Lecture Notes
Infinite Dimensional Optimization

Roland Herzog*

2024-11-17

*Interdisciplinary Center for Scientific Computing, Heidelberg University, 69120 Heidelberg, Germany (roland.herzog@iwr.uni-heidelberg.de, https://scoop.iwr.uni-heidelberg.de/team/roland-herzog).

Material for approximately 25 lectures, 14 weeks.

In these lecture notes we use colored markup for **definitions** and alerts.

Expert Knowledge: topic

A block like this contains further information that are not subject to examination.

Contents

Chapter 0 Introduction

We will consider in this class optimization problems of the following kind:

$$\begin{aligned} \text{Minimize} \quad & f(x), \quad \text{where } x \in X \\ \text{subject to} \quad & h(x) = 0. \end{aligned}$$

In this problem, $f: X \to \mathbb{R}$ is called the **objective function** and $h: X \to Y$ is the **equality constraint**. The **optimization variable** x is sought in some **optimization space** X.

Inequality constraints may be added to the above problem, either

- explicitly in the form $g(x) \leqslant 0$ or, more generally, in the form $g(x) \in K \subseteq Z$,
- or implicitly, by imposing $x \in C \subseteq X$ or allowing f to take values in $\mathbb{R} \cup \{\infty\}$.

Often, K is a cone and C is a convex set.

What are reasonable choices for the "spaces" X, Y, Z?

(1) To define the notion of global minimizers, no structure at all is required, so X, Y, Z can be general sets.

(2) To define the notion of local minimizers, the space X of optimization variables must carry a topology since we require the concept of neighborhoods.

(3) Statements about the existence of global minimizers build on notions of continuity and compactness.[1] Therefore, topological spaces are required for this purpose as well.

(4) To formulate first-order optimality conditions, we need to be able to differentiate. A convenient setting for this are normed linear spaces.

(5) For algorithmic purposes, derivatives need to be converted into directions, e. g., directions of largest/smallest directional derivatives over the unit sphere. For this purpose, normed linear spaces or even Hilbert spaces, are convenient.

Based on these considerations, we will consider only **normed linear spaces** over the field of real numbers $\mathbb{R}$ (§ 2).[2]

We may anticipate a couple of differences compared to optimization over finite-dimensional linear spaces, as well as a number of questions that we will may want to to answer throughout the course:

(1) Different norms on an infinite-dimensional linear space are, in general, not equivalent to each other.

(2) How do we differentiate functions defined on infinite-dimensional normed linear space?

[1] Compare, for instance, the Weierstrass extreme value theorem: a continuous function $f: X \to \mathbb{R}$ attains its minimum (and its maximum) on a compact set $K \subseteq X$; see also Theorem 5.1.

[2] We use the term **linear space** instead of the synonymous **vector space**.

(3) Can we formulate optimization algorithms on infinite-dimensional spaces?

(4) If so, then when and how do we discretize in order to realize them numerically?

§ 1 Motivating Examples

Example 1.1 (Brachistochrone problem)**.**
In a 1696 article, Johann Bernoulli posted the following problem:

> Given two points A and B in a vertical plane, what is the curve traced out by a point acted on only by gravity, which starts at A and reaches B in the shortest time?

This problem is known as the **Brachistochrone problem** (ancient Greek: *βράχιστος χρόνος*). In modern terms, it can be formulated as follows. Suppose that the points have coordinates $A = (0, 0)$ and $B = (b_1, b_2)$ with $b_2 \geqslant 0$. Let $g > 0$ denote the gravitational constant.

We are seeking a function $y\colon [0, b_1] \to \mathbb{R}$ whose graph defines the curve from A to B. Using the principle of conservation of (potential plus kinetic) energy, we may express the speed of the particle at horizontal position x in terms of its height $y(x)$. Skipping the details, this eventually leads to the following optimization problem:

$$\begin{aligned} \text{Minimize} \quad & f(y) \coloneqq \int_0^{b_1} \frac{\sqrt{1 + y'(x)^2}}{\sqrt{2\, g\, y(x)}} \,\mathrm{d}x, \quad \text{where } y \in X \\ \text{s.t.} \quad & y(0) = 0 \\ \text{and} \quad & y(b_1) = b_2 \\ \text{as well as} \quad & y \geqslant 0 \text{ on } [0, b_1]. \end{aligned} \tag{1.1}$$

Here X is a suitable vector space of functions $y\colon [0, b_1] \to \mathbb{R}$, e. g., $X = C^1(0, b_1) \cap C([0, b_1])$, the space of continuous functions on $[0, b_1]$ whose restriction to the open interval $(0, b_1)$ is continuously differentiable. An alternative is the **Sobolev space** $X = H^1(0, b_1)$ of square integrable functions with square integrable weak derivative on $(0, b_1)$.[3]

(**Quiz 1.1:** Does the gravitational constant impact optimal curves?) One can show that the (unique) minimizer of (1.1) satisfies a first-order necessary optimality condition, which comes in the form of a differential equation:

$$\frac{1}{2} \sqrt{\frac{1 + y'(x)^2}{y(x)^3}} + \frac{\mathrm{d}}{\mathrm{d}x} \frac{y'(x)}{\sqrt{y(x) \left(1 + y'(x)^2\right)}} = 0.$$

The solutions of this equation satisfy

$$y(x) \left(1 + y'(x)^2\right) = C \quad \text{in } (0, b_1) \tag{1.2}$$

for some $C > 0$, and it has infinite slope initially:

$$\lim_{x \searrow 0} y'(x) = \infty.$$

[3] We will introduce Sobolev spaces later; see § 2.6.

The unique solution is given by the curve

$$t \mapsto \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = C \begin{pmatrix} t - \sin(t) \\ 1 - \cos(t) \end{pmatrix} \quad \text{for } t \in [0, T], \tag{1.3}$$

where $C > 0$ and $T \in (0, 2\pi]$ are determined by the conditions $x(T) = b_1$ and $y(T) = b_2$.

This curve is a segment of a cycloid with radius C. △

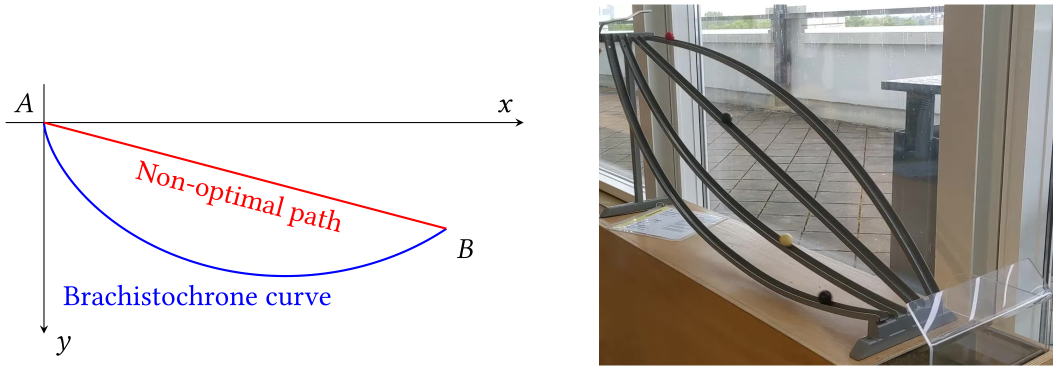

Figure 1.1: Some non-optimal curve $\gamma\colon [0, b_1] \to \mathbb{R}$ from A to B (left) as well as the unique global minimizer of the Brachistochrone problem (1.1), given by the segment of a cycloid (left). Image of an experimental device on display at Technoseum Mannheim (right), shot by Roland Herzog.

Remark 1.2 (on the Brachistochrone problem).
The first-order optimality condition of the Brachistochrone problem come in the form of a differential equation (1.2). This is typical for optimization problems whose variables are functions and whose objectives involve derivatives of those functions. As a result, minimizers may be more regular than suggested by the optimization space X. This is indeed the case in the Brachistochrone problem (1.1), where the unique minimizer turns out to be a $C^\infty(0, b_1)$-function. △

Expert Knowledge: The origins of the calculus of variations

The Brachistochrone problem belongs to a class of problems referred to as calculus of variations, where optimization variables are functions and objectives are typically integrals involving values of the function and its derivative(s). This term was coined in 1766 by Leonhard Euler. The first-order optimality conditions for calculus of variations problems are referred to as **Euler-Lagrange equations**.

Newton's problem of minimal resistance from 1687 is considered the first problem of this type, and the Brachistochrone problem (1696) is second. That problem attracted the attention of Johann Bernoulli's brother Jakob, as well as of Isaac Newton, Gottfried Leibniz, Ehrenfried Walther von Tschirnhaus and Guillaume de l'Hôpital, who all turned in solutions.

Example 1.3 (Fermat's principle in optics).

Suppose that $n\colon \mathbb{R}^2 \to \mathbb{R}_{>0}$ is the material dependent refractive index of an optical material. Let $\gamma\colon [0, b_1] \to \mathbb{R}$ denote a function whose graph defines a curve through this material. Then the optical length of this curve is defined by

$$\int_0^{b_1} n(x, \gamma(x)) \sqrt{1 + \gamma'(x)^2} \, \mathrm{d}x.$$

Fermat's principle stipulates that the path a ray of light will take minimizes the optical length. Suppose that the end points of that path are $A = (0, 0)$ and $B = (b_1, b_2)$. Then we obtain the following optimization problem:

$$\begin{aligned} \text{Minimize} \quad & f(\gamma) \coloneqq \int_0^{b_1} n(x, \gamma(x)) \sqrt{1 + \gamma'(x)^2} \, \mathrm{d}x, \quad \text{where } \gamma \in X \\ \text{s. t.} \quad & \gamma(0) = 0 \\ \text{and} \quad & \gamma(b_1) = b_2. \end{aligned} \tag{1.4}$$

In the particular case where the refractive index is piecewise constant on slabs, the unique global minimizer of (1.4) satisfies **Snell's law**, which states that the incident angles θ_+, θ_- (measured against the normal) of two neighboring slabs satisfy the relation $n_+ \sin(\theta_+) = n_- \sin(\theta_-)$, see Figure 1.2.

Similar as in Example 1.1, every minimizer satisfies a first-order optimality condition that amounts to a differential equation:

$$-\frac{n(x, \gamma(x)) \, \gamma'(x)}{\sqrt{1 + \gamma'(x)^2}} + n_y(x, \gamma(x)) \sqrt{1 + \gamma'(x)^2} = 0.$$

In this case, however, the discontinuous coefficient n may limit the regularity of an optimal path. Again, for piecewise constant refractive index, an optimal curve will be piecewise linear with discontinuous derivative at optical interfaces. △

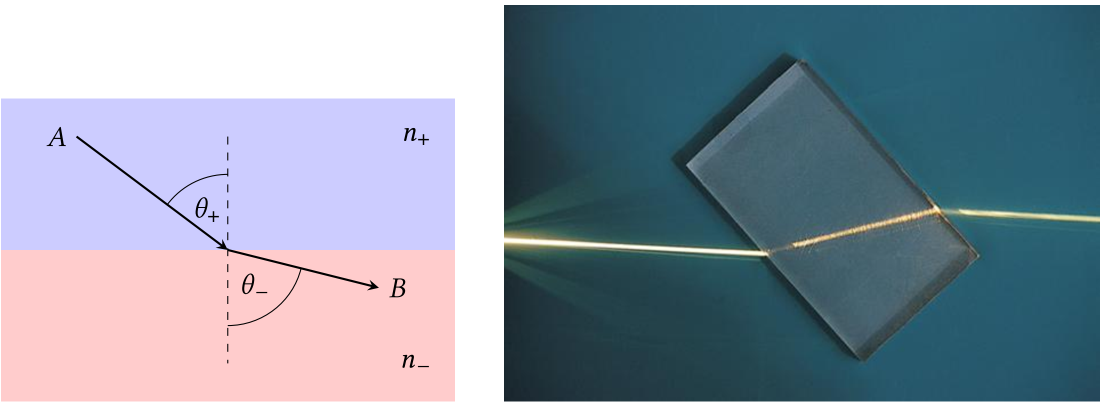

Figure 1.2: Illustratrion of Snell's law of refraction (left) as a special case of Example 1.3. Image (right) obtained from `https://en.wikipedia.org/wiki/Refraction`, released into the public domain by creator ajizai.

End of Class 1

Example 1.4 (signal denoising).

Suppose a signal $s\colon [0,T] \to \mathbb{R}$ is given.[4] In case the signal is noisy, we may formulate an optimization problem to try and find a denoised signal $y\colon [0,T] \to \mathbb{R}$:

$$\text{Minimize} \quad f(y) \coloneqq \int_0^T |y(t) - s(t)|^2 \, \mathrm{d}t + \beta \int_0^T |\dot{y}(t)|^2 \, \mathrm{d}t, \quad \text{where } y \in X. \tag{1.5}$$

The dot denotes the time derivative. A suitable function space for this problem is the Sobolev space $X = H^1(0,T)$.

The second term in the objective penalizes "fast variations" in the signal. The parameter $\beta > 0$ balances the two summands in the objective and thus determines the degree of denoising.

We will be able to show later that the first-order optimality conditions for (1.5) involve the second-order differential equation

$$-\beta\, \ddot{y}(t) + y(t) = s(t), \tag{1.6}$$

which shows that the minimizer will indeed be a smoothed version of the noisy signal s. More precisely, we can expect the solution to gain two orders of differentiation compared to the data s. In particular, the solution will not admit any discontinuities. Therefore, one often prefers a "less powerful" regularization term, such as the **total variation** of the function y. We will come back to this type of problem in the context of image denoising problems in Chapter 1. △

Example 1.5 (crane trolley optimal control problem)**.**
Consider a load on rope of length ℓ hanging from a crane trolley system (Figure 1.3). We denote the position of the trolley relative to the origin by s. The position of the load relative to the trolley is denoted by z. The trolley has mass M and the load has mass m. A controllable force u acts on the trolley.

This system is described by a second-order differential equation for the positions (s,z). It can be derived by working out Newton's law, force equals mass times acceleration. We convert it here to a first-order system of differential equations in terms of $x = (s, \dot{s}, z, \dot{z})$, where the dot denotes the time derivative. Assuming small angles θ, the differential equations can be taken as linear and the system reads

$$\begin{pmatrix} \dot{s} \\ \ddot{s} \\ \dot{z} \\ \ddot{z} \end{pmatrix} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -\frac{m}{M}\frac{g}{\ell} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -\frac{m+M}{M}\frac{g}{\ell} & 0 \end{bmatrix}}_{=:A} \begin{pmatrix} s \\ \dot{s} \\ z \\ \dot{z} \end{pmatrix} + \underbrace{\begin{bmatrix} 0 \\ \frac{1}{M} \\ 0 \\ \frac{1}{M} \end{bmatrix}}_{=:B} u \tag{1.7}$$

or, in short, $\dot{x} = A\,x + B\,u$. Notice that we have omitted the (t) argument everywhere for brevity.

