Analysis II
Lecture Notes

Joris Roos
UW Madison, Fall 2019

Last update: December 14, 2019

Contents

Disclaimer:

- This content is based on various sources, mainly *Principles of Mathematical Analysis* by Walter Rudin, various individual lecture notes by Andreas Seeger, and my own notes. For my own convenience, I will not reference sources individually throughout these notes.
- These notes are likely to contain typos, errors and imprecisions of all kinds. Possibly lots. Some might be deliberate, some less so. Don't 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. That usually helps. If you do notice a mistake or an inaccuracy, feel free to let me know.
- Thanks to the students of Math 522 for many useful questions and remarks that have improved these lecture notes.

Some recommended literature for further reading:

There are many books on mathematical analysis each of which likely has a large intersection with this course. Here are two very good ones:

- W. Rudin, *Principles of Mathematical Analysis*
- T. Apostol, *Mathematical analysis: A modern approach to advanced calculus*

For further reading on Fourier series and trigonometric polynomials, see:

- E. M. Stein, R. Shakarchi, *Fourier Analysis* (modern and very accessible for beginners)
- Y. Katznelson, *Introduction to Harmonic Analysis* (slightly more advanced)
- A. Zygmund, *Trigonometric Series* (a classic that continues to be relevant today)

We will sometimes dip into concepts from functional analysis. For instance, expositions of the Baire category theorem and its consequences are also contained in

- W. Rudin, *Real and Complex Analysis* (Chapter 5)
- E. M. Stein, R. Shakarchi, *Functional Analysis* (Chapter 4)

We roughly assume knowledge of the content of Rudin's book up to Chapter 7 up to (excluding) equicontinuity, but some of the material in previous chapters will also be repeated (everything related to compactness for instance).

CHAPTER 1

Review

Lecture 1 (Wednesday, Sep 4)

1. Metric spaces

DEFINITION 1.0 (Metric space). A non-empty set X equipped with a map $d : X \times X \to [0, \infty)$ is called a *metric space* if 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*.

We will use the following notations for (closed) balls in X:

$$B(x_0, r) = \{x \in X \,:\, d(x, x_0) < r\},\ \overline{B}(x_0, r) = \{x \in X \,:\, d(x, x_0) \leq r\}. \tag{1.1}$$

We write $\overline{B(x_0, r)}$ for the closure of $B(x_0, r)$. Note that $B(x_0, r) \subset \overline{B(x_0, r)} \subset \overline{B}(x_0, r)$, but each of these inclusions may be proper.

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

The most important examples of metric spaces for the purpose of this lecture are $\mathbb{R}, \mathbb{C}, \mathbb{R}^n$, subsets thereof and $C_b(X)$, the space of bounded continuous functions on a metric space X which will be introduced later.

DEFINITION 1.1 (Convergence). Let X be a metric space, $(x_n)_n \subset X$ a sequence and $x \in X$. We say that $(x_n)_n$ *converges* to x if for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n \geq N$ it holds that $d(x_n, x) < \varepsilon$.

DEFINITION 1.2 (Continuity). Let X, Y be metric spaces. A map $f : X \to Y$ is called *continuous at* $x \in X$ if for all $\varepsilon > 0$ there exists $\delta > 0$ such that if $d_X(x, y) < \delta$, then $d_Y(f(x), f(y)) < \varepsilon$. f is called *continuous* if it is continuous at every $x \in X$. We also write $f \in C(X, Y)$.

We assume familiarity with basic concepts of metric space topology except for compactness: open sets, closed sets, limit points, closure, completeness, dense sets, connected sets, etc. We will discuss compactness in metric spaces in detail in Section 2.

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 $C_b(X)$.

2. Uniform convergence

DEFINITION 1.3. A sequence $(f_n)_n$ 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 $x \in X$,

$$|f_n(x) - f(x)| < \varepsilon. \tag{1.2}$$

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:

$$\forall \varepsilon > 0 \, \exists N \in \mathbb{N} \, \forall x \in X \, \forall n \geq N \, : \, |f_n(x) - f(x)| < \varepsilon. \tag{1.3}$$

$$\forall \varepsilon > 0 \, \forall x \in X \, \exists N \in \mathbb{N} \, \forall n \geq N \, : \, |f_n(x) - f(x)| < \varepsilon. \tag{1.4}$$

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.

In the following we collect some important facts surrounding uniform convergence that will be used in this lecture. In case you are feeling a bit rusty on these concepts, all of these are good exercises to try and prove directly from first principles.

FACT 1.4. *A sequence $(f_n)_n$ 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 $x \in X$, $|f_n(x) - f_m(x)| < \varepsilon$.*

FACT 1.5. *If $(f_n)_n$ 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 $x \in X$.)

FACT 1.6. *If $(f_n)_n$ converges uniformly to f and each f_n is continuous, then f is continuous.*

Lecture 2 (Friday, Sep 6)

FACT 1.7. *Let X be a metric space. The space of bounded continuous functions $C_b(X)$ is a complete metric space with the* supremum metric

$$d_\infty(f,g) = \sup_{x\in X} |f(x) - g(x)|. \tag{1.5}$$

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

FACT 1.8. *Let $(f_n)_n \subset C_b(X)$ be a sequence. Then $(f_n)_n$ converges in $C_b(X)$ (with respect to d_∞) if and only if it converges uniformly to f for some $f \in C_b(X)$.*

FACT 1.9 (Weierstrass M-test). *Let $(f_n)_n$ be a sequence of functions on a metric space X such that there exists a sequence of non-negative real numbers $(M_n)_n$ with*

$$|f_n(x)| \leq M_n \tag{1.6}$$

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 (that is, the sequence of partial sums $(\sum_{n=1}^m f_n)_m$ converges uniformly).

FACT 1.10. *Suppose $(f_n)_n$ 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*

$$\lim_{n\to\infty} \int_a^b f_n = \int_a^b f. \tag{1.7}$$

EXERCISE 1.11. Remind yourself how to prove all these facts.

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

EXERCISE 1.12. Find an example for this. (*Hint:* Try trigonometric functions.)

3. Power series

A *power series* is a function of the form

$$f(x) = \sum_{n=0}^\infty c_n x^n \tag{1.8}$$

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*:

$$R = \Big(\limsup_{n\to\infty} |c_n|^{1/n} \Big)^{-1} \tag{1.9}$$

(with the convention that if $\limsup_{n\to\infty} |c_n|^{1/n} = 0$, then $R = \infty$.)

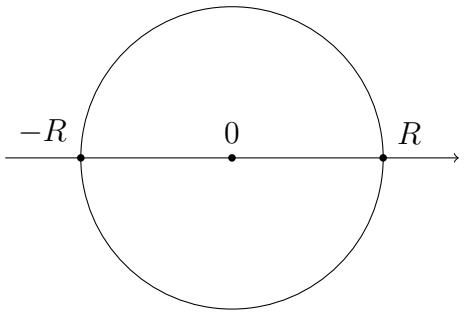

FIGURE 1. Radius of convergence

FACT 1.13. *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 1.14. Prove this. Why does uniform convergence not hold on $(-R, R)$? (Give an example.)

FACT 1.15. *If $f(x) = \sum_{n=0}^{\infty} c_n x^n$ has radius of convergence R, then f is differentiable on $(-R, R)$ and*

$$f'(x) = \sum_{n=1}^{\infty} n c_n x^{n-1} \tag{1.10}$$

for $|x| < R$.

EXAMPLE 1.16. The *exponential function* is a power series defined by

$$\exp(x) = \sum_{n=0}^{\infty} \frac{1}{n!} x^n. \tag{1.11}$$

The radius of convergence is $R = \infty$.

FACT 1.17. *The exponential function is differentiable and $\exp'(x) = \exp(x)$ for all $x \in \mathbb{R}$.*

FACT 1.18. *For all $x, y \in \mathbb{R}$ we have the functional equation*

$$\exp(x+y) = \exp(x)\exp(y). \tag{1.12}$$

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 1.19. The trigonometric functions can also be defined by power series:

$$\cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} \tag{1.13}$$

$$\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} \tag{1.14}$$

FACT 1.20. *The functions* sin *and* cos *are differentiable and*

$$\sin'(x) = \cos(x), \quad \cos'(x) = -\sin(x) \tag{1.15}$$

The trigonometric functions are related to the exponential function via complex numbers.

FACT 1.21 (Euler's identity). *For all* $x \in \mathbb{R}$,

$$e^{ix} = \cos(x) + i\sin(x), \tag{1.16}$$

$$\cos(x) = \frac{e^{ix} + e^{-ix}}{2}, \tag{1.17}$$

$$\sin(x) = \frac{e^{ix} - e^{-ix}}{2i}. \tag{1.18}$$

FACT 1.22 (Pythagorean theorem). *For all* $x \in \mathbb{R}$,

$$\cos(x)^2 + \sin(x)^2 = 1. \tag{1.19}$$

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

$$z = a + ib = re^{i\phi}. \tag{1.20}$$

The *complex conjugate* of z is defined by

$$\overline{z} = a - ib = re^{-i\phi} \tag{1.21}$$

The *absolute value* of z is defined by

$$|z| = \sqrt{a^2 + b^2} = r \tag{1.22}$$

We have

$$|z|^2 = z\overline{z}. \tag{1.23}$$

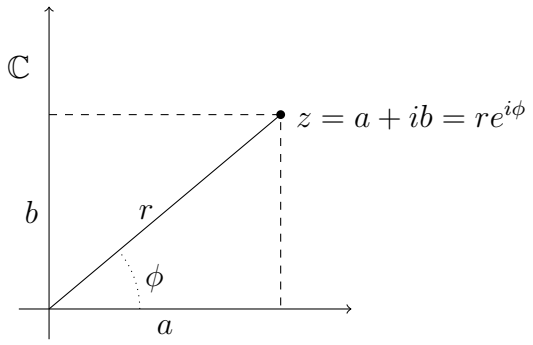

FIGURE 2. Polar and cartesian coordinates in the complex plane

Lecture 3 (Monday, Sep 9)

We finish the review section with a simple, but powerful theorem on the continuity of power series on the convergence boundary.

THEOREM 1.23 (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*

$$\lim_{x \to 1-} f(x) = \sum_{n=0}^{\infty} c_n. \tag{1.24}$$

(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 $a_{-1} = 0$):

$$\sum_{n=0}^{N} (a_n - a_{n-1}) b_n = a_0 b_0 + a_1 b_1 - a_0 b_1 + a_2 b_2 - a_1 b_2 + \cdots + a_N b_N - a_{N-1} b_N \tag{1.25}$$

(1.26)

$$= a_0(b_0 - b_1) + a_1(b_1 - b_2) + \cdots + a_{N-1}(b_{N-1} - b_N) + a_N b_N = a_N b_N + \sum_{n=0}^{N-1} a_n (b_n - b_{n+1})$$

PROOF. To apply summation by parts we set $s_n = \sum_{k=0}^{n} c_k$, $s_{-1} = 0$. Then

$$\sum_{n=0}^{N} c_n x^n = \sum_{n=0}^{N} (s_n - s_{n-1}) x^n = s_N x^N + (1 - x) \sum_{n=0}^{N-1} s_n x^n. \tag{1.27}$$

Let $0 < x < 1$. Then

$$f(x) = (1 - x) \sum_{n=0}^{\infty} s_n x^n \tag{1.28}$$

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

$$|s_n - s| < \varepsilon \tag{1.29}$$

for all $n > N$. Then,

$$|f(x) - s| = \left| (1 - x) \sum_{n=0}^{\infty} (s_n - s) x^n \right|, \tag{1.30}$$

because $(1 - x) \sum_{n=0}^{\infty} x^n = 1$. Now we use the triangle inequality and split the sum at $n = N$:

$$\leq (1 - x) \sum_{n=0}^{N} |s_n - s| x^n + (1 - x) \sum_{n=N+1}^{\infty} \overbrace{|s_n - s|}^{\leq \varepsilon} x^n \tag{1.31}$$

$$\leq (1 - x) \sum_{n=0}^{N} |s_n - s| x^n + \varepsilon. \tag{1.32}$$

By making x sufficiently close to 1 we can achieve that

$$(1-x)\sum_{n=0}^{N}|s_n-s|x^n\leq\varepsilon. \tag{1.33}$$

This concludes the proof. □

Abel's theorem provides a tool to evaluate convergent series.

EXAMPLE 1.24. Consider the power series

$$f(x)=\sum_{n=0}^{\infty}\frac{(-1)^n}{2n+1}x^{2n+1}. \tag{1.34}$$

The radius of convergence is $R=1$. This is the Taylor series at $x=0$ of the function arctan.

EXERCISE 1.25. (a) Prove that $f(x)$ really is the Taylor series at $x=0$ of arctan. (b) Prove using Taylor's theorem 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

$$\sum_{n=0}^{\infty}\frac{(-1)^n}{2n+1}=\lim_{x\to1-}\arctan(x)=\arctan(1)=\frac{\pi}{4}. \tag{1.35}$$

This is also known as *Leibniz' formula*.

4. Further exercises

EXERCISE 1.26. Prove or disprove convergence for each of the following series (a and b are real parameters and convergence may depend on their values).

$$\text{(i)}\sum_{n=2}^{\infty}\frac{1}{n^a(\log(n))^b}\qquad\text{(ii)}\sum_{n=3}^{\infty}(\log n)^{a\frac{\log n}{\log\log n}}\qquad\text{(iii)}\sum_{n=1}^{\infty}\left(e^{1/n}-\tfrac{n+1}{n}\right)$$

$$\text{(iv)}\sum_{n=1}^{\infty}\cos(\pi n)\sin(\pi n^{-1})\qquad\text{(v)}\sum_{n=2}^{\infty}\left(\left(1+\frac{1}{n}\right)^n-e\right)^2\qquad\text{(vi)}\sum_{n=1}^{\infty}\frac{1}{n(n^{1/n})^{100}}$$

$$\text{(vii)}\sum_{n=2}^{\infty}2^{-(\log(n))^a}\qquad\text{(viii)}\sum_{n=1}^{\infty}\left(\sum_{k=0}^{10n}(-1)^k\frac{n^k}{k!}\right)\qquad\text{(ix)}\sum_{n=1}^{\infty}\frac{1}{n^2(1-\cos(n))}$$

EXERCISE 1.27. 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 1.28. For which $x \in \mathbb{R}$ do the following series converge? On which sets do these series converge uniformly?

$$\text{(i)} \sum_{n=1}^{\infty} n^2 x^n \qquad \text{(ii)} \sum_{n=1}^{\infty} (3^{1/n} - 1)^n x^n \qquad \text{(iii)} \sum_{n=1}^{\infty} \tan(n^{-2}) e^{nx} \tag{1.36}$$

$$\text{(iv)} \sum_{n=1}^{\infty} \frac{x^n}{n^n} \quad \text{(v)} \sum_{n=1}^{\infty} \frac{\sin(nx)}{n^2} \quad \text{(vi)} \sum_{n=1}^{\infty} 2^{-n} \tan(\lfloor x \rfloor + 1/n) \tag{1.37}$$

EXERCISE 1.29. (i) Give an example of a sequence $(f_n)_n$ of continuously differentiable functions defined on $\mathbb{R}$, uniformly convergent on $\mathbb{R}$ such that the limit $\lim_{n\to\infty} f_n'(x)$ does not exist for any value of $x \in \mathbb{R}$.
(ii) Give an example of a sequence $(f_n)_n$ of continuously differentiable functions defined on $\mathbb{R}$, uniformly convergent on $\mathbb{R}$ to some function f such that f is not differentiable.
(iii) Give an example of a sequence $(f_n)_n$ of continuous bounded functions on $\mathbb{R}$ that converges pointwise to some function f such that f is unbounded and not continuous.

EXERCISE 1.30. Determine the value of the series $\sum_{n=1}^{\infty} \frac{(-1)^n}{n(n+1)}$.

EXERCISE 1.31. For a positive real number x define

$$f(x) = \sum_{n=0}^{\infty} \frac{1}{n(n+1) + x}. \tag{1.38}$$

(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.

EXERCISE 1.32. 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

$$|f^{(n)}(x)| \leq A^n n! \tag{1.39}$$

for $|x| < R$. Show that there exists $r > 0$ such that for every $|x| < r$ we have that

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n. \tag{1.40}$$

(That is, prove that the series on the right hand side converges and that the limit is $f(x)$.)

CHAPTER 2

Compactness in metric spaces

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 2.1. 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 $X \subset \bigcup_{i\in I} G_i$.

(*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.)

DEFINITION 2.2. 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 $i_1, \dots, i_m \in I$ such that $X \subset \bigcup_{j=1}^{m} G_{i_j}$. This is also called the *Heine-Borel property*.

DEFINITION 2.3. A subset $A \subset X$ is called *compact* if $(A, d|_{A\times A})$ is a compact metric space. Here $d|_{A\times A}$ denotes the restriction of d to $A \times A$.

Lecture 4 (Wednesday, Sep 11)

Review of relative topology. If we have a metric space X and a subset $A \subset X$, then A as a metric space with the metric $d|_{A\times A}$ comes with its own open sets: a set $U \subset A$ is open in A if and only if for every $x \in U$ there exists $\varepsilon > 0$ such that $B_A(x,\varepsilon) = \{y \in A : d(x,y) < \varepsilon\} \subset U$. 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: a set $U \subset A$ is open in A if and only if there exists $V \subset X$ open (in X) such that $U = V \cap A$ (see Chapter 2 in Rudin's book).

EXAMPLE 2.4. 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, $V = (-1, \frac{1}{2})$.

THEOREM 2.5 (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 2.6. A subset $A \subset X$ is called *relatively compact* or *precompact* if the closure $\overline{A} \subset X$ is compact.

EXAMPLES 2.7.
- 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

$$\overline{B}(x_0,r) = \{x \in X : d(x,x_0) \le r\} \subset X \tag{2.1}$$

  is *not* necessarily compact (examples later).

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

FACT 2.8. *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,

$$\{U_i : i \in I\} \cup \{X\backslash A\} \tag{2.2}$$

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 2.9. 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 $\overline{E} = X$).

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 2.10. *Let X, Y be metric spaces and assume that X is compact. If $f : X \to Y$ 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

$$B_x = B(x, \delta_x/2) = \{y \in X \,:\, d_X(x, y) < \delta_x/2\}. \tag{2.3}$$

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

$$\delta = \tfrac{1}{2} \min(\delta_{x_1}, \ldots, \delta_{x_m}). \tag{2.4}$$

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

$$d_X(x_i, y) \leq d_X(x_i, x) + d_X(x, y) \leq \tfrac{1}{2}\delta_{x_i} + \delta \leq \delta_{x_i}. \tag{2.5}$$

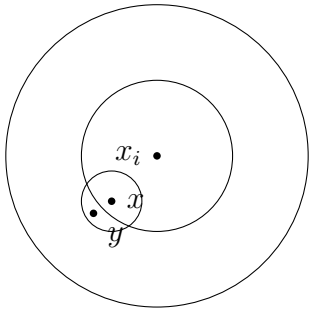

FIGURE 1. The balls B_{x_i}, $B(x_i, \delta_{x_i})$, $B(x, \delta)$.

Thus, by definition of δ_{x_i},

$$d_Y(f(x), f(y)) \leq d_Y(f(x), f(x_i)) + d_Y(f(x_i), f(y)) \leq \varepsilon/2 + \varepsilon/2 = \varepsilon. \tag{2.6}$$

□

THEOREM 2.11. *Let X, Y be metric spaces and assume that X is compact. If $f : X \to Y$ 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,

$$X \subset f^{-1}(f(X)) \subset \bigcup_{i\in I} f^{-1}(V_i) = \bigcup_{i\in I} U_i. \tag{2.7}$$

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,

$$X \subset \bigcup_{k=1}^{m} U_{i_k} \tag{2.8}$$

Consequently,

$$f(X) \subset \bigcup_{k=1}^{m} f(U_{i_k}) \subset \bigcup_{k=1}^{m} V_{i_k}. \tag{2.9}$$

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

Lecture 5 (Friday, Sep 13)

THEOREM 2.12. *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 2.11, $f(X) \subset \mathbb{R}$ is compact. By the Heine-Borel Theorem 2.5, 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 2.13. *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 2.43 below.

PROOF. Let $f \in C(X)$. Then $|f| : X \to [0, \infty)$ is also continuous. By Theorem 2.12 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. □

2. Sequential compactness and total boundedness

DEFINITION 2.14. 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 might have seen in Analysis I.

THEOREM 2.15 (Bolzano-Weierstrass). *Every bounded sequence in $\mathbb{R}$ has a convergent subsequence.*

DEFINITION 2.16. 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 2.17. 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

$$X \text{ totally bounded} \implies X \text{ bounded}. \tag{2.10}$$

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 2.18. *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 2.19.
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 2.19. 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 2.18. $\square$

EXAMPLE 2.20. Let ℓ^∞ be the space of bounded sequences $(a_n)_n \subset \mathbb{C}$ 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 $0 = (0, 0, \dots)$,

$$\overline{B}(0,1) = \{a \in \ell^\infty : |a_n| \leq 1 \, \forall n \in \mathbb{N}\} \tag{2.11}$$

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

$$e_n^{(k)} = \begin{cases} 0, & k \neq n, \\ 1, & k = n. \end{cases} \tag{2.12}$$

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 2.18 it is not compact.

EXAMPLE 2.21. Let ℓ^1 be the space of complex sequences $(a_n)_n \subset \mathbb{C}$ such that $\sum_n |a_n| < \infty$. We define a metric on ℓ^1 by

$$d(a,b) = \sum_n |a_n - b_n| \tag{2.13}$$

EXERCISE 2.22. Show that the closed and bounded set $\overline{B}(0,1) \in \ell^1$ is not compact.

Lecture 6 (Monday, Sep 16)

PROOF OF THEOREM 2.18. X compact $\Rightarrow$ X sequentially compact: Suppose that X is compact, but not sequentially compact. Then there exists a sequence $(x_n)_n \subset X$ without a convergent subsequence. Let $A = \{x_n : n \in \mathbb{N}\} \subset X$. Note that A must be an infinite set (otherwise $(x_n)_n$ has a constant subsequence). Since A has no limit 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.

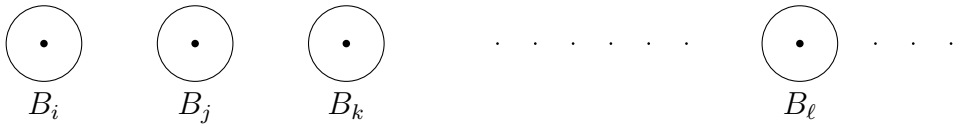

FIGURE 2.

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 $(p_n)_n$ has no convergent subsequence (no subsequence can be Cauchy). Contradiction! Thus, X is totally bounded.

Lecture 7 (Wednesday, Sep 18)

X totally bounded and complete $\Rightarrow$ X sequentially compact: Assume that X is totally bounded and complete. Let $(x_n)_n \subset X$ 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 $(x_n^{(0)})_n \supset (x_n^{(1)})_n \supset \dots$ of $(x_n)_n$ such that $(x_n^{(k)})_n$ is contained in a ball of radius 2^{-k}. Now let $a_n = x_n^{(n)}$. Then $(a_n)_n$ is a subsequence of $(x_n)_n$.

Claim: $(a_n)_n$ 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

$$d(a_m, a_n) \leq 2 \cdot 2^{-n} \leq 2^{-N+1} < \varepsilon, \tag{2.14}$$

because $a_n, a_m \in B_n$ and B_n is a ball of radius 2^{-n}. □

Since X is complete, the Cauchy sequence $(a_n)_n$ 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 $(p_n)_n$ has a convergent subsequence $(p_{n_k})_k$ 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

$$d(p, x) \leq d(p, p_{n_k}) + d(p_{n_k}, x) < \delta/2 + \delta/2 = \delta. \tag{2.15}$$

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

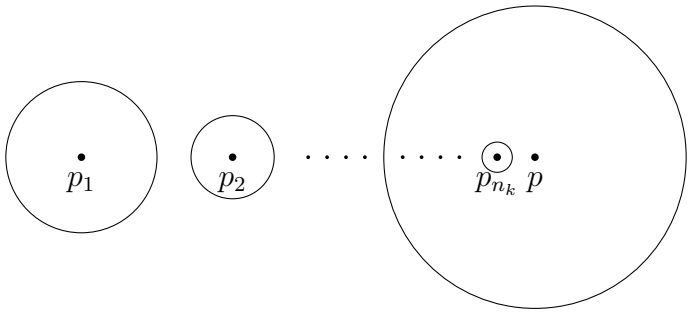

FIGURE 3.

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

$p_1, \ldots, p_M$ 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:

$$X \subset \bigcup_{j=1}^{M} B(p_j, \varepsilon) \subset \bigcup_{j=1}^{M} G_{i_j}. \tag{2.16}$$

□

COROLLARY 2.23. *Compact subsets of metric spaces are bounded and closed.*

COROLLARY 2.24. *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 2.25. Prove this.

Lecture 8 (Friday, Sep 20)

3. Equicontinuity and the Arzelà-Ascoli theorem

Let (K, d) be a compact metric space. By Corollary 2.13, continuous functions on K are automatically bounded. Thus, $C(K) = C_b(K)$ is a complete metric space with the supremum metric

$$d_\infty(f, g) = \sup_{x \in K} |f(x) - g(x)| \tag{2.17}$$

(see Fact 1.7). Convergence with respect to d_∞ is uniform convergence (see Fact 1.8). In this section we ask ourselves when a subset $\mathcal{F} \subset C(K)$ is compact.

EXAMPLE 2.26. Let $\mathcal{F} = \{f_n : n \in \mathbb{N}\} \subset C([0, 1])$, where

$$f_n(x) = x^n, \quad x \in [0, 1]. \tag{2.18}$$

$\mathcal{F}$ is not compact, because no subsequence of $(f_n)_n$ converges. This is because the pointwise limit

$$f(x) = \begin{cases} 0, & x \in [0, 1), \\ 1, & x = 1. \end{cases} \tag{2.19}$$

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

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

DEFINITION 2.27 (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 2.28. $\mathcal{F} \subset C(K)$ 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 2.16). We have

$$\mathcal{F} \text{ uniformly bounded } \Rightarrow \mathcal{F} \text{ pointwise bounded.} \tag{2.20}$$

The converse is false in general.

