Lecture Notes
Infinite Dimensional Optimization

Roland Herzog*

2024-10-28

*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 $C \subseteq X$.

[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:

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

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. △

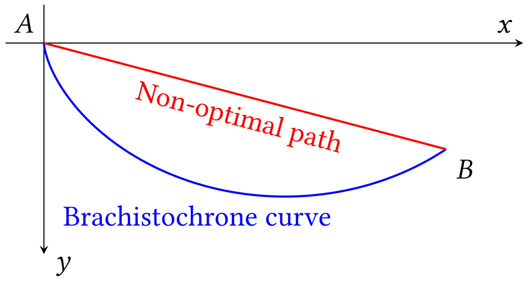

Figure 1.1: Some non-optimal curve $y\colon [0, b_1] \to \mathbb{R}$ from A to B (left) as well as the unique global minimizer of the Brachistochrone problem (1.1), given by the segment of a cycloid (left). Image of an experimental device on display at Technoseum Mannheim (right), shot by Roland Herzog.

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. △

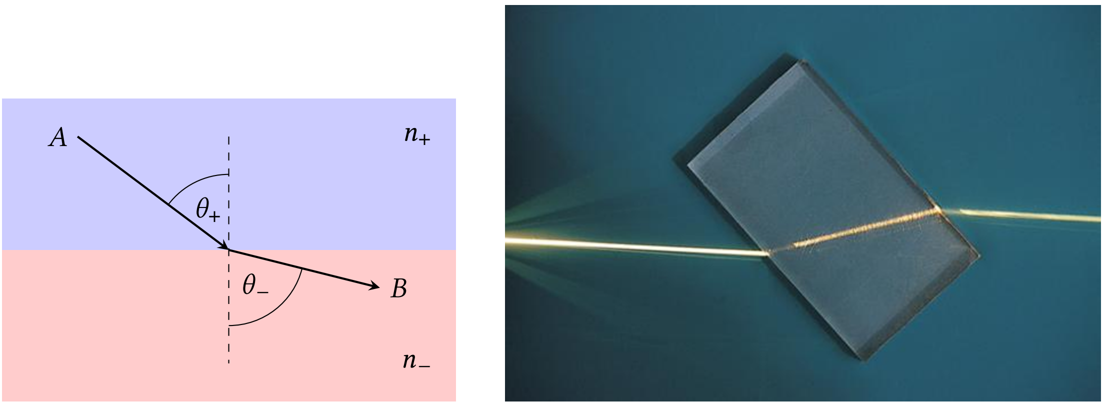

Figure 1.2: Illustratrion of Snell's law of refraction (left) as a special case of Example 1.3. Image (right) obtained from `https://en.wikipedia.org/wiki/Refraction`, released into the public domain by creator ajizai.

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

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

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. △

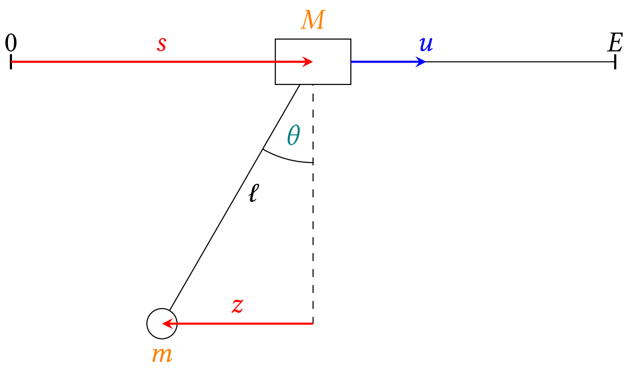

Figure 1.3: Illustration of the crane trolley problem (Example 1.5).

[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]

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

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 $f\colon V \to \mathbb{R}$ or, more generally, $f\colon V \to W$, we will require linear spaces to be normed.

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

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

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

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

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

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

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

[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 $\|x^{(k)} - x\| \longrightarrow 0$ in $\mathbb{R}$. We then write $x^{(k)} \longrightarrow x$ 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, $x^{(k)} \longrightarrow x$ 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 $x^{(k)} \longrightarrow x$.