We wish to steer the system from an initial state $x(0) = (0,0,0,0)^\intercal$ to a terminal state $x(T) = (E,0,0,0)^\intercal$ in as short a time T as possible. This leads us to the preliminary optimization problem

$$\begin{aligned} \text{Minimize} \quad & \int_0^T 1 \, \mathrm{d}t, \quad \text{where } (u,x,T) \in U \times X \times \mathbb{R} \\ \text{s.t.} \quad & \dot{x} = A\,x + B\,u \quad \text{in } [0,T] \\ \text{and} \quad & x(0) = (0,0,0,0)^\intercal \\ \text{and} \quad & x(T) = (E,0,0,0)^\intercal \\ \text{as well as} \quad & T > 0. \end{aligned} \tag{1.8}$$

[4]Think, for instance, of an audio signal sampled with a certain frequency, say, 48 kHz into a piecewise constant function.

This preliminary problem formulation has some issues. Due to the terminal time T being an optimization variable, we cannot fix function spaces for the **control** u and the **state** x since they depend on T.

There is, however, an easy remedy to this. We can renormalize the unknown time interval $[0, T]$ to the fixed interval $[0, 1]$. Replacing the unknowns x and u by their counterparts on the fixed interval, the dynamics need to be rescaled and the problem becomes

$$\begin{aligned} \text{Minimize} \quad & \int_0^1 T \, \mathrm{d}t, \quad \text{where } (u, x, T) \in U \times X \times \mathbb{R} \\ \text{s.t.} \quad & \dot{x} = \frac{1}{T}(A\,x + B\,u) \quad \text{in } [0, 1] \\ \text{and} \quad & x(0) = (0, 0, 0, 0)^\intercal \\ \text{and} \quad & x(1) = (E, 0, 0, 0)^\intercal \\ \text{as well as} \quad & T > 0. \end{aligned} \tag{1.9}$$

We can now fix suitable function spaces[5], e. g., $U = L^2(0, 1)$ and $X = H^1(0, 1)^4$. A problem such as (1.9), in which a **state** function x depends on the choice of the **control** function u through a differential equation, is termed an **optimal control problem**. We will see more of these in Chapter 2.

Unfortunately, problem (1.9) as stated will not have a solution. (**Quiz 1.2:** Can you see why?) We may fix this by imposing bounds on the control function, e. g., by adding the pointwise inequality constraints

$$u(t) \in [-u_{\max}, u_{\max}],$$

with some $u_{\max} > 0$ to problem (1.9), or by adding a cost term such as

$$\beta \int_0^1 |u(t)| \, \mathrm{d}t$$

to the objective. △

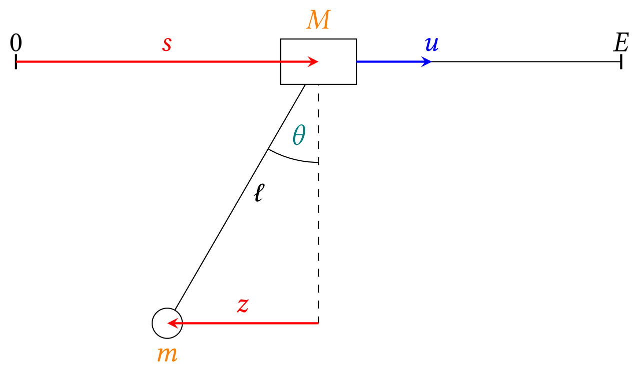

Figure 1.3: Illustration of the crane trolley problem (Example 1.5).

[5] Again, we will introduce these Lebesgue and Sobolev spaces later; see §§ 2.5 and 2.6.

§ 2 Normed Linear Spaces

In this section we recap the notion of a normed linear space. We will also introduce Lebesgue and Sobolev spaces as our prime examples of normed linear spaces.

Definition 2.1 (linear space)**.**
An algebraic structure $(V, +, \cdot)$ with two operations[6]

$$\begin{aligned} +&: V \times V \to V \quad &\text{(addition)} \\ \cdot&: \mathbb{R} \times V \to V \quad &\text{(S-multiplication)} \end{aligned}$$

is said to be a **linear space** over the field of real numbers $\mathbb{R}$ if

(*i*) $(V, +)$ is an Abelian group.

(*ii*) The S-multiplication satisfies the mixed distributive laws

$$\begin{aligned} \alpha\,(u + v) &= (\alpha\, u) + (\alpha\, v) \\ (\alpha + \beta)\, v &= (\alpha\, v) + (\beta\, v) \end{aligned}$$

as well as the mixed associative law

$$(\alpha\, \beta)\, v = \alpha\, (\beta\, v)$$

for all $\alpha, \beta \in \mathbb{R}$ and $u, v \in V$. Moreover, the neutral element $1 \in \mathbb{R}$ w.r.t. multiplication in $\mathbb{R}$ is also neutral w.r.t. S-multiplication:

$$1\, v = v. \qquad \triangle$$

All linear spaces will be over the field of real numbers $\mathbb{R}$ and we will not explicitly mention that. We already anticipated that in order to be able to differentiate functions $f: V \to \mathbb{R}$ or, more generally, $f: V \to W$, we will require linear spaces to be normed.

Definition 2.2 (normed linear space)**.**
Suppose that V is a linear space.

(*i*) A map $\|\cdot\|: V \to \mathbb{R}$ is said to be a **norm on V** if the following conditions hold:

$$\|u\| \geqslant 0, \quad \text{and } \|u\| = 0 \;\Rightarrow\; u = 0 \qquad \textbf{positive definiteness} \tag{2.1a}$$

$$\|\alpha\, u\| = |\alpha|\, \|u\| \qquad \textbf{absolute homogeneity} \tag{2.1b}$$

$$\|u + v\| \leqslant \|u\| + \|v\| \qquad \textbf{triangle inequality} \text{ or } \textbf{subadditivity} \tag{2.1c}$$

for all $u, v \in V$ and all $\alpha \in \mathbb{R}$.

(*ii*) The pair $(V, \|\cdot\|)$ is said to be a **(real) normed linear space** or **normed vector space**. $\triangle$

[6] The dot $\cdot$ for S-multiplication is usually not written, just as the multiplication symbol in $\mathbb{R}$ is usually not written.

Expert Knowledge: from topological to normed linear spaces

We have the inclusions

- Every normed linear space is a metric space.
- Every metric space is a topological space.

A topological space is defined by a collection of its subsets that are called the open sets. Topological spaces admit notions of convergence and limits, closure and compactness of sets, as well as notions of continuity of functions.

Metric spaces are spaces with a notion of distance. The metric induces a topology.

Normed spaces are spaces with a notion of length. The norm induces a metric.

We will not discuss general topological spaces in full generality but restrict ourselves to normed linear spaces.

§ 2.1 Open and Closed Sets

Definition 2.3 (balls, spheres, open sets, closed sets)**.**
Suppose that $(V, \|\cdot\|)$ is a normed linear space.

(i) For $\varepsilon > 0$, the set
$$B_\varepsilon(x) \coloneqq \{y \in V \mid \|y - x\| < \varepsilon\}$$
is said to be the **open ε-ball** about x of radius ε. In particular, $B_1(0)$ is termed the **open unit ball**.

(ii) A point $x \in E$ of a subset $E \subseteq V$ is said to be an **interior point** of E if there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq E$. The subset of interior points of E is called the **interior** of E and it is denoted by $\operatorname{int} E$.

(iii) A set $U \subseteq V$ is said to be **open** if every $x \in U$ is an interior point of U, i. e., if $\operatorname{int} U = U$.

(iv) A set $A \subseteq V$ is said to be **closed** if its complement $V \setminus A$ is open.

(v) For $\varepsilon > 0$, the set
$$\overline{B_\varepsilon(x)} \coloneqq \{y \in V \mid \|y - x\| \leqslant \varepsilon\}$$
is said to be the **closed ε-ball** about x of radius ε. In particular, $\overline{B_1(0)}$ is termed the **closed unit ball**.

(vi) The **closure** of a subset $E \subseteq V$ is
$$\operatorname{cl} E \coloneqq \bigcap \big\{A \subseteq V \,\big|\, A \text{ is closed and } E \subseteq A\big\}. \tag{2.2}$$

(vii) The **boundary** of a subset $E \subseteq V$ is $\partial E \coloneqq \operatorname{cl} E \setminus \operatorname{int} E$, i. e., the closure minus the interior of E.

$(viii)$ The set
$$\partial B_\varepsilon(x) \coloneqq \{y \in V \mid \|y - x\| = \varepsilon\}$$
is said to be the **ε-sphere** about x of radius ε. In particular, $\partial B_1(0)$ is termed the **unit sphere** of V. △

It is not difficult to show that the interior of a set is open and the closure of a set is closed. In fact, a set E is open if and only if $E = \operatorname{int} E$, and a set A is closed if and only if $A = \operatorname{cl} A$. Also, a set A is closed if and only if $A = \partial A$. The boundary of a set is also closed. (**Quiz 2.1:** Can you show this?)

The following result was inserted after the class.

Lemma 2.4 (characterization of the closure[7]).

Suppose that $(V, \|\cdot\|)$ is a normed linear space and $E \subseteq V$. Then

$$\begin{aligned} \operatorname{cl} E &= \{y \in V \mid \text{for any } \varepsilon > 0 \text{ there exists } x \in E \text{ such that } \|x - y\| < \varepsilon\} \\ &= \{y \in V \mid \text{for any } \varepsilon > 0, B_\varepsilon(y) \cap E \neq \emptyset\} \\ &= \{y \in V \mid \text{there exists a sequence } \big(x^{(k)}\big) \text{ in } E \text{ converging to } y\}. \end{aligned} \tag{2.3}$$

Proof. □

The following lemma **(inserted after the class)** confirms that the nomenclature and symbols related to balls and spheres is meaningful:

Lemma 2.5 (openness, closedness, boundary of balls and spheres).

Suppose that $(V, \|\cdot\|)$ is a normed linear space.

(i) Open balls $B_\varepsilon(x)$ are open sets.

(ii) Closed balls $\overline{B_\varepsilon(x)}$ are closed sets.

(iii) Open balls and closed balls are related via

$$\overline{B_\varepsilon(x)} = \operatorname{cl} B_\varepsilon(x) \quad \text{and} \quad B_\varepsilon(x) = \operatorname{int} \overline{B_\varepsilon(x)}. \tag{2.4}$$

(iv) Spheres and balls are related via

$$\partial B_\varepsilon(x) = \partial\big(B_\varepsilon(x)\big) = \partial\big(\overline{B_\varepsilon(x)}\big). \tag{2.5}$$

Proof. □

End of Class 2

End of Week 1

[7] We can read this result as "The closure of a set E consists of the **accumulation points** of E."

§ 2.2 Banach Spaces

Since norms furnish a linear space with a topology, they also bring about a notion of convergence.

Definition 2.6 (convergent sequence, Cauchy sequence)**.**
Suppose that $(V, \|\cdot\|)$ is a normed linear space.

(i) A sequence[8] $\big(x^{(k)}\big)$ in V is said to **converge to** $x \in V$ in case $\|x^{(k)} - x\| \to 0$ in $\mathbb{R}$. We then write $x^{(k)} \to x$ or $\lim_{k\to\infty} x^{(k)} = x$ and call x a **limit point** or **limit** of the sequence $\big(x^{(k)}\big)$.

In other words, $x^{(k)} \to x$ means: for every $\varepsilon > 0$ there exists an index k_ε such that $\|x^{(k)} - x\| < \varepsilon$ holds for all $k \geqslant k_\varepsilon$.

(ii) A sequence $\big(x^{(k)}\big)$ in V is said to **converge** if there exists some $x \in V$ such that $x^{(k)} \to x$.

(iii) A sequence $\big(x^{(k)}\big)$ in V is said to be a **Cauchy sequence** in V if, for every $\varepsilon > 0$, there exists an index k_ε such that $\|x^{(k)} - x^{(\ell)}\| < \varepsilon$ holds for all $k, \ell \geqslant k_\varepsilon$. △

Lemma 2.7 (properties of convergent sequences)**.**
Suppose that $(V, \|\cdot\|)$ is a normed linear space and that $\big(x^{(k)}\big)$ is a sequence in V.

(i) Suppose that $\big(x^{(k)}\big)$ converges. Then its limit is unique.

(ii) Suppose that $\big(x^{(k)}\big)$ converges. Then it is a Cauchy sequence.

Proof. This proof is addressed in homework problem 2.3. □

The converse of statement (ii) is not true in general. Therefore, spaces in which it is true deserve special mention:

Definition 2.8 (complete normed linear space, Banach space, complete subset)**.**
Suppose that $(V, \|\cdot\|)$ is a normed linear space.

(i) The space $(V, \|\cdot\|)$ is said to be **complete** or a **Banach space** if every Cauchy sequence in V converges.

(ii) A subset $A \subseteq V$ is said to be **complete** if every Cauchy sequence in A converges to a limit in A. △

The following result was inserted after the class.

Lemma 2.9 (in Banach spaces, completeness is closedness)**.**
Suppose that $(V, \|\cdot\|)$ is a Banach space. The $A \subseteq V$ is complete if and only if A is closed.

Proof. This proof is addressed in homework problem 2.2. □

The following result was inserted after the class.

[8] The exact index set of a sequence does not matter. We will allow any interval of the integers $\mathbb{Z}$ which is bounded below but not bounded above. In other words, any subset of $\mathbb{Z}$ of the form $\{k_0, k_0 + 1, k_0 + 2, \ldots\}$.

Lemma 2.10 (complete sets are closed).
Suppose that $(V, \|\cdot\|)$ is a normed linear space and $E \subseteq V$. If E is complete, then E is closed.

Proof. Suppose that $\big(x^{(k)}\big)$ is a sequence in E converging to some $x \in V$. Then this sequence is a Cauchy sequence in E. Since E is complete, $\big(x^{(k)}\big)$ converges to a limit $y \in E$. By uniqueness of the limit, we have $x = y \in E$. By the characterization (2.3) of the closure, we have $E = \operatorname{cl} E$. □

§ 2.3 Comparison of Norms

We wish to be able to compare two different norms on the same linear space. The following definition allows us to do that.

Definition 2.11 (partial ordering of norms).
Suppose that V is a linear space and that $\|\cdot\|_a$ and $\|\cdot\|_b$ are two norms on V.

(*i*) The norm $\|\cdot\|_a$ is said to be **weaker** than the norm $\|\cdot\|_b$ if there exists a constant $c > 0$ such that

$$\|x\|_a \leqslant c\,\|x\|_b \quad \text{holds for all } x \in V. \tag{2.6}$$

In this case, we also say that $\|\cdot\|_b$ is **stronger** than $\|\cdot\|_a$. We write $\|\cdot\|_a \lesssim \|\cdot\|_b$ or $\|\cdot\|_b \gtrsim \|\cdot\|_a$.

(*ii*) The norms $\|\cdot\|_a$ and $\|\cdot\|_b$ are said to be **equivalent** if both $\|\cdot\|_a \lesssim \|\cdot\|_b$ and $\|\cdot\|_b \lesssim \|\cdot\|_a$ hold, i. e., if there exist constants $c_1, c_2 > 0$ such that

$$c_1\,\|x\|_a \leqslant \|x\|_b \leqslant c_2\,\|x\|_a \quad \text{holds for all } x \in V. \tag{2.7}$$

△

The following result was corrected.[RH]

Lemma 2.12 (openness, closedness, completeness and the Cauchy property are preserved under weaker norms).
Suppose that V is a linear space and that $\|\cdot\|_a$ and $\|\cdot\|_b$ are two norms on V such that $\|\cdot\|_a \lesssim \|\cdot\|_b$. Then the following hold:

(*i*) For any open ball $B_\varepsilon^{\|\cdot\|_a}(x)$ in the weaker norm $\|\cdot\|_a$, there exists an open ball $B_\delta^{\|\cdot\|_b}(x)$ in the stronger norm $\|\cdot\|_b$ such that $B_\delta^{\|\cdot\|_b}(x) \subseteq B_\varepsilon^{\|\cdot\|_a}(x)$.