FACT 2.29. *If $(f_n)_n \subset C(K)$ 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

$$\sup_{x \in K} |f_n(x) - f_N(x)| \leq \varepsilon/3 \tag{2.21}$$

for $n \geq N$. By uniform continuity (using Theorem 2.10) there exists $\delta > 0$ such that

$$|f_k(x) - f_k(y)| \leq \varepsilon/3 \tag{2.22}$$

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

$$|f_n(x) - f_n(y)| \leq |f_n(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f_n(y)| \leq 3 \cdot \varepsilon/3 = \varepsilon. \tag{2.23}$$

□

FACT 2.30. *If $\mathcal{F} \subset C(K)$ is pointwise bounded and equicontinuous, then it is uniformly bounded.*

PROOF. Choose $\delta > 0$ such that

$$|f(x) - f(y)| \leq 1 \tag{2.24}$$

for all $d(x, y) < \delta$, $f \in \mathcal{F}$. Since K is totally bounded (by Theorem 2.18) 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

$$C := \max\{C(p_1), \ldots, C(p_m)\}. \tag{2.25}$$

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

$$|f(x)| \leq |f(p_j)| + |f(x) - f(p_j)| \leq C + 1, \tag{2.26}$$

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

THEOREM 2.31 (Arzelà-Ascoli). *Let K be a compact metric space. Then $\mathcal{F} \subset C(K)$ is relatively compact if and only if it is pointwise bounded and equicontinuous.*

COROLLARY 2.32. *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*

$$\mathcal{F}' = \{f' : f \in \mathcal{F}\} \tag{2.27}$$

is bounded.

Then $\mathcal{F}$ is relatively compact.

Lecture 9 (Monday, Sep 23)

EXAMPLE 2.33. 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

$$\Big|\sum_{n=0}^{\infty} c_n x^n\Big| \le \sum_{n=0}^{\infty} 2^{-n} = 2. \tag{2.28}$$

for all sequences $(c_n)_n$ with $|c_n| \le 1$ and for all $x \in [-1/2, 1/2]$. Similarly,

$$\mathcal{F}' = \Big\{ \sum_{n=1}^{\infty} n c_n x^{n-1} \,:\, |c_n| \le 1 \Big\} \tag{2.29}$$

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 2.26).

EXAMPLE 2.34. The set

$$\mathcal{F} = \Big\{ \sin(\pi n x) \,:\, n \in \mathbb{Z} \Big\} \subset C([0,1]) \tag{2.30}$$

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!

PROOF OF THEOREM 2.31. $\Longleftarrow$: Without loss of generality assume that $\mathcal{F}$ is closed. Assume that $\mathcal{F} \subset C(K)$ is pointwise bounded and equicontinuous. By Fact 2.30 it follows that $\mathcal{F}$ is uniformly bounded. By Exercise 2.9 there exists a countable, dense set $E \subset K$. Let $(f_n)_n \subset \mathcal{F}$ be a sequence.

Claim: There exists a subsequence $(f_{n_k})_k$ such that $(f_{n_k}(p))_k$ converges for all $p \in E$.

Proof of claim. This is again a diagonal subsequence argument. Let $E = \{p_1, p_2, \dots\}$. By the Bolzano-Weierstrass theorem (see Corollary 2.19 or Theorem 2.15) we have the following inductive claim: for all $j \in \mathbb{N}$ there is a subsequence $(f_{n_k^{(j)}})_k$ such that $(f_{n_k^{(j)}}(p_\ell))_k$ converges for all $1 \le \ell \le j$. Indeed, for $j = 1$ this is a direct consequence of Bolzano-Weierstrass because $(f_n(p_1))_n$ is just a bounded sequence of complex numbers. Assume we proved the claim up to some $j \ge 1$. Then by one more application of Bolzano-Weierstrass, there is a subsequence $(f_{n_k^{(j+1)}})_k$ of $(f_{n_k^{(j)}})_k$ (and therefore a subsequence of $(f_n)_n$) such that $(f_{n_k^{(j+1)}}(p_{j+1}))_k$ converges. Now we set $n_k = n_k^{(k)}$ and observe that $(f_{n_k}(p_j))_k$ converges for all j. $\square$

Let us set $g_k = f_{n_k}$ to simplify notation. We show that $(g_k)_k$ converges uniformly on K. Let $\varepsilon > 0$. By equicontinuity there exists $\delta > 0$ such that

$$|g_k(x) - g_k(y)| < \varepsilon/3 \tag{2.31}$$

for all $x, y \in K$ with $d(x,y) < \delta$ and all $k \in \mathbb{N}$. Since K is compact and E is dense, there exist $p_1, \dots, p_m \in E$ with

$$K \subset B(p_1, \delta) \cup \cdots \cup B(p_m, \delta). \tag{2.32}$$

Since $(g_k(p_j))_k$ converges, there exists $N_j \in \mathbb{N}$ such that

$$|g_k(p_j) - g_\ell(p_j)| < \varepsilon/3 \tag{2.33}$$

for all $k, \ell \geq N_j$. Set

$$N = \max\{N_1, \dots, N_m\}. \tag{2.34}$$

Then,

$$|g_k(p_j) - g_\ell(p_j)| < \varepsilon/3 \tag{2.35}$$

for all $k, \ell \geq N$ and all $j = 1, \dots, m$. Let $x \in K$. By (2.32) there exists $j \in \{1, \dots, m\}$ such that $x \in B(p_j, \delta)$. Then from (2.31) and (2.33) we have

$$|g_k(x) - g_\ell(x)| \leq |g_k(x) - g_k(p_j)| + |g_k(p_j) - g_\ell(p_j)| + |g_\ell(p_j) - g_\ell(x)| \leq \varepsilon. \tag{2.36}$$

Thus $(g_k)_k$ converges uniformly and therefore converges to a limit in $\overline{\mathcal{F}}$. By Theorem 2.18, this shows that $\mathcal{F}$ is relatively compact.

$\Longrightarrow$: Assume that $\mathcal{F}$ is relatively compact. Then it is bounded. Say that $\mathcal{F}$ is not equicontinuous. Then there exists $\varepsilon > 0$ and a sequence $(f_n)_n \subset \mathcal{F}$ and points $(x_n)_n, (y_n)_n \subset K$ with $d(x_n, y_n) < \frac{1}{n}$ such that

$$|f_n(x_n) - f_n(y_n)| \geq \varepsilon. \tag{2.37}$$

By (relative) compactness, $(f_n)_n$ has a uniformly convergent subsequence which we will also call $(f_n)_n$ for simplicity. Then f_n converges to some limit $f \in \overline{\mathcal{F}} \subset C(K)$. By uniform convergence, there exists $N \in \mathbb{N}$ such that

$$|f_n(x) - f(x)| < \varepsilon/3 \tag{2.38}$$

for all $n \geq N$ and $x \in K$. By uniform continuity of f (using Theorem 2.10) there exists $\delta > 0$ such that

$$|f(x) - f(y)| < \varepsilon/3 \tag{2.39}$$

for all $x, y \in K$ with $d(x, y) < \delta$. Let $n \geq \max\{N, \delta^{-1}\}$. Then by (2.38) and (2.39) we have

$$|f_n(x_n) - f_n(y_n)| \leq |f_n(x_n) - f(x_n)| + |f(x_n) - f(y_n)| + |f(y_n) - f_n(y_n)| < \varepsilon. \tag{2.40}$$

This contradicts (2.37). □

Lecture 10 (Wednesday, Sep 25)

Proof of Corollary 2.32. Using the mean value theorem we see that for all $x, y \in [a, b]$ there exists $\xi \in [a, b]$ such that

$$f(x) - f(y) = f'(\xi)(x - y). \tag{2.41}$$

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

$$|f'(\xi)| \leq C \tag{2.42}$$

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

$$|f(x) - f(x)| \leq C|x - y| \tag{2.43}$$

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

$$|f(x) - f(y)| \leq C|x - y| < C\delta = \varepsilon. \tag{2.44}$$

Therefore the claim follows from Theorem 2.31. □

EXAMPLE 2.35. Condition (i) from Corollary 2.32 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 $(f_n)_n$ is uniformly convergent, so by Fact 2.29, $\mathcal{F}$ is equicontinuous and hence relatively compact.

4. Further exercises

EXERCISE 2.36. 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 2.37. Let ℓ^1 denote space of all sequences $(a_n)_n$ 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

$$A = \{a \in \ell^1 \, : \, \sum_{n=1}^{\infty} |a_n| \leq 1\} \tag{2.45}$$

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

$$B = \{a \in \ell^1 \, : \, |a_n| \leq b_n \, \forall \, n \in \mathbb{N}\} \tag{2.46}$$

is compact.

EXERCISE 2.38. 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

$$A = \{a \in \ell^\infty \, : \, |a_n| \leq s_n \text{ for all } n\}. \tag{2.47}$$

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

EXERCISE 2.39. 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 2.40. 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

$$\lim_{t\to 0+} \omega(t) = \omega(0) = 0. \tag{2.48}$$

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

$$|f(x) - f(y)| \leq \omega(|x - y|). \tag{2.49}$$

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

EXERCISE 2.41. For $1 \leq p < \infty$ we denote by ℓ^p the space of sequences $(a_n)_n$ of complex numbers such that $\sum_{n=1}^{\infty} |a_n|^p < \infty$. Define a metric on ℓ^p by

$$d(a, b) = \left(\sum_{n\in\mathbb{N}} |a_n - b_n|^p \right)^{1/p}. \tag{2.50}$$

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

$$\sum_{n=N}^{\infty} |a_n|^p < \varepsilon \text{ for all } a \in \mathcal{F}. \tag{2.51}$$

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

$$d(f,g) = \sum_{j=0}^{k} \sup_{x\in[a,b]} |f^{(j)}(x) - g^{(j)}(x)|. \tag{2.52}$$

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

$$\iota : C^k([a,b]) \to C^\ell([a,b]) \text{ with } \iota(f) = f. \tag{2.53}$$

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 2.43. 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 $|f(x)| \le C_f$ 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 2.44. Consider $\mathcal{F} = \{f_N : N \in \mathbb{N}\} \subset C([0,1])$ with

$$f_N(x) = \sum_{n=0}^{N} b^{-n\alpha} \sin(b^n x), \tag{2.54}$$

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 $N \in \mathbb{N}$,

$$|f_N(x) - f_N(y)| \le c|x-y|^\alpha. \tag{2.55}$$

EXERCISE 2.45. 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 $a = (a_n)_n$ 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 3

Approximation theory

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 3.1 (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?).

Lecture 11 (Friday, Sep 27)

Let f be continuous on $[0,1]$. Define for $n = 1, 2, \ldots$:

$$B_n f(t) = \sum_{k=0}^{n} f\Big(\frac{k}{n}\Big)\binom{n}{k} t^k (1-t)^{n-k}. \tag{3.1}$$

$B_n f$ is a polynomial of degree n. We will show that $B_n f \to f$ uniformly on $[0,1]$. By the binomial theorem,

$$1 = (t + 1 - t)^n = \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k}. \tag{3.2}$$

Thus,

$$B_n f(t) - f(t) = \sum_{k=0}^{n} (f(k/n) - f(t))\binom{n}{k} t^k (1-t)^{n-k}. \tag{3.3}$$

Let $\varepsilon > 0$. By uniform continuity of f we choose $\delta > 0$ be such that $|f(t) - f(s)| \le \varepsilon/2$ for all $t, s \in [0,1]$ with $|t - s| \le \delta$. Now we write the sum on the right hand side of (3.3) as $\mathrm{I} + \mathrm{II}$, where

$$\mathrm{I} = \sum_{\substack{k=0,\\ |\frac{k}{n} - t| < \delta}}^{n} (f(k/n) - f(t))\binom{n}{k} t^k (1-t)^{n-k}, \tag{3.4}$$

$$\mathrm{II} = \sum_{\substack{k=0,\\ |\frac{k}{n} - t| \ge \delta}}^{n} (f(k/n) - f(t))\binom{n}{k} t^k (1-t)^{n-k}. \tag{3.5}$$

We estimate I and II separately. For I we have from uniform continuity that

$$|\mathrm{I}| \le \varepsilon/2 \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k} = \varepsilon/2. \tag{3.6}$$

Lecture 12 (Monday, Sep 30)

To estimate II we first compute the Bernstein polynomials for the monomials $1, t, t^2$.

LEMMA 3.2. *Let $g_m(t) = t^m$. Then*

$$B_n g_0(t) = 1 \tag{3.7}$$

$$B_n g_1(t) = t \tag{3.8}$$

$$B_n g_2(t) = t^2 + \frac{t - t^2}{n} \quad \textit{for } n \geq 2 \tag{3.9}$$

PROOF. We have

$$B_n g_0(t) = \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k} = (t + (1-t))^n = 1 \tag{3.10}$$

by the binomial theorem. Next,

$$\begin{aligned} B_n g_1(t) &= \sum_{k=0}^{n} \frac{k}{n} \binom{n}{k} t^k (1-t)^{n-k} = \sum_{k=1}^{n} \binom{n-1}{k-1} t^k (1-t)^{n-k} \\ &= t \sum_{k=0}^{n-1} \binom{n-1}{k} t^k (1-t)^{(n-1)-k} = t(t + (1-t))^{n-1} = t. \end{aligned} \tag{3.11}$$

To compute $B_n g_2$ we use that

$$\begin{aligned} \frac{k^2}{n^2} \binom{n}{k} &= \frac{k}{n} \binom{n-1}{k-1} = \frac{n-1}{n} \frac{k-1}{n-1} \binom{n-1}{k-1} + \frac{1}{n} \binom{n-1}{k-1} \\ &= \frac{n-1}{n} \binom{n-2}{k-2} + \frac{1}{n} \binom{n-1}{k-1}. \end{aligned} \tag{3.12}$$

Thus,

$$\begin{aligned} B_n g_2(t) &= \frac{n-1}{n} \sum_{k=2}^{n} \binom{n-2}{k-2} t^k (1-t)^{n-k} + \frac{1}{n} \sum_{k=1}^{n} \binom{n-1}{k-1} t^k (1-t)^{n-k} \\ &= \frac{n-1}{n} t^2 + \frac{1}{n} t = t^2 + \frac{t - t^2}{n}. \end{aligned} \tag{3.13}$$

□

As a consequence, we obtain the following:

LEMMA 3.3. *For all $t \in [0, 1]$,*

$$\sum_{k=0}^{n} (\tfrac{k}{n} - t)^2 \binom{n}{k} t^k (1-t)^{n-k} \leq \frac{1}{n}. \tag{3.14}$$

PROOF. From the previous lemma,

$$\begin{aligned} \sum_{k=0}^{n} (\tfrac{k}{n} - t)^2 \binom{n}{k} t^k (1-t)^{n-k} &= B_n g_2(t) - 2t B_n g_1(t) + t^2 B_n g_0(t) \\ &= t^2 + \frac{t - t^2}{n} - 2t^2 + t^2 = \frac{t - t^2}{n}. \end{aligned} \tag{3.15}$$

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 $n \geq N$,

$$|\mathrm{II}| \leq 2c \sum_{\substack{k=0,\\ |\frac{k}{n}-t|\geq\delta}}^{n} \binom{n}{k} t^k (1-t)^{n-k} \leq 2c\delta^{-2} \sum_{k=0}^{n} (\tfrac{k}{n}-t)^2 \binom{n}{k} t^k (1-t)^{n-k} \tag{3.16}$$

$$\leq 2c\delta^{-2}N^{-1} \leq \varepsilon/2.$$

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

$$|B_n f(t) - f(t)| \leq |\mathrm{I}| + |\mathrm{II}| \leq \varepsilon/2 + \varepsilon/2 = \varepsilon. \tag{3.17}$$

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 3.4. For two piecewise continuous functions f, g on an interval $[a,b]$ we define their *inner product* by

$$\langle f, g\rangle = \int_a^b f(x)\overline{g(x)}dx. \tag{3.18}$$

If $\langle f, g\rangle = 0$ we say that f and g are *orthogonal*. We define the *L^2-norm* of f by

$$\|f\|_2 = \left(\int_a^b |f(x)|^2 dx\right)^{1/2}. \tag{3.19}$$

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:

$$\langle f + \lambda g, h\rangle = \langle f, h\rangle + \lambda\langle g, h\rangle, \tag{3.20}$$

$$\langle h, f + \lambda g\rangle = \langle h, f\rangle + \overline{\lambda}\langle h, g\rangle. \tag{3.21}$$

- 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 3.5 (Cauchy-Schwarz inequality). *For two piecewise continuous functions f, g we have*

$$|\langle f, g\rangle| \leq \|f\|_2 \|g\|_2. \tag{3.22}$$

PROOF. For nonnegative real numbers x and y we have the elementary inequality

$$xy \leq \frac{x^2}{2} + \frac{y^2}{2}. \tag{3.23}$$

Thus we have

$$|\langle f, g\rangle| \leq \int_a^b |f(x)g(x)|dx \leq \tfrac{1}{2}\int_a^b |f(x)|^2 dx + \tfrac{1}{2}\int_a^b |g(x)|^2 dx. = \tfrac{1}{2}\langle f, f\rangle + \tfrac{1}{2}\langle g, g\rangle. \tag{3.24}$$

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

$$|\langle f, g\rangle| \leq \tfrac{\lambda^2}{2}\langle f, f\rangle + \tfrac{1}{2\lambda^2}\langle g, g\rangle. \tag{3.25}$$

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

$$|\langle f, g\rangle| \leq \sqrt{\langle f, f\rangle}\sqrt{\langle g, g\rangle}. \tag{3.26}$$

Note that one can arrive at this definition of λ in a systematic way: treat the right hand side of (3.25) as a function of λ and minimize it using calculus. □

COROLLARY 3.6 (Minkowski's inequality). *For two functions $f, g \in \mathrm{pc}([a, b])$,*

$$\|f + g\|_2 \leq \|f\|_2 + \|g\|_2. \tag{3.27}$$

PROOF. We may assume $\|f + g\|_2 \neq 0$ because otherwise there is nothing to prove. Then

$$\|f + g\|_2^2 = \int_a^b |f + g|^2 \leq \int_a^b |f + g||f| + \int_a^b |f + g||g| \tag{3.28}$$

$$\leq \|f + g\|_2\|f\|_2 + \|f + g\|_2\|g\|_2 = \|f + g\|_2(\|f\|_2 + \|g\|_2). \tag{3.29}$$

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 3.70.)

Lecture 13 (Wednesday, October 2)

DEFINITION 3.7. A sequence $(\phi_n)_n$ of piecewise continuous functions on $[a,b]$ is called an *orthonormal system on* $[a,b]$ if

$$\langle \phi_n, \phi_m \rangle = \int_a^b \phi_n(x)\overline{\phi_m(x)}dx = \begin{cases} 0, & \text{if } n \neq m, \\ 1, & \text{if } n = m. \end{cases} \tag{3.30}$$

(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 $\mathbf{1}_A$ the *characteristic function of* A. This is the function such that $\mathbf{1}_A(x) = 1$ when $x \in A$ and $\mathbf{1}_A(x) = 0$ when $x \notin A$.

EXAMPLE 3.8 (Disjoint support). Let $\phi_n(x) = \mathbf{1}_{[n,n+1)}$ and $N \in \mathbb{N}$. Then $(\phi_n)_{n=0,\ldots,N-1}$ is an orthonormal system on $[0,N]$.

EXAMPLES 3.9 (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 3.10 (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 $(r_n)_n$ 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,

$$f(x) = \sum_{n=1}^{N} c_n \phi_n(x). \tag{3.31}$$

Then there is an easy way to compute the coefficients c_n:

$$c_n = \langle f, \phi_n \rangle = \int_a^b f(x)\overline{\phi_n(x)}dx. \tag{3.32}$$

To prove this we multiply (3.31) by $\overline{\phi_m(x)}$ and integrate over x:

$$\int_a^b f(x)\overline{\phi_m(x)}dx = \sum_{n=1}^{N} c_n \int_a^b \phi_n(x)\overline{\phi_m(x)}dx = \sum_{n=1}^{N} c_n \langle \phi_n, \phi_m \rangle = c_m. \tag{3.33}$$

Notice that the formula $c_n = \langle f, \phi_n \rangle$ still makes sense if f is not of the form (3.31).

THEOREM 3.11. *Let $(\phi_n)_n$ be an orthonormal system on $[a,b]$. Let f be a piecewise continuous function. Consider*

$$s_N(x) = \sum_{n=1}^{N} \langle f, \phi_n \rangle \phi_n(x). \tag{3.34}$$

Denote the linear span of the functions $(\phi_n)_{n=1,\ldots,N}$ by X_N. Then

$$\|f - s_N\|_2 \leq \|f - g\|_2 \tag{3.35}$$

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 (3.35) 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).

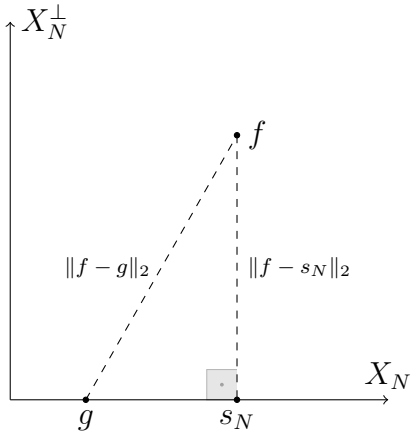

FIGURE 1. s_N is the orthogonal projection of f onto X_N.

Lecture 14 (Friday, October 4)

THEOREM 3.12 (Bessel's inequality). *If $(\phi_n)_n$ is an orthonormal system on $[a,b]$ and f a piecewise continuous function on $[a,b]$ then*

$$\sum_n |\langle f, \phi_n\rangle|^2 \leq \|f\|_2^2. \tag{3.36}$$

COROLLARY 3.13 (Riemann-Lebesgue lemma). *Let $(\phi_n)_{n=1,2,\dots}$ be an orthonormal system and f a piecewise continuous function. Then*

$$\lim_{n\to\infty} \langle f, \phi_n\rangle = 0. \tag{3.37}$$

This follows because the series $\sum_{n=1}^{\infty} |\langle f, \phi_n\rangle|^2$ converges as a consequence of Bessel's inequality.

DEFINITION 3.14. An orthonormal system $(\phi_n)_n$ is called *complete* if

$$\sum_n |\langle f, \phi_n\rangle|^2 = \|f\|_2^2 \tag{3.38}$$

for all f.

THEOREM 3.15. *Let $(\phi_n)_n$ be an orthonormal system on $[a,b]$. Let $(s_N)_N$ be as in Theorem 3.11. 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 3.11. Let $g \in X_N$ and write

$$g(x) = \sum_{n=1}^{N} b_n \phi_n(x). \tag{3.39}$$

Let us also write

$$c_n = \langle f, \phi_n\rangle. \tag{3.40}$$

We have

$$\langle f, g\rangle = \sum_{n=1}^{N} \overline{b_n}\langle f, \phi_n\rangle = \sum_{n=1}^{N} c_n \overline{b_n}. \tag{3.41}$$

Using that $(\phi_n)_n$ is orthonormal we get

$$\langle g, g\rangle = \Big\langle \sum_{n=1}^{N} b_n\phi_n, \sum_{m=1}^{N} b_m\phi_m \Big\rangle = \sum_{n=1}^{N}\sum_{m=1}^{N} b_n\overline{b_m}\langle \phi_n, \phi_m\rangle = \sum_{n=1}^{N} |b_n|^2. \tag{3.42}$$

Thus,

$$\langle f-g, f-g\rangle = \langle f, f\rangle - \langle f, g\rangle - \langle g, f\rangle + \langle g, g\rangle \tag{3.43}$$

$$= \langle f, f\rangle - \sum_{n=1}^{N} c_n\overline{b_n} - \sum_{n=1}^{N} \overline{c_n} b_n + \sum_{n=1}^{N} |b_n|^2 \tag{3.44}$$

$$= \langle f, f\rangle - \sum_{n=1}^{N} |c_n|^2 + \sum_{n=1}^{N} |b_n - c_n|^2 \tag{3.45}$$

We have

$$\begin{aligned}\langle f - s_N, f - s_N\rangle &= \langle f, f\rangle - \langle f, s_N\rangle - \langle s_N, f\rangle + \langle s_N, s_N\rangle \\ &= \langle f, f\rangle - 2\sum_{n=1}^{N} |c_n|^2 + \sum_{n=1}^{N} |c_n|^2 = \langle f, f\rangle - \sum_{n=1}^{N} |c_n|^2.\end{aligned} \tag{3.46}$$

Thus we have shown

$$\langle f - g, f - g\rangle = \langle f - s_N, f - s_N\rangle + \sum_{n=1}^{N} |b_n - c_n|^2 \tag{3.47}$$

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 $n = 1, \ldots, N$. □

Proof of Theorem 3.12. From the calculation in (3.46),

$$\langle f, f\rangle - \sum_{n=1}^{N} |c_n|^2 = \langle f - s_N, f - s_N\rangle \geq 0, \tag{3.48}$$

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 3.15. From (3.46),

$$\|f - s_N\|_2^2 = \langle f, f\rangle - \sum_{n=1}^{N} |\langle f, \phi_n\rangle|^2 \tag{3.49}$$

This converges to 0 as $N \to \infty$ if and only if $(\phi_n)_n$ is complete. □

Lecture 15 (Monday, October 7)

3. The Haar system

In this section we discuss an important example of an orthonormal system on $[0,1]$.

DEFINITION 3.16 (Dyadic intervals). For non-negative integers j,k with $0 \le j < 2^k$ we define

$$I_{k,j} = [2^{-k}j, 2^{-k}(j+1)) \subset [0,1]. \tag{3.50}$$

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 $\mathcal{D} = \bigcup_{k\ge 0} \mathcal{D}_k$ the set of all dyadic intervals on $[0,1]$.

DEFINITION 3.17. 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 3.18. 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 $I_r = [\frac{1}{2}+\frac{1}{8}, \frac{1}{2}+\frac{1}{4})$.

