# A “crash course” in nonsmooth convex optimization

Zev Woodstock
woodstock(at)zib.de*

## 1 Introduction

These notes are supplementary material to a “crash course” I am teaching in May of 2023. The topic is *proximity operators and nonsmooth convex optimization*. These notes are not meant to be used as a standalone resource. Please cite peer-reviewed material. The reference book for this class is *Convex Analysis and Monotone Operator Theory*, 2nd edition, by Heinz H. Bauschke and Patrick L. Combettes, published by Springer.

If unspecified, $\mathcal{H}$ is a real finite-dimensional vector space in Section 1 and a real finite-dimensional Hilbert space from Sections 2 onward (e.g., $\mathbb{R}^n$ with the Euclidean inner product is fine). While this class sticks to finite dimensions, virtually all of these results also apply to real (infinite-dimensional) Hilbert spaces, modulo minor adjustments detailed in the class book.

### 1.1 Optimization terminology and the extended real line

**Notation 1.1** We will work with the **extended real line**, i.e., $[-\infty, +\infty] := \mathbb{R} \cup \{-\infty, +\infty\}$. Algebra on this field follows most “natural” rules one could expect (e.g., for $x \in \mathbb{R}$, $x + \infty = \infty$). However, the following quantities are **undefined**:

- Any subtraction of infinities: “$+\infty - (+\infty)$”
- Zero times infinity: “$0 \cdot (\pm\infty)$”
- Any quotient of infinities: “$\pm\infty / \pm\infty, \pm\infty / \mp\infty, \ldots$”

As a result, if we work with extended-real-valued functions, we must be sure to avoid anything which is undefined (e.g., the objective function $f(x) + g(x)$ could be undefined if there exists $z$ such that $g(z) = -\infty$ and $f(z) = \infty$.)

*Please report typos/errors found in these notes. Homework solutions should be handed in to my office ZIB 3107.

**Definition 1.2** Given a real vector space $\mathcal{H}$, a function $f\colon \mathcal{H} \to [-\infty, +\infty]$, and a set $C \subset \mathcal{H}$, consider the following optimization problem.

$$\underset{x\in C}{\text{minimize}} \quad f(x) \tag{1}$$

We call $f$ the **objective function**. We call $C$ a **constraint**. For any $x \in C$, we say $x$ is **feasible**. Otherwise, for $x \in \mathbb{R}^n \setminus C$, $x$ is infeasible. If a point $x^* \in C$ satisfies

$$(\forall x \in C) \quad f(x^*) \le f(x), \tag{2}$$

we call $x^*$ a **solution** to the optimization problem (1).

For this class, we consider minimization; to maximize $f$, just use the objective function $-f$.

**Definition 1.3** For $I \subset [-\infty, +\infty]$, $a \in [-\infty, +\infty]$ is a **lower bound** (**upper bound**) if, for every $\xi \in I$, $a \le \xi$ ($a \ge \xi$). The **greatest lower bound**, or **infimum**, of the set $I$ is denoted $\inf I$. Analogously, the **least upper bound**, or **supremum**, of the set $I$ is denoted $\sup I$. In general, $\inf I, \sup I \in [-\infty, +\infty]$. If, additionally, $\inf I \in I$ ($\sup I \in I$), we call it the **minimum** (**maximum**), and denote it $\min I$ ($\max I$). In these cases, we say the infimum (supremum) is *attained*.

A few things to mention:

(i) For $I \neq \varnothing$, we have $\inf I \le \sup I$. For the empty set, $\inf \varnothing = +\infty$ and $\sup \varnothing = -\infty$.

(ii) While the $\inf$ and $\sup$ are always defined, $\max$ and $\min$ may not exist (e.g., consider $I = (0, 1)$ has $\inf I = 0$ and $\sup I = 1$. However, since $0, 1 \notin I$, neither $\max I$ nor $\min I$ exist.)

(iii) Let $f\colon \mathbb{R}^n \to [-\infty, +\infty]$. We adopt the notation that $\inf_{x\in C} f(x) = \inf\{f(x) \mid x \in C\}$.