(The stronger norm has the smaller balls and more open sets.)

(*ii*) If $U \subseteq V$ is open in the weaker norm $\|\cdot\|_a$, then U is open in the stronger norm $\|\cdot\|_b$.

(The stronger norm defines the finer topology.)

(*iii*) If $A \subseteq V$ is closed in the weaker norm $\|\cdot\|_a$, then A is closed in the stronger norm $\|\cdot\|_b$.

(*iv*) If $E \subseteq V$ is bounded in the stronger norm $\|\cdot\|_b$, then E is bounded in the weaker norm $\|\cdot\|_a$.

(*v*) If $K \subseteq V$ is totally bounded in the stronger norm $\|\cdot\|_b$, then K is totally bounded in the weaker norm $\|\cdot\|_a$.

(*vi*) If $K \subseteq V$ is compact in the stronger norm $\|\cdot\|_b$, then K is compact in the weaker norm $\|\cdot\|_a$.

(*vii*) If $\big(x^{(k)}\big)$ converges in the stronger norm $\|\cdot\|_b$, then $\big(x^{(k)}\big)$ converges in the weaker norm $\|\cdot\|_a$ (to the same limit point).

(*viii*) If $\big(x^{(k)}\big)$ is a Cauchy sequence in the stronger norm $\|\cdot\|_b$, then $\big(x^{(k)}\big)$ is a Cauchy sequence in the weaker norm $\|\cdot\|_a$.

Proof. This proof is addressed in homework problem 3.1. □

Theorem 2.13 (in finite-dimensional normed linear spaces, all norms are equivalent)**.**
Suppose that V is a finite-dimensional linear space. If $\|\cdot\|_a$ and $\|\cdot\|_b$ are two norms on V, then $\|\cdot\|_a$ and $\|\cdot\|_b$ are equivalent.

Proof. Suppose that $\{v^{(1)}, \dots, v^{(n)}\}$ is a basis of V. Then every $x \in V$ can be uniquely written as $x = \sum_{j=1}^n x_j \, v^{(j)}$. The map $x \mapsto \|x\|_\infty := \left\| \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right\|_\infty = \max\big\{|x_1|, \dots, |x_n|\big\}$ is a norm on V.

It is enough to prove that the norms $\|\cdot\|_a$ and $\|\cdot\|_\infty$ are equivalent norms on V since equivalence of norms is an equivalence relation.

Step 1: To show $\|\cdot\|_a \lesssim \|\cdot\|_\infty$, we estimate:

$$\begin{aligned} \|x\|_a &= \Big\| \sum_{j=1}^n x_j \, v^{(j)} \Big\|_a \\ &\leqslant \sum_{j=1}^n |x_j| \, \|v^{(j)}\|_a \\ &\leqslant \|x\|_\infty \sum_{j=1}^n \|v^{(j)}\|_a \\ &=: c \, \|x\|_\infty. \end{aligned}$$

Step 2: We show that $\|\cdot\|_\infty \lesssim \|\cdot\|_a$.

Suppose that this is not the case. Then there exists a sequence $\big(x^{(k)}\big)$ in V such that $\|x^{(k)}\|_\infty > k \, \|x^{(k)}\|_a$. We can assume that $\|x^{(k)}\|_\infty = 1$ holds. (**Quiz 2.2:** Why?)

On the other hand, for all $j = 1, \dots, n$, the j-th coefficients $\big\{x_j^{(k)} \,\big|\, k \in \mathbb{N}\big\}$ belong to the compact interval $[-1, 1]$. Therefore, we can find a subsequence $x^{(k^{(\ell)})}$ such that $x_j^{(k^{(\ell)})}$ converges to some x_j^* for all $j = 1, \dots, n$. Moreover, for at least one index $j_0 \in \{1, \dots, n\}$, we have $\big|x_{j_0}^{(k^{(\ell)})}\big| = 1$ for infinitely many indices $\ell \in \mathbb{N}$. We pass to this subsequence without re-labeling it. This shows $|x_{j_0}^*| = 1$ by continuity of the absolute value function.

We define $x^* := \sum_{j=1}^n x_j^* \, v^{(j)}$. The estimate

$$\begin{aligned} \|x^*\|_a &\leqslant \|x^* - k^{(\ell)}\|_a + \|k^{(\ell)}\|_a \\ &\leqslant c \, \|x^* - k^{(\ell)}\|_\infty + \frac{1}{k^{(\ell)}} \qquad \text{by step 1} \\ &\longrightarrow 0 + 0 \quad \text{as } \ell \longrightarrow \infty \end{aligned}$$

shows $x^* = 0$, i. e., all coefficients x_j^* are zero. This contradicts $|x_{j_0}^*| = 1$. □

Note: As a consequence of this theorem, we do not necessarily need to specify the norm when we talk about a finite-dimensional linear space. In particular, all norms on $\mathbb{R}$ are equivalent, with the absolute value $|\cdot|$ as the standard norm.

As a consequence of Theorem 2.13, we can show:

Lemma 2.14 (finite-dimensional subspaces are complete and thus closed).
Suppose that $(V, \|\cdot\|)$ is a normed linear space. Every finite-dimensional subspace $Y \subseteq V$ is complete and thus closed.

Proof. Suppose that $\{y^{(1)}, \ldots, y^{(n)}\}$ is a basis of Y. By Theorem 2.13, the norms $\|\cdot\|$ and $\|\cdot\|_\infty$ are equivalent on Y, where $\|x\|_\infty \coloneqq \max\{|x_1|, \ldots, |x_n|\}$ when $x = \sum_{j=1}^{n} x_j y^{(j)}$.

Suppose now that $(x^{(k)})$ is a Cauchy sequence in Y. The elements of $(x^{(k)})$ have a representation

$$x^{(k)} = \sum_{j=1}^{n} x_j^{(k)} y^{(j)}.$$

Then for any $j = 1, \ldots, n$, the sequence $\{x_j^{(k)}\}$ is a Cauchy sequence in $(\mathbb{R}, |\cdot|)$. Therefore, $x_j^{(k)} \to x_j^*$ for some $x_j^* \in \mathbb{R}$. We thus obtain

$$x^{(k)} = \sum_{j=1}^{n} x_j^{(k)} y^{(j)} \to \sum_{j=1}^{n} x_j^* y^{(j)} \in Y.$$

This shows that $(x^{(k)})$ converges in Y. Therefore, Y is a complete subset of V and thus closed by Lemma 2.10. $\square$

Note: In particular, if V itself is finite-dimensional, then it is complete and thus closed.

§ 2.4 Compactness

Compactness of sets plays a major role in topology, analysis, and also in optimization.

Definition 2.15 (compact, sequentially compact and totally bounded sets).
Suppose that $(V, \|\cdot\|)$ is a normed linear space and $E \subseteq V$ is some subset.

(*i*) A collection $(U_i)_{i \in I}$ of open subsets $U_i \subseteq V$ is said to be an **open cover** of E if $E \subseteq \bigcup_{i \in I} U_i$ holds.

(*ii*) A subset $K \subseteq V$ is said to be **compact** if every open cover $(U_i)_{i \in I}$ of K contains a finite subcover, i. e., there exist a finite number of indices $i_1, \ldots, i_N \in I$ such that $K \subseteq \bigcup_{j=1}^{N} U_{i_j}$.

(*iii*) A subset $K \subseteq V$ is said to be **sequentially compact** if every sequence $(x^{(k)})$ in K contains a convergent subsequence whose limit belongs to K.[9]

[9] Stated equivalently, $(x^{(k)})$ has an accumulation point in K.

(*iv*) A subset $K \subseteq V$ is said to be **totally bounded** if for any $\varepsilon > 0$, there exist finitely many $x^{(1)}, \dots, x^{(N)} \in K$ such that $\{B_\varepsilon(x^{(1)}), \dots, B_\varepsilon(x^{(N)})\}$ covers K. △

The verification of compactness via Definition 2.15 (*ii*) can be cumbersome. The following results can help.

Lemma 2.16 (compact sets are closed and bounded).
Suppose that $(V, \|\cdot\|)$ is a normed linear space and $K \subseteq V$ is a compact subset. Then K is closed and bounded.

Proof. We prove both properties independently.

Step 1: We show that K is closed.

The statement is true when $K = V$ (**Quiz 2.3:** Is it clear to you?), so suppose $K \subsetneq V$ from now on. Suppose that $z \in V \setminus K$ is a point of the complement of K. We need to show that there exists an open ball $B_\varepsilon(z) \subseteq V \setminus K$.

For any $x \in K$, define $\varepsilon_x := \frac{1}{2}\|x - z\|$. In view of $z \notin K$ and the positive definiteness of the norm, we have $\varepsilon_x > 0$. The open balls $B_{\varepsilon_x}(x)$ and $B_{\varepsilon_x}(z)$ are disjoint since for any point y in their intersection, the triangle inequality would imply the contradiction

$$\|x - z\| \leqslant \|x - y\| + \|y - z\| < \varepsilon_x + \varepsilon_x = \|x - z\|.$$

The sets $\{B_{\varepsilon_x}(x) \mid x \in K\}$ form an open cover of K. Since K is compact, finitely many of these suffice, say, those with center points $x^{(1)}, \dots, x^{(N)} \in K$. As we noticed above, $B_{\varepsilon_{x^{(j)}}}(x^{(j)})$ and $B_{\varepsilon_{x^{(j)}}}(z)$ are disjoint for all $j = 1, \dots, N$. Let $\varepsilon := \min\{\varepsilon_{x^{(1)}}, \dots, \varepsilon_{x^{(N)}}\}$. Then $B_\varepsilon(z)$ is disjoint from all $B_{\varepsilon_{x^{(j)}}}(x^{(j)})$ and hence from K.

Step 2: We show that K is bounded.

Fix $x \in V$ arbitrarily and consider the open balls $\{B_i(x) \mid i \in \mathbb{N}\}$. Since every element of K has a finite distance from the point x, this collection of open balls covers K. Since K is compact, a finite number of these suffice, say,

$$\{B_{i^{(1)}}(x), \dots, B_{i^{(N)}}(x)\}.$$

These being balls with the same center, one of them is largest, say, $B_{i^{(*)}}(x)$, which alone covers K. □

Theorem 2.17 (in normed linear spaces, the notions of compact and sequentially compact sets coincide).

Suppose that $(V, \|\cdot\|)$ is a normed linear space and $K \subseteq V$ is some subset. Then the following statements are equivalent:

(*i*) K is compact.

(*ii*) K is sequentially compact.

(*iii*) K is complete and totally bounded.

Proof. Statement (i) $\Rightarrow$ statement (ii): Suppose that $(x^{(n)})$ is a sequence in K that does not possess a convergent subsequence with limit in K. In other words, $(x^{(n)})$ does not have an accumulation point in K. Therefore, for any $x \in K$, there exists $\varepsilon_x > 0$ such that $x^{(k)} \in B_x(\varepsilon_x)$ holds only for finitely many indices k. The sets $\{B_{\varepsilon_x}(x) \mid x \in K\}$ form an open cover of K. By the compactness of K, there exists a finite subcover

$$\{B_{\varepsilon_{x^{(1)}}}(x^{(1)}), \dots, B_{\varepsilon_{x^{(N)}}}(x^{(N)})\}$$

of K. By construction, $x^{(k)} \in B_{\varepsilon_{x^{(i)}}}(x^{(i)})$ holds only for finitely many indices k. That is, $x^{(k)} \in \bigcup_{i=1}^N B_{\varepsilon_{x^{(i)}}}(x^{(i)})$ also holds only for finitely many indices k. Therefore, finally, $x^{(k)} \in K$ also holds only for finitely many indices k. This contradicts $(x^{(n)})$ being a sequence in K.

Statement (ii) $\Rightarrow$ statement (iii): Suppose now that K is sequentially compact. Then, by definition, every sequence in K contains a convergent subsequence whose limit belongs to K. In particular, this is true for any Cauchy sequence in K, hence K is complete.

To show that K is totally bounded, suppose that $\varepsilon > 0$. If $K = \emptyset$, nothing is to be done, so suppose $K \neq \emptyset$. Pick a point $x^{(1)} \in K$. In case $K \subseteq B_\varepsilon(x^{(1)})$, we are done. Otherwise, pick a point $x^{(2)} \in K \setminus B_\varepsilon(x^{(1)})$. In case $K \subseteq B_\varepsilon(x^{(1)}) \cup B_\varepsilon(x^{(2)})$, we are done. Otherwise, continue in the same way. If this process produced an infinite sequence $(x^{(k)})$, its members would satisfy $\|x^{(k)} - x^{(\ell)}\| \geqslant \varepsilon$ for all $k \neq \ell$. Therefore, this sequence in K cannot have a convergent subsequence, contradicting the assumption that K is sequentially compact. Consequently, the process above terminates after finitely many steps, showing $K \subseteq \bigcup_{i=1}^N B_{\varepsilon_{x^{(i)}}}(x^{(i)})$. That is, K is totally bounded.

Statement (iii) $\Rightarrow$ statement (i): We proceed by contradiction. Suppose that $(U_i)_{i \in I}$ is an open cover of K that does not possess a finite subcover.

Since K is totally bounded, K can be covered by a finite number of open balls of radius 1 with centers in K. For at least one of these, say, $B_1(x^{(0)})$, the intersection $B_1(x^{(0)}) \cap K$ cannot be covered by a finite subfamily of $(U_i)_{i \in I}$. (Otherwise, K itself could be covered by a finite subfamily of $(U_i)_{i \in I}$, which we assumed is not the case.)

Now consider $B_1(x^{(0)}) \cap K$. As a subset of K, this set is again totally bounded and thus can be covered by a finite number of open balls of radius 1/2 with centers in $B_1(x^{(0)}) \cap K$. Again, for at least one of these, say, $B_{1/2}(x^{(2)})$, the intersection $B_{1/2}(x^{(2)}) \cap K$ cannot be covered by a finite subfamily of $(U_i)_{i \in I}$. (Otherwise, $B_1(x^{(0)}) \cap K$ itself could be covered by a finite subfamily of $(U_i)_{i \in I}$, which we know is not the case.)

Repeating this process, we obtain a sequence of balls $B_{2^{-k}}(x^{(k)})$, for none of which $B_{2^{-k}}(x^{(k)}) \cap K$ is covered by a finite subfamily of $(U_i)_{i \in I}$. The centers satisfy $x^{(k+1)} \in B_{2^{-k}}(x^{(k)}) \cap K$. Therefore, the sequence $(x^{(k)})$ is a Cauchy sequence in K since $\|x^{(k)} - x^{(\ell)}\| < 2^{1-k}$ holds for all $\ell \geqslant k$. (**Quiz 2.4:** Can you fill in the details?) Since K was assumed to be a complete subset of V, this Cauchy sequence converges and its limit x^* belongs to K.