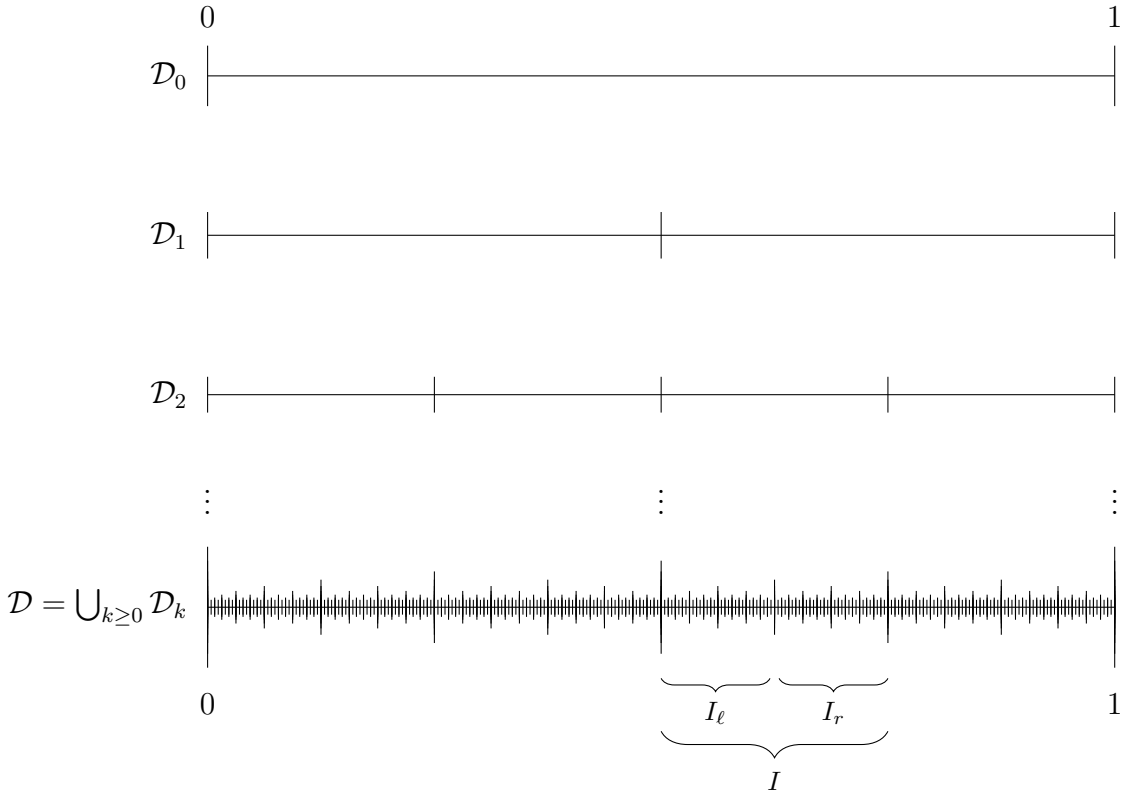

FIGURE 2. Dyadic intervals.

LEMMA 3.19. (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,*

$$[0,1) = \bigcup_{I \in \mathcal{D}_k} I. \tag{3.51}$$

EXERCISE 3.20. Prove this lemma.

EXERCISE 3.21. 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 3.22. For each $I \in \mathcal{D}$ we define the *Haar function* associated with it by

$$\psi_I = |I|^{-1/2}(\mathbf{1}_{I_\ell} - \mathbf{1}_{I_r}) \tag{3.52}$$

The countable set of functions given by

$$\mathcal{H} = \{\mathbf{1}_{[0,1]}\} \cup \{\psi_I \,:\, I \in \mathcal{D}\} \tag{3.53}$$

is called the *Haar system* on $[0,1]$.

EXAMPLE 3.23. The Haar function associated with the dyadic interval $I = [0, \frac{1}{2})$ is given by

$$\psi_{[0,\frac{1}{2}]} = \sqrt{2} \cdot (\mathbf{1}_{[0,\frac{1}{4})} - \mathbf{1}_{[\frac{1}{4},\frac{1}{2})}) \tag{3.54}$$

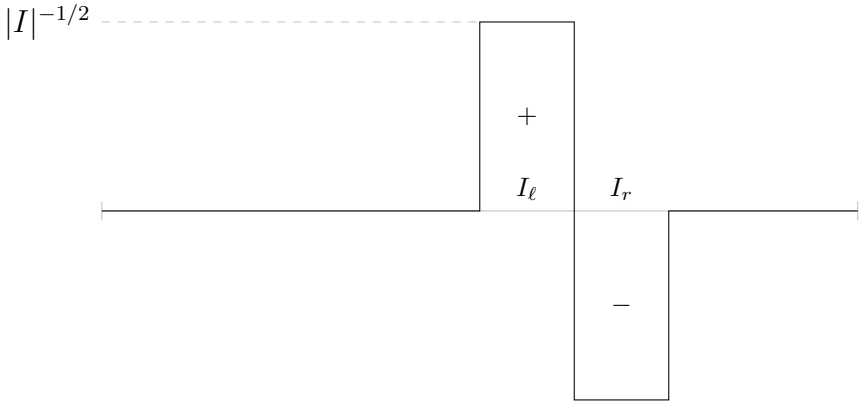

FIGURE 3. A Haar function ψ_I.

LEMMA 3.24. *The Haar system on $[0,1]$ is an orthonormal system.*

PROOF. Let $f \in \mathcal{H}$. If $f = \mathbf{1}_{[0,1]}$ then $\|f\|_2 = (\int_0^1 1^2)^{1/2} = 1$. Otherwise, $f = \psi_I$ for some $I \in \mathcal{D}$. Then by (3.52) and since I_ℓ and I_r are disjoint,

$$\|f\|_2^2 = \int_0^1 |\psi_I|^2 = |I|^{-1}\Big(\int_0^1 \mathbf{1}_{I_\ell} + \int_0^1 \mathbf{1}_{I_r}\Big) = 1. \tag{3.55}$$

Next let $f, g \in \mathcal{H}$ with $f \neq g$. Suppose that one of f, g equals $\mathbf{1}_{[0,1]}$, say $f = \mathbf{1}_{[0,1]}$. Then $g = \psi_J$ for some $J \in \mathcal{D}$ and thus

$$\langle f, g \rangle = \int_0^1 \psi_J = 0. \tag{3.56}$$

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 3.19 (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,

$$\langle \psi_I, \psi_J \rangle = \int \psi_I \cdot \psi_J = \pm |I|^{-1} \int_0^1 (\mathbf{1}_{I_\ell} - \mathbf{1}_{I_r}) = 0. \tag{3.57}$$

□

Let us write

$$\mathcal{D}_{<n} = \bigcup_{0 \leq k < n} \mathcal{D}_k \tag{3.58}$$

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 3.11, we define for every positive integer n, the orthogonal projection

$$\mathbf{E}_n f = \sum_{I \in \mathcal{D}_{<n}} \langle f, \psi_I \rangle \psi_I. \tag{3.59}$$

Definition 3.25. 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 3.26. *Let $\int_0^1 f = 0$. Then, for every $I \in \mathcal{D}_n$,*

$$\mathbf{E}_n f(x) = \langle f \rangle_I \quad \text{if } x \in I. \tag{3.60}$$

In other words,

$$\mathbf{E}_n f = \sum_{I \in \mathcal{D}_n} \langle f \rangle_I \mathbf{1}_I. \tag{3.61}$$

Theorem 3.27. *Suppose that $\int_0^1 f = 0$ and $f \in C([0,1])$. Then*

$$\mathbf{E}_n f \to f \quad \text{uniformly on } [0,1] \text{ as } n \to \infty. \tag{3.62}$$

Remark. If $f \in C([0,1])$ does not have mean zero then $\mathbf{E}_n f$ converges to $f - \langle f \rangle_{[0,1]}$.

Corollary 3.28. *The Haar system is complete in the sense of Definition 3.14. For every $f \in C([0,1])$ we have*

$$\|f\|_2^2 = |\langle f \rangle_{[0,1]}|^2 + \sum_{I \in \mathcal{D}} |\langle f, \psi_I \rangle|^2. \tag{3.63}$$

Exercise 3.29. By using Theorem 3.27, prove Corollary 3.28.

Proof of Theorem 3.26. Fix $n \geq 0$ and write $g = \mathbf{E}_n f$. 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 $\mathbf{E}_n f$ 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

$$\langle f \rangle_{I_\ell} = \langle g \rangle_{I_\ell} \quad \text{and} \quad \langle f \rangle_{I_r} = \langle g \rangle_{I_r}. \tag{3.64}$$

Since the Haar system is orthonormal and $I \in \mathcal{D}_{<n}$,

$$\langle g, \psi_I \rangle = \sum_{J \in \mathcal{D}_{<n}} \langle f, \psi_J \rangle \langle \psi_J, \psi_I \rangle = \langle f, \psi_I \rangle. \tag{3.65}$$

Compute

$$\int_{I_\ell} f - \int_{I_r} f = |I|^{1/2} \int f \cdot \psi_I = |I|^{1/2} \langle f, \psi_I \rangle \tag{3.66}$$

and by the same reasoning,

$$\int_{I_\ell} g - \int_{I_r} g = |I|^{1/2} \langle g, \psi_I \rangle. \tag{3.67}$$

Combining the last three displays we get

$$\int_{I_\ell} f - \int_{I_r} f = \int_{I_\ell} g - \int_{I_r} g. \tag{3.68}$$

By the inductive hypothesis we know that $\langle f \rangle_I = \langle g \rangle_I$, so

$$\int_{I_\ell} f + \int_{I_r} f = \int_{I_\ell} g + \int_{I_r} g. \tag{3.69}$$

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 3.27. Let $\varepsilon > 0$. By uniform continuity of f on $[0,1]$ (which follows from Theorem 2.10) 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 3.26,

$$|\mathbf{E}_n f(t) - f(t)| = |\langle f \rangle_I - f(t)| \leq |I|^{-1} \int_I |f(s) - f(t)| ds < \varepsilon. \tag{3.70}$$

□

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 $(\mathbf{E}_n f)_n$ are also called *dyadic martingale averages* of f and have wide applications in modern analysis and probability theory.

Exercise 3.30. Recall the functions $r_n(x) = \operatorname{sgn}(\sin(2^n \pi x))$ from Exercise 3.10.
(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 3.31. Define

$$\Delta_n f = \mathbf{E}_{n+1} f - \mathbf{E}_n f, \quad Sf = \big(\sum_{n \geq 1} |\Delta_n f|^2 \big)^{1/2}. \tag{3.71}$$

(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 finite linear combination of Haar functions f_m such that $\sup_{x \in [0,1]} |f_m(x)| \leq 1$ and $\sup_{x \in [0,1]} |Sf_m(x)| \geq m$.

Lecture 16 (Wednesday, October 9)

4. Trigonometric polynomials

In the following we will only be concerned with the *trigonometric system* on $[0,1]$:

$$\phi_n(x) = e^{2\pi inx} \quad (n \in \mathbb{Z}) \tag{3.72}$$

DEFINITION 3.32. A *trigonometric polynomial* is a function of the form

$$f(x) = \sum_{n=-N}^{N} c_n e^{2\pi inx} \quad (x \in \mathbb{R}), \tag{3.73}$$

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 1.21) we see that every trigonometric polynomial can also be written in the alternate form

$$f(x) = a_0 + \sum_{n=1}^{N} (a_n \cos(2\pi nx) + b_n \sin(2\pi nx)). \tag{3.74}$$

EXERCISE 3.33. Work out how the coefficients a_n, b_n in (3.74) are related to the c_n in (3.73).

Every trigonometric polynomial is *1-periodic*:

$$f(x) = f(x+1) \tag{3.75}$$

for all $x \in \mathbb{R}$.

FACT 3.34. $(e^{2\pi inx})_{n\in\mathbb{Z}}$ *forms an orthonormal system on* $[0,1]$. *In particular,*
(i) for all $n \in \mathbb{Z}$,

$$\int_0^1 e^{2\pi inx} dx = \begin{cases} 0, & \text{if } n \neq 0, \\ 1, & \text{if } n = 0. \end{cases} \tag{3.76}$$

(ii) if $f(x) = \sum_{n=-N}^{N} c_n e^{2\pi inx}$ *is a trigonometric polynomial, then*

$$c_n = \int_0^1 f(t) e^{-2\pi int} dt. \tag{3.77}$$

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 3.35. For a 1-periodic function $f \in$ pc and $n \in \mathbb{Z}$ we define the *nth Fourier coefficient* by

$$\widehat{f}(n) = \int_0^1 f(t) e^{-2\pi int} dt. \tag{3.78}$$

The series

$$\sum_{n=-\infty}^{\infty} \widehat{f}(n) e^{2\pi inx} \tag{3.79}$$

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 3.36. For a 1-periodic function $f \in \mathrm{pc}$ we define the *partial sums*

$$S_N f(x) = \sum_{n=-N}^{N} \widehat{f}(n) e^{2\pi i n x}. \tag{3.80}$$

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 3.11 tells us that

$$\|f - S_N f\|_2 \leq \|f - g\|_2 \tag{3.81}$$

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 3.37 (Convolution). For two 1-periodic functions $f, g \in \mathrm{pc}$ we define their *convolution* by

$$f * g(x) = \int_0^1 f(t) g(x-t) dt \tag{3.82}$$

Note that if $f, g \in \mathrm{pc}$ then $f * g \in \mathrm{pc}$.

EXAMPLE 3.38. 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 $[x - 1/N, x + 1/N]$.

FACT 3.39. *For* 1*-periodic functions* $f, g \in \mathrm{pc}$*,*

$$f * g = g * f. \tag{3.83}$$

PROOF. For $x \in [0, 1]$,

$$f*g(x) = \int_0^1 f(t)g(x-t)dt = \int_{x-1}^{x} f(x-t)g(t)dt = \int_{x-1}^{0} f(x-t)g(t)dt + \int_0^x f(x-t)g(t). \tag{3.84}$$

$$= \int_x^1 f(x-(t-1))g(t-1)dt + \int_0^x f(x-t)g(t)dt = g * f(x), \tag{3.85}$$

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:

$$S_N f(x) = \sum_{n=-N}^{N} \int_0^1 f(t) e^{-2\pi i n t} dt e^{2\pi i n x} = \int_0^1 f(t) \sum_{n=-N}^{N} e^{2\pi i n (x-t)} dt = f * D_N(x). \tag{3.86}$$

where

$$D_N(x) = \sum_{n=-N}^{N} e^{2\pi i n x}. \tag{3.87}$$

The sequence of functions $(D_N)_N$ is called *Dirichlet kernel*. The Dirichlet kernel can be written more explicitly.

FACT 3.40. *We have*

$$D_N(x) = \frac{\sin(2\pi(N+\frac{1}{2})x)}{\sin(\pi x)} \tag{3.88}$$

PROOF.

$$D_N(x) = \sum_{n=-N}^{N} e^{2\pi inx} = e^{-2\pi iNx} \sum_{n=0}^{2N} e^{2\pi inx} = e^{-2\pi iNx} \frac{e^{2\pi i(2N+1)x} - 1}{e^{2\pi ix} - 1} \tag{3.89}$$

$$= \frac{e^{2\pi i(N+\frac{1}{2})} - e^{-2\pi i(N+\frac{1}{2})x}}{e^{\pi ix} - e^{-\pi ix}} = \frac{\sin(2\pi(N+\frac{1}{2})x)}{\sin(\pi x)}. \tag{3.90}$$

□

Lecture 17 (Friday, October 11)

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

$$K_N(x) = \frac{1}{N+1}\sum_{n=0}^{N} D_n(x). \tag{3.91}$$

FACT 3.41. *We have*

$$K_N(x) = \frac{1}{2(N+1)}\frac{1-\cos(2\pi(N+1)x)}{\sin(\pi x)^2} = \frac{1}{N+1}\Big(\frac{\sin(\pi(N+1)x)}{\sin(\pi x)}\Big)^2 \tag{3.92}$$

PROOF. Using that $2\sin(x)\sin(y) = \cos(x-y) - \cos(x+y)$,

$$D_N(x) = \frac{\sin(2\pi(N+\frac{1}{2})x)}{\sin(\pi x)} = \frac{2\sin(\pi x)\sin(2\pi(N+\frac{1}{2})x)}{2\sin(\pi x)^2} = \frac{\cos(2\pi N x) - \cos(2\pi(N+1)x)}{2\sin(\pi x)^2}. \tag{3.93}$$

Thus,

$$\sum_{n=0}^{N} D_n(x) = \frac{1}{2\sin(\pi x)^2}\sum_{n=0}^{N}\cos(2\pi n x) - \cos(2\pi(n+1)x) = \frac{1-\cos(2\pi(N+1)x)}{2\sin(\pi x)^2} \tag{3.94}$$

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

$$\sigma_N f(x) = f * K_N(x). \tag{3.95}$$

THEOREM 3.42 (Fejér). *For every 1-periodic continuous function f,*

$$\sigma_N f \to f \tag{3.96}$$

uniformly on $\mathbb{R}$ as $N \to \infty$.

COROLLARY 3.43. *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:

$$\sigma_N f(x) = \int_0^1 f(t)\frac{1}{N+1}\sum_{n=0}^{N}\sum_{k=-n}^{n} e^{2\pi ik(x-t)}dt = \frac{1}{N+1}\sum_{n=0}^{N}\sum_{k=-n}^{n}\int_0^1 f(t)e^{-2\pi ikt}dt e^{2\pi ikx} \tag{3.97}$$

$$= \frac{1}{N+1}\sum_{n=0}^{N}\sum_{k=-n}^{n}\widehat{f}(k)e^{2\pi ikx} = \frac{1}{N+1}\sum_{k=-N}^{N}\sum_{n=|k|}^{N}\widehat{f}(k)e^{2\pi ikx} = \sum_{k=-N}^{N}(1-\tfrac{|k|}{N+1})\widehat{f}(k)e^{2\pi ikx}. \tag{3.98}$$

We will now derive Fejér's Theorem as a consequence of a more general principle.

DEFINITION 3.44 (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,

$$\sup_{x\in\mathbb{R}} |f * k_n(x) - f(x)| \to 0 \tag{3.99}$$

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 3.62). An approximation of unity is a sequence $(k_n)_n$ that *approximates unity*:

$$\lim_{n\to\infty} k_n * f = f \tag{3.100}$$

for every continuous, 1-periodic f.

THEOREM 3.45. *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*

$$\int_{-\delta}^{\delta} k_n(t)dt \to 1 \tag{3.101}$$

as $n \to \infty$.

Then $(k_n)_n$ is an approximation of unity.

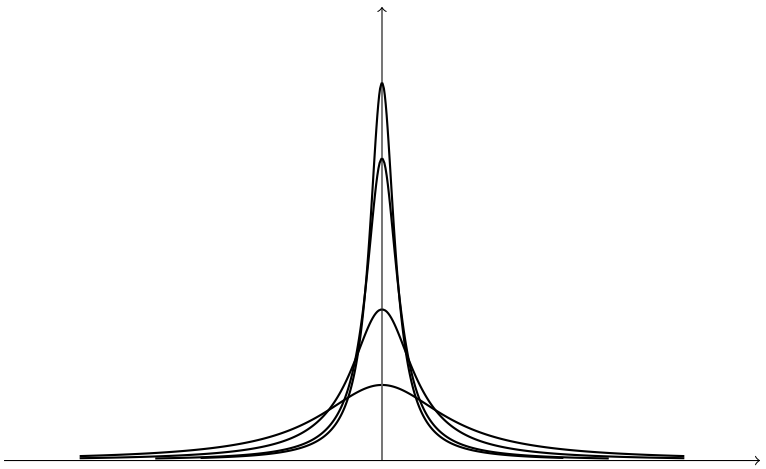

FIGURE 4. Approximation of unity

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:

$$\int_{\frac{1}{2}\geq|t|\geq\delta} k_n(t)dt \to 0 \tag{3.102}$$

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

$$|f(x-t) - f(x)| \leq \varepsilon/2 \tag{3.103}$$

for all $|t| < \delta$, $x \in \mathbb{R}$. Using Assumption (2),

$$f * k_n(x) - f(x) = \int_{-1/2}^{1/2} (f(x-t) - f(x))k_n(t)dt = A + B, \tag{3.104}$$

where

$$A = \int_{|t|\leq\delta} (f(x-t) - f(x))k_n(t)dt, \quad B = \int_{\frac{1}{2}\geq|t|\geq\delta} (f(x-t) - f(x))k_n(t)dt. \tag{3.105}$$

By 3.103 and Assumption (2),

$$|A| \leq \frac{\varepsilon}{2} \int_{|t|\leq\delta} k_n(t)dt \leq \frac{\varepsilon}{2}. \tag{3.106}$$

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 $n \geq N$,

$$\int_{\frac{1}{2}\geq|t|\geq\delta} k_n(t)dt \leq \frac{\varepsilon}{4C}. \tag{3.107}$$

Thus, if $n \geq N$,

$$|B| \leq 2C \int_{\frac{1}{2}\geq|t|\geq\delta} k_n(t)dt \leq \frac{\varepsilon}{2}. \tag{3.108}$$

This implies

$$|f * k_n(x) - f(x)| \leq \varepsilon/2 + \varepsilon/2 \leq \varepsilon \tag{3.109}$$

for $n \geq N$ and $x \in \mathbb{R}$. □

COROLLARY 3.46. *The Fejér kernel $(K_N)_N$ is an approximation of unity.*

PROOF. We verify the assumptions of Theorem 3.45. From (3.92) we see that $K_N \geq 0$. Also,

$$\int_{-1/2}^{1/2} K_N(t)dt = \frac{1}{N+1} \sum_{n=0}^{N} \sum_{k=-n}^{n} \int_{-1/2}^{1/2} e^{2\pi ikt}dt = \frac{1}{N+1} \sum_{n=0}^{N} 1 = 1. \tag{3.110}$$

Now we verify the last property. Let $\frac{1}{2} > \delta > 0$ and $|x| \geq \delta$. By (3.92),

$$K_N(x) \leq \frac{1}{N+1} \frac{1}{\sin(\pi\delta)^2} \tag{3.111}$$

Thus,

$$\int_{\frac{1}{2}\geq|t|\geq\delta} K_N(t)dt \leq \frac{1}{N+1} \frac{1}{\sin(\pi\delta)^2} \tag{3.112}$$

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 3.47. *Let f be a 1-periodic and continuous function. Then*

$$\lim_{N\to\infty} \|S_N f - f\|_2 = 0. \tag{3.113}$$

PROOF. Let $\varepsilon > 0$. By Fejér's theorem there exists a trigonometric polynomial p such that $|f(x) - p(x)| \le \varepsilon/2$ for all $x \in \mathbb{R}$. Then

$$\|f - p\|_2 = \Big(\int_0^1 |f(x) - p(x)|^2 dx \Big)^{1/2} \le \varepsilon/2. \tag{3.114}$$

Let N be the degree of p. Then $S_N p = p$ by Fact 3.34. Thus,

$$S_N f - f = S_N f - S_N p + S_N p - f = S_N(f - p) + p - f. \tag{3.115}$$

By Minkowski's inequality,

$$\|S_N f - f\|_2 \le \|S_N(f - p)\|_2 + \|p - f\|_2 \tag{3.116}$$

Bessel's inequality (Theorem 3.12) says that $\|S_N f\|_2 \le \|f\|_2$. Therefore,

$$\|S_N f - f\|_2 \le 2\|f - p\|_2 \le \varepsilon. \tag{3.117}$$

□

In view of Theorem 3.15 this means that the trigonometric system is complete.

Lecture 18 (Monday, October 14)

COROLLARY 3.48 (Parseval's theorem). *If f, g are 1-periodic, continuous functions, then*

$$\langle f, g\rangle = \sum_{n=-\infty}^{\infty} \widehat{f}(n)\overline{\widehat{g}(n)}. \tag{3.118}$$

In particular,

$$\|f\|_2^2 = \sum_{n=-\infty}^{\infty} |\widehat{f}(n)|^2. \tag{3.119}$$

PROOF. We have

$$\langle S_N f, g\rangle = \sum_{n=-N}^{N} \widehat{f}(n)\langle e^{2\pi inx}, g\rangle = \sum_{n=-N}^{N} \widehat{f}(n)\overline{\widehat{g}(n)}. \tag{3.120}$$

But $\langle S_N f, g\rangle \to \langle f, g\rangle$ as $N \to \infty$ because

$$|\langle S_N f, g\rangle - \langle f, g\rangle| = |\langle S_N f - f, g\rangle| \le \|S_N f - f\|_2 \|g\|_2 \to 0 \tag{3.121}$$

as $N \to \infty$. Here we have used the Cauchy-Schwarz inequality and the previous theorem. Equation (3.119) follows from putting $f = g$. □

Remark. Theorems 3.47 and Corollary 3.48 also hold for piecewise continuous and 1-periodic functions.

EXERCISE 3.49. (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

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}. \tag{3.122}$$

(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 3.50. *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,

$$S_N f(x) = \int_0^1 f(x - t) D_N(t) dt. \tag{3.123}$$

Also,

$$\int_0^1 D_N(t) dt = \sum_{n=-N}^{N} \int_0^1 e^{2\pi int} dt = 1. \tag{3.124}$$

Thus from Fact 3.40,

$$S_N f(x) - f(x) = \int_0^1 (f(x - t) - f(x)) D_N(t) dt \tag{3.125}$$

$$= \int_0^1 g(t) \sin(2\pi(N + \tfrac{1}{2})t) dt, \tag{3.126}$$

where

$$g(t) = \frac{f(x-t) - f(x)}{\sin(\pi t)}. \tag{3.127}$$

Differentiability of f at x implies that g is continuous at 0. Indeed,

$$\frac{f(x-t) - f(x)}{\sin(\pi t)} = \frac{f(x-t) - f(x)}{t} \frac{t}{\sin(\pi t)} \to f'(x) \frac{1}{\pi} \tag{3.128}$$

as $t \to 0$.

EXERCISE 3.51. 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 (3.126) and the Riemann-Lebesgue lemma (Corollary 3.13). □

EXERCISE 3.52. Show that there exists a constant $c > 0$ such that

$$\int_0^1 |D_N(x)| dx \geq c \log(2 + N) \tag{3.129}$$

holds for all $N = 0, 1, \ldots$.

EXERCISE 3.53. (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.

