. *If $(f_n)_{n\in\mathbb{N}} \subset C(K)$ is uniformly convergent (on K), then $\{f_n : n \in \mathbb{N}\}$ is equicontinuous.*

PROOF. Let $\varepsilon > 0$. By uniform convergence there exists $N \in \mathbb{N}$ such that

$$\sup_{x\in K} |f_n(x) - f_N(x)| \leq \varepsilon/3 \tag{1.60}$$

for $n \geq N$. By uniform continuity (using Theorem 1.53) there exists $\delta > 0$ such that

$$|f_k(x) - f_k(y)| \leq \varepsilon/3 \tag{1.61}$$

for all $x, y \in K$ with $d(x,y) < \delta$ and all $k = 1, \ldots, N$. Thus, for $n \geq N$ and $x, y \in K$ with $d(x,y) < \delta$ we have

$$|f_n(x) - f_n(y)| \leq |f_n(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f_n(y)| \leq 3 \cdot \varepsilon/3 = \varepsilon. \tag{1.62}$$

□

LEMMA 1.73. *If $\mathcal{F} \subset C(K)$ is pointwise bounded and equicontinuous, then it is uniformly bounded.*

PROOF. Choose $\delta > 0$ such that

$$|f(x) - f(y)| \leq 1 \tag{1.63}$$

for all $d(x,y) < \delta$, $f \in \mathcal{F}$. Since K is totally bounded (by Theorem 1.61) there exist $p_1, \ldots, p_m \in K$ such that the balls $B(p_j, \delta)$ cover K. By pointwise boundedness, for every $x \in K$ there exists $C(x)$ such that $|f(x)| \leq C(x)$ for all $f \in \mathcal{F}$. Set

$$C = \max\{C(p_1), \ldots, C(p_m)\}. \tag{1.64}$$

Then for $f \in \mathcal{F}$ and $x \in K$,

$$|f(x)| \leq |f(p_j)| + |f(x) - f(p_j)| \leq C + 1, \tag{1.65}$$

where j is chosen such that $x \in B(p_j, \delta)$. □

THEOREM 1.74 (Arzelà-Ascoli). *A subset $\mathcal{F}$ of $C(K)$ is totally bounded if and only if it is pointwise bounded and equicontinuous.*

PROOF OF NECESSITY. We show that if $\mathcal{F}$ is a totally bounded subset of $C(K)$ then $\mathcal{F}$ is pointwise bounded and equicontinuous.

By the definition of $\mathcal{F}$ totally bounded there are functions $f_1, \ldots, f_N$ in $\mathcal{F}$ so that for every $f \in \mathcal{F}$ there is an index $i \in \{1, \ldots, N\}$ with $\sup_{x\in K} |f_i(x) - f(x)| < \varepsilon/4$. Clearly for every $x \in K$,

$$|f(x)| \le |f_i(x)| + |f(x) - f_i(x)| \le \max_{i=1,\ldots,N} \sup_{x\in K} |f_i(x)| + \varepsilon/4 \tag{1.66}$$

so that $\mathcal{F}$ is pointwise bounded.

Now we show the equicontinuity of the family $\mathcal{F}$. By Theorem 1.53 each f_i is uniformly continuous. Thus for each i there exists a $\delta_i > 0$ such that $|f_i(x) - f_i(x')| < \varepsilon/2$ whenever $d_K(x, x') < \delta_i$. Let $\delta = \min\{\delta_1, \ldots, \delta_N\}$. Then $\delta > 0$ and we have $|f_i(x) - f_i(x')| < \varepsilon/2$ for every i whenever $d_K(x, x') < \delta$.

Now pick any $f \in \mathcal{F}$, and let i be so that $\sup_{x\in K} |f_i(x) - f(x)| < \varepsilon/4$, and let x, x' be so that $d_K(x, x') < \delta$. Then

$$\begin{aligned} |f(x) - f(x')| &\le |f(x) - f_i(x)| + |f_i(x) - f_i(x')| + |f_i(x') - f(x')| \\ &\le \varepsilon/4 + |f_i(x) - f_i(x')| + \varepsilon/4 < \varepsilon\,. \end{aligned}$$

□

PROOF OF SUFFICIENCY. We show that if $\mathcal{F} \subset C(K)$ is equicontinuous and pointwise bounded then $\mathcal{F}$ is totally bounded.

Fix $\varepsilon > 0$. We shall first find a finite collection $\mathcal{G}$ of functions in $\mathcal{B}(K)$ so that for every $f \in \mathcal{F}$ there exists a $g \in \mathcal{G}$ with $\sup_{x\in K} |f(x) - g(x)| < \varepsilon$.

Let $\delta > 0$ so that for all $f \in \mathcal{F}$ we have $|f(x) - f(x')| < \varepsilon/4$ whenever $|x - x'| < \delta$. Again we use the compactness of K and cover K with finitely many balls $B(x_i, \delta)$, $i = 1, \ldots, L$. There is M_i so that $|f(x_i)| \le M_i$ for all $f \in \mathcal{F}$. Let $M = 1 + \max_{i=1,\ldots,L} M_i$.

We now let $A_1 = B(x_1, \delta)$, and $A_i = B(x_i, \delta) \setminus \cup_{\nu=1}^{i-1} B(x_\nu, \delta)$, for $2 \le i \le L$. (Some of the A_i could be empty but that does not matter).

Let $\mathcal{Z}^L(M, \varepsilon)$ be the set of L-tuples $\vec{n}$ of integers $\vec{n} = (n_1, \ldots, n_L)$ with the property that $|n_i|\varepsilon/4 \le M$ for $i = 1, \ldots, L$. Note that $\mathcal{Z}^L(M, \varepsilon)$ is a finite set (indeed its cardinality is $\le (8M\varepsilon^{-1} + 1)^L$).

We now define a collection $\mathcal{G}$ of functions which are constant on the sets A_i (these are analogues of step functions). Namely given $\vec{n}$ in $\mathcal{Z}^L(M, \varepsilon)$ we let $g^{\vec{n}}$ be the unique function that takes the value $n_i\varepsilon/4$ on the set A_i (provided that that set is nonempty). Clearly the cardinality of $\mathcal{G}$ is not larger than the cardinality of $\mathcal{Z}^L(M, \varepsilon)$.

Let $f \in \mathcal{F}$. Consider an A_i which by construction is a subset of $B(x_i, \delta)$. Then $|f(x) - f(x_i)| < \varepsilon/4$ for all $x \in A_i$ (this condition is vacuous if A_i is empty). Now $|f(x_i)| \le M_i \le M$ and therefore there exists an integer n_i with the property that $-M \le n_i\varepsilon/4 \le M$ and $|f(x_i) - n_i\varepsilon/4| < \varepsilon/4$. Then we also have that for $i = 1, \ldots, L$ and for every $x \in A_i$,

$$|f(x) - n_i\varepsilon/4| \le |f(x) - f(x_i)| + |f(x_i) - n_i\varepsilon/4| < \varepsilon/4 + \varepsilon/4 = \varepsilon/2.$$

This implies that for this choice of $\vec{n} = (n_1, \ldots, n_L)$ we get $\sup_{x\in K} |f(x) - g^{\vec{n}}(x)| < \varepsilon/2$.

Finally, we need to find a finite cover of $\mathcal{F}$ with ε-balls centered at points in $\mathcal{F}$. Consider the subcollection $\widetilde{\mathcal{G}}$ of functions in $\mathcal{G}$ for which the ball of radius $\varepsilon/2$ centered at g contains a function in $\mathcal{F}$. Denote the functions in $\widetilde{\mathcal{G}}$ by $g_1, \ldots, g_N$. The balls

of radius $\varepsilon/2$ centered at $g_1, \dots g_N$ cover $\mathcal{F}$. For $i = 1, \dots, N$ pick $f_i \in \mathcal{F}$ so that $\sup_{x\in K} |g_i(x) - f_i(x)| < \varepsilon/2$. By the triangle inequality (for the norm in $\mathcal{B}(K)$ whose restriction to $C(K)$ is also the norm in $C(K)$) the ball of radius $\varepsilon/2$ centered at g_i is contained in the ball of radius ε centered at f_i. Thus the balls of radius ε centered at f_i, $i = 1, \dots, N$ cover the set $\mathcal{F}$. □

We get as a corollary of the theorem of Arzela-Ascoli a characterization of compactness.

COROLLARY 1.75. *A closed subset $\mathcal{F}$ of $C(K)$ is compact if and only if it is pointwise bounded and equicontinuous.*

PROOF. Recall that the space $C(K)$ is complete. Since we now assume that $\mathcal{F}$ is closed in $C(K)$ the metric space $\mathcal{F}$ is complete. Thus by the characterization of compactness ($\mathcal{F}$ compact $\iff$ $\mathcal{F}$ totally bounded and complete) the corollary follows from the theorem. □

COROLLARY 1.76. *An equicontinuous and bounded sequence $\{f_n\}$ of functions in $C(K)$ has a uniformly convergent subsequence.*

PROOF. The closure of $\mathcal{F} = \{f_n : n \in \mathbb{N}\}$ is bounded, complete, and equicontinuous, thus compact. By a part of the theorem on the characterization of compactness it is also sequentially compact, therefore f_n has a convergent subsequence. □

We now discuss a special case of the Arzelá-Ascoli theorem.

COROLLARY 1.77. *Let $\mathcal{F} \subset C([a, b])$ be such that*

(i) $\mathcal{F}$ is bounded (i.e. uniformly bounded),

(ii) every $f \in \mathcal{F}$ is continuously differentiable and

$$\mathcal{F}' = \{f' : f \in \mathcal{F}\} \tag{1.67}$$

is bounded.

Then $\mathcal{F}$ is totally bounded.

PROOF OF COROLLARY 1.77. Using the mean value theorem we see that for all $x, y \in [a, b]$ there exists $\xi \in [a, b]$ such that

$$f(x) - f(y) = f'(\xi)(x - y). \tag{1.68}$$

But since $\mathcal{F}'$ is bounded there exists $C > 0$ such that

$$|f'(\xi)| \le C \tag{1.69}$$

for all $f \in \mathcal{F}, \xi \in [a, b]$. Thus,

$$|f(x) - f(x)| \le C|x - y| \tag{1.70}$$

for all $x, y \in [a, b]$ and all $f \in \mathcal{F}$. This implies equicontinuity: for $\varepsilon > 0$ we set $\delta = C^{-1}\varepsilon$. Then for $x, y \in [a, b]$ with $|x - y| < \delta$ we have

$$|f(x) - f(y)| \le C|x - y| < C\delta = \varepsilon. \tag{1.71}$$

Therefore the claim follows from Theorem 1.74. □

EXAMPLE 1.78. Let $\mathcal{F} = \{x \mapsto \sum_{n=0}^{\infty} c_n x^n : |c_n| \le 1\} \subset C([-\frac{1}{2}, \frac{1}{2}])$. The set $\mathcal{F}$ is bounded, because

$$\Big| \sum_{n=0}^{\infty} c_n x^n \Big| \le \sum_{n=0}^{\infty} 2^{-n} = 2. \tag{1.72}$$

for all sequences $(c_n)_{n\in\mathbb{N}}$ with $|c_n| \leq 1$ and for all $x \in [-1/2, 1/2]$. Similarly,

$$\mathcal{F}' = \Big\{ \sum_{n=1}^{\infty} n c_n x^{n-1} \,:\, |c_n| \leq 1 \Big\} \tag{1.73}$$

is also bounded. Thus, $\mathcal{F} \subset C([-\frac{1}{2}, \frac{1}{2}])$ is relatively compact. However, note that the $\mathcal{F}$ interpreted as a subset of $C([0,1])$ (with the understanding that convergence at $x = 1$ is also assumed) is not relatively compact (it contains the set in Example 1.69).

EXAMPLE 1.79. The set

$$\mathcal{F} = \Big\{ \sin(\pi n x) \,:\, n \in \mathbb{Z} \Big\} \subset C([0,1]) \tag{1.74}$$

is bounded, but not relatively compact. Indeed, suppose it is. Then by Arzelà-Ascoli it is equicontinuous, so there exists $\delta > 0$ such that for all $n \in \mathbb{N}$ and for all $x, y \in [0,1]$ with $|x - y| < \delta$ we have $|\sin(\pi n x) - \sin(\pi n y)| < 1/2$. Set $x = 0$ and $y = 1/(2n)$ for $n > \delta^{-1}/2$. Then $|\sin(\pi n x) - \sin(\pi n y)| = 1$. Contradiction!

EXAMPLE 1.80. Condition (i) from Corollary 1.77 is necessary, because relatively compact sets are bounded. Condition (ii) however is not necessary. Consider for example $\mathcal{F} = \{f_n : n = 1, 2, \dots\} \subset C([0,1])$ with $f_n(x) = \sin(nx)/\sqrt{n}$. The set $\mathcal{F}$ is bounded, but $\mathcal{F}'$ is unbounded. But the sequence $(f_n)_{n\in\mathbb{N}}$ is uniformly convergent, so by Fact 1.72, $\mathcal{F}$ is equicontinuous and hence relatively compact.

4. Covering numbers and Minkowski dimension*

DEFINITION 1.81. Let E be a totally bounded subset of a metric space X, i.e. for every $\delta > 0$ it is contained in a finite collection of δ-balls.

For $\delta > 0$ let $\mathcal{N}(E, \delta)$ be the minimal number of δ-balls needed to cover E (the centers of these balls are not required to belong to E). This number is called the *δ-covering number of E*; note that it depends not only on E but also on the underlying metric space X and the given metric d. The function $\delta \mapsto \log \mathcal{N}(E, \delta)$ is called the *metric entropy function* of E.

The definition of $\mathcal{N}(E, \delta)$ is extended to sets that are not totally bounded if we allow the value ∞. If E is not totally bounded then there exists a δ_0 such that $\mathcal{N}(E, \delta) = \infty$ for $\delta < \delta_0$.

One is interested in the behavior of $\mathcal{N}(E, \delta)$ for small δ. For compact E this serves as a *quantitative measure of compactness.*

DEFINITION 1.82. Let E be totally bounded. The number

$$\overline{\dim}_M(E) = \limsup_{\delta \to 0+} \frac{\log \mathcal{N}(E, \delta)}{\log(\frac{1}{\delta})}$$

is called the *upper Minkowski dimension* (also known as *Box counting dimension* or *upper metric dimension of E.*) The expression

$$\underline{\dim}_M(E) = \liminf_{\delta \to 0+} \frac{\log \mathcal{N}(E, \delta)}{\log(\frac{1}{\delta})}$$

is called *lower Minkowski dimension* or lower box counting or metric metric dimension of E. If $\underline{\dim}(E) = \overline{\dim}(E) = \alpha$ we say that E has Minkowski dimension α.

EXAMPLE 1.83. Let $k \leq n$ and let E denote a k-dimensional box in $\mathbb{R}^n$:

(1.75)
$$E = [0,1]^k \times \{0\}^{n-k} = \{x \in \mathbb{R}^n \,:\, x_j \in [0,1] \text{ for } 1 \leq j \leq k,\ x_j = 0 \text{ for } k < j \leq n\}.$$

Then there exist constants $c, c' > 0$ such that

$$c'\delta^{-k} \leq \mathcal{N}(E,\delta) \leq c\delta^{-k} \tag{1.76}$$

for all $\delta \in (0,1)$. Hence E has Minkowski dimension k.

EXERCISE 1.84. Let $A \subset \mathbb{R}^n$ be a compact set. Show that there exists a constant $c \in (0,\infty)$ such that

$$\mathcal{N}(E,\delta) \leq c \cdot \delta^{-n} \tag{1.77}$$

holds for all $\delta > 0$. Hence $\overline{\dim}_M E \leq n$

EXERCISE 1.85. (i) Show that if we replace the natural log in the above definitions by another $\log_b$ with base $b > 1$ then the definitions of the dimensions do not change.

(ii) Let $\alpha > 0$. Suppose that for every $\varepsilon > 0$ there is a $\delta(\varepsilon) > 0$ and a positive constant $C_\varepsilon \geq 1$ such that $C_\varepsilon^{-1}\delta^{-\alpha+\varepsilon} \leq \mathcal{N}(E,\delta) \leq C_\varepsilon\delta^{-\alpha-\varepsilon}$ for $0 < \delta < \delta(\varepsilon)$. Show that E has Minkowski dimension α.

(iii) Let $E \subset X$ be totally bounded and let $\overline{E}$ be the closure of E. Then $\overline{E}$ is totally bounded and we have

$$\mathcal{N}(E,\delta) \leq \mathcal{N}(\overline{E},\delta) \leq \mathcal{N}(E,\delta') \text{ if } 0 < \delta' < \delta.$$

(iv) Define $N^{\text{cent}}(E,\delta)$ to be the minimal number of δ-balls with center in E needed to cover E. Show that

$$\mathcal{N}(E,\delta) \leq N^{\text{cent}}(E,\delta) \leq \mathcal{N}(E,\delta/2).$$

(v) Let $B_1,\ldots,B_M$ be balls of radius δ in X, so that each ball has nonempty intersection with the set E. For each $i = 1,\ldots,M$ denote by B_i^* the ball with same center as B_i and radius 3δ. Assume that the balls $B_1^*,\ldots,B_M^*$ are disjoint. Prove that $M \leq \mathcal{N}(E,\delta)$.

Remark: This can be an effective tool to prove lower bounds for the covering numbers.

EXERCISE 1.86. Consider the following metrics in $\mathbb{R}^n$.

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

(i) Let $E \subset \mathbb{R}^n$ and let $\mathcal{N}_1(E,\delta), \mathcal{N}_2(E,\delta), \mathcal{N}_\infty(E,\delta)$ be the metric entropy numbers of E associated with to the metrics d_1, d_2, d_∞, respectively. Show that

$$\mathcal{N}_\infty(E,\delta) \leq \mathcal{N}_2(E,\delta) \leq \mathcal{N}_1(E,\delta) \leq \mathcal{N}_2(E,\delta/\sqrt{n}) \leq \mathcal{N}_\infty(E,\delta/n).$$

(iii) Let $Q = [0,1]^n$ be the unit cube in $\mathbb{R}^n$. Show that Q has Minkowski dimension n (with respect to any of the metrics d_1, d_2, d_3).

(iv) Let f be a differentiable function on $[0,1]$ with bounded derivative. Let E be the set of all $x = (x_1,x_2) \in \mathbb{R}^2$ for which $0 \leq x_1 \leq 1$ and $x_2 = f(x_1)$. What is the Minkowski dimension of E?

(v) Let E be the set of all $x = (x_1,x_2) \in \mathbb{R}^2$ for which $0 \leq x_1 \leq 1$ and $x_2 = \sqrt{x_1}$. What is the Minkowski dimension of E?

EXERCISE 1.87. Let $\beta > 0$. Consider the subset E of $\mathbb{R}$ consisting of the numbers $n^{-\beta}$, for $n = 1, 2, \ldots$. Show that E has a Minkowski dimension and determine it.

Hint: It might help to try this first for the sequence $1/n$ which, perhaps counterintuitively, turns out to have Minkowski dimension $\frac{1}{2}$.

EXAMPLE 1.88. The Cantor middle third set $\mathfrak{C}$ is given as a the subset of $[0, 1]$ consisting of numbers of the form

$$\sum_{k=1}^{\infty} a_k 3^{-k} \text{ where } a_k \in \{0, 2\}. \tag{1.78}$$

It can be written as

$$\mathfrak{C} = [0, 1] \setminus \bigcup_{\ell=0}^{\infty} \bigcup_{k=0}^{3^\ell - 1} \left(\tfrac{3k+1}{3^{\ell+1}}, \tfrac{3k+2}{3^{\ell+1}}\right). \tag{1.79}$$

$\mathfrak{C}$ is a compact subset of $[0, 1]$, with the property that for each N there are 2^N disjoint closed intervals of length 3^{-N} which cover $\mathfrak{C}$.

EXERCISE 1.89. Show that $\mathfrak{C}$ has Minkowski dimension $\frac{\log 2}{\log 3}$.

EXERCISE 1.90. Let A be the space of functions $f : \mathbb{N} \to \mathbb{R}$ (aka sequences) so that $|f(n)| \leq 2^{-n}$ for all $n \in \mathbb{N}$. It is a subset of the space of bounded sequences with norm $\|f\|_\infty = \sup_{n\in\mathbb{N}} |f(n)|$ and associated metric d_∞. Show that for $\delta < 1/2$ the covering numbers $\mathcal{N}(A, \delta)$ satisfy the bounds

$$\mathcal{N}(A, \delta) \leq \left(\frac{1}{\delta}\right)^{C + \frac{1}{2}\log_2 \frac{1}{\delta}}$$

where C is independent of δ. *Hint:* It helps to work with $\delta = 2^{-M}$ where $M \in \mathbb{N}$.

Also provide a lower bound which shows that A does not have finite lower Minkowski dimension.

5. Oscillation as a quantification of discontinuity*

In this section let (X, d) be a metric space and $f : X \to \mathbb{R}$ be a function.

DEFINITION 1.91. (i) Let $f : X \to \mathbb{R}$. For each $x \in X$ and $\delta > 0$ we form the expressions

$$M_{f,\delta}(x) = \sup\{f(y) : d(x, y) < \delta, \quad y \in X\}$$
$$m_{f,\delta}(x) = \inf\{f(y) : d(x, y) < \delta, \quad y \in X\}$$

Observe that, for fixed $x \in X$, $m_{f,\delta}(x)$ increases in δ as δ decreases. Moreover $M(f, \delta)(x)$ decreases in δ as δ decreases. Thus $M_{f,\delta}(x) - m_{f,\delta}(x)$ is a nonnegative quantity which decreases as δ decreases. Hence the limit as $\delta \to 0+$ exists.

DEFINITION 1.92. We call the quantity

$$\operatorname{osc}_f(x) = \lim_{\delta\to 0+} M_{f,\delta}(x) - m_{f,\delta}(x) \tag{1.80}$$

the oscillation of f at x.

The number $\operatorname{osc}_f(x)$ can be used to quantify discontinuities:

LEMMA 1.93. *Let $f : X \to \mathbb{R}$ be a bounded function. Then f is continuous at x if and only if $\operatorname{osc}_f(x) = 0$.*

PROOF. This is a consequence of the definition of continuity. □

LEMMA 1.94. *Let $f : X \to \mathbb{R}$ be a bounded function. Then for every $\gamma \geq 0$ the set $\{x : \mathrm{osc}_f(x) \geq \gamma\}$ is closed.*

PROOF. The conclusion is shown by proving that the complement

$$\Omega_\gamma = \{x : \mathrm{osc}_f(x) < \gamma\} \tag{1.81}$$

is open. Let $x \in \Omega_\gamma$ and choose ε such that $0 < \varepsilon < \gamma - \mathrm{osc}_f(x)$. By the definition of $\mathrm{osc}_f(x)$ we can pick $\delta > 0$ such that $M_{f,\delta}(x) - m_{f,\delta}(x) < \mathrm{osc}_f(x) + \varepsilon$. If $d(y, x) < \delta/2$ and $d(z, y) < \delta/2$ then $d(z, x) < \delta$ and thus $M_{f,\delta/2}(y) \leq M_{f,\delta}(x)$ and $m_{f,\delta/2}(y) \geq m_{f,\delta}(x)$. Hence

$$\mathrm{osc}_f(y) \leq M_{f,\delta/2}(y) - m_{f,\delta/2}(y) \leq M_{f,\delta}(x) - m_{f,\delta}(x) < \mathrm{osc}_f(x) + \varepsilon < \gamma$$

so that $B(x, \delta/2) \subset \Omega_\gamma$. Hence x is an interior point of Ω_γ and since x was chosen arbitrarily in Ω_γ this set is open. □

EXERCISE 1.95. Define $f : [-10, 10] \to \mathbb{R}$ by $f(x) = -4x$ for $x \leq 0$, $f(x) = \sin(\pi/x)$ for $0 < x < 3/2$, $f(x) = \cos(\pi/x)$ for $x \geq 3/2$. Determine $\mathrm{osc}_f(x)$ for all $x \in [-10, 10]$.

EXERCISE 1.96. Consider Thomae's function $f : [0, 1] \to \mathbb{R}$, defined by

$$f(x) = \begin{cases} 0 & \text{if } x \in [0,1] \setminus \mathbb{Q}, \\ \frac{1}{n} & \text{if } x = \frac{m}{n} \text{ with } \gcd(m, n) = 1. \end{cases} \tag{1.82}$$

Find $\mathrm{osc}_f(x)$ for all $x \in [0, 1]$.

6. Further exercises

EXERCISE 1.97. Let (X, d) be a metric space and $A \subset X$ a subset.
(i) Show that A is totally bounded if and only if $\overline{A}$ is totally bounded.
(ii) Assume that X is complete. Show that A is totally bounded if and only if A is relatively compact. Which direction is still always true if X is not complete?

EXERCISE 1.98. Let ℓ^1 denote space of all sequences $(a_n)_{n\in\mathbb{N}}$ of complex numbers such that $\sum_{n=1}^\infty |a_n| < \infty$, equipped with the metric $d(a, b) = \sum_{n=1}^\infty |a_n - b_n|$.
(i) Prove that

$$A = \{a \in \ell^1 \,:\, \sum_{n=1}^\infty |a_n| \leq 1\} \tag{1.83}$$

is bounded and closed, but not compact.
(ii) Let $b \in \ell^1$ with $b_n \geq 0$ for all $n \in \mathbb{N}$. Show that

$$B = \{a \in \ell^1 \,:\, |a_n| \leq b_n \,\forall\, n \in \mathbb{N}\} \tag{1.84}$$

is compact.

EXERCISE 1.99. Recall that ℓ^∞ is the metric space of bounded sequences of complex numbers equipped with the supremum metric $d(a, b) = \sup_{n\in\mathbb{N}} |a_n - b_n|$. Let $s \in \ell^\infty$ be a sequence of non–negative real numbers that converges to zero. Let

$$A = \{a \in \ell^\infty \,:\, |a_n| \leq s_n \text{ for all } n\}. \tag{1.85}$$

Prove that $A \subset \ell^\infty$ is compact.

EXERCISE 1.100. For each of the following subsets of $C([0,1])$ prove or disprove compactness:
(i) $A_1 = \{f \in C([0,1]) : \max_{x\in[0,1]} |f(x)| \leq 1\}$,
(ii) $A_2 = A_1 \cap \{p : p \text{ polynomial of degree } \leq d\}$ (where $d \in \mathbb{N}$ is given)
(iii) $A_3 = A_1 \cap \{f : f \text{ is a power series with infinite radius of convergence}\}$

EXERCISE 1.101. Let $\mathcal{F} \subset C([a,b])$ be a bounded set. Assume that there exists a function $\omega : [0,\infty) \to [0,\infty)$ such that

$$\lim_{t\to 0+} \omega(t) = \omega(0) = 0. \tag{1.86}$$

and for all $x, y \in [a,b]$, $f \in \mathcal{F}$,

$$|f(x) - f(y)| \leq \omega(|x-y|). \tag{1.87}$$

Show that $\mathcal{F} \subset C([a,b])$ is relatively compact.

EXERCISE 1.102. For $1 \leq p < \infty$ we denote by ℓ^p the space of sequences $(a_n)_{n\in\mathbb{N}}$ of complex numbers such that $\sum_{n=1}^{\infty} |a_n|^p < \infty$. Define a metric on ℓ^p by

$$d(a,b) = \left(\sum_{n\in\mathbb{N}} |a_n - b_n|^p \right)^{1/p}. \tag{1.88}$$

The purpose of this exercise is to prove a theorem of Fréchet that characterizes compactness in ℓ^p. Let $\mathcal{F} \subset \ell^p$.
(i) Assume that $\mathcal{F}$ is bounded and *equisummable* in the following sense: for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$\sum_{n=N}^{\infty} |a_n|^p < \varepsilon \text{ for all } a \in \mathcal{F}. \tag{1.89}$$

Then show that $\mathcal{F}$ is totally bounded.
(ii) Conversely, assume that $\mathcal{F}$ is totally bounded. Then show that it is equisummable in the above sense.
Hint: Mimick the proof of Arzelà-Ascoli.

EXERCISE 1.103. Let $C^k([a,b])$ denote the space of k-times continuously differentiable functions on $[a,b]$ endowed with the metric

$$d(f,g) = \sum_{j=0}^{k} \sup_{x\in[a,b]} |f^{(j)}(x) - g^{(j)}(x)|. \tag{1.90}$$

Let $0 \leq \ell < k$ be integers and consider the canonical embedding map

$$\iota : C^k([a,b]) \to C^\ell([a,b]) \text{ with } \iota(f) = f. \tag{1.91}$$

Prove that if $B \subset C^k([a,b])$ is bounded, then the image $\iota(B) = \{\iota(f) : f \in B\} \subset C^\ell([a,b])$ is relatively compact. *Hint:* Use the Arzelà-Ascoli theorem.

EXERCISE 1.104. Let X be a metric space. Assume that for every continuous function $f : X \to \mathbb{C}$ there exists a constant $C_f > 0$ such that $|f(x)| \leq C_f$ for all $x \in X$. Show that X is compact. *Hint:* Assume that X is not sequentially compact and construct an unbounded continuous function on X.

EXERCISE 1.105. Consider $\mathcal{F} = \{f_N : N \in \mathbb{N}\} \subset C([0,1])$ with

$$f_N(x) = \sum_{n=0}^{N} b^{-n\alpha} \sin(b^n x), \tag{1.92}$$

where $0 < \alpha < 1$ and $b > 1$ are fixed.
(a) Show that $\mathcal{F}$ is relatively compact in $C([0,1])$.
(b) Show that $\mathcal{F}'$ is not a bounded subset of $C([0,1])$.
(c) Show that there exists $c > 0$ such that for all $x, y \in \mathbb{R}$ and $N \in \mathbb{N}$,

$$|f_N(x) - f_N(y)| \leq c|x - y|^\alpha. \tag{1.93}$$

EXERCISE 1.106. Suppose (X, d) is a metric space with a countable dense subset, i.e. a set $A = \{x_1, x_2, \dots\} \subset X$ with $\overline{A} = X$. Let ℓ^∞ denote the metric space of bounded sequences $a = (a_n)_{n\in\mathbb{N}}$ of real numbers with metric $d_\infty(a, b) = \sup_{n\in\mathbb{N}} |a_n - b_n|$. Show that there exists a map $\iota : X \to \ell^\infty$ with $d_\infty(\iota(x), \iota(y)) = d(x, y)$ for every $x, y \in X$ (in other words, X can be isometrically embedded into ℓ^∞).

CHAPTER 2

Linear operators and derivatives

1. Bounded linear operators

Let $\mathbb{K}$ denote either one of the fields $\mathbb{R}$ or $\mathbb{C}$. Let X be a vector space over $\mathbb{K}$.

DEFINITION 2.1. A map $\|\cdot\| : X \to [0,\infty)$ is called a *norm* if for all $x, y \in X$ and $\lambda \in \mathbb{K}$,

$$\|\lambda x\| = |\lambda| \cdot \|x\|, \quad \|x+y\| \le \|x\| + \|y\|, \quad \|x\| = 0 \Leftrightarrow x = 0. \tag{2.1}$$

A $\mathbb{K}$-vector space equipped with a norm is called a *normed vector space* . On every normed vector space we have a natural metric space structure defined by

$$d(x,y) = \|x-y\|. \tag{2.2}$$

A normed vector space which is also complete as a metric space is called *Banach space* .

EXAMPLES 2.2.

- $\mathbb{R}^n$ with the Euclidean norm is a Banach space.
- $\mathbb{R}^n$ with the norm $\|x\| = \sup_{i=1,\dots,n} |x_i|$ is also a Banach space.
- If K is a compact metric space, then $C(K)$ is a Banach space with the *supremum norm* $\|f\|_\infty = \sup_{x\in K} |f(x)|$.
- The space of continuous functions on $[0,1]$ equipped with the L^2-norm $\|f\|_2 = (\int_0^1 |f(x)|^2 dx)^{1/2}$ is a normed vector space, but not a Banach space (why?).

EXAMPLE 2.3. The set of bounded sequences $(a_n)_{n\in\mathbb{N}}$ of complex numbers equipped with the ℓ^∞*-norm*,

$$\|a\|_\infty = \sup_{n=1,2,\dots} |a_n| \tag{2.3}$$

is a Banach space. As a metric space, ℓ^∞ conincides with $C_b(\mathbb{N})$.

EXERCISE 2.4.
Define $\ell^1 = \{(a_n)_{n\in\mathbb{N}} \subset \mathbb{C} : \sum_{n=1}^\infty |a_n| < \infty\}$. We equip ℓ^1 with the norm defined by

$$\|a\|_1 = \sum_{n=1}^\infty |a_n|. \tag{2.4}$$

Prove that this defines a Banach space.

EXERCISE 2.5. Define $\ell^2 = \{(a_n)_{n\in\mathbb{N}} \subset \mathbb{C} : \sum_{n=1}^\infty |a_n|^2 < \infty\}$. We equip ℓ^2 with the norm defined by

$$\|a\|_2 = \big(\sum_{n=1}^\infty |a_n|^2\big)^{1/2}. \tag{2.5}$$

Prove that this is really a norm and that ℓ^2 is complete.

Let X, Y be normed vector spaces. Recall that a map $T : X \to Y$ is called *linear* if

$$T(x + \lambda y) = Tx + \lambda Ty \tag{2.6}$$

for every $x, y \in X, \lambda \in \mathbb{K}$. Here we adopt the convention that whenever T is a linear map we write Tx instead of $T(x)$ (unless brackets are necessary because of operator precedence).

DEFINITION 2.6. A linear map $T : X \to Y$ is called *bounded* if there exists $C > 0$ such that $\|Tx\|_Y \leq C\|x\|_X$ for all $x \in X$.

Linear maps between normed vector spaces are also referred to as *linear operators* .

LEMMA 2.7. *Let $T : X \to Y$ be a linear map. The following are equivalent:*

(i) T is bounded
(ii) T is continuous
(iii) T is continuous at 0
(iv) $\sup_{\|x\|_X=1} \|Tx\|_Y < \infty$

PROOF. (i) ⇒ (ii): By assumption and linearity, for $x, y \in X$,

$$\|Tx - Ty\|_Y = \|T(x - y)\|_Y \leq C\|x - y\|_X. \tag{2.7}$$

This implies continuity.
(ii) ⇒ (iii): There is nothing to prove.
(iii) ⇒ (iv): By continuity at 0 there exists $\delta > 0$ such that for $x \in X$ with $\|x\|_X \leq \delta$ we have $\|Tx\|_Y \leq 1$. Let $x \in X$ with $\|x\|_X = 1$. Then $\|\delta x\|_X = \delta$, so

$$\|T(\delta x)\|_Y \leq 1 \tag{2.8}$$

By linearity of T, $\|Tx\|_Y \leq \delta^{-1}$. Thus, $\sup_{\|x\|_X=1} \|Tx\|_Y \leq \delta^{-1} < \infty$.
(iv) ⇒ (i): Let $x \in X$ with $x \neq 0$. Let $C = \sup_{\|x\|_X=1} \|Tx\|_Y < \infty$. Then

$$\Big\| \frac{x}{\|x\|_X} \Big\|_X = 1. \tag{2.9}$$

Thus,

$$\Big\| T\Big(\frac{x}{\|x\|_X} \Big) \Big\|_Y \leq C. \tag{2.10}$$

By linearity of T this implies

$$\|Tx\|_Y \leq C\|x\|_X. \tag{2.11}$$

□

DEFINITION 2.8. By $L(X, Y)$ we denote the space of bounded linear maps $T : X \to Y$. For every $T \in L(X, Y)$ we define its *operator norm* by

$$\|T\|_{\text{op}} = \sup_{x \neq 0} \frac{\|Tx\|_Y}{\|x\|_X}. \tag{2.12}$$

We also denote $\|T\|_{\text{op}}$ by $\|T\|_{X \to Y}$.

One should think of $\|T\|_{\text{op}}$ as the best (i.e. smallest) constant $C > 0$ for which

$$\|Tx\|_Y \leq C\|x\|_X \tag{2.13}$$

holds. We have by definition that

$$\|Tx\|_Y \leq \|T\|_{\text{op}}\|x\|_X. \tag{2.14}$$

Observe that by linearity of T and homogeneity of the norm,

$$\|T\|_{\mathrm{op}} = \sup_{\|x\|_X=1} \|Tx\|_Y = \sup_{\|x\|_X\leq 1} \|Tx\|_Y. \tag{2.15}$$

EXERCISE 2.9. Show that $L(X,Y)$ endowed with the operator norm forms a normed vector space (i.e. show that $\|\cdot\|_{\mathrm{op}}$ is a norm).

EXAMPLE 2.10. Let $A \in \mathbb{R}^{n\times m}$ be a real $n \times m$ matrix. We view A as a linear map $\mathbb{R}^m \to \mathbb{R}^n$: for $x \in \mathbb{R}^m$, $A(x) = A \cdot x \in \mathbb{R}^n$. Let us equip $\mathbb{R}^n$ and $\mathbb{R}^m$ with the corresponding $\|\cdot\|_\infty$ norms. Consider the operator norm $\|A\|_{\infty\to\infty} = \sup_{\|x\|_\infty=1} \|Ax\|_\infty$ with respect to these normed spaces:

$$\|Ax\|_\infty = \max_{i=1,\dots,n} \Big|\sum_{j=1}^m A_{ij}x_j\Big| \leq \big(\max_{i=1,\dots,n} \sum_{j=1}^m |A_{ij}|\big)\|x\|_\infty. \tag{2.16}$$

This implies $\|A\|_{\infty\to\infty} \leq \max_{i=1,\dots,n} \sum_{j=1}^m |A_{ij}|$. On the other hand, for given $i = 1,\dots,n$ we choose $x \in \mathbb{R}^m$ with $x_j = |A_{ij}|/A_{ij}$ if $A_{ij} \neq 0$ and $x_j = 0$ if $A_{ij} = 0$. Then $\|x\|_\infty \leq 1$ and

$$\|A\|_{\infty\to\infty} \geq \|Ax\|_\infty = \sum_{j=1}^m |A_{ij}|. \tag{2.17}$$

Since i was arbitrary, we get $\|A\|_{\infty\to\infty} \geq \max_{i=1,\dots,n} \sum_{j=1}^m |A_{ij}|$. Altogether we proved

$$\|A\|_{\infty\to\infty} = \max_{i=1,\dots,n} \sum_{j=1}^m |A_{ij}|. \tag{2.18}$$

EXERCISE 2.11. Let $A \in \mathbb{R}^{n\times m}$. For $x \in \mathbb{R}^n$ we define $\|x\|_1 = \sum_{i=1}^n |x_i|$.
(i) Determine the value of $\|A\|_{1\to 1} = \sup_{\|x\|_1=1} \|Ax\|_1$ (that is, find a formula for $\|A\|_{1\to 1}$ involving only finitely many computations in terms of the entries of A).
(ii) Do the same for $\|A\|_{1\to\infty} = \sup_{\|x\|_1=1} \|Ax\|_\infty$.

EXERCISE 2.12. Let $A \in \mathbb{R}^{n\times n}$. Define $\|x\|_2 = (\sum_{i=1}^n |x_i|^2)^{1/2}$ (Euclidean norm) and $\|A\|_{2\to 2} = \sup_{\|x\|_2=1} \|Ax\|_2$. Observe that AA^T is a symmetric $n \times n$ matrix and hence has only non-negative eigenvalues. Denote the largest eigenvalue of AA^T by ρ. Prove that $\|A\|_{2\to 2} = \sqrt{\rho}$. *Hint:* First consider the case that A is symmetric. Use that symmetric matrices are orthogonally diagonalizable.

EXERCISE 2.13. Let $A \in \mathbb{R}^{n\times n}$ and define

$$\|A\|_{\mathrm{HS}} = \Big(\sum_{i=1}^n \sum_{j=1}^n |A_{ij}|^2\Big)^{1/2}. \tag{2.19}$$

This is a norm on $\mathbb{R}^{n\times n}$. Prove the following properties for all $A, B \in \mathbb{R}^{n\times n}$:
(i) $\|AB\|_{\mathrm{HS}} \leq \|A\|_{\mathrm{HS}}\|B\|_{\mathrm{HS}}$
(ii) $\|A\|_{\mathrm{HS}} = \sqrt{\mathrm{trace}(AA^T)}$
(iii) $\|UA\|_{\mathrm{HS}} = \|A\|_{\mathrm{HS}}$ for all orthogonal $U \in \mathbb{R}^{n\times n}$
(iv) $\|A\|_{2\to 2} \leq \|A\|_{\mathrm{HS}} \leq \sqrt{n}\|A\|_{2\to 2}$ *Hint:* First do Exercise 2.12.

EXAMPLE 2.14. Let $A \in \mathbb{R}^{n\times n}$ be an invertible $n \times n$ matrix and $b \in \mathbb{R}^n$. Say we want to solve the linear system

$$Ax = b \tag{2.20}$$

for x. Of course, $x = A^{-1}b$. However, A^{-1} is expensive to compute if n is large, so other methods are desirable for solving linear equations. Let

$$F(x) = \lambda(Ax - b) + x \tag{2.21}$$

for some constant $\lambda \neq 0$ that we may choose freely. Then $Ax = b$ if and only if x is a fixed point of F. Moreover,

$$\|F(x)-F(y)\| = \|\lambda A(x-y)+x-y\| = \|(\lambda A+I)(x-y)\| \leq \|\lambda A+I\|_{\mathrm{op}}\|x-y\|. \tag{2.22}$$

Suppose that λ happens to be such that $\|\lambda A + I\|_{\mathrm{op}} < 1$. Then $F : \mathbb{R}^n \to \mathbb{R}^n$ is a contraction, so we can compute the solution to the equation by the iteration $x_{n+1} = F(x_n)$.

2. Equivalence of norms

DEFINITION 2.15. Two norms $\|\cdot\|_a$ and $\|\cdot\|_b$ on a vector space X are called *equivalent* if there exist constants $c, C > 0$ such that

$$c\|x\|_a \leq \|x\|_b \leq C\|x\|_a \tag{2.23}$$

for all $x \in X$.

EXERCISE 2.16. Prove that equivalent norms generate the same topologies: if $\|\cdot\|_a$ and $\|\cdot\|_b$ are equivalent then a set $U \subset X$ is open with respect to $\|\cdot\|_a$ if and only if it is open with respect to $\|\cdot\|_b$.

EXERCISE 2.17. Show that equivalence of norms forms an equivalence relation on the space of norms. That is, if we write $n_1 \sim n_2$ to denote that two norms n_1, n_2 are equivalent, then prove that $n_1 \sim n_1$ (reflexivity), $n_1 \sim n_2 \Rightarrow n_2 \sim n_1$ (symmetry) and $n_1 \sim n_2$, $n_2 \sim n_3 \Rightarrow n_1 \sim n_3$ (transitivity).

THEOREM 2.18. *Let X be a finite-dimensional $\mathbb{K}$-vector space. Then all norms on X are equivalent.*

PROOF. Let $\{b_1, \ldots, b_n\}$ be a basis. Then for every $x \in X$ we can write $x = \sum_{i=1}^n x_i b_i$ with uniquely determined coefficients $x_i \in \mathbb{K}$. Then $\|x\|_* = \max_i |x_i|$ defines a norm on X. Let $\|\cdot\|$ be any norm on X. Since equivalence of norms is an equivalence relation, it suffices to show that $\|\cdot\|_*$ and $\|\cdot\|$ are equivalent. We have

$$\|x\| \leq \sum_{i=1}^n |x_i|\|b_i\| \leq (\max_{j=1,\ldots,n} |x_j|) \sum_{i=1}^n \|b_i\| = C\|x\|_*, \tag{2.24}$$

where $C = \sum_{i=1}^n \|b_i\| \in (0, \infty)$. Now define

$$S = \{x \in X \,:\, \|x\|_* = 1\}. \tag{2.25}$$

We claim that this is a compact set with respect to $\|\cdot\|_*$. Indeed, define the canonical isomorphism $\phi : \mathbb{K}^n \to X$, $(x_1, \ldots, x_n) \mapsto \sum_{i=1}^n x_i b_i$. This is a continuous map (where we equip $\mathbb{K}^n$ with the Euclidean metric, say) and $S = \phi(K)$, where $K = \{x \in \mathbb{K}^n : \max_i |x_i| = 1\}$ is compact by the Heine-Borel Theorem (see Corollary 1.62). Thus S is compact by Theorem 1.54.