(iv) It is common in optimization literature to abuse notation, and use

$$\min_{x\in C} f(x) \tag{3}$$

to describe the optimization problem (1). Technically, $\min_{x\in C} f(x)$ is not an optimization problem – it is the optimal value of the objective function at a solution, which may or may not exist.

The following theorem is often used as a tool to ensure that a solution to an optimization problem exists. Regretfully, this class does not have enough time to detail the topics of compact/closed/lsc. However, since the following theorem is referenced a few times in the class, I will provide its statement here.[1]

**Theorem 1.4 (Weierstraß)** *Let $f\colon \mathcal{H} \to [-\infty, +\infty]$ be lower semicontinuous and let $C$ be a compact subset of $\mathcal{H}$. Suppose that $C \cap \operatorname{dom} f \neq \varnothing$. Then $f$ achieves its infimum over $C$.*

[1] Write me if you are interested in learning more about existence of solutions to optimization problems! For unbounded problems, analytic notions of "coercivity" and "recession cones" can also yield existence results.

**Definition 1.5** Let $f\colon \mathcal{H} \to [-\infty, +\infty]$. We will use the following terms.

(i) The **domain** of $f$ is

$$\operatorname{dom} f = \{x \in \mathcal{H} \mid f(x) < +\infty\} \tag{4}$$

(ii) The **epigraph** of $f$ is

$$\operatorname{epi} f = \{(x, \xi) \in \mathcal{H} \times \mathbb{R} \mid f(x) \leq \xi\} \tag{5}$$

(iii) The function $f$ is **proper** if $\operatorname{dom} f \neq \varnothing$ and it never outputs the value $-\infty$ (i.e., $-\infty \notin f(\mathcal{H})$).

(iv) The function $f$ is **lower semicontinuous** (sometimes abbreviated "lsc") at $x \in \mathcal{H}$ if, for every sequence $(x_n)_{n\in\mathbb{N}}$ satisfying $x_n \to x$, we have $f(x) \leq \liminf f(x_n)$

For this class, we will predominantly consider proper and lsc functions. A few things to note about the lsc assumption: (1) every continuous function is lsc, and (2) lower semicontinuity basically allows for a jump-discontinuity to occur at $x \in \mathcal{H}$, but requires that $f$ takes the lowest possible limiting value at $x$ (cf. the figures drawn in class, or here[2]).

## 1.2 Inner product and norms

**Definition 1.6** Let $\mathcal{H}$ be a real finite-dimensional vector space. A **scalar product** (sometimes called **inner product**) is a function $\langle \cdot \mid \cdot \rangle\colon \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ which satisfies the following properties.

(i) $(\forall x \in \mathcal{H} \setminus \{0\}) \quad \langle x \mid x \rangle > 0$

(ii) $(\forall x, y \in \mathcal{H}) \quad \langle x \mid y \rangle = \langle y \mid x \rangle$

(iii) $(\forall x, y, z \in \mathcal{H})(\forall \alpha \in \mathbb{R}) \quad \langle \alpha x + y \mid z \rangle = \alpha \langle x \mid z \rangle + \langle y \mid z \rangle$

**Exercise 1.7** Let $\mathbf{0} \in \mathcal{H}$ be the zero element of $\mathcal{H}$. Show that, for every $x \in \mathcal{H}$, $\langle \mathbf{0} \mid x \rangle = 0$.

**Exercise 1.8** Consider $\mathcal{H} = \mathbb{R}^n$. For two vectors $x, y \in \mathbb{R}^n$, the *dot product* is given by $\langle x \mid y \rangle = x^\top y$. Show that the dot product on $\mathbb{R}^n$ is a scalar product.

**Exercise 1.9** Consider the vector space of matrices $\mathbb{R}^{n\times n}$. For two matrices $A = (a_{i,j})_{1\leq i,j\leq n}$ and $B = (b_{i,j})_{1\leq i,j\leq n}$, the *Frobenius inner product* is given by

$$\langle A \mid B \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j} b_{i,j} \tag{6}$$

Show (6) is an inner product.