Lecture 19 (Wednesday, October 16)

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 3.54 (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 $f, g \in \mathcal{A}$, $c \in \mathbb{C}$,*

$$f + g \in \mathcal{A}, f \cdot g \in \mathcal{A}, c \cdot f \in \mathcal{A}, \overline{f} \in \mathcal{A}. \tag{3.130}$$

(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 3.55. 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 3.56. 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 3.57. *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 3.58. 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 3.59. *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 3.57 there exist $c_1, \ldots, c_n \in \mathbb{R}$ such that

$$|\sum_{i=1}^{n} c_i y^i - |y|| \leq \varepsilon. \tag{3.131}$$

for all $y \in [-a, a]$. By Condition (1) we have that

$$g = \sum_{i=1}^{n} c_i f^i \in \overline{\mathcal{A}}. \tag{3.132}$$

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 3.60. *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

$$\min(f, g) = \frac{f+g}{2} - \frac{|f-g|}{2}, \ \max(f, g) = \frac{f+g}{2} + \frac{|f-g|}{2}. \tag{3.133}$$

Thus, Condition (1) and Lemma 3.60 imply that $\min(f, g), \max(f, g) \in \overline{\mathcal{A}}$. □

LEMMA 3.61. *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

$$u_i(x) = g(x)h_i(x) - g(x_{1-i})h_i(x). \tag{3.134}$$

Then $u_i(x_{1-i}) = 0$ and $u_i(x_i) \neq 0$ for $i = 0, 1$. Set

$$f(x) = \frac{c_0 u_0(x)}{u_0(x_0)} + \frac{c_1 u_1(x)}{u_1(x_1)}. \tag{3.135}$$

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 3.81). 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.

Lecture 20 (Friday, October 18)

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 3.61 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,

$$h_y(t) > f(t) - \varepsilon. \tag{3.136}$$

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

$$g_x = \max(h_{y_1}, \ldots, h_{y_m}). \tag{3.137}$$

By Lemma 3.60, $g_x \in \overline{\mathcal{A}}$. $\square$

By continuity of g_x there exists an open ball U_x such that

$$|g_x(t) - f(t)| < \varepsilon \tag{3.138}$$

for $t \in U_x$. In particular,

$$g_x(t) < f(t) + \varepsilon. \tag{3.139}$$

$(U_x)_{x \in K}$ is an open cover of K which has a finite subcover by $U_{x_1}, \ldots, U_{x_n}$. Then let

$$h = \min(g_{x_1}, \ldots, g_{x_n}). \tag{3.140}$$

By Lemma 3.60 we have $h \in \overline{\mathcal{A}}$. Also,

$$f(t) - \varepsilon < h(t) < f(t) + \varepsilon \tag{3.141}$$

for all $t \in K$. That is,

$$|f(t) - h(t)| < \varepsilon \tag{3.142}$$

for all $t \in K$. This proves that $f \in \overline{\mathcal{A}}$.

6. Further exercises

Exercise 3.62. 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 3.63. Give an alternative proof of Weierstrass' theorem by using Fejér's theorem and then approximating the resulting trigonometric polynomials by truncated Taylor expansions.

Exercise 3.64. 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 3.65 (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

$$\langle f, g \rangle_{L^2(w)} = \int_a^b f(x)\overline{g(x)}w(x)dx \tag{3.143}$$

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 3.66. 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

$$\int_0^1 p_n(x)p_m(x)w(x)dx = \begin{cases} 1, & \text{if } n = m, \\ 0, & \text{if } n \neq m \end{cases} \tag{3.144}$$

for all non-negative integers n, m.

EXERCISE 3.67 (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

$$\int_{-1}^1 T_n(x)T_m(x)\frac{dx}{\sqrt{1-x^2}} \tag{3.145}$$

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 3.68. 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 2.12.

EXERCISE 3.69. 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 3.70 (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

$$d(a,b) = \lim_{n\to\infty} d(a_n, b_n), \tag{3.146}$$

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

$$L^2(a,b) = \overline{L^2_c(a,b)}. \tag{3.147}$$

(i) Define an inner product on $L^2(a,b)$ by

$$\langle f,g\rangle = \lim_{n\to\infty}\int_a^b f_n(x)\overline{g_n(x)}dx, \tag{3.148}$$

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

$$\sum_{n=1}^{\infty} |\langle f,\phi_n\rangle|^2 \leq \|f\|_2^2 \tag{3.149}$$

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

$$f_N = \sum_{n=1}^{N} c_n\phi_n \in L^2(a,b). \tag{3.150}$$

Show that $(f_N)_N$ converges in $L^2(a,b)$ if and only if

$$\sum_{n=1}^{\infty} |c_n|^2 < \infty. \tag{3.151}$$

EXERCISE 3.71. 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 3.72. 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 $\widehat{f'}(n) = 2\pi in\widehat{f}(n)$.
(iv) Let $y \in \mathbb{R}$ and set $f_y(x) = f(x+y)$. Show that $\widehat{f_y}(n) = e^{2\pi iny}\widehat{f}(n)$.
(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 3.73 (Legendre polynomials). Define $p_n(x) = \frac{d^n}{dx^n}[(1-x^2)^n]$ for $n = 0,1,\dots$ and

$$\phi_n(x) = p_n(x)\cdot\Big(\int_{-1}^{1} p_n(t)^2dt\Big)^{-1/2}. \tag{3.152}$$

Show that $(\phi_n)_{n=0,1,\dots}$ is a complete orthonormal system on $[-1,1]$.

EXERCISE 3.74. Let f be 1-periodic and k times continuously differentiable. Prove that there exists a constant $c > 0$ such that

$$|\widehat{f}(n)| \leq c|n|^{-k} \quad \text{for all } n \in \mathbb{Z}. \tag{3.153}$$

Hint: What can you say about the Fourier coefficients of $f^{(k)}$?

EXERCISE 3.75. 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

$$\sum_{n=1}^{\infty} \frac{\widehat{f}(n)}{n} < \infty. \tag{3.154}$$

(ii) Show that there does not exist a 1-periodic continuous function f such that

$$\widehat{f}(n) = \frac{\operatorname{sgn}(n)}{\log|n|} \quad \text{for all } |n| \geq 2. \tag{3.155}$$

Here $\operatorname{sgn}(n) = 1$ if $n > 0$ and $\operatorname{sgn}(n) = -1$ if $n < 0$.

EXERCISE 3.76. Suppose that f is a 1-periodic function such that there exists $c > 0$ and $\alpha \in (0, 1]$ such that

$$|f(x) - f(y)| \leq c|x - y|^{\alpha} \tag{3.156}$$

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 3.77. Let $f \in C([0, 1])$ and $\mathcal{A} \subset C([0, 1])$ dense. Suppose that

$$\int_0^1 f(x)\overline{a(x)}dx = 0 \tag{3.157}$$

for all $a \in \mathcal{A}$. Show that $f = 0$.
Hint: Show that $\int_0^1 |f(x)|^2 dx = 0$.

EXERCISE 3.78. 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 3.79. Prove that

$$-\frac{1}{2} = \sum_{n=1}^{\infty} (-1)^n \frac{\sin(n)}{n}. \tag{3.158}$$

EXERCISE 3.80. 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 3.81 (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 3.82 (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

$$\mu(x) = \sum_{i=1}^{m} \sigma((Wx)_i + b_i) v_i \text{ for all } x \in K. \tag{3.159}$$

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.

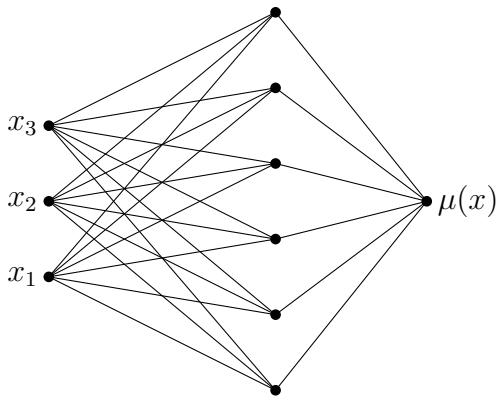

FIGURE 5. Visualization of μ when $n = 3$ and $m = 6$.

EXERCISE 3.83. 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

$$L_N(x) = \sum_{j=0}^{N} f(x_k) \prod_{j=0, j\neq k}^{N} \frac{x - x_j}{x_k - x_j}. \tag{3.160}$$

(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

$$f(x) - L_N(x) = \frac{f^{(N+1)}(\xi)}{(N + 1)!} \prod_{k=0}^{N} (x - x_k). \tag{3.161}$$

(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 4

Linear operators and derivatives

Lecture 21 (Monday, October 21)

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

DEFINITION 4.1. A map $\|\cdot\| : X \to [0,\infty)$ is called a *norm* if for all $x, y \in X$ and $\lambda \in \mathbb{K}$,

$$\|\lambda x\| = |\lambda| \cdot \|x\|, \quad \|x+y\| \le \|x\| + \|y\|, \quad \|x\| = 0 \Leftrightarrow x = 0. \tag{4.1}$$

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

$$d(x,y) = \|x-y\|. \tag{4.2}$$

A complete normed vector space is called *Banach space*.

EXAMPLES 4.2.
- $\mathbb{R}^n$ 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 4.3. The set of bounded sequences $(a_n)_{n\in\mathbb{N}}$ of complex numbers equipped with the *ℓ^∞-norm*,

$$\|a\|_\infty = \sup_{n=1,2,\dots} |a_n| \tag{4.3}$$

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

EXERCISE 4.4.
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

$$\|a\|_1 = \sum_{n=1}^\infty |a_n|. \tag{4.4}$$

Prove that this defines a Banach space.

EXERCISE 4.5. 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

$$\|a\|_2 = \big(\sum_{n=1}^\infty |a_n|^2\big)^{1/2}. \tag{4.5}$$

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

$$T(x + \lambda y) = Tx + \lambda Ty \tag{4.6}$$

for every $x, y \in X, \lambda \in \mathbb{K}$. 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 4.6. 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 4.7. *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) $\Rightarrow$ (ii): By assumption and linearity, for $x, y \in X$,

$$\|Tx - Ty\|_Y = \|T(x - y)\|_Y \leq C\|x - y\|_X. \tag{4.7}$$

This implies continuity.
(ii) $\Rightarrow$ (iii): There is nothing to prove.
(iii) $\Rightarrow$ (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

$$\|T(\delta x)\|_Y \leq 1 \tag{4.8}$$

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

$$\Big\| \frac{x}{\|x\|_X} \Big\|_X = 1. \tag{4.9}$$

Thus,

$$\Big\| T\Big(\frac{x}{\|x\|_X} \Big) \Big\|_Y \leq C. \tag{4.10}$$

By linearity of T this implies

$$\|Tx\|_Y \leq C\|x\|_X. \tag{4.11}$$

□

DEFINITION 4.8. 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

$$\|T\|_{\mathrm{op}} = \sup_{x \neq 0} \frac{\|Tx\|_Y}{\|x\|_X}. \tag{4.12}$$

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

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

$$\|Tx\|_Y \leq C\|x\|_X \tag{4.13}$$

holds. We have by definition that

$$\|Tx\|_Y \leq \|T\|_{\mathrm{op}}\|x\|_X. \tag{4.14}$$

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

$$\|T\|_{\mathrm{op}} = \sup_{\|x\|_X=1} \|Tx\|_Y = \sup_{\|x\|_X\leq 1} \|Tx\|_Y. \tag{4.15}$$

EXERCISE 4.9. 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 4.10. 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:

$$\|Ax\|_\infty = \max_{i=1,\dots,n} \Big|\sum_{j=1}^m A_{ij}x_j\Big| \leq \big(\max_{i=1,\dots,n} \sum_{j=1}^m |A_{ij}|\big)\|x\|_\infty. \tag{4.16}$$

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

$$\|A\|_{\infty\to\infty} \geq \|Ax\|_\infty = \sum_{j=1}^m |A_{ij}|. \tag{4.17}$$

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

$$\|A\|_{\infty\to\infty} = \max_{i=1,\dots,n} \sum_{j=1}^m |A_{ij}|. \tag{4.18}$$

EXERCISE 4.11. 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 $\|A\|_{1\to\infty} = \sup_{\|x\|_1=1} \|Ax\|_\infty$ and $\|A\|_{\infty\to 1} = \sup_{\|x\|_\infty=1} \|Ax\|_1$.

EXERCISE 4.12. 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.

Lecture 22 (Wednesday, October 23)

1. Equivalence of norms

DEFINITION 4.13. Two norms $\|\cdot\|_a$ and $\|\cdot\|_b$ on a vector space X are called *equivalent* if there exist constants $c, C > 0$ such that

$$c\|x\|_a \leq \|x\|_b \leq C\|x\|_a \tag{4.19}$$

for all $x \in X$.

EXERCISE 4.14. 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 4.15. 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 4.16. *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

$$\|x\| \leq \sum_{i=1}^n |x_i|\|b_i\| \leq (\max_{j=1,\ldots,n} |x_j|) \sum_{i=1}^n \|b_i\| = C\|x\|_*, \tag{4.20}$$

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

$$S = \{x \in X : \|x\|_* = 1\}. \tag{4.21}$$

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 2.19). Thus S is compact by Theorem 2.11.
Next note that the function $x \mapsto \|x\|$ is continuous with respect to the $\|\cdot\|_*$ norm. This is because by the triangle inequality and (4.20),

$$|\|x\| - \|y\|| \leq \|x - y\| \leq C\|x - y\|_*. \tag{4.22}$$

Thus by Theorem 2.12, $x \mapsto \|x\|$ attains its infimum on the compact set S and therefore there exists $c > 0$ such that

$$\|y\| \geq c \tag{4.23}$$

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 (4.23) with $y = \frac{x}{\|x\|_*}$ gives

$$\|x\| \geq c\|x\|_*. \tag{4.24}$$

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 3.64).

COROLLARY 4.17. *If X is finite-dimensional then every linear map $T : X \to Y$ is bounded.*

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

$$\|Tx\|_Y \leq \sum_{i=1}^n |c_i| \|Tx_i\|_Y \leq C \max_{i=1,\dots,n} |c_i|, \tag{4.25}$$

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 4.18. 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 $T : X \to Y$, $(Ta)_n = na_n$ 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

$$\|Te^{(k)}\| = k \tag{4.26}$$

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

EXERCISE 4.19. 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.

Optional topic (not relevant for exams)

2. Dual spaces*

THEOREM 4.20. *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 $(T_n)_n \subset L(X,Y)$ be a Cauchy sequence. Then for every $x \in X$, $(T_n x)_n \subset Y$ 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 $(T_n)_n$ 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 $x \in X$,

$$\|Tx\|_Y \leq \|Tx - T_n x\|_Y + \|T_n x\|_Y \leq \|Tx - T_n x\|_Y + M\|x\|_X. \tag{4.27}$$

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

$$\|T_n x - Tx\|_Y \leq \varepsilon \|x\|_X \tag{4.28}$$

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

$$\|T_n x - T_m x\|_Y \leq \tfrac{\varepsilon}{2}\|x\|_X \tag{4.29}$$

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

$$\|T_{m_x} x - Tx\|_Y \leq \tfrac{\varepsilon}{2}\|x\|_X. \tag{4.30}$$

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

$$\|T_n x - Tx\|_Y \leq \|T_n x - T_{m_x} x\|_Y + \|T_{m_x} x - Tx\|_Y \leq \varepsilon \|x\|_X. \tag{4.31}$$

□

DEFINITION 4.21. 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 4.22. *Dual spaces of normed vector spaces are Banach spaces.*

PROOF. This follows from Theorem 4.20 because $\mathbb{K}$ (which is $\mathbb{R}$ or $\mathbb{C}$) is complete. □

THEOREM 4.23. *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,

$$f(x) = \sum_{i=1}^n c_i f(x_i) = \sum_{i=1}^n f(x_i) f_i(x), \tag{4.32}$$

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

$$\sum_{i=1}^n b_i f_i = 0 \tag{4.33}$$

for some coefficients $(b_i)_{i=1,\dots,n} \subset \mathbb{K}$. Then for every $j \in \{1,\dots,n\}$, $b_j = \sum_{i=1}^n b_i f_i(x_j) = 0$. Thus, $\{f_1,\dots,f_n\}$ 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 $i = 1,\dots,n$. □

Optional topic (not relevant for exams)

3. Sequential ℓ^p spaces*

DEFINITION 4.24. 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

$$\|x\|_p = \Big(\sum_{n=1}^\infty |x_n|^p\Big)^{1/p}. \tag{4.34}$$

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.

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

$$\Big|\sum_{n=1}^\infty x_n y_n\Big| \le \|x\|_p \|y\|_{p'} \tag{4.35}$$

LEMMA 4.26 (Young's inequality). *For $a, b \ge 0$ and $p \in (1,\infty)$ we have the elementary inequality*

$$ab \le \frac{a^p}{p} + \frac{b^{p'}}{p'}. \tag{4.36}$$

PROOF. Recall that log is a concave function. Thus, for $u, v \ge 0$ and $t \in [0,1]$,

$$t\log(u) + (1-t)\log(v) \le \log(tu + (1-t)v). \tag{4.37}$$

The left hand side equals $\log(u^t v^{1-t})$. Now let $u = a^p$, $v = b^{p'}$, $t = \frac{1}{p}$. Then the claim follows from applying the exponential function on both sides of the inequality. □

PROOF OF HÖLDER'S INEQUALITY. If $p \in \{1,\infty\}$, the inequality is trivial. So we assume $p \in (1,\infty)$. By Young's inequality,

$$\sum_{n=1}^\infty |x_n y_n| \le \frac{1}{p}\sum_{n=1}^\infty |x_n|^p + \frac{1}{p'}\sum_{n=1}^\infty |y_n|^{p'}. \tag{4.38}$$

Let $\lambda > 0$. Replacing x_n by λx_n and y_n by $\lambda^{-1} y_n$ we obtain

$$\sum_{n=1}^\infty |x_n y_n| \le \frac{\lambda^p}{p}\sum_{n=1}^\infty |x_n|^p + \frac{\lambda^{-p'}}{p'}\sum_{n=1}^\infty |y_n|^{p'} = \lambda^p A + \lambda^{-p'} B, \tag{4.39}$$

where $A = \frac{1}{p}\|x\|_p^p$ and $B = \frac{1}{p'}\|y\|_{p'}^{p'}$. Without loss of generality we may assume that $A \ne 0$. We choose λ such that this inequality is strongest. This turns out to be when $\lambda = (\frac{p'B}{pA})^{\frac{1}{p+p'}}$. Plugging this into (4.39) implies the claim. □

THEOREM 4.27 (Minkowski's inequality). *Let $p \in [1,\infty]$. For $x, y \in \ell^p$,*

$$\|x + y\|_p \le \|x\|_p + \|y\|_p. \tag{4.40}$$

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

$$\|x+y\|_p^p \leq \sum_{n=1}^{\infty} |x_n||x_n+y_n|^{p-1} + \sum_{n=1}^{\infty} |y_n||x_n+y_n|^{p-1} \tag{4.41}$$

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

$$\leq \|x\|_p \|x+y\|_{p'(p-1)}^{p-1} + \|y\|_p \|x+y\|_{p'(p-1)}^{p-1} \tag{4.42}$$

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

$$\|x+y\|_p^p = (\|x\|_p + \|y\|_p)\|x+y\|_p^{p-1}. \tag{4.43}$$

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 4.28. *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

$$x_k^{(n)} = \begin{cases} \frac{|v(e_k)|^{p'}}{v(e_k)} & \text{if } k \leq n,\ v(e_k) \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4.44}$$

We have on the one hand

$$v(x^{(n)}) = \sum_{n=1}^{n} |v(e_k)|^{p'}. \tag{4.45}$$

And on the other hand

$$|v(x^{(n)})| \leq \|v\|_{\mathrm{op}} \|x^{(n)}\|_p = \|v\|_{\mathrm{op}} \Big(\sum_{k=1}^{n} |v(e_k)|^{p'} \Big)^{1/p}. \tag{4.46}$$

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

$$\Big(\sum_{k=1}^{n} |v(e_k)|^{p'} \Big)^{\frac{1}{p'}} \leq \|v\|_{\mathrm{op}}. \tag{4.47}$$

Letting $n \to \infty$ this implies that

$$\|\phi(v)\|_{p'} = \Big(\sum_{n=1}^{\infty} |v(e_n)|^{p'} \Big)^{1/p'} \leq \|v\|_{\mathrm{op}}, \tag{4.48}$$

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

$$\|\phi(v)\|_{p'} \leq \|v\|_{\mathrm{op}} \tag{4.49}$$

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 4.29. *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 4.3 and Exercise 4.4.

EXERCISE 4.30. Show that $\ell^p \subsetneq \ell^q$ if $1 \leq p < q \leq \infty$.

Lecture 23 (Friday, October 25)

4. 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

$$\lim_{h\to 0} \frac{|f(x+h)-f(x)-Th|}{|h|} = 0. \tag{4.50}$$

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

$$\mathbb{R} \longrightarrow L(\mathbb{R},\mathbb{R}), T \longmapsto (x \mapsto T \cdot x) \tag{4.51}$$

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 4.31. 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

$$\lim_{h\to 0} \frac{\|F(x+h)-F(x)-Th\|_Y}{\|h\|_X} = 0. \tag{4.52}$$

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 (4.52). Then

$$\|Th - \widetilde{T}h\|_Y \le \|F(x+h)-F(x)-Th\|_Y + \|F(x+h)-F(x)-\widetilde{T}h\|_Y \tag{4.53}$$

Thus, by (4.52),

$$\frac{\|Th-\widetilde{T}h\|_Y}{\|h\|_X} \longrightarrow 0 \quad \text{as } h \to 0 \tag{4.54}$$

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

$$\|Th - \widetilde{T}h\|_Y \le \varepsilon \|h\|_X \tag{4.55}$$

if $\|h\|_X \le \delta$. By homogeneity of norms we argue that the inequality (4.55) 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,

$$\|Th_0 - \widetilde{T}h_0\|_Y \le \varepsilon \|h_0\|_X = \varepsilon\delta. \tag{4.56}$$

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

$$\|Th - \widetilde{T}h\|_Y \le \varepsilon \|h\|_X \tag{4.57}$$

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

$$f(h) = O(g(h)) \quad \text{as } h \to 0 \tag{4.58}$$

to mean

$$\limsup_{h\to 0} \frac{\|f(h)\|}{\|g(h)\|} < \infty. \tag{4.59}$$

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

$$\|f(h)\| \leq C\|g(h)\| \tag{4.60}$$

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

- Little-o: Write

$$f(h) = o(g(h)) \quad \text{as } h \to 0 \tag{4.61}$$

to mean

$$\lim_{h\to 0} \frac{\|f(h)\|}{\|g(h)\|} = 0. \tag{4.62}$$

Comments.

- O and o are *not* functions and (4.58), (4.61) 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 (4.60), 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

$$F(x + h) = F(x) + Th + o(\|h\|) \quad (\text{as } h \to 0). \tag{4.63}$$

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.

Lecture 24 (Monday, October 28)

EXAMPLE 4.32. 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

$$F(x+h) = (x_1 + h_1)\cos(x_2 + h_2) = x_1 \cos(x_2 + h_2) + h_1 \cos(x_2 + h_2) \tag{4.64}$$

From Taylor's theorem we have that

$$\cos(t + \varepsilon) = \cos(t) - \sin(t)\varepsilon + O(\varepsilon^2) \quad \text{as } \varepsilon \to 0 \tag{4.65}$$

Thus,

$$F(x+h) = x_1 \cos(x_2) - x_1 \sin(x_2)h_2 + O(\|h\|^2) + h_1 \cos(x_2) - h_1 \sin(x_2)h_2 + O(\|h\|^2) \tag{4.66}$$

$$F(x+h) - F(x) = h_1 \cos(x_2) - x_1 \sin(x_2)h_2 + O(\|h\|^2) \tag{4.67}$$

This implies

$$F(x+h) = F(x) + Th + o(\|h\|), \tag{4.68}$$

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

$$DF|_x h = h_1 \cos(x_2) - x_1 \sin(x_2)h_2. \tag{4.69}$$

EXAMPLE 4.33. 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

$$F(f+h)(x) - F(f)(x) = \int_0^x (f(t)+h(t))^2 dt - \int_0^x f(t)^2 dt = 2\int_0^x f(t)h(t)dt + \int_0^x h(t)^2 dt \tag{4.70}$$

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

$$\|T(h)\|_\infty \leq 2\int_0^1 |f(t)h(t)|dt \leq C\|h\|_\infty, \tag{4.71}$$

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

$$F(f+h)(x) - F(f)(x) - T(h)(x) = \int_0^x h(t)^2 dt. \tag{4.72}$$

Thus

$$\|F(f+h) - F(f) - Th\|_\infty \leq \sup_{x\in[0,1]} \Big| \int_0^x h(t)^2 dt \Big| \tag{4.73}$$

$$\leq \int_0^1 |h(t)|^2 dt \leq \sup_{x\in[0,1]} |h(x)|^2 = \|h\|_\infty^2 \tag{4.74}$$

This implies

$$\frac{1}{\|h\|_\infty}\|F(f+h) - F(f) - Th\|_\infty \leq \|h\|_\infty \to 0 \tag{4.75}$$

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 4.34 (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*

$$D(f \circ g)|_x h = Df|_{g(x)} Dg|_x h \tag{4.76}$$

for all $h \in X_1$.