This implies that x^* belongs to some member of the family $(U_i)_{i \in I}$, say, $x \in U_{i^*}$. Since U_{i^*} is open, there exists $\varepsilon > 0$ such that

$$x^* \in B_\varepsilon(x^*) \subseteq U_{i^*}$$

holds. We can find an index $N \in \mathbb{N}$ such that $2^{-N} < \varepsilon/2$ and

$$\|x^{(N)} - x^*\| < \frac{\varepsilon}{2}$$

holds. Consequently, for any $y \in B_{2^{-N}}(x^{(N)})$, we have

$$\|y - x^*\| \leqslant \|y - x^{(N)}\| + \|x^{(N)} - x^*\| < 2^{-N} + \frac{\varepsilon}{2} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

which means that we have

$$B_{2^{-N}}(x^{(N)}) \subseteq B_\varepsilon(x^*) \subseteq U_{i^*}.$$

This, however, contradicts the fact that for none of the balls $B_{2^{-k}}(x^{(k)})$, the intersection $B_{2^{-k}}(x^{(k)}) \cap K$ can be covered by a finite subfamily of $(U_i)_{i \in I}$.

Consequently, the assumption that there exists an open cover $(U_i)_{i \in I}$ of K that does not possess a finite subcover, cannot be true. This shows that K is compact. □

The notion of compactness is very strong in infinite-dimensional normed linear spaces. As a consequence, only "few" sets are compact.

Theorem 2.18 (compactness of the unit ball).
Suppose that $(V, \|\cdot\|)$ is a normed linear space. Then the following statements are equivalent:

(i) The closed unit ball $\overline{B_1(0)}$ is compact.

(ii) The unit sphere $\partial B_1(0)$ is compact.

(iii) $\dim(V)$ is finite.

Notice that this theorem holds independently of which particular norm is chosen on the linear space V!

The proof of Theorem 2.18 uses the following result:

Lemma 2.19 (Riesz lemma).
Suppose that $(V, \|\cdot\|)$ is a normed linear space. Moreover, let $Y \subsetneq V$ be a closed proper subspace of V. Then for any $\theta \in (0, 1)$, there exists $x_\theta \in V$ of unit norm $\|x_\theta\| = 1$ such that

$$\theta \leqslant \|x_\theta - y\| \quad \text{for all } y \in Y. \tag{2.8}$$

Note: Read this as: "You can find a vector x_θ on the unit sphere that is at least the distance θ away from any point in the subspace Y." This result is sometimes written equivalently as[10]

$$\theta \leqslant \operatorname{dist}_Y(x_\theta) \leqslant 1.$$

Proof. Pick any $v \in V \setminus Y$ and define $R := \inf\{\|v - y\| \mid y \in Y\}$. By Lemma 2.4, $\operatorname{dist}_Y(x) = 0$ if and only if $x \in \operatorname{cl} Y$. Therefore, we have $R = \operatorname{dist}_Y(v) > 0$. Due to $\theta < 1$, we can find $y_\theta \in Y$ such that

$$0 < \|v - y_\theta\| \leqslant \frac{R}{\theta} \tag{2.9}$$

holds. We define

$$x_\theta := \frac{v - y_\theta}{\|v - y_\theta\|}.$$

[10] The **distance** of a point x to a set Y in a normed linear space is defined as $\operatorname{dist}_Y(x) := \inf\{\|x - y\| \mid y \in Y\}$.

Then we have $\|x_\theta\| = 1$ and, for any $y \in Y$,

$$\begin{aligned} \|x_\theta - y\| &= \left\| \frac{v - y_\theta}{\|v - y_\theta\|} - y \right\| \\ &= \frac{1}{\|v - y_\theta\|} \Big\| v - \big(\underbrace{y_\theta + \|v - y_\theta\| \, y}_{\in Y} \big) \Big\| \\ &\geqslant \frac{R}{\|v - y_\theta\|}. \end{aligned}$$

Together with (2.9), this proves (2.8). □

End of Class 4

Proof of Theorem 2.18:
Item (*i*) ⇒ item (*iii*): When the closed unit ball $\overline{B_1(0)}$ is compact, then it is also totally bounded by Theorem 2.17. Thus, it can be covered by finitely many balls of radius $1/2$:

$$\overline{B_1(0)} \subseteq \bigcup_{i=1}^{N} B_{1/2}(y^{(i)}).$$

Define $Y := \operatorname{span}\{y^{(1)}, \ldots, y^{(N)}\}$. Then by Lemma 2.14, Y is a closed subspace of V.

Suppose that $Y \subsetneq V$ is a *proper* subspace. The Riesz lemma 2.19 then implies that there exists $x_\theta \in V$ of unit norm such that $\operatorname{dist}_Y(x_\theta) \geqslant \theta := \frac{3}{4}$. Moreover, x_θ belongs to one of the covering balls, say, $B_{1/2}(y^{(j)})$. Therefore, we have

$$\operatorname{dist}_Y(x_\theta) \leqslant \|x_\theta - y^{(j)}\| < \frac{1}{2},$$

which contradicts $\operatorname{dist}_Y(x_\theta) \geqslant \frac{3}{4}$. Therefore, $Y = V$ and $\dim(V)$ is finite.

Item (*ii*) ⇒ item (*iii*): The proof is the same as above.

Item (*iii*) ⇒ item (*i*): The closed unit ball $\overline{B_1(0)}$ is a clearly a closed subset of V. Suppose that $\dim(V) = n \in \mathbb{N}_0$ and that $\{v^{(1)}, \ldots, v^{(n)}\}$ is a basis of V. Then V is complete by Lemma 2.14. When we show that $\overline{B_1(0)}$ is totally bounded w.r.t. $\|\cdot\|$, then it is compact by Theorem 2.17. By the equivalence of norms (Theorem 2.13), we may equivalently show that $\overline{B_1(0)}$ is totally bounded w.r.t. $\|\cdot\|_\infty$.

Suppose that $\|\cdot\|_\infty \leqslant c \, \|\cdot\|$ holds for $c > 0$. Consider $\varepsilon > 0$. We claim that

$$\overline{B_1(0)} \subseteq \overline{B_c^{\|\cdot\|_\infty}(0)} \subseteq \bigcup_{\substack{q \in \varepsilon \mathbb{Z}^n \\ \|q\|_\infty \leqslant c + \varepsilon/2}} B_\varepsilon^{\|\cdot\|_\infty}\Big(\sum_{j=1}^{n} q_j \, v^{(j)} \Big),$$

holds. Notice that the right-hand side is a finite union of open balls of radius ε. The first inequality is clear. For the second inequality, consider a point $x \in \overline{B_c^{\|\cdot\|_\infty}(0)}$, whose coordinates x_j then satisfy $|x_j| \leqslant c$. For $j = 1, \ldots, n$, find $q_j \in \varepsilon \mathbb{Z}$ closest to x_j. This implies $|x_j - q_j| \leqslant \varepsilon/2$ and thus

$$\Big\| x - \sum_{j=1}^{n} q_j \, v^{(j)} \Big\|_\infty \leqslant \frac{\varepsilon}{2} < \varepsilon.$$

In other words, x belongs to the open ball $B_\varepsilon^{\|\cdot\|_\infty}\Big(\sum_{j=1}^n q_j\, v^{(j)}\Big)$. Due to $|x_j| \leqslant c$, we will have $|q_j| \leqslant c+\varepsilon/2$. This proves the claim.

Item (iii) $\Rightarrow$ item (ii): The proof is the same as above. □

Note: The proof item (iii) $\Rightarrow$ item (i) can be easily extended to show that every bounded set in a finite-dimensional normed linear space is totally bounded.

Remark 2.20 (there is nothing special about *unit* balls)**.**
For any $r > 0$, the closed ball $\overline{B_r(0)}$ is compact if and only if $\dim(V)$ is finite. The same holds for spheres. △

§ 2.5 Lebesgue Spaces

Literature: Rudin, 1987, Chapter 3

Lebesgue spaces are prominent examples of Banach spaces. All references to a measure will mean the Lebesgue measure on $\mathbb{R}^d$. We will state results in this subsection without proof.

Definition 2.21 (Lebesgue spaces)**.**
Suppose that $\Omega \subseteq \mathbb{R}^n$ is an open set and $p \in [1, \infty)$.

- (i) A measurable function $f\colon \Omega \to \mathbb{R}$ is said to be **Lebesgue integrable of index p** or simply **p-integrable** if $|f|^p$ is integrable on Ω.
- (ii) A measurable function $f\colon \Omega \to \mathbb{R}$ is said to be **essentially bounded** if it is bounded except on a set of measure zero.
- (iii) Two measurable functions $f, g\colon \Omega \to \mathbb{R}$ are said to be **equivalent** if they coincide except on a set of measure zero.
- (iv) The **Lebesgue space** $L^p(\Omega)$ is defined as the set of equivalence classes[11] of measurable functions $f\colon \Omega \to \mathbb{R}$ that are Lebesgue integrable of index p:
$$L^p(\Omega) \coloneqq \big\{[f] \,\big|\, f\colon \Omega \to \mathbb{R} \text{ is Lebesgue integrable of index } p\big\}. \tag{2.10}$$
- (v) The **Lebesgue space** $L^\infty(\Omega)$ is defined as the set of equivalence classes of measurable functions $f\colon \Omega \to \mathbb{R}$ that are essentially bounded:
$$L^\infty(\Omega) \coloneqq \big\{[f] \,\big|\, f\colon \Omega \to \mathbb{R} \text{ is essentially bounded}\big\}. \tag{2.11}$$
△

It is customary to denote the equivalence class of a function f by f itself. We will do so from now on.

[11] The construction is that of a quotient space: we begin with the vector space of p-integrable functions and factor out the subspace of functions which are almost everywhere zero. Recall that "**almost everywhere**" means "except on a set of measure zero".

Theorem 2.22 (Lebesgue spaces as Banach spaces).
Suppose that $\Omega \subseteq \mathbb{R}^n$ is an open set.

(*i*) For $p \in [1, \infty)$, the Lebesgue space $L^p(\Omega)$ is a Banach space when equipped with the norm

$$\|f\|_{L^p(\Omega)} := \Big(\int_\Omega |f|^p\Big)^{1/p}. \tag{2.12}$$

(*ii*) The Lebesgue space $L^\infty(\Omega)$ is a Banach space when equipped with the norm

$$\|f\|_{L^\infty(\Omega)} := \operatorname*{ess\,sup}_{x\in\Omega}|f(x)| := \inf\big\{M \geqslant 0 \,\big|\, |f(x)| \leqslant M \text{ for almost all } x \in \Omega\big\}. \tag{2.13}$$

(*iii*) For any $p \in [1, \infty]$, the triangle inequality $\|f+g\|_{L^p(\Omega)} \leqslant \|f\|_{L^p(\Omega)}+\|g\|_{L^p(\Omega)}$ for all $f, g \in L^p(\Omega)$ is called the **Minkowski inequality**.

Example 2.23 (functions in L^p).

(*i*) On $\Omega = \mathbb{R}$, non-zero constant functions belong to $L^\infty(\mathbb{R})$ but not to any $L^p(\mathbb{R})$ with $p < \infty$.

(*ii*) On $\Omega = (-1, 1)$, the absolute power function $x \mapsto |x|^\alpha$ belongs to $L^p((-1, 1))$ if and only if $\alpha\, p > -1$.[12] For instance, the inverse square root function $x \mapsto 1/\sqrt{|x|} = x^{-1/2}$ belongs to $L^p((-1, 1))$ if and only if $p < 2$.

(*iii*) More generally, on the open unit ball $\Omega = B_1(0) \subseteq \mathbb{R}^d$, the function $x \mapsto |x|^\alpha$ belongs to $L^p(\Omega)$ if and only if $\alpha\, p > -d$ holds.[13] △

Lemma 2.24 (Hölder's inequality).
Suppose that $\Omega \subseteq \mathbb{R}^d$ is an open set. Moreover, let $p, q \in [1, \infty]$ be such that $\frac{1}{p} + \frac{1}{q} = 1$.[14] For all $f \in L^p(\Omega)$ and $g \in L^q(\Omega)$, the product $f\, g$ belongs to $L^1(\Omega)$, and the estimate

$$\|f\, g\|_{L^1(\Omega)} \leqslant \|f\|_{L^p(\Omega)}\, \|g\|_{L^q(\Omega)} \tag{2.14}$$

holds. Inequality (2.14) is known as **Hölder inequality**.

Lemma 2.25 (comparison of norms on Lebesgue spaces).
Suppose that $\Omega \subseteq \mathbb{R}^d$ is an open and bounded set. For $1 \leqslant p \leqslant q \leqslant \infty$, the space $L^q(\Omega)$ is a subspace of $L^p(\Omega)$ (and a proper subspace if $1 \leqslant p < q \leqslant \infty$). Moreover, the L^q-norm is stronger than the L^p-norm:

$$\|f\|_{L^p(\Omega)} \leqslant |\Omega|^{\frac{q-p}{pq}}\, \|f\|_{L^q(\Omega)} \quad \text{for all } f \in L^q(\Omega), \tag{2.15}$$

where $|\Omega|$ denotes the Lebesgue measure (d-dimensional volume) of Ω. When $q = \infty$, the expression $\frac{q-p}{pq}$ is to be understood as $1/p$ (for $p < \infty$) or as 0 (for $p = \infty$).

Note: Lemma 2.25 states that the higher the index of a Lebesgue space on a bounded domain, the smaller the space and the stronger the norm.

[12] With the convention that $\alpha\,\infty = \infty$ for $\alpha > 0$ and $\alpha\,\infty = -\infty$ for $\alpha < 0$ as well as $0\,\infty = 0$.
[13] Here $|\cdot|$ denotes the Euclidean norm on $\mathbb{R}^d$.
[14] Such numbers p, q are called **conjugate exponents**. The convention here is that $1/\infty = 0$ so that 1 and ∞ are conjugate.

Example 2.26 (comparison of norms on Lebesgue spaces).
Suppose that $\Omega \subseteq \mathbb{R}^d$ is an open and bounded set.

(i) $\|f\|_{L^1(\Omega)} \leqslant |\Omega|^{1/2} \|f\|_{L^2(\Omega)}$ for all $f \in L^2(\Omega)$.

(ii) $\|f\|_{L^2(\Omega)} \leqslant |\Omega|^{1/2} \|f\|_{L^\infty(\Omega)}$ for all $f \in L^\infty(\Omega)$.

(iii) $\|f\|_{L^1(\Omega)} \leqslant |\Omega| \|f\|_{L^\infty(\Omega)}$ for all $f \in L^\infty(\Omega)$. △

End of Class 5

End of Week 3

§ 2.6 Sobolev Spaces

Lebesgue spaces are not sufficient to deal with optimization problems whose objective functions involve derivatives of the unknown, as is the case in the Brachistochrone problem (Example 1.1), Fermat's principle in optics (Example 1.3), the signal denoising problem (Example 1.4), and the optimal control example (Example 1.5). Sobolev spaces are the natural setting for such problems. In brief, they consist of functions whose derivatives up to a certain order are in a Lebesgue space. The notion of derivative is meant in a *weak sense*.