Next note that the function $x \mapsto \|x\|$ is continuous with respect to the $\|\cdot\|_*$ norm. This is because by the triangle inequality and (2.24),

$$|\|x\| - \|y\|| \leq \|x - y\| \leq C\|x - y\|_*. \tag{2.26}$$

Thus by Theorem 1.55, $x \mapsto \|x\|$ attains its infimum on the compact set S and therefore there exists $c > 0$ such that

$$\|y\| \geq c \tag{2.27}$$

for all $y \in S$. For $x \in X, x \neq 0$ we have $\frac{x}{\|x\|_*} \in S$ and thus by homogeneity of norms, using (2.27) with $y = \frac{x}{\|x\|_*}$ gives

$$\|x\| \geq c\|x\|_*. \tag{2.28}$$

Thus we proved that $\|\cdot\|$ and $\|\cdot\|_*$ are equivalent norms. □

In contrast, two given norms on an infinite-dimensional vector space are generally not equivalent. For example, the supremum norm and the L^2-norm on $C([0,1])$ are not equivalent (as a consequence of Exercise 4.64).

COROLLARY 2.19. *If X is finite-dimensional then every linear map $T : X \to Y$ is bounded.*

PROOF. Let $\{x_1, \ldots, x_n\} \subset X$ be a basis. Then for $x = \sum_{i=1}^n c_i x_i$ with $c_i \in \mathbb{K}$,

$$\|Tx\|_Y \leq \sum_{i=1}^n |c_i| \|Tx_i\|_Y \leq C \max_{i=1,\ldots,n} |c_i|, \tag{2.29}$$

where $C = \sum_{i=1}^n \|Tx_i\|_Y$. By equivalence of norms we may assume that $\max_i |c_i|$ is the norm on X. □

This is not true if X is infinite-dimensional.

EXAMPLE 2.20. Let X be the set of sequences of complex numbers $(a_n)_{n\in\mathbb{N}}$ such that $\sup_{n\in\mathbb{N}} n|a_n| < \infty$ and let Y be the space of bounded complex sequences. Then $X \subset Y$. Equip both spaces with the norm $\|a\| = \sup_{n\in\mathbb{N}} |a_n|$. The map $T : X \to Y$, $(Ta)_{n\in\mathbb{N}} = na_n$ is not bounded: let $e_n^{(k)} = 1$ if $k = n$ and $e_n^{(k)} = 0$ if $k \neq n$. Then $e^{(k)} \in X$ and $Te^{(k)} = ke^{(k)}$ and $\|e^{(k)}\| = 1$. So

$$\|Te^{(k)}\| = k \tag{2.30}$$

for every $k \in \mathbb{N}$ and therefore $\sup_{\|x\|=1} \|Tx\| = \infty$.

EXERCISE 2.21. Let X be the set of continuously differentiable functions on $[0,1]$ and let $Y = C([0,1])$. We consider X and Y as normed vector spaces with the norm $\|f\| = \sup_{x\in[0,1]} |f(x)|$. Define a linear map $T : X \to Y$ by $Tf = f'$. Show that T is not bounded.

3. Dual spaces*

THEOREM 2.22. *Let X be a normed vector space and Y a Banach space. Then $L(X,Y)$ is a Banach space (with the operator norm).*

PROOF. Let $(T_n)_{n\in\mathbb{N}} \subset L(X,Y)$ be a Cauchy sequence. Then for every $x \in X$, $(T_n x)_{n\in\mathbb{N}} \subset Y$ is Cauchy and by completeness of Y it therefore converges to some limit which we call Tx. This defines a linear operator $T : X \to Y$. We claim that T is bounded. Since $(T_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, it is a bounded sequence. Thus there exists $M > 0$ such that $\|T_n\|_{\text{op}} \leq M$ for all $n \in \mathbb{N}$. We have for $x \in X$,

$$\|Tx\|_Y \leq \|Tx - T_n x\|_Y + \|T_n x\|_Y \leq \|Tx - T_n x\|_Y + M\|x\|_X. \tag{2.31}$$

Letting $n \to \infty$ we get $\|Tx\|_Y \leq M\|x\|_X$. So T is bounded with $\|T\|_{\text{op}} \leq M$. It remains to show that $T_n \to T$ in $L(X,Y)$. That is, for all $\varepsilon > 0$ we need to find $N \in \mathbb{N}$ such that

$$\|T_n x - Tx\|_Y \leq \varepsilon\|x\|_X \tag{2.32}$$

for all $n \geq N$ and $x \in X$. Since $(T_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, there exists $N \in \mathbb{N}$ such that

$$\|T_n x - T_m x\|_Y \leq \tfrac{\varepsilon}{2}\|x\|_X \tag{2.33}$$

for all $n, m \geq N$ and $x \in X$. Fix $x \in X$. Then there exists $m_x \geq N$ such that

$$\|T_{m_x} x - Tx\|_Y \leq \tfrac{\varepsilon}{2}\|x\|_X. \tag{2.34}$$

Then if $n \geq N$ and $x \in X$,

$$\|T_n x - Tx\|_Y \leq \|T_n x - T_{m_x} x\|_Y + \|T_{m_x} x - Tx\|_Y \leq \varepsilon\|x\|_X. \tag{2.35}$$

□

DEFINITION 2.23. Let X be a normed vector space. Elements of $L(X,\mathbb{K})$ are called *bounded linear functionals* . $L(X,\mathbb{K})$ is called the *dual space* of X and denoted X'.

COROLLARY 2.24. *Dual spaces of normed vector spaces are Banach spaces.*

PROOF. This follows from Theorem 2.22 because $\mathbb{K}$ (which is $\mathbb{R}$ or $\mathbb{C}$) is complete. □

THEOREM 2.25. *If X is finite-dimensional, then X' is isomorphic to X.*

PROOF. Let $\{x_1, \ldots, x_n\} \subset X$ be a basis. Then we can define a corresponding *dual basis* of X' as follows: let $f_i \in X'$, $i \in \{1, \ldots, n\}$ be the linear map given by $f_i(x_i) = 1$ and $f_i(x_j) = 0$ for $j \neq i$. Then we claim that $\{f_1, \ldots, f_n\}$ is a basis of X'. Indeed, let $f \in X'$. For $x \in X$ we can write $x = \sum_{i=1}^n c_i x_i$ with uniquely determined $c_i \in \mathbb{K}$. Then by linearity,

$$f(x) = \sum_{i=1}^n c_i f(x_i) = \sum_{i=1}^n f(x_i) f_i(x), \tag{2.36}$$

because $f_i(x) = c_i$. Thus, the linear span of $\{f_1, \ldots, f_n\}$ is X'. On the other hand, suppose

$$\sum_{i=1}^n b_i f_i = 0 \tag{2.37}$$

for some coefficients $(b_i)_{i=1,\ldots,n} \subset \mathbb{K}$. Then for every $j \in \{1, \ldots, n\}$, $b_j = \sum_{i=1}^n b_i f_i(x_j) = 0$. Thus, $\{f_1, \ldots, f_n\}$ is linearly independent. Thus, X' and X are isomorphic since

they have the same dimension. We can define an isomorphism $\phi : X \to X'$ by $x_i \mapsto f_i$ for $i = 1, \dots, n$. □

4. Sequential ℓ^p spaces*

DEFINITION 2.26. Let $1 \le p < \infty$. Then we define ℓ^p as the set of all sequences $(x_n)_{n=1,2,\dots} \subset \mathbb{C}$ such that $\sum_{n=1}^{\infty} |x_n|^p < \infty$. The *$\ell^p$-norm* is defined as

$$\|x\|_p = \Big(\sum_{n=1}^{\infty} |x_n|^p \Big)^{1/p}. \tag{2.38}$$

If $p \in [1, \infty]$ then the number $p' \in [1, \infty]$ such that $\frac{1}{p} + \frac{1}{p'} = 1$ is called the *Hölder dual exponent* of p.

REMARK. The definition of ℓ^p extends to values of $p < 1$, but $\|x\|_p$ does not define a norm if $p < 1$.

Our first goal is to show that $\|\cdot\|_p$ really is a norm. To do that we need the following generalization of the Cauchy-Schwarz inequality.

THEOREM 2.27 (Hölder's inequality). *Let $p \in [1, \infty]$ and $x \in \ell^p, y \in \ell^{p'}$. Then*

$$\Big| \sum_{n=1}^{\infty} x_n y_n \Big| \le \|x\|_p \|y\|_{p'} \tag{2.39}$$

We need an auxiliary Lemma which generalizes the usual inequality for two non-negative numbers a, b

$$\sqrt{ab} \le \frac{a+b}{2} \tag{2.40}$$

comparing the geometrical mean of a, b (i.e. the sidelength of the square whose area equals the area of the rectangle with sides a and b) with the arithmetical mean (the number half way between a and b).

LEMMA 2.28. *Let $a, b \ge 0$.*
(i) Let $0 < \vartheta < 1$. Then

$$a^{1-\vartheta} b^{\vartheta} \le (1-\vartheta)a + \vartheta b. \tag{2.41}$$

(ii) (Young's inequality) Let $p \in (1, \infty)$. Then

$$ab \le \frac{a^p}{p} + \frac{b^{p'}}{p'}. \tag{2.42}$$

PROOF. Clearly the inequality holds if one of a, b is 0. Also Check that if the inequality is true for some a, b then it is also true for ta, tb where $t > 0$.

Assume now $0 < b \le a$ and let $s = b/a$. Then the stated inequality is equivalent with $s^{\vartheta} \le (1-\vartheta) + \vartheta s$ for $0 \le s \le 1$. Set $f(s) = 1 - \vartheta + \vartheta s - s^{\vartheta}$. Then $f(1) = 0$ and $f'(s) < 0$ for $0 < s < 1$, thus $f(s) \ge 0$ for $0 \le s \le 1$ which implies the desired iequality. The case $0 < a \le b$ is shown in the same way (in fact follows from the previous case by interchanging a, b and replacing ϑ by $1 - \vartheta$). This proves part (i).

For part (ii) set $x = a^p$, $y = b^{p'}$, $\vartheta = 1 - 1/p$ and observe that the inequality is then equivalent with $x^{1-\vartheta} y^{\vartheta} \le (1-\vartheta)x + \vartheta y$ which holds by part (i). □

PROOF OF HÖLDER'S INEQUALITY. Observe that the inequality is true if either x or y are 0. Check that if the inequality is true for some choice of x and y then it is also true for sx, ty with $s > 0$, $t > 0$. Finally If $p \in \{1, \infty\}$, the inequality is trivial. So we assume $p \in (1, \infty)$.

By Young's inequality,

$$\sum_{n=1}^{\infty} |x_n y_n| \leq \frac{1}{p} \sum_{n=1}^{\infty} |x_n|^p + \frac{1}{p'} \sum_{n=1}^{\infty} |y_n|^{p'}. \tag{2.43}$$

Observe that this yields the asserted inequality when $\|x\|_p = 1$ and $\|y\|_{p'} = 1$.

Also we have $\|\frac{x}{\|x\|_p}\|_p = 1$, $\|\frac{y}{\|y\|_{p'}} = 1$, and since the assertion holds for $x/\|x\|_p$ and $y/\|y|_{p'}$ it holds also for x and y. □

THEOREM 2.29 (Minkowski's inequality). *Let $p \in [1, \infty]$. For $x, y \in \ell^p$,*

$$\|x + y\|_p \leq \|x\|_p + \|y\|_p. \tag{2.44}$$

PROOF. If $p \in \{1, \infty\}$ the inequality is trivial. Thus we assume $p \in (1, \infty)$. If $\|x + y\|_p = 0$, the inequality is also trivial, so we can assume $\|x + y\|_p > 0$. Now we write

$$\|x + y\|_p^p \leq \sum_{n=1}^{\infty} |x_n||x_n + y_n|^{p-1} + \sum_{n=1}^{\infty} |y_n||x_n + y_n|^{p-1} \tag{2.45}$$

Using Hölder's inequality on both sums we obtain that this is

$$\leq \|x\|_p \|x + y\|_{p'(p-1)}^{p-1} + \|y\|_p \|x + y\|_{p'(p-1)}^{p-1} \tag{2.46}$$

We have $p'(p-1) = \frac{p}{p-1}(p-1) = p$, so we have proved that

$$\|x + y\|_p^p = (\|x\|_p + \|y\|_p)\|x + y\|_p^{p-1}. \tag{2.47}$$

Dividing by $\|x + y\|_p^{p-1}$ gives the claim. □

We conclude that $\|\cdot\|_p$ is a norm and ℓ^p a normed vector space.

THEOREM 2.30. *Let $p \in (1, \infty)$. The dual space $(\ell^p)'$ is isometrically isomorphic to $\ell^{p'}$.*

PROOF. By e_k we denote the sequence which is 1 at position k and 0 everywhere else.

Then we define a map $\phi : (\ell^p)' \to \ell^{p'}$ by $\phi(v) = (v(e_k))_k$. Clearly, this is a linear map. First we need to show that $\phi(v) \in \ell^{p'}$. Let $v \in (\ell^p)'$. For each n we define $x^{(n)} \in \ell^p$ by

$$x_k^{(n)} = \begin{cases} \frac{|v(e_k)|^{p'}}{v(e_k)} & \text{if } k \leq n,\ v(e_k) \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.48}$$

We have on the one hand

$$v(x^{(n)}) = \sum_{n=1}^{n} |v(e_k)|^{p'}. \tag{2.49}$$

And on the other hand

$$|v(x^{(n)})| \leq \|v\|_{\mathrm{op}} \|x^{(n)}\|_p = \|v\|_{\mathrm{op}} \Big(\sum_{k=1}^{n} |v(e_k)|^{p'} \Big)^{1/p}. \tag{2.50}$$

Here we have used that $p(p'-1) = p(\frac{p}{p-1} - 1) = \frac{p}{p-1} = p'$. Combining these two we get

$$\Big(\sum_{k=1}^{n} |v(e_k)|^{p'}\Big)^{\frac{1}{p'}} \leq \|v\|_{\mathrm{op}}. \tag{2.51}$$

Letting $n \to \infty$ this implies that

$$\|\phi(v)\|_{p'} = \Big(\sum_{n=1}^{\infty} |v(e_n)|^{p'}\Big)^{1/p'} \leq \|v\|_{\mathrm{op}}, \tag{2.52}$$

so $\phi(v) \in \ell^{p'}$. The calculation also shows that ϕ is bounded. It is easy to check that ϕ is injective. We show that it is surjective: let $x \in \ell^{p'}$. Then define $v \in (\ell^p)'$ by $v(y) = \sum_{n=1}^{\infty} x_n y_n$. By Hölder's inequality, v is well-defined. We have $v(e_k) = x_k$, so $\phi(v) = x$. Thus ϕ is an isomorphism. It remains to show that ϕ is an isometry. We have already seen that

$$\|\phi(v)\|_{p'} \leq \|v\|_{\mathrm{op}} \tag{2.53}$$

We leave it to the reader to verify the other inequality. □

Remark. It can be shown similarly that $(\ell^1)' = \ell^\infty$. However, the dual of ℓ^∞ is not ℓ^1.

COROLLARY 2.31. *ℓ^p is a Banach space for all $p \in (1, \infty)$.*

Remark. ℓ^1 and ℓ^∞ are also Banach spaces as we saw in Example 2.3 and Exercise 2.4.

EXERCISE 2.32. (i) Let $0 < p < \infty$. For $x \in \mathbb{R}^n$ set $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$. Prove that $\|x\|_{p_2} \leq \|x\|_{p_1}$ if $p_1 \leq p_2$.

Hint: First do this with the additional condition that $\|x\|_{p_1} = 1$.

(ii) Show that $\ell^p \subsetneq \ell^q$ if $1 \leq p < q \leq \infty$.

(iii) Show that this inclusion extends to values of $p_1, p_2 \in (0, \infty)$.

5. Derivatives

Recall that a function f on an interval (a, b) is called differentiable at $x \in (a, b)$ if $\lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$ exists. In other words, if there exists a number $T \in \mathbb{R}$ such that

$$\lim_{h\to 0} \frac{|f(x+h) - f(x) - Th|}{|h|} = 0. \tag{2.54}$$

In that case we denote that real number T by $f'(x)$. A real number can be understood as a linear map $\mathbb{R} \to \mathbb{R}$:

$$\mathbb{R} \longrightarrow L(\mathbb{R}, \mathbb{R}), T \longmapsto (x \mapsto T \cdot x) \tag{2.55}$$

That is, the linear map associated with a real number T is given by multiplication with T. Interpreting the derivative at a given point as a linear map, we can formulate the definition in the general setting of normed vector spaces.

DEFINITION 2.33. Let X, Y be normed vector spaces and $U \subset X$ open. A map $F : U \to Y$ is called *Fréchet differentiable* (we also say *differentiable*) at $x \in U$ if there exists $T \in L(X, Y)$ such that

$$\lim_{h\to 0} \frac{\|F(x+h) - F(x) - Th\|_Y}{\|h\|_X} = 0. \tag{2.56}$$

In that case we call T the *(Fréchet) derivative* of F at x and write $T = DF(x)$ or $T = DF|_x$. F is called *(Fréchet) differentiable* if it is differentiable at every point

$x \in U$. When $X = \mathbb{R}^n$ we also use the following terminology: F is *totally differentiable* and $DF(x)$ is the *total derivative* of F at x.

Before we move on we need to verify that $DF(x)$ is well-defined. That is, that T is uniquely determined by F and x. Suppose $T, \widetilde{T} \in L(X,Y)$ both satisfy (2.56). Then

$$\|Th - \widetilde{T}h\|_Y \leq \|F(x+h) - F(x) - Th\|_Y + \|F(x+h) - F(x) - \widetilde{T}h\|_Y \tag{2.57}$$

Thus, by (2.56),

$$\frac{\|Th - \widetilde{T}h\|_Y}{\|h\|_X} \longrightarrow 0 \quad \text{as } h \to 0 \tag{2.58}$$

In other words, for all $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\|Th - \widetilde{T}h\|_Y \leq \varepsilon \|h\|_X \tag{2.59}$$

if $\|h\|_X \leq \delta$. By homogeneity of norms we argue that the inequality (2.59) must hold for all $h \in X$: let $h \in X$, $h \neq 0$ be arbitrary. Then let $h_0 = \delta \frac{h}{\|h\|_X}$. By homogeneity of norms we have $\|h_0\|_X = \delta$. Thus,

$$\|Th_0 - \widetilde{T}h_0\|_Y \leq \varepsilon \|h_0\|_X = \varepsilon\delta. \tag{2.60}$$

Multiplying both sides by $\delta^{-1}\|h\|_X$ and using homogeneity of norms and linearity of T, we obtain

$$\|Th - \widetilde{T}h\|_Y \leq \varepsilon \|h\|_X \tag{2.61}$$

for all $h \in X$ (it is trivial for $h = 0$). Since $\varepsilon > 0$ was arbitrary (and is independent of h), this implies $\|Th - \widetilde{T}h\|_Y = 0$, so $Th = \widetilde{T}h$ for all h. Thus $T = \widetilde{T}$.

Reminder: Big-O and little-o notation. Let f, g be maps between normed vector spaces X, Y, Z: $f : U \to Y, g : U \to Z$, $U \subset X$ open neighborhood of 0.

- Big-O: We write

$$f(h) = O(g(h)) \quad \text{as } h \to 0 \tag{2.62}$$

to mean

$$\limsup_{h \to 0} \frac{\|f(h)\|}{\|g(h)\|} < \infty. \tag{2.63}$$

This is equivalent to saying that there exists a $C > 0$ and $\delta > 0$ such that

$$\|f(h)\| \leq C\|g(h)\| \tag{2.64}$$

for all h with $0 < \|h\| < \delta$.

- Little-o: Write

$$f(h) = o(g(h)) \quad \text{as } h \to 0 \tag{2.65}$$

to mean

$$\lim_{h \to 0} \frac{\|f(h)\|}{\|g(h)\|} = 0. \tag{2.66}$$

Comments.

- O and o are *not* functions and (2.62), (2.65) are *not* equations!
- This is an abuse of the inequality sign: it would be more accurate to define $O(g)$ as the class of functions that satisfy (2.64), say to write $f \in O(g)$.

- One can think of (say) $O(g)$ as a placeholder for a function which may change at every occurrence of the symbol $O(g)$ but always satisfies the respective condition that it is dominated by a constant times $\|g(h)\|$ if $\|h\|$ is small.
- For brevity, we may sometimes not write out the phrase "as $h \to 0$".
- There is nothing special about letting h tend to 0 in this definition. We can also define $o(g)$, $O(g)$ with respect to another limit, for instance, say, as $\|h\| \to \infty$.
- If $f(h) = o(g(h))$, then $f(h) = O(g(h))$, but generally not vice versa.
- If $f(h) = O(\|h\|^k)$, then $f(h) = o(\|h\|^{k-\varepsilon})$ for every $\varepsilon > 0$.

- $f(h) = o(1)$ is equivalent to saying that $f(h) \to 0$ as $h \to 0$.

We can use little-o notation to restate the definition of derivatives in an equivalent way: F is Fréchet differentiable at x if and only if there exists $T \in L(X, Y)$ such that

$$F(x+h) = F(x) + Th + o(\|h\|) \quad (\text{as } h \to 0). \tag{2.67}$$

The derivative map $T = DF|_x$ provides a linear approximation to $F(x+h)$ when $\|h\|$ is small. Thus, in the same way as in the one-dimensional setting, the derivative is a way to describe how the values of F change around a fixed point x.

EXAMPLE 2.34. Let $F : \mathbb{R}^2 \to \mathbb{R}$ be given by $F(x_1, x_2) = x_1 \cos(x_2)$. We claim that F is totally differentiable at every $x = (x_1, x_2) \in \mathbb{R}^2$. Indeed, let $x \in \mathbb{R}^2$ and $h = (h_1, h_2) \in \mathbb{R}^2 \backslash \{0\}$. Then

$$F(x+h) = (x_1 + h_1)\cos(x_2 + h_2) = x_1 \cos(x_2 + h_2) + h_1 \cos(x_2 + h_2) \tag{2.68}$$

From Taylor's theorem we have that

$$\cos(t + \varepsilon) = \cos(t) - \sin(t)\varepsilon + O(\varepsilon^2) \quad \text{as } \varepsilon \to 0 \tag{2.69}$$

Thus,

$$F(x+h) = x_1 \cos(x_2) - x_1 \sin(x_2)h_2 + O(\|h\|^2) + h_1 \cos(x_2) - h_1 \sin(x_2)h_2 + O(\|h\|^2) \tag{2.70}$$

$$F(x+h) - F(x) = h_1 \cos(x_2) - x_1 \sin(x_2)h_2 + O(\|h\|^2) \tag{2.71}$$

This implies

$$F(x+h) = F(x) + Th + o(\|h\|), \tag{2.72}$$

where we have set $Th = h_1 \cos(x_2) - x_1 \sin(x_2)h_2$ (this is a linear map $\mathbb{R}^2 \to \mathbb{R}$). So we have proven that F is differentiable at x and

$$DF|_x h = h_1 \cos(x_2) - x_1 \sin(x_2)h_2. \tag{2.73}$$

EXAMPLE 2.35. Let $F : C([0,1]) \to C([0,1])$ be given by $F(f)(x) = \int_0^x f(t)^2 dt$. Then F is Fréchet differentiable at every $f \in C([0,1])$. Indeed, we compute

$$F(f+h)(x) - F(f)(x) = \int_0^x (f(t)+h(t))^2 dt - \int_0^x f(t)^2 dt = 2\int_0^x f(t)h(t)dt + \int_0^x h(t)^2 dt \tag{2.74}$$

Set $T(h)(x) = 2\int_0^x f(t)h(t)dt$. This is a bounded linear map:

$$\|T(h)\|_\infty \leq 2\int_0^1 |f(t)h(t)|dt \leq C\|h\|_\infty, \tag{2.75}$$

where $C = 2\int_0^1 |f(t)|dt$. We have

$$F(f+h)(x) - F(f)(x) - T(h)(x) = \int_0^x h(t)^2 dt. \tag{2.76}$$

Thus

$$\|F(f+h) - F(f) - Th\|_\infty \leq \sup_{x\in[0,1]} \Big| \int_0^x h(t)^2 dt \Big| \tag{2.77}$$

$$\leq \int_0^1 |h(t)|^2 dt \leq \sup_{x\in[0,1]} |h(x)|^2 = \|h\|_\infty^2 \tag{2.78}$$

This implies

$$\frac{1}{\|h\|_\infty} \|F(f+h) - F(f) - Th\|_\infty \leq \|h\|_\infty \to 0 \tag{2.79}$$

as $h \to 0$. Thus F is Fréchet differentiable at f and $DF|_f(h) = 2\int_0^x f(t)h(t)dt$.

We go on to discuss some of the familiar properties of derivatives. It follows directly from the definition that $DF|_x$ is linear in F. That is, if $F : U \to Y, G : U \to Y$ are differentiable at $x \in U$ and $\lambda \in \mathbb{R}$, then the function $F + \lambda G : U \to Y$ defined by $(F + \lambda G)(x) = F(x) + \lambda G(x)$ is differentiable at x and $D(F + \lambda G)|_x = DF|_x + \lambda DG|_x$.

THEOREM 2.36 (Chain rule). *Let X_1, X_2, X_3 be normed vector spaces and $U_1 \subset X_1, U_2 \subset X_2$ open. Let $x \in U_1$ and $g : U_1 \to X_2$, $f : U_2 \to X_3$ such that g is Fréchet differentiable at x, $g(U_1) \subset U_2$ and f is Fréchet differentiable at $g(x)$. Then the function $f \circ g : U_1 \to X_3$ defined by $(f \circ g)(x) = f(g(x))$ is Fréchet differentiable at x and*

$$D(f \circ g)|_x h = Df|_{g(x)} Dg|_x h \tag{2.80}$$

for all $h \in X_1$.

PROOF. Let $x, x + h \in U_1$. We write

$$f(g(x+h)) - f(g(x)) - Df|_{g(x)} Dg|_x h \tag{2.81}$$

$$= f(g(x)+k) - f(g(x)) - Df|_{g(x)} k + Df|_{g(x)}(g(x+h) - g(x) - Dg|_x h), \tag{2.82}$$

where $k = g(x+h) - g(x)$. Using the triangle inequality and that $Df|_{g(x)}$ is a bounded linear map we obtain

$$\|f(g(x+h)) - f(g(x)) - Df|_{g(x)} Dg|_x h\|_{X_3} \tag{2.83}$$

$$\leq \|f(g(x)+k) - f(g(x)) - Df|_{g(x)} k\|_{X_3} + \|Df|_{g(x)}\|_{\mathrm{op}} \|g(x+h) - g(x) - Dg|_x h\|_{X_2} \tag{2.84}$$

We have

$$\|k\|_{X_2} = \|g(x+h) - g(x)\|_{X_2} \leq \|Dg|_x\|_{\mathrm{op}} \|h\|_{X_1} + o(\|h\|_{X_1}). \tag{2.85}$$

Dividing by $\|h\|_{X_1}$ on both sides, (2.84) implies

$$\frac{1}{\|h\|_{X_1}} \|f(g(x+h)) - f(g(x)) - Df|_{g(x)} Dg|_x h\|_{X_3}$$
$$\leq \frac{\|k\|_{X_2}}{\|h\|_{X_1}} \frac{\|f(g(x)+k) - f(g(x)) - Df|_{g(x)} k\|_{X_3}}{\|k\|_{X_2}} + o(1), \text{ as } h \to 0.$$

By (2.85),

$$\frac{\|k\|_{X_2}}{\|h\|_{X_1}} \leq \|Dg|_x\|_{\mathrm{op}} + 1 \tag{2.86}$$

if $\|h\|_{X_1}$ is small enough. In particular, $k \to 0$ as $h \to 0$. Since f is differentiable at $g(x)$ we have that

$$\frac{\|f(g(x)+k) - f(g(x)) - Df|_{g(x)}k\|_{X_3}}{\|k\|_{X_2}} \tag{2.87}$$

converges to 0 as $h \to 0$ (since then $k \to 0$). □

THEOREM 2.37 (Product rule). *Let X be a normed vector space, $U \subset X$ open and assume that $F, G : U \to \mathbb{R}$ are differentiable at $x \in U$. Then the function $F \cdot G : U \to \mathbb{R}$, $(F \cdot G)(x) = F(x)G(x)$ is also differentiable at x and*

$$D(F \cdot G)\,|_x = F(x) \cdot DG\,|_x + G(x) \cdot DF\,|_x\,. \tag{2.88}$$

EXERCISE 2.38. Prove this.

DEFINITION 2.39. Let X, Y be normed vector spaces, $U \subset X$ open, $F : U \to Y$. Let $v \in X$ with $v \neq 0$. If the limit

$$\lim_{h \to 0} \frac{F(x+hv) - F(x)}{h} \in Y \qquad (h \in \mathbb{K} \setminus \{0\}) \tag{2.89}$$

exists, then it is called the *directional derivative* (or *Gâteaux derivative*) of F at x in direction v and denoted $D_v F|_x$.

THEOREM 2.40. *Let X, Y be normed vector spaces, $U \subset X$ open and $F : U \to Y$ Fréchet differentiable at $x \in U$. Then for every $v \in X$, $v \neq 0$, the directional derivative $D_v F|_x$ exists and*

$$D_v F|_x = DF|_x v. \tag{2.90}$$

PROOF. By definition

$$F(x+hv) - F(x) - DF|_x(hv) = o(h) \quad \text{as } h \to 0. \tag{2.91}$$

Therefore,

$$\frac{F(x+hv) - F(x)}{h} = DF|_x v + o(1) \quad \text{as } h \to 0. \tag{2.92}$$

In other words, the limit as $h \to 0$ exists and equals $DF|_x v$. □

EXAMPLE 2.41. Consider $F : \mathbb{R}^2 \to \mathbb{R}$, $F(x) = x_1^2 + x_2^2$ (where $x = (x_1, x_2) \in \mathbb{R}^2$). Let $e_1 = (1,0)$, $e_2 = (0,1)$. Then the directional derivatives $D_{e_1} F|_x$ and $D_{e_2} F|_x$ exist at every point $x \in \mathbb{R}^2$ and

$$D_{e_1} F|_x = 2x_1, \quad D_{e_2} F|_x = 2x_2. \tag{2.93}$$

Also, $DF|_x$ exists at every x and we can compute it using $D_{e_1} F|_x$ and $D_{e_2} F|_x$: let $v \in \mathbb{R}^2$ and write $v = v_1 e_1 + v_2 e_2$ where $v_1, v_2 \in \mathbb{R}$. Then

$$DF|_x v = v_1 DF|_x e_1 + v_2 DF|_x e_2 \tag{2.94}$$

By Theorem 2.40 this equals

$$v_1 D_{e_1} F|_x + v_2 D_{e_2} F|_x = 2x_1 v_1 + 2x_2 v_2. \tag{2.95}$$

Remark. The converse of Theorem 2.40 is not true!

EXAMPLE 2.42. Let $F : \mathbb{R}^2 \to \mathbb{R}$ be defined by $F(x) = \frac{x_1^3}{x_1^2+x_2^2}$ if $x \neq 0$ and $F(0) = 0$. Then all directional derivatives $D_v F|_0$ for $v \neq 0$ exist: for $v = (v_1, v_2)$,

$$F(hv) - F(0) = h\frac{v_1^3}{v_1^2 + v_2^2}, \tag{2.96}$$

so $D_v F|_0 = \frac{v_1^3}{v_1^2+v_2^2}$. But F is not totally differentiable at 0, otherwise we would have by linearity of the total derivative,

$$D_v F|_0 = DF|_0 v = v_1 D_{e_1} F|_0 + v_2 D_{e_2} F|_0 = v_1, \tag{2.97}$$

which is false.

6. Further exercises

EXERCISE 2.43. For $x, y \in \mathbb{R}^n$ define

$$\rho_p(x, y) = \sum_{i=1}^{n} |x_i - y_i|^p.$$

(i) Let $0 < p \leq 1$. Prove that ρ_p is a metric on $\mathbb{R}^n$.
(ii) Let $1 < p < \infty$. Prove that ρ_p is not a metric on $\mathbb{R}^n$.
(iii) Let $0 < p < 1$, $n \geq 2$. Prove that $\rho_p^{1/p}$ is not a metric on $\mathbb{R}^n$.
(iv) Let $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$. Prove that if $0 < p < 1$, $n \geq 2$, neither $\|x\|_p$ nor $\|x\|_p^p$ define a norm on $\mathbb{R}^n$.

EXERCISE 2.44. Let $x \in \mathbb{R}^n$. Define $\|x\|_p = \Big(\sum_{i=1}^n |x_i|^p\Big)^{1/p}$ for $0 < p < \infty$ and $\|x\|_\infty = \max_{i=1,\dots,n} |x_i|$.
(i) Show that $\lim_{p\to\infty} \|x\|_p = \|x\|_\infty$.
(ii) Show that $\lim_{p\to 0} \|x\|_p$ exists and determine its value (we also allow ∞ as a limit).

EXERCISE 2.45. Let $C([0, 1])$ be the space of continuous real-valued functions on the interval $[0, 1]$. Assume $1 \leq p < \infty$.
(i) Show that for $1 \leq p < \infty$ the expression

$$\|f\|_p = \Big(\int_0^1 |f(t)|^p dt\Big)^{1/p}$$

defines a norm on $C([0, 1])$. You may choose to do part (iv) below first.
(ii) Let $\alpha < 1/p$ and define

$$f_n(t) = \begin{cases} t^{-\alpha} & \text{for } n^{-1} \leq t \leq 1, \\ n^\alpha & \text{for } 0 \leq t < n^{-1}. \end{cases}$$

Show that $\{f_n\}_{n=0}^\infty$ is a Cauchy sequence in $C([0, 1])$, with respect to the norm $\|\cdot\|_p$. Show that it is not a Cauchy sequence with respect to the usual sup-norm on $C([0, 1])$.
(iii) Show that $C([0, 1])$ with norm $\|\cdot\|_p$ is not complete.
(iv) Let $\mathcal{R}([0, 1])$ be the space of Riemann integrable functions defined on the interval $[0, 1]$. Show that if f is Riemann integrable then $|f|^p$ is also Riemann integrable. Show that $\|f\|_p$ defines a seminorm on $\mathcal{R}([0, 1])$ but not a norm.
(v) Show for $f, g \in \mathcal{R}([0, 1])$, $1 < p < \infty$, $\frac{1}{p} + \frac{1}{p'} = 1$,

$$\Big|\int_0^1 f(t)g(t)dt\Big| \leq \|f\|_p \|g\|_{p'}.$$

(vi) Show for $f \in \mathcal{R}([0,1])$,

$$\|f\|_{p_1} \leq \|f\|_{p_2} \text{ if } p_1 \leq p_2\,.$$

EXERCISE 2.46. Let $C(\mathbb{R})$ be the set of continuous functions on $\mathbb{R}$. Let $w(t) = \frac{t}{1+t}$ for $t \geq 0$. Define

$$d(f,g) = \sum_{k=0}^{\infty} 2^{-k} w(\sup_{x\in[-k,k]} |f(x) - g(x)|). \tag{2.98}$$

(i) Show that d is a well-defined metric.
(ii) Show that $C(\mathbb{R})$ is complete with this metric.
(iii) Show that there exists no norm $\|\cdot\|$ on $C(\mathbb{R})$ such that $d(f,g) = \|f - g\|$.

EXERCISE 2.47. Consider the space ℓ^1 of absolutely summable sequences of complex numbers. Let $p, q \in [1,\infty]$ with $p \neq q$. Then $\|\cdot\|_p$ and $\|\cdot\|_q$ are norms on ℓ^1 (recall that $\|a\|_p = (\sum_{n=1}^{\infty} |a_n|^p)^{1/p}$ for $p \in [1,\infty)$ and $\|a\|_\infty = \sup_{n\in\mathbb{N}} |a_n|$). Show that $\|\cdot\|_p$ and $\|\cdot\|_q$ are not equivalent.

EXERCISE 2.48. Let X be the space of continuous functions on $[0,1]$ equipped with the norm $\|f\| = \int_0^1 |f(t)|dt$. Define a linear map $T : X \to X$ by

$$Tf(x) = \int_0^x f(t)dt. \tag{2.99}$$

Show that T is well-defined and bounded and determine the value of $\|T\|_{\text{op}}$.

EXERCISE 2.49. Let X, Y be normed vector spaces and $F : X \to Y$ a map.
(i) Show that F is continuous if it is Fréchet differentiable.
(ii) Prove that F is Fréchet differentiable if it is linear and bounded.

EXERCISE 2.50. Let V, W be normed vector spaces and let $T : V \to W$ be a bounded linear transformation. Show that T is differentiable everywhere and compute the derivative DT_v for all $v \in V$.

EXERCISE 2.51. Let $X = C([0,1])$ be the Banach space of continuous functions on $[0,1]$ (with the supremum norm) and define a map $F : X \to X$ by

$$F(f)(s) = \int_0^s \cos(f(t)^2)dt, \quad s \in [0,1]. \tag{2.100}$$

(i) Show that F is Fréchet differentiable and compute the Fréchet derivative $DF\,|_f$ for each $f \in X$.
(ii) Show that $FX = \{F(f) : f \in X\} \subset X$ is relatively compact.

EXERCISE 2.52. Let $\mathbb{R}^{n\times n}$ denote the space of real $n \times n$ matrices equipped with the matrix norm $\|A\| = \sup_{\|x\|=1} \|Ax\|$. Define

$$F : \mathbb{R}^{n\times n} \longrightarrow \mathbb{R}^{n\times n}, \qquad F(A) = A^2. \tag{2.101}$$

Show that F is totally differentiable and compute $DF\,|_A$.

EXERCISE 2.53. (i) Is there a constant C such that for all continuous functions f on $[0,2]$ the inequality

$$\int_0^2 |f(t)|dt \leq C \max_{0\leq x\leq 2} |f(x)|$$

holds? Is there a constant C such that for all continuous functions f on $[0,2]$ the reverse inequality

$$\max_{0\le x\le 1}|f(x)|\le C\int_0^2|f(t)|dt$$

holds? The expressions on the both sides of the above inequalities define norms on $C([0,1])$. Are these equivalent norms?

(ii) True or false: There is a constant C_n such that for all polynomials P of degree $\le n$ we have

$$\max_{0\le x\le 2}|P(x)|\le C_n\int_0^2|P(t)|dt\,.$$

What about the analogous question concerning the inequality

$$\max_{0\le x\le 200}|P(x)|\le C_n\int_0^{10^{-10}}|P(t)|dt\,?$$

CHAPTER 3

Differential calculus in $\mathbb{R}^n$

In this section we study the differential calculus of maps $f : U \to \mathbb{R}^m$, $U \subset \mathbb{R}^n$ open. We shall use the Euclidean norms on $\mathbb{R}^n$ and $\mathbb{R}^m$, i.e. $\|x\| = (\sum_{j=1}^n |x_j|^2)^{1/2}$ for $x \in \mathbb{R}^n$ and $\|y\| = (\sum_{i=1}^m |y_i|^2)^{1/2}$ for $y \in \mathbb{R}^m$. In this setting we refer to the Fréchet derivative as *total derivative* . Whenever we speak of *functions* in this section, we mean real-valued functions.

DEFINITION 3.1. By e_k we denote the kth unit vector in $\mathbb{R}^n$. Then the directional derivative in the direction e_k is called *kth partial derivative* and denoted by $\partial_k f(x)$ or $\partial_{x_k} f(x)$ (if it exists).

If f is totally differentiable at a point $x = (x_1, \dots, x_n) \in U$, then we can compute its total derivative in terms of the partial derivatives by using (2.90):

$$Df|_x h = \sum_{j=1}^n h_j Df|_x e_j = \sum_{j=1}^n h_j \partial_j f(x) \in \mathbb{R}^m. \tag{3.1}$$

By definition, $Df|_x$ is a linear map $\mathbb{R}^n \to \mathbb{R}^m$. It is therefore given by multiplication with a real $m \times n$ matrix. We will denote this matrix also by $Df|_x$ and call it the *Jacobian matrix of f at x* . From (2.90) we conclude that the jth column vector of this matrix is given by $\partial_j f(x) \in \mathbb{R}^m$. Therefore the Jacobian matrix is given by

$$Df|_x = (\partial_j f_i(x))_{i,j} = \begin{pmatrix} \partial_1 f_1(x) & \cdots & \partial_n f_1(x) \\ \vdots & \ddots & \vdots \\ \partial_1 f_m(x) & \cdots & \partial_n f_m(x) \end{pmatrix} \in \mathbb{R}^{m \times n}, \tag{3.2}$$

where $f(x) = (f_1(x), \dots, f_m(x)) \in \mathbb{R}^m$. If $m = 1$, then the *gradient* of f at x is defined as[1]

$$\nabla f(x) = Df|_x^T = \begin{pmatrix} \partial_1 f(x) \\ \vdots \\ \partial_n f(x) \end{pmatrix} \in \mathbb{R}^n. \tag{3.3}$$

(Note that $n \times 1$ matrices are identified with vectors in $\mathbb{R}^n$: $\mathbb{R}^{n \times 1} = \mathbb{R}^n$.)

EXAMPLE 3.2. Let $F : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by $F(x) = (x_1 x_2 \sin(x_3), x_2^2 - e^{x_1})$. Then F is totally differentiable and the Jacobian is given by

$$DF|_x = \begin{pmatrix} x_2 \sin(x_3) & x_1 \sin(x_3) & x_1 x_2 \cos(x_3) \\ -e^{x_1} & 2x_2 & 0 \end{pmatrix}. \tag{3.4}$$

Recall that a set $A \subset \mathbb{R}^n$ is called *convex* if $tx + (1-t)y \in A$ for every $x, y \in A$, $t \in [0, 1]$.

[1]Here $M^T \in \mathbb{R}^{n \times m}$ denotes the transpose of the matrix $M \in \mathbb{R}^{m \times n}$.

THEOREM 3.3 (Mean value theorem). *Let $U \subset \mathbb{R}^n$ be open and convex. Suppose that $f : U \to \mathbb{R}$ is totally differentiable on U. Then, for every $x, y \in U$, there exists $\xi \in U$ such that*

$$f(x) - f(y) = Df|_\xi(x - y) \tag{3.5}$$

and there exists $t \in [0, 1]$ such that $\xi = tx + (1 - t)y$.

The idea of the proof is to apply the one-dimensional mean value theorem to the function restricted to the line passing through x and y.

PROOF. If $x = y$ there is nothing to show. Let $x \neq y$. Define $g : [0, 1] \to \mathbb{R}$ by $g(t) = f(tx + (1 - t)y)$. The function g is continuous on $[0, 1]$ and differentiable on $(0, 1)$. By the one-dimensional mean value theorem there exists $t_0 \in [0, 1]$ such that $g(1) - g(0) = g'(t_0)$. By the chain rule,

$$g'(t_0) = Df|_{t_0x+(1-t_0)y}(x - y). \tag{3.6}$$

□

COROLLARY 3.4. *Under the assumptions of the previous theorem: if $Df|_x = 0$ for all $x \in U$, then f is constant.*

EXERCISE 3.5. Show that the conclusion of the corollary also holds under the weaker assumption that U is open and connected (rather than convex). *Hint:* Consider overlapping open balls along a continuous path connecting two given points in U.

DEFINITION 3.6. A map $f : U \to \mathbb{R}^m$, $U \subset \mathbb{R}^n$ open, is called *continuously differentiable* (on U) if it is totally differentiable on U and the map $U \to L(\mathbb{R}^n, \mathbb{R}^m)$, $x \mapsto Df|_x$ is continuous. We denote the collection of continuously differentiable maps by $C^1(U, \mathbb{R}^m)$. If $m = 1$ we also write $C^1(U, \mathbb{R}) = C^1(U)$.

Remark. For $f : U \to \mathbb{R}$, continuity of the map $U \to \mathbb{R}^n$, $x \mapsto \nabla f(x)$ is equivalent to continuity of the map $U \to L(\mathbb{R}^n, \mathbb{R})$, $x \mapsto Df|_x$.

THEOREM 3.7. *Let $U \subset \mathbb{R}^n$ be open. Let $f : U \to \mathbb{R}$. Then $f \in C^1(U)$ if and only if $\partial_j f(x)$ exists for every $j \in \{1, \dots, n\}$ and $x \mapsto \partial_j f(x)$ is continuous on U for $j \in \{1, \dots, n\}$.*

Remark. Without additional assumptions (such as continuity of $x \mapsto \partial_j f(x)$), existence of partial derivatives does not imply total differentiability.

EXERCISE 3.8. Let $F : \mathbb{R}^2 \to \mathbb{R}$ be defined by $F(x) = \frac{x_1 x_2}{x_1^2 + x_2^2}$ if $x \neq 0$ and $F(0) = 0$.
(i) Show that the partial derivatives $\partial_1 F(x)$, $\partial_2 F(x)$ exist for every $x \in \mathbb{R}^2$.
(ii) Show that F is not continuous at $(0, 0)$.
(iii) Determine at which points F is totally differentiable.