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

$$f(g(x + h)) - f(g(x)) - Df|_{g(x)} Dg|_x h \tag{4.77}$$

$$= f(g(x) + k) - f(g(x)) - Df|_{g(x)} k + Df|_{g(x)}(g(x + h) - g(x) - Dg|_x h), \tag{4.78}$$

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

$$\|f(g(x + h)) - f(g(x)) - Df|_{g(x)} Dg|_x h\|_{X_3} \tag{4.79}$$

$$\leq \|f(g(x)+k)-f(g(x))-Df|_{g(x)}k\|_{X_3}+\|Df|_{g(x)}\|_{\mathrm{op}}\|g(x+h)-g(x)-Dg|_x h\|_{X_2} \tag{4.80}$$

We have

$$\|k\|_{X_2} = \|g(x + h) - g(x)\|_{X_2} \leq \|Dg|_x\|_{\mathrm{op}}\|h\|_{X_1} + o(\|h\|_{X_1}). \tag{4.81}$$

Dividing by $\|h\|_{X_1}$ on both sides, (4.80) implies

$$\frac{1}{\|h\|_{X_1}}\|f(g(x+h))-f(g(x))-Df|_{g(x)}Dg|_x h\|_{X_3} \leq \frac{\|k\|_{X_2}}{\|h\|_{X_1}}\frac{\|f(g(x)+k)-f(g(x))-Df|_{g(x)}k\|_{X_3}}{\|k\|_{X_2}}+o(1) \tag{4.82}$$

(as $h \to 0$). By (4.81),

$$\frac{\|k\|_{X_2}}{\|h\|_{X_1}} \leq \|Dg|_x\|_{\mathrm{op}} + 1 \tag{4.83}$$

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

$$\frac{\|f(g(x) + k) - f(g(x)) - Df|_{g(x)} k\|_{X_3}}{\|k\|_{X_2}} \tag{4.84}$$

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

THEOREM 4.35 (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*

$$D(F \cdot G)\,|_x = F(x) \cdot DG\,|_x + G(x) \cdot DF\,|_x\,. \tag{4.85}$$

EXERCISE 4.36. Prove this.

Lecture 25 (Friday, November 1)

DEFINITION 4.37. 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

$$\lim_{h\to 0} \frac{F(x+hv) - F(x)}{h} \in Y \qquad (h \in \mathbb{K} \setminus \{0\}) \tag{4.86}$$

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 4.38. *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*

$$D_v F|_x = DF|_x v. \tag{4.87}$$

PROOF. By definition

$$F(x+hv) - F(x) - DF|_x(hv) = o(h) \quad \text{as } h \to 0. \tag{4.88}$$

Therefore,

$$\frac{F(x+hv) - F(x)}{h} = DF|_x v + o(1) \quad \text{as } h \to 0. \tag{4.89}$$

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

EXAMPLE 4.39. 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

$$D_{e_1} F|_x = 2x_1, \quad D_{e_2} F|_x = 2x_2. \tag{4.90}$$

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

$$DF|_x v = v_1 DF|_x e_1 + v_2 DF|_x e_2 \tag{4.91}$$

By Theorem 4.38 this equals

$$v_1 D_{e_1} F|_x + v_2 D_{e_2} F|_x = 2x_1 v_1 + 2x_2 v_2. \tag{4.92}$$

Remark. The converse of Theorem 4.38 is not true!

EXAMPLE 4.40. 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 $v = (v_1, v_2)$,

$$F(hv) - F(0) = h \frac{v_1^3}{v_1^2 + v_2^2}, \tag{4.93}$$

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,

$$D_v F|_0 = DF|_0 v = v_1 D_{e_1} F|_0 + v_2 D_{e_2} F|_0 = v_1, \tag{4.94}$$

which is false.

5. Further exercises

EXERCISE 4.41. Let $x \in \mathbb{R}^n$. Define $\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$ 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 4.42. 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

$$d(f,g) = \sum_{k=0}^{\infty} 2^{-k} w(\sup_{x\in[-k,k]} |f(x) - g(x)|). \tag{4.95}$$

(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 4.43. 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 4.44. 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

$$Tf(x) = \int_0^x f(t)dt. \tag{4.96}$$

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

EXERCISE 4.45. 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 4.46. 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

$$F(f)(s) = \int_0^s \cos(f(t)^2)dt, \;\; s \in [0,1]. \tag{4.97}$$

(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 4.47. 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

$$F : \mathbb{R}^{n\times n} \longrightarrow \mathbb{R}^{n\times n},\ A \longmapsto A^2. \tag{4.98}$$

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

CHAPTER 5

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

Lecture 26 (Monday, November 4)

In this section we study the differential calculus of maps $f : U \to \mathbb{R}^m$, $U \subset \mathbb{R}^n$ open. 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 5.1. 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 (4.87):

$$Df|_x h = \sum_{j=1}^{n} h_j Df|_x e_j = \sum_{j=1}^{n} h_j \partial_j f(x) \in \mathbb{R}^m. \tag{5.1}$$

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 (4.87) 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

$$Df|_x = (\partial_j f_i(x))_{i,j} = \begin{pmatrix} \partial_1 f_1(x) & \cdots & \partial_n f_1(x) \\ \vdots & \ddots & \vdots \\ \partial_1 f_m(x) & \cdots & \partial_n f_m(x) \end{pmatrix} \in \mathbb{R}^{m \times n}, \tag{5.2}$$

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]

$$\nabla f(x) = Df|_x^T = \begin{pmatrix} \partial_1 f(x) \\ \vdots \\ \partial_n f(x) \end{pmatrix} \in \mathbb{R}^n. \tag{5.3}$$

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

EXAMPLE 5.2. 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

$$DF|_x = \begin{pmatrix} x_2 \sin(x_3) & x_1 \sin(x_3) & x_1 x_2 \cos(x_3) \\ -e^{x_1} & 2x_2 & 0 \end{pmatrix}. \tag{5.4}$$

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 5.3 (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*

$$f(x) - f(y) = Df|_\xi(x - y) \tag{5.5}$$

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,

$$g'(t_0) = Df|_{t_0x+(1-t_0)y}(x - y). \tag{5.6}$$

□

COROLLARY 5.4. *Under the assumptions of the previous theorem: if $Df|_x = 0$ for all $x \in U$, then f is constant.*

EXERCISE 5.5. 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 5.6. 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 5.7. *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 5.8. 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 4.38 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

$$f(x+h) - f(x) = f(x+v_n) - f(x+v_{n-1}) + f(x+v_{n-1}) - f(x+v_{n-2}) + \cdots + f(x+v_1) - f(x+v_0) \tag{5.7}$$

$$= \sum_{j=1}^{n} (f(x+v_j) - f(x+v_{j-1})). \tag{5.8}$$

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

$$f(x+v_j) - f(x+v_{j-1}) = f(x+v_{j-1}+h_j e_j) - f(x+v_{j-1}) = \partial_j f(x+v_{j-1}+t_j h_j e_j) h_j. \tag{5.9}$$

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

$$|\partial_j f(y) - \partial_j f(x)| \le \varepsilon/n \quad \text{for all } j = 1, \dots, n, \tag{5.10}$$

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

$$\left| f(x+h) - f(x) - \sum_{j=1}^{n} h_j \partial_j f(x) \right| \le \sum_{j=1}^{n} \left| f(x+v_j) - f(x+v_{j-1}) - h_j \partial_j f(x) \right| \tag{5.11}$$

$$= \sum_{j=1}^{n} |h_j| |\partial_j f(x + v_{j-1} + t_j h_j e_j) - \partial_j f(x)| \le \sum_{j=1}^{n} |h_j| \frac{\varepsilon}{n} \le \varepsilon \|h\|. \tag{5.12}$$

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

$$Df|_x h = \sum_{j=1}^{n} h_j \partial_j f(x), \tag{5.13}$$

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

$$f(x) - f(y) = \int_0^1 Df|_{tx+(1-t)y}(x-y)\,dt. \tag{5.14}$$

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

$$f(x) - f(y) = g(1) - g(0) = \int_0^1 g'(s)\,ds = \int_0^1 Df|_{tx+(1-t)y}(x-y)\,dt. \tag{5.15}$$

□

THEOREM 5.10 (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*

$$\|F(x) - F(y)\| \le \|DF|_\xi\|_{\mathrm{op}} \|x - y\|, \tag{5.16}$$

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

PROOF. Write $F = (F_1, \dots, F_m)$. Then by Theorem 5.9

$$F_i(x) - F_i(y) = \int_0^1 DF_i|_{tx+(1-t)y}(x-y)\,dt. \tag{5.17}$$

This implies

$$F(x) - F(y) = \int_0^1 DF|_{tx+(1-t)y}(x-y)dt. \tag{5.18}$$

By the triangle inequality, we have

$$\|F(x) - F(y)\| \leq \int_0^1 \|DF|_{tx+(1-t)y}\|_{\mathrm{op}} dt \|x-y\| \tag{5.19}$$

The map $[0,1] \to \mathbb{R}$, $t \mapsto \|DF|_{tx+(1-t)y}\|_{\mathrm{op}}$ 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

$$\|F(x) - F(y)\| \leq \|DF|_\xi\|_{\mathrm{op}} \|x-y\|. \tag{5.20}$$

□

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

$$F(x) - F(y) = DF|_\xi(x-y). \tag{5.21}$$

Exercise 5.11. 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 $F(x) - F(y) = DF|_\xi(x-y)$.

Lecture 27 (Wednesday, November 6)

1. The contraction principle

The contraction principle is a powerful tool in analysis. It is most naturally described in the setting of metric spaces.

DEFINITION 5.12. Let (X, d) be a metric space. A map $\varphi : X \to X$ is called a *contraction* (of X) if there exists a constant $c \in (0, 1)$ such that

$$d(\varphi(x), \varphi(y)) \leq c \cdot d(x, y) \tag{5.22}$$

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

Remark. Contractions are continuous.

EXAMPLE 5.13. 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. Indeed, by the mean value theorem (Theorem 5.10), for every $x, y \in U$ there exists $\xi \in U$ such that

$$\|F(x) - F(y)\| \leq \|DF|_\xi\|_{\text{op}} \|x - y\| \leq c\|x - y\|. \tag{5.23}$$

THEOREM 5.14 (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

$$0 \leq d(x_0, x_1) = d(\varphi(x_0), \varphi(x_1)) \leq c \cdot d(x_0, x_1), \tag{5.24}$$

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

$$x_{n+1} = \varphi(x_n). \tag{5.25}$$

We claim that $(x_n)_n$ is a Cauchy sequence. Indeed, by induction we see that

$$d(x_{n+1}, x_n) \leq cd(x_n, x_{n-1}) \leq c^2 d(x_{n-1}, x_{n-2}) \leq \cdots \leq c^n d(x_1, x_0). \tag{5.26}$$

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

$$d(x_m, x_n) \leq d(x_m, x_{m-1}) + d(x_{m-1}, x_{m-2}) + \cdots + d(x_{n+1}, x_n) \tag{5.27}$$

$$= \sum_{i=n}^{m-1} d(x_{i+1}, x_i) \leq \sum_{i=n}^{m-1} c^i d(x_1, x_0) \leq d(x_1, x_0) \sum_{i=n}^{\infty} c^i = c^n \frac{d(x_1, x_0)}{1 - c}. \tag{5.28}$$

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

$$\varphi(x_*) = \varphi(\lim_{n \to \infty} x_n) = \lim_{n \to \infty} \varphi(x_n) = \lim_{n \to \infty} x_{n+1} = x_*. \tag{5.29}$$

□

Remarks. 1. The theorem is false 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, but has no fixed point.
2. 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 $(x_n)_n$ defined in the proof satisfies the inequality

$$d(x_n, x_*) \leq \frac{c^n}{1-c} d(x_0, x_1), \tag{5.30}$$

so speed of convergence depends only on the parameter $c \in (0,1)$ and the quality of the initial guess $x_0 \in X$.
3. 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.

EXAMPLE 5.15. 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

$$Ax = b \tag{5.31}$$

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

$$F(x) = \lambda(Ax - b) + x \tag{5.32}$$

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,

$$\|F(x) - F(y)\| = \|\lambda A(x-y) + x - y\| = \|(\lambda A + I)(x-y)\| \leq \|\lambda A + I\|_{\mathrm{op}} \|x - y\|. \tag{5.33}$$

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. Inverse function theorem and implicit 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 5.16. Let $E \subset \mathbb{R}^n$ be open. 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 5.17 (Inverse function theorem). *Let $E \subset \mathbb{R}^n$ be open, $f \in C^1(E, \mathbb{R}^n)$ and $a \in E$. Assume that $Df|_a$ is invertible. Then f is locally invertible at a in some open neighborhood $U \subset E$ of a and $f|_U^{-1} \in C^1(f(U), U)$ with*

$$D(f|_U^{-1})|_{f(a)} = (Df|_a)^{-1}. \tag{5.34}$$

Lecture 28 (Friday, November 8)

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

$$\varphi_y(x) = x + Df|_a^{-1}(y - f(x)) \quad (x \in E) \tag{5.35}$$

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

$$D\varphi_y|_x = I - Df|_a^{-1}Df|_x = Df|_a^{-1}(Df|_a - Df|_x). \tag{5.36}$$

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

$$\|Df|_a - Df|_x\|_{\text{op}} \leq \frac{1}{2\lambda} \quad \text{for } x \in U. \tag{5.37}$$

Then for $x, x' \in U$

$$\|\varphi_y(x) - \varphi_y(x')\| \leq \|D\varphi_y|_\xi\|_{\text{op}}\|x - x'\| \tag{5.38}$$

$$\leq \|Df|_a^{-1}\|_{\text{op}}\|Df|_a - Df|_\xi\|_{\text{op}}\|x - x'\| \leq \frac{1}{2}\|x - x'\|. \tag{5.39}$$

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 $B_r(x_0) \subset U$ (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

$$\|\varphi_y(x_0) - x_0\| = \|Df|_a^{-1}(y - y_0)\| \leq \lambda\varepsilon. \tag{5.40}$$

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

$$\|\varphi_y(x) - x_0\| \leq \|\varphi_y(x) - \varphi_y(x_0)\| + \|\varphi_y(x_0) - x_0\| \tag{5.41}$$

$$\leq \tfrac{1}{2}\|x - x_0\| + \tfrac{r}{2} \leq \tfrac{r}{2} + \tfrac{r}{2} = r. \tag{5.42}$$

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

$$\|B - A\| \cdot \|A^{-1}\| < 1. \tag{5.43}$$

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 (5.43) 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,

$$\|x\| = \|A^{-1}Ax\| \leq \|A^{-1}\| \cdot \|Ax\| \leq \|A^{-1}\|(\|(A-B)x\| + \|Bx\|) \tag{5.44}$$

$$\leq \|A^{-1}\| \cdot \|B - A\| \cdot \|x\| + \|A^{-1}\|\|Bx\|, \tag{5.45}$$

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

Lecture 29 (Monday, November 11)

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,

$$\|Df|_a^{-1}\|\|Df|_x - Df|_a\| \leq \tfrac{1}{2}. \tag{5.46}$$

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

$$\|h\| \leq \|h - Df|_a^{-1}k\| + \|Df|_a^{-1}k\| \quad \text{and} \tag{5.47}$$

$$h - Df|_a^{-1}k = h + Df|_a^{-1}(f(x) - f(x + h)) = \varphi_y(x + h) - \varphi_y(x), \tag{5.48}$$

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

$$g(y + k) - g(y) - Df|_x^{-1}k = x + h - x - Df|_x^{-1}k \tag{5.49}$$

$$= h - Df|_x^{-1}(f(x + h) - f(x)) = -Df|_x^{-1}(f(x + h) - f(x) - Df|_x h) \quad \text{and so} \tag{5.50}$$

$$\frac{1}{\|k\|}\|g(y + k) - g(y) - Df|_x^{-1}k\| \leq \|Df|_x^{-1}\|\frac{\|f(x + h) - f(x) - Df|_x h\|}{\|h\|}\frac{\|h\|}{\|k\|} \tag{5.51}$$

$$\leq \|Df|_x^{-1}\|\frac{\|f(x + h) - f(x) - Df|_x h\|}{\|h\|}2\lambda \longrightarrow 0 \quad \text{as } k \to 0. \tag{5.52}$$

Therefore g is differentiable at y with $Dg|_y = Df|_x^{-1}$.

It remains to show that Dg is continuous. To show this we need another lemma.

LEMMA 5.19. *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 $g \in C^1(V, U)$. □

EXERCISE 5.20. Let $f \in C^1(E, \mathbb{R}^n)$ and assume that $Df|_x$ is invertible for all $x \in E$. Prove that $f(U)$ is open for every open set $U \subset E$.

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

EXAMPLE 5.21. 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

$$Df|_x = \begin{pmatrix} e^{x_2}\cos(x_1) & e^{x_2}\sin(x_1) \\ -e^{x_2}\sin(x_1) & e^{x_2}\cos(x_1) \end{pmatrix}. \tag{5.53}$$

Thus $\det Df|_x = e^{2x_2}(\cos(x_1)^2 + \sin(x_1)^2) = e^{2x_2} \neq 0$, so by Theorem 5.17, f is locally invertible at every point $x \in \mathbb{R}^2$. f is not bijective: it is not injective because, for instance, $f(0, 0) = f(2\pi, 0)$.

Lecture 30 (Wednesday, November 13)

We will now use the inverse function theorem to prove a significant generalization concerning equations of the form $f(x,y)=0$, where y is given and we want to solve for x. Let $E \subset \mathbb{R}^n \times \mathbb{R}^m$ open, $f: E \to \mathbb{R}^n$ differentiable at $p=(a,b) \in \mathbb{R}^n \times \mathbb{R}^m$. Then $Df|_p$ is a $n \times (n+m)$ matrix:

$$Df|_p = \left(\begin{array}{ccc|ccc} \partial_1 f_1|_p & \cdots & \partial_n f_1|_p & \partial_{n+1} f_1|_p & \cdots & \partial_{n+m} f_1|_p \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \partial_1 f_n|_p & \cdots & \partial_n f_n|_p & \partial_{n+1} f_n|_p & \cdots & \partial_{n+m} f_n|_p \end{array}\right) \in \mathbb{R}^{n\times(n+m)} \tag{5.54}$$

We denote the left $n \times n$ submatrix by $D_x f|_p$ and the right $n \times m$ submatrix by $D_y f|_p$. Note that $D_x f|_{(a,b)}$ is the Jacobian matrix of the differentiable map $x \mapsto f(x,b)$ at $x=a$ (b is fixed).

THEOREM 5.22 (Implicit function theorem). *Let $f \in C^1(E,\mathbb{R}^n)$, $(a,b) \in E \subset \mathbb{R}^n \times \mathbb{R}^m$ with $f(a,b)=0$. Assume that $D_x f|_{(a,b)} \in \mathbb{R}^{n\times n}$ is invertible. Then there exist open sets $U \subset E$ and $W \subset \mathbb{R}^m$ with $(a,b) \in U$, $b \in W$ such that for every $y \in W$ there exists a unique x such that $(x,y) \in U$ and $f(x,y)=0$. Write $x=g(y)$. Then W can be chosen such that $g \in C^1(W,\mathbb{R}^n)$, $g(b)=a$, $(g(y),y) \in U$ and $f(g(y),y)=0$ for $y \in W$. Moreover,*

$$Dg|_b = -D_x f|_{(a,b)}^{-1} D_y f|_{(a,b)}. \tag{5.55}$$

(Note that this equation makes sense, because $Dg|_b \in \mathbb{R}^{n\times m}$, $D_x f|_{(a,b)}^{-1} \in \mathbb{R}^{n\times n}$, $D_y f|_{(a,b)} \in \mathbb{R}^{n\times m}$.)

Remark. The equation (5.55) can be obtained from differentiating the equation

$$f(g(y),y)=0 \tag{5.56}$$

with respect to y using the chain rule (this is called *implicit differentiation*).

PROOF. Define $F(x,y)=(f(x,y),y)$ for $(x,y) \in E$. Then $F \in C^1(E,\mathbb{R}^n \times \mathbb{R}^m)$. We would like to apply the inverse function theorem to F. For $h \in \mathbb{R}^n, k \in \mathbb{R}^m$ with $(a+h,b+k) \in E$,

(5.57)

$$F(a+h,b+k)-F(a,b) = (f(a+h,b+k)-f(a,b),k) = (Df|_{(a,b)}(h,k)+o(\|(h,k)\|),k)$$

$$= (Df|_{a,b}(h,k),k)+o(\|(h,k)\|) \tag{5.58}$$

and thus $DF|_{(a,b)}(h,k)=(Df|_{(a,b)}(h,k),k)$. We claim that $DF|_{(a,b)} \in \mathbb{R}^{(n+m)\times(n+m)}$ is invertible. It suffices to show that $DF|_{(a,b)}$ is injective. Let $DF|_{(a,b)}(h,k)=0$. Then $k=0$ and $Df|_{(a,b)}(h,0)=0$. Thus, $D_x f|_{(a,b)} h=0$, so $h=0$ because $D_x f|_{(a,b)}$ is injective.

By the inverse function theorem there exist open sets $U,V \subset \mathbb{R}^{n+m}$ with $(a,b) \in U$, $(0,b) \in V$ such that $F|_U : U \to V$ is bijective. Let $W=\{y \in \mathbb{R}^m : (0,y) \in V\}$. W is open because V is open. Then, if $y \in W$, there exists a unique $(x,y) \in U$ such that $F(x,y)=(0,y)$, so $f(x,y)=0$. Define $g(y)=x$. We need to show that $g \in C^1(W,\mathbb{R}^n)$. We have $F(g(y),y)=(0,y)$. Let $G=F|_U^{-1}: V \to U$. Then $G(0,y)=(g(y),y)$, so g is C^1 because G is C^1. Set $\phi(y)=(g(y),y)$. Then $D\phi|_y=(Dg|_y,I)$. Also, $f(\phi(y))=0$. By the chain rule,

$$0 = Df|_{\phi(y)} D\phi|_y = D_x f|_{\phi(y)} Dg|_y + D_y f|_{\phi(y)}, \tag{5.59}$$

so $Dg|_y = -D_x f|^{-1}_{\phi(y)} D_y f|_{\phi(y)}$. Setting $y = b$ we obtain

$$Dg|_b = -D_x f|^{-1}_{(a,b)} D_y f|_{(a,b)}. \tag{5.60}$$

□

EXAMPLE 5.23. While we used the inverse function theorem in the proof of the implicit function theorem, the inverse function theorem is also a consequence of the implicit function theorem. Say $E \subset \mathbb{R}^n$, $f \in C^1(E, \mathbb{R}^n)$ and $a \in E$ such that $Df|_a$ is invertible.

Define $F : E \times \mathbb{R}^n \to \mathbb{R}^n$ by $F(x, y) = f(x) - y$ and set $b = f(a)$. Then $F(a, b) = 0$. Also, $D_x F|_{(a,b)} = Df|_a$ is invertible, so by the implicit function theorem there exist open sets $\Omega \subset E \times \mathbb{R}^n$, $W \subset \mathbb{R}^n$, $f(a) \in W$ and $g \in C^1(W, \mathbb{R}^n)$ with $g(b) = a$, $(g(y), y) \in \Omega$ for $y \in W$ and

$$F(g(y), y) = 0 \quad \text{and} \quad Dg|_{f(a)} = Df|^{-1}_a. \tag{5.61}$$

But $F(g(y), y) = 0$ is equivalent to $f(g(y)) = y$. Define $U = g(W) = \{g(y) : y \in W\}$. Then $a \in U$ since $g(f(a)) = g(b) = a$ and $b \in W$.

Also $U \subset E$ since if $x \in U$, then $x = g(y)$ for $y \in W$, so $(x, y) \in \Omega \subset E \times \mathbb{R}^n$. Similarly, we see that U is open because Ω is open. Let $x \in U$. Then there exists $y \in W$ such that $x = g(y)$. Also, $f(x) = f(g(y)) = y$ and therefore $g(f(x)) = g(y) = x$. Thus $f|_U$ is invertible and $f|^{-1}_U = g$.

EXAMPLE 5.24. Let $f : \mathbb{R}^2 \times \mathbb{R}^3 \to \mathbb{R}^2$ be given by

$$f(x, y) = \begin{pmatrix} x_1^2 y_1 + x_2 \cos(y_2) - y_3 \\ e^{-x_2} + \sin(y_1) + x_1 y_2 y_3 - 1 - \sin(1) \end{pmatrix} \quad (x \in \mathbb{R}^2, y \in \mathbb{R}^3). \tag{5.62}$$

Then

$$Df(x, y) = \begin{pmatrix} 2x_1 y_1 & \cos(y_2) & x_1^2 & -x_2 \sin(y_2) & -1 \\ y_2 y_3 & -e^{-x_2} & \cos(y_1) & x_1 y_3 & x_1 y_2 \end{pmatrix} \tag{5.63}$$

Set $a = (1, 0)$, $b = (1, 0, 1)$. Then $f(a, b) = 0$ and

$$D_x f|_{(a,b)} = \begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix}, \tag{5.64}$$

then $\det D_x f|_{(a,b)} = -2 \neq 0$, so $D_x f|_{(a,b)}$ is invertible. We have

$$D_y f|_{(a,b)} = \begin{pmatrix} 1 & 0 & -1 \\ \cos(1) & 1 & 0 \end{pmatrix}. \tag{5.65}$$

Thus there exists $U \subset \mathbb{R}^2 \times \mathbb{R}^3$ open, $(a, b) \in U$, $W \subset \mathbb{R}^3$, $b \in W$, $g \in C^1(W, \mathbb{R}^2)$ such that

$$f(g(y), y) = 0 \quad \text{for } y \in W \text{ and} \tag{5.66}$$

$$Dg|_b = -D_x f|^{-1}_{(a,b)} D_y f|_{(a,b)} = \tfrac{1}{2} \begin{pmatrix} -1 - \cos(1) & -1 & 1 \\ 2\cos(1) & 2 & 0 \end{pmatrix}. \tag{5.67}$$

3. Ordinary differential equations

In this section we study *initial value problems* of the form

$$\begin{cases} y'(t) = F(t, y(t)) \\ y(t_0) = y_0, \end{cases} \tag{5.68}$$

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 $t \in E$. 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

$$y' = F(t, y). \tag{5.69}$$

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.

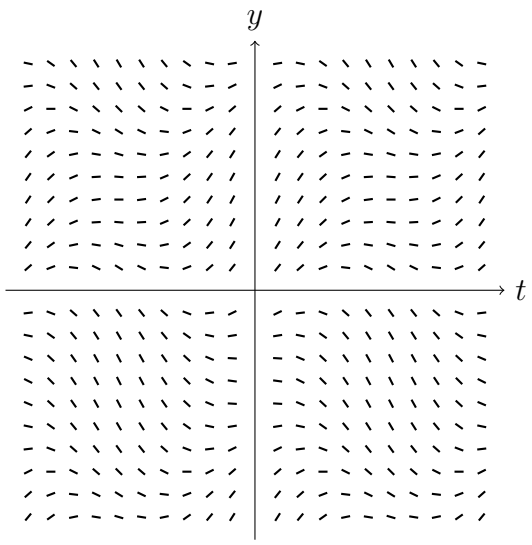

FIGURE 1. Visualization of $F(t, y)$.