[2] https://en.wikipedia.org/wiki/Semi-continuity#/media/File:Lower_semi.svg

**Proposition 1.10 (Cauchy-Schwarz)** *For every* $x, y \in \mathcal{H}$,

$$\langle x \mid y \rangle^2 \leq \langle x \mid x \rangle \langle y \mid y \rangle. \tag{7}$$

*Proof.* If $y = 0$, (7) holds. Now suppose that $y \neq 0$. By Definition 1.6, $\langle y \mid y \rangle > 0$. Set $\alpha = \langle x \mid y \rangle / \langle y \mid y \rangle$. First, we find

$$\begin{aligned} 0 &\leq \langle x - \alpha y \mid x - \alpha y \rangle & (8)\\ &= \langle x \mid x \rangle - 2\alpha \langle x \mid y \rangle + \alpha^2 \langle y \mid y \rangle & (9)\\ &= \langle x \mid x \rangle - 2\alpha \langle x \mid y \rangle + \alpha \langle x \mid y \rangle & (10)\\ &= \langle x \mid x \rangle - \alpha \langle x \mid y \rangle. & (11) \end{aligned}$$

Rearranging the inequality, we find that

$$\begin{aligned} & \frac{\langle x \mid y \rangle^2}{\langle y \mid y \rangle} = \alpha \langle x \mid y \rangle \leq \langle x \mid x \rangle & (12)\\ \Leftrightarrow \quad & \langle x \mid y \rangle^2 \leq \langle y \mid y \rangle \langle x \mid x \rangle. & (13) \end{aligned}$$

□

**Definition 1.11** Let $\mathcal{H}$ be a real finite-dimensional vector space. A function $\| \cdot \| \colon \mathcal{H} \to \mathbb{R}$ is a **norm** if the following hold.

(i) $(\forall x \in \mathcal{H}) \quad \|x\| = 0 \Rightarrow x = 0$

(ii) $(\forall x, y \in \mathcal{H}) \quad \|x + y\| \leq \|x\| + \|y\|$

(iii) $(\forall x \in \mathcal{H})(\forall \alpha \in \mathcal{H}) \quad \|\alpha x\| = |\alpha| \|x\|$

A norm is a way to measure magnitude of vectors, or the distance from one vector to another $\|x - y\|$.

**Exercise 1.12** Let $\mathcal{H}$ be a real finite-dimensional vector space, and let $\langle \cdot \mid \cdot \rangle$ be a scalar product on $\mathcal{H}$. Show that the norm defined by

$$\| \cdot \| \colon \mathcal{H} \to \mathbb{R} \colon x \mapsto \sqrt{\langle x \mid x \rangle} \tag{14}$$

satisfies the properties in Definition 1.11.

The **Euclidean norm** on $\mathbb{R}^n$, given by $(\xi_1, \ldots, \xi_n) \mapsto \sqrt{\xi_1^2 + \cdots + \xi_n^2}$, arises from the dot product. Exercise 1.12 yields the following formulation of the Cauchy-Schwarz inequality

$$(\forall x, y \in \mathcal{H}) \quad \langle x \mid y \rangle \leq \|x\| \|y\|. \tag{C-S}$$

While the actual definition can get quite technical, for our class, when we say "**Hilbert space**", we are referring to the finite-dimensional vector space $\mathcal{H}$, equipped with a scalar product $\langle \cdot \mid \cdot \rangle$ and a norm who arises from the scalar product via $\| \cdot \| = \sqrt{\langle \cdot \mid \cdot \rangle}$. Some examples are $\mathbb{R}^n$ under the Euclidean inner product, or the space of real $n \times m$ matrices under the Frobenius inner product.

**Exercise 1.13** Let $(x_1, x_2, x_3) \in \mathbb{R}^3$. Show that

$$2x_1 - x_2^4 + 6x_3 \le 4\sqrt{x_1^2 + x_2^8 + 9x_3^2}. \tag{15}$$

Can the coefficient $4$ in (15) be reduced?