PROOF. Let $f \in C^1(U)$. Then $\partial_j f(x)$ exists by Theorem 2.40 and $x \mapsto \partial_j f(x)$ is continuous because it can be written as the composition of the continuous maps $x \mapsto \nabla f(x)$ and $\pi_j : \mathbb{R}^n \to \mathbb{R}, x \mapsto x_j$: $\partial_j f(x) = (\pi_j \circ \nabla f)(x)$.
Conversely, assume that $\partial_j f(x)$ exists for every $x \in U$, $j \in \{1, \dots, n\}$ and $x \mapsto \partial_j f(x)$ is continuous. Let $x \in U$. Write $h = \sum_{j=1}^n h_j e_j$ and define $v_k = \sum_{j=1}^k h_j e_j$ for $1 \leq k \leq n$ and $v_0 = 0$. Then, if $\|h\|$ is small enough so that $x + h \in U$, then

$$f(x+h) - f(x) = f(x+v_n) - f(x+v_{n-1}) + f(x+v_{n-1}) - f(x+v_{n-2}) + \cdots + f(x+v_1) - f(x+v_0) \tag{3.7}$$

$$= \sum_{j=1}^{n} (f(x+v_j) - f(x+v_{j-1})). \tag{3.8}$$

By the one-dimensional mean value theorem there exists $t_j \in [0,1]$ such that

$$f(x+v_j) - f(x+v_{j-1}) = f(x+v_{j-1}+h_j e_j) - f(x+v_{j-1}) = \partial_j f(x+v_{j-1}+t_j h_j e_j) h_j. \tag{3.9}$$

By continuity of $\partial_j f$, for every $\varepsilon > 0$ exists $\delta > 0$ such that

$$|\partial_j f(y) - \partial_j f(x)| \le \varepsilon/n \quad \text{for all } j = 1, \dots, n, \tag{3.10}$$

whenever $y \in U$ is such that $\|x - y\| \le \delta$. We may choose δ small enough so that $x + h \in U$ whenever $\|h\| \le \delta$. Then, if $\|h\| \le \delta$ (then also $\|v_j\| \le \delta, \|v_{j-1} + t_j h_j e_j\| \le \delta$) we get

$$\left| f(x+h) - f(x) - \sum_{j=1}^{n} h_j \partial_j f(x) \right| \le \sum_{j=1}^{n} \left| f(x+v_j) - f(x+v_{j-1}) - h_j \partial_j f(x) \right| \tag{3.11}$$

$$= \sum_{j=1}^{n} |h_j| |\partial_j f(x + v_{j-1} + t_j h_j e_j) - \partial_j f(x)| \le \sum_{j=1}^{n} |h_j| \frac{\varepsilon}{n} \le \varepsilon \|h\|. \tag{3.12}$$

Therefore, $f(x+h) - f(x) - Df|_x h = o(h)$, where

$$Df|_x h = \sum_{j=1}^{n} h_j \partial_j f(x), \tag{3.13}$$

so f is differentiable at x. Also, $x \mapsto \nabla f(x)$ is continuous, because the $\partial_j f$ are continuous. □

To conclude this introductory section, we discuss some variants of the mean value theorem that will be useful later.

THEOREM 3.9 (Mean value theorem, integral version). *Let $U \subset \mathbb{R}^n$ be open and convex and $f \in C^1(U)$. Then for every $x, y \in U$,*

$$f(x) - f(y) = \int_0^1 Df|_{tx+(1-t)y}(x-y)\,dt. \tag{3.14}$$

PROOF. Let $g(t) = f(tx + (1-t)y)$. By the fundamental theorem of calculus and the chain rule,

$$f(x) - f(y) = g(1) - g(0) = \int_0^1 g'(s)\,ds = \int_0^1 Df|_{tx+(1-t)y}(x-y)\,dt. \tag{3.15}$$

□

THEOREM 3.10 (Mean value theorem, vector-valued case). *Let $U \subset \mathbb{R}^n$ be open and convex and $F \in C^1(U, \mathbb{R}^m)$. Then for every $x, y \in U$ there exists $\theta \in [0,1]$ such that*

$$\|F(x) - F(y)\| \le \|DF|_\xi\|_{\mathrm{op}} \|x - y\|, \tag{3.16}$$

where $\xi = \theta x + (1-\theta) y$.

PROOF. Write $F = (F_1, \dots, F_m)$. Then by Theorem 3.9

$$F_i(x) - F_i(y) = \int_0^1 DF_i|_{tx+(1-t)y}(x-y)\,dt. \tag{3.17}$$

This implies

$$F(x) - F(y) = \int_0^1 DF|_{tx+(1-t)y}(x-y)dt. \tag{3.18}$$

By the triangle inequality, we have

$$\|F(x) - F(y)\| \leq \int_0^1 \|DF|_{tx+(1-t)y}\|_{\text{op}} dt \|x-y\| \tag{3.19}$$

The map $[0,1] \to \mathbb{R}$, $t \mapsto \|DF|_{tx+(1-t)y}\|_{\text{op}}$ is continuous (because F is C^1) and therefore assumes its supremum at some point $\theta \in [0,1]$. Define $\xi = \theta x + (1-\theta)y$. Then

$$\|F(x) - F(y)\| \leq \|DF|_\xi\|_{\text{op}} \|x-y\|. \tag{3.20}$$

□

Remark. If $m \geq 2$ and $F : U \to \mathbb{R}^m$ is C^1 and $x, y \in U$, then it is *not* necessarily true that there exists $\xi \in U$ such that $F(x) - F(y) = DF|_\xi(x-y)$.

EXERCISE 3.11. Find a C^1 map $F : \mathbb{R} \to \mathbb{R}^2$ and points $x, y \in \mathbb{R}$ such that there does not exist $\xi \in \mathbb{R}$ such that $F(x) - F(y) = DF|_\xi(x-y)$.

EXERCISE 3.12. Let $U \subset \mathbb{R}^n$ be open and convex and $F : U \to U$ a differentiable map. If there exists $c \in (0,1)$ such that $\|DF|_x\|_{\text{op}} \leq c$ for all $x \in U$, then F is a contraction of U.

1. Inverse function theorem

In this section we will see how the contraction principle can be applied to find (local) inverses of maps between open sets in $\mathbb{R}^n$, in other words to solve equations of the form $f(x) = y$.

DEFINITION 3.13. Let $E \subset \mathbb{R}^n$. We say that a map $f : E \to \mathbb{R}^n$ is *locally invertible* at $a \in E$ if there exist open sets $U, V \subset \mathbb{R}^n$ such that $U \subset E$, $a \in U$, $f(a) \in V$ and a function $g : V \to U$ such that $g(f(x)) = x$ for all $x \in U$ and $f(g(y)) = y$ for all $y \in V$. In that case we call g a *local inverse* of f (at a) and denote it by $f|_U^{-1}$ (this is consistent with usual notation of inverse functions, because the restriction $f|_U$ of f to U is an invertible map $U \to V$).

THEOREM 3.14 (Inverse function theorem). *Let $E \subset \mathbb{R}^n$ be open and let $f : E \to \mathbb{R}^n$ be differentiable on E. Let $a \in E$ and assume that $Df|_a$ is invertible and that $x \mapsto DF_x$ is continuous at a. Then f is locally invertible at a in some open neighborhood $U \subset E$ of a with $(f|_U)^{-1}$ differentiable on $V = f(U)$, and we have for all $x \in U$*

$$D(f|_U^{-1})|_{f(x)} = (Df|_x)^{-1}. \tag{3.21}$$

(ii) If $f \in C^1(E < \mathbb{R}^n)$ the $(f|U)^{-1} \in C^1(V, \mathbb{R}^n)$

PROOF. We want to apply the contraction principle. For fixed $y \in \mathbb{R}^n$, consider the map

$$\varphi_y(x) = x + Df|_a^{-1}(y - f(x)) \quad (x \in E) \tag{3.22}$$

Then $f(x) = y$ if and only if x is a fixed point of φ_y. Calculate

$$D\varphi_y|_x = I - Df|_a^{-1} Df|_x = Df|_a^{-1}(Df|_a - Df|_x). \tag{3.23}$$

Let $\lambda = \|Df|_a^{-1}\|_{\text{op}}$. By continuity of Df at a, there exists an open ball $U \subset E$ such that

$$\|Df|_a - Df|_x\|_{\text{op}} \leq \frac{1}{2\lambda} \quad \text{for } x \in U. \tag{3.24}$$

Then for $x, x' \in U$

$$\|\varphi_y(x) - \varphi_y(x')\| \leq \|D\varphi_y|_\xi\|_{\text{op}}\|x - x'\| \tag{3.25}$$

$$\leq \|Df|_a^{-1}\|_{\text{op}}\|Df|_a - Df|_\xi\|_{\text{op}}\|x - x'\| \leq \tfrac{1}{2}\|x - x'\|. \tag{3.26}$$

Note that this doesn't show that φ_y is a contraction, because $\varphi_y(U)$ may not be contained in U. However, it does show that φ_y has at most one fixed point (by the same argument used to show uniqueness in the Banach fixed point theorem). This already implies that f is injective on U: for every $y \in \mathbb{R}^n$ we have $f(x) = y$ for at most one $x \in U$. Let $V = f(U)$. Then $f|_U : U \to V$ is a bijection and has an inverse $g : V \to U$.

Claim. V is open.

Proof of claim. Let $y_0 \in V$. We need to show that there exists an open ball around y_0 that is contained in V. Since $V = f(U)$ there exists $x_0 \in U$ such that $f(x_0) = y_0$. Let $r > 0$ be small enough so that $\overline{B_r(x_0)} \subset U$ (possible because U is open). Let $\varepsilon > 0$ and $y \in B_\varepsilon(y_0)$. We will demonstrate that if $\varepsilon > 0$ is small enough, then φ_y maps $\overline{B_r(x_0)}$ into itself. First note

$$\|\varphi_y(x_0) - x_0\| = \|Df|_a^{-1}(y - y_0)\| \leq \lambda\varepsilon. \tag{3.27}$$

Hence, choosing $\varepsilon \leq \frac{r}{2\lambda}$, we get for $x \in \overline{B_r(x_0)}$ that

$$\|\varphi_y(x) - x_0\| \leq \|\varphi_y(x) - \varphi_y(x_0)\| + \|\varphi_y(x_0) - x_0\| \tag{3.28}$$

$$\leq \tfrac{1}{2}\|x - x_0\| + \tfrac{r}{2} \leq \tfrac{r}{2} + \tfrac{r}{2} = r. \tag{3.29}$$

Thus $\varphi_y(x) \in \overline{B_r(x_0)}$. This proves $\varphi_y(\overline{B_r(x_0)}) \subset \overline{B_r(x_0)}$, so φ_y is a contraction of $\overline{B_r(x_0)}$. By the Banach fixed point theorem, φ_y must have a unique fixed point $x \in \overline{B_r(x_0)}$. So by definition of φ_y we have $f(x) = y$, so $y \in f(U) = V$. Therefore we have shown that $B_\varepsilon(y_0) \subset V$, so V is open. □

It remains to show that $g \in C^1(V, U)$ and $Dg|_{f(a)} = Df|_a^{-1}$. We use the following lemma.

Lemma 3.15. *Let $A, B \in \mathbb{R}^{n\times n}$ such that A is invertible and*

$$\|B - A\| \cdot \|A^{-1}\| < 1. \tag{3.30}$$

Then B is invertible. (Here $\|\cdot\|$ denotes the matrix norm, which is just the operator norm: $\|A\| = \sup_{\|x\|=1}\|Ax\|$.)

In other words, if a matrix A is invertible and B is a "small" perturbation of A ("small" in the sense that (3.30) holds), then B is also invertible.

Proof. It suffices to show that B is injective. Let $x \neq 0$. Then we need to show $Bx \neq 0$. Indeed,

$$\|x\| = \|A^{-1}Ax\| \leq \|A^{-1}\| \cdot \|Ax\| \leq \|A^{-1}\|(\|(A - B)x\| + \|Bx\|) \tag{3.31}$$

$$\leq \|A^{-1}\| \cdot \|B - A\| \cdot \|x\| + \|A^{-1}\|\|Bx\|, \tag{3.32}$$

which implies $\|A^{-1}\|\|Bx\| \geq (1 - \|A^{-1}\| \cdot \|B - A\|)\|x\| > 0$, so $Bx \neq 0$. □

Let $y \in V$. We show that g is totally differentiable at y. There exists $x \in U$ such that $f(x) = y$ and from the above,

$$\|Df|_a^{-1}\|\|Df|_x - Df|_a\| \leq \tfrac{1}{2}. \tag{3.33}$$

By the lemma, $Df|_x$ is invertible. Let k be such that $y + k \in V$. Then there exists h such that $y + k = f(x + h)$. We have

$$\|h\| \leq \|h - Df|_a^{-1}k\| + \|Df|_a^{-1}k\| \quad \text{and} \tag{3.34}$$

$$h - Df|_a^{-1}k = h + Df|_a^{-1}(f(x) - f(x + h)) = \varphi_y(x + h) - \varphi_y(x), \tag{3.35}$$

so $\|h - Df|_a^{-1}k\| \leq \frac{1}{2}\|h\|$. Therefore, $\|h\| \leq 2\lambda\|k\| \to 0$ as $\|k\| \to 0$. Now we compute

$$g(y + k) - g(y) - Df|_x^{-1}k = x + h - x - Df|_x^{-1}k \tag{3.36}$$

$$= h - Df|_x^{-1}(f(x + h) - f(x)) = -Df|_x^{-1}(f(x + h) - f(x) - Df|_x h) \quad \text{and so} \tag{3.37}$$

$$\frac{1}{\|k\|}\|g(y + k) - g(y) - Df|_x^{-1}k\| \leq \|Df|_x^{-1}\|\frac{\|f(x + h) - f(x) - Df|_x h\|}{\|h\|}\frac{\|h\|}{\|k\|} \tag{3.38}$$

$$\leq \|Df|_x^{-1}\|\frac{\|f(x + h) - f(x) - Df|_x h\|}{\|h\|}2\lambda \longrightarrow 0 \quad \text{as } k \to 0. \tag{3.39}$$

Therefore g is differentiable at y with $Dg|_y = Df|_x^{-1}$. This finishes the proof of part (i) of Theorem 3.14.

To prove part (ii) we assume f is of class C^1, and it remains to show that Dg is continuous. To show this we need another lemma.

LEMMA 3.16. *Let* $\mathrm{GL}(n)$ *denote the space of real invertible* $n \times n$ *matrices (equipped with some norm). The map* $\mathrm{GL}(n) \to \mathrm{GL}(n)$ *defined by* $A \mapsto A^{-1}$ *is continuous.*

This lemma follows because the entries of A^{-1} are rational functions with non-vanishing denominator in terms of the entries of A (by Cramer's rule).

Since $Dg|_y = Df|_x^{-1}$ and compositions of continuous maps are continuous (Df is continuous by assumption), we have that Dg must be continuous, so $g \in C^1(V, U)$. □

Remark. If f is locally invertible at every point, it is not necessarily (globally) invertible (that is, bijective).

EXAMPLE 3.17. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $f(x) = (e^{x_2}\sin(x_1), e^{x_2}\cos(x_1))$. Then

$$Df|_x = \begin{pmatrix} e^{x_2}\cos(x_1) & e^{x_2}\sin(x_1) \\ -e^{x_2}\sin(x_1) & e^{x_2}\cos(x_1) \end{pmatrix}. \tag{3.40}$$

Thus $\det Df|_x = e^{2x_2}(\cos(x_1)^2 + \sin(x_1)^2) = e^{2x_2} \neq 0$, so by Theorem 3.14, f is locally invertible at every point $x \in \mathbb{R}^2$. f is not bijective: it is not injective because, for instance, $f(0, 0) = f(2\pi, 0)$.

2. Implicit function theorem

We will now use the inverse function theorem to prove a significant generalization called the *implicit function theorem*. Let $E \subset \mathbb{R}^n \times \mathbb{R}^m$ be an open set and $f : E \to \mathbb{R}^m$ a C^1 map. Consider the zero set of f given by

$$\mathcal{Z} = \{(x, y) \in E \,:\, f(x, y) = 0\}. \tag{3.41}$$

It is natural to ask when $\mathcal{Z}$ is locally the graph of a function. The implicit function theorem gives a satisfactory answer. More precisely, given a point $(x_0, y_0) \in \mathcal{Z}$ we ask whether there exists an open neighborhood of (x_0, y_0) so that $\mathcal{Z}$ intersected with that neighborhood is given as the graph of a C^1 function in the sense that there exists g so that $f(x, g(x)) = 0$ for x close to x_0. Another way to think of this is that we would like to solve the system of equations given by $f(x, y) = 0$ for y, when x is given (this seems reasonable since there are m equations and m unknowns).

THEOREM 3.18 (Implicit function theorem). *Let $E \subset \mathbb{R}^n \times \mathbb{R}^m$ be open, $f \in C^1(E, \mathbb{R}^m)$ and $(x_0, y_0) \in \mathcal{Z}$ such that the matrix $D_y f|_{(x_0,y_0)} \in \mathbb{R}^{m\times m}$ is invertible. Then there exist open neighborhoods U, V of x_0, y_0, respectively and a C^1 function $g : U \to V$ so that*

$$\mathcal{Z} \cap (U \times V) = \{(x, g(x)) \,:\, x \in U\}. \tag{3.42}$$

In other words, $U \times V \subset E$ and $f(x, g(x)) = 0$ for all $x \in U$. Moreover,

$$Dg|_{x_0} = -(D_y F|_{(x_0,y_0)})^{-1} D_x F|_{(x_0,y_0)}. \tag{3.43}$$

Here, $D_x f|_{(x_0,y_0)} \in \mathbb{R}^{m\times n}$ denotes the Jacobian matrix of the function $x \mapsto f(x, y_0)$ at x_0, and $D_y f|_{(x_0,y_0)} \in \mathbb{R}^{m\times m}$ the Jacobian matrix of the function $y \mapsto f(x_0, y)$ at y_0.

It is instructive to observe that the relation (3.43) follows from an application of the chain rule when taking derivatives on both sides of the identity

$$f(x, g(x)) = 0 \tag{3.44}$$

with respect to x. This is also known as *implicit differentiation*.

The formula (3.43) is especially useful in cases when it is difficult or impossible to determine the implicit function g algebraically.

PROOF. The proof is an application of the inverse function theorem, Theorem 3.14. Define a map $F : E \to \mathbb{R}^n \times \mathbb{R}^m$ by

$$F(x, y) = (x, f(x, y)). \tag{3.45}$$

Then F is C^1 and $DF|_{(x_0,y_0)}$ is given by the $(n + m) \times (n + m)$ block matrix

$$\begin{pmatrix} I_n & 0 \\ D_x f|_{(x_0,y_0)} & D_y f|_{(x_0,y_0)}, \end{pmatrix} \tag{3.46}$$

where I_n denotes the $n \times n$ identity matrix. Thus $\det DF|_{(x_0,y_0)} = \det D_y f|_{(x_0,y_0)} \neq 0$, so $DF|_{(x_0,y_0)}$ is invertible. By Theorem 3.14, F is therefore locally invertible at (x_0, y_0). As a consequence, there exist an open neighborhood U' of x_0 and an open neighborhood V of y_0 so that $U' \times V \subset E$, $F(U \times V) \subset \mathbb{R}^n \times \mathbb{R}^m$ is open and $F|_{U'\times V}$ is invertible with a C^1 inverse

$$G : F(U' \times V) \to U' \times V. \tag{3.47}$$

Let $U = \{x \in U' : (x, 0) \in F(U' \times V)\} \subset \mathbb{R}^n$. Then U is open, because $F(U' \times V)$ is open. Also, $x_0 \in U$. For $x \in U$ we can now *define* $g(x)$ by $G(x, 0) = (x, g(x))$. Then $g(x) \in V$ and

$$(x, f(x, g(x))) = F(x, g(x)) = F(G(x, 0)) = (x, 0), \tag{3.48}$$

so $f(x, g(x)) = 0$ for all $x \in U$. Moreover, g is C^1 and satisfies (3.43). □

EXAMPLE 3.19. Let $n = m = 1$ and $f(x, y) = x^2 + y^2 - 1$. Then $\mathcal{Z}$ is the unit circle around the origin, which is locally a graph at every point with $y_0 \neq 0$. Coincidentally,

$$D_y f|_{(x,y)} = 2y \neq 0 \text{ if and only if } y \neq 0. \tag{3.49}$$

In this case an implicit function g can be determined explicitly: if say $(x_0, y_0) = (0, 1)$, then $g : (-1, 1) \to \mathbb{R}$ with

$$g(x) = \sqrt{1 - x^2} \tag{3.50}$$

is C^1 and satisfies $f(x, g(x)) = 0$.

EXAMPLE 3.20. Let $n = m = 1$ and $f(x, y) = x^2 - y^3$. Then $\mathcal{Z}$ is a cubic curve with a *cusp singularity* at the origin. In this case, $\mathcal{Z}$ is (globally) the graph of the function $g : \mathbb{R} \to \mathbb{R}$ with $g(x) = |x|^{2/3}$. However,

$$D_y f|_{(x,y)} = -3y^2. \tag{3.51}$$

so the implicit function theorem does not apply at the cusp $(x_0, y_0) = (0, 0) \in \mathcal{Z}$. This is consistent with the fact that g is not C^1 at zero.

EXAMPLE 3.21. Let $n = m = 1$ and $f(x, y) = (y - x)(y + x)$. Then $\mathcal{Z}$ is locally the graph of a function at every point except for the origin, where it has a *self-intersection*.

3. Ordinary differential equations

In this section we study *initial value problems* of the form

$$\begin{cases} y'(t) = F(t, y(t)) \\ y(t_0) = y_0, \end{cases} \tag{3.52}$$

where $E \subset \mathbb{R} \times \mathbb{R}$ is open, $(t_0, y_0) \in E$ and $F \in C(E)$ are given. We say that a differentiable function $y : I \to \mathbb{R}$ defined on some open interval $I \subset \mathbb{R}$ that includes the point $t_0 \in I$ is a *solution* to the initial value problem if $(t, y(t)) \in E$ for all $t \in I$ and $y(t_0) = y_0$ and $y'(t) = F(t, y(t))$ for all $t \in I$. The equation $y'(t) = F(t, y(t))$ is a *first order ordinary differential equation* . We also write this differential equation in short form as

$$y' = F(t, y). \tag{3.53}$$

Geometric interpretation. At each point $(t, y) \in E$ imagine a small line segment with slope $F(t, y)$. We are looking for a function such that its graph has the slope $F(t, y)$ at each point (t, y) on the graph of the function.

EXAMPLE 3.22. Consider the equation $y' = \frac{y}{t}$. The solutions of this equation are of the form $y(t) = ct$ for $c \in \mathbb{R}$.

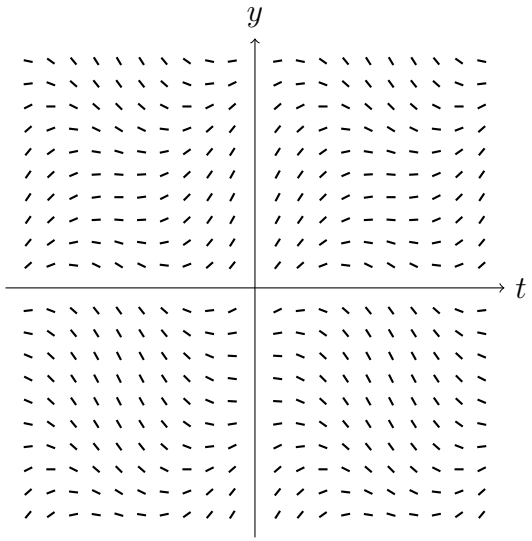

FIGURE 1. Visualization of $F(t, y)$.

EXAMPLE 3.23. Sometimes we can solve initial value problems by computing an explicit expression for y. Recall for instance that solving differential equations of the form $y' = f(t)g(y)$ is easy (by *separation of variables*). Consider for instance

$$\begin{cases} y'(t) = \frac{t}{y(t)} \\ y(t_0) = y_0 \end{cases} \tag{3.54}$$

for $(t_0, y_0) \in (0, \infty) \times (0, \infty)$. Then $y(t) = \sqrt{t^2 + y_0^2 - t_0^2}$. Note that if $y_0^2 - t_0^2 \geq 0$, then y is defined on $I = (0, \infty)$. But if $y_0^2 - t_0^2 < 0$, then y is only defined on $I = (\sqrt{t_0^2 - y_0^2}, \infty) \ni t_0$.

In general, however it is not easy to find a solution. It may also happen that the solution is not expressible in terms of elementary functions. Try for instance, to solve the initial value problem

$$\begin{cases} y'(t) = e^{y(t)^2 t^2} \sin(t + y(t)), \\ y(1) = 5. \end{cases} \tag{3.55}$$

THEOREM 3.24 (Picard-Lindelöf). *Let $E \subset \mathbb{R} \times \mathbb{R}$ be open, $(t_0, y_0) \in E$, $F \in C(E)$. Let $a > 0$ and $b > 0$ be small enough such that*

$$R = \{(t, y) \in \mathbb{R}^2 : |t - t_0| \leq a, |y - y_0| \leq b\} \subset E. \tag{3.56}$$

Let $M = \sup_{(t,y)\in R} |F(t, y)| < \infty$. Assume that there exists $c \in (0, \infty)$ such that

$$|F(t, y) - F(t, u)| \leq c|y - u| \tag{3.57}$$

for all $(t, y), (t, u) \in R$. Define $a_* = \min(a, b/M)$ and let $I = [t_0 - a_*, t_0 + a_*]$. Then there exists a unique solution $y : I \to \mathbb{R}$ to the initial value problem

$$\begin{cases} y'(t) = F(t, y(t)), \\ y(t_0) = y_0. \end{cases} \tag{3.58}$$

Remarks. 1. If F satisfies condition (3.57), we also say that F is *Lipschitz continuous in the second variable* . Note that the solution interval I guaranteed by the theorem is independent on the Lipschitz constant.
2. The condition (3.57) follows if F is differentiable in the second variable and $|\partial_y F(t, y)| \leq$

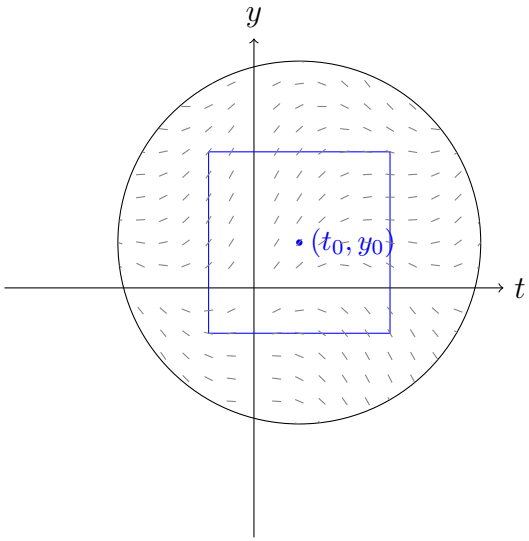

FIGURE 2. Visualization of $F(t, y)$.

c for every $(t, y) \in R$ (by the mean value theorem).
3. By the fundamental theorem of calculus, the initial value problem (3.58) is equivalent to the integral equation

$$y(t) = y_0 + \int_{t_0}^{t} F(s, y(s))ds. \tag{3.59}$$

COROLLARY 3.25. *Let $E \subset \mathbb{R} \times \mathbb{R}$ open, $(t_0, y_0) \in E$, $F \in C^1(E)$. Then there exists an interval $I \subset \mathbb{R}$ and a unique differentiable function $y : I \to \mathbb{R}$ such that $(t, y(t)) \in E$ for all $t \in I$ and y solves* (3.58).

This is true because (3.57) follows from the mean value theorem and continuity of the second derivative $\partial_y F$.

PROOF OF THEOREM 3.24. Let $J = [y_0 - b, y_0 + b]$. It suffices to show that there exists a unique continuous function $y : I \to J$ such that

$$y(t) = y_0 + \int_{t_0}^{t} F(s, y(s))ds \tag{3.60}$$

(that is, y is a solution of the integral equation). Let

$$\mathcal{Y} = \{y : I \to J \ : \ y \text{ continuous on } I\}. \tag{3.61}$$

For every $y \in \mathcal{Y}$, $t \mapsto F(t, y(t))$ is a well-defined continuous function on I. Define

$$Ty(t) = y_0 + \int_{t_0}^{t} F(s, y(s))ds. \tag{3.62}$$

Claim. $T\mathcal{Y} \subset \mathcal{Y}$.

PROOF OF CLAIM. Let $y \in \mathcal{Y}$. Then Ty is a continuous function on I. It remains to show that $Ty(t) \in J$ for all $t \in I$. Recalling that $|F(t, y)| \leq M$ for all $(t, y) \in R$ we obtain:

$$|Ty(t) - y_0| \leq \int_{t_0}^{t} |F(s, y(s))|ds \leq |t_0 - t|M \leq Ma_* \leq b, \tag{3.63}$$

where we used that $a_* = \min(a, b/M) \le b/M$. □

To apply the contraction principle we need to equip $\mathcal{Y}$ with a metric such that $T : \mathcal{Y} \to \mathcal{Y}$ is a contraction and $\mathcal{Y}$ is complete. We could be tempted to try the usual supremum metric $d_\infty(g_1, g_2) = \sup_{t\in I} |g_1(t) - g_2(t)|$. Then $\mathcal{Y} \subset C(I)$ is closed, so $(\mathcal{Y}, d_\infty)$ is a complete metric space. However, T will not necessarily be a contraction[2] with respect to d_∞. Instead, we define the metric

$$d_*(g_1, g_2) = \sup_{t\in I} e^{-2c|t-t_0|}|g_1(t) - g_2(t)|. \tag{3.64}$$

Then $d_*(g_1, g_2) \le d_\infty(g_1, g_2) \le e^{2ca_*} d_*(g_1, g_2)$. In other words, d_* and d_∞ are equivalent metrics. This implies that $(\mathcal{Y}, d_*)$ is still complete.

Claim. $T : \mathcal{Y} \to \mathcal{Y}$ is a contraction with respect to d_*.

Proof of claim. For $g_1, g_2 \in \mathcal{Y}$, $t \in I$ we have by (3.57),

$$|Tg_1(t) - Tg_2(t)| = \Big| \int_{t_0}^t (F(s, g_1(s)) - F(s, g_2(s)))ds \Big| \tag{3.65}$$

$$\le c \int_{t_0}^t |g_1(s) - g_2(s)|ds. \tag{3.66}$$

Let us assume that $t \in [t_0, t_0 + a_*]$. Then

$$|Tg_1(t) - Tg_2(t)| \le c \int_{t_0}^t |g_1(s) - g_2(s)|ds \le c \int_{t_0}^t e^{2c(s-t_0)} d_*(g_1, g_2)ds \tag{3.67}$$

$$= c d_*(g_1, g_2) \frac{1}{2c}(e^{2c(t-t_0)} - 1) \le \tfrac{1}{2} d_*(g_1, g_2) e^{2c|t-t_0|} \tag{3.68}$$

Similarly, for $t \in [t_0 - a_*, t_0]$ we also have

$$|Tg_1(t) - Tg_2(t)| \le \tfrac{1}{2} d_*(g_1, g_2) e^{2c|t-t_0|}. \tag{3.69}$$

Thus,

$$e^{-2c|t-t_0|}|Tg_1(t) - Tg_2(t)| \le \tfrac{1}{2} d_*(g_1, g_2) \tag{3.70}$$

holds for all $t \in I$, so $d_*(Tg_1, Tg_2) \le \frac{1}{2} d_*(g_1, g_2)$. □

By the Banach fixed point theorem, there exists a unique $y \in \mathcal{Y}$ such that $Ty = y$, i.e. a unique solution to the initial value problem (3.58). □

Remarks. 1. The proof is constructive. That is, it tells us how to compute the solution. This is because the proof of the Banach fixed point theorem is constructive. Indeed, construct a sequence $(y_n)_{n\ge 0} \subset \mathcal{Y}$ by $y_0(t) = y_0$ and

$$y_n(t) = y_0 + \int_{t_0}^t F(s, y_{n-1}(s))ds \quad \text{for } n = 1, 2, \dots \tag{3.71}$$

Then $(y_n)_{n\ge 0}$ converges uniformly on I to the solution y. This method is called *Picard iteration* .

2. Note that the length of the existence interval I does not depend on the size of the constant c in (3.57).

[2]For the supremum metric to give rise to a contraction we would need to make the interval I smaller.

EXAMPLE 3.26. Consider the initial value problem

$$\begin{cases} y'(t) = \frac{e^t \sin(t+y(t))}{ty(t)-1}, \\ y(1) = 5. \end{cases} \tag{3.72}$$

Let $F(t,y) = \frac{e^t \sin(t+y)}{ty-1}$. We need to choose a rectangle R around the point $(1,5)$ where we have control over $|F(t,y)|$ and $|\partial_y F(t,y)|$. Thus we need to stay away from the set of (t,y) such that $ty-1=0$. Say,

$$R = \{(t,y) \,:\, |t-1| \le \tfrac{1}{2},\, |y-5| \le 1\}. \tag{3.73}$$

Then for $(t,y) \in R$:

$$|ty-1| \ge (1-\tfrac{1}{2})(5-1) - 1 = 1. \tag{3.74}$$

Also, $|e^t \sin(t+y)| \le e^{3/2}$. Setting $M = e^{3/2}$, we obtain

$$|F(t,y)| \le M \quad \text{for all } (t,y) \in R. \tag{3.75}$$

Compute

$$\partial_y F(t,y) = \frac{e^t \cos(t+y)}{ty-1} - t\frac{e^t \sin(t+y)}{(ty-1)^2}. \tag{3.76}$$

For $(t,y) \in R$ we estimate

$$|\partial_y F(t,y)| \le \Big|\frac{e^t \cos(t+y)}{ty-1}\Big| + \Big|t\frac{e^t \sin(t+y)}{(ty-1)^2}\Big| \le c, \tag{3.77}$$

where we have set $c = e^{3/2} + \frac{3}{2}e^{3/2}$. Then the number a_* from Theorem 3.24 is $a_* = \min(a, b/M) = \min(\frac{1}{2}, 1/e^{3/2}) = e^{-3/2}$. So the theorem yields the existence and uniqueness of a solution the the initial value problem (3.72) in the interval $I = [1-e^{-3/2}, 1+e^{-3/2}]$. We can also compute that solution by Picard iteration: let $y_0(t) = 5$ and

$$y_n(t) = 5 + \int_1^t \frac{e^s \sin(s + y_{n-1}(s))}{s y_{n-1}(s) - 1} ds. \tag{3.78}$$

The sequence $(y_n)_{n\in\mathbb{N}}$ converges uniformly on I to the solution y.

EXAMPLE 3.27. Sometimes one can extend solutions beyond the interval obtained from the Picard-Lindelöf theorem. Consider the initial value problem

$$\begin{cases} y'(t) = \cos(y(t)^2 - 2t^3) \\ y(0) = 1 \end{cases} \tag{3.79}$$

We claim that there exists a unique solution $y : \mathbb{R} \to \mathbb{R}$. To prove this it suffices to demonstrate the existence of a unique solution on the interval $[-L, L]$ for every $L > 0$. To do this we invoke the Picard-Lindelöf theorem. Set

$$R = \{(t,y) \in \mathbb{R}^2 \,:\, |t| \le L, |y-1| \le L\}. \tag{3.80}$$

Let $F(t,y) = \cos(y^2 - 2t^3)$. Then

$$|F(t,y)| \le 1 \quad \text{for all } (t,y) \in \mathbb{R}^2. \tag{3.81}$$

We have $\partial_y F(t,y) = -2y \sin(y^2 - 2t^3)$, so $|\partial_y F(t,y)| \le 2|y| \le 2(L+1)$ for all $(t,y) \in R$. Then by Theorem 3.24, there exists a unique solution to (3.79) on $I = [-L, L]$.

EXAMPLE 3.28. If the Lipschitz condition (3.57) fails, then the initial value problem may have more than one solution. Consider

$$\begin{cases} y'(t) = |y(t)|^{1/2}, \\ y(0) = 0. \end{cases} \tag{3.82}$$

The function $y \mapsto |y|^{1/2}$ is not Lipschitz continuous in any neighborhood of 0: for $y > 0$ its derivative $\frac{1}{2}y^{-1/2}$ is unbounded as $y \to 0$. The function $y_1(t) = 0$ solves the initial value problem (3.82). The function

$$y_2(t) = \begin{cases} t^2/4, & \text{if } t > 0, \\ 0, & \text{if } t \leq 0 \end{cases} \tag{3.83}$$

also does.

Existence of a solution still holds without the assumption (3.57). We will prove this as a consequence of the Arzelá-Ascoli theorem.

THEOREM 3.29 (Peano existence theorem). *Let $E \subset \mathbb{R} \times \mathbb{R}$ open, $(t_0, y_0) \in E$, $F \in C(E)$,*

$$R = \{(t, y) \,:\, |t - t_0| \leq a,\, |y - y_0| \leq b\} \subset E. \tag{3.84}$$

Let $M = \sup_{(t,y)\in R} |F(t, y)| < \infty$. Define $a_* = \min(a, b/M)$ and let $I = [t_0 - a_*, t_0 + a_*]$. Then there exists a solution $y : I \to \mathbb{R}$ to the initial value problem

$$\begin{cases} y'(t) = F(t, y(t)), \\ y(t_0) = y_0. \end{cases} \tag{3.85}$$

COROLLARY 3.30. *Let $E \subset \mathbb{R} \times \mathbb{R}$ open, $(t_0, y_0) \in E$, $F \in C(E)$. Then there exists an interval $I \subset \mathbb{R}$ and a differentiable function $y : I \to \mathbb{R}$ such that $(t, y(t)) \in E$ for all $t \in I$ and y solves* (3.58).

PROOF. It suffices to produce a solution to the integral equation

$$y(t) = y_0 + \int_{t_0}^{t} F(s, y(s))ds. \tag{3.86}$$

To avoid some technicalities we will only present the proof under the additional assumption that

$$|F(t, y)| \leq M \tag{3.87}$$

holds for $|t - t_0| \leq a$ and *all* $y \in \mathbb{R}$. Then we may choose b arbitrarily large and thus $a_* = a$. We also restrict our attention to the interval $[t_0, t_0 + a]$, which we denote by I. The construction is similar on the other half, $[t_0 - a, t_0]$. Let $\mathcal{P}$ be a partition of $[t_0, t_0 + a]$: $\mathcal{P} = \{t_0 < t_1 < \cdots < t_N = t_0 + a\}$ of $[t_0, t_0 + a]$. We let $\Delta\mathcal{P} = \max_{0\leq k\leq N-1}(t_{k+1} - t_k)$ denote the *fineness* of $\mathcal{P}$. We try to build an approximate solution given as a piecewise linear function. The function $y_\mathcal{P} : [t_0, t_0 + a] \to \mathbb{R}$ shall be defined as follows: let $y_\mathcal{P}(t_0) = y_0$ and for $t \in (t_k, t_{k+1}]$ we define $y_\mathcal{P}(t)$ recursively by

$$y_\mathcal{P}(t) = y_\mathcal{P}(t_k) + F(t_k, y_\mathcal{P}(t_k))(t - t_k). \tag{3.88}$$

Claim 1. For $t, t' \in [t_0, t_0 + a]$,

$$|y_\mathcal{P}(t) - y_\mathcal{P}(t')| \leq M|t - t'|. \tag{3.89}$$

PROOF OF CLAIM. In this proof we will write $y_{\mathcal{P}}$ as y for brevity. Say $t' \in [t_k, t_{k+1}], t \in [t_\ell, t_{\ell+1}]$, $k \leq \ell$. If $k = \ell$, then by (3.87),

$$|y(t) - y(t')| = |F(t_k, y(t_k))(t - t')| \leq M|t - t'|. \tag{3.90}$$

If $k < \ell$, then

$$|y(t) - y(t')| = |y(t) - y(t_\ell) + \sum_{j=k+1}^{\ell-1} (y(t_{j+1}) - y(t_j)) + y(t_{k+1}) - y(t')| \tag{3.91}$$

$$\leq |y(t) - y(t_\ell)| + \sum_{j=k+1}^{\ell-1} |y(t_{j+1}) - y(t_j)| + |y(t_{k+1}) - y(t')| \tag{3.92}$$

$$\leq M(t - t_\ell) + \sum_{j=k+1}^{\ell-1} M(t_{j+1} - t_j) + M(t_{k+1} - t') = M(t - t'). \tag{3.93}$$

□

Define $g_{\mathcal{P}}(t) = F(t_k, y_{\mathcal{P}}(t_k))$ for $t \in (t_k, t_{k+1}]$. Then $g_{\mathcal{P}}$ is a step function and $y'_{\mathcal{P}}(t) = g_{\mathcal{P}}(t)$ for $t \in (t_k, t_{k+1})$.

Let $\varepsilon > 0$. F is uniformly continuous on R, because R is compact (Theorem 1.53). Thus there exists $\delta = \delta(\varepsilon) > 0$ such that

$$|F(t, y) - F(t', y')| \leq \varepsilon \tag{3.94}$$

for all $(t, y), (t', y') \in R$ with $\|(t, y) - (t', y')\| \leq 100\delta$.

Claim 2. Suppose that $\Delta\mathcal{P} \leq \delta(\varepsilon) \min(1, M^{-1})$. Then we have for all $t \in [t_0, t_0 + a]$ that

$$y_{\mathcal{P}}(t) = y_0 + \int_{t_0}^{t} g_{\mathcal{P}}(s)ds \quad \text{and} \quad |g_{\mathcal{P}}(s) - F(s, y_{\mathcal{P}}(s))| \leq \varepsilon \text{ if } s \in (t_{k-1}, t_k). \tag{3.95}$$

PROOF OF CLAIM. We will write y instead of $y_{\mathcal{P}}$ and g instead of $g_{\mathcal{P}}$ in this proof. First we have for $t = t_k$:

$$y(t_k) - y_0 = y(t_k) - y(t_0) = \sum_{j=1}^{k} y(t_j) - y(t_{j-1}) \tag{3.96}$$

$$= \sum_{j=1}^{k} F(t_{j-1}, y(t_{j-1}))(t_j - t_{j-1}) = \sum_{j=1}^{k} \int_{t_{j-1}}^{t_j} g(s)ds = \int_{t_0}^{t_k} g(s)ds. \tag{3.97}$$

Similarly, for $t \in (t_k, t_{k+1})$:

$$y(t) - y(t_k) = F(t_k, y(t_k))(t - t_k) = \int_{t_k}^{t} g(s)ds. \tag{3.98}$$

Thus,

$$y(t) = y(t_k) + \int_{t_k}^{t} g(s)ds = y_0 + \int_{t_0}^{t_k} g(s)ds + \int_{t_k}^{t} g(s)ds = y_0 + \int_{t_0}^{t} g(s)ds. \tag{3.99}$$

Let $s \in (t_{k-1}, t_k)$. Then

$$|g(s) - F(s, y(s))| = |F(t_{k-1}, y(t_{k-1})) - F(s, y(s))|. \tag{3.100}$$

We have

$$|y(t_{k-1}) - y(s)| \le M|t_{k-1} - s| \le M(t_k - t_{k-1}) \le M \cdot \Delta\mathcal{P} \le \delta. \tag{3.101}$$

Also, $|t_{k-1} - s| \le t_k - t_{k-1} \le \Delta\mathcal{P} \le \delta$. Thus,

$$\|(t_{k-1}, y(t_{k-1})) - (s, y(s))\| \le 100\delta. \tag{3.102}$$

By (3.94),

$$|g(s) - F(s, y(s))| = |F(t_{k-1}, y(t_{k-1})) - F(s, y(s))| \le \varepsilon. \tag{3.103}$$

□

Claim 3. Suppose that $\Delta\mathcal{P} \le \delta(\varepsilon)\min(1, M^{-1})$. Then it holds for all $t \in [t_0, t_0 + a]$ that

$$|y_{\mathcal{P}}(t) - (y_0 + \int_{t_0}^{t} F(s, y_{\mathcal{P}}(s))ds)| \le \varepsilon a. \tag{3.104}$$

Proof of claim. By Claim 2, the left hand side equals

$$\Big| \int_{t_0}^{t} (g_{\mathcal{P}}(s) - F(s, y_{\mathcal{P}}(s)))ds \Big| \le \int_{t_0}^{t} |g_{\mathcal{P}}(s) - F(s, y_{\mathcal{P}}(s))|ds. \tag{3.105}$$

Claim 2 implies that this is no larger than $\varepsilon(t - t_0) \le \varepsilon a$. □

Claim 3 says that $y_{\mathcal{P}}$ is almost a solution if the partition $\mathcal{P}$ is sufficiently fine. In the final step we use a compactness argument to obtain an honest solution.

Claim 4. The set $\mathcal{F} = \{y_{\mathcal{P}} : \mathcal{P} \text{ partition}\} \subset C([t_0, t_0 + a])$ is relatively compact.