EXAMPLE 5.25. Consider the equation $y' = \frac{y}{t}$. The solutions of this equation are of the form $y(t) = ct$ for $c \in \mathbb{R}$.

EXAMPLE 5.26. 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

$$\begin{cases} y'(t) = \frac{t}{y(t)} \\ y(t_0) = y_0 \end{cases} \tag{5.70}$$

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

$$\begin{cases} y'(t) = e^{y(t)^2 t^2} \sin(t + y(t)), \\ y(1) = 5. \end{cases} \tag{5.71}$$

Lecture 31 (Friday, November 15)

THEOREM 5.27 (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*

$$R = \{(t, y) \in \mathbb{R}^2 : |t - t_0| \leq a, |y - y_0| \leq b\} \subset E. \tag{5.72}$$

Let $M = \sup_{(t,y)\in R} |F(t, y)| < \infty$. Assume that there exists $c \in (0, \infty)$ such that

$$|F(t, y) - F(t, u)| \leq c|y - u| \tag{5.73}$$

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

$$\begin{cases} y'(t) = F(t, y(t)), \\ y(t_0) = y_0. \end{cases} \tag{5.74}$$

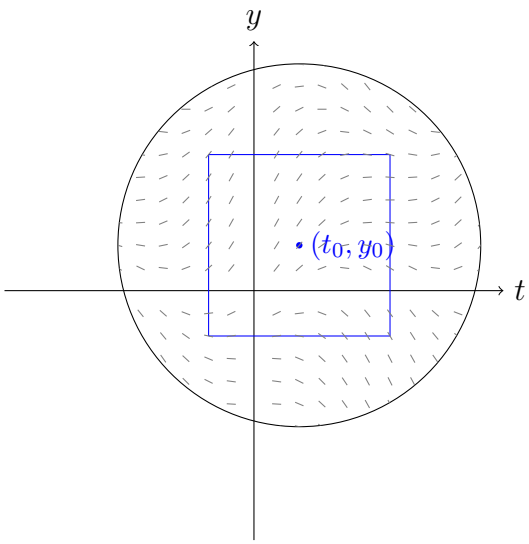

FIGURE 2. Visualization of $F(t, y)$.

Remarks. 1. If F satisfies condition (5.73), we also say that F is *Lipschitz continuous in the second variable.*
2. The condition (5.73) follows if F is differentiable in the second variable and $|\partial_y F(t, y)| \leq c$ for every $(t, y) \in R$ (by the mean value theorem).
3. By the fundamental theorem of calculus, the initial value problem (5.74) is equivalent to the integral equation

$$y(t) = y_0 + \int_{t_0}^{t} F(s, y(s))ds. \tag{5.75}$$

COROLLARY 5.28. *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* (5.74).

This is true because (5.73) follows from the mean value theorem and continuity of the second derivative $\partial_y F$.

PROOF OF THEOREM 5.27. 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

$$y(t) = y_0 + \int_{t_0}^{t} F(s, y(s))ds \tag{5.76}$$

(that is, y is a solution of the integral equation). Let

$$\mathcal{Y} = \{y : I \to J \,:\, y \text{ continuous on } I\}. \tag{5.77}$$

For every $y \in \mathcal{Y}$, $t \mapsto F(t, y(t))$ is a well-defined continuous function on I. Define

$$Ty(t) = y_0 + \int_{t_0}^{t} F(s, y(s))ds. \tag{5.78}$$

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 $|F(t, y)| \le M$ for all $(t, y) \in R$ we obtain:

$$|Ty(t) - y_0| \le \int_{t_0}^{t} |F(s, y(s))|ds \le |t_0 - t|M \le Ma_* \le b, \tag{5.79}$$

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

$$d_*(g_1, g_2) = \sup_{t\in I} e^{-2c|t-t_0|}|g_1(t) - g_2(t)|. \tag{5.80}$$

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 (5.73),

$$|Tg_1(t) - Tg_2(t)| = \Big| \int_{t_0}^{t} (F(s, g_1(s)) - F(s, g_2(s)))ds \Big| \tag{5.81}$$

$$\le c \int_{t_0}^{t} |g_1(s) - g_2(s)|ds. \tag{5.82}$$

Let us assume that $t \in [t_0, t_0 + a_*]$. Then

$$|Tg_1(t) - Tg_2(t)| \le c \int_{t_0}^{t} |g_1(s) - g_2(s)|ds \le c \int_{t_0}^{t} e^{2c(s-t_0)} d_*(g_1, g_2)ds \tag{5.83}$$

$$= cd_*(g_1, g_2)\frac{1}{2c}(e^{2c(t-t_0)} - 1) \le \tfrac{1}{2} d_*(g_1, g_2)e^{2c|t-t_0|} \tag{5.84}$$

Similarly, for $t \in [t_0 - a_*, t_0]$ we also have

$$|Tg_1(t) - Tg_2(t)| \le \tfrac{1}{2} d_*(g_1, g_2)e^{2c|t-t_0|}. \tag{5.85}$$

[2]For the supremum metric to give rise to a contraction we would need to make the interval I smaller.

Thus,

$$e^{-2c|t-t_0|}|Tg_1(t) - Tg_2(t)| \le \tfrac{1}{2} d_*(g_1, g_2) \tag{5.86}$$

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 (5.74). □

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

$$y_n(t) = y_0 + \int_{t_0}^{t} F(s, y_{n-1}(s))ds \quad \text{for } n = 1, 2, \dots \tag{5.87}$$

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 (5.73).

Example 5.29. Consider the initial value problem

$$\begin{cases} y'(t) = \frac{e^t \sin(t+y(t))}{ty(t)-1}, \\ y(1) = 5. \end{cases} \tag{5.88}$$

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,

$$R = \{(t, y) \,:\, |t - 1| \le \tfrac{1}{2},\ |y - 5| \le 1\}. \tag{5.89}$$

Then for $(t, y) \in R$:

$$|ty - 1| \ge (1 - \tfrac{1}{2})(5 - 1) - 1 = 1. \tag{5.90}$$

Also, $|e^t \sin(t + y)| \le e^{3/2}$. Setting $M = e^{3/2}$, we obtain

$$|F(t, y)| \le M \quad \text{for all } (t, y) \in R. \tag{5.91}$$

Compute

$$\partial_y F(t, y) = \frac{e^t \cos(t + y)}{ty - 1} - t\frac{e^t \sin(t + y)}{(ty - 1)^2}. \tag{5.92}$$

For $(t, y) \in R$ we estimate

$$|\partial_y F(t, y)| \le \left|\frac{e^t \cos(t + y)}{ty - 1}\right| + \left|t\frac{e^t \sin(t + y)}{(ty - 1)^2}\right| \le c, \tag{5.93}$$

where we have set $c = e^{3/2} + \frac{3}{2}e^{3/2}$. Then the number a_* from Theorem 5.27 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 (5.88) 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

$$y_n(t) = 5 + \int_1^t \frac{e^s \sin(s + y_{n-1}(s))}{sy_{n-1}(s) - 1} ds. \tag{5.94}$$

The sequence $(y_n)_n$ converges uniformly on I to the solution y.

EXAMPLE 5.30. Sometimes one can extend solutions beyond the interval obtained from the Picard-Lindelöf theorem. Consider the initial value problem

$$\begin{cases} y'(t) = \cos(y(t)^2 - 2t^3) \\ y(0) = 1 \end{cases} \tag{5.95}$$

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

$$R = \{(t, y) \in \mathbb{R}^2 \,:\, |t| \le L, |y - 1| \le L\}. \tag{5.96}$$

Let $F(t, y) = \cos(y^2 - 2t^3)$. Then

$$|F(t, y)| \le 1 \quad \text{for all } (t, y) \in \mathbb{R}^2. \tag{5.97}$$

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 5.27, there exists a unique solution to (5.95) on $I = [-L, L]$.

EXAMPLE 5.31. If the Lipschitz condition (5.73) fails, then the initial value problem may have more than one solution. Consider

$$\begin{cases} y'(t) = |y(t)|^{1/2}, \\ y(0) = 0. \end{cases} \tag{5.98}$$

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 (5.98). The function

$$y_2(t) = \begin{cases} t^2/4, & \text{if } t > 0, \\ 0, & \text{if } t \le 0 \end{cases} \tag{5.99}$$

also does.

Lecture 32 (Monday, November 18)

Existence of a solution still holds without the assumption (5.73). We will prove this as a consequence of the Arzelá-Ascoli theorem.

THEOREM 5.32 (Peano existence theorem). *Let $E \subset \mathbb{R} \times \mathbb{R}$ open, $(t_0, y_0) \in E$, $F \in C(E)$,*

$$R = \{(t, y) : |t - t_0| \leq a, |y - y_0| \leq b\} \subset E. \tag{5.100}$$

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

$$\begin{cases} y'(t) = F(t, y(t)), \\ y(t_0) = y_0. \end{cases} \tag{5.101}$$

COROLLARY 5.33. *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* (5.74).

PROOF. It suffices to produce a solution to the integral equation

$$y(t) = y_0 + \int_{t_0}^{t} F(s, y(s))ds. \tag{5.102}$$

To avoid some technicalities we will only present the proof under the additional assumption that

$$|F(t, y)| \leq M \tag{5.103}$$

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

$$y_{\mathcal{P}}(t) = y_{\mathcal{P}}(t_k) + F(t_k, y_{\mathcal{P}}(t_k))(t - t_k). \tag{5.104}$$

Claim 1. For $t, t' \in [t_0, t_0 + a]$,

$$|y_{\mathcal{P}}(t) - y_{\mathcal{P}}(t')| \leq M|t - t'|. \tag{5.105}$$

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 (5.103),

$$|y(t) - y(t')| = |F(t_k, y(t_k))(t - t')| \leq M|t - t'|. \tag{5.106}$$

If $k < \ell$, then

$$|y(t) - y(t')| = |y(t) - y(t_\ell) + \sum_{j=k+1}^{\ell-1} (y(t_{j+1}) - y(t_j)) + y(t_{k+1}) - y(t')| \tag{5.107}$$

$$\leq |y(t) - y(t_\ell)| + \sum_{j=k+1}^{\ell-1} |y(t_{j+1}) - y(t_j)| + |y(t_{k+1}) - y(t')| \tag{5.108}$$

$$\leq M(t-t_\ell)+\sum_{j=k+1}^{\ell-1} M(t_{j+1}-t_j)+M(t_{k+1}-t')=M(t-t'). \tag{5.109}$$

□

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 2.10). Thus there exists $\delta=\delta(\varepsilon)>0$ such that

$$|F(t,y)-F(t',y')|\leq\varepsilon \tag{5.110}$$

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

$$y_{\mathcal{P}}(t)=y_0+\int_{t_0}^{t} g_{\mathcal{P}}(s)ds \quad\text{and}\quad |g_{\mathcal{P}}(s)-F(s,y_{\mathcal{P}}(s))|\leq\varepsilon \text{ if } s\in(t_{k-1},t_k). \tag{5.111}$$

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 $t=t_k$:

$$y(t_k)-y_0=y(t_k)-y(t_0)=\sum_{j=1}^{k} y(t_j)-y(t_{j-1}) \tag{5.112}$$

$$=\sum_{j=1}^{k} F(t_{j-1},y(t_{j-1}))(t_j-t_{j-1})=\sum_{j=1}^{k}\int_{t_{j-1}}^{t_j} g(s)ds=\int_{t_0}^{t_k} g(s)ds. \tag{5.113}$$

Similarly, for $t\in(t_k,t_{k+1})$:

$$y(t)-y(t_k)=F(t_k,y(t_k))(t-t_k)=\int_{t_k}^{t} g(s)ds. \tag{5.114}$$

Thus,

$$y(t)=y(t_k)+\int_{t_k}^{t} g(s)ds=y_0+\int_{t_0}^{t_k} g(s)ds+\int_{t_k}^{t} g(s)ds=y_0+\int_{t_0}^{t} g(s)ds. \tag{5.115}$$

Let $s\in(t_{k-1},t_k)$. Then

$$|g(s)-F(s,y(s))|=|F(t_{k-1},y(t_{k-1}))-F(s,y(s))|. \tag{5.116}$$

We have

$$|y(t_{k-1})-y(s)|\leq M|t_{k-1}-s|\leq M(t_k-t_{k-1})\leq M\cdot\Delta\mathcal{P}\leq\delta. \tag{5.117}$$

Also, $|t_{k-1}-s|\leq t_k-t_{k-1}\leq\Delta\mathcal{P}\leq\delta$. Thus,

$$\|(t_{k-1},y(t_{k-1}))-(s,y(s))\|\leq 100\delta. \tag{5.118}$$

By (5.110),

$$|g(s)-F(s,y(s))|=|F(t_{k-1},y(t_{k-1}))-F(s,y(s))|\leq\varepsilon. \tag{5.119}$$

□

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

$$|y_{\mathcal{P}}(t) - (y_0 + \int_{t_0}^{t} F(s, y_{\mathcal{P}}(s))ds)| \le \varepsilon a. \tag{5.120}$$

Proof of claim. By Claim 2, the left hand side equals

$$\Big| \int_{t_0}^{t} (g_{\mathcal{P}}(s) - F(s, y_{\mathcal{P}}(s)))ds \Big| \le \int_{t_0}^{t} |g_{\mathcal{P}}(s) - F(s, y_{\mathcal{P}}(s))|ds. \tag{5.121}$$

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,

$$|y_{\mathcal{P}}(t) - y_{\mathcal{P}}(t')| \le M|t - t'|. \tag{5.122}$$

This implies that $\mathcal{F}$ is equicontinuous. It is also bounded:

$$|y_{\mathcal{P}}(t)| \le |y_{\mathcal{P}}(t_0)| + |y_{\mathcal{P}}(t) - y_{\mathcal{P}}(t_0)| \le |y_0| + M|t - t_0| \le |y_0| + Ma. \tag{5.123}$$

Thus the claim follows from the Arzelà-Ascoli theorem (Theorem 2.31). □

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 $(y_{\mathcal{P}_n})_n \subset \mathcal{F} \subset \overline{\mathcal{F}}$ 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 (5.102). Let us denote that subsequence by $(y_n)_n$. 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,

$$\int_{t_0}^{t} F(s, y_n(s))ds \longrightarrow \int_{t_0}^{t} F(s, y(s))ds \quad \text{as } n \to \infty. \tag{5.124}$$

On the other hand, by Claim 3 we get

$$|y_n(t) - (y_0 + \int_{t_0}^{t} F(s, y_n(s))ds)| \le \frac{a}{n} \longrightarrow 0 \quad \text{as } n \to \infty. \tag{5.125}$$

Therefore, y solves the integral equation (5.102). □

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:

$$\begin{cases} y'(t) = F(t, y(t)), \\ y(t_0) = y_0 \end{cases} \tag{5.126}$$

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 $d \ge 1$ and consider the d-th order ordinary differential equation given by

$$y^{(d)}(t) = F(t, y(t), y'(t), \dots, y^{(d-1)}(t)) \tag{5.127}$$

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 $Y = (Y_1, \dots, Y_d)$ solves the system

$$\begin{cases} Y_1'(t) = Y_2(t) \\ Y_2'(t) = Y_3(t) \\ \vdots \\ Y_{d-1}'(t) = Y_d(t) \\ Y_d'(t) = F(t, Y(t)) \end{cases} \tag{5.128}$$

then Y_d is a solution to (5.127).

Lecture 33 (Wednesday, November 20)

4. Higher order derivatives and Taylor's theorem

DEFINITION 5.34. Let $U \subset \mathbb{R}^n$ be open and $f : U \to \mathbb{R}$. We define the *partial derivatives of second order* as

$$\partial_{ij} f = \partial_{x_i x_j} f = \partial_{x_i}(\partial_{x_j} f) \quad \text{for } i, j \in \{1, \dots, n\} \tag{5.129}$$

(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 $i, j \in \{1, \dots, n\}$, then we say that $f \in C^2(U)$.

THEOREM 5.35 (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*

$$\partial_{x_j x_i} f(x_0) = \partial_{x_i x_j} f(x_0). \tag{5.130}$$

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

$$\partial_1 f(a, b + k) - \partial_1 f(a, b). \tag{5.131}$$

This leads us to consider the quantity

$$\Delta(a, b, h, k) = (f(a + h, b + k) - f(a, b + k)) - (f(a + h, b) - f(a, b)). \tag{5.132}$$

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

$$\Delta(a, b, h, k) = g(b + k) - g(b) = g'(\eta)k = k(\partial_2 f(a + h, \eta) - \partial_2 f(a, \eta)) \tag{5.133}$$

Since $\partial_{12} f$ exists at every point in U, another application of the mean value theorem yields

$$\Delta(a, b, h, k) = hk \partial_{12} f(\xi, \eta), \tag{5.134}$$

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),

$$|\partial_{12} f(a, b) - \partial_{12} f(x, y)| \leq \varepsilon \tag{5.135}$$

whenever $\|(a, b) - (x, y)\|$ is small enough. Thus, for small enough h and k we have

$$|\partial_{12} f(a, b) - \frac{\Delta(a, b, h, k)}{hk}| \leq \varepsilon. \tag{5.136}$$

Letting $h \to 0$ and using that $\partial_1 f$ exists at every point this inequality implies

$$\left|\partial_{12} f(a, b) - \frac{\partial_1 f(a, b + k) - \partial_1 f(a, b)}{k}\right| \leq \varepsilon. \tag{5.137}$$

In other words, $\partial_{21} f(a, b)$ exists and equals $\partial_{12} f(a, b)$. □

This is not true without the assumption that $\partial_{x_i x_j} f$ is continuous at x.

EXERCISE 5.36. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x,y) = \begin{cases} xy\frac{x^2-y^2}{x^2+y^2}, & \text{if } (x,y) \neq (0,0), \\ 0, & \text{if } (x,y) = (0,0). \end{cases} \tag{5.138}$$

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 5.37. *If $f \in C^2(U)$, then $\partial_{x_i x_j} f = \partial_{x_j x_i} f$ for every $i, j \in \{1, \ldots, n\}$.*

DEFINITION 5.38. 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 $j \in \{1, \ldots, k\}$ and $i_1, \ldots, i_j \in \{1, \ldots, n\}$, 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 5.39. *If $f \in C^k(U)$ and $\pi : \{1, \ldots, k\} \to \{1, \ldots, k\}$ is a bijection, then*

$$\partial_{i_1} \cdots \partial_{i_k} f = \partial_{i_{\pi(1)}} \cdots \partial_{i_{\pi(k)}} f \tag{5.139}$$

for all $i_1, \ldots, i_k \in \{1, \ldots, n\}$.

Lecture 34 (Friday, November 22)

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

$$\partial^\alpha f = \partial_{x_1}^{\alpha_1} \cdots \partial_{x_n}^{\alpha_n} f, \tag{5.140}$$

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

$$x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n} \tag{5.141}$$

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 5.40 (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*

$$f(x+y) = \sum_{|\alpha| \le k} \frac{\partial^\alpha f(x)}{\alpha!} y^\alpha + \sum_{|\alpha| = k+1} \frac{\partial^\alpha f(\xi)}{\alpha!} y^\alpha. \tag{5.142}$$

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:

$$\sum_{|\alpha| \le k} \frac{\partial^\alpha f(x)}{\alpha!} y^\alpha = \sum_{\substack{\alpha_1, \dots, \alpha_n \ge 0, \\ \alpha_1 + \cdots + \alpha_n \le k}} \frac{\partial_{x_1}^{\alpha_1} \cdots \partial_{x_n}^{\alpha_n} f(x)}{\alpha_1! \cdots \alpha_n!} y_1^{\alpha_1} \cdots y_n^{\alpha_n}. \tag{5.143}$$

PROOF OF THEOREM 5.40. 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 $m = 1, \dots, k + 1$,

$$g^{(m)}(t) = \sum_{|\alpha| = m} \frac{m!}{\alpha!} \partial^\alpha f(x + ty) y^\alpha \tag{5.144}$$

PROOF OF CLAIM. We first show by induction on m that

$$g^{(m)}(t) = \sum_{i_1, \dots, i_m = 1}^n \partial_{i_1} \cdots \partial_{i_m} f(x + ty) y_{i_1} \cdots y_{i_m}. \tag{5.145}$$

Indeed, for $m = 1$, by the chain rule,

$$g'(t) = \sum_{i=1}^n \partial_i f(x + ty) y_i. \tag{5.146}$$

Suppose we have shown it for m. Then

$$g^{(m+1)}(t) = \frac{d}{dt} g^{(m)}(t) = \frac{d}{dt} \sum_{i_1, \dots, i_m = 1}^n \partial_{i_1} \cdots \partial_{i_m} f(x + ty) y_{i_1} \cdots y_{i_m}. \tag{5.147}$$

By the chain rule this equals

(5.148)
$$= \sum_{i_1,\dots,i_m=1}^{n} \sum_{i=1}^{n} \partial_{i_1} \cdots \partial_{i_m} \partial_i f(x+ty) y_{i_1} \cdots y_{i_m} y_i = \sum_{i_1,\dots,i_{m+1}=1}^{n} \partial_{i_1} \cdots \partial_{i_{m+1}} f(x+ty) y_{i_1} \cdots y_{i_{m+1}}.$$

It remains to show that

$$\sum_{i_1,\dots,i_m=1}^{n} \partial_{i_1} \cdots \partial_{i_m} f(x+ty) y_{i_1} \cdots y_{i_m} = \sum_{|\alpha|=m} \frac{m!}{\alpha!} \partial^\alpha f(x+ty) y^\alpha. \tag{5.149}$$

This follows because for a given $\alpha = (\alpha_1, \dots, \alpha_n)$ with $|\alpha| = m$ there are

$$\frac{m!}{\alpha!} = \frac{m!}{\alpha_1! \cdots \alpha_n!} = \binom{m}{\alpha_1} \binom{m-\alpha_1}{\alpha_2} \cdots \binom{m-\alpha_1-\cdots-\alpha_{n-1}}{\alpha_n} \tag{5.150}$$

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

$$g(t) = \sum_{m=0}^{k} \frac{g^{(m)}(0)}{m!} t^m + \frac{g^{(k+1)}(\theta)}{(k+1)!} t^{k+1} \tag{5.151}$$

From the claim we see that this equals

$$\sum_{m=0}^{k} \frac{1}{m!} \sum_{|\alpha|=m} \frac{m!}{\alpha!} \partial^\alpha f(x) y^\alpha t^m + \frac{1}{(k+1)!} \sum_{|\alpha|=k+1} \frac{(k+1)!}{\alpha!} \partial^\alpha f(x+\theta y) y^\alpha t^{k+1} \tag{5.152}$$

$$= \sum_{|\alpha|\le k} \frac{\partial^\alpha f(x)}{\alpha!} (ty)^\alpha + \sum_{|\alpha|=k+1} \frac{\partial^\alpha f(\xi)}{\alpha!} (ty)^\alpha, \tag{5.153}$$

where we have set $\xi = x + \theta y$. Letting $t = 1$ we obtain the claim. □

Corollary 5.41. *If $E \subset \mathbb{R}^n$ is open and $f \in C^k(E)$, then for every $x \in E$,*

$$f(x+y) = \sum_{|\alpha|\le k} \frac{\partial^\alpha f(x)}{\alpha!} y^\alpha + o(\|y\|^k) \quad \text{as } y \to 0. \tag{5.154}$$

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

(5.155)
$$f(x+y) = \sum_{|\alpha|\le k-1} \frac{\partial^\alpha f(x)}{\alpha!} y^\alpha + \sum_{|\alpha|=k} \frac{\partial^\alpha f(x+\theta y)}{\alpha!} y^\alpha = \sum_{|\alpha|\le k} \frac{\partial^\alpha f(x)}{\alpha!} y^\alpha + \sum_{|\alpha|=k} \frac{\partial^\alpha f(x+\theta y) - \partial^\alpha f(x)}{\alpha!} y^\alpha$$

for some $\theta \in [0,1]$. Since $\partial^\alpha f$ is continuous for every $|\alpha| = k$, it holds that

$$|\partial^\alpha f(x+\theta y) - \partial^\alpha f(x)| \to 0 \quad \text{as } y \to 0. \tag{5.156}$$

Also $|y^\alpha| = |y_1|^{\alpha_1} \cdots |y_n|^{\alpha_n} \le \|y\|^{\alpha_1+\cdots+\alpha_n} = \|y\|^{|\alpha|}$, so

$$\sum_{|\alpha|=k} \frac{\partial^\alpha f(x+\theta y) - \partial^\alpha f(x)}{\alpha!} y^\alpha = o(\|y\|^k). \tag{5.157}$$

□

DEFINITION 5.42. 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

$$D^2 f|_x = (\partial_i \partial_j f(x))_{i,j=1,\dots,n} = \begin{pmatrix} \partial_1^2 f(x) & \cdots & \partial_1 \partial_n f(x) \\ \vdots & \ddots & \vdots \\ \partial_n \partial_1 f(x) & \cdots & \partial_n^2 f(x) \end{pmatrix} \in \mathbb{R}^{n \times n}. \tag{5.158}$$

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 5.35 the Hessian matrix is symmetric.