Derivatives of multivariate functions are conveniently described using multi-indices.

Definition 2.27 (multi-index).
Suppose $d \in \mathbb{N}$.

(i) A **multi-index** of length d is a tuple $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$.

(ii) The **order** of a multi-index $\alpha = (\alpha_1, \ldots, \alpha_d)$ is defined as $|\alpha| \coloneqq \alpha_1 + \ldots + \alpha_d$.

(iii) We associate with a multi-index $\alpha = (\alpha_1, \ldots, \alpha_d)$ the derivative operator $D^\alpha \coloneqq \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}$. The **order** of D^α is defined as $|\alpha|$.

(iv) In particular, we have $D^{(0,\ldots,0)} = \mathrm{id}$ and

$$D_i \coloneqq \frac{\partial}{\partial x_i} = D^{(0,\ldots,0,1,0,\ldots,0)}$$

for $i = 1, \ldots, d$. △

Definition 2.28 (function spaces $C^k(\Omega)$, $C_c^k(\Omega)$ and $C^k(\operatorname{cl}\Omega)$).
Suppose that $\Omega \subseteq \mathbb{R}^d$ is an open set.

(i) For $f \colon \Omega \to \mathbb{R}$, the set

$$\operatorname{supp} f \coloneqq \operatorname{cl}\{x \in \Omega \mid f(x) \neq 0\} \tag{2.16}$$

is called the **support** of f.

(ii) For $k \in \mathbb{N}_0$, the set of all **k-times continuously partially differentiable functions** on Ω is denoted by $C^k(\Omega)$. This means that all partial derivatives of order $\leqslant k$ exist and are continuous functions on Ω.

Moreover, $C^\infty(\Omega) \coloneqq \bigcap_{k \in \mathbb{N}_0} C^k(\Omega)$ is the set of all **infinitely often continuously partially differentiable functions** on Ω.

(*iii*) For $k \in \mathbb{N}_0$, the set $C_c^k(\Omega)$ consists of all **functions** $f \in C^k(\Omega)$ **with compact support**, i. e., $\operatorname{supp} f$ is a compact subset of Ω.

Moreover, $C_c^\infty(\Omega) := \bigcap_{k \in \mathbb{N}_0} C_c^k(\Omega)$ is the set of all **infinitely often continuously partially differentiable functions** on Ω **with compact support**.

(*iv*) For $k \in \mathbb{N}_0$, $C^k(\operatorname{cl}\Omega)$ denotes the set of all k-times continuously partially differentiable functions $f \colon \Omega \to \mathbb{R}$ such that all partial derivatives of order $\leqslant k$ extend continuously to $\operatorname{cl}\Omega$. △

Note: Given $k \in \mathbb{N}_0$ and $f \in C^k(\Omega)$, the support of all partial derivatives $D^\alpha f$ of order $|\alpha| \leqslant k$ is contained in the support of f.

Lemma 2.29 (properties of derivatives)**.**
Suppose that $\Omega \subseteq \mathbb{R}^d$ is an open set. Moreover, suppose that $\alpha \in \mathbb{N}_0^d$ is a multi-index of order $m \in \mathbb{N}_0$ and $k \geqslant m$. Then the following holds:

(*i*) The derivative operator D^α is well-defined as a map

$$D^\alpha \colon C^k(\Omega) \to C^{k-m}(\Omega).$$

(*ii*) The order of differentiation does not matter, i. e., for any decomposition of the multi-index $\alpha = \beta + \gamma$, we have

$$D^\alpha f = D^\beta(D^\gamma f) = D^\gamma(D^\beta f)$$

for all $f \in C^k(\Omega)$.

Proof. Statement (*i*) follows immediately from the fact that higher-order partial derivatives are derivatives of lower-order partial derivatives. Statement (*ii*) is a consequence of the commutativity of partial derivatives by Schwarz' theorem. □

Example 2.30 (function spaces $C^k(\Omega)$ and $C^k(\operatorname{cl}\Omega)$)**.**

(*i*) For $\Omega = (0, 1)$, the function $x \mapsto 1/x$ belongs to $C^\infty(\Omega)$ but not to $C(\operatorname{cl}\Omega)$ since it does not extend continuously to 0.

(*ii*) For $\Omega = (0, 1)$, the function $x \mapsto \sqrt{x}$ belongs to $C^\infty(\Omega)$ and to $C(\operatorname{cl}\Omega)$ but not to $C^1(\operatorname{cl}\Omega)$ since the derivative $1/(2\sqrt{x})$ does not extend continuously to 0. △

Lemma 2.31 (integration by parts)**.**
Suppose that $\Omega \subseteq \mathbb{R}^d$ is an open set and $f \in C^1(\Omega)$. Then for any $i = 1, \ldots, d$, we have

$$\int_\Omega (D_i f)\, g \,\mathrm{d}x = -\int_\Omega f\, (D_i g) \,\mathrm{d}x \quad \text{for all } g \in C_c^1(\Omega). \tag{2.17}$$

Note: The supports of both integrands are compact subsets of Ω and the integrands are continuous, so that the integrals are well-defined.

Proof. Suppose that $C = (a_1, b_1) \times \cdots \times (a_d, b_d) \subseteq \mathbb{R}^d$ is an open and bounded box containing the compact set $\operatorname{supp} g$. Define

$$\Phi(x) \coloneqq \begin{cases} f(x)\, g(x) & \text{if } x \in \Omega, \\ 0 & \text{if } x \in \mathbb{R}^d \setminus \Omega. \end{cases}$$

Then by the product rule, $\Phi \in C^1(\mathbb{R}^d)$ and $\operatorname{supp} \Phi \subseteq \operatorname{supp} g \subseteq C$ and thus also $\operatorname{supp} D_1\Phi \subseteq C$.

For notational convenience, we consider only the case $i = 1$. By the fundamental theorem of calculus, we have

$$\int_{a_1}^{b_1} D_1\Phi(x_1, x_2, \ldots, x_d) \;\mathrm{d}x_1 = \Phi(b_1, x_2, \ldots, x_d) - \Phi(a_1, x_2, \ldots, x_d)$$

for any $x_2, \ldots, x_d \in \mathbb{R}$. Plugging in the definition of Φ, this amounts to

$$\begin{aligned} &\int_{a_1}^{b_1} (D_1 f)(x_1, x_2, \ldots, x_d)\, g(x_1, x_2, \ldots, x_d) \;\mathrm{d}x_1 + \int_{a_1}^{b_1} f(x_1, x_2, \ldots, x_d)\, (D_1 g)(x_1, x_2, \ldots, x_d) \,\mathrm{d}x_1 \\ &\qquad = f(b_1, x_2, \ldots, x_d)\; \underbrace{g(b_1, x_2, \ldots, x_d)}_{=0} - f(a_1, x_2, \ldots, x_d)\; \underbrace{g(a_1, x_2, \ldots, x_d)}_{=0}. \end{aligned}$$

Notice that the right-hand side is zero since the points where g is being evaluated are outside of $\operatorname{supp} g$. We now see that

$$\begin{aligned} &\int_\Omega (D_1 f)\, g \;\mathrm{d}x + \int_\Omega f\, (D_1 g) \;\mathrm{d}x \\ &\qquad = \int_C (D_1 f)\, g \;\mathrm{d}x + \int_C f\, (D_1 g) \;\mathrm{d}x \\ &\qquad = \int_{a_d}^{b_d} \cdots \int_{a_1}^{b_1} (D_1 f)(x_1, x_2, \ldots, x_d)\, g(x_1, x_2, \ldots, x_d) \;\mathrm{d}x_1 \cdots \mathrm{d}x_d \\ &\qquad\quad + \int_{a_d}^{b_d} \cdots \int_{a_1}^{b_1} f(x_1, x_2, \ldots, x_d)\, (D_1 g)(x_1, x_2, \ldots, x_d) \;\mathrm{d}x_1 \cdots \mathrm{d}x_d \quad \text{by Fubini's theorem} \\ &\qquad = 0, \end{aligned}$$

which concludes the proof. □

By induction, we can generalize Lemma 2.31 to higher-order derivatives:

Corollary 2.32 (integration by parts for higher-order derivatives)**.**
Suppose that $\Omega \subseteq \mathbb{R}^d$ is an open set and $f \in C^k(\Omega)$. Then for any multi-index $\alpha \in \mathbb{N}_0^d$ of order $k \in \mathbb{N}_0$, we have

$$\int_\Omega (D^\alpha f)\, g \;\mathrm{d}x = (-1)^{|\alpha|} \int_\Omega f\, (D^\alpha g) \;\mathrm{d}x \quad \text{for all } g \in C_c^k(\Omega). \tag{2.18}$$

Formula (2.18) describes properties of classical derivatives for sufficiently smooth functions. These properties serve as a motivation for the definition of a more general notion of derivative, applicable to a much larger class of functions.

Definition 2.33 (weak derivative)**.**
Suppose that $\Omega \subseteq \mathbb{R}^d$ is an open set.

(*i*) For any $A \subseteq \Omega$, the **characteristic function** $\chi_A \colon \Omega \to \mathbb{R}$ is defined as $\chi_A(x) = 1$ if $x \in A$ and $\chi_A(x) = 0$ if $x \notin A$.

(*ii*) The set $L^1_{\text{loc}}(\Omega)$ denotes the set of all (equivalence classes of) functions $f \colon \Omega \to \mathbb{R}$ such that $f\,\chi_K \in L^1(\Omega)$ for all compact subsets $K \subseteq \Omega$.

(*iii*) Suppose that $f \in L^1_{\text{loc}}(\Omega)$ and $\alpha \in \mathbb{N}_0^d$ is a multi-index. A function $w \in L^1_{\text{loc}}(\Omega)$ is called the α-th **weak derivative** of f if

$$\int_\Omega f\, D^\alpha \varphi \,\mathrm{d}x = (-1)^{|\alpha|} \int_\Omega w\, \varphi \,\mathrm{d}x \quad \text{for all } \varphi \in C_c^\infty(\Omega) \tag{2.19}$$

holds. In this case, we write $w = D^\alpha f$. △

Note: For any $\varphi \in C_c^\infty(\Omega)$, both v and $D^\alpha v$ have compact support in Ω. The function class $L^1_{\text{loc}}(\Omega)$ is therefore a natural setting so that the integrals in (2.19) are well-defined.

Remark 2.34 (weak derivative).

(*i*) The α-th weak derivative $D^\alpha f$ of a function $f \in L^1_{\text{loc}}(\Omega)$ is unique (if it exists).

(*ii*) The existence of a weak derivative $D^\alpha f$ does not imply the existence of weak derivatives $D^{\alpha'} f$ for multi-indices $\alpha' \leqslant \alpha$.

(*iii*) If both $D^\alpha f$ and $D^{\alpha+\beta} f$ exist, then $D^{\alpha+\beta} f = D^\beta(D^\alpha f)$.

(*iv*) If both $D^\alpha f$ and $D^\beta(D^\alpha f)$ exist, then $D^{\alpha+\beta} f = D^\beta(D^\alpha f)$. △

Example 2.35 (weak derivative).
The function $f \colon \Omega := (-1, 1) \to \mathbb{R}$ defined by $f(x) = |x|$ has the weak first-order derivative

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

But f does not have a weak second-order derivative in $L^1_{\text{loc}}(\Omega)$. △

We can now define the Sobolev spaces.

Definition 2.36 (Sobolev spaces).
Suppose that $\Omega \subseteq \mathbb{R}^d$ is an open set. The **Sobolev space** of differentiability index $k \in \mathbb{N}_0$ and index $p \in [1, \infty]$ is defined as

$$W^{k,p}(\Omega) := \left\{ f \in L^p(\Omega) \,\middle|\, D^\alpha f \in L^p(\Omega) \text{ for all } |\alpha| \leqslant k \right\}. \tag{2.20}$$

△

We re-iterate that the elements of a Sobolev space are actually equivalence classes of functions but we continue to use the simplified notation. For $k = 0$, the Sobolev spaces agree with the Lebesgue spaces: $W^{0,p}(\Omega) = L^p(\Omega)$

Remark 2.37 (alternative definition of Sobolev spaces).
For $k \in \mathbb{N}_0$ and $p \in [1, \infty)$, Sobolev spaces can be defined alternatively via a process of completion:

$$H^{k,p}(\Omega) \coloneqq \operatorname{cl}\big(C^\infty(\Omega) \cap W^{k,p}(\Omega)\big) \text{ w.r.t. the norm } \|\cdot\|_{W^{k,p}(\Omega)}.$$

The paper Meyers, Serrin, 1964 with the title "$H = W$" shows that $H^{k,p}(\Omega) = W^{k,p}(\Omega)$; see also Adams, Fournier, 2003, Theorem 3.17. △

Theorem 2.38 (Sobolev spaces as Banach spaces).
Suppose that $\Omega \subseteq \mathbb{R}^n$ is an open set.

(i) For $k \in \mathbb{N}_0$ and $p \in [1, \infty)$, the Sobolev space $W^{k,p}(\Omega)$ is a Banach space when equipped with the norm

$$\|f\|_{W^{k,p}(\Omega)} \coloneqq \Big(\sum_{|\alpha| \leqslant k} \int_\Omega |D^\alpha f|^p \Big)^{1/p} = \Big(\sum_{|\alpha| \leqslant k} \|D^\alpha f\|_{L^p(\Omega)}^p \Big)^{1/p}. \tag{2.21}$$

(ii) The Sobolev space $W^{k,\infty}(\Omega)$ is a Banach space when equipped with the norm

$$\|f\|_{W^{k,\infty}(\Omega)} \coloneqq \max_{|\alpha| \leqslant k} \operatorname*{ess\,sup}_{x \in \Omega} |D^\alpha f(x)| = \max_{|\alpha| \leqslant k} \|D^\alpha f\|_{L^\infty(\Omega)}. \tag{2.22}$$

Example 2.39 (Sobolev spaces).
By Example 2.35, the function defined by $f(x) = |x|$ on $\Omega = (-1, 1)$ belongs to $W^{1,\infty}(\Omega)$. However, it does not belong to any $W^{2,p}(\Omega)$ for any $p \in [1, \infty]$. △

End of Class 6

§ 3 Inner Product Spaces

In this section we introduce the notion of an inner product space, which is a concept more specific than a normed linear space.

Definition 3.1 (inner product space).
Suppose that V is a linear space.

(i) A map $(\cdot,\cdot)\colon V \times V \to \mathbb{R}$ is said to be an **inner product on V** if the following conditions hold:

$$(\alpha_1 u_1 + \alpha_2 u_2, v) = \alpha_1 (u_1, v) + \alpha_2 (u_2, v) \quad \textbf{linearity in the first argument} \tag{3.1a}$$
$$(u, \alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 (u, v_1) + \alpha_2 (u, v_2) \quad \textbf{linearity in the second argument} \tag{3.1b}$$
$$(u, v) = (v, u) \quad \textbf{symmetry} \tag{3.1c}$$
$$(u, u) \geqslant 0, \quad \text{and } (u, u) = 0 \;\Rightarrow\; u = 0 \quad \textbf{positive definiteness} \tag{3.1d}$$

for all $u, u_1, u_2, v, v_1, v_2 \in V$ and all $\alpha_1, \alpha_2 \in \mathbb{R}$.