Proof of claim. By Claim 1,

$$|y_{\mathcal{P}}(t) - y_{\mathcal{P}}(t')| \le M|t - t'|. \tag{3.106}$$

This implies that $\mathcal{F}$ is equicontinuous. It is also bounded:

$$|y_{\mathcal{P}}(t)| \le |y_{\mathcal{P}}(t_0)| + |y_{\mathcal{P}}(t) - y_{\mathcal{P}}(t_0)| \le |y_0| + M|t - t_0| \le |y_0| + Ma. \tag{3.107}$$

Thus the claim follows from the Arzelà-Ascoli theorem (Theorem 1.74). □

For $n \in \mathbb{N}$, choose a partition $\mathcal{P}_n$ with $\Delta\mathcal{P}_n \le \delta(1/n)\min(1, M^{-1})$. By compactness of $\overline{\mathcal{F}}$, the sequence $(y_{\mathcal{P}_n})_{n\in\mathbb{N}} \subset \mathcal{F} \subset \overline{\mathcal{F}}$ has a convergent subsequence that converges to some limit $y \in C([t_0, t_0 + a])$. It remains to show that y is a solution to the integral equation (3.86). Let us denote that subsequence by $(y_n)_{n\in\mathbb{N}}$. By (uniform) continuity of F, we have that $F(s, y_n(s)) \to F(s, y(s))$ uniformly in $s \in [t_0, t]$ as $n \to \infty$. Thus,

$$\int_{t_0}^{t} F(s, y_n(s))ds \longrightarrow \int_{t_0}^{t} F(s, y(s))ds \quad \text{as } n \to \infty. \tag{3.108}$$

On the other hand, by Claim 3 we get

$$|y_n(t) - (y_0 + \int_{t_0}^{t} F(s, y_n(s))ds)| \le \frac{a}{n} \longrightarrow 0 \quad \text{as } n \to \infty. \tag{3.109}$$

Therefore, y solves the integral equation (3.86). □

The theory for ordinary differential equations that we have developed turns out to be far more general.

Systems of first-order ordinary differential equations. The proofs of the Picard-Lindelöf theorem and the Peano existence theorem can easily be extended to apply to systems of differential equations:

$$\begin{cases} y'(t) = F(t, y(t)), \\ y(t_0) = y_0 \end{cases} \tag{3.110}$$

for $F : E \to \mathbb{R}^m$, $E \subset \mathbb{R} \times \mathbb{R}^m$ open, $(t_0, y_0) \in E$.

Higher-order differential equations. Let $d \geq 1$ and consider the d-th order ordinary differential equation given by

$$y^{(d)}(t) = F(t, y(t), y'(t), \ldots, y^{(d-1)}(t)) \tag{3.111}$$

for some $F : E \to \mathbb{R}$, $E \subset \mathbb{R} \times \mathbb{R}^d$ open. We can transform this equation into a system of d first-order equations: if $Y = (Y_1, \ldots, Y_d)$ solves the system

$$\begin{cases} Y_1'(t) = Y_2(t) \\ Y_2'(t) = Y_3(t) \\ \vdots \\ Y_{d-1}'(t) = Y_d(t) \\ Y_d'(t) = F(t, Y(t)) \end{cases} \tag{3.112}$$

then Y_d is a solution to (3.111).

4. Higher order derivatives and Taylor's theorem

DEFINITION 3.31. Let $U \subset \mathbb{R}^n$ be open and $f : U \to \mathbb{R}$. We define the *partial derivatives of second order* as

$$\partial_{ij} f = \partial_{x_i x_j} f = \partial_{x_i}(\partial_{x_j} f) \quad \text{for } i, j \in \{1, \ldots, n\} \tag{3.113}$$

(if $\partial_{x_j} f$, $\partial_{x_i}(\partial_{x_j} f)$ exist). If $\partial_i f$ and $\partial_{ij} f$ exist and are continuous for all $i, j \in \{1, \ldots, n\}$, then we say that $f \in C^2(U)$.

THEOREM 3.32 (Schwarz). *Let $U \subset \mathbb{R}^n$ open, $f : U \to \mathbb{R}$ such that $\partial_{x_i} f, \partial_{x_j} f$, $\partial_{x_i x_j} f$ exist at every point in U and $\partial_{x_i x_j} f$ is continuous at some point $x_0 \in U$. Then $\partial_{x_j x_i} f(x_0)$ exists and*

$$\partial_{x_j x_i} f(x_0) = \partial_{x_i x_j} f(x_0). \tag{3.114}$$

PROOF. Without loss of generality assume that $n = 2$, $i = 1$, $j = 2$. Let f be as in the theorem and $x_0 = (a, b) \in U$ and $(h, k) \in \mathbb{R}^2 \backslash \{0\}$ such that $(a + h, b + k)$ are contained in an open ball around x_0 that is contained in U. We want to show that $\partial_{21} f$ exists, so we need to study the expression

$$\partial_1 f(a, b + k) - \partial_1 f(a, b). \tag{3.115}$$

This leads us to consider the quantity

$$\Delta(a, b, h, k) = (f(a + h, b + k) - f(a, b + k)) - (f(a + h, b) - f(a, b)). \tag{3.116}$$

Define $g(y) = f(a + h, y) - f(a, y)$. Since $\partial_2 f$ exists at every point in U, the mean value theorem implies that there exists $\eta = \eta_{h,k}$ contained in the closed interval with endpoints b and $b + k$ such that

$$\Delta(a, b, h, k) = g(b + k) - g(b) = g'(\eta)k = k(\partial_2 f(a + h, \eta) - \partial_2 f(a, \eta)) \tag{3.117}$$

Since $\partial_{12}f$ exists at every point in U, another application of the mean value theorem yields

$$\Delta(a,b,h,k) = hk\partial_{12}f(\xi,\eta), \tag{3.118}$$

where $\xi = \xi_h$ is in the closed interval with endpoints a and $a+h$.

Let $\varepsilon > 0$. Since $\partial_{12}f$ is continuous at (a,b),

$$|\partial_{12}f(a,b) - \partial_{12}f(x,y)| \le \varepsilon \tag{3.119}$$

whenever $\|(a,b)-(x,y)\|$ is small enough. Thus, for small enough h and k we have

$$|\partial_{12}f(a,b) - \frac{\Delta(a,b,h,k)}{hk}| \le \varepsilon. \tag{3.120}$$

Letting $h \to 0$ and using that $\partial_1 f$ exists at every point this inequality implies

$$\left|\partial_{12}f(a,b) - \frac{\partial_1 f(a,b+k) - \partial_1 f(a,b)}{k}\right| \le \varepsilon. \tag{3.121}$$

In other words, $\partial_{21}f(a,b)$ exists and equals $\partial_{12}f(a,b)$. □

This is not true without the assumption that $\partial_{x_ix_j}f$ is continuous at x.

EXERCISE 3.33. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x,y) = \begin{cases} xy\frac{x^2-y^2}{x^2+y^2}, & \text{if } (x,y) \neq (0,0), \\ 0, & \text{if } (x,y) = (0,0). \end{cases} \tag{3.122}$$

Show that $\partial_x\partial_y f$ and $\partial_y\partial_x f$ exist at every point in $\mathbb{R}^2$, but that $\partial_x\partial_y f(0,0) \neq \partial_y\partial_x f(0,0)$.

COROLLARY 3.34. *If $f \in C^2(U)$, then $\partial_{x_ix_j}f = \partial_{x_jx_i}f$ for every $i,j \in \{1,\dots,n\}$.*

DEFINITION 3.35. Let $U \subset \mathbb{R}^n$ be an open set and $f : U \to \mathbb{R}$. Let $k \in \mathbb{N}$. If all partial derivatives of f up to order k exist, i.e. for all $j \in \{1,\dots,k\}$ and $i_1,\dots,i_j \in \{1,\dots,n\}$, the $\partial_{i_1}\cdots\partial_{i_j}f$ exist, and are continuous, then we write $f \in C^k(E)$ and say that f is k *times continuously differentiable* .

COROLLARY 3.36. *If $f \in C^k(U)$ and $\pi : \{1,\dots,k\} \to \{1,\dots,k\}$ is a bijection, then*

$$\partial_{i_1}\cdots\partial_{i_k}f = \partial_{i_{\pi(1)}}\cdots\partial_{i_{\pi(k)}}f \tag{3.123}$$

for all $i_1,\dots,i_k \in \{1,\dots,n\}$.

Multiindex notation. In order to make formulas involving higher order derivatives shorter and more readable, we introduce *multiindex notation* . A *multiindex of order k* is a vector $\alpha = (\alpha_1,\dots,\alpha_n) \in \mathbb{N}_0^n = \{0,1,2,\dots\}^n$ such that $\sum_{i=1}^n \alpha_i = k$. We write $|\alpha| = \sum_{i=1}^n \alpha_i$. For every multiindex α we introduce the notation

$$\partial^\alpha f = \partial_{x_1}^{\alpha_1}\cdots\partial_{x_n}^{\alpha_n}f, \tag{3.124}$$

where $\partial_{x_i}^{\alpha_i}$ is short for $\partial_{x_i}\cdots\partial_{x_i}$ (α_i times). For $x = (x_1,\dots,x_n) \in \mathbb{R}^n$ we also write

$$x^\alpha = x_1^{\alpha_1}\cdots x_n^{\alpha_n} \tag{3.125}$$

and $\alpha! = \alpha_1!\cdots\alpha_n!$. Moreover, for $\alpha,\beta \in \mathbb{N}_0^n$, $\alpha \le \beta$ means that $\alpha_i \le \beta_i$ for every $i \in \{1,\dots,n\}$.

With this notation, we can state Taylor's theorem in $\mathbb{R}^n$ quite succinctly.

THEOREM 3.37 (Taylor). *Let $U \subset \mathbb{R}^n$ be open and convex, $f \in C^{k+1}(U)$ and $x, x+y \in U$. Then there exists $\xi \in U$ such that*

$$f(x+y) = \sum_{|\alpha|\le k} \frac{\partial^\alpha f(x)}{\alpha!} y^\alpha + \sum_{|\alpha|=k+1} \frac{\partial^\alpha f(\xi)}{\alpha!} y^\alpha. \tag{3.126}$$

Moreover, ξ takes the form $\xi = x + \theta y$ for some $\theta \in [0,1]$.

Remark. Without multiindex notation the statement of this theorem would look much more messy:

$$\sum_{|\alpha|\le k} \frac{\partial^\alpha f(x)}{\alpha!} y^\alpha = \sum_{\substack{\alpha_1,\dots,\alpha_n \ge 0,\\ \alpha_1+\cdots+\alpha_n \le k}} \frac{\partial_{x_1}^{\alpha_1}\cdots\partial_{x_n}^{\alpha_n} f(x)}{\alpha_1!\cdots\alpha_n!} y_1^{\alpha_1}\cdots y_n^{\alpha_n}. \tag{3.127}$$

PROOF OF THEOREM A.17. The idea is to apply Taylor's theorem in one dimension to the function $g : [0,1] \to \mathbb{R}$ given by $g(t) = f(x+ty)$. Let us compute the derivatives of g.

Claim. For $m = 1, \dots, k+1$,

$$g^{(m)}(t) = \sum_{|\alpha|=m} \frac{m!}{\alpha!} \partial^\alpha f(x+ty) y^\alpha \tag{3.128}$$

PROOF OF CLAIM. We first show by induction on m that

$$g^{(m)}(t) = \sum_{i_1,\dots,i_m=1}^{n} \partial_{i_1}\cdots\partial_{i_m} f(x+ty) y_{i_1}\cdots y_{i_m}. \tag{3.129}$$

Indeed, for $m = 1$, by the chain rule,

$$g'(t) = \sum_{i=1}^{n} \partial_i f(x+ty) y_i. \tag{3.130}$$

Suppose we have shown it for m. Then

$$g^{(m+1)}(t) = \frac{d}{dt} g^{(m)}(t) = \frac{d}{dt} \sum_{i_1,\dots,i_m=1}^{n} \partial_{i_1}\cdots\partial_{i_m} f(x+ty) y_{i_1}\cdots y_{i_m}. \tag{3.131}$$

By the chain rule this equals

$$= \sum_{i_1,\dots,i_m=1}^{n} \sum_{i=1}^{n} \partial_{i_1}\cdots\partial_{i_m}\partial_i f(x+ty) y_{i_1}\cdots y_{i_m} y_i = \sum_{i_1,\dots,i_{m+1}=1}^{n} \partial_{i_1}\cdots\partial_{i_{m+1}} f(x+ty) y_{i_1}\cdots y_{i_{m+1}}. \tag{3.132}$$

It remains to show that

$$\sum_{i_1,\dots,i_m=1}^{n} \partial_{i_1}\cdots\partial_{i_m} f(x+ty) y_{i_1}\cdots y_{i_m} = \sum_{|\alpha|=m} \frac{m!}{\alpha!} \partial^\alpha f(x+ty) y^\alpha. \tag{3.133}$$

This follows because for a given $\alpha = (\alpha_1, \dots, \alpha_n)$ with $|\alpha| = m$ there are

$$\frac{m!}{\alpha!} = \frac{m!}{\alpha_1!\cdots\alpha_n!} = \binom{m}{\alpha_1}\binom{m-\alpha_1}{\alpha_2}\cdots\binom{m-\alpha_1-\cdots-\alpha_{n-1}}{\alpha_n} \tag{3.134}$$

many tuples $(i_1, \dots, i_m) \in \{1, \dots, n\}^m$ such that i appears exactly α_i times among the i_js. In other words, this is the number of ways to sort m pairwise different marbles into n numbered bins such that bin number i contains exactly α_i marbles. □

By the one-dimensional Taylor theorem, there exists a $\theta \in [0,1]$ such that

$$g(t) = \sum_{m=0}^{k} \frac{g^{(m)}(0)}{m!} t^m + \frac{g^{(k+1)}(\theta)}{(k+1)!} t^{k+1} \tag{3.135}$$

From the claim we see that this equals

$$\sum_{m=0}^{k} \frac{1}{m!} \sum_{|\alpha|=m} \frac{m!}{\alpha!} \partial^\alpha f(x) y^\alpha t^m + \frac{1}{(k+1)!} \sum_{|\alpha|=k+1} \frac{(k+1)!}{\alpha!} \partial^\alpha f(x+\theta y) y^\alpha t^{k+1} \tag{3.136}$$

$$= \sum_{|\alpha|\le k} \frac{\partial^\alpha f(x)}{\alpha!} (ty)^\alpha + \sum_{|\alpha|=k+1} \frac{\partial^\alpha f(\xi)}{\alpha!} (ty)^\alpha, \tag{3.137}$$

where we have set $\xi = x + \theta y$. Letting $t = 1$ we obtain the claim. □

COROLLARY 3.38. *If $E \subset \mathbb{R}^n$ is open and $f \in C^k(E)$, then for every $x \in E$,*

$$f(x+y) = \sum_{|\alpha|\le k} \frac{\partial^\alpha f(x)}{\alpha!} y^\alpha + o(\|y\|^k) \quad \text{as } y \to 0. \tag{3.138}$$

PROOF. Let $x \in E$ and $\delta > 0$ be small enough so that $U = B_\delta(x) \subset E$. By Taylor's theorem we have for every y with $x + y \in U$ that

(3.139)

$$f(x+y) = \sum_{|\alpha|\le k-1} \frac{\partial^\alpha f(x)}{\alpha!} y^\alpha + \sum_{|\alpha|=k} \frac{\partial^\alpha f(x+\theta y)}{\alpha!} y^\alpha = \sum_{|\alpha|\le k} \frac{\partial^\alpha f(x)}{\alpha!} y^\alpha + \sum_{|\alpha|=k} \frac{\partial^\alpha f(x+\theta y) - \partial^\alpha f(x)}{\alpha!} y^\alpha$$

for some $\theta \in [0,1]$. Since $\partial^\alpha f$ is continuous for every $|\alpha| = k$, it holds that

$$|\partial^\alpha f(x+\theta y) - \partial^\alpha f(x)| \to 0 \quad \text{as } y \to 0. \tag{3.140}$$

Also $|y^\alpha| = |y_1|^{\alpha_1} \cdots |y_n|^{\alpha_n} \le \|y\|^{\alpha_1+\cdots+\alpha_n} = \|y\|^{|\alpha|}$, so

$$\sum_{|\alpha|=k} \frac{\partial^\alpha f(x+\theta y) - \partial^\alpha f(x)}{\alpha!} y^\alpha = o(\|y\|^k). \tag{3.141}$$

□

DEFINITION 3.39. Let $E \subset \mathbb{R}^n$ be open and $f \in C^2(E)$. We define the *Hessian matrix* of f at $x \in E$ by

$$D^2 f|_x = (\partial_i \partial_j f(x))_{i,j=1,\dots,n} = \begin{pmatrix} \partial_1^2 f(x) & \cdots & \partial_1 \partial_n f(x) \\ \vdots & \ddots & \vdots \\ \partial_n \partial_1 f(x) & \cdots & \partial_n^2 f(x) \end{pmatrix} \in \mathbb{R}^{n\times n}. \tag{3.142}$$

We call $\det D^2 f|_x$ the *Hessian determinant* of f at $x \in E$.

Sometimes the term *Hessian* is used for both, the matrix and its determinant. By Theorem 3.32 the Hessian matrix is symmetric.

COROLLARY 3.40. *Let $E \subset \mathbb{R}^n$ be open, $f \in C^2(E)$ and $x \in E$. Then*

$$f(x+y) = f(x) + \langle \nabla f(x), y\rangle + \tfrac{1}{2}\langle y, D^2 f|_x y\rangle + o(\|y\|^2) \quad \text{as } y \to 0. \tag{3.143}$$

(Here $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$ denotes the inner product of two vectors $x, y \in \mathbb{R}^n$.)

PROOF. By Corollary 3.38,

$$f(x+y) = f(x) + \sum_{|\alpha|=1} \frac{\partial^\alpha f(x)}{\alpha!} y^\alpha + \sum_{|\alpha|=2} \frac{\partial^\alpha f(x)}{\alpha!} y^\alpha + o(\|y\|^2) \quad \text{as } y \to 0. \tag{3.144}$$

We have

$$\sum_{|\alpha|=1} \frac{\partial^\alpha f(x)}{\alpha!} y^\alpha = \sum_{i=1}^n \partial_i f(x) y_i = \langle \nabla f(x), y \rangle. \tag{3.145}$$

If $|\alpha| = 2$ then either $\alpha = 2e_i$ for some $i \in \{1, \dots, n\}$ or $\alpha = e_i + e_j$ for some $1 \le i < j \le n$. Thus,

$$\sum_{|\alpha|=2} \frac{\partial^\alpha f(x)}{\alpha!} y^\alpha = \tfrac{1}{2} \sum_{i=1}^n \partial_i^2 f(x) y_i^2 + \sum_{1 \le i < j \le n} \partial_i \partial_j f(x) y_i y_j = \tfrac{1}{2} \sum_{i,j=1}^n \partial_i \partial_j f(x) y_i y_j \tag{3.146}$$

$$= \tfrac{1}{2} \sum_{i=1}^n y_i (D^2 f|_x y)_i = \tfrac{1}{2} \langle y, D^2 f|_x y \rangle. \tag{3.147}$$

□

5. Local extrema

Let $E \subset \mathbb{R}^n$ be an open set and $f : E \to \mathbb{R}$ a function.

DEFINITION 3.41. A point $a \in E$ is called a *local maximum* if there exists an open set $U \subset E$ with $a \in U$ such that $f(a) \ge f(x)$ for all $x \in U$. It is called a *strict local maximum* if $f(a) > f(x)$ for all $x \in U$, $x \ne a$. We define the terms *local minimum* , *strict local minimum* accordingly. A point is called a *(strict) local extremum* if it is a (strict) local maximum or a (strict) local minimum.

THEOREM 3.42. *Suppose the partial derivative $\partial_i f$ exists on E. Then, if f has a local extremum at $a \in E$, then $\partial_i f(a) = 0$.*

PROOF. Let $\delta > 0$ be such that $a + te_i \in E$ for all $|t| \le \delta$. Define $g : (-\delta, \delta) \to \mathbb{R}$ by $g(t) = f(a + te_i)$. By the chain rule, g is differentiable and $g'(t) = \partial_i f(a + te_i)$. Also, 0 is a local extremum of g so by Analysis I, $0 = g'(0) = \partial_i f(a)$. □

COROLLARY 3.43. *If f is differentiable at a and a is a local extremum, then $\nabla f(a) = 0$.*

Remark. $\nabla f(a) = 0$ is not a sufficient condition for a to be a local extremum. Think of saddle points.

DEFINITION 3.44. If $a \in E$ is such that $\nabla f(a) = 0$, then we call a a *critical point* of f.

Recall from linear algebra: A matrix $A \in \mathbb{R}^{n \times n}$ is called *positive definite* if $\langle x, Ax \rangle > 0$ for all $x \in \mathbb{R}^n \backslash \{0\}$ and *positive semidefinite* if $\langle x, Ax \rangle \ge 0$ for all $x \in \mathbb{R}^n$. We also write $A > 0$ to express that A is positive definite and $A \ge 0$ to express that A is positive semidefinite. The terms *negative definite, negative semidefinite* are defined accordingly. A is *indefinite* if it is not positive semidefinite and not negative semidefinite. Every real symmetric matrix has real eigenvalues and there is an orthonormal basis of eigenvectors (spectral theorem). A real symmetric matrix is positive definite if and only if all eigenvalues are positive.

THEOREM 3.45. *Let $f \in C^2(E)$ and $a \in E$ with $\nabla f(a) = 0$. Then*

(1) if $D^2 f|_a > 0$, then a is a strict local minimum of f,
(2) if $D^2 f|_a < 0$, then a is a strict local maximum of f,
(3) if $D^2 f|_a$ is indefinite, then a is not a local extremum of f.

Remark. If $D^2 f|_x$ is only positive semidefinite or negative semidefinite, then we need more information to be able to decide whether or not a is a local extremum.

PROOF. We write $A = D^2 f|_a$. Let $\varepsilon > 0$. By Corollary 3.40 there exists $\delta > 0$ such that for all y with $\|y\| \leq \delta$ we have

$$f(a+y) = f(a) + \tfrac{1}{2}\langle y, Ay\rangle + r(y) \tag{3.148}$$

with $|r(y)| \leq \varepsilon \|y\|^2$.
(1): Let A be positive definite. Let $S = \{y \in \mathbb{R}^n : \|y\| = 1\}$. S is compact, so the continuous map $y \mapsto \langle y, Ay\rangle$ attains its minimum on S. That is, there exists $y_0 \in S$ such that

$$\langle y_0, Ay_0\rangle \leq \langle y, Ay\rangle \tag{3.149}$$

for all $y \in S$. Define $\alpha = \langle y_0, Ay_0\rangle$. Since $y_0 \neq 0$ and A is positive definite, $\alpha > 0$. Let $y \in \mathbb{R}^n, y \neq 0$. Then $\frac{y}{\|y\|} \in S$, so

$$\alpha \leq \langle \frac{y}{\|y\|}, A\frac{y}{\|y\|}\rangle = \frac{1}{\|y\|^2}\langle y, Ay\rangle. \tag{3.150}$$

Thus, $\langle y, Ay\rangle \geq \alpha\|y\|^2$ for all $y \in \mathbb{R}^n$. Now we set $\varepsilon = \frac{\alpha}{4}$. Then

$$f(a+y) \geq f(a) + \tfrac{1}{2}\langle y, Ay\rangle - \tfrac{\alpha}{4}\|y\|^2 \geq f(a) + \tfrac{\alpha}{2}\|y\|^2 - \tfrac{\alpha}{4}\|y\|^2 = f(a) + \tfrac{\alpha}{4}\|y\|^2 > f(a) \tag{3.151}$$

if $y \neq 0$, $\|y\| \leq \delta$. Therefore a is a local minimum.
(2): Follows from (1) by replacing f by $-f$.
(3): Let A be indefinite. We need to show that in every open neighborhood of a there exist points y', y'' such that

$$f(y'') < f(a) < f(y'). \tag{3.152}$$

Since A is not negative semidefinite there exists $\xi \in \mathbb{R}^n$ such that $\alpha = \langle \xi, A\xi\rangle > 0$. Then, for $t \in \mathbb{R}$ small enough such that $|t\xi| \leq \delta$ we have

$$f(a+t\xi) = f(a) + \tfrac{1}{2}\langle t\xi, At\xi\rangle + r(t\xi) = f(a) + \tfrac{1}{2}\alpha t^2 + r(t\xi). \tag{3.153}$$

Let $\varepsilon > 0$ be such that $|r(t\xi)| \leq \frac{\alpha}{4}t^2$ for all $|t\xi| \leq \delta$ (recall that δ depends on ε). Then $f(a+t\xi) \geq f(a) + \frac{1}{4}\alpha t^2 > f(a)$. Similarly, since A is also not positive semidefinite, there exists $\eta \in \mathbb{R}^n$ such that $\langle \eta, A\eta\rangle < 0$ and for small enough t, $f(a+t\eta) < f(a)$. □

EXAMPLES 3.46. (1) Let $f(x,y) = c + x^2 + y^2$ for $c \in \mathbb{R}$. Then

$$D^2 f|_0 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} > 0 \tag{3.154}$$

and 0 is a strict local minimum of f (even a global minimum).

(2) Let $f(x,y) = c + x^2 - y^2$ for $c \in \mathbb{R}$. Then

$$D^2 f|_0 = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} \tag{3.155}$$

is indefinite and 0 is not a local extremum of f.

(3) Let $f_1(x,y) = x^2 + y^4$, $f_2(x,y) = x^2$, $f_3(x,y) = x^2 + y^3$. Then

$$D^2 f_i|_0 = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \geq 0, \tag{3.156}$$

but f_1 has a strict local minimum at 0, f_2 has a (non-strict) local minimum at 0 and f_3 has no local extremum at 0.

6. Optimization and convexity*

In applications it is often desirable to minimize a given function $f : E \to \mathbb{R}$, i.e. to find $x_* \in E$ such that $f(x_*) \leq f(x)$ for all $x \in E$. We call such a point x_* a *global minimum* of f. We say that x_* is a *strict global minimum* if $f(x_*) < f(x)$ for all $x \neq x_*$.

Example 3.47 (Linear regression). Say we are given finitely many points

$$(x_1, y_1), \ldots, (x_N, y_N) \in \mathbb{R}^n \times \mathbb{R}. \tag{3.157}$$

Suppose for instance that these represent measurements or observations of some physical system. For example, x_i could represent a point in space and y_i the corresponding air pressure measurement. We are looking to discover a "hidden relation" between the x and y coordinates. That is, we are looking for a function $F : \mathbb{R}^n \to \mathbb{R}$ such that $F(x_i)$ is (at least roughly) y_i. One way this is done is *linear regression* . Here we search only among F that take the form

$$F_{a,b}(x) = \langle x, a \rangle + b \tag{3.158}$$

with some parameters $a \in \mathbb{R}^n, b \in \mathbb{R}$. That is, we are trying to "model" the hidden relation by an affine linear function. The task is now to find the parameters a, b such that $F_{a,b}$ "fits best" to the given data set. To make this precise we introduce the error function

$$E(a,b) = \sum_{i=1}^{N} (F_{a,b}(x_i) - y_i)^2. \tag{3.159}$$

The problem of linear regression is to find the parameters (a, b) such that $E(a, b)$ is minimal.

One approach to minimizing a function $f : E \to \mathbb{R}$ is to solve the equation $\nabla f(x) = 0$, i.e. to find all critical points. By Corollary 3.43 we know that every minimum must be a critical point. However it is often difficult to solve that equation, so more practical methods are needed.

Gradient descent. Choose $x_0 \in \mathbb{R}^n$ arbitrary and let

$$x_{n+1} = x_n - \alpha_n \nabla f(x_n) \tag{3.160}$$

where $\alpha_n > 0$ is a small enough number to be determined later. The idea of this iteration is to keep moving into the direction where f decreases the fastest. Sometimes this simple process successfully converges to a minimum and sometimes it doesn't, depending on f, x_0 and α_n. What we can say from the definition is that, if $f \in C^1(E)$ and $(x_n)_{n \in \mathbb{N}}$ converges, then the limit is a critical point of f. The following lemma gives some more hope.

Lemma 3.48. *Let $f \in C^1(E)$. Then, for every $x \in E$ and small enough $\alpha > 0$,*

$$f(x - \alpha \nabla f(x)) \leq f(x). \tag{3.161}$$

PROOF. By the definition of total derivatives,

$$f(x-\alpha\nabla f(x)) = f(x)+\langle\nabla f(x), -\alpha\nabla f(x)\rangle+o(\alpha) = f(x)-\alpha\|\nabla f(x)\|^2+o(\alpha) \tag{3.162}$$

which is $\leq f(x)$ provided that $\alpha > 0$ is small enough. □

Remark. Note that the smallness of α in this lemma depends on the point x. Also, this result is not enough to prove anything about the convergence of gradient descent.

We will see that gradient descent works well if f is a convex function.

DEFINITION 3.49. Let $E \subset \mathbb{R}^n$ be convex. A function $f : E \to \mathbb{R}$ is called *convex* if

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \tag{3.163}$$

for all $x, y \in E$, $t \in [0,1]$. f is called *strictly convex* if

$$f(tx + (1-t)y) < tf(x) + (1-t)f(y) \tag{3.164}$$

for all $x \neq y \in E$ and $t \in (0,1)$.

THEOREM 3.50. *Let $E \subset \mathbb{R}^n$ be open and convex and $f \in C^1(E)$. Then f is convex if and only if*

$$f(u+v) \geq f(u) + \langle\nabla f(u), v\rangle \tag{3.165}$$

for all $u, u+v \in E$.

PROOF. $\underline{\Rightarrow:}$ Fix $u, u+v \in E$. By convexity, for $t \in [0,1]$,

$$f(u+tv) = f((1-t)u + t(u+v)) \leq (1-t)f(u) + tf(u+v). \tag{3.166}$$

By definition of the derivative,

$$f(u+tv) = f(u) + t\nabla f(u)^T v + r(t), \tag{3.167}$$

where $\lim_{t\to 0} \frac{r(t)}{t} = 0$. Thus,

$$f(u) + t\langle\nabla f(u), v\rangle + r(t) \leq (1-t)f(u) + tf(u+v) \tag{3.168}$$

which implies

$$f(u) + \langle\nabla f(u), v\rangle - f(u+v) \leq \frac{-r(t)}{t} \to 0 \quad \text{as } t \to 0. \tag{3.169}$$

Therefore $f(u) + \langle\nabla f(u), v\rangle \leq f(u+v)$.

$\underline{\Leftarrow:}$ Let $x, y \in E$, $t \in [0,1]$. Let $u = tx + (1-t)y$ and $v = x - u$. Then the assumption implies

$$f(x) \geq f(u) + \langle\nabla f(u), x-u\rangle. \tag{3.170}$$

On the other hand, letting $v = y - u$, the assumption implies

$$f(y) \geq f(u) + \langle\nabla f(u), y-u\rangle. \tag{3.171}$$

Therefore

$$tf(x)+(1-t)f(y) \geq t(f(u)+\langle\nabla f(u), x-u\rangle)+(1-t)(f(u)+\langle\nabla f(u), y-u\rangle) \tag{3.172}$$

$$= f(u)+\langle\nabla f(u), t(x-u)+(1-t)(y-u)\rangle = f(u)+\langle\nabla f(u), tx+(1-t)y-u\rangle. \tag{3.173}$$

Recalling that $u = tx + (1-t)y$, we get

$$tf(x) + (1-t)f(y) \geq f(u) = f(tx + (1-t)y). \tag{3.174}$$

□

THEOREM 3.51. *Let $E \subset \mathbb{R}^n$ be open and convex and $f \in C^2(E)$. Then*
(1) f is convex if and only if $D^2 f|_x \geq 0$ for all $x \in E$,
(2) f is strictly convex if $D^2 f|_x > 0$ for all $x \in E$.

PROOF. We only prove (1). The proof of (2) is very similar. Let f be convex. By Taylor's theorem, for $u, u+tv \in E$,

$$f(u+tv) = f(u) + t\langle \nabla f(u), v\rangle + \tfrac{1}{2}t^2 \langle D^2 f|_u v, v\rangle + o(t^2) \tag{3.175}$$

and by Theorem 3.50,

$$f(u+tv) \geq f(u) + t\langle \nabla f(u), v\rangle. \tag{3.176}$$

Combining these two pieces of information we obtain

$$\tfrac{1}{2}t^2 \langle D^2 f|_u v, v\rangle + o(t^2) \geq 0 \tag{3.177}$$

which implies $\langle D^2 f|_u v, v\rangle \geq 0$ for all $v \in \mathbb{R}^n$.
Conversely, assume that $D^2 f|_u \geq 0$ for all $u \in E$. By Taylor's theorem, for all $u, u+v \in E$ exists $\xi \in E$ such that

$$f(u+v) = f(u) + \langle \nabla f(u), v\rangle + \tfrac{1}{2}\langle D^2 f|_\xi v, v\rangle \geq f(u) + \langle \nabla f(u), v\rangle. \tag{3.178}$$

Therefore f is convex by Theorem 3.50. □

Remark. If f is strictly convex, then it does not follow that $D^2 f|_x > 0$ for all x.

EXAMPLE 3.52. Let $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^4$. Then $D^2 f|_x = f''(x) = 12x^2$ which is 0 at $x = 0$, but f is strictly convex.

THEOREM 3.53. *Let $E \subset \mathbb{R}^n$ be open and convex and $f \in C^2(E)$. Then*
(1) If f is convex, then every critical point of f is a global minimum.
(2) If f is strictly convex, then f has at most one critical point.

Remarks. 1. Convex functions may have more than one critical point. For instance, the constant function $f \equiv 0$ is convex.
2. Conclusion (1) implies that if f is convex and gradient descent converges, then it converges to a global minimum.

PROOF. (1): Let $\nabla f(x_*) = 0$. Then by Taylor's theorem, for every $x \in E$ there exists $\xi \in E$ such that

$$f(x) = f(x_*) + \underbrace{\langle \nabla f(x_*), x - x_*\rangle}_{=0} + \tfrac{1}{2}\underbrace{\langle D^2 f|_\xi (x - x_*), x - x_*\rangle}_{\geq 0} \geq f(x_*). \tag{3.179}$$

(2): Let $x_1, x_2 \in E$ be critical points of f. By (1), they are global minima. This implies $f(x_1) = f(x_2)$. If $x_1 \neq x_2$, then by strict convexity,

$$f(x_1) = \frac{f(x_1) + f(x_2)}{2} > f\Big(\frac{x_1 + x_2}{2}\Big). \tag{3.180}$$

This is a contradiction to x_1 being a global minimum. Therefore $x_1 = x_2$. □

EXAMPLE 3.54. If $\|\cdot\|$ is a norm on $\mathbb{R}^n$, then the function $x \mapsto \|x\|$ is convex:

$$\|tx + (1-t)y\| \leq t\|x\| + (1-t)\|y\| \tag{3.181}$$

by the triangle inequality. Also, this function has a unique global minimum at $x = 0$.

LEMMA 3.55. *Let $I \subset \mathbb{R}$, $E \subset \mathbb{R}^n$ be convex and suppose that*
(1) $f : E \to I$ is convex, and

(2) $g : I \to \mathbb{R}$ is convex and nondecreasing.
Then the function $h : E \to \mathbb{R}$ given by $h = g \circ f$ is convex.

PROOF. By convexity of f and since g is nondecreasing,

$$h(tx + (1-t)y) = g(f(tx + (1-t)y)) \leq g(tf(x) + (1-t)f(y)). \tag{3.182}$$

Since g is convex this is

$$\leq tg(f(x)) + (1-t)g(f(y)) = th(x) + (1-t)h(y). \tag{3.183}$$

□

COROLLARY 3.56. *If $\|\cdot\|$ is a norm on $\mathbb{R}^n$, then the function $x \mapsto \|x\|^2$ is convex.*

EXAMPLE 3.57. Recall the error function from linear regression (Example 3.47):

$$E(a,b) = \sum_{i=1}^{N} (\langle a, x_i \rangle + b - y_i)^2 \tag{3.184}$$

We claim that $E : \mathbb{R}^{n+1} \to \mathbb{R}$ is a convex function. We first rewrite $E(a,b)$ into a different form. Define a $N \times (n+1)$ matrix M and a vector $v \in \mathbb{R}^{n+1}$ by

$$M = \begin{pmatrix} x_{11} & \cdots & x_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ x_{N1} & \cdots & x_{Nn} & 1 \end{pmatrix} \in \mathbb{R}^{N \times (n+1)}, \quad v = \begin{pmatrix} a_1 \\ \vdots \\ a_n \\ b \end{pmatrix} \in \mathbb{R}^{n+1}, \tag{3.185}$$

where $x_i = (x_{i1}, \ldots, x_{in}) \in \mathbb{R}^n$ for $i = 1, \ldots, N$ and $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$. Then

$$E(a,b) = E(v) = \sum_{i=1}^{N} ((Mv)_i - y_i)^2 = \|Mv - y\|^2, \tag{3.186}$$

where $\|c\| = \left(\sum_{i=1}^{N} |c_i|^2 \right)^{1/2}$.

Let us rename variables and consider

$$E(x) = \|Mx - y\|^2 \tag{3.187}$$

for $x \in \mathbb{R}^n, M \in \mathbb{R}^{N \times n}, y \in \mathbb{R}^N$. Let $F : \mathbb{R}^N \to \mathbb{R}$ be defined by $F(y) = \|y\|^2$ and $G : \mathbb{R}^n \to \mathbb{R}^N$, $G(x) = Mx - y$. We have

$$\partial_i F(y) = 2y_i, \text{ so } DF|_y = 2y^T \in \mathbb{R}^{1 \times N}. \tag{3.188}$$

and $DG|_x = M \in \mathbb{R}^{N \times n}$. Therefore, by the chain rule we obtain

$$DE|_x = 2(Mx - y)^T M = 2(Mx)^T M - 2y^T M = 2x^T M^T M - 2y^T M \in \mathbb{R}^{1 \times n}. \tag{3.189}$$

Therefore,

$$D^2 E|_x = (\partial_i DE|_x)_{i=1,\ldots,n} = (2(M^T M)_i)_{i=1,\ldots,n} = 2M^T M. \tag{3.190}$$

Notice that $M^T M$ is positive semidefinite because

$$\langle M^T Mx, x \rangle = \langle Mx, Mx \rangle = \|Mx\|^2 \geq 0. \tag{3.191}$$

Therefore E is convex by Theorem 3.51.

EXAMPLE 3.58. Convex functions do not necessarily have a critical point. For instance the function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x$ is convex, because $D^2 f|_x = f''(x) = 0$ for all $x \in \mathbb{R}$. But $\nabla f(x) = f'(x) = 1 \neq 0$ for all $x \in \mathbb{R}$.
It is also not enough to assume strict convexity. For instance, the function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = e^x$ is strictly convex, because $f''(x) = e^x > 0$. But $f'(x) = e^x > 0$ for all $x \in \mathbb{R}$.

This motivates us to consider a stronger notion of convexity.

DEFINITION 3.59. Let $E \subset \mathbb{R}^n$ be convex and open. Let $f \in C^2(E)$. We say that f is *strongly convex* if there exists $\beta > 0$ such that

$$\langle D^2 f|_x y, y\rangle \geq \beta \|y\|^2 \tag{3.192}$$

for all $x \in E$, $y \in \mathbb{R}^n$.

Remarks. 1. f is strongly convex if and only if there exists $\beta > 0$ such that $D^2 f|_x - \beta I \geq 0$ for all $x \in E$. This follows directly from the definition using that $\beta\|y\|^2 = \langle \beta I y, y\rangle$. The condition $D^2 f|_x - \beta I \geq 0$ is equivalent to the smallest eigenvalue of $D^2 f|_x$ being $\geq \beta$. Yet another equivalent way of stating this is saying that the function $g(x) = f(x) - \frac{\beta}{2}\|x\|^2$ is convex. This is because $D^2 g|_x = D^2 f|_x - \beta I$.
2. If f is strongly convex, then f is strictly convex (by Theorem 3.51).
3. If f is strictly convex, then f is not necessarily strongly convex. For example consider $f : \mathbb{R} \to \mathbb{R}$, $f(x) = e^x$. For every $\beta > 0$ there exists $x \in \mathbb{R}$ such that $e^x < \beta$ because $e^x \to 0$ as $x \to -\infty$.

The following exercise shows that the assumption of strong convexity is not as restrictive as it may seem at first sight: strictly convex functions are strongly convex when restricted to compact sets.

EXERCISE 3.60. Suppose that $f \in C^2(\mathbb{R}^n)$ is strictly convex. Let $K \subset \mathbb{R}^n$ be compact and convex. Show that there exist $\beta_-, \beta_+ > 0$ such that

$$\beta_-\|y\|^2 \leq \langle D^2 f|_x y, y\rangle \leq \beta_+\|y\|^2 \tag{3.193}$$

for all $x \in K$ and $y \in \mathbb{R}^n$. (In particular, f is strongly convex on K.)
Hint: Consider the minimal eigenvalue of $D^2 f|_x$ as a function of x.

THEOREM 3.61. *Let $E \subset \mathbb{R}^n$ be open and convex. Let $f \in C^2(E)$. Then f is strongly convex if and only if there exists $\gamma > 0$ such that*

$$f(u+v) \geq f(u) + \langle \nabla f(u), v\rangle + \gamma\|v\|^2 \tag{3.194}$$

for every $u, u+v \in E$.

PROOF. $\underline{\Rightarrow:}$ Let $\beta > 0$ be such that $g(x) = f(x) - \frac{\beta}{2}\|x\|^2$ is convex. Then by Theorem 3.50,

$$g(u+v) \geq g(u) + \langle \nabla g(u), v\rangle = f(u) - \tfrac{\beta}{2}\|u\|^2 + \langle \nabla f(u) - \beta u, v\rangle \tag{3.195}$$

On the other hand,

$$g(u+v) = f(u+v) - \tfrac{\beta}{2}\|u+v\|^2 \tag{3.196}$$

Thus,

$$f(u+v) \geq f(u) + \langle \nabla f(u), v\rangle + \tfrac{\beta}{2}(\|u+v\|^2 - \|u\|^2 - 2\langle u, v\rangle) = f(u) + \langle \nabla f(u), v\rangle + \tfrac{\beta}{2}\|v\|^2. \tag{3.197}$$

$\underline{\Leftarrow:}$ This follows in the same way from the converse direction of Theorem 3.50. □

THEOREM 3.62. *Let $f \in C^2(\mathbb{R}^n)$ be strongly convex. Then for every $c \in \mathbb{R}$, the sublevel set*

$$B = \{x \in \mathbb{R}^n \,:\, f(x) \le c\} \tag{3.198}$$

is bounded.

PROOF. By Theorem 3.61 we have

$$f(x) \ge f(0) + \langle \nabla f(0), x\rangle + \gamma\|x\|^2. \tag{3.199}$$

Therefore, $\lim_{\|x\|\to\infty} f(x) = \infty$. Suppose that B is unbounded. Then there would exist a sequence $(x_n)_{n\ge1} \subset B$ such that $\lim_{n\to\infty} \|x_n\| = \infty$. But $f(x_n) \le c$, so $f(x_n) \not\to \infty$ as $n \to \infty$. Contradiction! □

THEOREM 3.63. *Let $f \in C^2(\mathbb{R}^n)$ be strongly convex. Then there exists a unique global minimum of f.*

PROOF. By the previous theorem, the set $B = \{x \in \mathbb{R}^n \,:\, f(x) \le f(0)\}$ is bounded. Thus, there exists $R > 0$ such that $B \subset B_R = \{x \in \mathbb{R}^n \,:\, \|x\| \le R\}$. B_R is compact, so f attains its minimum on $B_R(0)$ at some point $x_* \in B_R$. Then $f(x_*) \le f(x)$ for all $x \in B_R$. It remains to show $f(x_*) \le f(x)$ for all $x \notin B_R$. If $x \notin B_R$, then $x \notin B$, so $f(x) > f(0)$. Also, $0 \in B_R$, so $f(x_*) \le f(0) < f(x)$. □

We conclude this discussion by proving that gradient descent converges for strongly convex functions.

THEOREM 3.64. *Let $f \in C^2(\mathbb{R}^n)$ be strongly convex and $x_0 \in \mathbb{R}^n$. Define*

$$x_{n+1} = x_n - \alpha\nabla f(x_n) \quad \text{for } n \ge 0. \tag{3.200}$$

If α is small enough, then $(x_n)_{n\in\mathbb{N}}$ converges to the global minimum x_* of f.

Remark. The restriction to f defined on $\mathbb{R}^n$ is only for convenience (the same is true for Theorems 3.62 and 3.63).

LEMMA 3.65. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric and positive definite matrix. Then the matrix norm $\|A\|_{\mathrm{op}} = \sup_{x\ne0} \frac{\|Ax\|}{\|x\|}$ is equal to the largest eigenvalue of A.*
(Here $\|x\| = (\sum_{i=1}^n |x_i|^2)^{1/2}$ is the Euclidean norm.)