COROLLARY 5.43. *Let $E \subset \mathbb{R}^n$ be open, $f \in C^2(E)$ and $x \in E$. Then*

$$f(x+y) = f(x) + \langle \nabla f(x), y \rangle + \tfrac{1}{2} \langle y, D^2 f|_x y \rangle + o(\|y\|^2) \quad \text{as } y \to 0. \tag{5.159}$$

(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 5.41,

$$f(x+y) = f(x) + \sum_{|\alpha|=1} \frac{\partial^\alpha f(x)}{\alpha!} y^\alpha + \sum_{|\alpha|=2} \frac{\partial^\alpha f(x)}{\alpha!} y^\alpha + o(\|y\|^2) \quad \text{as } y \to 0. \tag{5.160}$$

We have

$$\sum_{|\alpha|=1} \frac{\partial^\alpha f(x)}{\alpha!} y^\alpha = \sum_{i=1}^n \partial_i f(x) y_i = \langle \nabla f(x), y \rangle. \tag{5.161}$$

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,

$$\sum_{|\alpha|=2} \frac{\partial^\alpha f(x)}{\alpha!} y^\alpha = \tfrac{1}{2} \sum_{i=1}^n \partial_i^2 f(x) y_i^2 + \sum_{1 \le i < j \le n} \partial_i \partial_j f(x) y_i y_j = \tfrac{1}{2} \sum_{i,j=1}^n \partial_i \partial_j f(x) y_i y_j \tag{5.162}$$

$$= \tfrac{1}{2} \sum_{i=1}^n y_i (D^2 f|_x y)_i = \tfrac{1}{2} \langle y, D^2 f|_x y \rangle. \tag{5.163}$$

□

Lecture 35 (Monday, November 25)

5. Local extrema

Let $E \subset \mathbb{R}^n$ be an open set and $f : E \to \mathbb{R}$ a function.

DEFINITION 5.44. 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 $f(a) \geq f(x)$ for all $x \in U$. It is called a *strict local maximum* if $f(a) > f(x)$ for all $x \in U$, $x \neq a$. 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 5.45. *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 $|t| \leq \delta$. 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 5.46. *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 5.47. 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 $\langle x, Ax\rangle \geq 0$ for all $x \in \mathbb{R}^n$. We also write $A > 0$ to express that A is positive definite and $A \geq 0$ 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 5.48. *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 5.43 there exists $\delta > 0$ such that for all y with $\|y\| \leq \delta$ we have

$$f(a + y) = f(a) + \tfrac{1}{2}\langle y, Ay\rangle + r(y) \tag{5.164}$$

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

$$\langle y_0, Ay_0\rangle \leq \langle y, Ay\rangle \tag{5.165}$$

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

$$\alpha \leq \langle \frac{y}{\|y\|}, A\frac{y}{\|y\|} \rangle = \frac{1}{\|y\|^2} \langle y, Ay \rangle. \tag{5.166}$$

Thus, $\langle y, Ay \rangle \geq \alpha \|y\|^2$ for all $y \in \mathbb{R}^n$. Now we set $\varepsilon = \frac{\alpha}{4}$. Then

$$f(a+y) \geq f(a) + \tfrac{1}{2}\langle y, Ay \rangle - \tfrac{\alpha}{4}\|y\|^2 \geq f(a) + \tfrac{\alpha}{2}\|y\|^2 - \tfrac{\alpha}{4}\|y\|^2 = f(a) + \tfrac{\alpha}{4}\|y\|^2 > f(a) \tag{5.167}$$

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

$$f(y'') < f(a) < f(y'). \tag{5.168}$$

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

$$f(a + t\xi) = f(a) + \tfrac{1}{2}\langle t\xi, At\xi \rangle + r(t\xi) = f(a) + \tfrac{1}{2}\alpha t^2 + r(t\xi). \tag{5.169}$$

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 5.49. (1) Let $f(x, y) = c + x^2 + y^2$ for $c \in \mathbb{R}$. Then

$$D^2 f|_0 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} > 0 \tag{5.170}$$

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

$$D^2 f|_0 = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} \tag{5.171}$$

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

$$D^2 f_i|_0 = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \geq 0, \tag{5.172}$$

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.

Optional topic (not relevant for exams)

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 5.50 (Linear regression). Say we are given finitely many points

$$(x_1, y_1), \dots, (x_N, y_N) \in \mathbb{R}^n \times \mathbb{R}. \tag{5.173}$$

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

$$F_{a,b}(x) = \langle x, a\rangle + b \tag{5.174}$$

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

$$E(a, b) = \sum_{i=1}^{N} (F_{a,b}(x_i) - y_i)^2. \tag{5.175}$$

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 5.46 we know that every minimum must be a critical points. 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

$$x_{n+1} = x_n - \alpha_n \nabla f(x_n) \tag{5.176}$$

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 $(x_n)_n$ converges, then the limit is a critical point of f. The following lemma gives some more hope.

LEMMA 5.51. *Let $f \in C^1(E)$. Then, for every $x \in E$ and small enough $\alpha > 0$,*

$$f(x - \alpha \nabla f(x)) \le f(x). \tag{5.177}$$

PROOF. By the definition of total derivatives,

$$f(x-\alpha\nabla f(x)) = f(x)+\langle \nabla f(x), -\alpha \nabla f(x)\rangle + o(\alpha) = f(x) - \alpha \|\nabla f(x)\|^2 + o(\alpha) \tag{5.178}$$

which is $\le f(x)$ 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 5.52. Let $E \subset \mathbb{R}^n$ be convex. A function $f : E \to \mathbb{R}$ is called *convex* if

$$f(tx + (1-t)y) \le tf(x) + (1-t)f(y) \tag{5.179}$$

for all $x, y \in E$, $t \in [0, 1]$. f is called *strictly convex* if

$$f(tx + (1-t)y) < tf(x) + (1-t)f(y) \tag{5.180}$$

for all $x \neq y \in E$ and $t \in (0,1)$.

THEOREM 5.53. *Let $E \subset \mathbb{R}^n$ be open and convex and $f \in C^1(E)$. Then f is convex if and only if*

$$f(u+v) \geq f(u) + \langle \nabla f(u), v \rangle \tag{5.181}$$

for all $u, u+v \in E$.

PROOF. $\underline{\Rightarrow}$: Fix $u, u+v \in E$. By convexity, for $t \in [0,1]$,

$$f(u+tv) = f((1-t)u + t(u+v)) \leq (1-t)f(u) + tf(u+v). \tag{5.182}$$

By definition of the derivative,

$$f(u+tv) = f(u) + t\nabla f(u)^T v + r(t), \tag{5.183}$$

where $\lim_{t\to 0} \frac{r(t)}{t} = 0$. Thus,

$$f(u) + t\langle \nabla f(u), v \rangle + r(t) \leq (1-t)f(u) + tf(u+v) \tag{5.184}$$

which implies

$$f(u) + \langle \nabla f(u), v \rangle - f(u+v) \leq \frac{-r(t)}{t} \to 0 \quad \text{as } t \to 0. \tag{5.185}$$

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

$$f(x) \geq f(u) + \langle \nabla f(u), x - u \rangle. \tag{5.186}$$

On the other hand, letting $v = y - u$, the assumption implies

$$f(y) \geq f(u) + \langle \nabla f(u), y - u \rangle. \tag{5.187}$$

Therefore

$$tf(x) + (1-t)f(y) \geq t(f(u) + \langle \nabla f(u), x-u \rangle) + (1-t)(f(u) + \langle \nabla f(u), y-u \rangle) \tag{5.188}$$

$$= f(u) + \langle \nabla f(u), t(x-u) + (1-t)(y-u) \rangle = f(u) + \langle \nabla f(u), tx + (1-t)y - u \rangle. \tag{5.189}$$

Recalling that $u = tx + (1-t)y$, we get

$$tf(x) + (1-t)f(y) \geq f(u) = f(tx + (1-t)y). \tag{5.190}$$

□

THEOREM 5.54. *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 $u, u+tv \in E$,

$$f(u+tv) = f(u) + t\langle \nabla f(u), v \rangle + \tfrac{1}{2}t^2 \langle D^2 f|_u v, v \rangle + o(t^2) \tag{5.191}$$

and by Theorem 5.53,

$$f(u+tv) \geq f(u) + t\langle \nabla f(u), v \rangle. \tag{5.192}$$

Combining these two pieces of information we obtain

$$\tfrac{1}{2}t^2 \langle D^2 f|_u v, v \rangle + o(t^2) \geq 0 \tag{5.193}$$

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

$$f(u+v) = f(u) + \langle \nabla f(u), v\rangle + \tfrac{1}{2}\langle D^2 f|_\xi v, v\rangle \geq f(u) + \langle \nabla f(u), v\rangle. \tag{5.194}$$

Therefore f is convex by Theorem 5.53. □

Remark. If f is strictly convex, then it does not follow that $D^2 f|_x > 0$ for all x.

EXAMPLE 5.55. 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 5.56. *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

$$f(x) = f(x_*) + \underbrace{\langle \nabla f(x_*), x - x_*\rangle}_{=0} + \tfrac{1}{2}\underbrace{\langle D^2 f|_\xi (x - x_*), x - x_*\rangle}_{\geq 0} \geq f(x_*). \tag{5.195}$$

(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,

$$f(x_1) = \frac{f(x_1) + f(x_2)}{2} > f\Big(\frac{x_1 + x_2}{2}\Big). \tag{5.196}$$

This is a contradiction to x_1 being a global minimum. Therefore $x_1 = x_2$. □

EXAMPLE 5.57. If $\|\cdot\|$ is a norm on $\mathbb{R}^n$, then the function $x \mapsto \|x\|$ is convex:

$$\|tx + (1-t)y\| \leq t\|x\| + (1-t)\|y\| \tag{5.197}$$

by the triangle inequality. Also, this function has a unique global minimum at $x = 0$.

LEMMA 5.58. *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,

$$h(tx + (1-t)y) = g(f(tx + (1-t)y)) \leq g(tf(x) + (1-t)f(y)). \tag{5.198}$$

Since g is convex this is

$$\leq tg(f(x)) + (1-t)g(f(y)) = th(x) + (1-t)h(y). \tag{5.199}$$

□

COROLLARY 5.59. *If $\|\cdot\|$ is a norm on $\mathbb{R}^n$, then the function $x \mapsto \|x\|^2$ is convex.*

EXAMPLE 5.60. Recall the error function from linear regression (Example 5.50):

$$E(a,b) = \sum_{i=1}^{N} (\langle a, x_i \rangle + b - y_i)^2 \tag{5.200}$$

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

$$M = \begin{pmatrix} x_{11} & \cdots & x_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ x_{N1} & \cdots & x_{Nn} & 1 \end{pmatrix} \in \mathbb{R}^{N \times (n+1)}, \quad v = \begin{pmatrix} a_1 \\ \vdots \\ a_n \\ b \end{pmatrix} \in \mathbb{R}^{n+1}, \tag{5.201}$$

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

$$E(a,b) = E(v) = \sum_{i=1}^{N} ((Mv)_i - y_i)^2 = \|Mv - y\|^2, \tag{5.202}$$

where $\|c\| = \left(\sum_{i=1}^{N} |c_i|^2 \right)^{1/2}$.

Let us rename variables and consider

$$E(x) = \|Mx - y\|^2 \tag{5.203}$$

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

$$\partial_i F(y) = 2y_i, \text{ so } DF|_y = 2y^T \in \mathbb{R}^{1 \times N}. \tag{5.204}$$

and $DG|_x = M \in \mathbb{R}^{N \times n}$. Therefore, by the chain rule we obtain

$$DE|_x = 2(Mx - y)^T M = 2(Mx)^T M - 2y^T M = 2x^T M^T M - 2y^T M \in \mathbb{R}^{1 \times n}. \tag{5.205}$$

Therefore,

$$D^2 E|_x = (\partial_i DE|_x)_{i=1,\ldots,n} = (2(M^T M)_i)_{i=1,\ldots,n} = 2M^T M. \tag{5.206}$$

Notice that $M^T M$ is positive semidefinite because

$$\langle M^T Mx, x \rangle = \langle Mx, Mx \rangle = \|Mx\|^2 \geq 0. \tag{5.207}$$

Therefore E is convex by Theorem 5.54.

EXAMPLE 5.61. 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 5.62. 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

$$\langle D^2 f|_x y, y \rangle \geq \beta \|y\|^2 \tag{5.208}$$

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 5.54).
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 5.63. 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

$$\beta_-\|y\|^2 \leq \langle D^2 f|_x y, y\rangle \leq \beta_+\|y\|^2 \tag{5.209}$$

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 5.64. *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*

$$f(u+v) \geq f(u) + \langle \nabla f(u), v\rangle + \gamma\|v\|^2 \tag{5.210}$$

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 5.53,

$$g(u+v) \geq g(u) + \langle \nabla g(u), v\rangle = f(u) - \tfrac{\beta}{2}\|u\|^2 + \langle \nabla f(u) - \beta u, v\rangle \tag{5.211}$$

On the other hand,

$$g(u+v) = f(u+v) - \tfrac{\beta}{2}\|u+v\|^2 \tag{5.212}$$

Thus,

$$f(u+v) \geq f(u) + \langle \nabla f(u), v\rangle + \tfrac{\beta}{2}(\|u+v\|^2 - \|u\|^2 - 2\langle u, v\rangle) = f(u) + \langle \nabla f(u), v\rangle + \tfrac{\beta}{2}\|v\|^2. \tag{5.213}$$

$\underline{\Leftarrow:}$ This follows in the same way from the converse direction of Theorem 5.53. □

THEOREM 5.65. *Let $f \in C^2(\mathbb{R}^n)$ be strongly convex. Then for every $c \in \mathbb{R}$, the* sublevel set

$$B = \{x \in \mathbb{R}^n \,:\, f(x) \leq c\} \tag{5.214}$$

is bounded.

PROOF. By Theorem 5.64 we have

$$f(x) \geq f(0) + \langle \nabla f(0), x\rangle + \gamma\|x\|^2. \tag{5.215}$$

Therefore, $\lim_{\|x\|\to\infty} f(x) = \infty$. Suppose that B is unbounded. Then there would exist a sequence $(x_n)_{n\geq 1} \subset B$ such that $\lim_{n\to\infty} \|x_n\| = \infty$. But $f(x_n) \leq c$, so $f(x_n) \not\to \infty$ as $n \to \infty$. Contradiction! □

THEOREM 5.66. *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 $B = \{x \in \mathbb{R}^n : f(x) \leq f(0)\}$ is bounded. Thus, there exists $R > 0$ such that $B \subset B_R = \{x \in \mathbb{R}^n : \|x\| \leq R\}$. B_R is compact, so f attains its minimum on $B_R(0)$ at some point $x_* \in B_R$. Then $f(x_*) \leq f(x)$ for all $x \in B_R$. It remains to show $f(x_*) \leq f(x)$ 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 $f(x_*) \leq f(0) < f(x)$. □

We conclude this discussion by proving that gradient descent converges for strongly convex functions.

THEOREM 5.67. *Let $f \in C^2(\mathbb{R}^n)$ be strongly convex and $x_0 \in \mathbb{R}^n$. Define*

$$x_{n+1} = x_n - \alpha \nabla f(x_n) \quad \text{for } n \geq 0. \tag{5.216}$$

If α is small enough, then $(x_n)_n$ 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 5.65 and 5.66).

LEMMA 5.68. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric and positive definite matrix. Then the matrix norm $\|A\|_{\text{op}} = \sup_{x\neq 0} \frac{\|Ax\|}{\|x\|}$ 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 $\{v_1, \ldots, v_n\}$ be an orthonormal basis of eigenvectors corresponding to eigenvalues $\lambda_1, \ldots, \lambda_n$, respectively. Then

$$\|Ax\| = \Big\| \sum_{i=1}^n x_i A v_i \Big\| = \Big\| \sum_{i=1}^n x_i \lambda_i v_i \Big\| \tag{5.217}$$

which by orthogonality is equal to $\left(\sum_{i=1}^n |x_i|^2 \lambda_i^2 \right)^{1/2}$ (use that $\|x\| = (\langle x, x\rangle)^{1/2}$). Thus

$$\|Ax\| = (\sum_{i=1}^n |x_i|^2 \lambda_i^2)^{1/2} \leq \max_{i=1,\ldots,n} \lambda_i \left(\sum_{i=1}^n |x_i|^2 \right)^{1/2} = \max_{i=1,\ldots,n} \lambda_i \|x\|. \tag{5.218}$$

Let $\max_{i=1,\ldots,n} \lambda_i = \lambda_{i_0}$. We have shown that $\|A\| \leq \lambda_{i_0}$. On the other hand,

$$\|Av_{i_0}\| = \lambda_{i_0} \|v_{i_0}\| = \lambda_{i_0}, \tag{5.219}$$

so $\|A\| = \sup_{\|x\|=1} \|Ax\| \geq \|Av_{i_0}\| = \lambda_{i_0}$. □

PROOF OF THEOREM 5.67. 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

$$DT|_x = I - \alpha D^2 f|_x. \tag{5.220}$$

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

$$\langle D^2 f|_x y, y\rangle \leq \gamma \|y\|^2 \tag{5.221}$$

for all $y \in \mathbb{R}^n$ and $x \in B_R$. By strong convexity,

$$\beta\|y\|^2 \leq \langle D^2 f|_x y, y\rangle \leq \gamma\|y\|^2. \tag{5.222}$$

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

$$[1 - \tfrac{\gamma}{2\gamma}, 1 - \tfrac{\beta}{2\gamma}] = [\tfrac{1}{2}, 1 - \tfrac{\beta}{2\gamma}] \subset (0, 1). \tag{5.223}$$

Set $c = 1 - \frac{\beta}{2\gamma}$. By Lemma 5.68, we have

$$\|I - \alpha D^2 f|_x\| \leq c < 1. \tag{5.224}$$

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 $T(x_*) = x_*$,

$$\|T(x) - x_*\| = \|T(x) - T(x_*)\| \leq c\|x - x_*\| \leq cR \leq R. \tag{5.225}$$

□

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

Exercise 5.70. 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 5.71. A function $f : \mathbb{R}^n \to \mathbb{R}$ is called *homogeneous of degree* $\alpha \in \mathbb{R}$ if $f(\lambda x) = \lambda^\alpha f(x)$ for all $\lambda > 0$ and $x \in \mathbb{R}^n$. Suppose that f is differentiable. Then show that f is homogeneous of degree α if and only if

$$\sum_{i=1}^{n} x_i \partial_i f(x) = \alpha f(x) \tag{5.226}$$

for all $x \in \mathbb{R}^n$. *Hint:* Consider the function $g(\lambda) = f(\lambda x) - \lambda^\alpha f(x)$.

Exercise 5.72. Define $F : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$F(x, y) = (x^4 - y^4, e^{xy} - e^{-xy}). \tag{5.227}$$

(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 5.73. 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

$$f(x, y, z) = \left(\frac{x}{a + x + y + z}, \frac{y}{a + x + y + z}, \frac{z}{a + x + y + z}\right). \tag{5.228}$$

(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 5.74. 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 5.75. 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 5.76. Let $(t_0, y_0) \in \mathbb{R}^2$, $c \in \mathbb{R}$ and define $Y_0(t) = y_0$,

$$Y_n(t) = y_0 + c\int_{t_0}^{t} sY_{n-1}(s)ds. \tag{5.229}$$

Compute $Y_n(t)$ and $Y(t) = \lim_{n\to\infty} Y_n(t)$. Which initial value problem does Y solve?

EXERCISE 5.77. Consider the initial value problem

$$\begin{cases} y'(t) = e^{y(t)^2} - \frac{1}{ty(t)}, \\ y(1) = 1. \end{cases} \tag{5.230}$$

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 5.78. Consider the initial value problem

$$\begin{cases} y'(t) = t + \sin(y(t)), \\ y(2) = 1. \end{cases} \tag{5.231}$$

Find the largest interval $I \subseteq \mathbb{R}$ containing $t_0 = 2$ such that the problem has a unique solutions y in I.

EXERCISE 5.79. 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 $0 \le k \le a/h$.
Define a function y_h recursively by setting $y_h(t_0) = y_0$ and

$$y_h(t) = y_h(t_k) + (t - t_k)F(t_k, y_h(t_k)) \tag{5.232}$$

for $t \in (t_k, t_{k+1}]$ for integers $0 \le k \le a/h$.
(i) From the proof of Peano's theorem (Theorem 5.32) 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,

$$|y(t) - y_h(t)| \le Ch. \tag{5.233}$$

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 5.80. Let us improve the approximation from Exercise 5.79. In the context of that exercise, define a piecewise linear function y_h^* recursively by setting $y_h^*(t_0) = y_0$ and

$$y_h^*(t) = y_h^*(t_k) + (t - t_k)G(t_k, y_h^*(t_k), h), \tag{5.234}$$

for $t \in (t_k, t_{k+1}]$ for integers $0 \le k \le a/h$, where

$$G(t,y,h) = \tfrac{1}{2}(F(t,y) + F(t+h, y+hF(t,y))). \tag{5.235}$$

Prove that there exists a constant $C > 0$ such that for all $t \in I$ and $h > 0$ sufficiently small,

$$|y(t) - y_h^*(t)| \le Ch^2. \tag{5.236}$$

EXERCISE 5.81. For a function $f : [a,b] \to \mathbb{R}$ define

$$\mathscr{I}(f) = \int_a^b (1 + f'(t)^2)^{1/2} dt. \tag{5.237}$$

Let $\mathcal{A} = \{f \in C^2([a,b]) \,:\, f(a) = c, f(b) = d\}$. Determine $f_* \in \mathcal{A}$ such that

$$\mathscr{I}(f_*) = \inf_{f \in \mathcal{A}} \mathscr{I}(f). \tag{5.238}$$

What is the geometric meaning of $\mathscr{I}(f)$ and $\inf_{f\in\mathcal{A}} \mathscr{I}(f)$?

EXERCISE 5.82. 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 $\alpha \in \mathbb{N}_0^n$,

$$\partial^\alpha (f \cdot g)(x) = \sum_{\beta \in \mathbb{N}_0^n \,:\, \beta \le \alpha} \binom{\alpha}{\beta} \partial^\beta f(x) \partial^{\alpha-\beta} g(x) \tag{5.239}$$

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 5.83. 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 5.84. Determine the Taylor expansion of the function

$$f : (0,\infty) \times (0,\infty) \to \mathbb{R},\ f(x,y) = \frac{x-y}{x+y} \tag{5.240}$$

at the point $(x,y) = (1,1)$ up to order 2.

EXERCISE 5.85. Show that every continuous function $f : [a,b] \to [a,b]$ has a fixed point.

EXERCISE 5.86. Let X be a real Banach space. Let $B = \{x \in X \,:\, \|x\| \le 1\}$ 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 5.87. Determine the local minima and maxima of the function

$$f : \mathbb{R}^2 \to \mathbb{R},\ \ f(x,y) = (4x^2 + y^2) e^{-x^2 - 4y^2}. \tag{5.241}$$

EXERCISE 5.88. 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

$$f(x+y) = \sum_{|\alpha| \leq k} c_\alpha y^\alpha + o(\|y\|^k) \tag{5.242}$$

as $y \to 0$ and

$$f(x+y) = \sum_{|\alpha| \leq k} \widetilde{c}_\alpha y^\alpha + o(\|y\|^k) \tag{5.243}$$

as $y \to 0$. Show that $c_\alpha = \widetilde{c}_\alpha$ for all $|\alpha| \leq k$.

EXERCISE 5.89. 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 5.90. 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 5.91. 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 5.92. 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 5.93. Prove that the point x_* from Exercise 5.92 can be computed using gradient descent: that is, if $x_0 \in \mathbb{R}^n$ arbitrary and

$$x_{n+1} = x_n - \alpha \nabla f(x_n) \tag{5.244}$$

for $n = 0, 1, 2, \ldots$, then the sequence $(x_n)_n$ converges to x_* for all starting points $x_0 \in \mathbb{R}^n$, provided that α is chosen sufficiently small.

EXERCISE 5.94. Let $\mathcal{D} \subset \mathbb{R}^2$ be a finite set. Define a function $E : \mathbb{R}^3 \to \mathbb{R}$ by

$$E(a, b, c) = \sum_{x \in \mathcal{D}} (ax_1^2 + bx_1 + c - x_2)^2. \tag{5.245}$$

(1) Show that E is convex.
(2) Does there exist a set $\mathcal{D}$ such that E is strongly convex? Proof or counterexample.

EXERCISE 5.95. (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 5.96. (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 5.97. 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 5.98. 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

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - \gamma t(1-t)\|x-y\|^2 \tag{5.246}$$

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 5.99. (Recall Exercise 3.82 as motivation for this exercise.) Fix a function $\sigma \in C^1(\mathbb{R})$ and define for $x \in \mathbb{R}^n, W \in \mathbb{R}^{m\times n}, v \in \mathbb{R}^m$,

$$\mu(x, W, v) = \sum_{i=1}^{m} \sigma((Wx)_i)v_i \tag{5.247}$$

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

$$E(W, v) = \sum_{j=1}^{N} (\mu(x_i, W, v) - y_i)^2. \tag{5.248}$$

Is E necessarily convex? (Proof or counterexample.)

CHAPTER 6

The Baire category theorem*

Lecture 36 (Wednesday, November 27)

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.

Remarks. 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: let $X = \mathbb{Q}$ 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 $(x_n)_n$ 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 $(U_n)_n$ 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.

Lecture 37 (Monday, December 2)

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 $(f_k)_k \subset \mathcal{A}_n$ 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 $(t_k)_k \subset [0,1]$ 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 $(t_k)_k$ 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 2.44.