(ii) The pair $(V, (\cdot,\cdot))$ is said to be a **(real) inner product space**.

(iii) Two vectors $u, v \in V$ are said to be **orthogonal** if $(u, v) = 0$. △

In brief, an inner product is a bilinear form on V that is symmetric and positive definite.

An inner product induces a norm on the linear space under consideration:

Lemma 3.2 (inner product induces norm)**.**
Suppose that $(V, (\cdot,\cdot))$ is an inner product space. Then

$$\|u\| \coloneqq \sqrt{(u,u)} \tag{3.2}$$

defines a norm on V.

Proof. The proof is part of homework problem 4.2. □

Definition 3.3 (Hilbert space)**.**
An inner product space $(V, (\cdot,\cdot))$ is said to be a **Hilbert space** if the norm induced by the inner product is complete (see Definition 2.8). △

Note: In other words, a Hilbert space is a Banach space whose norm is induced by an inner product.

Example 3.4 (Hilbert space)**.**

(i) In $\mathbb{R}^n$, inner products are in a bijective correspondence with symmetric positive definite matrices. Every inner product on $\mathbb{R}^n$ has the form

$$(u,v)_M = u^\intercal M\, v$$

for some symmetric positive definite matrix $M \in \mathbb{R}^{n\times n}$. The induced norm is then given by

$$\|u\|_M = \sqrt{u^\intercal M\, u}.$$

(ii) Every finite-dimensional inner product space is a Hilbert space.

(iii) Suppose that $\Omega \subseteq \mathbb{R}^d$ is an open set. The Lebesgue space $L^2(\Omega)$ carries the inner product

$$(f,g)_{L^2(\Omega)} \coloneqq \int_\Omega f\, g \,\mathrm{d}x, \tag{3.3}$$

which induces the norm (2.12) for $p = 2$. Since $L^2(\Omega)$ is complete, it is a Hilbert space.

(iv) More generally, suppose that $\Omega \subseteq \mathbb{R}^d$ is an open set and $k \in \mathbb{N}_0$. The Sobolev space $W^{k,2}(\Omega)$ carries the inner product

$$(f,g)_{W^{k,2}(\Omega)} \coloneqq \sum_{|\alpha|\leqslant k} \int_\Omega (D^\alpha f)\,(D^\alpha g) \,\mathrm{d}x, \tag{3.4}$$

which induces the norm (2.21) for $p = 2$. Since $W^{k,2}(\Omega)$ is complete, it is a Hilbert space. It is customary to denote the inner product space $W^{k,2}(\Omega)$ by $H^k(\Omega)$. In particular, $H^0(\Omega) = L^2(\Omega)$. △

Lemma 3.5 (Cauchy-Schwarz inequality).
Suppose that $(V, (\cdot,\cdot))$ is an inner product space. Then for all $u, v \in V$, we have

$$|(u, v)| \leqslant \|u\| \, \|v\|. \tag{3.5}$$

Equality holds if and only if u and v are linearly dependent, i. e., $\alpha\, u + \beta\, v = 0$ and not both α and β are zero.

Proof. When $v = 0$, then (3.5) holds with equality, and $\{u, v\}$ is linearly dependent.

For the rest of the proof, assume $v \neq 0$. For $\beta \in \mathbb{R}$ we have

$$\begin{aligned} 0 &\leqslant (u - \beta\, v, u - \beta\, v) && \text{due to positive definiteness} \\ &= (u, u) - 2\,\beta\,(u, v) + \beta^2\,(v, v) && \text{due to bilinearity and symmetry.} \end{aligned}$$

Here $(v, v) > 0$ due to positive definiteness, and we set $\beta := \frac{(u,v)}{(v,v)}$. This implies

$$\begin{aligned} 0 &\leqslant (u, u) - 2\,\frac{(u, v)}{(v, v)}\,(u, v) + \frac{(u, v)^2}{(v, v)} \\ &= (u, u) - \frac{(u, v)^2}{(v, v)}. \end{aligned}$$

Multiplication by $(v, v) > 0$ yields (3.5).

We have to investigate when equality holds in (3.5). We can continue to assume $v \neq 0$. When $\{u, v\}$ is linearly dependent, then we have $u = \delta v$ for some $\delta \in \mathbb{R}$. Bilinearity then implies $(u, u) = \delta^2\,(v, v)$ and

$$(u, v)^2 = \big(\delta\,(v, v)\big)^2 = \delta^2\,(v, v)^2 = (u, u)\,(v, v),$$

hence equality (3.5).

Conversely, suppose that equality holds in (3.5), i. e., $(u, v)^2 = (u, u)\,(v, v)$. Setting $\beta := \frac{(u,v)}{(v,v)}$ and applying the same manipulations as above, we find that

$$\begin{aligned} 0 &= (u, u) - \frac{(u, v)^2}{(v, v)} \\ &= (u, u) - 2\,\beta\,(u, v) + \beta^2\,(v, v) \\ &= (u - \beta\, v, u - \beta\, v). \end{aligned}$$

The positive definiteness implies $u - \beta\, v = 0$, and thus $\{u, v\}$ is linearly dependent. □

§ 4 Continuous Functions

The continuity of functions between normed linear spaces can be defined via sequences.

Definition 4.1 (continuity).
Suppose that X and Y are normed linear spaces. A map $F\colon X \to Y$ is said to be **continuous at $x \in X$** if for all sequences $\big(x^{(k)}\big)$ in X with $x^{(k)} \to x$ in X, we have $F(x^{(k)}) \to F(x)$ in Y. It is said to be **continuous (on X)** if it is continuous at every $x \in X$. △

Lemma 4.2 (equivalent definition of continuity)**.**
Suppose that X and Y are normed linear spaces. A map $F\colon X \to Y$ is continuous if and only if for all $x \in X$ and $\varepsilon > 0$, there exists $\delta > 0$ such that $\|x - y\|_X < \delta$ implies $\|F(x) - F(y)\|_Y < \varepsilon$.

Proof. □

§ 4.1 Linear Operators

Linear maps between normed linear spaces are of particular importance.

Definition 4.3 (linear operator, bounded linear operator)**.**
Suppose that X and Y are normed linear spaces.

(*i*) A function $A\colon X \to Y$ is said to be a **linear map** or **linear operator** if

$$A(\alpha_1\, x_1 + \alpha_2\, x_2) = \alpha_1\, A(x_1) + \alpha_2\, A(x_2) \tag{4.1}$$

holds for all $x_1, x_2 \in X$ and all $\alpha_1, \alpha_2 \in \mathbb{R}$.

(*ii*) A linear operator $A\colon X \to Y$ is said to be **bounded** if there exists $C \geqslant 0$ such that

$$\|A(x)\|_Y \leqslant C\, \|x\|_X \quad \text{for all } x \in X. \tag{4.2}$$

The number

$$\|A\|_{\mathcal{L}(X,Y)} \coloneqq \inf\big\{C \geqslant 0 \,\big|\, (4.2) \text{ holds}\big\} \tag{4.3}$$

is called the **operator norm** of A.

△

Note: It is easy to see that the interval $\{C \geqslant 0 \,|\, (4.2) \text{ holds}\}$ is closed, and thus the infimum in (4.3) is actually a minimum.

Lemma 4.4 (alternative definitions of the operator norm)**.**
Suppose that X and Y are normed linear spaces and $A\colon X \to Y$ is a bounded linear operator. The operator norm satisfies

$$\|A\|_{\mathcal{L}(X,Y)} = \sup_{\|x\|_X = 1} \|A(x)\|_Y = \sup_{\|x\|_X \leqslant 1} \|A(x)\|_Y = \sup_{x \neq 0} \frac{\|A(x)\|_Y}{\|x\|_X}. \tag{4.4}$$

Lemma 4.5 (boundedness is continuity)**.**
Suppose that X and Y are normed linear spaces and $A\colon X \to Y$ is a linear operator. Then the following statements are equivalent:

(*i*) A is continuous at 0.

(*ii*) A is continuous on X.

(*iii*) A is Lipschitz continuous.

(*iv*) A is bounded.

Proof. The proof is part of homework problem 5.3. □

End of Class 7

End of Week 4

Convergence in the operator norm implies pointwise convergence:

Lemma 4.6 (convergence in the operator norm implies pointwise convergence)**.**
Suppose that X and Y are normed linear spaces and $\big(A^{(k)}\big)$ is a sequence of bounded linear operators $X \to Y$. If $A^{(k)}$ converges to $A \in \mathcal{L}(X, Y)$ in the operator norm, then $A^{(k)}(x)$ converges to $A(x)$ for all $x \in X$.

Proof. The proof is part of homework problem 5.2. □

Lemma 4.7 (existence of unbounded operators)**.**
Suppose that X and Y are normed linear spaces with $\dim(Y) \geqslant 1$. Then the following statements are equivalent:

(i) X is finite-dimensional.

(ii) Every linear operator $A\colon X \to Y$ is continuous.

Proof. Statement (i) $\Rightarrow$ statement (ii): Suppose that $\dim(X) = n \in \mathbb{N}_0$ and that $\{v^{(1)}, \ldots, v^{(n)}\}$ is a basis of X. For any $x \in X$, we can write $x = \sum_{j=1}^n x_j\, v^{(j)}$ and thus $A(x) = \sum_{j=1}^n x_j\, A(v^{(j)})$. We estimate

$$\|A(x)\|_Y = \Big\|\sum_{j=1}^n x_j\, A(v^{(j)})\Big\|_Y \leqslant \|x\|_\infty \sum_{j=1}^n \|A(v^{(j)})\|_Y =: C\,\|x\|_\infty,$$

where $C \geqslant 0$ is a constant. By Theorem 2.13, the norms $\|\cdot\|_\infty$ and $\|\cdot\|_X$ are equivalent, and thus A is continuous.

$\neg$ Statement (i) $\Rightarrow$ $\neg$ statement (ii): Suppose that X is infinite-dimensional, i. e., at least of countable dimension. Suppose that $(v^{(i)})_{i \in I}$ is a basis for X. Without loss of generality, $\mathbb{N} \subseteq I$. Pick a non-zero element $y \in Y$ and define the linear operator $A\colon X \to Y$ by $A(v^{(k)}) = k\,\|v^{(k)}\|_X\, y$ for $k \in \mathbb{N}$, and $A(v^{(i)}) = 0$ for $i \in I \setminus \mathbb{N}$. Then A is not bounded since $\|A(x^{(k)})\|_Y = k\,\|y\|_Y$ for all $k \in \mathbb{N}$. □

The set of all linear operators $X \to Y$ forms itself a linear space, which we denote by $L(X, Y)$. Addition and scalar multiplication are defined pointwise. The subset of *bounded* linear operators forms a subspace:

Theorem 4.8 (subspace of bounded linear operators)**.**
Suppose that X and Y are normed linear spaces.

(i) The set of all bounded linear operators $X \to Y$ is a linear subspace of the space of all linear operators $X \to Y$. We denote it by $\mathcal{L}(X, Y)$.

(ii) The operator norm (4.3) is a norm on $\mathcal{L}(X, Y)$.

(iii) If Y is a Banach space, then $\mathcal{L}(X, Y)$ is a Banach space.

(*iv*) If $\mathcal{L}(X,Y)$ is a Banach space and $\dim(X) \geqslant 1$, then Y is a Banach space.

Proof. Statement (*i*) and statement (*ii*): We use the subspace criterion to show that $\mathcal{L}(X,Y)$ is a linear subspace of $L(X,Y)$. The zero operator is bounded, so $\mathcal{L}(X,Y)$ is nonempty. With $A \in \mathcal{L}(X,Y)$, we have $\alpha\, A \in \mathcal{L}(X,Y)$ since

$$\|\alpha A\|_{\mathcal{L}(X,Y)} = \sup_{\|x\|_X=1} \|\alpha A(x)\|_Y = \sup_{\|x\|_X=1} |\alpha|\, \|A(x)\|_Y = |\alpha| \sup_{\|x\|_X=1} \|A(x)\|_Y.$$

This proves the absolute homogeneity of the operator norm. Also, for $A, B \in \mathcal{L}(X,Y)$, we have

$$\begin{aligned} \|A+B\|_{\mathcal{L}(X,Y)} &= \sup_{\|x\|_X=1} \|A(x)+B(x)\|_Y \\ &\leqslant \sup_{\|x\|_X=1} \|A(x)\|_Y + \|B(x)\|_Y \\ &\leqslant \sup_{\|x\|_X=1} \|A(x)\|_Y + \sup_{\|x\|_X=1} \|B(x)\|_Y \\ &= \|A\|_{\mathcal{L}(X,Y)} + \|B\|_{\mathcal{L}(X,Y)} \end{aligned}$$

and thus $A + B \in \mathcal{L}(X,Y)$ and the triangle inequality holds. Finally, $\|A\|_{\mathcal{L}(X,Y)} \geqslant 0$ is clear, and $\|A\|_{\mathcal{L}(X,Y)} = 0$ implies $\|A(x)\|_Y = 0$ for all $x \in X$, and thus $A = 0$, the zero element of $\mathcal{L}(X,Y)$.

Statement (*iii*): Suppose that Y is a Banach space and that $\big(A^{(k)}\big)$ is a Cauchy sequence in $\mathcal{L}(X,Y)$. That is, for every $\varepsilon > 0$, there exists an index k_ε such that $\|A^{(k)} - A^{(\ell)}\| < \varepsilon$ holds for all $k, \ell \geqslant k_\varepsilon$.

Step 1: We construct the candidate $A\colon X \to Y$ for the limit of $A^{(k)}$.

For any fixed $x \in X$, we have

$$\|A^{(k)}(x) - A^{(\ell)}(x)\|_Y = \big\|[A^{(k)} - A^{(\ell)}](x)\big\|_Y \leqslant \|A^{(k)} - A^{(\ell)}\|_{\mathcal{L}(X,Y)}\, \|x\|_X.$$

Therefore, the sequence $\big(A(x)^{(k)}\big)$ is Cauchy in Y. Since Y is complete, we can define the pointwise limit $A(x) \coloneqq \lim_{k\to\infty} A_k(x)$.

Step 2: We show that A is linear.

For any $x, y \in X$ and $\alpha, \beta \in \mathbb{R}$, we have

$$\begin{aligned} A(\alpha\, x + \beta\, y) &= \lim_{k\to\infty} A^{(k)}(\alpha\, x + \beta\, y) && \text{by definition of } A \\ &= \lim_{k\to\infty} \big[\alpha\, A^{(k)}(x) + \beta\, A^{(k)}(y)\big] && \text{by linearity of } A^{(k)} \\ &= \alpha \lim_{k\to\infty} A^{(k)}(x) + \beta \lim_{k\to\infty} A^{(k)}(y) && \text{by linearity of the limit, all limits exist} \\ &= \alpha\, A(x) + \beta\, A(y) && \text{by definition of } A. \end{aligned}$$

Step 3: We show that A is bounded.