PROOF. Let $\{v_1, \dots, v_n\}$ be an orthonormal basis of eigenvectors corresponding to eigenvalues $\lambda_1, \dots, \lambda_n$, respectively. Then

$$\|Ax\| = \Big\| \sum_{i=1}^n x_i A v_i \Big\| = \Big\| \sum_{i=1}^n x_i \lambda_i v_i \Big\| \tag{3.201}$$

which by orthogonality is equal to $\Big(\sum_{i=1}^n |x_i|^2 \lambda_i^2 \Big)^{1/2}$ (use that $\|x\| = (\langle x, x\rangle)^{1/2}$). Thus

$$\|Ax\| = (\sum_{i=1}^n |x_i|^2\lambda_i^2)^{1/2} \le \max_{i=1,\dots,n} \lambda_i \left(\sum_{i=1}^n |x_i|^2 \right)^{1/2} = \max_{i=1,\dots,n} \lambda_i \|x\|. \tag{3.202}$$

Let $\max_{i=1,\dots,n} \lambda_i = \lambda_{i_0}$. We have shown that $\|A\| \le \lambda_{i_0}$. On the other hand,

$$\|Av_{i_0}\| = \lambda_{i_0}\|v_{i_0}\| = \lambda_{i_0}, \tag{3.203}$$

so $\|A\| = \sup_{\|x\|=1} \|Ax\| \ge \|Av_{i_0}\| = \lambda_{i_0}$. □

PROOF OF THEOREM 3.64. Let $\alpha > 0$. Define $T(x) = x - \alpha\nabla f(x)$. Then $x_{n+1} = T(x_n)$. We want T to be a contraction. For $R > 0$ define $B_R = \{x \in \mathbb{R}^n : \|x - x_*\| \leq R\}$. Let $R > 0$ be large enough such that $x_0 \in B_R$.
Claim. If α is small enough, then T is a contraction of B_R.

PROOF OF CLAIM. x_* is a global minimum of f, so $\nabla f(x_*) = 0$. Thus, $T(x_*) = x_*$. We have

$$DT|_x = I - \alpha D^2 f|_x. \tag{3.204}$$

The largest eigenvalue of $D^2 f|_x$ is a continuous function of x which is bounded on the compact set B_R. Therefore there exists $\gamma > 0$ such that

$$\langle D^2 f|_x y, y\rangle \leq \gamma\|y\|^2 \tag{3.205}$$

for all $y \in \mathbb{R}^n$ and $x \in B_R$. By strong convexity,

$$\beta\|y\|^2 \leq \langle D^2 f|_x y, y\rangle \leq \gamma\|y\|^2. \tag{3.206}$$

In other words, the eigenvalues of $D^2 f|_x$ are contained in the interval $[\beta, \gamma]$ for all $x \in B_R$. Let $\alpha \leq \frac{1}{2\gamma}$. Then the eigenvalues of $I - \alpha D^2 f|_x$ are contained in

$$[1 - \tfrac{\gamma}{2\gamma}, 1 - \tfrac{\beta}{2\gamma}] = [\tfrac{1}{2}, 1 - \tfrac{\beta}{2\gamma}] \subset (0, 1). \tag{3.207}$$

Set $c = 1 - \frac{\beta}{2\gamma}$. By Lemma 3.65, we have

$$\|I - \alpha D^2 f|_x\| \leq c < 1. \tag{3.208}$$

Therefore, $\|T(x) - T(y)\| \leq c\|x - y\|$ for all $x, y \in B_R$. It remains to show that $T(B_R) \subset B_R$. Let $x \in B_R$. Then since $T(x_*) = x_*$,

$$\|T(x) - x_*\| = \|T(x) - T(x_*)\| \leq c\|x - x_*\| \leq cR \leq R. \tag{3.209}$$

□

The claim now follows from the contraction principle (more precisely, from the same argument used to prove the Banach fixed point theorem). □

7. Further exercises

EXERCISE 3.66. Let $U \subseteq \mathbb{R}^n$ be open and convex and $f : U \to \mathbb{R}$ differentiable such that $\partial_1 f(x) = 0$ for all $x \in U$.
(i) Show that the value of $f(x)$ for $x = (x_1, \ldots, x_n) \in U$ does not depend on x_1.
(ii) Does (i) still hold if we assume that U is connected instead of convex? Give a proof or counterexample.

EXERCISE 3.67. A function $f : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ is called *homogeneous of degree* $\alpha \in \mathbb{R}$ if $f(\lambda x) = \lambda^\alpha f(x)$ for all $\lambda > 0$ and $x \in \mathbb{R}^n \setminus \{0\}$. Suppose that f is differentiable in $\mathbb{R}^n \setminus \{0\}$. Then show that f is homogeneous of degree α if and only if

$$\sum_{i=1}^n x_i \partial_i f(x) = \alpha f(x) \tag{3.210}$$

for all $x \in \mathbb{R}^n \setminus \{0\}$. *Hint:* Consider the function $g(\lambda) = f(\lambda x) - \lambda^\alpha f(x)$.

EXERCISE 3.68. Define $F : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$F(x, y) = (x^4 - y^4, e^{xy} - e^{-xy}). \tag{3.211}$$

(i) Compute the Jacobian of F.
(ii) Let $p_0 \in \mathbb{R}^2$ and $p_0 \neq (0, 0)$. Show that there exist open neighborhoods $U, V \subset \mathbb{R}^2$ of p_0 and $F(p_0)$, respectively and a function $G : V \to U$ such that $G(F(p)) = p$ for all $p \in U$ and $F(G(p)) = p$ for all $p \in V$.
(iii) Compute $DG\big|_{F(p_0)}$.
(iv) Is F a bijective map?

EXERCISE 3.69. Let $a \in \mathbb{R}$, $a \neq 0$ and $E = \{(x, y, z) \in \mathbb{R}^3 : a + x + y + z \neq 0\}$ and $f : E \to \mathbb{R}^3$ defined by

$$f(x, y, z) = \left(\frac{x}{a + x + y + z}, \frac{y}{a + x + y + z}, \frac{z}{a + x + y + z} \right). \tag{3.212}$$

(i) Compute the Jacobian determinant of f (that is, the determinant of the Jacobian matrix).
(ii) Show that f is one-to-one and compute its inverse f^{-1}.

EXERCISE 3.70. Prove that there exists $\delta > 0$ such that for all square matrices $A \in \mathbb{R}^{n\times n}$ with $\|A - I\| < \delta$ (where I denotes the identity matrix) there exists $B \in \mathbb{R}^{n\times n}$ such that $B^2 = A$.

EXERCISE 3.71. Look at each of the following as an equation to be solved for $x \in \mathbb{R}$ in terms of parameter $y, z \in \mathbb{R}$. Notice that $(x, y, z) = (0, 0, 0)$ is a solution for each of these equations. For each one, prove that it can be solved for x as a C^1-function of y, z in a neighborhood of $(0, 0, 0)$.

(a) $\cos(x)^2 - e^{\sin(xy)^3 + x} = z^2$
(b) $(x^2 + y^3 + z^4)^2 = \sin(x - y + z)$
(c) $x^7 + ye^z x^3 - x^2 + x = \log(1 + y^2 + z^2)$

EXERCISE 3.72. Let $(t_0, y_0) \in \mathbb{R}^2$, $c \in \mathbb{R}$ and define $Y_0(t) = y_0$,

$$Y_n(t) = y_0 + c \int_{t_0}^{t} sY_{n-1}(s)ds. \tag{3.213}$$

Compute $Y_n(t)$ and $Y(t) = \lim_{n\to\infty} Y_n(t)$. Which initial value problem does Y solve?

EXERCISE 3.73. Consider the initial value problem

$$\begin{cases} y'(t) = e^{y(t)^2} - \frac{1}{ty(t)}, \\ y(1) = 1. \end{cases} \tag{3.214}$$

Find an interval $I = (1 - h, 1 + h)$ such that this problem has a unique solution y in I. Give an explicit estimate for h (it does not need to be best possible).

EXERCISE 3.74. Consider the initial value problem

$$\begin{cases} y'(t) = t + \sin(y(t)), \\ y(2) = 1. \end{cases} \tag{3.215}$$

Find the largest interval $I \subseteq \mathbb{R}$ containing $t_0 = 2$ such that the problem has a unique solutions y in I.

EXERCISE 3.75. Let F be a smooth function on $\mathbb{R}^2$ (i.e. partial derivatives of all orders exist everywhere and are continuous) and suppose that the initial value problem $y' = F(t, y)$, $y(t_0) = y_0$ has a unique solution y on the interval $I = [t_0, t_0 + a]$ with y smooth on I. Let $h > 0$ be sufficiently small and define $t_k = t_0 + kh$ for integers $0 \leq k \leq a/h$.
Define a function y_h recursively by setting $y_h(t_0) = y_0$ and

$$y_h(t) = y_h(t_k) + (t - t_k)F(t_k, y_h(t_k)) \tag{3.216}$$

for $t \in (t_k, t_{k+1}]$ for integers $0 \leq k \leq a/h$.
(i) From the proof of Peano's theorem (Theorem 3.29) it follows that $y_h \to y$ uniformly on I as $h \to 0$. Prove the following stronger statement: there exists a constant $C > 0$ such that for all $t \in I$ and $h > 0$ sufficiently small,

$$|y(t) - y_h(t)| \leq Ch. \tag{3.217}$$

Hint: The left hand side is zero if $t = t_0$. Use Taylor expansion to study how the error changes as t increases from t_k to t_{k+1}.
(ii) Let $F(t, y) = \lambda y$ with $\lambda \in \mathbb{R}$ a parameter. Explicitly determine y, y_h and a value for C in (i).

EXERCISE 3.76. Let us improve the approximation from Exercise 3.75. In the context of that exercise, define a piecewise linear function y_h^* recursively by setting $y_h^*(t_0) = y_0$ and

$$y_h^*(t) = y_h^*(t_k) + (t - t_k)G(t_k, y_h^*(t_k), h), \tag{3.218}$$

for $t \in (t_k, t_{k+1}]$ for integers $0 \leq k \leq a/h$, where

$$G(t, y, h) = \tfrac{1}{2}(F(t, y) + F(t + h, y + hF(t, y))). \tag{3.219}$$

Prove that there exists a constant $C > 0$ such that for all $t \in I$ and $h > 0$ sufficiently small,

$$|y(t) - y_h^*(t)| \leq Ch^2. \tag{3.220}$$

EXERCISE 3.77. For a function $f : [a, b] \to \mathbb{R}$ define

$$\mathcal{I}(f) = \int_a^b (1 + f'(t)^2)^{1/2} dt. \tag{3.221}$$

Let $\mathcal{A} = \{f \in C^2([a, b]) : f(a) = c, f(b) = d\}$. Determine $f_* \in \mathcal{A}$ such that

$$\mathcal{I}(f_*) = \inf_{f \in \mathcal{A}} \mathcal{I}(f). \tag{3.222}$$

What is the geometric meaning of $\mathcal{I}(f)$ and $\inf_{f \in \mathcal{A}} \mathcal{I}(f)$?

EXERCISE 3.78. Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be smooth functions (that is, all partial derivatives exist to arbitrary orders and are continuous). Show that for all multiindices $\alpha \in \mathbb{N}_0^n$,

$$\partial^\alpha (f \cdot g)(x) = \sum_{\beta \in \mathbb{N}_0^n : \beta \leq \alpha} \binom{\alpha}{\beta} \partial^\beta f(x) \partial^{\alpha - \beta} g(x) \tag{3.223}$$

for all $x \in \mathbb{R}^n$, where $\binom{\alpha}{\beta} = \frac{\alpha!}{\beta!(\alpha - \beta)!} = \frac{\alpha_1! \cdots \alpha_n!}{\beta_1! \cdots \beta_n! (\alpha_1 - \beta_1)! \cdots (\alpha_n - \beta_n)!}$.

EXERCISE 3.79. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be such that $\partial_1 \partial_2 f$ exists everywhere. Does it follow that $\partial_1 f$ exists? Give a proof or counterexample.

EXERCISE 3.80. Determine the Taylor expansion of the function

$$f : (0,\infty) \times (0,\infty) \to \mathbb{R},\ f(x,y) = \frac{x-y}{x+y} \tag{3.224}$$

at the point $(x,y) = (1,1)$ up to order 2.

EXERCISE 3.81. Show that every continuous function $f : [a,b] \to [a,b]$ has a fixed point.

EXERCISE 3.82. Let X be a real Banach space. Let $B = \{x \in X : \|x\| \leq 1\}$ and $\partial B = \{x \in X : \|x\| = 1\}$. Show that the following are equivalent:
(i) every continuous map $f : B \to B$ has a fixed point
(ii) there exists no continuous map $r : B \to \partial B$ such that $r(b) = b$ for all $b \in \partial B$.

EXERCISE 3.83. Determine the local minima and maxima of the function

$$f : \mathbb{R}^2 \to \mathbb{R},\ \ f(x,y) = (4x^2 + y^2)e^{-x^2-4y^2}. \tag{3.225}$$

EXERCISE 3.84. Let $E \subset \mathbb{R}^n$ be open, $f : E \to \mathbb{R}$ and $x \in E$. Assume that for y in a neighborhood of 0 we have

$$f(x+y) = \sum_{|\alpha| \leq k} c_\alpha y^\alpha + o(\|y\|^k) \tag{3.226}$$

as $y \to 0$ and

$$f(x+y) = \sum_{|\alpha| \leq k} \widetilde{c}_\alpha y^\alpha + o(\|y\|^k) \tag{3.227}$$

as $y \to 0$. Show that $c_\alpha = \widetilde{c}_\alpha$ for all $|\alpha| \leq k$.

EXERCISE 3.85. Let $D = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$. Determine the maximum and minimum values of the function $f : D \to \mathbb{R}$, $f(x,y) = 4x^2 - 3xy$.

EXERCISE 3.86. Let $f \in C^2(\mathbb{R}^n)$ and suppose that the Hessian of f is positive definite at every point. Show that $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ is an injective map.

EXERCISE 3.87. Let $f \in C^2(\mathbb{R}^n)$ be strongly convex. Show that $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ is a diffeomorphism (that is, show that it is differentiable, bijective and that its inverse is differentiable).

EXERCISE 3.88. Let $f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + c$ with $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n, c \in \mathbb{R}$. Assume that A is symmetric and positive definite. Show that f has a unique global minimum at some point x_* and determine $f(x_*)$ in terms of A, b, c.

EXERCISE 3.89. Prove that the point x_* from Exercise 3.88 can be computed using gradient descent: that is, if $x_0 \in \mathbb{R}^n$ arbitrary and

$$x_{n+1} = x_n - \alpha \nabla f(x_n) \tag{3.228}$$

for $n = 0, 1, 2, \ldots$, then the sequence $(x_n)_{n\in\mathbb{N}}$ converges to x_* for all starting points $x_0 \in \mathbb{R}^n$, provided that α is chosen sufficiently small.

EXERCISE 3.90. Let $\mathcal{D} \subset \mathbb{R}^2$ be a finite set. Define a function $E : \mathbb{R}^3 \to \mathbb{R}$ by

$$E(a,b,c) = \sum_{x\in\mathcal{D}} (ax_1^2 + bx_1 + c - x_2)^2. \tag{3.229}$$

(1) Show that E is convex.

(2) Does there exist a set $\mathcal{D}$ such that E is strongly convex? Proof or counterexample.

EXERCISE 3.91. (a) Find a convex function that is not bounded from below.
(b) Find a strictly convex function that is not bounded from below.
(c) If a function is strictly convex and bounded from below, does it necessarily have a critical point? (Proof or counterexample.)

EXERCISE 3.92. (a) Give an example of a convex function that is not continuous.
(b) Let $f : (a, b) \to \mathbb{R}$. Show that if f is convex, then f is continuous.

EXERCISE 3.93. Construct a strictly convex function $f : \mathbb{R} \to \mathbb{R}$ such that f is not differentiable at x for every $x \in \mathbb{Q}$.

EXERCISE 3.94. Let $f \in C^2(\mathbb{R}^n)$. Recall that we defined f to be *strongly convex* if there exists $\beta > 0$ such that $\langle D^2 f|_x y, y\rangle \geq \beta\|y\|^2$ for every $x, y \in \mathbb{R}^n$. Show that f is strongly convex if and only if there exists $\gamma > 0$ such that

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - \gamma t(1-t)\|x-y\|^2 \tag{3.230}$$

for all $x, y \in \mathbb{R}^n, t \in [0, 1]$.
(Consequently, that condition can serve as an alternative definition of strong convexity, which is also valid if f is not C^2.)

EXERCISE 3.95. (Recall Exercise 4.82 as motivation for this exercise.) Fix a function $\sigma \in C^1(\mathbb{R})$ and define for $x \in \mathbb{R}^n, W \in \mathbb{R}^{m\times n}, v \in \mathbb{R}^m$,

$$\mu(x, W, v) = \sum_{i=1}^{m} \sigma((Wx)_i)v_i \tag{3.231}$$

Given a finite set of points $\mathcal{D} = \{(x_1, y_1), \ldots, (x_N, y_N) \in \mathbb{R}^n \times \mathbb{R}\}$ define

$$E(W, v) = \sum_{j=1}^{N} (\mu(x_i, W, v) - y_i)^2. \tag{3.232}$$

Is E necessarily convex? (Proof or counterexample.)

CHAPTER 4

Approximation of functions

In this section we want to study different ways to approximate continuous functions.

Let X be a normed vector space of functions (say, continuous functions on $[0,1]$) and $A \subset X$ some subspace of it (say, polynomials). Let $f \in X$ be arbitrary. Our goal is to 'approximate' the function f by functions g in A. We measure the quality of approximation by the error in norm, i.e. $\|f-g\|$.
The most basic question in this context is:

Can we make $\|f-g\|$ arbitrarily small?

More precisely, we are asking if A is *dense* in X. Recall that $A \subset X$ is called *dense* if $\overline{A} = X$. That is, if for every $f \in X$ and every $\varepsilon > 0$ there exists a $g \in A$ such that $\|f-g\| \leq \varepsilon$.

1. Polynomial approximation

THEOREM 4.1 (Weierstrass). *For every continuous function f on $[a,b]$ there exists a sequence of polynomials that converges uniformly to f.*

In other words, the theorem says that the set $A = \{p : p \text{ polynomial}\}$ is dense in $C([a,b])$.

There are many proofs of this theorem in the literature. We present a proof using *Bernstein polynomials* . Without loss of generality we consider only the interval $[a,b] = [0,1]$ (why are we allowed to do that?).

Let f be continuous on $[0,1]$. Define for $n = 1, 2, \ldots$:

$$B_n f(t) = \sum_{k=0}^{n} f\Big(\frac{k}{n}\Big)\binom{n}{k} t^k (1-t)^{n-k}. \tag{4.1}$$

$B_n f$ is a polynomial of degree n. We will show that $B_n f \to f$ uniformly on $[0,1]$. By the binomial theorem,

$$1 = (t+1-t)^n = \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k}. \tag{4.2}$$

Thus,

$$B_n f(t) - f(t) = \sum_{k=0}^{n} (f(k/n) - f(t))\binom{n}{k} t^k (1-t)^{n-k}. \tag{4.3}$$

Let $\varepsilon > 0$. By uniform continuity of f we choose $\delta > 0$ be such that $|f(t) - f(s)| \leq \varepsilon/2$ for all $t, s \in [0,1]$ with $|t-s| \leq \delta$. Now we write the sum on the right hand side of

(4.3) as I + II, where

$$\mathrm{I} = \sum_{\substack{k=0,\\ |\frac{k}{n}-t|<\delta}}^{n} (f(k/n) - f(t))\binom{n}{k} t^k (1-t)^{n-k}, \tag{4.4}$$

$$\mathrm{II} = \sum_{\substack{k=0,\\ |\frac{k}{n}-t|\geq\delta}}^{n} (f(k/n) - f(t))\binom{n}{k} t^k (1-t)^{n-k}. \tag{4.5}$$

We estimate I and II separately. For I we have from uniform continuity that

$$|\mathrm{I}| \leq \varepsilon/2 \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k} = \varepsilon/2. \tag{4.6}$$

To estimate II we first compute the Bernstein polynomials for the monomials $1, t, t^2$.

LEMMA 4.2. *Let $g_m(t) = t^m$. Then*

$$B_n g_0(t) = 1 \tag{4.7}$$

$$B_n g_1(t) = t \tag{4.8}$$

$$B_n g_2(t) = t^2 + \frac{t - t^2}{n} \quad \textit{for } n \geq 2 \tag{4.9}$$

PROOF. We have

$$B_n g_0(t) = \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k} = (t + (1-t))^n = 1 \tag{4.10}$$

by the binomial theorem. Next,

$$\begin{aligned} B_n g_1(t) &= \sum_{k=0}^{n} \frac{k}{n}\binom{n}{k} t^k (1-t)^{n-k} = \sum_{k=1}^{n} \binom{n-1}{k-1} t^k (1-t)^{n-k} \\ &= t \sum_{k=0}^{n-1} \binom{n-1}{k} t^k (1-t)^{(n-1)-k} = t(t + (1-t))^{n-1} = t. \end{aligned} \tag{4.11}$$

To compute $B_n g_2$ we use that

$$\begin{aligned} \frac{k^2}{n^2}\binom{n}{k} = \frac{k}{n}\binom{n-1}{k-1} &= \frac{n-1}{n}\frac{k-1}{n-1}\binom{n-1}{k-1} + \frac{1}{n}\binom{n-1}{k-1} \\ &= \frac{n-1}{n}\binom{n-2}{k-2} + \frac{1}{n}\binom{n-1}{k-1}. \end{aligned} \tag{4.12}$$

Thus,

$$\begin{aligned} B_n g_2(t) &= \frac{n-1}{n} \sum_{k=2}^{n} \binom{n-2}{k-2} t^k (1-t)^{n-k} + \frac{1}{n} \sum_{k=1}^{n} \binom{n-1}{k-1} t^k (1-t)^{n-k} \\ &= \frac{n-1}{n} t^2 + \frac{1}{n} t = t^2 + \frac{t - t^2}{n}. \end{aligned} \tag{4.13}$$

□

As a consequence, we obtain the following:

LEMMA 4.3. *For all* $t \in [0,1]$,

$$\sum_{k=0}^{n} (\tfrac{k}{n} - t)^2 \binom{n}{k} t^k (1-t)^{n-k} \leq \frac{1}{n}. \tag{4.14}$$

PROOF. From the previous lemma,

$$\sum_{k=0}^{n} (\tfrac{k}{n} - t)^2 \binom{n}{k} t^k (1-t)^{n-k} = B_n g_2(t) - 2tB_n g_1(t) + t^2 B_n g_0(t) \tag{4.15}$$

$$= t^2 + \frac{t-t^2}{n} - 2t^2 + t^2 = \frac{t-t^2}{n}.$$

Since $t \in [0,1]$ we have $0 \leq t - t^2 = t(1-t) \leq 1$. □

Now we are ready to estimate II. First note that f is bounded, so there exists $c > 0$ such that $|f(x)| \leq c$ for all $x \in [0,1]$. Choose $N \in \mathbb{N}$ such that $2c\delta^{-2}N^{-1} \leq \varepsilon/2$. Then for all $n \geq N$,

$$|\mathrm{II}| \leq 2c \sum_{\substack{k=0,\\ |\frac{k}{n}-t| \geq \delta}}^{n} \binom{n}{k} t^k (1-t)^{n-k} \leq 2c\delta^{-2} \sum_{k=0}^{n} (\tfrac{k}{n} - t)^2 \binom{n}{k} t^k (1-t)^{n-k} \tag{4.16}$$

$$\leq 2c\delta^{-2}N^{-1} \leq \varepsilon/2.$$

In the second inequality we have used that $\delta^{-2}|\frac{k}{n} - t|^2 \leq 1$. Thus if $n \geq N$ and $t \in [0,1]$, then

$$|B_n f(t) - f(t)| \leq |\mathrm{I}| + |\mathrm{II}| \leq \varepsilon/2 + \varepsilon/2 = \varepsilon. \tag{4.17}$$

This concludes the proof of Weierstrass' theorem.

2. Orthonormal systems

In the previous section we studied approximation of continuous functions in the supremum norm, $\|f\|_\infty = \sup_{x \in [a,b]} |f(x)|$. In this section we turn our attention to another important norm, the L^2 norm.

DEFINITION 4.4. For two piecewise continuous functions f, g on an interval $[a,b]$ we define their *inner product* by

$$\langle f, g \rangle = \int_a^b f(x)\overline{g(x)}dx. \tag{4.18}$$

If $\langle f, g \rangle = 0$ we say that f and g are *orthogonal* . We define the L^2*-norm* of f by

$$\|f\|_2 = \left(\int_a^b |f(x)|^2 dx \right)^{1/2}. \tag{4.19}$$

If $\|f\|_2 = 1$ then we say that f is L^2*-normalized* .

Note: Some comments are in order regarding the term 'piecewise continuous'. For our purposes we call a function f, defined on an interval $[a,b]$, *piecewise continuous* if $\lim_{x \to x_0} f(x)$ exists at every point x_0 and is different from $f(x_0)$ at at most finitely many points. We denote this class of functions by $\mathrm{pc}([a,b])$. Piecewise continuous functions are Riemann integrable.

The inner product has the following properties (for functions f, g, h and $\lambda \in \mathbb{C}$):

- Sesquilinearity:

$$\langle f+\lambda g, h\rangle = \langle f, h\rangle + \lambda\langle g, h\rangle, \tag{4.20}$$

$$\langle h, f+\lambda g\rangle = \langle h, f\rangle + \overline{\lambda}\langle h, g\rangle. \tag{4.21}$$

- Antisymmetry: $\langle f, g\rangle = \overline{\langle g, f\rangle}$
- Positivity: $\langle f, f\rangle \geq 0$ (and > 0 unless f is zero except at possibly finitely many points)

THEOREM 4.5 (Cauchy-Schwarz inequality). *For two piecewise continuous functions f, g we have*

$$|\langle f, g\rangle| \leq \|f\|_2 \|g\|_2. \tag{4.22}$$

PROOF. For nonnegative real numbers x and y we have the elementary inequality

$$xy \leq \frac{x^2}{2} + \frac{y^2}{2}. \tag{4.23}$$

Thus we have

$$|\langle f, g\rangle| \leq \int_a^b |f(x)g(x)|dx \leq \tfrac{1}{2}\int_a^b |f(x)|^2 dx + \tfrac{1}{2}\int_a^b |g(x)|^2 dx. = \tfrac{1}{2}\langle f, f\rangle + \tfrac{1}{2}\langle g, g\rangle. \tag{4.24}$$

Now we note that for every $\lambda > 0$, replacing f by λf and g by $\lambda^{-1} g$ does not change the left hand side of this inequality. Thus we have for every $\lambda > 0$ that

$$|\langle f, g\rangle| \leq \tfrac{\lambda^2}{2}\langle f, f\rangle + \tfrac{1}{2\lambda^2}\langle g, g\rangle. \tag{4.25}$$

Now we choose λ so that this inequality is as strong as possible: $\lambda^2 = \sqrt{\frac{\langle g,g\rangle}{\langle f,f\rangle}}$ (we may assume that $\langle f, f\rangle \neq 0$ because otherwise there is nothing to show). Then

$$|\langle f, g\rangle| \leq \sqrt{\langle f, f\rangle}\sqrt{\langle g, g\rangle}. \tag{4.26}$$

Note that one can arrive at this definition of λ in a systematic way: treat the right hand side of (4.25) as a function of λ and minimize it using calculus. □

COROLLARY 4.6 (Minkowski's inequality). *For two functions $f, g \in \mathrm{pc}([a, b])$,*

$$\|f+g\|_2 \leq \|f\|_2 + \|g\|_2. \tag{4.27}$$

PROOF. We may assume $\|f+g\|_2 \neq 0$ because otherwise there is nothing to prove. Then

$$\|f+g\|_2^2 = \int_a^b |f+g|^2 \leq \int_a^b |f+g||f| + \int_a^b |f+g||g| \tag{4.28}$$

$$\leq \|f+g\|_2\|f\|_2 + \|f+g\|_2\|g\|_2 = \|f+g\|_2(\|f\|_2 + \|g\|_2). \tag{4.29}$$

Dividing by $\|f+g\|_2$ we obtain $\|f+g\|_2 \leq \|f\|_2 + \|g\|_2$. □

This is the triangle inequality for $\|\cdot\|_2$. This makes $d(f, g) = \|f-g\|_2$ a metric on say, the set of continuous functions. Unfortunately, the resulting metric space is not complete. (Its *completion* is a space called $L^2([a, b])$, see Exercise 4.70.)

DEFINITION 4.7. A sequence $(\phi_n)_n$ of piecewise continuous functions on $[a, b]$ is called an *orthonormal system on* $[a, b]$ if

$$\langle \phi_n, \phi_m\rangle = \int_a^b \phi_n(x)\overline{\phi_m(x)}dx = \begin{cases} 0, & \text{if } n \neq m, \\ 1, & \text{if } n = m. \end{cases} \tag{4.30}$$

(The index n may run over the natural numbers, or the integers, a finite set of integers, or more generally any countable set. We will write $\sum_n$ to denote a sum over all the indices. In proofs we will always adopt the interpretation that the index n runs over $1, 2, 3, \ldots$. This is no loss of generality.)

Notation: For a set A we denote by $\mathbb{1}_A$ the *characteristic function of* A . This is the function such that $\mathbb{1}_A(x) = 1$ when $x \in A$ and $\mathbb{1}_A(x) = 0$ when $x \notin A$.

EXAMPLE 4.8 (Disjoint support). Let $\phi_n(x) = \mathbb{1}_{[n,n+1)}$ and $N \in \mathbb{N}$. Then $(\phi_n)_{n=0,\ldots,N-1}$ is an orthonormal system on $[0, N]$.

EXAMPLES 4.9 (Trigonometric functions). The following are orthonormal systems on $[0, 1]$:
1. $\phi_n(x) = e^{2\pi i n x}$
2. $\phi_n(x) = \sqrt{2}\cos(2\pi n x)$
3. $\phi_n(x) = \sqrt{2}\sin(2\pi n x)$

EXERCISE 4.10 (Rademacher functions). For $n = 0, 1, \ldots$ and $x \in [0, 1]$ we define $r_n(x) = \operatorname{sgn}(\sin(2^n \pi x))$. Show that $(r_n)_{n\in\mathbb{N}}$ is an orthonormal system on $[0, 1]$.

Let $(\phi_n)_n$ be an orthonormal system and let f be a finite linear combination of the functions $(\phi_n)_n$. Say,

$$f(x) = \sum_{n=1}^{N} c_n \phi_n(x). \tag{4.31}$$

Then there is an easy way to compute the coefficients c_n:

$$c_n = \langle f, \phi_n \rangle = \int_a^b f(x)\overline{\phi_n(x)}dx. \tag{4.32}$$

To prove this we multiply (4.31) by $\overline{\phi_m(x)}$ and integrate over x:

$$\int_a^b f(x)\overline{\phi_m(x)}dx = \sum_{n=1}^{N} c_n \int_a^b \phi_n(x)\overline{\phi_m(x)}dx = \sum_{n=1}^{N} c_n \langle \phi_n, \phi_m \rangle = c_m. \tag{4.33}$$

Notice that the formula $c_n = \langle f, \phi_n \rangle$ still makes sense if f is not of the form (4.31).

THEOREM 4.11. *Let $(\phi_n)_n$ be an orthonormal system on $[a, b]$. Let f be a piecewise continuous function. Consider*

$$s_N(x) = \sum_{n=1}^{N} \langle f, \phi_n \rangle \phi_n(x). \tag{4.34}$$

Denote the linear span of the functions $(\phi_n)_{n=1,\ldots,N}$ by X_N. Then

$$\|f - s_N\|_2 \leq \|f - g\|_2 \tag{4.35}$$

holds for all $g \in X_N$ with equality if and only if $g = s_N$.

In other words, the theorem says that among all functions of the form $\sum_{n=1}^{N} c_n \phi_n(x)$, the function s_N defined by the coefficients $c_n = \langle f, \phi_n \rangle$ is the best "L^2-approximation" to f in the sense that (4.35) holds.

This can be interpreted geometrically: the function s_N is the *orthogonal projection* of f onto the subspace X_N. As in Euclidean space, the orthogonal projection is characterized by being the point in X_N that is closest to f and it is uniquely determined by this property (see Figure 1).

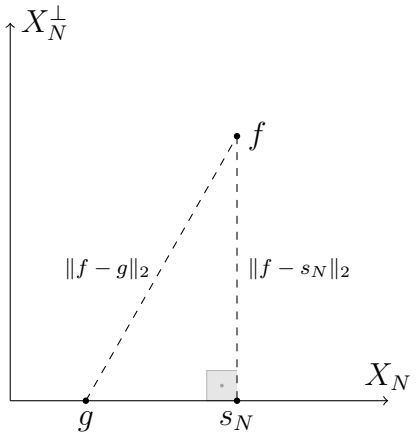

FIGURE 1. s_N is the orthogonal projection of f onto X_N.

THEOREM 4.12 (Bessel's inequality). *If $(\phi_n)_n$ is an orthonormal system on $[a,b]$ and f a piecewise continuous function on $[a,b]$ then*

$$\sum_n |\langle f, \phi_n\rangle|^2 \leq \|f\|_2^2. \tag{4.36}$$

COROLLARY 4.13 (Riemann-Lebesgue lemma). *Let $(\phi_n)_{n=1,2,\dots}$ be an orthonormal system and f a piecewise continuous function. Then*

$$\lim_{n\to\infty} \langle f, \phi_n\rangle = 0. \tag{4.37}$$

This follows because the series $\sum_{n=1}^\infty |\langle f, \phi_n\rangle|^2$ converges as a consequence of Bessel's inequality.

DEFINITION 4.14. An orthonormal system $(\phi_n)_n$ is called *complete* if

$$\sum_n |\langle f, \phi_n\rangle|^2 = \|f\|_2^2 \tag{4.38}$$

for all f.

THEOREM 4.15. *Let $(\phi_n)_n$ be an orthonormal system on $[a,b]$. Let $(s_N)_N$ be as in Theorem 4.11. Then $(\phi_n)_n$ is complete if and only if $(s_N)_N$ converges to f in the L^2-norm (that is, $\lim_{N\to\infty} \|f - s_N\|_2 = 0$) for every piecewise continuous f on $[a,b]$.*

We will later see that the orthonormal system $\phi_n(x) = e^{2\pi i n x}$ ($n \in \mathbb{Z}$) on $[0,1]$ is complete.

PROOF OF THEOREM 4.11. Let $g \in X_N$ and write

$$g(x) = \sum_{n=1}^N b_n \phi_n(x). \tag{4.39}$$

Let us also write

$$c_n = \langle f, \phi_n\rangle. \tag{4.40}$$

We have

$$\langle f, g\rangle = \sum_{n=1}^N \overline{b_n}\langle f, \phi_n\rangle = \sum_{n=1}^N c_n \overline{b_n}. \tag{4.41}$$

Using that $(\phi_n)_n$ is orthonormal we get

$$\langle g, g\rangle = \Big\langle \sum_{n=1}^{N} b_n\phi_n, \sum_{m=1}^{N} b_m\phi_m \Big\rangle = \sum_{n=1}^{N}\sum_{m=1}^{N} b_n\overline{b_m}\langle\phi_n,\phi_m\rangle = \sum_{n=1}^{N}|b_n|^2. \tag{4.42}$$

Thus,

$$\langle f-g, f-g\rangle = \langle f,f\rangle - \langle f,g\rangle - \langle g,f\rangle + \langle g,g\rangle \tag{4.43}$$

$$= \langle f,f\rangle - \sum_{n=1}^{N} c_n\overline{b_n} - \sum_{n=1}^{N}\overline{c_n}b_n + \sum_{n=1}^{N}|b_n|^2 \tag{4.44}$$

$$= \langle f,f\rangle - \sum_{n=1}^{N}|c_n|^2 + \sum_{n=1}^{N}|b_n-c_n|^2 \tag{4.45}$$

We have

$$\begin{aligned}\langle f-s_N, f-s_N\rangle &= \langle f,f\rangle - \langle f,s_N\rangle - \langle s_N,f\rangle + \langle s_N,s_N\rangle \\ &= \langle f,f\rangle - 2\sum_{n=1}^{N}|c_n|^2 + \sum_{n=1}^{N}|c_n|^2 = \langle f,f\rangle - \sum_{n=1}^{N}|c_n|^2.\end{aligned} \tag{4.46}$$

Thus we have shown

$$\langle f-g, f-g\rangle = \langle f-s_N, f-s_N\rangle + \sum_{n=1}^{N}|b_n-c_n|^2 \tag{4.47}$$

which implies the claim since $\sum_{n=1}^{N}|b_n-c_n|^2 \geq 0$ with equality if and only if $b_n = c_n$ for all $n = 1,\dots,N$. □

Proof of Theorem 4.12. From the calculation in (4.46),

$$\langle f,f\rangle - \sum_{n=1}^{N}|c_n|^2 = \langle f-s_N, f-s_N\rangle \geq 0, \tag{4.48}$$

so $\sum_{n=1}^{N}|c_n|^2 \leq \|f\|_2^2$ for all N. Letting $N\to\infty$ this proves the claim (in particular, the series $\sum_{n=1}^{\infty}|c_n|^2$ converges). □

Proof of Theorem 4.15. From (4.46),

$$\|f-s_N\|_2^2 = \langle f,f\rangle - \sum_{n=1}^{N}|\langle f,\phi_n\rangle|^2 \tag{4.49}$$

This converges to 0 as $N\to\infty$ if and only if $(\phi_n)_n$ is complete. □

3. The Haar system

In this section we discuss an important example of an orthonormal system on $[0,1]$.

Definition 4.16 (Dyadic intervals). For non-negative integers j,k with $0 \leq j < 2^k$ we define

$$I_{k,j} = [2^{-k}j, 2^{-k}(j+1)) \subset [0,1]. \tag{4.50}$$

The interval $I_{k,j}$ is called a *dyadic interval* and k is called its *generation* . We denote by $\mathcal{D}_k$ the set of all dyadic intervals of generation k and by $\mathcal{D} = \bigcup_{k\geq 0}\mathcal{D}_k$ the set of all dyadic intervals on $[0,1]$.

DEFINITION 4.17. Each dyadic interval $I \in \mathcal{D}$ with $|I| = 2^{-k}$ can be split in the middle into its *left child* and *right child* , which are again dyadic intervals that we denote by I_ℓ and I_r, respectively.

EXAMPLE 4.18. The interval $I = [\frac{1}{2}, \frac{1}{2} + \frac{1}{4})$ is a dyadic interval and its left and right children are given by $I_\ell = [\frac{1}{2}, \frac{1}{2} + \frac{1}{8})$ and $I_r = [\frac{1}{2} + \frac{1}{8}, \frac{1}{2} + \frac{1}{4})$.

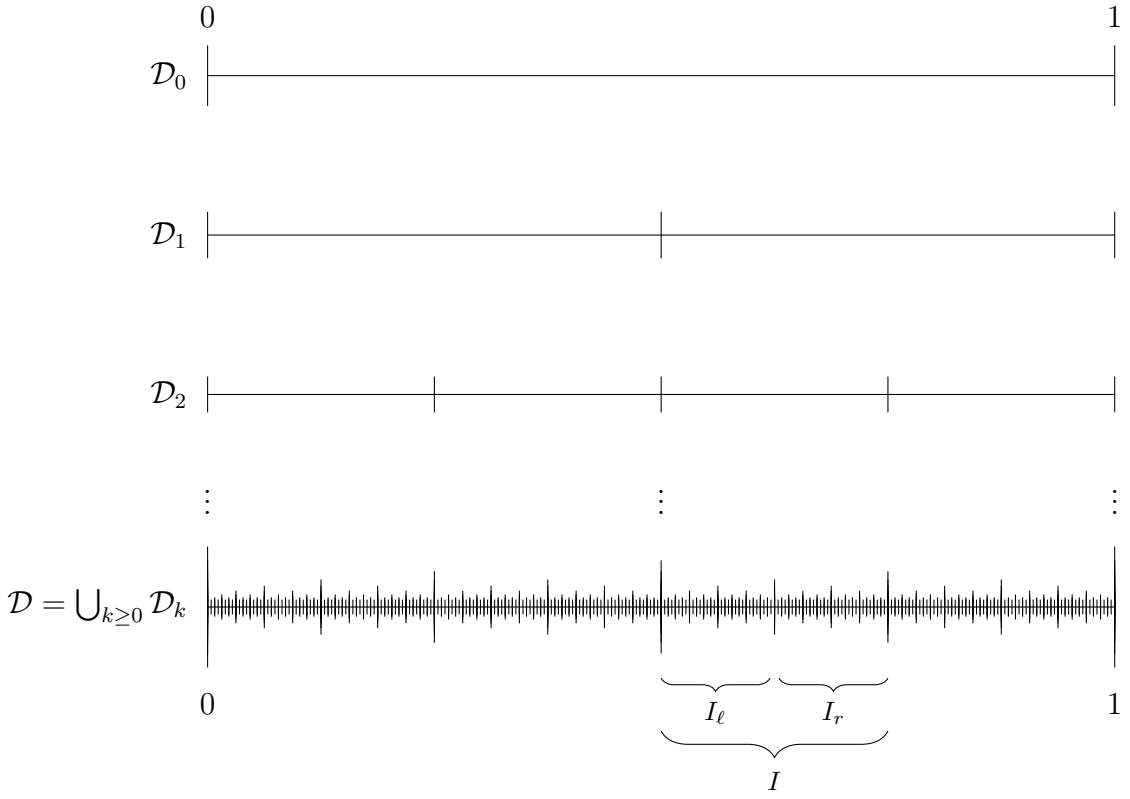

FIGURE 2. Dyadic intervals.

LEMMA 4.19. *(1) Two dyadic intervals are either disjoint or contained in each other. That is, for every $I, J \in \mathcal{D}$ at least one of the following is true: $I \cap J = \emptyset$ or $I \subset J$ or $J \subset I$.*

(2) For every $k \geq 0$ the dyadic intervals of generation k are a partition of $[0, 1)$. That is,

$$[0, 1) = \bigcup_{I \in \mathcal{D}_k} I. \tag{4.51}$$

EXERCISE 4.20. Prove this lemma.

EXERCISE 4.21. Let $J \subset [0, 1]$ be any interval. Show that there exists $I \in \mathcal{D}$ such that $|I| \leq |J|$ and $3I \supset J$. (Here $3I$ denotes the interval with three times the length of I and the same center as I.)

DEFINITION 4.22. For each $I \in \mathcal{D}$ we define the *Haar function* associated with it by

$$\psi_I = |I|^{-1/2}(\mathbb{1}_{I_\ell} - \mathbb{1}_{I_r}) \tag{4.52}$$

The countable set of functions given by

$$\mathcal{H} = \{\mathbb{1}_{[0,1]}\} \cup \{\psi_I \,:\, I \in \mathcal{D}\} \tag{4.53}$$

is called the *Haar system* on $[0,1]$.

EXAMPLE 4.23. The Haar function associated with the dyadic interval $I = [0, \frac{1}{2})$ is given by

$$\psi_{[0,\frac{1}{2}]} = \sqrt{2} \cdot (\mathbb{1}_{[0,\frac{1}{4})} - \mathbb{1}_{[\frac{1}{4},\frac{1}{2})}) \tag{4.54}$$

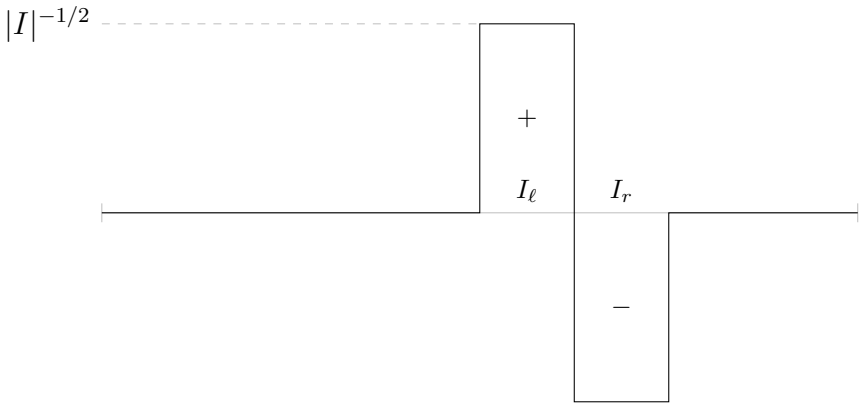

FIGURE 3. A Haar function ψ_I.