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, $X \setminus C_f$ 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 $X \setminus C_f$ 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 define the *oscillation of f at x* by

$$\omega_f(x) = \inf_{\varepsilon>0} \omega_f(B(x, \varepsilon)). \tag{6.11}$$

Then we have

$$x \in C_f \iff \omega_f(x) = 0 \tag{6.12}$$

and we can write the set of continuity of f as

$$C_f = \bigcap_{n=1}^{\infty} \{x \in X \,:\, \omega_f(x) < \tfrac{1}{n}\}. \tag{6.13}$$

We are done if we can show that $U_n = \{x \in X : \omega_f(x) < \frac{1}{n}\}$ is open for every $n \in \mathbb{N}$. Let $x_0 \in U_n$. Then $\omega_f(x_0) < \frac{1}{n}$. 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,

$$\omega_f(x) \leq \omega_f(B(x, \varepsilon/2)) \leq \omega_f(B(x_0, \varepsilon)) < \tfrac{1}{n}. \tag{6.14}$$

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. □

Remark. 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.*

Lecture 38 (Wednesday, December 4)

3. 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 6.24 (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*

$$\sup_{T\in\mathcal{F}} \|Tx\|_Y < \infty \text{ for all } x \in X \quad \Longleftrightarrow \quad \sup_{T\in\mathcal{F}} \|T\|_{\mathrm{op}} < \infty. \tag{6.15}$$

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

$$A_n = \{x \in X \,:\, \sup_{T\in\mathcal{F}} \|Tx\|_Y \leq n\} \subset X. \tag{6.16}$$

A_n is a closed set: if $(x_k)_k \subset A_n$ is a sequence with $x_k \to x \in X$, then since T is continuous, $\|Tx\|_Y = \lim_{k\to\infty} \|Tx_k\|_Y \leq n$ 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

$$X = \bigcup_{n=1}^{\infty} A_n. \tag{6.17}$$

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

$$\overline{B}(x_0, \varepsilon) \subset A_{n_0}. \tag{6.18}$$

Let $x \in X$ be such that $\|x\|_X \leq \varepsilon$. Then for all $T \in \mathcal{F}$,

$$\|Tx\|_Y = \|T(x_0 - x) - Tx_0\|_Y \leq \|T(x_0 - x)\|_Y + \|Tx_0\|_Y \leq 2n_0. \tag{6.19}$$

Now we use the usual scaling trick: let $x \in X$ satisfy $\|x\|_X = 1$. Then

$$\|Tx\|_Y = \varepsilon^{-1}\|T(\varepsilon x)\|_Y \leq 2\varepsilon^{-1} n_0. \tag{6.20}$$

This implies

$$\sup_{T\in\mathcal{F}} \|T\|_{\mathrm{op}} = \sup_{T\in\mathcal{F}} \sup_{\|x\|_X=1} \|Tx\|_Y \leq 2\varepsilon^{-1} n_0 < \infty. \tag{6.21}$$

□

EXAMPLE 6.25. If X is not complete, then the conclusion of the theorem may fail. For instance, let X be the space of all sequences $(x_n)_n \subset \mathbb{R}$ 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

$$|\ell_n(x)| = |nx_n| \leq n\|x\|_\infty. \tag{6.22}$$

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

$$\sup_{n\in\mathbb{N}} |\ell_n(x)| = \max\{|\ell_n(x)| \,:\, n = 1, \ldots, N_x\} < \infty. \tag{6.23}$$

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,

$$\sup_{n\in\mathbb{N}} \|\ell_n\|_{\text{op}} = \infty. \tag{6.24}$$

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 6.26. *Let X be a Banach space and Y a normed vector space. Suppose $(T_n)_n \subset L(X,Y)$ is such that $(T_n x)_n$ 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 $(T_n x)_n$ 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

$$\|Tx\|_Y = \lim_{n\to\infty} \|T_n x\|_Y \leq C\|x\|_X. \tag{6.25}$$

□

Remark. Note that in the context of Corollary 6.26 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 $L(X,Y)$.

Lecture 39 (Friday, December 6)

3.1. 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

$$S_N f(x) = \sum_{n=-N}^{N} c_n e^{2\pi i n x} = f * D_N(x), \tag{6.26}$$

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 6.27. *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 6.28. *There exists a constant $c \in (0, \infty)$ such that for every $N \in \mathbb{N}$,*

$$\int_0^1 |D_N(x)|dx \geq c \log(N). \tag{6.27}$$

PROOF. Since $|\sin(x)| \leq |x|$,

$$\int_0^1 |D_N(x)|dx = \int_0^1 \frac{|\sin(2\pi(N+\frac{1}{2})x)|}{|\sin(\pi x)|}dx \geq \pi^{-1}\int_0^1 \frac{|\sin(2\pi(N+\frac{1}{2})x)|}{x}dx. \tag{6.28}$$

Changing variables $2\pi(N+\frac{1}{2})x \mapsto x$ we see that the right hand side of this display equals

$$\pi^{-1}\int_0^{\pi(2N+1)} \frac{|\sin(x)|}{x}dx = \pi^{-1}\sum_{k=0}^{2N}\int_{\pi k}^{\pi(k+1)} \frac{|\sin(x)|}{x}dx. \tag{6.29}$$

We have that

$$\sum_{k=0}^{2N}\int_{\pi k}^{\pi(k+1)} \frac{|\sin(x)|}{x}dx \geq \sum_{k=0}^{2N}\int_{\pi k+\frac{\pi}{2}-\frac{\pi}{100}}^{\pi k+\frac{\pi}{2}+\frac{\pi}{100}} \frac{|\sin(x)|}{x}dx \geq c\sum_{k=0}^{2N}\int_{\pi k+\frac{\pi}{2}-\frac{\pi}{100}}^{\pi k+\frac{\pi}{2}+\frac{\pi}{100}} \frac{dx}{x}. \tag{6.30}$$

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,

$$\int_{\pi k+\frac{\pi}{2}-\frac{\pi}{100}}^{\pi k+\frac{\pi}{2}+\frac{\pi}{100}} \frac{dx}{x} \geq \tfrac{\pi}{50}\cdot\frac{1}{\pi k+\frac{\pi}{2}+\frac{\pi}{100}} \geq \tfrac{1}{50}\cdot\frac{1}{k+1}. \tag{6.31}$$

Thus,

$$\sum_{k=0}^{2N}\int_{\pi k+\frac{\pi}{2}-\frac{\pi}{100}}^{\pi k+\frac{\pi}{2}+\frac{\pi}{100}} \frac{dx}{x} \geq \tfrac{1}{50}\sum_{k=0}^{2N}\frac{1}{k+1} \geq \tfrac{1}{50}\sum_{k=0}^{2N}\int_{k+1}^{k+2}\frac{dx}{x} = \tfrac{1}{50}\int_1^{2N+2}\frac{dx}{x} = \tfrac{1}{50}\log(2N+2), \tag{6.32}$$

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

$$T_N f = S_N f(x_0). \tag{6.33}$$

LEMMA 6.29. *For every $N \in \mathbb{N}$, $T_N : C(\mathbb{T}) \to \mathbb{C}$ is a bounded linear map and*

$$\|T_N\|_{\text{op}} = \|D_N\|_1. \tag{6.34}$$

(Here $\|D_N\|_1 = \int_0^1 |D_N(x)|dx$.)

PROOF. For every $f \in C(\mathbb{T})$ we have

$$|T_N f| = |f * D_N(x_0)| \le \int_0^1 |f(x_0 - t)D_N(t)|dt \le \|f\|_\infty \int_0^1 |D_N(t)|dt = \|f\|_\infty \|D_N\|_1. \tag{6.35}$$

Therefore, T_N is bounded and $\|T_N\|_{\text{op}} \le \|D_N\|_1$. To prove the lower bound we let

$$f(x) = \text{sgn}(D_N(x_0 - x)). \tag{6.36}$$

While f is not a continuous function, it can be approximated by continuous functions as the following exercise shows.

EXERCISE 6.30. Show that for every $\varepsilon > 0$ there exists $g \in C(\mathbb{T})$ such that $|g(t)| \le 1$ for all $t \in \mathbb{R}$ and

$$\int_0^1 |f(t) - g(t)|dt \le \frac{\varepsilon}{2N+1} \tag{6.37}$$

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

$$|T_N f| = |f * D_N(x_0)| = \Big| \int_0^1 \text{sgn}(D_N(t))D_N(t)dt \Big| = \int_0^1 |D_N(t)|dt = \|D_N\|_1. \tag{6.38}$$

Moreover,

$$|T_N g| \ge |T_N f| - |T_N(f - g)|, \tag{6.39}$$

The error term $|T_N(f - g)|$ can be estimated as follows:

$$|T_N(f-g)| \le \int_0^1 |D_N(x_0-t)||f(t)-g(t)|dt \le \|D_N\|_\infty \int_0^1 |f(t)-g(t)|dt \le (2N+1)\frac{\varepsilon}{2N+1} = \varepsilon. \tag{6.40}$$

so

$$\|T_N\|_{\text{op}} \ge |T_N g| \ge \|D_N\|_1 - \varepsilon. \tag{6.41}$$

Since $\varepsilon > 0$ was arbitrary, this implies $\|T_N\|_{\text{op}} \ge \|D_N\|_1$. □

Armed with this knowledge, we can now reveal Corollary 6.27 as a direct consequence of Theorem 6.24. Indeed, we have that

$$\|T_N\|_{\text{op}} = \|D_N\|_1 \ge c \log(N) \tag{6.42}$$

[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.

and therefore

$$\sup_{N\in\mathbb{N}} \|T_N\|_{\mathrm{op}} = \infty. \tag{6.43}$$

So by Theorem 6.24 there must exist an $f \in C(\mathbb{T})$ such that

$$\sup_{N\in\mathbb{N}} |T_N f| = \infty. \tag{6.44}$$

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 6.27 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 $\mathbb{R} \to \mathbb{R}/\mathbb{Z} = \mathbb{T}$.

Lecture 40 (Monday, December 9)

4. Kakeya sets*

DEFINITION 6.31. We call a compact set $A \subset \mathbb{R}^n$ a *Kakeya set* if A contains a unit line segment in every direction. That is, if for every $v \in \mathbb{R}^n$ with $\|v\| = 1$ there exists $x \in A$ such that $x + tv \in A$ for all $t \in [0, 1]$.

(Note that this is only an interesting concept if $n \geq 2$.)

EXAMPLE 6.32. Consider the unit disk $A = \{x \in \mathbb{R}^2 : \|x\| \leq 1\}$. Clearly $0+tv \in A$ for every $t \in [0, 1]$ and $v \in \mathbb{R}^2$ with $\|v\| = 1$, so A is a Kakeya set in $\mathbb{R}^2$. The area of the unit disk is $\pi/4$.

EXAMPLE 6.33. Let A be the compact set the boundary of which is the *deltoid curve* defined by $\gamma(t) = (\frac{1}{2}\cos(t) + \frac{1}{4}\cos(2t), \frac{1}{2}\sin(t) - \frac{1}{4}\sin(2t))$ for $t \in \mathbb{R}$. It can be seen that A is a Kakeya set and has area $\pi/8$ (draw a picture).

Do there exist Kakeya sets in $\mathbb{R}^2$ with even smaller area? What is the smallest possible "area" or "volume" of a Kakeya set in $\mathbb{R}^n$?

While we are not going to attempt a rigorous definition of the notion of "volume" for an arbitrary subset of $\mathbb{R}^n$ at this point (this leads to a subject of its own, called *measure theory*), we can easily make rigorous what we mean by a subset of "zero volume".

DEFINITION 6.34. A set in $A \subset \mathbb{R}^n$ is called a *Lebesgue null set* (or *of Lebesgue measure zero*) if for every $\varepsilon > 0$ there exist $(x_1, r_1), (x_2, r_2), \ldots$ with $x_i \in \mathbb{R}^n$ and $r_i > 0$ such that

$$A \subset \bigcup_{i=1}^{\infty} B(x_i, r_i) \text{ and } \sum_{i=1}^{\infty} r_i^n \leq \varepsilon. \tag{6.45}$$

In other words, A is a Lebesgue null set if it can be covered by countably many balls the combined volume of which can be made arbitrarily small. Intuitively, Lebesgue null sets are sets of "volume zero".

The surprising answer to our question on the smallest possible volume of Kakeya sets is that Kakeya sets may have volume zero.

THEOREM 6.35. *Let $n \geq 2$. There exists a compact set $K \subset \mathbb{R}^n$ such that K is a Kakeya set and a Lebesgue null set.*

Remark. Many explicit constructions of such sets have been described in the literature. The first example (for the case $n = 2$) was given by Besicovitch in 1926. Therefore such sets are also called *Besicovitch sets*.

We will give a non-constructive proof using the Baire category theorem. This proof first appeared in *T. W. Körner: Besicovitch via Baire, Stud. Math. 158 (2003), no. 1, 65–78.*

To simplify the exposition we only consider the case $n = 2$, but the method can be extended to any $n \geq 3$. To apply the Baire category theorem, we need to work in a complete metric space. Let $\mathcal{K}$ denote the set of all non-empty compact subsets of $\mathbb{R}^2$. We need to define a metric on $\mathcal{K}$. For a point $x \in \mathbb{R}^n$ and a set $A \in \mathcal{K}$ we define

$$d(x, A) = \inf_{a \in A} \|a - x\|. \tag{6.46}$$

For $A, B \in \mathcal{K}$ we define

$$d(A,B) = \max(\sup_{a\in A} d(a,B), \sup_{b\in B} d(A,b)). \tag{6.47}$$

This is called *Hausdorff metric*.

EXERCISE 6.36. Show that d is a metric on $\mathcal{K}$ and that $(\mathcal{K}, d)$ is a complete metric space.

Consider the set of all $P \in \mathcal{K}$ with $P \subset [-1,1] \times [0,1]$ which are of the form

$$P = \bigcup_{i\in I} \ell_i, \tag{6.48}$$

where I is some index set and for each $i \in I$, there exist $x_1, x_2 \in [-1,1]$ such that ℓ_i is the line segment connecting the point $(x_1, 0)$ to the point $(x_2, 1)$. We define $\mathcal{P} \subset \mathcal{K}$ to be the set of all such P such that additionally for every $|v| \leq \frac{1}{2}$ there exist $x_1, x_2 \in [-1,1]$ such that $x_2 - x_1 = v$ and the line segment connecting $(x_1, 0)$ to $(x_2, 1)$ is contained in P.

This definition ensures that sets $P \in \mathcal{P}$ are "almost" Kakeya sets in the sense that while they do not contain a line segment in *every* direction, they *do* contain a line segment pointing in every direction that makes a sufficently small angle with the y-axis. We can always produce a true Kakeya set from such a P by taking a finite union of some rotated copies of P.

EXERCISE 6.37. Show that $\mathcal{P}$ is a closed subset of $\mathcal{K}$ (with respect to the Hausdorff metric d).

This implies in particular that $(\mathcal{P}, d|_{\mathcal{P}\times\mathcal{P}})$ is a complete metric space. Thus, we are done if we can show that there exists a set $P \in \mathcal{P}$ that has Lebesgue measure zero. We will actually show the following stronger result.

THEOREM 6.38 (Körner). *The set*

$$\mathcal{B} = \{P \in \mathcal{P} \,:\, P \text{ is a Lebesgue null set}\} \subset \mathcal{P} \tag{6.49}$$

is comeager.

The Baire category theorem says that comeager subsets of complete metric spaces are dense and in particular, non-empty.

To prove Theorem 6.38 it suffices to show that $\mathcal{B}$ contains a countable intersection of open dense sets.

Let $v \in [0,1]$ and $\varepsilon > 0$. Then we define $\mathcal{P}(v,\varepsilon) \subset \mathcal{P}$ to be the set of all $P \in \mathcal{P}$ such that there exist finitely many intervals $I_1, \ldots, I_N$ such that if $y \in [0,1] \cap [v-\varepsilon, v+\varepsilon]$, then

$$\{x \,:\, (x,y) \in P\} \subset \bigcup_{j=1}^{N} I_j \quad \text{and} \quad \sum_{j=1}^{N} |I_j| < 100\varepsilon. \tag{6.50}$$

LEMMA 6.39. *$\mathcal{P}(v,\varepsilon) \subset \mathcal{P}$ is open and dense.*

This is the main ingredient of the argument. Please refer to Lemma 2.4 in Körner's paper for the proof.

Now we show how this allows us to complete the proof of Theorem 6.38. Suppose that

$$P \in \bigcap_{n\in\mathbb{N}} \bigcap_{r=0}^{n} \mathcal{P}(\tfrac{r}{n}, \tfrac{1}{n}). \tag{6.51}$$

Then by definition of Lebesgue null sets and (6.50),

$$\{x : (x, y) \in P\} \subset [-1, 1] \tag{6.52}$$

is a Lebesgue null set for every $y \in [0, 1]$. This implies that P is a Lebesgue null set[3]. Therefore,

$$\bigcap_{n\in\mathbb{N}} \bigcap_{r=0}^{n} \mathcal{P}(\tfrac{r}{n}, \tfrac{1}{n}) \subset \mathcal{B}, \tag{6.53}$$

so $\mathcal{B}$ contains a countable intersection of open dense sets and is therefore comeager.

4.1. Box counting dimension. Consider a compact subset $A \subset \mathbb{R}^n$. Let $\delta \in (0, 1)$ and define $N_\delta(A)$ to be the minimum number of balls of radius δ required to cover the set A (it is clear that $N_\delta(A)$ is finite because $A \subset \mathbb{R}^n$ is totally bounded). We are interested in the rate of growth of the number $N_\delta(A)$ as δ tends to zero.

EXAMPLE 6.40. Let $k \leq n$ and let A denote a k-dimensional box in $\mathbb{R}^n$:

$$A = [0, 1]^k \times \{0\}^{n-k} = \{x \in \mathbb{R}^n : x_j \in [0, 1] \text{ for } 1 \leq j \leq k,\ x_j = 0 \text{ for } k < j \leq n\}. \tag{6.54}$$

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

$$c'\delta^{-k} \leq N_\delta(A) \leq c\delta^{-k} \tag{6.55}$$

for all $\delta \in (0, 1)$.

EXERCISE 6.41. Let $A \subset \mathbb{R}^n$ be a compact set. Show that there exists a constant $c \in (0, \infty)$ such that

$$N_\delta(x) \leq c \cdot \delta^{-n} \tag{6.56}$$

holds for all $\delta > 0$.

DEFINITION 6.42. Let $A \subset \mathbb{R}^n$ be a compact set. The *upper box counting dimension* of A is defined as

$$\overline{\dim}(A) = \limsup_{\delta\to 0} \frac{\log(N_\delta(A))}{\log(1/\delta)}. \tag{6.57}$$

Similarly, the *lower box counting dimension* of A is defined as

$$\underline{\dim}(A) = \liminf_{\delta\to 0} \frac{\log(N_\delta(A))}{\log(1/\delta)}. \tag{6.58}$$

We always have

$$0 \leq \underline{\dim}(A) \leq \overline{\dim}(A) \leq n. \tag{6.59}$$

The first two of these inequalities follow directly from the definitions and the third inequality follows from Exercise 6.41. Each of these inequalities may be strict.
If $\underline{\dim}(A) = \overline{\dim}(A) = d$, then we say that d is the *box counting dimension* (or *Minkowski dimension*) of A and write

$$\dim(A) = d. \tag{6.60}$$

[3]This is not obvious directly from the definitions. For the purpose of this discussion, we will take this implication for granted. It follows directly from properties of the Lebesgue integral, more precisely, Fubini's theorem. Intuitively, if every "horizontal slice" of a subset of the plane has zero length in $\mathbb{R}$, then that subset of the plane has zero area.

The numbers $\underline{\dim}(A), \overline{\dim}(A)$ do not depend on the norm on $\mathbb{R}^n$ used to form the balls that appear in the definition of $N_\delta(A)$ (because the number $N_\delta(A)$ only changes by a multiplicative constant when swapping out norms). The balls in the maximum norm on $\mathbb{R}^n$ defined by $\|x\|_\infty = \max_{i=1,\dots,n} |x_i|$ look like boxes. This motivates the term "box counting dimension".

This notion of dimension conincides with our intuition about dimension. For instance, the set A from Example 6.40 which we referred to as a "k-dimensional box" actually has box counting dimension k. Note that there is no reason why the box counting dimension of some given set A should always be an integer. In fact, there are lots of compact sets with a non-integer box counting dimension. We refer to such sets as *fractals* (because they have fractional dimension).

EXAMPLE 6.43. Maybe the simplest example of a fractal is the Cantor set $\mathfrak{C} \subset [0,1]$ (see (6.1)). In the iterative construction of the Cantor set, at the kth step we arrive at a disjoint union of 2^k closed intervals each of which has length 3^{-k}. Thus

$$N_{\frac{1}{2}\cdot 3^{-k}}(\mathfrak{C}) = 2^k. \tag{6.61}$$

Similarly, $N_\delta(\mathfrak{C}) \approx 2^k$ where $\delta \in (0,1)$ and k is such that $3^{-k-1} < \delta \leq 3^{-k}$. This shows that

$$0 < \dim(\mathfrak{C}) = \frac{\log(2)}{\log(3)} < 1. \tag{6.62}$$

Dimension is related the notion of Lebesgue null sets in the following way.

LEMMA 6.44. *If $A \subset \mathbb{R}^n$ is a compact set such that $\underline{\dim}(A) < n$, then A is a Lebesgue null set.*

PROOF. Let $\nu = \frac{1}{2}(n - \underline{\dim}(A)) > 0$. By assumption, there exists a sequence $(\delta_m)_m$ such that $\delta_m \to 0$ and for each m there exist open balls $(B(x_{m,j}, \delta_m))_{j=1,\dots,N_m}$ covering A, where $N_m \leq \delta_m^{-(n-\nu)}$. Then

$$\sum_{j=1}^{N_m} \delta_m^n = N_m \delta_m^n \leq \delta_m^\nu \longrightarrow 0 \text{ as } m \to \infty. \tag{6.63}$$

□

EXAMPLE 6.45. It is not true that a Lebesgue null set in $\mathbb{R}^n$ necessarily has box counting dimension n: take the set $A = \mathbb{Q} \cap [0,1] \subset \mathbb{R}$ (A is not compact, but it still makes sense to speak of its box counting dimension). It is not hard to show that $\dim(A) = 1$.

In view of this fact and the existence of Besicovitch sets, it is a natural instance of our original question about the smallest possible "size" of a Kakeya set to ask whether there exist Besicovitch sets in $\mathbb{R}^n$ that have a box counting dimension strictly smaller than n. It is conjectured that the answer is 'no' for all n.

Kakeya conjecture. Let $K \subset \mathbb{R}^n$ be a Kakeya set. Then $\dim(K) = n$.

This is known to hold if $n = 2$ (and trivial if $n = 1$), but still widely open if $n \geq 3$. See Exercise 6.53 below for a walkthrough to a simple proof that $\dim(K) \geq \frac{n+1}{2}$. Wolff (1995) proved that $\dim(K) \geq \frac{n+2}{2}$. The currently best known results are as follows:

- $n = 2$: $\dim(K) = 2$ (Davies 1971)
- $n = 3$: $\dim(K) \geq \frac{5}{2} + 10^{-10}$ (Katz-Łaba-Tao 1999)
- $n = 4$: $\dim(K) \geq 3 + 10^{-10}$ (Łaba-Tao 2000)
- $4 < n < 24$: $\dim(K) \geq (2 - \sqrt{2})(n - 4) + 3$ (Katz-Tao 2001)
- $n \geq 24$: $\dim(K) \geq n/\alpha + (\alpha - 1)/\alpha$, where $\alpha \in (1, 2)$ is such that $\alpha^3 - 4\alpha + 2 = 0$ (Katz-Tao 2001)

The Kakeya conjecture has many surprising connections to other open problems in mathematics, in particular Fourier analysis.

5. Further exercises

EXERCISE 6.46. We define the subset $A \subset \mathbb{R}$ as follows: $x \in A$ if and only if there exists $c > 0$ such that

$$|x - j2^{-k}| \geq c2^{-k} \tag{6.64}$$

holds for all $j \in \mathbb{Z}$ and integers $k \geq 0$. Show that A is meager and dense.

EXERCISE 6.47. Show that the set A from Exercise 6.46 is a Lebesgue null set.

EXERCISE 6.48. Let (X, d) be a complete metric space without isolated points. Prove that X cannot be countable.

EXERCISE 6.49. (i) Show that if X is a normed vector space and $U \subset X$ a proper subspace, then U has empty interior.
(ii) Let

$$X = \{P : \mathbb{R} \to \mathbb{R} \mid P \text{ is a polynomial}\}. \tag{6.65}$$

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 6.50. Consider $X = C([-1, 1])$ with the usual norm $\|f\|_\infty = \sup_{t\in[-1,1]} |f(t)|$. Let

$$A_+ = \{f \in X \,:\, f(t) = f(-t) \quad \forall t \in [-1, 1]\}, \tag{6.66}$$

$$A_- = \{f \in X \,:\, f(t) = -f(-t) \quad \forall t \in [-1, 1]\}. \tag{6.67}$$

(i) Show that A_+ and A_- are meager.
(ii) Is $A_+ + A_- = \{f + g \,:\, f \in A_+, g \in A_-\}$ meager?

EXERCISE 6.51. 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 6.52. 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 6.53. Let $0 < \delta \ll 1$ (say $\delta < \frac{1}{10}$) and $n \geq 2$. A *δ-tube* is a rectangular box in $\mathbb{R}^n$ of dimensions $1 \times \delta \times \cdots \times \delta$. We call a collection of δ-tubes *δ-separated* if every two distinct tubes make an angle of at least δ. Let $K \subset \mathbb{R}^n$ be a Kakeya set and denote by $K(\delta)$ its δ-neighborhood:

$$K(\delta) = \{x \in \mathbb{R}^n \,:\, \operatorname{dist}(x, K) \leq \delta\}. \tag{6.68}$$

Then $K(\delta)$ must contain a δ-tube in every direction. Let $\mathcal{T}_\delta$ denote a maximal δ-separated collection of δ-tubes contained in $K(\delta)$ (then $\mathcal{T}_\delta$ contains roughly δ^{1-n} many δ-tubes). If $A \subset \mathbb{R}^n$ is a finite union of δ-tubes, then we denote by $\mathrm{vol}(A)$ the volume of A.

(i) Prove that there must exist a point $x \in K(\delta)$ such that the number of tubes $T \in \mathcal{T}_\delta$ such that $x \in T$ is at least $c/\mathrm{vol}(\cup \mathcal{T}_\delta)$, where $c > 0$ is a constant depending only on the dimension, n.

(ii) Conclude from (i) that there exists $c > 0$ such that for every $\delta \in (0, \frac{1}{10})$:

$$\mathrm{vol}(\cup \mathcal{T}_\delta) \geq c \cdot \delta^{\frac{n-1}{2}} \tag{6.69}$$

(iii) Conclude from (ii) that

$$\underline{\dim}(K) \geq \tfrac{n+1}{2}. \tag{6.70}$$

(iv) Suppose that for every $\varepsilon > 0$ there exists $c_\varepsilon > 0$ such that for every $\delta \in (0, \frac{1}{10})$ we have

$$\mathrm{vol}(\cup \mathcal{T}_\delta) \geq c_\varepsilon \delta^\varepsilon. \tag{6.71}$$

Show that this would imply the Kakeya conjecture: $\dim(K) = n$.

(Here $\cup \mathcal{T}_\delta = \bigcup_{T \in \mathcal{T}_\delta} T$.)