(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 $(x^{(k)})$ is a sequence in E converging to some $x \in V$. Then this sequence is a Cauchy sequence in E. Since E is complete, $(x^{(k)})$ 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/more open balls.)

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

(*viii*) If $(x^{(k)})$ is a Cauchy sequence in the stronger norm $\|\cdot\|_b$, then $(x^{(k)})$ 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 $x \mapsto \|x\|_\infty := \left\| \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right\|_\infty = \max\{|x_1|, \dots, |x_n|\}$ 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:

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

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

Suppose that this is not the case. Then there exists a sequence $(x^{(k)})$ 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 $\{x_j^{(k)} \mid k \in \mathbb{N}\}$ 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 $|x_{j_0}^{(k^{(\ell)})}| = 1$ 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

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

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 $\|x\|_\infty := \max\{|x_1|, \ldots, |x_n|\}$ 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. □

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 $x^{(1)}, \ldots, x^{(N)} \in K$ such that $\{B_\varepsilon(x^{(1)}), \ldots, B_\varepsilon(x^{(N)})\}$ 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 $x^{(1)}, \ldots, x^{(N)} \in K$. As we noticed above, $B_{\varepsilon_{x^{(j)}}}(x^{(j)})$ and $B_{\varepsilon_{x^{(j)}}}(z)$ are disjoint for all $j = 1, \ldots, N$. Let $\varepsilon := \min\{\varepsilon_{x^{(1)}}, \ldots, \varepsilon_{x^{(N)}}\}$. 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,

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

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 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 $\big(x^{(n)}\big)$ is a sequence in K that does not possess a convergent subsequence with limit in K. In other words, $\big(x^{(n)}\big)$ 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

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

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 $\big(x^{(n)}\big)$ 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 $\big(x^{(k)}\big)$, 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 $\big(U_i\big)_{i\in I}$ 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 $\big(U_i\big)_{i\in I}$. (Otherwise, K itself could be covered by a finite subfamily of $\big(U_i\big)_{i\in I}$, 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 $\big(U_i\big)_{i\in I}$. (Otherwise, $B_1(x^{(0)}) \cap K$ itself could be covered by a finite subfamily of $\big(U_i\big)_{i\in I}$, 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 $\big(U_i\big)_{i\in I}$. The centers satisfy $x^{(k+1)} \in B_{2^{-k}}(x^{(k)}) \cap K$. Therefore, the sequence $\big(x^{(k)}\big)$ 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 $\big(U_i\big)_{i\in I}$, 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 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 $R \coloneqq \inf\{\|v - y\| \mid y \in Y\}$. 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

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

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

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

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

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

End of Class 4

Proof of Theorem 2.18:
Item $(i) \Rightarrow$ 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 $Y \coloneqq \operatorname{span}\{y^{(1)}, \ldots, y^{(N)}\}$. 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 $\operatorname{dist}_Y(x_\theta) \geqslant \theta \coloneqq \frac{3}{4}$. 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 $(ii) \Rightarrow$ item (iii): The proof is the same as above.

Item $(iii) \Rightarrow$ 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 $x \in \overline{B_1(0)}$, 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. We will state results in this subsection without proof.

Definition 2.21 (Lebesgue spaces)**.**
Suppose that $\Omega \subset \mathbb{R}^n$ 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 $\Omega \subset \mathbb{R}^n$ 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

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

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

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

(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 $B_1(0) \subset \mathbb{R}^d$, the function $x \mapsto |x|^\alpha$ belongs to $L^p(B_1(0))$ if and only if $\alpha\, p > -d$ holds.[13] $\triangle$

Lemma 2.24 (Hölder's inequality)**.**
Suppose that $\Omega \subset \mathbb{R}^d$ 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 $\Omega \subset \mathbb{R}^d$ is an open and bounded set. For $1 \leqslant p \leqslant q \leqslant \infty$, the space $L^q(\Omega)$ is a subspace of $L^p(\Omega)$. 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 $1/q$ (for $q < \infty$) or as 0 (for $q = \infty$).

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 $\Omega \subset \mathbb{R}^d$ 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

Chapter 1 Convex Infinite-Dimensional Optimization with Applications in Imaging

Chapter 2 Optimal Control of Partial Differential Equations

Bibliography

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