LEMMA 4.24. *The Haar system on* $[0,1]$ *is an orthonormal system.*

PROOF. Let $f \in \mathcal{H}$. If $f = \mathbb{1}_{[0,1]}$ then $\|f\|_2 = (\int_0^1 1^2)^{1/2} = 1$. Otherwise, $f = \psi_I$ for some $I \in \mathcal{D}$. Then by (4.52) and since I_ℓ and I_r are disjoint,

$$\|f\|_2^2 = \int_0^1 |\psi_I|^2 = |I|^{-1}\Big(\int_0^1 \mathbb{1}_{I_\ell} + \int_0^1 \mathbb{1}_{I_r}\Big) = 1. \tag{4.55}$$

Next let $f, g \in \mathcal{H}$ with $f \neq g$. Suppose that one of f, g equals $\mathbb{1}_{[0,1]}$, say $f = \mathbb{1}_{[0,1]}$. Then $g = \psi_J$ for some $J \in \mathcal{D}$ and thus

$$\langle f, g\rangle = \int_0^1 \psi_J = 0. \tag{4.56}$$

It remains to treat the case that $f = \psi_I$ and $g = \psi_J$ for $I, J \in \mathcal{D}$ with $I \neq J$. By Lemma 4.19 (i), I and J are either disjoint or contained in each other. If I and J are disjoint, then $\langle \psi_I, \psi_J\rangle = 0$. Otherwise they are contained in each other, say $I \subsetneq J$. Then ψ_J is constant on the set where ψ_I is different from zero. Thus,

$$\langle \psi_I, \psi_J\rangle = \int \psi_I \cdot \psi_J = \pm|I|^{-1} \int_0^1 (\mathbb{1}_{I_\ell} - \mathbb{1}_{I_r}) = 0. \tag{4.57}$$

□

Let us write

$$\mathcal{D}_{<n} = \bigcup_{0 \le k < n} \mathcal{D}_k \tag{4.58}$$

to denote the set of dyadic intervals of generation less than n. We want to study how continuous functions can be approximated by linear combinations of Haar functions.

Let $f \in C([0,1])$. Motivated by Theorem 4.11, we define for every positive integer n, the orthogonal projection

$$\mathbb{E}_n f = \sum_{I \in \mathcal{D}_{<n}} \langle f, \psi_I \rangle \psi_I. \tag{4.59}$$

DEFINITION 4.25. For a function f on $[0,1]$ and an interval $I \subset [0,1]$ we write $\langle f \rangle_I = |I|^{-1} \int_I f$ to denote the *average* or the *mean* of f on I.

THEOREM 4.26. *Let $\int_0^1 f = 0$. Then, for every $I \in \mathcal{D}_n$,*

$$\mathbb{E}_n f(x) = \langle f \rangle_I \quad \text{if } x \in I. \tag{4.60}$$

In other words,

$$\mathbb{E}_n f = \sum_{I \in \mathcal{D}_n} \langle f \rangle_I \mathbb{1}_I. \tag{4.61}$$

THEOREM 4.27. *Suppose that $\int_0^1 f = 0$ and $f \in C([0,1])$. Then*

$$\mathbb{E}_n f \to f \quad \text{uniformly on } [0,1] \text{ as } n \to \infty. \tag{4.62}$$

Remark. If $f \in C([0,1])$ does not have mean zero then $\mathbb{E}_n f$ converges to $f - \langle f \rangle_{[0,1]}$.

COROLLARY 4.28. *The Haar system is complete in the sense of Definition 4.14. For every $f \in C([0,1])$ we have*

$$\|f\|_2^2 = |\langle f \rangle_{[0,1]}|^2 + \sum_{I \in \mathcal{D}} |\langle f, \psi_I \rangle|^2. \tag{4.63}$$

EXERCISE 4.29. By using Theorem 4.27, prove Corollary 4.28.

PROOF OF THEOREM 4.26. Fix $n \geq 0$ and write $g = \mathbb{E}_n f$. We prove something seemingly stronger.

Claim. For every dyadic interval $I \in \mathcal{D}_n$, we have $\langle f \rangle_I = \langle g \rangle_I$.

This implies the statement in the theorem because $\mathbb{E}_n f$ is constant on dyadic intervals of generation n.

To prove the claim we perform an induction on $I \in \mathcal{D}_n$. To begin with, the claim holds for $I = [0,1)$ because $\int_0^1 f = 0$. Now suppose that it is true for some interval $I \in \mathcal{D}_{<n}$. It suffices to show that it also holds for I_ℓ and I_r, i.e. that

$$\langle f \rangle_{I_\ell} = \langle g \rangle_{I_\ell} \quad \text{and} \quad \langle f \rangle_{I_r} = \langle g \rangle_{I_r}. \tag{4.64}$$

Since the Haar system is orthonormal and $I \in \mathcal{D}_{<n}$,

$$\langle g, \psi_I \rangle = \sum_{J \in \mathcal{D}_{<n}} \langle f, \psi_J \rangle \langle \psi_J, \psi_I \rangle = \langle f, \psi_I \rangle. \tag{4.65}$$

Compute

$$\int_{I_\ell} f - \int_{I_r} f = |I|^{1/2} \int f \cdot \psi_I = |I|^{1/2} \langle f, \psi_I \rangle \tag{4.66}$$

and by the same reasoning,

$$\int_{I_\ell} g - \int_{I_r} g = |I|^{1/2} \langle g, \psi_I \rangle. \tag{4.67}$$

Combining the last three displays we get

$$\int_{I_\ell} f - \int_{I_r} f = \int_{I_\ell} g - \int_{I_r} g. \tag{4.68}$$

By the inductive hypothesis we know that $\langle f\rangle_I = \langle g\rangle_I$, so

$$\int_{I_\ell} f + \int_{I_r} f = \int_{I_\ell} g + \int_{I_r} g. \tag{4.69}$$

Adding the previous two displays gives $\langle f\rangle_{I_\ell} = \langle g\rangle_{I_\ell}$ and subtracting them gives $\langle f\rangle_{I_r} = \langle g\rangle_{I_r}$. This concludes the proof. □

Proof of Theorem 4.27. Let $\varepsilon > 0$. By uniform continuity of f on $[0,1]$ (which follows from Theorem 1.53) we may choose $\delta > 0$ such that $|f(t) - f(s)| < \varepsilon$ whenever $t, s \in [0,1]$ are such that $|t - s| < \delta$. Let $N \in \mathbb{N}$ be large enough so that $2^{-N} < \delta$ and $n \geq N$. Let $t \in [0,1]$ and $I \in \mathcal{D}_n$ such that $t \in I$. Then by Theorem 4.26,

$$|\mathbb{E}_n f(t) - f(t)| = |\langle f\rangle_I - f(t)| \leq |I|^{-1} \int_I |f(s) - f(t)| ds < \varepsilon. \tag{4.70}$$

□

Remark. This result goes back to A. Haar's 1910 article *Zur Theorie der orthogonalen Funktionensysteme* in *Math. Ann. 69 (1910), no. 3, p. 331–371.* The functions $(\mathbb{E}_n f)_n$ are also called *dyadic martingale averages* of f and have wide applications in modern analysis and probability theory.

Exercise 4.30. Recall the functions $r_n(x) = \operatorname{sgn}(\sin(2^n \pi x))$ from Exercise 4.10.
(i) Show that every r_n for $n \geq 1$ can be written as a finite linear combination of Haar functions and determine the coefficients of this linear combination.
(ii) Show that the orthonormal system on $[0,1]$ given by $(r_n)_n$ is not complete.

Exercise 4.31. Define

$$\Delta_n f = \mathbb{E}_{n+1} f - \mathbb{E}_n f, \quad Sf = \big(\sum_{n\geq 1} |\Delta_n f|^2\big)^{1/2}. \tag{4.71}$$

(i) Assume that $\int_0^1 f = 0$. Prove that $\|Sf\|_2 = \|f\|_2$.
(ii) Show that for every $m \in \mathbb{N}$ there exists a function f_m that is a finite linear combination of Haar functions such that $\sup_{x\in[0,1]} |f_m(x)| \leq 1$ and $\sup_{x\in[0,1]} |Sf_m(x)| \geq m$.

4. Trigonometric polynomials

In the following we will only be concerned with the *trigonometric system* on $[0,1]$:

$$\phi_n(x) = e^{2\pi i n x} \ \ (n \in \mathbb{Z}) \tag{4.72}$$

Definition 4.32. A *trigonometric polynomial* is a function of the form

$$f(x) = \sum_{n=-N}^{N} c_n e^{2\pi i n x} \quad (x \in \mathbb{R}), \tag{4.73}$$

where $N \in \mathbb{N}$ and $c_n \in \mathbb{C}$. If c_N or c_{-N} is non-zero, then N is called the *degree* of f.

From Euler's identity (see Fact A.12) we see that every trigonometric polynomial can also be written in the alternate form

$$f(x) = a_0 + \sum_{n=1}^{N} (a_n \cos(2\pi n x) + b_n \sin(2\pi n x)). \tag{4.74}$$

Exercise 4.33. Work out how the coefficients a_n, b_n in (4.74) are related to the c_n in (4.73).

Every trigonometric polynomial is *1-periodic* :

$$f(x) = f(x+1) \tag{4.75}$$

for all $x \in \mathbb{R}$.

LEMMA 4.34. $(e^{2\pi inx})_{n\in\mathbb{Z}}$ *forms an orthonormal system on* $[0,1]$. *In particular,*
(i) for all $n \in \mathbb{Z}$,

$$\int_0^1 e^{2\pi inx} dx = \begin{cases} 0, & \text{if } n \neq 0, \\ 1, & \text{if } n = 0. \end{cases} \tag{4.76}$$

(ii) if $f(x) = \sum_{n=-N}^{N} c_n e^{2\pi inx}$ *is a trigonometric polynomial, then*

$$c_n = \int_0^1 f(t) e^{-2\pi int} dt. \tag{4.77}$$

One goal in this section is to show that this orthonormal system is in fact complete.

We denote by pc the space of piecewise continuous, 1-periodic functions $f : \mathbb{R} \to \mathbb{C}$ (let us call a 1-periodic function piecewise continuous, if its restriction to $[0,1]$ is piecewise continuous in the sense defined in the beginning of this section).

DEFINITION 4.35. For a 1-periodic function $f \in$ pc and $n \in \mathbb{Z}$ we define the *nth Fourier coefficient* by

$$\widehat{f}(n) = \int_0^1 f(t) e^{-2\pi int} dt. \tag{4.78}$$

The series

$$\sum_{n=-\infty}^{\infty} \widehat{f}(n) e^{2\pi inx} \tag{4.79}$$

is called the *Fourier series* of f.

The question of when the Fourier series of a function f converges and in what sense it represents the function f is a very subtle issue and we will only scratch the surface in this lecture.

DEFINITION 4.36. For a 1-periodic function $f \in$ pc we define the *partial sums*

$$S_N f(x) = \sum_{n=-N}^{N} \widehat{f}(n) e^{2\pi inx}. \tag{4.80}$$

Remark. Note that since $(\phi_n)_n$ is an orthonormal system, $S_N f$ is exactly the orthogonal projection of f onto the space of trigonometric polynomials of degree $\leq N$. In particular, Theorem 4.11 tells us that

$$\|f - S_N f\|_2 \leq \|f - g\|_2 \tag{4.81}$$

holds for all trigonometric polynomials g of degree $\leq N$. That is, $S_N f$ is the best approximation to f in the L^2-norm among all trigonometric polynomials of degree $\leq N$.

DEFINITION 4.37 (Convolution). For two 1-periodic functions $f, g \in$ pc we define their *convolution* by

$$f * g(x) = \int_0^1 f(t) g(x-t) dt \tag{4.82}$$

Note that if $f, g \in \mathrm{pc}$ then $f * g \in \mathrm{pc}$.

EXAMPLE 4.38. Suppose f is a given 1-periodic function and g is a 1-periodic function, non-negative and $\int_0^1 g = 1$. Then $(f * g)(x)$ can be viewed as a *weighted average* of f around x with weight profile g. For instance, if $g = 2N\mathbf{1}_{[-1/N,1/N]}$, then $(f * g)(x)$ is the average value of f in the interval $[x - 1/N, x + 1/N]$.

LEMMA 4.39. *For 1-periodic functions $f, g \in \mathrm{pc}$,*

$$f * g = g * f. \tag{4.83}$$

PROOF. For $x \in [0, 1]$,

$$f*g(x) = \int_0^1 f(t)g(x-t)dt = \int_{x-1}^x f(x-t)g(t)dt = \int_{x-1}^0 f(x-t)g(t)dt + \int_0^x f(x-t)g(t). \tag{4.84}$$

$$= \int_x^1 f(x-(t-1))g(t-1)dt + \int_0^x f(x-t)g(t)dt = g * f(x), \tag{4.85}$$

where in the last step we used that $f(x - (t - 1)) = f(x - t)$ and $g(t - 1) = g(t)$ by periodicity. □

It turns out that the partial sum $S_N f$ can be written in terms of a convolution:

$$S_N f(x) = \sum_{n=-N}^{N} \int_0^1 f(t)e^{-2\pi int}dt e^{2\pi inx} = \int_0^1 f(t) \sum_{n=-N}^{N} e^{2\pi in(x-t)}dt = f * D_N(x). \tag{4.86}$$

where

$$D_N(x) = \sum_{n=-N}^{N} e^{2\pi inx}. \tag{4.87}$$

The sequence of functions $(D_N)_N$ is called *Dirichlet kernel* . The Dirichlet kernel can be written more explicitly.

LEMMA 4.40. *We have*

$$D_N(x) = \frac{\sin(2\pi(N + \frac{1}{2})x)}{\sin(\pi x)} \tag{4.88}$$

PROOF.

$$D_N(x) = \sum_{n=-N}^{N} e^{2\pi inx} = e^{-2\pi iNx} \sum_{n=0}^{2N} e^{2\pi inx} = e^{-2\pi iNx} \frac{e^{2\pi i(2N+1)x} - 1}{e^{2\pi ix} - 1} \tag{4.89}$$

$$= \frac{e^{2\pi i(N+\frac{1}{2})} - e^{-2\pi i(N+\frac{1}{2})x}}{e^{\pi ix} - e^{-\pi ix}} = \frac{\sin(2\pi(N + \frac{1}{2})x)}{\sin(\pi x)}. \tag{4.90}$$

□

We would like to approximate continuous functions by trigonometric polynomials. If f is only continuous it may happen that $S_N f(x)$ does not converge. However, instead of $S_N f$ we may also consider their arithmetic means. We define the Fejér kernel by

$$K_N(x) = \frac{1}{N+1} \sum_{n=0}^{N} D_n(x). \tag{4.91}$$

LEMMA 4.41. *We have*

$$K_N(x) = \frac{1}{2(N+1)} \frac{1-\cos(2\pi(N+1)x)}{\sin(\pi x)^2} = \frac{1}{N+1} \Big(\frac{\sin(\pi(N+1)x)}{\sin(\pi x)} \Big)^2 \tag{4.92}$$

PROOF. Using that $2\sin(x)\sin(y) = \cos(x-y) - \cos(x+y)$,

$$D_N(x) = \frac{\sin(2\pi(N+\frac{1}{2})x)}{\sin(\pi x)} = \frac{2\sin(\pi x)\sin(2\pi(N+\frac{1}{2})x)}{2\sin(\pi x)^2} = \frac{\cos(2\pi N x) - \cos(2\pi(N+1)x)}{2\sin(\pi x)^2}. \tag{4.93}$$

Thus,

$$\sum_{n=0}^{N} D_n(x) = \frac{1}{2\sin(\pi x)^2} \sum_{n=0}^{N} \cos(2\pi n x) - \cos(2\pi(n+1)x) = \frac{1-\cos(2\pi(N+1)x)}{2\sin(\pi x)^2} \tag{4.94}$$

The claim now follows from the formula $1-\cos(2x) = 2\sin(x)^2$. □

As a consequence of this explicit formula we see that $K_N(x) \geq 0$ for all $x \in \mathbb{R}$ which is not at all obvious from the initial definition. We define

$$\sigma_N f(x) = f * K_N(x). \tag{4.95}$$

THEOREM 4.42 (Fejér). *For every 1-periodic continuous function f,*

$$\sigma_N f \to f \tag{4.96}$$

uniformly on $\mathbb{R}$ as $N \to \infty$.

COROLLARY 4.43. *Every 1-periodic continuous function can be uniformly approximated by trigonometric polynomials.*

Remark. There is nothing special about the period 1 here. By considering the orthonormal system $(L^{-1/2} e^{\frac{2\pi}{L} inx})_{n\in\mathbb{Z}}$ we obtain a similar result for L-periodic functions.

This follows from Fejér's theorem because $\sigma_N f$ is a trigonometric polynomial:

$$\sigma_N f(x) = \int_0^1 f(t) \frac{1}{N+1} \sum_{n=0}^{N} \sum_{k=-n}^{n} e^{2\pi i k(x-t)} dt = \frac{1}{N+1} \sum_{n=0}^{N} \sum_{k=-n}^{n} \int_0^1 f(t) e^{-2\pi i k t} dt e^{2\pi i k x} \tag{4.97}$$

$$= \frac{1}{N+1} \sum_{n=0}^{N} \sum_{k=-n}^{n} \widehat{f}(k) e^{2\pi i k x} = \frac{1}{N+1} \sum_{k=-N}^{N} \sum_{n=|k|}^{N} \widehat{f}(k) e^{2\pi i k x} = \sum_{k=-N}^{N} (1 - \tfrac{|k|}{N+1}) \widehat{f}(k) e^{2\pi i k x}. \tag{4.98}$$

We will now derive Fejér's Theorem as a consequence of a more general principle.

DEFINITION 4.44 (Approximation of unity). A sequence of 1-periodic continuous functions $(k_n)_n$ is called *approximation of unity* if for all 1-periodic continuous functions f we have that $f * k_n$ converges uniformly to f on $\mathbb{R}$. That is,

$$\sup_{x\in\mathbb{R}} |f * k_n(x) - f(x)| \to 0 \tag{4.99}$$

as $n \to \infty$.

Remark. There is no *unity* for the convolution of functions. More precisely, there exists no continuous function k such that $k * f = f$ for all continuous, 1-periodic f (this is the content of Exercise 4.62). An approximation of unity is a sequence $(k_n)_n$ that *approximates unity*:

$$\lim_{n\to\infty} k_n * f = f \tag{4.100}$$

for every continuous, 1-periodic f.

THEOREM 4.45. *Let $(k_n)_n$ be a sequence of 1-periodic continuous functions such that*

(1) $k_n(x) \geq 0$
(2) $\int_{-1/2}^{1/2} k_n(t)dt = 1$.
(3) For all $1/2 \geq \delta > 0$ *we have*

$$\int_{-\delta}^{\delta} k_n(t)dt \to 1 \tag{4.101}$$

as $n \to \infty$.

Then $(k_n)_n$ is an approximation of unity.

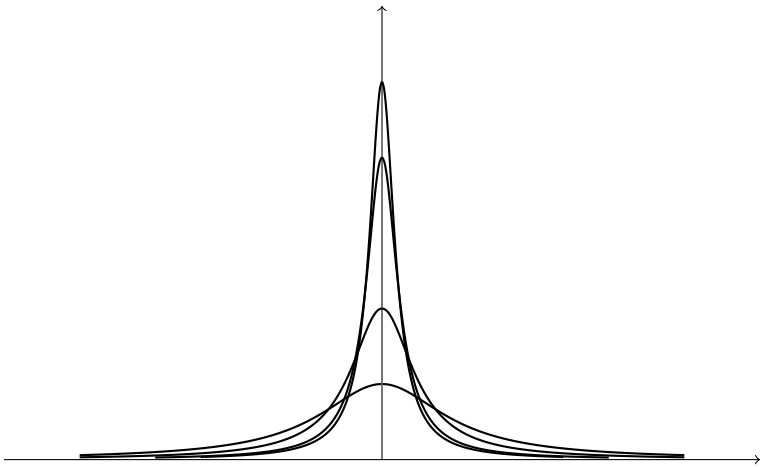

FIGURE 4. Approximation of unity

Assumption (3) is a precise way to express the idea that the "mass" of k_n concentrates near the origin. Keeping in mind Assumption (2), Assumption (3) can be rewritten equivalently as:

$$\int_{\frac{1}{2}\geq|t|\geq\delta} k_n(t)dt \to 0 \tag{4.102}$$

PROOF. Let f be 1-periodic and continuous. By continuity, f is bounded and uniformly continuous on $[-1/2, 1/2]$. By periodicity, f is also bounded and uniformly continuous on all of $\mathbb{R}$. Let $\varepsilon > 0$. By uniform continuity there exists $\delta > 0$ such that

$$|f(x-t) - f(x)| \leq \varepsilon/2 \tag{4.103}$$

for all $|t| < \delta$, $x \in \mathbb{R}$. Using Assumption (2),

$$f * k_n(x) - f(x) = \int_{-1/2}^{1/2} (f(x-t) - f(x))k_n(t)dt = A + B, \tag{4.104}$$

where

$$A = \int_{|t|\leq\delta} (f(x-t) - f(x))k_n(t)dt, \quad B = \int_{\frac{1}{2}\geq|t|\geq\delta} (f(x-t) - f(x))k_n(t)dt. \tag{4.105}$$

By 4.103 and Assumption (2),

$$|A| \leq \frac{\varepsilon}{2} \int_{|t|\leq\delta} k_n(t)dt \leq \frac{\varepsilon}{2}. \tag{4.106}$$

Since f is bounded there exists $C > 0$ such that $|f(x)| \leq C$ for all $x \in \mathbb{R}$. for all $0 < \delta < \frac{1}{2}$. Let N be large enough so that for all $n \geq N$,

$$\int_{\frac{1}{2}\geq|t|\geq\delta} k_n(t)dt \leq \frac{\varepsilon}{4C}. \tag{4.107}$$

Thus, if $n \geq N$,

$$|B| \leq 2C \int_{\frac{1}{2}\geq|t|\geq\delta} k_n(t)dt \leq \frac{\varepsilon}{2}. \tag{4.108}$$

This implies

$$|f * k_n(x) - f(x)| \leq \varepsilon/2 + \varepsilon/2 \leq \varepsilon \tag{4.109}$$

for $n \geq N$ and $x \in \mathbb{R}$. □

COROLLARY 4.46. *The Fejér kernel $(K_N)_N$ is an approximation of unity.*

PROOF. We verify the assumptions of Theorem 4.45. From (4.92) we see that $K_N \geq 0$. Also,

$$\int_{-1/2}^{1/2} K_N(t)dt = \frac{1}{N+1} \sum_{n=0}^{N} \sum_{k=-n}^{n} \int_{-1/2}^{1/2} e^{2\pi ikt}dt = \frac{1}{N+1} \sum_{n=0}^{N} 1 = 1. \tag{4.110}$$

Now we verify the last property. Let $\frac{1}{2} > \delta > 0$ and $|x| \geq \delta$. By (4.92),

$$K_N(x) \leq \frac{1}{N+1} \frac{1}{\sin(\pi\delta)^2} \tag{4.111}$$

Thus,

$$\int_{\frac{1}{2}\geq|t|\geq\delta} K_N(t)dt \leq \frac{1}{N+1} \frac{1}{\sin(\pi\delta)^2} \tag{4.112}$$

which converges to 0 as $N \to \infty$. □

Therefore we have proven Fejér's theorem. Note that although the Dirichlet kernel also satisfies Assumptions (2) and (3), it is *not* an approximation of unity. In other words, if f is continuous then it is *not* necessarily true that $S_N f \to f$ uniformly. However, we can use Fejér's theorem to show that $S_N f \to f$ in the L^2-norm.

THEOREM 4.47. *Let f be a 1-periodic and continuous function. Then*

$$\lim_{N\to\infty} \|S_N f - f\|_2 = 0. \tag{4.113}$$

PROOF. Let $\varepsilon > 0$. By Fejér's theorem there exists a trigonometric polynomial p such that $|f(x) - p(x)| \leq \varepsilon/2$ for all $x \in \mathbb{R}$. Then

$$\|f - p\|_2 = \Big(\int_0^1 |f(x) - p(x)|^2 dx \Big)^{1/2} \leq \varepsilon/2. \tag{4.114}$$

Let N be the degree of p. Then $S_N p = p$ by Fact 4.34. Thus,

$$S_N f - f = S_N f - S_N p + S_N p - f = S_N(f - p) + p - f. \tag{4.115}$$

By Minkowski's inequality,

$$\|S_N f - f\|_2 \leq \|S_N(f - p)\|_2 + \|p - f\|_2 \tag{4.116}$$

Bessel's inequality (Theorem 4.12) says that $\|S_N f\|_2 \leq \|f\|_2$. Therefore,

$$\|S_N f - f\|_2 \leq 2\|f - p\|_2 \leq \varepsilon. \tag{4.117}$$

□

In view of Theorem 4.15 this means that the trigonometric system is complete.

COROLLARY 4.48 (Parseval's theorem). *If f, g are 1-periodic, continuous functions, then*

$$\langle f, g \rangle = \sum_{n=-\infty}^{\infty} \widehat{f}(n)\overline{\widehat{g}(n)}. \tag{4.118}$$

In particular,

$$\|f\|_2^2 = \sum_{n=-\infty}^{\infty} |\widehat{f}(n)|^2. \tag{4.119}$$

PROOF. We have

$$\langle S_N f, g \rangle = \sum_{n=-N}^{N} \widehat{f}(n)\langle e^{2\pi i n x}, g \rangle = \sum_{n=-N}^{N} \widehat{f}(n)\overline{\widehat{g}(n)}. \tag{4.120}$$

But $\langle S_N f, g \rangle \to \langle f, g \rangle$ as $N \to \infty$ because

$$|\langle S_N f, g \rangle - \langle f, g \rangle| = |\langle S_N f - f, g \rangle| \leq \|S_N f - f\|_2 \|g\|_2 \to 0 \tag{4.121}$$

as $N \to \infty$. Here we have used the Cauchy-Schwarz inequality and the previous theorem. Equation (4.119) follows from putting $f = g$. □

Remark. Theorems 4.47 and Corollary 4.48 also hold for piecewise continuous and 1-periodic functions.

EXERCISE 4.49. (i) Let f be the 1-periodic function such that $f(x) = x$ for $x \in [0, 1)$. Compute the Fourier coefficient $\widehat{f}(n)$ for every $n \in \mathbb{Z}$ and use Parseval's theorem to derive the formula

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}. \tag{4.122}$$

(ii) Using Parseval's theorem for a suitable 1-periodic function, determine the value of $\sum_{n=1}^{\infty} \frac{1}{n^4}$.

While the Fourier series of a continuous function does not necessarily converge pointwise, we can obtain pointwise convergence easily if we impose additional conditions.

THEOREM 4.50. *Let f be a 1-periodic continuous function and let $x \in \mathbb{R}$. Assume that f is differentiable at x. Then $S_N f(x) \to f(x)$ as $N \to \infty$.*

PROOF. By definition,

$$S_N f(x) = \int_0^1 f(x-t) D_N(t) dt. \tag{4.123}$$

Also,

$$\int_0^1 D_N(t) dt = \sum_{n=-N}^{N} \int_0^1 e^{2\pi i n t} dt = 1. \tag{4.124}$$

Thus from Fact 4.40,

$$S_N f(x) - f(x) = \int_0^1 (f(x-t) - f(x)) D_N(t) dt \tag{4.125}$$

$$= \int_0^1 g(t) \sin(2\pi(N + \tfrac{1}{2})t) dt, \tag{4.126}$$

where

$$g(t) = \frac{f(x-t) - f(x)}{\sin(\pi t)}. \tag{4.127}$$

Differentiability of f at x implies that g is continuous at 0. Indeed,

$$\frac{f(x-t) - f(x)}{\sin(\pi t)} = \frac{f(x-t) - f(x)}{t} \frac{t}{\sin(\pi t)} \to f'(x) \frac{1}{\pi} \tag{4.128}$$

as $t \to 0$.

EXERCISE 4.51. Show that $\phi_n(x) = \sqrt{2} \sin(2\pi(n + \frac{1}{2})x)$ with $n = 1, 2, \ldots$ defines an orthonormal system on $[0, 1]$.

With this exercise, the claim follows from (4.126) and the Riemann-Lebesgue lemma (Corollary 4.13). □

EXERCISE 4.52. Show that there exists a constant $c > 0$ such that

$$\int_0^1 |D_N(x)| dx \geq c \log(2 + N) \tag{4.129}$$

holds for all $N = 0, 1, \ldots$.

EXERCISE 4.53. (i) Let $(a_k)_k$ be a sequence of complex numbers with limit L. Prove that

$$\lim_{n \to \infty} \frac{a_1 + \cdots + a_n}{n} = L$$

Given the sequence a_k, form the partial sums $s_n = \sum_{k=1}^{n} a_k$ and let

$$\sigma_N = \frac{s_1 + \cdots + s_N}{N}.$$

σ_N is called the Nth *Cesàro mean* of the sequence s_k or the Nth Cesàro sum of the series $\sum_{k=1}^{\infty} a_k$. If σ_N converges to a limit S we say that the series $\sum_{k=1}^{\infty} a_k$ is Cesàro summable to S.
(ii) Prove that if $\sum_{k=1}^{\infty} a_k$ is summable to S (i.e. by definition converges with sum S) then $\sum_{k=1}^{\infty} a_k$ is Cesàro summable to S.
(iii) Prove that the sum $\sum_{k=1}^{\infty} (-1)^{k-1}$ does not converge but is Cesàro summable to some limit S and determine S.

5. The Stone-Weierstrass Theorem

We have seen two different classes of continuous functions that are rich enough to enable uniform approximation of arbitrary continuous functions: polynomials and trigonometric polynomials. In other words, we have shown that polynomials are dense in $C([a,b])$ and trigonometric polynomials are dense in $C(\mathbb{R}/\mathbb{Z})$ (space of continuous and 1-periodic functions). The Stone-Weierstrass theorem gives a sufficient criterion for a subset of $C(K)$ to be dense (where K is a compact metric space). Both, Fejér's and Weierstrass' theorems are consequences of this more general theorem.

THEOREM 4.54 (Stone-Weierstrass). *Let K be a compact metric space and $\mathcal{A} \subset C(K)$. Assume that $\mathcal{A}$ satisfies the following conditions:*

(1) $\mathcal{A}$ is a self-adjoint algebra : for $f, g \in \mathcal{A}$, $c \in \mathbb{C}$,

$$f + g \in \mathcal{A}, f \cdot g \in \mathcal{A}, c \cdot f \in \mathcal{A}, \overline{f} \in \mathcal{A}. \tag{4.130}$$

(2) $\mathcal{A}$ separates points : for all $x, y \in K$ with $x \neq y$ there exists $f \in \mathcal{A}$ such that $f(x) \neq f(y)$.

(3) $\mathcal{A}$ vanishes nowhere : for all $x \in K$ there exists $f \in \mathcal{A}$ such that $f(x) \neq 0$.

Then $\mathcal{A}$ is dense in $C(K)$ (that is, $\overline{\mathcal{A}} = C(K)$).

EXERCISE 4.55. Let K be a compact metric space. Show that if a subset $\mathcal{A} \subset C(K)$ does not separate points or does not vanish nowhere, then $\mathcal{A}$ is not dense.

EXERCISE 4.56. Let $\mathcal{A} \subset C([1,2])$ be the set of all polynomials of the form $p(x) = \sum_{k=0}^{n} c_k x^{2k+1}$ where $c_k \in \mathbb{C}$ and n a non-negative integer. Show that $\mathcal{A}$ is dense, but not an algebra.

Before we begin the proof of the Stone-Weierstrass theorem we first need some preliminary lemmas.

LEMMA 4.57. *For every $a > 0$ there exists a sequence of polynomials $(p_n)_n$ with real coefficients such that $p_n(0) = 0$ for all n and $\sup_{x \in [-a,a]} |p_n(x) - |x|| \to 0$ as $n \to \infty$.*

PROOF. From Weierstrass' theorem we get that there exists a sequence of polynomials q_n that converges uniformly to $f(x) = |x|$ on $[-a,a]$. Now set $p_n(x) = q_n(x) - q_n(0)$. □

EXERCISE 4.58. Work out an explicit sequence of polynomials $(p_n)_n$ that converges uniformly to $x \mapsto |x|$ on $[-1,1]$.

Let $\mathcal{A} \subset C(K)$ satisfy conditions (1),(2),(3). Observe that then also $\overline{\mathcal{A}}$ satisfies (1), (2), (3).

We may assume without loss of generality that we are dealing with real-valued functions (otherwise split functions into real and imaginary parts $f = g + ih$ and go through the proof for both parts).

LEMMA 4.59. *If $f \in \overline{\mathcal{A}}$, then $|f| \in \overline{\mathcal{A}}$.*

PROOF. Let $\varepsilon > 0$ and $a = \max_{x \in K} |f(x)|$. By Lemma 4.57 there exist $c_1, \ldots, c_n \in \mathbb{R}$ such that

$$|\sum_{i=1}^{n} c_i y^i - |y|| \leq \varepsilon. \tag{4.131}$$

for all $y \in [-a, a]$. By Condition (1) we have that

$$g = \sum_{i=1}^{n} c_i f^i \in \overline{\mathcal{A}}. \tag{4.132}$$

Then $|g(x) - |f(x)|| \leq \varepsilon$ for all $x \in K$. Thus, $|f|$ can be uniformly approximated by functions in $\overline{\mathcal{A}}$. But $\overline{\mathcal{A}}$ is closed, so $|f| \in \overline{\mathcal{A}}$. □

LEMMA 4.60. *If $f_1, \ldots, f_m \in \overline{\mathcal{A}}$, then $\min(f_1, \ldots, f_m) \in \overline{\mathcal{A}}$ and $\max(f_1, \ldots, f_m) \in \overline{\mathcal{A}}$.*

PROOF. It suffices to show the claim for $m = 2$ (the general case then follows by induction). Let $f, g \in \overline{\mathcal{A}}$. We have

$$\min(f, g) = \frac{f+g}{2} - \frac{|f-g|}{2}, \ \max(f, g) = \frac{f+g}{2} + \frac{|f-g|}{2}. \tag{4.133}$$

Thus, Condition (1) and Lemma 4.60 imply that $\min(f, g), \max(f, g) \in \overline{\mathcal{A}}$. □

LEMMA 4.61. *For every $x_0, x_1 \in K$, $x_0 \neq x_1$ and $c_0, c_1 \in \mathbb{R}$ there exists $f \in \overline{\mathcal{A}}$ such that $f(x_i) = c_i$ for $i = 0, 1$.*

In other words, any two points in $K \times \mathbb{R}$ that could lie on the graph of a function in $\overline{\mathcal{A}}$ do lie on the graph of a function in $\overline{\mathcal{A}}$.

PROOF. By Conditions (2) and (3) there exist $g, h_0, h_1 \in \overline{\mathcal{A}}$ such that $g(x_0) \neq g(x_1)$ and $h_i(x_i) \neq 0$ for $i = 0, 1$. Set

$$u_i(x) = g(x)h_i(x) - g(x_{1-i})h_i(x). \tag{4.134}$$

Then $u_i(x_{1-i}) = 0$ and $u_i(x_i) \neq 0$ for $i = 0, 1$. Set

$$f(x) = \frac{c_0 u_0(x)}{u_0(x_0)} + \frac{c_1 u_1(x)}{u_1(x_1)}. \tag{4.135}$$

Then $f(x_0) = c_0$ and $f(x_1) = c_1$ and $f \in \overline{\mathcal{A}}$ by Condition (1). □

This lemma can be seen as a baby version of the full theorem: the statement extends to finitely many points. So we can use it to find a function in $\overline{\mathcal{A}}$ that matches a given function f in any given collection of finitely many points (see Exercise 4.81). Thus, if K was finite, we would already be done. If K is not finite, we need to exploit compactness. Let us now get to the details.

Fix $f \in C(K)$ and let $\varepsilon > 0$.

Claim: For every $x \in K$ there exists $g_x \in \overline{\mathcal{A}}$ such that $g_x(x) = f(x)$ and $g_x(t) > f(t) - \varepsilon$ for $t \in K$.

Proof of Claim. Let $y \in K$. By Lemma 4.61 there exists $h_y \in \overline{\mathcal{A}}$ such that $h_y(x) = f(x)$ and $h_y(y) = f(y)$. By continuity of h_y there exists an open ball B_y around y such that $|h_y(t) - f(t)| < \varepsilon$ for all $t \in B_y$. In particular,

$$h_y(t) > f(t) - \varepsilon. \tag{4.136}$$

Observe that $(B_y)_{y \in K}$ is an open cover of K. Since K is compact, we can find a finite subcover by $B_{y_1}, \ldots, B_{y_m}$. Set

$$g_x = \max(h_{y_1}, \ldots, h_{y_m}). \tag{4.137}$$

By Lemma 4.60, $g_x \in \overline{\mathcal{A}}$. □

By continuity of g_x there exists an open ball U_x such that

$$|g_x(t) - f(t)| < \varepsilon \tag{4.138}$$

for $t \in U_x$. In particular,

$$g_x(t) < f(t) + \varepsilon. \tag{4.139}$$

$(U_x)_{x\in K}$ is an open cover of K which has a finite subcover by $U_{x_1}, \ldots, U_{x_n}$. Then let

$$h = \min(g_{x_1}, \ldots, g_{x_n}). \tag{4.140}$$

By Lemma 4.60 we have $h \in \overline{\mathcal{A}}$. Also,

$$f(t) - \varepsilon < h(t) < f(t) + \varepsilon \tag{4.141}$$

for all $t \in K$. That is,

$$|f(t) - h(t)| < \varepsilon \tag{4.142}$$

for all $t \in K$. This proves that $f \in \overline{\mathcal{A}}$.

6. Further exercises

Exercise 4.62. Show that there exists no continuous 1-periodic function g such that $f * g = f$ holds for all continuous 1-periodic functions f.
Hint: Use the Riemann-Lebesgue lemma.

Exercise 4.63. Give an alternative proof of Weierstrass' theorem by using Fejér's theorem and then approximating the resulting trigonometric polynomials by truncated Taylor expansions.

Exercise 4.64. Find a sequence of continuous functions $(f_n)_n$ on $[0,1]$ and a continuous function f on $[0,1]$ such that $\|f_n - f\|_2 \to 0$, but $f_n(x)$ does not converge to $f(x)$ for any $x \in [0,1]$.

Exercise 4.65 (Weighted L^2 norms). Fix a function $w \in C([a,b])$ that is non-negative and does not vanish identically. Let us define another inner product by

$$\langle f, g\rangle_{L^2(w)} = \int_a^b f(x)\overline{g(x)}w(x)dx \tag{4.143}$$

and a corresponding norm $\|f\|_{L^2(w)} = \langle f, f\rangle_{L^2(w)}^{1/2}$. Similarly, we say that $(\phi_n)_n$ is an orthonormal system by asking that $\langle \phi_n, \phi_m\rangle_{L^2(w)}$ is 1 if $n = m$ and 0 otherwise. Verify that all theorems in Section 2 continue to hold when $\langle\cdot,\cdot\rangle$, $\|\cdot\|_2$ are replaced by $\langle\cdot,\cdot\rangle_{L^2(w)}$, $\|\cdot\|_{L^2(w)}$, respectively.

Exercise 4.66. Let $w \in C([0,1])$ be such that $w(x) \geq 0$ for all $x \in [0,1]$ and $w \not\equiv 0$. Prove that there exists a sequence of real-valued polynomials $(p_n)_n$ such that p_n is of degree n and

$$\int_0^1 p_n(x)p_m(x)w(x)dx = \begin{cases} 1, & \text{if } n = m, \\ 0, & \text{if } n \neq m \end{cases} \tag{4.144}$$

for all non-negative integers n, m.

EXERCISE 4.67 (Chebyshev polynomials). Define a sequence of polynomials $(T_n)_n$ by $T_0(x) = 1$, $T_1(x) = x$ and the recurrence relation $T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x)$ for $n \geq 2$.
(i) Show that $T_n(x) = \cos(nt)$ if $x = \cos(t)$.
Hint: Use that $2\cos(a)\cos(b) = \cos(a+b) + \cos(a-b)$ for all $a, b \in \mathbb{C}$.
(ii) Compute

$$\int_{-1}^{1} T_n(x)T_m(x)\frac{dx}{\sqrt{1-x^2}} \tag{4.145}$$

for all non-negative integers n, m.
(iii) Prove that $|T_n(x)| \leq 1$ for $x \in [-1, 1]$ and determine when there is equality.

EXERCISE 4.68. Let d be a positive integer and $f \in C([a, b])$. Denote by P_d the set of polynomials with real coefficients of degree $\leq d$. Prove that there exists a polynomial $p_* \in P_d$ such that $\|f - p_*\|_\infty = \inf_{p \in P_d} \|f - p\|_\infty$.
Hint: Find a way to apply Theorem 1.55.

EXERCISE 4.69. Let f be smooth on $[0, 1]$ (that is, arbitrarily often differentiable).
(i) Let p be a polynomial such that $|f'(x) - p(x)| \leq \varepsilon$ for all $x \in [0, 1]$. Construct a polynomial q such that $|f(x) - q(x)| \leq \varepsilon$ for all $x \in [0, 1]$.
(ii) Prove that there exists a sequence of polynomials $(p_n)_n$ such that $(p_n^{(k)})_n$ converges uniformly on $[0, 1]$ to $f^{(k)}$ for all $k = 0, 1, 2, \dots$.

EXERCISE 4.70 (The space L^2). Let (X, d) be a metric space. Recall that the *completion* $\overline{X}$ of X is defined as follows: for two Cauchy sequences $(a_n)_n$, $(b_n)_n$ in X we say that $(a_n)_n \sim (b_n)_n$ if $\lim_{n\to\infty} d(a_n, b_n) = 0$. Then $\sim$ is an equivalence relation on the space of Cauchy sequences and we define $\overline{X}$ as the set of equivalence classes. We identify X with a subset of $\overline{X}$ by identifying $x \in X$ with the equivalence class of the constant sequence $(x, x, \dots)$. We make $\overline{X}$ a metric space by defining

$$d(a, b) = \lim_{n\to\infty} d(a_n, b_n), \tag{4.146}$$

where $(a_n)_n, (b_n)_n$ are representatives of $a, b \in \overline{X}$, respectively. Then $\overline{X}$ is a complete metric space. Let us denote by $L^2_c(a, b)$ the metric space of continuous functions on $[a, b]$ equipped with the metric $d(f, g) = \|f - g\|_2$, where $\|f\|_2 = (\int_a^b |f|^2)^{1/2}$. Define

$$L^2(a, b) = \overline{L^2_c(a, b)}. \tag{4.147}$$

(i) Define an inner product on $L^2(a, b)$ by

$$\langle f, g\rangle = \lim_{n\to\infty} \int_a^b f_n(x)\overline{g_n(x)}dx, \tag{4.148}$$

for $f, g \in L^2(a, b)$ with $(f_n)_n, (g_n)_n$ being representatives of f, g, respectively. Show that this is well-defined: that is, show that the limit on the right hand side exists and is independent of the representatives $(f_n)_n, (g_n)_n$ and that $\langle \cdot, \cdot \rangle$ is an inner product.
Hint: Use the Cauchy-Schwarz inequality on $L^2_c(a, b)$.

For $f \in L^2(a, b)$ we define $\|f\|_2 = \langle f, f\rangle^{1/2}$. Let $(\phi_n)_{n=1,2,\dots}$ be an orthonormal system in $L^2(a, b)$ (that is, $\langle \phi_n, \phi_m \rangle = 0$ if $n \neq m$ and $= 1$ if $n = m$).

(ii) Prove Bessel's inequality: for every $f \in L^2(a,b)$ it holds that

$$\sum_{n=1}^{\infty} |\langle f, \phi_n \rangle|^2 \leq \|f\|_2^2 \tag{4.149}$$

Hint: Use the same proof as seen for $L^2_c(a,b)$ in the lecture!

(iii) Let $(c_n)_n \subset \mathbb{C}$ be a sequence of complex numbers and let

$$f_N = \sum_{n=1}^{N} c_n \phi_n \in L^2(a,b). \tag{4.150}$$

Show that $(f_N)_N$ converges in $L^2(a,b)$ if and only if

$$\sum_{n=1}^{\infty} |c_n|^2 < \infty. \tag{4.151}$$

EXERCISE 4.71. Let f be the 1-periodic function such that $f(x) = |x|$ for $x \in [-1/2, 1/2]$. Determine explicitly a sequence of trigonometric polynomials $(p_N)_N$ such that $p_N \to f$ uniformly as $N \to \infty$.