For any $x \in X$, we have

$$\begin{aligned} \|A(x)\|_Y &\leqslant \|A(x) - A^{(k)}(x)\|_Y + \|A^{(k)}(x)\|_Y && \text{by the triangle inequality} \\ &\leqslant \|A(x) - A^{(k)}(x)\|_Y + \|A^{(k)}\|_{\mathcal{L}(X,Y)}\, \|x\|_X. \end{aligned}$$

Since every Cauchy sequence is bounded (**Quiz 4.1:** Can you prove it?), we have

$$\leqslant \|A(x) - A^{(k)}(x)\|_Y + C\,\|x\|_X.$$

By letting $k \to \infty$, we find that $\|A(x)\|_Y \leqslant C\,\|x\|_X$, with C independent of x. That is, A is bounded.

Step 4: We show that $A^{(k)} \to A$ in $\mathcal{L}(X, Y)$.

Let $\varepsilon > 0$. Since $\big(A^{(k)}\big)$ is Cauchy, there exists k_ε such that $\|A^{(k)} - A^{(\ell)}\|_{\mathcal{L}(X,Y)} < \varepsilon$ for all $k, \ell \geqslant k_\varepsilon$. Now let $x \in X$ be arbitrary. We estimate

$$\begin{aligned}\|A^{(k)}(x) - A^{(\ell)}(x)\|_Y &\leqslant \|A^{(k)} - A^{(\ell)}\|_{\mathcal{L}(X,Y)}\,\|x\|_X \\ &\leqslant \varepsilon\,\|x\|_X \quad \text{for all } k, \ell \geqslant k_\varepsilon.\end{aligned}$$

Passing to the limit $\ell \to \infty$, we obtain

$$\|A^{(k)}(x) - A(x)\|_Y \leqslant \varepsilon\,\|x\|_X \quad \text{for all } k \geqslant k_\varepsilon.$$

This shows $\|A^{(k)} - A\|_{\mathcal{L}(X,Y)} \leqslant \varepsilon$ for all $k \geqslant k_\varepsilon$, i. e., $A^{(k)} \to A$ in $\mathcal{L}(X, Y)$.

Statement (*iv*): Suppose that $\mathcal{L}(X, Y)$ is a Banach space and $\dim(X) \geqslant 1$. Then there exists a non-zero bounded linear map $f\colon X \to \mathbb{R}$. (**Quiz 4.2:** How do we see this?) In particular, we have $f(x_0) = 1$ for some $x_0 \in X$.

Now define a family $(A_y)_{y\in Y}$ of bounded linear operators $X \to Y$ by

$$A_y(x) \coloneqq f(x)\,y \quad \text{for all } x \in X.$$

Notice that $y \mapsto A_y$ is a linear map $Y \to \mathcal{L}(X, Y)$. Every A_y is indeed bounded since

$$\|A_y(x)\|_Y = |f(x)|\,\|y\|_Y \leqslant \|f\|_{\mathcal{L}(X,\mathbb{R})}\,\|y\|_Y\,\|x\|_X$$

and thus $\|A_y\|_{\mathcal{L}(X,Y)} \leqslant \|f\|_{\mathcal{L}(X,\mathbb{R})}\,\|y\|_Y$. Suppose now that $\big(y^{(k)}\big)$ is a Cauchy sequence in Y. Then

$$\|A_{y^{(k)}} - A_{y^{(\ell)}}\|_{\mathcal{L}(X,Y)} = \|A_{y^{(k)} - y^{(\ell)}}\|_{\mathcal{L}(X,Y)} \leqslant \|f\|_{\mathcal{L}(X,\mathbb{R})}\,\|y^{(k)} - y^{(\ell)}\|_Y$$

and therefore, $A_{y^{(k)}}$ is a Cauchy sequence in $\mathcal{L}(X, Y)$. Since $\mathcal{L}(X, Y)$ is complete, there exists a limit $A \in \mathcal{L}(X, Y)$. But this and Lemma 4.6 imply

$$y^{(k)} = A_{y^{(k)}}(x_0) \to A(x_0) \in Y,$$

and thus $\big(y^{(k)}\big)$ converges, i. e., Y is complete. □

End of Class 8

End of Week 5

§ 4.2 Continuous Embeddings

Definition 4.9 (continuous embedding, isomorphism)**.**
Suppose that X and Y are normed linear spaces.

(i) An injective linear map $A\colon X \to Y$ that is also bounded is said to be a **continuous embedding** of X into Y. In this case, the space X is said to be **continuously embedded** into Y.

(ii) A bijective linear map $A\colon X \to Y$ that is also bounded and whose inverse is bounded is said to be an **isomorphism** of X onto Y. In this case, the spaces X and Y are said to be **isomorphic**.

(iii) An isomorphism $A\colon X \to Y$ such that $\|A(x)\|_Y = \|x\|_X$ for all $x \in X$ is said to be an **isometric isomorphism** or an **isometry** of X onto Y. In this case, the spaces X and Y are said to be **isometric**. △

Remark 4.10 (continuous embedding, isomorphism)**.**

(i) In many cases, $X \subseteq Y$ algebraically as a subspace, and we consider the linear inclusion map $i\colon X \to Y$ with $i(x) = x$, which is clearly injective. Notice that the inclusion map is continuous if and only if $\|x\|_Y = \|i(x)\|_Y \leqslant C\,\|x\|_X$, i. e., if and only if $\|\cdot\|_Y$ is weaker on X than $\|\cdot\|_X$. We denote the continuous embedding of X into Y by $X \hookrightarrow Y$.

(ii) A surjective linear map $A\colon X \to Y$ is an isomorphism if and only if there exist constants $c, C > 0$ such that

$$c\,\|x\|_X \leqslant \|A(x)\|_Y \leqslant C\,\|x\|_X \quad \text{for all } x \in X$$

holds.

(iii) A surjective linear map $A\colon X \to Y$ is an isometry if and only if

$$\|x\|_X = \|A(x)\|_Y \quad \text{for all } x \in X$$

holds.

(iv) Two isomorphic normed linear spaces X and Y cannot be distinguished in terms of their structure, up to the equivalence of norms. Two isometric normed linear spaces X and Y cannot be distinguished at all. △

Example 4.11 (continuous embeddings)**.**
Suppose that $\Omega \subseteq \mathbb{R}^d$ is an open and bounded set. Then we have the following continuous embeddings of Sobolev spaces:

$$\begin{array}{ccccccccc}
L^\infty(\Omega) & \hookrightarrow & \cdots & \hookrightarrow & L^2(\Omega) & \hookrightarrow & \cdots & \hookrightarrow & L^1(\Omega) \\
\uparrow & & & & \uparrow & & & & \uparrow \\
W^{1,\infty}(\Omega) & \hookrightarrow & \cdots & \hookrightarrow & W^{1,2}(\Omega) & \hookrightarrow & \cdots & \hookrightarrow & W^{1,1}(\Omega) \\
\uparrow & & & & \uparrow & & & & \uparrow \\
W^{2,\infty}(\Omega) & \hookrightarrow & \cdots & \hookrightarrow & W^{2,2}(\Omega) & \hookrightarrow & \cdots & \hookrightarrow & W^{2,1}(\Omega) \\
\vdots & & & & \vdots & & & & \vdots
\end{array}$$

The inclusions in horizontal direction rely on the boundedness of Ω, while the inclusions in vertical direction hold for any open set Ω. Moreover, there are further embeddings in "north-westerly" direction due to the Sobolev embedding theorem, which allow differentiability to be traded for higher integrability indices. △

§ 4.3 The Dual Space

Definition 4.12 (algebraic and topological dual spaces)**.**

Suppose that X is a normed linear space.

(*i*) The **algebraic dual space** of X is the linear space

$$X' \coloneqq L(X, \mathbb{R}) \tag{4.5}$$

of all linear maps $X \to \mathbb{R}$, also known as **linear functionals** on X.

(*ii*) The **topological dual space** of X is the linear space

$$X^* \coloneqq \mathcal{L}(X, \mathbb{R}) \tag{4.6}$$

of continuous (bounded) linear functionals on X. △

Clearly, X^* is a linear subspace of X'. It is, in fact a proper subspace, if and only if X is infinite-dimensional (Lemma 4.7). Since $\mathbb{R}$ is complete, X^* is always a Banach space by Theorem 4.8. Since we use the absolute value as the norm on $\mathbb{R}$, the dual space X^* is equipped with the operator norm

$$\|f\|_{X^*} = \sup_{\|x\|_X=1} |f(x)|.$$

Given $f \in X^*$ and $x \in X$, we often use the notation

$$\langle f\,, x\rangle_{X^*,X} \coloneqq f(x).$$

The bracket $\langle \cdot\,, \cdot\rangle_{X^*,X}$ is a bilinear form on $X^* \times X$ and it is called the **dual pairing** of X and X^*. In the future, we will often simply say **dual space** instead of **topological dual space** since we will not use the algebraic dual space much.

Example 4.13 (dual spaces)**.**

(*i*) Suppose that $\Omega \subseteq \mathbb{R}^d$ is an open and bounded set. Moreover, let $p, q \in [1, \infty)$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Then the dual space of $L^p(\Omega)$ is isometrically isomorphic to $L^q(\Omega)$.
In this representation of $L^p(\Omega)^*$, the dual pairing is given by

$$\langle f\,, g\rangle \coloneqq \int_\Omega f\,g \,\mathrm{d}x \quad \text{for } f \in L^p(\Omega) \text{ and } g \in L^q(\Omega). \tag{4.7}$$

(*ii*) The dual space of $L^\infty(\Omega)$ does not have a similarly simple representation. It is isometrically isomorphic to the space of finitely additive signed measures on Ω that are absolutely continuous w.r.t. the Lebesgue measure; see for instance Dunford, Schwartz, 1988, Theorem IV.8.16. △

§ 4.4 The Dual Space of a Hilbert Space

The ability to represent the dual of a normed linear space as concretely as for L^p spaces is a rather special property of a normed linear space. However, it is always possible for Hilbert spaces.

Theorem 4.14 (Riesz representation theorem)**.**
Suppose that H is a Hilbert space. Then the dual space H^* of H is isometrically isomorphic to H itself, via the isomorphism

$$\Phi\colon H \ni u \mapsto (u, \cdot)_H \in H^*. \tag{4.8}$$

Moreover, the norm of H^* (i. e., the operator norm of $f \in \mathcal{L}(H, \mathbb{R})$) is induced by the inner product

$$(f, g)_{H^*} := \big(\Phi^{-1}(f), \Phi^{-1}(g)\big)_H = \langle f\,, \Phi^{-1}(g)\rangle_{H^*,H} = \langle g\,, \Phi^{-1}(f)\rangle_{H^*,H}. \tag{4.9}$$

Proof. We break the proof down into several steps.

Step 1: We show that $\Phi\colon H \to H'$ is linear.

First of all, $\Phi(u) \in H'$ for all $u \in H$ since $\Phi(u) = (u, \cdot)_H$ and the inner product is linear in the second argument.

For $u, v, w \in H$ and $\alpha, \beta \in \mathbb{R}$, we have

$$\begin{aligned} \langle \Phi(\alpha\, u + \beta\, v)\,, w\rangle &= (\alpha\, u + \beta\, v, w)_H && \text{by definition of } \Phi \\ &= \alpha\,(u, w)_H + \beta\,(v, w)_H && \text{by linearity of the inner product in the first argument} \\ &= \alpha\,\langle \Phi(u)\,, w\rangle + \beta\,\langle \Phi(v)\,, w\rangle && \text{by definition of } \Phi. \end{aligned}$$

This shows $\Phi(\alpha\, u + \beta\, v) = \alpha\,\Phi(u) + \beta\,\Phi(v)$, so Φ is linear.

Step 2: We show that $\Phi\colon H \to H^*$ holds.

For $u \in H$ and $v \in H$, we have

$$|\langle \Phi(u)\,, v\rangle| = |(u, v)_H| \le \|u\|_H\,\|v\|_H \quad \text{by the Cauchy-Schwarz inequality.}$$

Therefore, $\Phi(u)$ is a bounded linear functional on H with $\|\Phi(u)\|_{H^*} \le \|u\|_H$.

Step 3: We show that $\|\Phi(u)\|_{H^*} = \|u\|_H$ for all $u \in H$.

For $u \in H$, we have

$$|\langle \Phi(u)\,, u\rangle| = |(u, u)_H| = (u, u)_H = \|u\|_H^2,$$

which shows $\|\Phi(u)\|_{H^*} \ge \|u\|_H$.

Step 4: We show that Φ is surjective.[15] (By Remark 4.10 *(iii)* this implies that Φ is an isometric isomorphism.)

Suppose that $f \in H^*$ is given. When $f = 0$, we can simply choose $u = 0$ since $\Phi(0) = 0$, which holds for any linear map. Now suppose $f \ne 0$. Consider the kernel (nullspace) of f,

$$\ker(f) := \{u \in H \mid f(u) = 0\}.$$

[15] This is the main step in the proof, where the completeness of H is crucial.

It is not difficult to see that $\ker(f)$ is a closed subspace of H, and it is not equal to H since $f \neq 0$. One can show that, as a consequence, there exists $v \in H$ such that $f(v) \neq 0$ that is orthogonal to $\ker(f)$.[16] Without loss of generality, we can assume that $\|v\|_H = 1$.

We now choose $u \coloneqq f(v)\, v$ and show $\Phi(u) = f$, so that Φ is surjective. Indeed, we have

$$\|u\|_H = \|f(v)\, v\|_H = |f(v)|\, \|v\|_H = |f(v)|$$

and

$$f(u) = f(f(v)\, v) = f(v)\, f(v) = |f(v)|^2 = \|u\|_H^2.$$

For any $w \in H$, this implies

$$\begin{aligned}
\langle \Phi(u)\, , w\rangle &= (u, w)_H && \text{by definition of } \Phi \\
&= \Big(u, w - \frac{f(w)}{\|u\|_H^2}\, u\Big)_H + \Big(u, \frac{f(w)}{\|u\|_H^2}\, u\Big)_H \\
&= \Big(u, w - \frac{f(w)}{\|u\|_H^2}\, u\Big)_H + \frac{f(w)}{\|u\|_H^2}(u, u)_H \\
&= \Big(u, w - \frac{f(w)}{\|u\|_H^2}\, u\Big)_H + f(w).
\end{aligned}$$

The second factor in the inner product belongs to $\ker(f)$, since

$$\begin{aligned}
f\Big(w - \frac{f(w)}{\|u\|_H^2}\, u\Big) &= f(w) - \frac{f(w)}{\|u\|_H^2} f(u) && \text{by linearity of } f \\
&= f(w) - \frac{f(w)}{\|u\|_H^2} \|u\|_H^2 && \text{since } f(u) = \|u\|_H^2 \\
&= 0.
\end{aligned}$$

But since v is orthogonal to $\ker(f)$, so is $u = f(v)\, v$. This proves

$$\langle \Phi(u)\, , w\rangle = f(w) \quad \text{for all } w \in H,$$

whence $\Phi(u) = f$.

Step 5: We show that (4.9) defines an inner product that induces the norm of H^*.