EXERCISE 4.72. Let f, g be continuous, 1-periodic functions.
(i) Show that $\widehat{f * g}(n) = \widehat{f}(n)\widehat{g}(n)$.
(ii) Show that $\widehat{f \cdot g}(n) = \sum_{m \in \mathbb{Z}} \widehat{f}(n-m)\widehat{g}(m)$.
(iii) If f is continuously differentiable, prove that $\widehat{f'}(n) = 2\pi i n \widehat{f}(n)$.
(iv) Let $y \in \mathbb{R}$ and set $f_y(x) = f(x+y)$. Show that $\widehat{f_y}(n) = e^{2\pi i n y}\widehat{f}(n)$.
(v) Let $m \in \mathbb{Z}$, $m \neq 0$ and set $f_m(x) = f(mx)$. Show that $\widehat{f_m}(n)$ equals $\widehat{f}(\frac{n}{m})$ if m divides n and zero otherwise.

EXERCISE 4.73 (Legendre polynomials). Define $p_n(x) = \frac{d^n}{dx^n}[(1-x^2)^n]$ for $n = 0, 1, \ldots$ and

$$\phi_n(x) = p_n(x) \cdot \Big(\int_{-1}^{1} p_n(t)^2 dt \Big)^{-1/2}. \tag{4.152}$$

Show that $(\phi_n)_{n=0,1,\ldots}$ is a complete orthonormal system on $[-1,1]$.

EXERCISE 4.74. Let f be 1-periodic and k times continuously differentiable. Prove that there exists a constant $c > 0$ such that

$$|\widehat{f}(n)| \leq c|n|^{-k} \quad \text{for all } n \in \mathbb{Z}. \tag{4.153}$$

Hint: What can you say about the Fourier coefficients of $f^{(k)}$?

EXERCISE 4.75. Let f be 1-periodic and continuous.
(i) Suppose that $\widehat{f}(n) = -\widehat{f}(-n) \geq 0$ holds for all $n \geq 0$. Prove that

$$\sum_{n=1}^{\infty} \frac{\widehat{f}(n)}{n} < \infty. \tag{4.154}$$

(ii) Show that there does not exist a 1-periodic continuous function f such that

$$\widehat{f}(n) = \frac{\operatorname{sgn}(n)}{\log|n|} \quad \text{for all } |n| \geq 2. \tag{4.155}$$

Here $\operatorname{sgn}(n) = 1$ if $n > 0$ and $\operatorname{sgn}(n) = -1$ if $n < 0$.

EXERCISE 4.76. Suppose that f is a 1-periodic function such that there exists $c > 0$ and $\alpha \in (0, 1]$ such that

$$|f(x) - f(y)| \leq c|x - y|^\alpha \tag{4.156}$$

holds for all $x, y \in \mathbb{R}$. Show that the sequence of partial sums $S_N f(x) = \sum_{n=-N}^{N} \widehat{f}(n) e^{2\pi i n x}$ converges uniformly to f as $N \to \infty$.

EXERCISE 4.77. Let $f \in C([0, 1])$ and $\mathcal{A} \subset C([0, 1])$ dense. Suppose that

$$\int_0^1 f(x)\overline{a(x)}dx = 0 \tag{4.157}$$

for all $a \in \mathcal{A}$. Show that $f = 0$.
Hint: Show that $\int_0^1 |f(x)|^2 dx = 0$.

EXERCISE 4.78. Let $f \in C([-1, 1])$ and $a \in [-1, 1]$. Show that for every $\varepsilon > 0$ there exists a polynomial p such that $p(a) = f(a)$ and $|f(x) - p(x)| < \varepsilon$ for all $x \in [-1, 1]$.

EXERCISE 4.79. Prove that

$$-\frac{1}{2} = \sum_{n=1}^{\infty} (-1)^n \frac{\sin(n)}{n}. \tag{4.158}$$

EXERCISE 4.80. Suppose $f \in C([1, \infty))$ and $\lim_{x\to+\infty} f(x) = a$. Show that f can be uniformly approximated on $[1, \infty)$ by functions of the form $g(x) = p(1/x)$, where p is a polynomial.

EXERCISE 4.81 (Stone-Weierstrass for finite sets). Let K be a finite set and $\mathcal{A}$ a family of functions on K that is an algebra (i.e. closed under taking finite linear combinations and products), separates points and vanishes nowhere. Give a purely algebraic proof that $\mathcal{A}$ must then already contain every function on K. (That means your proof is not allowed to use the concept of an inequality. In particular, you are not allowed to use any facts about metric spaces such as the Stone-Weierstrass theorem.)
Hint: Take a close look at the proof of Stone-Weierstrass.

EXERCISE 4.82 (Uniform approximation by neural networks). Let $\sigma(t) = e^t$ for $t \in \mathbb{R}$. Fix $n \in \mathbb{N}$ and let $K \subset \mathbb{R}^n$ be a compact set. As usual, let $C(K)$ denote the space of real-valued continuous functions on K. Define a class of functions $\mathcal{N} \subset C(K)$ by saying that $\mu \in \mathcal{N}$ iff there exist $m \in \mathbb{N}$, $W \in \mathbb{R}^{m\times n}$, $v, b \in \mathbb{R}^m$ such that

$$\mu(x) = \sum_{i=1}^{m} \sigma((Wx)_i + b_i)v_i \text{ for all } x \in K. \tag{4.159}$$

Prove that $\mathcal{N}$ is dense in $C(K)$.
Remark. This is a special case of a well-known result of G. Cybenko, *Approximation by Superpositions of a Sigmoidal Function* in Math. Control Signals Systems (1989). As a real-world motivation for this problem, note that a function $\mu \in \mathcal{N}$ can be interpreted as a neural network with a single hidden layer, see Figure 5. Consequently, in this problem you are asked to show that every continuous function can be uniformly approximated by neural networks of this form.

EXERCISE 4.83. Let f be a continuous function on $[0, 1]$ and N a positive integer. Define $x_k = \frac{k}{N}$ for $k = 0, \ldots, N$. Define

$$L_N(x) = \sum_{j=0}^{N} f(x_k) \prod_{j=0, j\neq k}^{N} \frac{x - x_j}{x_k - x_j}. \tag{4.160}$$

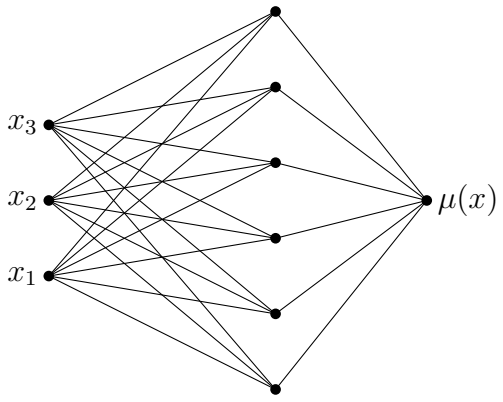

FIGURE 5. Visualization of μ when $n = 3$ and $m = 6$.

(i) Show that $f(x_k) = L_N(x_k)$ for all $k = 0, \ldots, N$ and that L_N is the unique polynomial of degree $\leq N$ with this property.
(ii) Suppose $f \in C^{N+1}([0,1])$. Show that for every $x \in [0,1]$ there exists $\xi \in [0,1]$ such that

$$f(x) - L_N(x) = \frac{f^{(N+1)}(\xi)}{(N+1)!} \prod_{k=0}^{N} (x - x_k). \tag{4.161}$$

(iii) Show that L_N does not necessarily converge to f uniformly on $[0,1]$. (Find a counterexample.)
(iv) Suppose f is given by a power series with infinite convergence radius. Does L_N necessarily converge to f uniformly on $[0,1]$?
Remark. The polynomials L_N are also known as *Lagrange interpolation polynomials* .

CHAPTER 5

From Riemann to Lebesgue*

1. Lebesgue null sets

For a compact interval $I = [c, d]$ we call $d - c$ the length of I, also denoted by $\ell(I)$.

DEFINITION 5.1. A set $E \subset [a, b]$ is called a Lebesgue null set if for every $\varepsilon > 0$ there is a sequence $(I_n)_{n\in\mathbb{N}}$ of intervals such that $E \subset \cup_{n\in\mathbb{N}} I_n$ and $\sum_{n=1}^{\infty} \ell(I_n) < \varepsilon$.

LEMMA 5.2. *Countable unions of Lebesgue null sets are Lebesgue null sets.*

PROOF. Let E_k, $k \in \mathbb{N}$ be Lebesgue nullsets. For each k find a countable family of intervals $\{I_{k,n}\}_{n=1}^{\infty}$ such that $\sum_{n=1}^{\infty} \ell(I_{k,n}) < \varepsilon 2^{-k-1}$ and $E_k \subset \cup_{n=1}^{\infty} I_{k,n}$.

The family of intervals $\{I_{k,n}\}_{(k,n)\in\mathbb{N}^2}$ is countable (and thus cn be arranged in a series) and we have

$$\sum_{k=1}^{\infty}\sum_{n=1}^{\infty} \ell(I_{k,n}) < \sum_{k=1}^{\infty} \varepsilon 2^{-k-1} = \varepsilon/2. \tag{5.1}$$

□

EXERCISE 5.3. Which theorems about series with nonnegative terms have been used in this proof?

DEFINITION 5.4. A set $E \subset [a, b]$ has content zero if for every $\varepsilon > 0$ there is a finite set of intervals $I_1, \ldots, I_N$ such that $E \subset \cup_{n=1}^{N} I_n$ and $\sum_{n=1}^{N} \ell(I_n) < \varepsilon$.

Note that any set of content zero is a Lebesgue null set, but the converse is not true (see Exercise 5.8 below).

LEMMA 5.5. *Let $\{I_\nu\}_{\nu=1}^{N}$ be a finite collection of intervals such that $[a, b] \subset \cup_{\nu=1}^{N} I_j$. Then $\sum_{\nu=1}^{N} \ell(I_\nu) \geq b - a$. In particular $[a, b]$ does not have content zero.*

PROOF. Let $J_\nu := \overline{I}_\nu \cap [a, b]$. Arrange the finite set formed by all the endpoint of those intervals in increasing order, written as $a \leq x_1 \leq \cdots < x_M = b$. Then every interval $[x_{i-1}, x_i]$ is contained in at least J_ν.

Define inductively sets of indices $\mathcal{J}_\nu$. For $\nu = 1$ set

$$\mathcal{J}_1 = \big\{i \in \{1, \ldots, M\} : [x_{i-1}, x_i] \subset J_1\big\}. \tag{5.2}$$

For any $\nu > 1$ we are either in the situation that $\mathcal{J}_1 \cup \cdots \cup \mathcal{J}_{\nu-1}$ contains all $i \in \{1, \ldots, M\}$ (then we stop the construction) or, if not, then we form

$$\mathcal{J}_\nu = \big\{i \in \{1, \ldots, M\} : [x_{i-1}, x_i] \subset J_\nu \text{ and } [x_{i-1}, x_i] \subsetneq J_l \text{ for } l \leq \nu - 1\ \big\}. \tag{5.3}$$

The construction stops after K steps, where $K \leq N$. Note that each index i is in exactly one family $\mathcal{J}_\nu$ and also for each ν we have

$$\sum_{i\in\mathcal{J}_\nu} (x_i - x_{i-1}) \leq \ell(J_\nu). \tag{5.4}$$

Consequently

$$b-a=\sum_{i=1}^{N}(x_i-x_{i-1})=\sum_{\nu=1}^{K}\sum_{i\in\mathcal{J}_\nu}(x_i-x_{i-1})\leq\sum_{\nu=1}^{K}\ell(J_\nu)\leq\sum_{\nu=1}^{N}\ell(I_\nu). \tag{5.5}$$ □

LEMMA 5.6. *Let E be a compact Lebesgue null set. Then E has content zero.*

PROOF. Let $\varepsilon>0$. Since E is a null set there is a countable family $\{I_\nu\}_{\nu\in\mathbb{N}}$ of closed intervals such that $\sum_{\nu=1}^{\infty}\ell(I_\nu)<\varepsilon/2$. Write $I_\nu=[a_\nu,b_\nu]$ and form the slightly larger open intervals $\widetilde{I}_\nu=(a_\nu-\varepsilon 2^{-\nu-2},b_\nu+\varepsilon 2^{-\nu-2})$ so that $\ell(\widetilde{I}_\nu)=\ell(I_\nu)+\varepsilon 2^{-\nu-1}$ and thus

$$\sum_{\nu=1}^{\infty}\ell(\widetilde{I}_\nu)\leq\sum_{\nu=1}^{\infty}\ell(I_\nu)+\sum_{\nu=1}^{\infty}\varepsilon 2^{-\nu-1}<\varepsilon/2+\varepsilon/2=\varepsilon. \tag{5.6}$$

Since E is compact we may choose finitely many $\widetilde{I}_{\nu_1},\dots,\widetilde{I}_{\nu_M}$ such that $E\subset\cup_{l=1}^{M}\widetilde{I}_{\nu_l}$ and $\sum_{l=1}^{M}\ell(\widetilde{I}_{\nu_l})\leq\sum_{\nu=1}\ell(\widetilde{I}_\nu)<\varepsilon$. Hence E has content zero. □

COROLLARY 5.7. *Let $a<b$. Then $[a,b]$ is not a Lebesgue null set.*

PROOF. This is an immediate consequence of from Lemma 5.5 together with Lemma 5.6. □

EXERCISE 5.8. Let E be the set of rational numbers in $[a,b]$. Show that E is a Lebesgue null set but E is not of content zero.

The Lebesgue null sets are usually called *sets of Lebesgue measure zero* . We avoid this terminology here because we have not defined Lebesgue measure here and indeed have not identified the class of sets on which it can be defined (the so called Lebesgue measurable sets). A substitute for Lebesgue measure which can be defined on all subsets of $\mathbb{R}$ is Lebesgue outer measure:

DEFINITION 5.9. For a subset of $\mathbb{R}$ the Lebesgue outer measure $\lambda_*(E)$ of E is defined as the quantity $\lambda_*(E)=\inf\sum_{n=1}^{\infty}\ell(I_n)$ where the infimum is taken over all countable collections $\{I_n\}_{n\in\mathbb{N}}$ of intervals which have the property that $E\subset\cup_{n=1}^{\infty}I_n$.

With this definition, the Lebesgue null sets are simply the sets of Lebesgue outer measure zero.

2. Lebesgue's Characterization of the Riemann integral

We can now formulate the main theorem of this chapter.

THEOREM 5.10. *Let $f:[a,b]\to\mathbb{R}$ be a bounded function. Then f is Riemann integrable if and only if the set of discontinuities of f,*

$$D_f:=\{x\in[a,b]:\ f \text{ is not continuous at } x\},$$

is a Lebesgue null set.

The following lemma linking oscillation to lower and upper sums is very helpful in the proof of Theorem 5.10.

LEMMA 5.11. *Let $f:[a,b]\to\mathbb{R}$ be a bounded function and assume that $\operatorname{osc}_f(x)<\gamma$ for all $x\in[a,b]$. Then there is a partition P of $[a,b]$ such that $U(f,P)-L(f,P)<\gamma(b-a)$.*

PROOF. By definition of $\mathrm{osc}_f(x)$ we can find a $\delta_x > 0$ such that

$$M_{f,2\delta_x}(x) - m_{f,2\delta_x}(x) < \gamma. \tag{5.7}$$

Since $[a.b]$ is compact we find x_1, ..., x_N such that $[a,b]$ is contained in the union of the intervals $(x_i - \delta_{x_i}, x_i + \delta_{x_i})$. Consider the finite set consisting of the a, b the x_i, the corresponding point $x_i - \delta_{x_i}$ and $x_i + \delta_{x_i}$ and then discard those point which do not lie in $[a,b]$. The resulting set P is a partition of $[a,b]$ with nodes $a = t_0 < \cdots < t_M = b$ and if t_{i-1}, t_i are consecutive nodes in this partition then

$$\sup\{f(t) : t \in [t_{i-1}, t_i]\} - \inf\{f(t) : t \in [t_{i-1}, t_i]\} < \gamma.$$

Hence

$$U(f,P) - L(f,P) < \gamma \sum_{i=1}^{M} (t_i - t_{i-1}) = \gamma(b-a)$$

and the lemma is proved. □

PROOF OF THEOREM 5.10. *Part 1: Set of discontinuities is a null set $\implies$ f is Riemann integrable.* By Lemma 4 it suffices to construct, for given $\varepsilon > 0$, a partition $\mathcal{P}$ such that

$$U(f,\mathcal{P}) - L(f,\mathcal{P}) < \varepsilon. \tag{5.8}$$

The function f is bounded and thus there is $C > 0$ such that $|f(x)| \le C$ for $x \in [a,b]$.

Now let $\varepsilon_1 \ll \varepsilon$ depending on ε; we will see (only at the end) that

$$\varepsilon_1 = \frac{\varepsilon}{2C + b - a}$$

is an appropriate choice. Consider the set

$$D(\varepsilon_1) = \{x \in [a,b] : \mathrm{osc}_f(x) \ge \varepsilon_1\}. \tag{5.9}$$

$D(\varepsilon_1)$ is a Lebesgue null set since $D(\varepsilon_1) \subset D_f$ and D_f is a Lebesgue null set. Also, $D(\varepsilon_1)$ is a closed subset of $[a,b]$, and thus compact and thus has *content zero.*

Thus there is a *finite* collection $\{I_\nu\}_{\nu=1}^N$ of closed intervals such that $\sum_{\nu=1}^N \ell(I_\nu) < \varepsilon_1$ and $D(\varepsilon_1) \subset \cup_{\nu=1}^N (I_\nu)^\circ$ (where $(I_\nu)^\circ$ denotes the interior of I_ν).

We may choose a partition $P = \{a = x_0 < \cdots < x_N = b\}$ such that each index i belongs to (at least) one of the following sets:

$$\mathcal{J}_1 = \{i : [x_{i-1}, x_i] \subset I_\nu \text{ for some } \nu \text{ in } [1,N]]\}$$
$$\mathcal{J}_2 = \{i : [x_{i-1}, x_i] \cap D(\varepsilon_1) = \emptyset\}.$$

Regarding the intervals $[x_{i-1}, x_i]$ with $i \in \mathcal{J}_1$ we have

$$\sum_{i \in \mathcal{J}_1} (x_i - x_{i-1}) \le \sum_{\nu=1}^{N} \sum_{\substack{i: \\ [x_{i-1},x_i] \subset I_\nu}} (x_i - x_{i-1}) \le \sum_{\nu=1}^{N} \ell(I_\nu) < \varepsilon_1. \tag{5.10}$$

We observe that for all $i \in \mathcal{J}_2$ we have $\mathrm{osc}_f(x) < \varepsilon_1$ for all $x \in [x_i, x_{i+1}]$. Thus by Lemma 5.11, we find a partition P_i of $[x_{i-1}, x_i]$, labeled $\{x_i = x_{i,0}, \ldots, x_{i,N_i} := x_{i+1}\}$,

such that with

$$U^i(f,P_i) := \sum_{j=1}^{N_i}(x_{i,j}-x_{i,j-1}) \sup_{[x_{i,0},x_{i,N_i}]} f(x),$$

$$L^i(f,P_i) := \sum_{j=1}^{N_i}(x_{i,j}-x_{i,j-1}) \inf_{[x_{i,0},x_{i,N_i}]} f(x)$$

we have

$$U^i(f,P_i) - L^i(f,P_i) < \varepsilon_1(x_i - x_{i-1}). \tag{5.11}$$

Now the desired partition is defined by

$$\mathcal{P} = \{x_i : i \in \mathcal{J}_1\} \cup \bigcup_{i\in\mathcal{J}_2}\{x_{i,0},\dots,x_{i,N_i}\} \tag{5.12}$$

and we can split and then estimate

$$\begin{aligned} U(f,\mathcal{P}) - L(f,\mathcal{P}) &= \sum_{i\in\mathcal{J}_1}\Big(\sup_{[x_{i-1},x_i]} f - \inf_{[x_{i-1},x_i]} f\Big)(x_i-x_{i-1}) + \sum_{i\in\mathcal{J}_2}(U^i(f,P_i)-L^i(f,P_i)) \\ &\le 2C\sum_{i\in\mathcal{J}_1}(x_i-x_{i-1}) + \sum_{i\in\mathcal{J}_2}(U^i(f,P_i)-L^i(f,P_i)). \end{aligned}$$

By (5.10) and (5.11) we get

$$U(f,\mathcal{P}) - L(f,\mathcal{P}) < 2C\varepsilon_1 + \varepsilon_1\sum_{i\in\mathcal{J}_2}(x_i-x_{i-1}) \le 2C\varepsilon_1 + (b-a)\varepsilon_1. \tag{5.13}$$

In view of our choice $\varepsilon_1 = \varepsilon/(2C+b-a)$ we have proved the desired inequality (5.8).

Part 2: f is Riemann integrable $\implies$ Set of discontinuities is a null set.

For each $n \in \mathbb{N}$ we define $D^n = \{x \in [a,b] : \operatorname{osc}_f(x) \ge 1/n\}$. Observe that $D_f = \cup_{n=1}^{\infty} D^n$, by Lemma 1.93. Thus by Lemma 5.2 it suffices to show that each D^n is a Lebesgue null set.

Fix $n \in \mathbb{N}$. Since f is Riemann integrable there exists, by Lemma 4 a partition $P = \{x_0 < \cdots < x_N\}$ of $[a,b]$ such that

$$U(f,P) - L(f,P) < \varepsilon/n.$$

Let $\mathcal{J}$ be the set of indices i for which $I_i := [x_{i-1},x_i]$ contains a point in D^n, so that $D^n \subset \cup_{i\in\mathcal{J}} I_i$. Clearly we have

$$M_i(f) - m_i(f) \ge \inf_{x\in[x_{i-1},x_i]} \operatorname{osc}_f(x) \ge \frac{1}{n} \text{ if } i \in \mathcal{J}. \tag{5.14}$$

Hence

$$\begin{aligned} \sum_{i\in\mathcal{J}} \ell(I_i) &= \sum_{I\in\mathcal{J}} \frac{(M_i(f)-m_i(f))(x_i-x_{i-1})}{M_i(f)-m_i(f)} \\ &\le \frac{1}{1/n}\sum_{i=1}^{N}(M_i(f)-m_i(f))(x_i-x_{i-1}) < n\cdot\frac{\varepsilon}{n} = \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary we have proved that each D^n is a Lebesgue null set, and thus D_f is a Lebesgue null set. □

CHAPTER 6

The Baire category theorem*

Let (X, d) be a metric space. Recall that the *interior* A^o of a set $A \subset X$ is the set of interior points of A, i.e. the set of all $x \in A$ such that there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subset A$. A set $A \subset X$ is *dense* if $\overline{A} = X$. Note that A is dense if and only if for all non-empty open sets $U \subset X$ we have $A \cap U \neq \emptyset$.

DEFINITION 6.1. A set $A \subset X$ is called *nowhere dense* if its closure has empty interior. In other words, if $\overline{A}^o = \emptyset$. Equivalently, A is nowhere dense if and only if $\overline{A}$ contains no non-empty open set.

Note that 1. A closed set $A \subset X$ has empty interior if and only if $A^c = X \setminus A$ is open and dense. (This is because A is closed if and only if A^c is open and A has empty interior if and only if A^c is dense.)
2. A is nowhere dense if and only if A^c contains an open dense set.
3. A is nowhere dense if and only if A is contained in a closed set with empty interior.

EXAMPLE 6.2. The Cantor set

$$\mathfrak{C} = [0,1] \setminus \bigcup_{\ell=0}^{\infty} \bigcup_{k=0}^{3^\ell - 1} \left(\tfrac{3k+1}{3^{\ell+1}}, \tfrac{3k+2}{3^{\ell+1}}\right) \tag{6.1}$$

is a closed subset of $[0, 1]$ and has empty interior. Therefore, it is nowhere dense.

LEMMA 6.3. *Suppose $A_1, \dots, A_n \subset X$ are nowhere dense sets. Then $\bigcup_{k=1}^n A_k$ is nowhere dense.*

PROOF. Without loss of generality let $n = 2$. We need to show that $\overline{A_1} \cup \overline{A_2}$ has empty interior. Equivalently, setting $U_k = \overline{A_k}^c$ for $k = 1, 2$. We show that $U_1 \cap U_2$ is dense. Let $U \subset X$ be a non-empty open set. Then $V_1 = U \cap U_1$ is open and non-empty, because U_1 is dense. Since U_2 is also dense, $V_1 \cap U_2 = U \cap (U_1 \cap U_2)$ is non-empty, so $U_1 \cap U_2$ is dense. □

Also, a subset of a nowhere dense set is nowhere dense and the closure of a nowhere dense set is nowhere dense.

However, countable unions of nowhere dense sets are not necessarily nowhere dense sets.

EXAMPLE 6.4. Enumerate the rationals as $\mathbb{Q} = \{q_1, q_2, \dots\}$. For every $k = 1, 2, \dots$, the set $A_k = \{q_k\}$ is nowhere dense in $\mathbb{R}$. But $\mathbb{Q} = \bigcup_{k=1}^\infty A_k \subset \mathbb{R}$ is not nowhere dense (it is dense!).

DEFINITION 6.5. A set $A \subset X$ is called *meager* (or of *first category*) in X if it is the countable union of nowhere dense sets. A is called *comeager* (or *residual* or of *second category*) if A^c is meager.

The above example shows that $\mathbb{Q} \subset \mathbb{R}$ is meager. In fact, every countable subset of $\mathbb{R}$ is meager (because single points are nowhere dense in $\mathbb{R}$).

By definition, countable unions of meager sets are meager. The choice of the word "meager" suggests that meager sets are somehow "small" or "negligible". But how "large" can meager sets be? For example, can X be meager? That is, can we write the entire metric space X as a countable union of nowhere dense subsets? The Baire category theorem will show that the answer is no, if X is complete.

THEOREM 6.6 (Baire category theorem). *In a complete metric space, meager sets have empty interior. Equivalently, countable intersections of open dense sets are dense.*

COROLLARY 6.7. *Let X be a complete metric space and $A \subset X$ a meager set. Then $A \neq X$. In other words, X is not a meager subset of itself.*

EXAMPLE 6.8. The conclusion of the Baire category theorem fails if we drop the assumption that X is complete. Consider $X = \mathbb{Q}$ with the metric inherited from $\mathbb{R}$ (so $d(p, q) = |p - q|$). Then X is a meager subset of itself because it is countable and single points are nowhere dense in X (X has no isolated points). But the interior of X is non-empty, because X is open in X.

EXAMPLE 6.9. Not every set with empty interior is meager: consider the irrational numbers $A = \mathbb{R} \setminus \mathbb{Q}$. A has empty interior, because $A^c = \mathbb{Q}$ is dense. It is not meager, because otherwise $\mathbb{R} = A \cup A^c$ would be meager, which contradicts the Baire category theorem.

EXERCISE 6.10. Another notion of "smallness" is the following:
Definition. A set $A \subset \mathbb{R}$ is called a *Lebesgue null set* if for every $\varepsilon > 0$ there exist intervals $I_1, I_2, \ldots$ such that

$$A \subset \bigcup_{j=1}^{\infty} I_j \text{ and } \sum_{j=1}^{\infty} |I_j| \leq \varepsilon. \tag{6.2}$$

(Here $|I|$ denotes the length of the interval I.)
Give an example of a comeager Lebesgue null set. (Recall that a set is called *comeager* if its complement is meager.)
(This implies in particular that Lebesgue null sets are not necessarily meager and meager sets are not necessarily Lebesgue null sets.)

For the proof of Theorem 6.6 we will need the following lemma.

LEMMA 6.11. *Let X be complete and $A_1 \supset A_2 \supset \cdots$ a decreasing sequence of non-empty closed sets in X such that*

$$\operatorname{diam} A_n = \sup_{x,y \in A_n} d(x, y) \longrightarrow 0 \tag{6.3}$$

as $n \to \infty$. Then $\bigcap_{n=1}^{\infty} A_n$ is non-empty.

PROOF OF LEMMA 6.11. For every $n \geq 1$ we choose $x_n \in A_n$. Then $(x_n)_{n \in \mathbb{N}}$ is a Cauchy sequence, because for all $n \geq m$ we have $d(x_n, x_m) \leq \operatorname{diam} A_m \to 0$ as $m \to \infty$. Since X is complete, there exists $x \in X$ such that $\lim_{n \to \infty} x_n = x$. Let $N \in \mathbb{N}$. Then A_N contains the sequence $(x_n)_{n \geq N}$ and since A_N is closed, it must also contain the limit of this sequence, so $x \in A_N$. This proves that $x \in \bigcap_{N=1}^{\infty} A_N$. □

PROOF OF THEOREM 6.6. Let $(U_n)_{n\in\mathbb{N}}$ be open dense sets. We need to show that $\bigcap_{n=1}^{\infty} U_n$ is dense. Let $U \subset X$ be open and non-empty. It suffices to show that $U \cap \bigcap_{n=1}^{\infty} U_n$ is non-empty. Since U_1 is open and dense, $U \cap U_1$ is open and non-empty. Choose a closed ball $\overline{B}(x_1, r_1) \subset U \cap U_1$ with $r_1 \in (0, 1)$. Then $B(x_1, r_1) \cap U_2$ is open and non-empty (because U_2 is dense), so we can choose a closed ball $\overline{B}(x_2, r_2) \subset B(x_1, r_1) \cap U_2$ with $r_2 \in (0, \frac{1}{2})$. Iterating this process, we obtain a sequence of closed balls $(\overline{B}(x_n, r_n))_n$ such that $\overline{B}(x_n, r_n) \subset B(x_{n-1}, r_{n-1}) \cap U_n$ and $r_n \in (0, \frac{1}{n})$. By Lemma 6.11 there exists a point x contained in $\bigcap_{n=1}^{\infty} \overline{B}(x_n, r_n)$. Since $\overline{B}(x_n, r_n) \subset U \cap U_n$ for all $n \geq 1$, we have $x \in U \cap \bigcap_{n=1}^{\infty} U_n$. □

The Baire category theorem has a number of interesting consequences.

1. Nowhere differentiable continuous functions*

THEOREM 6.12. *Let $\mathcal{A} \subset C([0, 1])$ be the set of all functions that are differentiable at at least one point in $[0, 1]$. Then $\mathcal{A}$ is meager.*

PROOF. For $n \in \mathbb{N}$ we define $\mathcal{A}_n$ to be the set of all $f \in C([0, 1])$ such that there exists $t \in [0, 1]$ such that

$$\Big|\frac{f(t+h) - f(t)}{h}\Big| \leq n \tag{6.4}$$

holds for all $h \in \mathbb{R}$ with $t + h \in [0, 1]$. Then

$$\mathcal{A} \subset \bigcup_{n=1}^{\infty} \mathcal{A}_n. \tag{6.5}$$

It suffices to show that each $\mathcal{A}_n$ is nowhere dense. We first prove that $\mathcal{A}_n$ is closed. Let $(f_k)_{k\in\mathbb{N}} \subset \mathcal{A}_n$ be a sequence that converges to some $f \in C([0, 1])$. We show that $f \in \mathcal{A}_n$. Indeed, by assumption, there exists $(t_k)_{k\in\mathbb{N}} \subset [0, 1]$ such that

$$\Big|\frac{f_k(t_k + h) - f_k(t_k)}{h}\Big| \leq n \tag{6.6}$$

holds for all $k \geq 1$ if $t_k + h \in [0, 1]$. By the Bolzano-Weierstrass theorem, we may assume without loss of generality that $(t_k)_{k\in\mathbb{N}}$ converges to some $t \in [0, 1]$ (by passing to a subsequence). Then, by continuity of f,

$$\Big|\frac{f(t+h) - f(t)}{h}\Big| = \lim_{k\to\infty} \Big|\frac{f_k(t_k + h) - f_k(t_k)}{h}\Big| \leq n. \tag{6.7}$$

Therefore, $f \in \mathcal{A}_n$ and $\mathcal{A}_n$ is closed. Also, $\mathcal{A}_n$ has empty interior. Indeed, one can see that $C([0, 1]) \setminus \mathcal{A}_n$ is dense because every $f \in C([0, 1])$ can be uniformly approximated by a function that has arbitrarily large slope (think of "sawtooth" functions).

EXERCISE 6.13. Provide the details of this argument: show that $\mathcal{A}_n$ has empty interior.

□

The Baire category theorem implies that $\mathcal{A}$ has empty interior. In other words, the set of nowhere differentiable functions $C([0, 1]) \setminus \mathcal{A}$ is dense. In this sense, it is "generic" behavior for continuous functions to be nowhere differentiable. In particular, we can conclude that there exists $f \in C([0, 1]) \setminus \mathcal{A}$ (so f is nowhere differentiable) without actually constructing such a function. On the other hand, one can also give explicit examples of nowhere differentiable functions.

EXAMPLE 6.14 (Weierstrass' function). Consider the function $f \in C([0,1])$ defined as

$$f(x) = \sum_{n=0}^{\infty} b^{-n\alpha} \sin(b^n x), \tag{6.8}$$

where $0 < \alpha < 1$ and $b > 1$ are fixed. The function f is indeed continuous because the series is uniformly convergent. In fact, f is the uniform limit of the sequence of functions $(f_N)_N$ considered in Exercise 1.105.

EXERCISE 6.15. Show that f is nowhere differentiable.

2. Sets of continuity*

DEFINITION 6.16. Let X, Y be metric spaces and $f : X \to Y$ a map. The set

$$C_f = \{x \in X \,:\, f \text{ is continuous at } x\} \subset X \tag{6.9}$$

is called the *set of continuity of* f . Similarly, $D_f = X \setminus C_f$ is called the *set of discontinuity of* f .

EXAMPLE 6.17. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = 1$ if x is rational and $f(x) = 0$ if x is irrational. Then $C_f = \emptyset$.

EXAMPLE 6.18. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x$ if x is rational and $f(x) = 0$ if x is irrational. Then $C_f = \{0\}$.

EXAMPLE 6.19. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined as follows: we set $f(0) = 1$ and if $x \in \mathbb{Q} \setminus \{0\}$, then we let $f(x) = 1/q$, where $x = \frac{p}{q}$, where $p \in \mathbb{Z}$, $q \in \mathbb{N}$ and the greatest common divisor of p and q is one. If $x \notin \mathbb{Q}$, then we let $f(x) = 0$. We claim that $C_f = \mathbb{R} \setminus \mathbb{Q}$. Indeed, say $x \in \mathbb{R} \setminus \mathbb{Q}$ and $p_n/q_n \to x$ a rational approximation. Then $q_n \to \infty$ (otherwise, it must converge and then x would be rational). This implies that f is continuous at x. On the other hand, say $x \in \mathbb{Q}$. Set $x_n = x + \frac{\sqrt{2}}{n}$. Then $x_n \notin \mathbb{Q}$ because $\sqrt{2} \notin \mathbb{Q}$, so $f(x_n) = 0$ for all n, so $\lim_{n\to\infty} f(x_n) = 0$, but $f(x) \neq 0$. Hence f is not continuous at x.

It is natural to ask which subsets of X arise as the set of continuity of some function on X. For instance, does there exist a function $f : \mathbb{R} \to \mathbb{R}$ such that $C_f = \mathbb{Q}$?

DEFINITION 6.20. A set $A \subset X$ is called an *F_σ-set* if it is a countable union of closed sets. A set $G \subset X$ is called a *G_δ-set* if it is a countable intersection of open sets.

These names are motivated historically. The F in F_σ is for *fermé* which is French for *closed* . On the other hand, the G in G_δ is for *Gebiet* which is German for *region* .

EXAMPLES 6.21. 1. Every open set is a G_δ-set and every closed set is an F_σ-set.
2. Let $x \in X$. Then $\{x\}$ is a G_δ-set: it is the intersection of the open balls $B(x, 1/n)$.
3. $\mathbb{Q} \subset \mathbb{R}$ is an F_σ set, because $\mathbb{Q} = \bigcup_{q\in\mathbb{Q}} \{q\}$ (a countable union of closed sets).

THEOREM 6.22. *Let X and Y be metric spaces and $f : X \to Y$ a map. Then $C_f \subset X$ is a G_δ-set and D_f is an F_σ-set.*

PROOF. Let $f : X \to Y$ be given. It suffices to show that C_f is a G_δ-set. For every $S \subset X$ we define the *oscillation of f on S* by

$$\omega_f(S) = \sup_{x,x'\in S} d_Y(f(x), f(x')) = \operatorname{diam} f(S). \tag{6.10}$$

For a point $x \in X$ we recall definition of *oscillation of f at x* by

$$\operatorname{osc}_f(x) = \inf_{\varepsilon>0} \omega_f(B(x,\varepsilon)). \tag{6.11}$$

This coincides with the definition 1.91. Then we have

$$x \in C_f \iff \operatorname{osc}_f(x) = 0 \tag{6.12}$$

and we can write the set of continuity of f as

$$C_f = \bigcap_{n=1}^{\infty} \{x \in X : \operatorname{osc}_f(x) < \tfrac{1}{n}\}. \tag{6.13}$$

We are done if we can show that $U_n = \{x \in X : \operatorname{osc}_f(x) < \frac{1}{n}\}$ is open for every $n \in \mathbb{N}$. Let $x_0 \in U_n$. Then $\operatorname{osc}_f(x_0) < \frac{1}{n}$. Therefore, there exists $\varepsilon > 0$ such that $\omega_f(B(x_0,\varepsilon)) < \frac{1}{n}$. Let $x \in B(x_0, \varepsilon/2)$. Then by the triangle inequality, $B(x,\varepsilon/2) \subset B(x_0,\varepsilon)$. Therefore,

$$\operatorname{osc}_f(x) \le \omega_f(B(x,\varepsilon/2)) \le \omega_f(B(x_0,\varepsilon)) < \tfrac{1}{n}. \tag{6.14}$$

Thus, $B(x_0,\varepsilon/2) \subset U_n$ and so U_n is open. □

As a sample application of the Baire category theorem we now answer one of our previous questions negatively:

LEMMA 6.23. *$\mathbb{Q} \subset \mathbb{R}$ is not a G_δ-set. Consequently, there exists no function $f : \mathbb{R} \to \mathbb{R}$ such that $C_f = \mathbb{Q}$.*

PROOF. Suppose $\mathbb{Q}$ is a G_δ-set. Then $\mathbb{R} \setminus \mathbb{Q}$ is an F_σ-set and therefore can be written as a countable union of closed sets $A_1, A_2, \ldots$. Since $\mathbb{R} \setminus \mathbb{Q}$ has empty interior (its complement $\mathbb{Q}$ is dense), $A_n \subset \mathbb{R} \setminus \mathbb{Q}$ also has empty interior for every n. Thus A_n is nowhere dense, so $\mathbb{R} \setminus \mathbb{Q}$ is meager. But then $\mathbb{R} = \mathbb{Q} \cup (\mathbb{R} \setminus \mathbb{Q})$ must be meager, which contradicts the Baire category theorem. □

Observe that an F_σ-set is either meager or has non-empty interior: suppose $A \subset X$ is an F_σ-set with empty interior. Then it is a countable union of closed sets with empty interior and therefore meager. Similarly, a G_δ-set is either comeager or not dense.

REMARK. It is natural to ask if the converse of Theorem 6.22 is true in the following sense: given a G_δ-set $G \subset X$, can we find a function $f : X \to \mathbb{R}$ such that $C_f = G$? This cannot hold in general: suppose X contains an isolated point, that is X contains an open set of the form $\{x\}$. Then necessarily $x \in C_f$, but x is not necessarily contained in every possible G_δ-set. However, this turns out to be the only obstruction: if X contains no isolated points, then for every G_δ-set $G \subset X$ one can find $f : X \to \mathbb{R}$ such that $C_f = G$. For a very short proof of this, see *S. S. Kim: A Characterization of the Set of Points of Continuity of a Real Function. Amer. Math. Monthly 106 (1999), no. 3, 258—259.*

3. Baire functions*

Consider again the Dirichlet function D

$$D(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q}. \end{cases} \tag{6.15}$$

It is natural to ask whether D is the pointwise limit of a sequence of continuous functions. The answer turns out to be no.

We start with a definition.

DEFINITION 6.24. Let X be a metric space.

A function $f : X \to \mathbb{R}$ is a *Baire-1 function* if there is a sequence of continuous functions $f_n : X \to \mathbb{R}$ such that $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in X$.

Clearly a Baire-1 function does not have to be continuous everywhere in X. However the following theorem shows that f will be continuous on a residual set.

THEOREM 6.25. *Let X be a complete metric space and let $f : X \to \mathbb{R}$ be a Baire-1 function. Then the set $C_f = \{x : f \text{ is continuous at } x\}$ is a dense G_δ-set.*

In the proof of this theorem we shall apply the Baire category theorem twice. The first application is used in the following lemma.

LEMMA 6.26. *Let Y be a complete metric space let $f_n : Y \to \mathbb{R}$ be continuous functions on Y converging pointwise to f. Then for every $\alpha > 0$ there is an $N \in \mathbb{N}$ and an open ball B such that*

$$|f_n(x) - f(x)| \le \alpha \tag{6.16}$$

for all $n \ge N$ and all $x \in B$.

PROOF. Let

$$A_n = \{x \in Y : |f_n(x) - f_k(x)| \le \alpha \text{ for all } k > n\}. \tag{6.17}$$

First observe that A_n is a closed set; indeed the set $E_{n,k} = \{x : |f_n(x) - f_k(x)| \le \alpha\}$ is closed since $|f_n - f_k|$ is continuous and we have $A_n = \cap_{k=n+1}^{\infty} E_{n,k}$.

Since f_n converges pointwise to f we have that $|f_n(x) - f(x)| \le \alpha$ for $x \in A_n$, and $Y = \cup_{n=1}^{\infty} A_n$. By Baire's theorem there exists $n \in \mathbb{N}$ such that A_n is not nowhere dense and since A_n is closed there is an open ball $B \subset A_n$. We have thus proved the assertion. □

PROOF OF THEOREM 6.25, CONCLUSION. We recall the definition of oscillation $\omega_f(A)$ and the definition of oscillation $\mathrm{osc}_f(x)$ in the previous section. Consider the open sets $W_M = \{x : \mathrm{osc}_f(x) < 1/M\}$ and we have $C_f = \cap_{M=1}^{\infty} W_M$. We show that W_M is dense in X. Let B_0 be any open ball; we need to show that its intersection with W_M is not empty. Let B_1 be an open ball such that $\overline{B_1} \subset B_0$. We apply Lemma 6.26 with the complete metric space $\overline{B}_1$, and $\alpha = (4M)^{-1}$ and find an open ball B_2 (which is an open ball in X contained in B_0) and $n \in \mathbb{N}$ such that $|f_n(x) - f(x)| \le (4M)^{-1}$ for all $x \in B_2$. Since f_n is continuous we find an open ball $B_3 \subset B_2$ such that $|f_n(x) - f_n(y)| < (4M)^{-1}$ for all $x, y \in B_3$. Hence

$$|f(x) - f(y)| \le |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| \le \tfrac{3}{4M} < \tfrac{1}{M} \tag{6.18}$$

for $x, y \in B_3$ and therefore $\mathrm{osc}_f(x) < 1/M$ for all $x \in B_3$. Thus $B_3 \subset W_M$.

We have identified the G_δ-set C_f as a countable intersection of dense open sets, hence it is dense set by Baire's theorem, and thus a residual set. □

REMARK. Baire considered a hierarchy of increasing classes that we refer to as Baire-n classes: One defines the Baire-0 class as the class of continuous functions on X. Then for $n \ge 1$ the Baire-n class consists of the pointwise limits of sequences of functions in the Baire-$(n-1)$ class.

As an illustration consider an enumeration $\{r_n\}$ of the rational numbers and define the function D_n so that $D_n(r_k) = 1$ for $1 \le k \le n$ and $D_n(x) = 0$ elsewhere. Verify that the functions D_n are Baire-1 and that D_n converges pointwise to the Dirichlet function

in (6.15); this identifies D as a Baire-2 function which by Theorem 6.25 is not Baire-1. Alternatively one can also use the formula $D(x) = \lim_{j\to\infty}(\lim_{m\to\infty}(\cos(j!\pi x))^{2m})$ to show that D is Baire-2.

4. The uniform boundedness principle*

The following theorem is one of the cornerstones of functional analysis and is a direct application of the Baire category theorem.

THEOREM 6.27 (Banach-Steinhaus). *Let X be a Banach space and Y a normed vector space. Let $\mathcal{F} \subset L(X,Y)$ be a family of bounded linear operators. Then*

$$\sup_{T\in\mathcal{F}} \|Tx\|_Y < \infty \text{ for all } x \in X \quad \Longleftrightarrow \quad \sup_{T\in\mathcal{F}} \|T\|_{\mathrm{op}} < \infty. \tag{6.19}$$

In other words, a family of bounded linear operators is uniformly bounded if and only if it is pointwise bounded.

This theorem is also called the uniform boundedness principle.

PROOF. In the '$\Leftarrow$' direction there is nothing to show. Let us prove '$\Rightarrow$'. Suppose that $\sup_{T\in\mathcal{F}} \|Tx\|_Y < \infty$ for all $x \in X$. Define

$$A_n = \{x \in X \,:\, \sup_{T\in\mathcal{F}} \|Tx\|_Y \le n\} \subset X. \tag{6.20}$$

A_n is a closed set: if $(x_k)_{k\in\mathbb{N}} \subset A_n$ is a sequence with $x_k \to x \in X$, then since T is continuous, $\|Tx\|_Y = \lim_{k\to\infty} \|Tx_k\|_Y \le n$ for all $T \in \mathcal{F}$, so $x \in A_n$. Also, the assumption $\sup_{T\in\mathcal{F}} \|Tx\|_Y < \infty$ for all $x \in X$ implies that

$$X = \bigcup_{n=1}^{\infty} A_n. \tag{6.21}$$

By the Baire category theorem, X is not meager. Thus, there exists $n_0 \in \mathbb{N}$ such that A_{n_0} has non-empty interior. This means that there exists $x_0 \in A_{n_0}$ and $\varepsilon > 0$ such that

$$\overline{B}(x_0, \varepsilon) \subset A_{n_0}. \tag{6.22}$$

Let $x \in X$ be such that $\|x\|_X \le \varepsilon$. Then for all $T \in \mathcal{F}$,

$$\|Tx\|_Y = \|T(x_0 - x) - Tx_0\|_Y \le \|T(x_0 - x)\|_Y + \|Tx_0\|_Y \le 2n_0. \tag{6.23}$$

Now we use the usual scaling trick: let $x \in X$ satisfy $\|x\|_X = 1$. Then

$$\|Tx\|_Y = \varepsilon^{-1}\|T(\varepsilon x)\|_Y \le 2\varepsilon^{-1} n_0. \tag{6.24}$$

This implies

$$\sup_{T\in\mathcal{F}} \|T\|_{\mathrm{op}} = \sup_{T\in\mathcal{F}} \sup_{\|x\|_X=1} \|Tx\|_Y \le 2\varepsilon^{-1} n_0 < \infty. \tag{6.25}$$

□

EXAMPLE 6.28. If X is not complete, then the conclusion of the theorem may fail. For instance, let X be the space of all sequences $(x_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ such that at most finitely many of the x_n are non-zero. Equip X with the norm $\|x\|_\infty = \sup_{n\in\mathbb{N}} |x_n|$. Define $\ell_n : X \to \mathbb{R}$ by $\ell_n(x) = nx_n$. ℓ_n is a bounded linear map because

$$|\ell_n(x)| = |nx_n| \le n\|x\|_\infty. \tag{6.26}$$

For every $x \in X$ there exists $N_x \in \mathbb{N}$ such that $x_n = 0$ for all $n > N_x$.This implies that

$$\sup_{n\in\mathbb{N}} |\ell_n(x)| = \max\{|\ell_n(x)| \,:\, n = 1, \dots, N_x\} < \infty. \tag{6.27}$$

But $\|\ell_n\|_{\text{op}} \geq n$ because $|\ell_n(e_n)| = n$ (where e_n denotes the sequence such that $e_n(m) = 0$ for every $m \neq n$ and $e_n(n) = 1$). Thus,

$$\sup_{n\in\mathbb{N}} \|\ell_n\|_{\text{op}} = \infty. \tag{6.28}$$

Remark. In the proof we only needed that X is not meager. This is true if X is complete, but it may also be true for an incomplete space.

As a first application of the uniform boundedness principle we prove that the pointwise limit of a sequence of bounded linear operators on a Banach space must be a bounded linear operator.

COROLLARY 6.29. *Let X be a Banach space and Y a normed vector space. Suppose $(T_n)_{n\in\mathbb{N}} \subset L(X,Y)$ is such that $(T_n x)_{n\in\mathbb{N}}$ converges to some Tx for every $x \in X$. Then $T \in L(X,Y)$.*

PROOF. Linearity of T follows from linearity of limits. It remains to show that T is bounded. Let $x \in X$. Since $(T_n x)_{n\in\mathbb{N}}$ converges, we have $\sup_n \|T_n x\|_Y < \infty$ (convergent sequences are bounded). By the Banach-Steinhaus theorem, there exists $C \in (0,\infty)$ such that $\|T_n\|_{\text{op}} \leq C$ for every n. Let $x \in X$. Then

$$\|Tx\|_Y = \lim_{n\to\infty} \|T_n x\|_Y \leq C\|x\|_X. \tag{6.29}$$

□

Remark. Note that in the context of Corollary 6.29 it does not follow that $T_n \to T$ in $L(X,Y)$. For instance, let $T_n : \ell^1 \to \ell^1$ and $T_n(x) = x_n e_n$. Then $T_n(x) \to 0$ as $n \to \infty$ for every $x \in \ell^1$, but $\|T_n\|_{\text{op}} = 1$ for every $n \in \mathbb{N}$, so T_n does not converge to 0 in $L(X,Y)$.

4.1. An application to Fourier series. Recall that for a 1-periodic continuous function $f : \mathbb{R} \to \mathbb{C}$ we defined the partial sums of its Fourier series by

$$S_N f(x) = \sum_{n=-N}^{N} c_n e^{2\pi i n x} = f * D_N(x), \tag{6.30}$$

where $c_n = \int_0^1 f(t)e^{-2\pi i t n}dt$ and $D_N(x) = \sum_{n=-N}^{N} e^{2\pi i x n} = \frac{\sin(2\pi(N+\frac{1}{2})x)}{\sin(\pi x)}$ is the Dirichlet kernel (see Section 4).

The uniform boudnedness principle directly implies the following:

COROLLARY 6.30. *Let $x_0 \in \mathbb{R}$. There exists a 1-periodic continuous function f such that the sequence $(S_N f(x_0))_N \subset \mathbb{C}$ does not converge. That is, the Fourier series of f does not converge at x_0.*

In particular, this means that the Dirichlet kernels do not form an approximation of unity. To see why this is a consequence of the uniform boudnedness principle, we first need to take another close look at the partial sums.

LEMMA 6.31. *There exists a constant $c \in (0,\infty)$ such that for every $N \in \mathbb{N}$,*

$$\int_0^1 |D_N(x)|dx \geq c\log(N). \tag{6.31}$$

PROOF. Since $|\sin(x)| \leq |x|$,

$$\int_0^1 |D_N(x)|dx = \int_0^1 \frac{|\sin(2\pi(N+\frac{1}{2})x)|}{|\sin(\pi x)|}dx \geq \pi^{-1}\int_0^1 \frac{|\sin(2\pi(N+\frac{1}{2})x)|}{x}dx. \tag{6.32}$$

Changing variables $2\pi(N+\frac{1}{2})x \mapsto x$ we see that the right hand side of this display equals

$$\pi^{-1}\int_0^{\pi(2N+1)} \frac{|\sin(x)|}{x}dx = \pi^{-1}\sum_{k=0}^{2N}\int_{\pi k}^{\pi(k+1)} \frac{|\sin(x)|}{x}dx. \tag{6.33}$$

We have that

$$\sum_{k=0}^{2N}\int_{\pi k}^{\pi(k+1)} \frac{|\sin(x)|}{x}dx \geq \sum_{k=0}^{2N}\int_{\pi k+\frac{\pi}{2}-\frac{\pi}{100}}^{\pi k+\frac{\pi}{2}+\frac{\pi}{100}} \frac{|\sin(x)|}{x}dx \geq c\sum_{k=0}^{2N}\int_{\pi k+\frac{\pi}{2}-\frac{\pi}{100}}^{\pi k+\frac{\pi}{2}+\frac{\pi}{100}} \frac{dx}{x}. \tag{6.34}$$

Here we have used that $|\sin(x)| \geq c$ for some positive number c whenever $|x|$ is at most $\frac{\pi}{100}$ away from $\pi k + \frac{\pi}{2}$ for some integer $k \in \mathbb{Z}$ (indeed, $|\sin(x)| \geq \sin(\pi/2 - \pi/100) > 0$ for such x). Since $x \mapsto 1/x$ is a decreasing function,

$$\int_{\pi k+\frac{\pi}{2}-\frac{\pi}{100}}^{\pi k+\frac{\pi}{2}+\frac{\pi}{100}} \frac{dx}{x} \geq \tfrac{\pi}{50}\cdot\frac{1}{\pi k+\frac{\pi}{2}+\frac{\pi}{100}} \geq \tfrac{1}{50}\cdot\frac{1}{k+1}. \tag{6.35}$$

Thus,

$$\sum_{k=0}^{2N}\int_{\pi k+\frac{\pi}{2}-\frac{\pi}{100}}^{\pi k+\frac{\pi}{2}+\frac{\pi}{100}} \frac{dx}{x} \geq \tfrac{1}{50}\sum_{k=0}^{2N}\frac{1}{k+1} \geq \tfrac{1}{50}\sum_{k=0}^{2N}\int_{k+1}^{k+2}\frac{dx}{x} = \tfrac{1}{50}\int_1^{2N+2}\frac{dx}{x} = \tfrac{1}{50}\log(2N+2), \tag{6.36}$$

which implies the claim. □

Let us denote the space of 1-periodic continuous functions $f : \mathbb{R} \to \mathbb{C}$ by $C(\mathbb{T})$ (here $\mathbb{T} = \mathbb{R}/\mathbb{Z} = S^1$ is the unit circle, which is a compact metric space[1]). Then $C(\mathbb{T})$ is a Banach space. Fix $x_0 \in \mathbb{R}$. We can define a linear map $T_N : C(\mathbb{T}) \to \mathbb{C}$ by

$$T_N f = S_N f(x_0). \tag{6.37}$$

LEMMA 6.32. *For every $N \in \mathbb{N}$, $T_N : C(\mathbb{T}) \to \mathbb{C}$ is a bounded linear map and*

$$\|T_N\|_{\mathrm{op}} = \|D_N\|_1. \tag{6.38}$$

(Here $\|D_N\|_1 = \int_0^1 |D_N(x)|dx$.)

PROOF. For every $f \in C(\mathbb{T})$ we have

$$|T_N f| = |f * D_N(x_0)| \leq \int_0^1 |f(x_0-t)D_N(t)|dt \leq \|f\|_\infty \int_0^1 |D_N(t)|dt = \|f\|_\infty \|D_N\|_1. \tag{6.39}$$

Therefore, T_N is bounded and $\|T_N\|_{\mathrm{op}} \leq \|D_N\|_1$. To prove the lower bound we let

$$f(x) = \mathrm{sgn}(D_N(x_0 - x)). \tag{6.40}$$

While f is not a continuous function, it can be approximated by continuous functions as the following exercise shows.

EXERCISE 6.33. Show that for every $\varepsilon > 0$ there exists $g \in C(\mathbb{T})$ such that $|g(t)| \leq 1$ for all $t \in \mathbb{R}$ and

$$\int_0^1 |f(t) - g(t)|dt \leq \frac{\varepsilon}{2N+1} \tag{6.41}$$

[1]The metric being the quotient metric inherited from $\mathbb{R}$ or the subspace metric induced by the inclusion $S^1 \subset \mathbb{R}^2$. These metrics are equivalent.

Hint: Modify the function f in a small enough neighborhood of each discontinuity; g can be chosen to be a piecewise linear function.

So let $\varepsilon > 0$ and choose $g \in C(\mathbb{T})$ as in the exercise. We have

$$|T_N f| = |f * D_N(x_0)| = \Big| \int_0^1 \operatorname{sgn}(D_N(t)) D_N(t) dt \Big| = \int_0^1 |D_N(t)| dt = \|D_N\|_1. \tag{6.42}$$

Moreover,

$$|T_N g| \geq |T_N f| - |T_N(f-g)|, \tag{6.43}$$

The error term $|T_N(f-g)|$ can be estimated as follows:

(6.44)

$$|T_N(f-g)| \leq \int_0^1 |D_N(x_0-t)||f(t)-g(t)|dt \leq \|D_N\|_\infty \int_0^1 |f(t)-g(t)|dt \leq (2N+1)\frac{\varepsilon}{2N+1} = \varepsilon.$$

so

$$\|T_N\|_{\text{op}} \geq |T_N g| \geq \|D_N\|_1 - \varepsilon. \tag{6.45}$$

Since $\varepsilon > 0$ was arbitrary, this implies $\|T_N\|_{\text{op}} \geq \|D_N\|_1$. □

Armed with this knowledge, we can now reveal Corollary 6.30 as a direct consequence of Theorem 6.27. Indeed, we have that

$$\|T_N\|_{\text{op}} = \|D_N\|_1 \geq c \log(N) \tag{6.46}$$

and therefore

$$\sup_{N \in \mathbb{N}} \|T_N\|_{\text{op}} = \infty. \tag{6.47}$$

So by Theorem 6.27 there must exist an $f \in C(\mathbb{T})$ such that

$$\sup_{N \in \mathbb{N}} |T_N f| = \infty. \tag{6.48}$$

In otherwords, $(S_N f(x_0))_N$ does not converge.

Remark. Continuous functions with divergent Fourier series can also be constructed explicitly. The conclusion of Corollary 6.30 can be strengthened significantly: for every Lebesgue null set $A \subset \mathbb{T}$ [2] there exists a continuous function whose Fourier series diverges on A (see *J.-P. Kahane, Y. Katznelson: Sur les ensembles de divergence des séries trigonométriques, Studia Math. 26 (1966), 305–306.*).
On the other hand, L. Carleson proved in 1966 that the Fourier series of a continuous function must always converge *almost everywhere* (that is, everywhere except possibly on a Lebesgue null set). This is a very deep result in Fourier analysis which is difficult to prove (see *M. Lacey, C. Thiele: A proof of boundedness of the Carleson operator, Math. Res. Lett. 7 (2000), no. 4, 361—370* for a very elegant proof).

[2] See Exercise 6.10 for a definition on $\mathbb{R}$; Lebesgue null sets of $\mathbb{T}$ are precisely the images of Lebesgue null sets on $\mathbb{R}$ under the canonical quotient map $\mathbb{R} \to \mathbb{R}/\mathbb{Z} = \mathbb{T}$.

5. Further exercises

EXERCISE 6.34. We define the subset $A \subset \mathbb{R}$ as follows: $x \in A$ if and only if there exists $c > 0$ such that

$$|x - j2^{-k}| \geq c2^{-k} \tag{6.49}$$

holds for all $j \in \mathbb{Z}$ and integers $k \geq 0$. Show that A is meager and dense.

EXERCISE 6.35. Let (X, d) be a complete metric space without isolated points. Prove that X cannot be countable.

EXERCISE 6.36. (i) Show that if X is a normed vector space and $U \subset X$ a proper subspace, then U has empty interior.
(ii) Let

$$X = \{P : \mathbb{R} \to \mathbb{R} \,|\, P \text{ is a polynomial}\}. \tag{6.50}$$

Use the Baire category theorem to prove that there exists no norm $\|\cdot\|$ on X such that $(X, \|\cdot\|)$ is a Banach space.
(iii) Let X be an infinite dimensional Banach space. Prove that X cannot have a countable (linear-algebraic) basis.

EXERCISE 6.37. Consider $X = C([-1, 1])$ with the usual norm $\|f\|_\infty = \sup_{t\in[-1,1]} |f(t)|$. Let

$$A_+ = \{f \in X \,:\, f(t) = f(-t) \quad \forall t \in [-1, 1]\}, \tag{6.51}$$

$$A_- = \{f \in X \,:\, f(t) = -f(-t) \quad \forall t \in [-1, 1]\}. \tag{6.52}$$

(i) Show that A_+ and A_- are meager.
(ii) Is $A_+ + A_- = \{f + g \,:\, f \in A_+, g \in A_-\}$ meager?

EXERCISE 6.38. Construct a function $f : \mathbb{R} \to \mathbb{R}$ such that f is continuous at every $x \in \mathbb{Z}$ and discontinuous at every $x \notin \mathbb{Z}$.

EXERCISE 6.39. For every interval (open, half-open or closed) $I \subset \mathbb{R}$ give an example of a function $f : \mathbb{R} \to \mathbb{R}$ such that f is continuous on I and discontinuous on $\mathbb{R} \setminus I$.

EXERCISE 6.40*. Let $f : \mathbb{R} \to \mathbb{R}$ be a smooth function so that for every $x \in \mathbb{R}$ there exists $n \geq 0$ with $f^{(n)}(x) = 0$. Prove that f is a polynomial.

APPENDIX A

Review

1. Series

Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of complex numbers. Recall that we say that the *series* $\sum_{n=1}^{\infty} a_n$ *converges* if the sequence of partial sums $(\sum_{n=1}^{N} a_n)_{N\in\mathbb{N}}$ converges. In that case, the symbol $\sum_{n=1}^{\infty} a_n$ represents the limit of this sequence. If the summands are non-negative (that is, $a_n \geq 0$ for all $n \in \mathbb{N}$), then we also write

$$\sum_{n=1}^{\infty} a_n < \infty \tag{A.1}$$

to denote that the series $\sum_{n=1}^{\infty} a_n$ converges. The series $\sum_{n=1}^{\infty} a_n$ is said to *converge absolutely* if the series $\sum_{n=1}^{\infty} |a_n|$ converges.

Similarly, given a sequence of functions $(f_n)_{n\in\mathbb{N}}$ on a metric space X we say that $\sum_{n=1}^{\infty} f_n$ *converges uniformly*, if the sequence of partial sums $(\sum_{n=1}^{N} f_n)_{N\in\mathbb{N}}$ converges uniformly.

We will also sometimes consider *doubly infinite series* of the form $\sum_{n=-\infty}^{\infty} a_n$ for a sequence of complex numbers $(a_n)_{n\in\mathbb{Z}}$. Such a series is considered convergent if each of the series $\sum_{n=0}^{\infty} a_{-n}$ and $\sum_{n=1}^{\infty} a_n$ converges (and its value is in this case the sum of the values of these two series).

LEMMA A.1 (Weierstrass M-test). *Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of functions on a metric space X such that there exists a sequence of non-negative real numbers $(M_n)_{n\in\mathbb{N}}$ with*

$$|f_n(x)| \leq M_n \tag{A.2}$$

for all $n = 1, 2, \ldots$ and all $x \in X$. Assume that $\sum_{n=1}^{\infty} M_n$ converges. Then the series $\sum_{n=1}^{\infty} f_n$ converges uniformly.

PROOF. Let $s_m(x) = \sum_{k=1}^{m} f_k(x)$. For $\ell < m$ we observe the estimate,

$$|s_m(x) - s_\ell(x)| = \Big| \sum_{k=\ell+1}^{m} f_k(x) \Big| \leq \sum_{k=\ell+1}^{m} |f_k(x)| \leq \sum_{k=\ell+1}^{m} M_k. \tag{A.3}$$

Since $\sum_k M_k$ converges there is, given $\varepsilon > 0$ an N_ε such that $\sum_{k=N_1}^{N_2} M_k < \varepsilon$ provided that N_1, N_2 are greater than N_ε (why?). Use this fact and the displayed estimate to conclude the proof. □

LEMMA A.2. *Suppose $(f_n)_{n\in\mathbb{N}}$ is a sequence of Riemann integrable functions on the interval $[a, b]$ which uniformly converges to some limit f on $[a, b]$. Then f is Riemann integrable and*

$$\lim_{n\to\infty} \int_a^b f_n = \int_a^b f. \tag{A.4}$$

PROOF. Let $a < b$ and recall that for two Riemann integrable functions h_1, h_2 on the interval $[a, b]$ which satisfy $h_1(x) \leq h_2(x)$ for all $x \in [a, b]$ we also have $\int_a^b h_1 \leq \int_a^b h_2$ (one proves this by considering first the corresponding inequalities for Riemann upper and lower sums). Apply this fact together with the linearity of the integral to get

$$\Big| \int_a^b f_n - \int_a^b f \Big| = \Big| \int_a^b f_n - f \Big| \leq \int_a^b |f_n - f| \leq (b-a) \sup_{[a,b]} |f_n - f|. \tag{A.5}$$

If f_n converges to f uniformly then the right hand side converges to 0. □

EXERCISE A.3. Prove some of these facts yourself.

2. Power series

A *power series* is a function of the form

$$f(x) = \sum_{n=0}^{\infty} c_n x^n \tag{A.6}$$

where $c_n \in \mathbb{C}$ are some complex coefficients.

To a power series we can associate a number $R \in [0, \infty]$ called its *radius of convergence* such that

- $\sum_{n=0}^{\infty} c_n x^n$ converges for every $|x| < R$,
- $\sum_{n=0}^{\infty} c_n x^n$ diverges for every $|x| > R$.

On the convergence boundary $|x| = R$, the series may converge or diverge. The number R can be computed by the *Cauchy-Hadamard formula* :

$$R = \Big(\limsup_{n \to \infty} |c_n|^{1/n} \Big)^{-1} \tag{A.7}$$

(with the convention that if $\limsup_{n\to\infty} |c_n|^{1/n} = 0$, then $R = \infty$.)

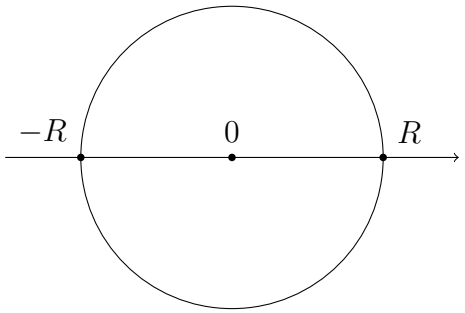

FIGURE 1. Radius of convergence

LEMMA A.4. *A power series with radius of convergence R converges uniformly on $[-R + \varepsilon, R - \varepsilon]$ for every $0 < \varepsilon < R$. Consequently, power series are continuous on $(-R, R)$.*

EXERCISE A.5. Prove this. Uniform convergence does not necessarily hold on $(-R, R)$; give an example.

LEMMA A.6. *If $f(x) = \sum_{n=0}^{\infty} c_n x^n$ has radius of convergence R, then f is differentiable on $(-R, R)$ and*

$$f'(x) = \sum_{n=1}^{\infty} n c_n x^{n-1} \tag{A.8}$$

for $|x| < R$.

EXAMPLE A.7. The *exponential function* is a power series defined by

$$\exp(x) = \sum_{n=0}^{\infty} \frac{1}{n!} x^n. \tag{A.9}$$

The radius of convergence is $R = \infty$.

LEMMA A.8. *The exponential function is differentiable and* $\exp'(x) = \exp(x)$ *for all* $x \in \mathbb{R}$.

LEMMA A.9. *For all $x, y \in \mathbb{R}$ we have the functional equation*

$$\exp(x + y) = \exp(x) \exp(y). \tag{A.10}$$

It also makes sense to speak of $\exp(z)$ for $z \in \mathbb{C}$ since the series converges absolutely. We also write e^x instead of $\exp(x)$.

EXAMPLE A.10. The trigonometric functions can also be defined by power series:

$$\cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} \tag{A.11}$$

$$\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} \tag{A.12}$$

LEMMA A.11. *The functions* sin *and* cos *are differentiable and*

$$\sin'(x) = \cos(x), \quad \cos'(x) = -\sin(x) \tag{A.13}$$

The trigonometric functions are related to the exponential function via complex numbers.

LEMMA A.12 (Euler's identity). *For all $x \in \mathbb{R}$,*

$$e^{ix} = \cos(x) + i \sin(x), \tag{A.14}$$

$$\cos(x) = \frac{e^{ix} + e^{-ix}}{2}, \tag{A.15}$$

$$\sin(x) = \frac{e^{ix} - e^{-ix}}{2i}. \tag{A.16}$$

LEMMA A.13 (Pythagorean theorem). *For all $x \in \mathbb{R}$,*

$$\cos(x)^2 + \sin(x)^2 = 1. \tag{A.17}$$

Let us also recall basic properties of complex numbers at this point: For every complex number $z \in \mathbb{C}$ there exist $a, b \in \mathbb{R}$, $r \geq 0$ and $\phi \in [0, 2\pi)$ such that

$$z = a + ib = re^{i\phi}. \tag{A.18}$$

The *complex conjugate* of z is defined by

$$\overline{z} = a - ib = re^{-i\phi} \tag{A.19}$$

The *absolute value* of z is defined by

$$|z| = \sqrt{a^2 + b^2} = r \tag{A.20}$$

We have

$$|z|^2 = z\overline{z}. \tag{A.21}$$

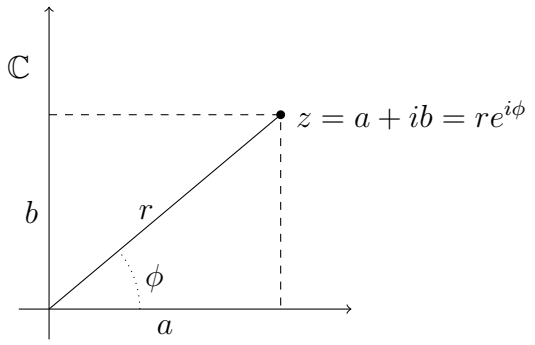

FIGURE 2. Polar and cartesian coordinates in the complex plane

We finish the review section with a simple, but powerful theorem on the continuity of power series on the convergence boundary.

THEOREM A.14 (Abel). *Let $f(x) = \sum_{n=0}^{\infty} c_n x^n$ be a power series with radius of convergence $R = 1$. Assume that $\sum_{n=0}^{\infty} c_n$ converges. Then*

$$\lim_{x \to 1-} f(x) = \sum_{n=0}^{\infty} c_n. \tag{A.22}$$

(In particular, the limit exists.)

The key idea for the proof is *Abel summation* , also referred to as *summation by parts* . The precise formula can be derived simply by reordering terms (we say that $a_{-1} = 0$):

$$\sum_{n=0}^{N} (a_n - a_{n-1}) b_n = a_0 b_0 + a_1 b_1 - a_0 b_1 + a_2 b_2 - a_1 b_2 + \cdots + a_N b_N - a_{N-1} b_N \tag{A.23}$$

(A.24)

$$= a_0(b_0 - b_1) + a_1(b_1 - b_2) + \cdots + a_{N-1}(b_{N-1} - b_N) + a_N b_N = a_N b_N + \sum_{n=0}^{N-1} a_n (b_n - b_{n+1})$$

PROOF. To apply summation by parts we set $s_n = \sum_{k=0}^{n} c_k$, $s_{-1} = 0$. Then

$$\sum_{n=0}^{N} c_n x^n = \sum_{n=0}^{N} (s_n - s_{n-1}) x^n = s_N x^N + (1 - x) \sum_{n=0}^{N-1} s_n x^n. \tag{A.25}$$

Let $0 < x < 1$. Then

$$f(x) = (1-x)\sum_{n=0}^{\infty} s_n x^n \tag{A.26}$$

Let $s = \sum_{n=0}^{\infty} c_n$. By assumption, $s_n \to s$. Let $\varepsilon > 0$ and choose $N \in \mathbb{N}$ such that

$$|s_n - s| < \varepsilon \tag{A.27}$$

for all $n > N$. Then,

$$|f(x) - s| = \Big|(1-x)\sum_{n=0}^{\infty}(s_n - s)x^n\Big|, \tag{A.28}$$

because $(1-x)\sum_{n=0}^{\infty} x^n = 1$. Now we use the triangle inequality and split the sum at $n = N$:

$$\leq (1-x)\sum_{n=0}^{N}|s_n - s|x^n + (1-x)\sum_{n=N+1}^{\infty}\overbrace{|s_n - s|}^{\leq\varepsilon}x^n \tag{A.29}$$

$$\leq (1-x)\sum_{n=0}^{N}|s_n - s|x^n + \varepsilon. \tag{A.30}$$

By making x sufficiently close to 1 we can achieve that

$$(1-x)\sum_{n=0}^{N}|s_n - s|x^n \leq \varepsilon. \tag{A.31}$$

This concludes the proof. □

Abel's theorem provides a tool to evaluate convergent series.

Example A.15. Consider the power series

$$f(x) = \sum_{n=0}^{\infty}\frac{(-1)^n}{2n+1}x^{2n+1}. \tag{A.32}$$

The radius of convergence is $R = 1$. This is the Taylor series at $x = 0$ of the function arctan.

Exercise A.16. (a) Prove that $f(x)$ really is the Taylor series at $x = 0$ of arctan. (b) Prove that $\arctan(x)$ is represented by its Taylor series at $x = 0$ for every $|x| < 1$, i.e. that $f(x) = \arctan(x)$ for $|x| < 1$.

It follows from the alternating series test that $\sum_{n=0}^{\infty}\frac{(-1)^n}{2n+1}$ converges. Thus, Abel's theorem implies that

$$\sum_{n=0}^{\infty}\frac{(-1)^n}{2n+1} = \lim_{x\to 1-}\arctan(x) = \arctan(1) = \frac{\pi}{4}. \tag{A.33}$$

This is also known as *Leibniz' formula* .

3. Taylor's theorem

THEOREM A.17. *Let I be an interval and let $f \in C^{n+1}(I)$, i.e all derivatives of f up to order $n+1$ are continuous in I. Fix $a \in I$. Then for all $x \in I$.*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k + R_n(x,a) \tag{A.34}$$

where

$$\begin{aligned} R_n(x,a) &= \int_a^x \frac{(x-t)^n}{n!} f^{(n+1)}(t)dt \\ &= \frac{(x-a)^{n+1}}{n!} \int_0^1 (1-s)^n f^{(n+1)}(a+s(x-a))ds \end{aligned} \tag{A.35}$$

PROOF. We first observe that the second version and the first version of the remainder term are equivalent by changing variables (via the substitution $t = a + s(x-a)$, $dt = (x-a)ds$; note that t ranges from a to x as s ranges from 0 to 1).

For $n = 0$ the formula reads

$$f(x) = f(a) + \int_a^x f'(t)dt \tag{A.36}$$

which just follows from the fundamental theorem of calculus.

We also find that by integration by parts for $f \in C^{(n+2)}(I)$

$$\begin{aligned} \int_a^x (x-t)^n f^{(n+1)}(t)\,dt &= \Big[\frac{-(x-t)^{n+1}}{n+1} f^{(n+1)}(t)\Big]_a^x - \int_a^x \frac{-(x-t)^{n+1}}{n+1} f^{(n+2)}(t)dt \\ &= \frac{(x-a)^{n+1}}{n+1} f^{(n+1)}(a) + \int_a^x \frac{(x-t)^{n+1}}{n+1} f^{(n+2)}(t)dt. \end{aligned}$$

which shows

$$R_n(x,a) = \frac{(x-a)^{n+1}}{n+1} f^{(n+1)}(a) + R_{n+1}(x,a) \tag{A.37}$$

and establishes the induction step of the proof of the formula. To be precise if $(*_N)$ denotes the statement that

$$f(x) = \sum_{k=0}^{N} \frac{f^{(k)}(a)}{k!}(x-a)^k + R_N(x,a) \tag{A.38}$$

holds for all $f \in C^{(N+1)}(I)$ then $(*)_N$ implies $(*_{N+1})$ for all $N = 0, 1, 2, \ldots$. □

THEOREM A.18. *Let f be as in Theorem A.17 and let R_n as in* (A.35). *Let*

$$M_{n+1} = \max\{|f^{(n+1)}(a+s(x-a))| : 0 \le s \le 1\}. \tag{A.39}$$

Then

$$|R_n(x,a)| \le \frac{M_{n+1}}{(n+1)!}|x-a|^{n+1}. \tag{A.40}$$

PROOF. We have

$$\begin{aligned} |R_n(x,a)| &\le \frac{|x-a|^{n+1}}{n!} \int_0^1 (1-s)^n |f^{(n+1)}(a+s(x-a))|ds \\ &\le \frac{|x-a|^{n+1}}{n!} M_{n+1} \int_0^1 (1-s)^n ds = \frac{|x-a|^{n+1}}{(n+1)!} M_{n+1} \end{aligned}$$

□

THEOREM A.19. *Let f be as in Theorem A.17 and let R_n as in* (A.35). *There is ξ between a and x such that*

$$R_n(x,a) = \frac{|x-a|^{n+1}}{(n+1)!} f^{(n+1)}(\xi). \tag{A.41}$$

PROOF. Let

$$m = \min\{f^{(n+1)}(a+s(x-a)) : 0 \le s \le 1\},$$
$$M = \max\{f^{(n+1)}(a+s(x-a)) : 0 \le s \le 1\}.$$

We estimate

$$\int_0^1 (1-s)^n m\, ds \le \int_0^1 (1-s)^n f^{(n+1)}(a+s(x-a))ds \le \int_0^1 (1-s)^n M\, ds \tag{A.42}$$

and hence

$$m \le (n+1)\int_0^1 (1-s)^n f^{(n+1)}(a+s(x-a))ds \le M. \tag{A.43}$$

By the intermediate value theorem for continuous functions there is $\sigma \in [0,1]$ such that

$$f^{(n+1)}(a+\sigma(x-a)) = (n+1)\int_0^1 (1-s)^n f^{(n+1)}(a+s(x-a))ds. \tag{A.44}$$

If we set $\xi = a+\sigma(x-a)$ so that ξ is on the line segment connecting a to x we get the claimed statement from (A.35) and (A.44). □

4. The Riemann integral

We recall some definitions. In what follows $a, b \in \mathbb{R}$, with $a < b$ are given. In this section we recall basic definitions which lead to the definition of Riemann integrable functions on $[a,b]$, and the Riemann integral of such functions.

DEFINITION A.20. (i) A partition $P = \{x_0, \dots, x_n\}$ of $[a,b]$ is a finite subset of $[a,b]$ which includes the points a and b and is ordered in the following way:

$$a = x_0 < ... < x_i < x_{i+1} < ... < x_n = b. \tag{A.45}$$

(ii) If P, P' are partitions of $[a,b]$ with $P \subset P'$ then P' is called a refinement of P.

DEFINITION A.21. Given a partition $P = \{a = x_0 < \cdots < x_n = b\}$ of $[a,b]$ and a *bounded* function $f : [a,b] \to \mathbb{R}$ define

$$m_i(f) = \inf_{t\in[x_{i-1},x_i]} f(t),$$
$$M_i(f) = \sup_{t\in[x_{i-1},x_i]} f(t).$$

(i) The expression

$$L(f,P) = \sum_{i=1}^n m_i(f)(x_i - x_{i-1}) \tag{A.46}$$

is called the lower sum of f with respect to the partition P.

(ii) The expression

$$U(f,P) = \sum_{i=1}^n M_i(f)(x_i - x_{i-1}) \tag{A.47}$$

is called the upper sum of f with respect to the partition P.

LEMMA A.22. *Let P, P' be partitions of $[a,b]$, let $f:[a,b]\to\mathbb{R}$ be bounded, and let P' be a refinement of P. Then*

$$(b-a)\inf_{[a,b]} f \le L(f,P)\le L(f,P')\le U(f,P')\le U(f,P)\le (b-a)\sup_{[a,b]} f. \tag{A.48}$$

COROLLARY A.23. *Let P_1, P_2 be partitions of $[a,b]$. Then $L(f,P_1)\le U(f,P_2)$.*

DEFINITION A.24. Let $f:[a,b]\to\mathbb{R}$ be a bounded function. The numbers

$$\underline{\mathcal{I}}_a^b(f):=\sup_P L(f,P),\qquad \overline{\mathcal{I}}_a^b(f):=\inf_P U(f,P) \tag{A.49}$$

are called the *lower and upper Riemann-Darboux integrals of f* on the interval $[a,b]$, respectively. Here the sup and inf are taken over all partitions of $[a,b]$.

LEMMA A.25. *Let $f:[a,b]\to\mathbb{R}$ be bounded. Then*

$$(b-a)\inf_{[a,b]} f \le \underline{\mathcal{I}}_a^b(f)\le \overline{\mathcal{I}}_a^b(f)\le (b-a)\sup_{[a,b]} f. \tag{A.50}$$

We are now ready to define the concept of Riemann integrable functions and the Riemann integral of such functons.

DEFINITION A.26. (i) Let $f:[a,b]\to\mathbb{R}$ be bounded. f is called Riemann integrable if $\underline{\mathcal{I}}_a^b(f)=\overline{\mathcal{I}}_a^b(f)$.

(ii) If f is Riemann integrable the number $\underline{\mathcal{I}}_a^b(f)=\overline{\mathcal{I}}_a^b(f)$ is called the Riemann integral of f, denoted by $\int_{[a,b]} f$ or by $\int_a^b f$ (or even by $\int_a^b f(t)dt$ ...)

LEMMA A.27. *Let $f:[a,b]\to\mathbb{R}$ be a bounded function. Then f is Riemann integrable if and only if for every $\varepsilon>0$ there is a partition P of $[a,b]$ such that*

$$U(f,P)-L(f,P)<\varepsilon. \tag{A.51}$$

PROOF. Suppose f is Riemann integrable. Then there are partitions P_1, P_2 of $[a,b]$ such that $L(f,P_1)\ge\int_a^b f-\varepsilon/2$, $U(f,P_2)\le\int_a^b f+\varepsilon/2$ and thus $U(f,P_2)-L(f,P_1)<\varepsilon$. Let P be the refinement $P_1\cup P_2$. Then $U(f,P_2)\ge U(f,P)\ge L(f,P)\ge L(f,P_1)$ and hence $U(f,P)-L(f,P)<\varepsilon$.

Vice versa assume that for every ϵ there is a partition P_ε of $[a,b]$ such that

$$U(f,P_\varepsilon)-L(f,P_\varepsilon)<\varepsilon. \tag{A.52}$$

Then $\overline{\mathcal{I}}_a^b(f)-\underline{\mathcal{I}}_a^b(f)\le U(f,P_\varepsilon)-L(f,P_\varepsilon)<\varepsilon$, and since ε was arbitrary we conclude $\overline{\mathcal{I}}_a^b(f)=\underline{\mathcal{I}}_a^b(f)$. Hence f is Riemann integrable. □

THEOREM A.28. *If $f:[a,b]\to\mathbb{R}$ is continuous in $[a,b]$ then f is Riemann integrable.*

PROOF. Recall that a continuous function on a compact set is uniformly continuous. Hence given $\varepsilon>0$ there exists $\delta_\varepsilon>0$ such that $|f(x)-f(\tilde{x})|<\varepsilon/(b-a)$ provided that $|x-\tilde{x}|<\delta$. Let N be such that $(b-a)/N<\delta$ and choose the partition $P=\{x_j:=a+j\frac{b-a}{N}, j=0,\dots N\}$. Let $I_j=[x_{j-1},x_j]$, $j=1,\dots,N$. Then

$$(M_i(f)-m_i(f)=(\sup_{I_j} f-\inf_{I_j} f)<\varepsilon/(b-a) \tag{A.53}$$

for $i = 1, \dots, N$ so that

$$U(f,P) - L(f,P) = \sum_{i=1}^{N} M_i f(x_j - x_{j-1}) - \sum_{i=1}^{N} m_i(f)(x_j - x_{j-1})$$
$$= \sum_{i=1}^{N} (M_j(f) - m_i(f))(x_j - x_{j-1}) \le \sum_{i=1}^{N} \frac{\varepsilon}{b-a}(x_j - x_{j-1}) = \frac{\varepsilon}{b-a}(b-a) = \varepsilon.$$

We can apply Lemma to see that f is Riemann integrable. □

THEOREM A.29. *(i) Let f and g be Riemann integrable functions on $[a,b]$ and suppose that $f(x) \le g(x)$ for all $x \in [a,b]$. Then $\int_a^b f \le \int_a^b g$.*

(ii) Let f be Riemann integrable on $[a,b]$. Then

$$\Big| \int_a^b f \Big| \le (b-a) \sup_{[a,b]} |f| \tag{A.54}$$

EXERCISE A.30. Let $f : [a,b] \to \mathbb{R}$ be a bounded function. Under each of the following hypotheses on f show that f is Riemann integrable.

(i) There is a point $c \in [a,b]$ such that f is continuous on $[a,b] \setminus \{c\}$.

(ii) f is continuous except possibly at a finite number of points in $[a,b]$.

(iii) f is continuous in $[a,b] \setminus \{c_k : k \in \mathbb{N}\}$, where $(c_k)_{k\in\mathbb{N}}$ is a *convergent* sequence of points in $[a,b]$,

EXERCISE A.31. Let $f : [0,1] \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} x^2 & \text{if } x \in \mathbb{Q} \\ x & \text{if } x \in [0,1] \setminus \mathbb{Q}. \end{cases} \tag{A.55}$$

Compute $\overline{\mathcal{I}}_0^1(f)$ and $\underline{\mathcal{I}}_0^1(f)$.

5. Further exercises

EXERCISE A.32. Prove or disprove convergence for each of the following series (a and b are real parameters and convergence may depend on their values).

$$\text{(i)} \sum_{n=2}^{\infty} \frac{1}{n^a (\log(n))^b} \qquad \text{(ii)} \sum_{n=3}^{\infty} (\log n)^{a \frac{\log n}{\log\log n}} \qquad \text{(iii)} \sum_{n=1}^{\infty} \left(e^{1/n} - \tfrac{n+1}{n}\right)$$

$$\text{(iv)} \sum_{n=1}^{\infty} \cos(\pi n)\sin(\pi n^{-1}) \qquad \text{(v)} \sum_{n=2}^{\infty} \left(\left(1+\frac{1}{n}\right)^n - e\right)^2 \qquad \text{(vi)} \sum_{n=1}^{\infty} \frac{1}{n(n^{1/n})^{100}}$$

$$\text{(vii)} \sum_{n=2}^{\infty} 2^{-(\log(n))^a} \qquad \text{(viii)} \sum_{n=1}^{\infty} \Big(\sum_{k=0}^{10n} (-1)^k \frac{n^k}{k!} \Big) \qquad \text{(ix)} \sum_{n=1}^{\infty} \frac{1}{n^2(1-\cos(n))}$$

EXERCISE A.33. Prove or disprove convergence for each of the following sequences and in case of convergence, determine the limit:

(i) $a_n = \sqrt{n^4 + \cos(n^2)} - n^2$

(ii) $a_n = n^2 + \frac{1}{2}n - \sqrt{n^4 + n^3}$

(iii) $a_n = \sum_{k=n}^{n^2} \frac{1}{k}$

(iv) $a_n = n \sum_{k=0}^{\infty} \frac{1}{n^2+k^2}$

(v) $a_0 = 1$, $a_{n+1} = \frac{a_n}{2} + \frac{1}{a_n}$

(vi) $a_n = \prod_{k=2}^{n} \frac{k^2-1}{k^2}$

EXERCISE A.34. For which $x \in \mathbb{R}$ do the following series converge? On which sets do these series converge uniformly?

$$\text{(i)} \sum_{n=1}^{\infty} n^2 x^n \qquad \text{(ii)} \sum_{n=1}^{\infty} (3^{1/n} - 1)^n x^n \qquad \text{(iii)} \sum_{n=1}^{\infty} \tan(n^{-2}) e^{nx} \tag{A.56}$$

$$\text{(iv)} \sum_{n=1}^{\infty} \frac{x^n}{n^n} \quad \text{(v)} \sum_{n=1}^{\infty} \frac{\sin(nx)}{n^2} \quad \text{(vi)} \sum_{n=1}^{\infty} 2^{-n} \tan(\lfloor x \rfloor + 1/n) \tag{A.57}$$

EXERCISE A.35. (i) Define f by setting $f(x) = x$ for $x \geq 0$ and $f(x) = 0$ for $x < 0$. Then f is not differentiable at $x = 0$. Construct an example of a sequence $(f_n)_{n\in\mathbb{N}}$ of continuously differentiable functions defined on $\mathbb{R}$, uniformly convergent on $\mathbb{R}$ to f.
(ii) Let $f_n(x) = n^{-1/2} \sin nx$. Show that f_n converges uniformly on $\mathbb{R}$, but for every $x \in \mathbb{R}$, the sequence $(f_n'(x))_{n\in\mathbb{N}}$ does not have a limit.

EXERCISE A.36. Give an example of a sequence $(f_n)_{n\in\mathbb{N}}$ of continuous bounded functions on $\mathbb{R}$ that converges pointwise to some function f such that f is unbounded and not continuous.

EXERCISE A.37. Determine the value of the series $\sum_{n=1}^{\infty} \frac{(-1)^n}{n(n+1)}$.

EXERCISE A.38. For a positive real number x define

$$f(x) = \sum_{n=0}^{\infty} \frac{1}{n(n+1) + x}. \tag{A.58}$$

(i) Show that $f : (0, \infty) \to (0, \infty)$ is a well-defined and continuous function.
(ii) Prove that there exists a unique $x_0 \in (0, \infty)$ such that $f(x_0) = 2\pi$.
(iii) Determine the value of x_0. *Hint:* Recall Leibniz' formula from Example A.15.

EXERCISE A.39. Let $f : \mathbb{R} \to \mathbb{R}$ be a smooth function (i.e. derivatives of all orders exist). Assume that there exist $A > 0, R > 0$ such that

$$|f^{(n)}(x)| \leq A^n n! \tag{A.59}$$

for $|x| < R$. Show that there exists $r > 0$ such that for every $|x| < r$ we have that

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n. \tag{A.60}$$

(That is, prove that the series on the right hand side converges and that the limit is $f(x)$.)