First of all, we have by definition of Φ and the symmetry of $(\cdot, \cdot)_H$ that

$$\langle f\, , \Phi^{-1}(g)\rangle_{H^*,H} = \big(\Phi^{-1}(f), \Phi^{-1}(g)\big)_H = \big(\Phi^{-1}(g), \Phi^{-1}(f)\big)_H = \langle g\, , \Phi^{-1}(f)\rangle_{H^*,H}$$

and so the equalities in (4.9) hold. Defining now

$$(f, g)_{H^*} \coloneqq \big(\Phi^{-1}(f), \Phi^{-1}(g)\big)_H$$

and the linearity of Φ^{-1} then show that $(\cdot, \cdot)_{H^*}$ is a symmetric bilinear form on H^*. it is also positive definite since Φ^{-1} is a bijection. □

End of Class 9

[16] The proof would require more machinery, including the parallelogram identity for inner products and subsequently the existence of orthogonal projections onto closed and convex subsets (in particular, onto closed subspaces) in Hilbert spaces.

§ 5 Existence Theorems for Global Minimizers

In this section we will discuss sufficient conditions for minimizers of optimization problems in normed linear spaces to exist. We begin with the well known

Theorem 5.1 (Weierstrass extreme value theorem)**.**
Suppose that V is a normed linear space and $K \subseteq V$ is compact. Moreover, suppose that $f : K \to \mathbb{R}$ is continuous. Then $f(K) \subseteq \mathbb{R}$ is compact. As a consequence, f attains its minimum (and its maximum) on K.

Proof. We will show that $f(K)$ is sequentially compact, which is equivalent to compactness due to Theorem 2.17. Suppose that $\big(r^{(k)}\big)$ is a sequence in $f(K)$. That is, there exists a sequence $\big(x^{(k)}\big)$ in K such that $f(x^{(k)}) = r^{(k)}$. Since K is compact, there exists a subsequence $\big(x^{(k^{(\ell)})}\big)$ such that $x^{(k^{(\ell)})} \to x^* \in K$ as $\ell \to \infty$. Due to the continuity of f (Definition 4.1), $f\big(x^{(k^{(\ell)})}\big) \to f(x^*)$, and since $x^* \in K$, we have $f(x^*) \in f(K)$. This shows that $f(K)$ is sequentially compact.

As a compact set, $f(K) \subseteq \mathbb{R}$ is closed and bounded, i. e., $\inf\{f(x) \mid x \in K\}$ and $\sup\{f(x) \mid x \in K\}$ are finite. Due to the closedness, inf and sup are actually attained. □

So Weierstrass' theorem is the same as in $V = \mathbb{R}^n$. However, it is rarely applicable in infinite-dimensional normed linear spaces V, because the choice of compact subsets $K \subseteq V$ is quite limited. This is hinted at by the fact that even unit balls in infinite-dimensional normed linear spaces are not compact (Theorem 2.18). For instance, in $L^p(\Omega)$, one can precisely characterize the compact subsets.

Theorem 5.2 (compact subsets of $L^p(\Omega)$, **Kolmogorov-Riesz theorem**)**.**
Suppose that $\Omega \subseteq \mathbb{R}^d$ is an open and bounded set and $K \subseteq L^p(\Omega)$. Then the following statements are equivalent:

(i) K is compact in $L^p(\Omega)$.

(ii) K is closed, bounded and **equicontinuous**.

For a proof, see for instance Adams, Fournier, 2003, Theorem 2.32. The definition of equicontinuity makes use of the shift-operator $\tau_h : L^p(\Omega) \to L^p(\Omega)$ for $h \in \mathbb{R}^n$, defined by $f \mapsto \tau_h f := f(\cdot + h)\, \chi_\Omega$.[17] **Equicontinuity** means that for any $\varepsilon > 0$, there exists $\delta > 0$ such that $\|\tau_h f - f\|_{L^p(\Omega)} < \varepsilon$ for all $f \in K$ and all $h \in \mathbb{R}^d$ with $\|h\|_2 < \delta$.

Example 5.3 (non-compactness of a L^p-functions with bound constraints)**.**
Suppose that $\Omega \subseteq \mathbb{R}^d$ is an open and bounded set. Moreover, let $a, b \in \mathbb{R}$ be such that $a < b$. Then the set

$$A := \{f \in L^p(\Omega) \mid a \leqslant f(x) \leqslant b \text{ for a.a. } x \in \Omega\} \tag{5.1}$$

is closed and bounded in $L^p(\Omega)$, but it not compact.

[17] It is easy to see that τ_h indeed maps $L^p(\Omega)$ into itself and has operator norm $\leqslant 1$.

To see this, we discuss for simplicity the case where $\Omega = (0,1) \subseteq \mathbb{R}$ is an open and bounded interval. Consider the sequence $\big(f^{(k)}\big)$ defined by

$$f^{(k)}(x) := \begin{cases} 0 & \text{if the } k\text{-th binary digit (after the decimal) of } x \text{ is } 0, \\ 1 & \text{if the } k\text{-th binary digit (after the decimal) of } x \text{ is } 1. \end{cases}$$

In other words, $f^{(k)}$ is the characteristic function of a union of disjoint intervals of length 2^{-k-1}. Then we have

$$\|f^{(k)} - f^{(\ell)}\|_{L^p(\Omega)}^p = \frac{1}{2} \quad \text{for all } k \neq \ell.$$

Therefore, no subsequence of $\big(f^{(k)}\big)$ is a Cauchy sequence. △

We would need to add further conditions to the functions in (5.1) to obtain a compact subset of $L^p(\Omega)$. Some possibilities are monotonicity (for $\Omega \subseteq \mathbb{R}$), convexity or concavity, or additional smoothness (such as $f \in W^{1,1}(\Omega)$).

The following example is a demonstration that global minimizers may fail to exist in infinite-dimensional normed linear spaces in the absence of compactness.

Example 5.4 (non-existence of global minimizers[18]).

On $\Omega = \mathbb{R}$, consider the function $g \in L^2(\Omega)$ defined by $g(x) := \exp(-x^2)$ and the problem

$$\begin{aligned} \text{Minimize} \quad & J(f) := \int_\Omega f(x)\, g(x) \;\mathrm{d}x \\ \text{subject to} \quad & f \geqslant 0 \quad \text{a.e. in } \Omega \\ \text{and} \quad & \|f\|_{L^2(\Omega)} = 1. \end{aligned}$$

This problem has the feasible set

$$F := \{f \in L^2(\Omega) \mid f \geqslant 0 \text{ a.e. in } \Omega \text{ and } \|f\|_{L^2(\Omega)} = 1\},$$

which is not compact. The objective $f \mapsto J(f)$ is continuous on $L^2(\Omega)$ (**Quiz 5.1:** Why?) and bounded below by 0. In fact, for all $f \in F$, we have $J(f) > 0$.

Considering the sequence of characteristic functions $f^{(k)} = \chi_{[k,k+1]}$ shows

$$J(f^{(k)}) = \int_k^{k+1} \exp(-x^2) \;\mathrm{d}x \longrightarrow 0 \quad \text{as } k \longrightarrow \infty.$$

Therefore, the infimum of J on F is 0, but it is not attained. △

As a remedy, we may resort to a different topology on normed linear spaces. Broadly speaking, when we have fewer open sets and thus fewer open covers of a set, we have a better chance of compactness.

[18] example communicated by Gerd Wachsmuth (BTU Cottbus)

§ 5.1 The Weak Topology on a Normed Linear Space

Definition 5.5 (weakly open sets, weakly convergent sequences)**.**

Suppose that V is a normed linear space.

(*i*) A set $U \subseteq V$ is said to be **weakly open** if for all $x \in U$ there exist $\varepsilon > 0$, $n \in \mathbb{N}$ and $f_1, \ldots, f_n \in V^*$ such that

$$\left\{ y \in V \,\middle|\, |\langle f_i \, , y - x \rangle| < \varepsilon \text{ for } i = 1, \ldots, n \right\} \subseteq U. \tag{5.2}$$

(*ii*) A sequence $\left(x^{(k)}\right)$ in V is said to be **weakly convergent** to $x \in V$ if for all $f \in V^*$, we have

$$\lim_{k \to \infty} \langle f \, , x^{(k)} \rangle = \langle f \, , x \rangle.$$

In this case we write $x^{(k)} \rightharpoonup x$.

$\triangle$

The collection of weakly open sets in V is called the **weak topology** on $(V, \|\cdot\|_V)$. For a clearer distinction, we may refer to the norm topology on V as the **strong topology**. Similarly, we may speak of **strongly convergent sequences**.

One can show that the weak limit of a sequence is unique.

Theorem 5.6 (weak topology in finite-dimensional normed linear spaces)**.**

Suppose that V is a finite-dimensional normed linear space. Then the weak topology on V coincides with the strong topology.

Proof. We will show below in Theorem 5.8 that every weakly open set in V is open in the strong topology. Therefore, we only need to show that every strongly open set is weakly open. So suppose that $U \subseteq V$ is strongly open and $x \in U$. Then there exists $r > 0$ such that $B_r(x) \subseteq U$. By Definition 5.5, we need to show that there exist $\varepsilon > 0$, $n \in \mathbb{N}$ and $f_1, \ldots, f_n \in V^*$ such that (5.2)

$$U' \coloneqq \left\{ y \in V \,\middle|\, |\langle f_i \, , y - x \rangle| < \varepsilon \text{ for } i = 1, \ldots, n \right\} \subseteq U$$

holds.

Suppose that $\{v^{(1)}, \ldots, v^{(n)}\}$ is a basis of V and that $x = \sum_{i=1}^n x_i \, v^{(i)}$. We denote by f_i the coordinate map $V \ni x \mapsto x_i \in \mathbb{R}$, which is linear and, thanks to the finite dimensionality of V, continuous

(Lemma 4.7). For any $y \in V$, we find

$$\begin{aligned}
\|y - x\|_V &= \Big\| \sum_{i=1}^n (y_i - x_i)\, v^{(i)} \Big\|_V \\
&= \Big\| \sum_{i=1}^n \langle f_i\,, y - x\rangle\, v^{(i)} \Big\|_V \\
&\leqslant \sum_{i=1}^n \big\| \langle f_i\,, y - x\rangle\, v^{(i)} \big\|_V \\
&\leqslant \sum_{i=1}^n |\langle f_i\,, y - x\rangle| \max\big\{\|v^{(i)}\|_V \,\big|\, i = 1, \ldots, n\big\} \\
&= C \sum_{i=1}^n |\langle f_i\,, y - x\rangle|.
\end{aligned}$$

Consequently, when we choose

$$U' \coloneqq \big\{ y \in V \,\big|\, |\langle f_i\,, y - x\rangle| < \varepsilon \text{ for } i = 1, \ldots, n \big\}$$

with $\varepsilon \coloneqq \frac{r}{C\,n}$, then we have $U' \subseteq B_r(x) \subseteq U$. □

Note: In particular, the strong and weak topologies on $\mathbb{R}$ coincide. In general, the finite dimension is sufficient, but not necessary for the weak and strong topologies to coincide. A prominent example if the space ℓ^1 of absolutely summable sequences.

Remark 5.7 (weak topology)**.**

(i) The norm on V enters Definition 5.5 only through the dual space V^*. (Recall that the norm determines which linear functionals are continuous.)

(ii) When $\|\cdot\|_a$ and $\|\cdot\|_b$ are equivalent norms on V, then both induce the same weak topology on V.

(iii) The weak topology is not, in general, induced by a norm. Therefore, there is in general no notion of “distance” in the weak topology.

(iv) Our definition of weakly convergent sequences is compatible with the general notion of convergence in topological spaces, i. e., for all weakly open neighborhoods U of the limit x, there exists $k_0 \in \mathbb{N}$ such that $x^{(k)} \in U$ for all $k \geqslant k_0$. △

Theorem 5.8 (relation between the weak and strong topologies)**.**
Suppose that V is a normed linear space.

(i) Every weakly open set in V is open in the strong topology.

(ii) Every strongly convergent sequence is weakly convergent (to the same limit).

(iii) Suppose that $f\colon V \longrightarrow \mathbb{R}$ is **weakly continuous**, i. e., continuous in the weak topology.[19] Then f is continuous in the strong topology as well.

[19] This means that pre-images of (weakly) open sets in $\mathbb{R}$ are weakly open in V.

Proof. Statement (i): Suppose that $U' \subseteq V$ is weakly open and $x \in U'$. Then there exist $\varepsilon > 0$, $n \in \mathbb{N}$ and $f_1, \ldots, f_n \in V^*$ such that (5.2) holds. We set

$$r := \min\left\{\frac{\varepsilon}{2\,\|f_i\|_{X^*} + 1} \,\middle|\, i = 1, \ldots, n\right\}. \tag{5.3}$$

Then we have for $y \in B_r(x)$:

$$\begin{aligned}
|\langle f_i\,, y - x\rangle|_X &\leqslant \|f_i\|_{X^*}\,\|y - x\|_X \\
&\leqslant \|f_i\|_{X^*}\, r \\
&\leqslant \frac{\varepsilon\,\|f_i\|_{X^*}}{2\,\|f_i\|_{X^*} + 1} \\
&\leqslant \frac{\varepsilon}{2}.
\end{aligned}$$

This implies $B_r(x) \subseteq U'$, so U' is open in the strong topology.

Statement (ii): Suppose that $\|x^{(k)} - x\|_V \to 0$ as $k \to \infty$. When $f \in X^*$, then this implies

$$\langle f\,, x^{(k)} - x\rangle \leqslant \|f\|_{X^*}\,\|x^{(k)} - x\|_X \to 0 \quad \text{as } k \to \infty,$$

so $\lim_{k\to\infty}\langle f\,, x^{(k)}\rangle = \langle f\,, x\rangle$, which means $x^{(k)} \rightharpoonup x$.

Statement (iii): Suppose that $U \subseteq \mathbb{R}$ is open. Then $f^{-1}(U)$ is weakly open, so $f^{-1}(U)$ is open in the strong topology as well. This means that f is strongly continuous. □

Lemma 5.9 ((weak) continuity of linear functionals)**.**
Suppose that V is a normed linear space and $f\colon V \to \mathbb{R}$ a linear functional, i. e., $f \in L(V, \mathbb{R})$. Then the following statements are equivalent:

(i) f is continuous in the strong topology, i. e., $f \in \mathcal{L}(V, \mathbb{R}) = V^*$.

(ii) f is weakly continuous.

Remark 5.10 (further properties)**.**
Suppose that V is a normed linear space.

(i) The weak topology on V is the weakest topology that makes all linear functionals in V^* continuous.

(ii) Weakly convergent sequences are bounded.[20]

(iii) Suppose that $x^{(k)} \rightharpoonup x$ in V and $f^{(k)} \to f$ in V^*. Then $\langle f^{(k)}\,, x^{(k)}\rangle \to \langle f\,, x\rangle$. △

End of Class 10

End of Week 6

[20]This follows from the Banach-Steinhaus theorem (uniform boundedness principle).

Chapter 1 Convex Infinite-Dimensional Optimization with Applications in Imaging

Chapter 2 Optimal Control of Partial Differential Equations

Bibliography

Adams, R.; J. Fournier (2003). *Sobolev Spaces*. 2nd ed. New York: Academic Press.

Dunford, N.; J. T. Schwartz (1988). *Linear Operators. Part I: General Theory*. Wiley Classics Library. John Wiley & Sons, Inc., New York.

Meyers, N.; J. Serrin (1964). "H=W". *Proceedings of the National Academy of Sciences* 51, pp. 1055–1056. DOI: 10.1073/pnas.51.6.1055.

Rudin, W. (1987). *Real and Complex Analysis*. McGraw-Hill.