Optimization for machine learning
Lecture Notes on Convexity
Master MASH & IASD

Vincent Duval

November 13, 2023

Foreword

The class of convex functions is particularly important in optimization. The reader is probably familiar with them, at least in a smooth setting. The beauty of convexity is that it provides a simple framework which allows to seamlessly handle non-smooth functions and constraints in optimization problems. In that respect, the treatise by R.T. Rockafellar [Roc97] is a masterpiece, and we mostly follow its presentation (see Appendix C for more references).

The key idea to handle constraints is to consider extended-valued functions, which may take the value $+\infty$ at points which are not feasible. Doing analysis with such functions requires a little care, and it is very often more convenient to work with their epigraphs instead, switching from the study of functions to the study of sets. The class of convex sets is special, as they have many nice properties, which convey as many interesting features to the class of convex functions.

These lecture notes are intended as a course material for the "Optimization for machine learning" class. Please note that they are at an early stage, and as a consequence, they are probably riddled with typos.

They are aimed at an active reader who is equipped with some paper and a pencil: not all the details of the proofs are given, but I have tried to provide the main clues which help completing the proofs.

On the other hand, those notes are intended to serve as a reference, and it is not necessary to learn everything that is written to complete the curriculum. The most important results and definitions, which should be known "by heart" are indicated by the symbol ♡. Conversely some parts, indicated by (♠), contain discussions or results that can be omitted in a first reading. They are not part of the curriculum, they usually give some pointers or remarks to go beyond the framework of the course.

Contents

1 Extended-valued functions **9**
- 1.1 Extended-valued functions, graphs and epigraphs 9
  - 1.1.1 Extended-valued functions. 9
  - 1.1.2 Proper functions 10
  - 1.1.3 Epigraph of a function 11
- 1.2 How to ensure the existence of a minimizer? 11
  - 1.2.1 The Weierstrass theorem 11
  - 1.2.2 Lower semi-continuity 13
  - 1.2.3 Coercive functions 16
  - 1.2.4 Main existence result 17
  - 1.2.5 Some remarks on the infinite-dimensional case (♠) . . 18
- 1.3 How to prove the uniqueness of the minimizer? 18

2 Convex sets **21**
- 2.1 Definiton and elementary examples 21
- 2.2 Operations that preserve convexity 22
  - 2.2.1 Intersection 22
  - 2.2.2 Cartesian product 23
  - 2.2.3 Affine maps 23
  - 2.2.4 Sum of convex sets 23
  - 2.2.5 Interior, closure 24
- 2.3 Convex hulls 24
- 2.4 Relative interior, extreme points, and faces 26
  - 2.4.1 Relative interior of a convex set 26
  - 2.4.2 Extreme points 27
- 2.5 Projection onto a convex set and consequences 29
  - 2.5.1 The projection theorem 29
  - 2.5.2 The separating hyperplane theorem 30

3 Convex (extended-valued) functions **31**
- 3.1 Definitions 31
  - 3.1.1 Geometric definition 31
  - 3.1.2 Definition with inequalities 33

3.2 Some consequences of the definition . . . 34
3.2.1 Level sets of a convex function . . . 34
3.2.2 Convex functions which take the value $-\infty$ . . . 34
3.3 Operations that preserve convexity . . . 34
3.4 Convex functions and continuity . . . 35
3.5 Strictly convex functions . . . 36
3.5.1 Definition . . . 36
3.5.2 Properties . . . 36
3.6 Strongly convex functions . . . 37
3.6.1 Definition and characterization . . . 37
3.6.2 Why are strongly convex functions useful? . . . 37

4 Convexity, continuity and differentiability **39**
4.1 Convexity on the real line . . . 39
4.1.1 Convex functions . . . 39
4.1.2 Strictly convex functions . . . 44
4.2 Convexity in dimension $p \geq 2$ . . . 45
4.2.1 Convex functions . . . 45
4.2.2 Strictly convex functions in dimension $p \geq 2$ . . . 47
4.2.3 Strongly convex functions . . . 47

5 Subgradients and conjugate functions **49**
5.1 Convex enveloppe of a function . . . 50
5.2 Subgradients of a function . . . 52
5.2.1 Definition . . . 52
5.2.2 Fermat's rule . . . 53
5.2.3 Digression: the relative interior and polyhedral convex functions . . . 53
5.2.4 Subdifferential and operations . . . 54
5.3 Conjugate functions . . . 55
5.3.1 The convex conjugate . . . 55
5.3.2 The double convex conjugate . . . 56
5.4 Fenchel-Rockafellar duality . . . 57

A Reminder on topology **59**
A.1 Interior, closure . . . 59
A.2 Continuity . . . 60

B Lower and upper limits of sequences and functions. **61**
B.1 Lower and upper limits of sequences . . . 61

C References **65**

Index **66**

Notation

The typical notation throughout the notes is the following. We want to optimize a function f defined on some vector space X. We shall always assume that $X = \mathbb{R}^p$ for some $p \in \mathbb{N}^*$.

Vocabulary. We say that a function $f \colon \mathbb{R} \to \mathbb{R}$ is increasing (resp. strictly increasing) if for all $x, y \in \mathbb{R}$, $x < y$ implies $f(x) \leq f(y)$ (resp. $f(x) < f(y)$). Similarly, we say that f is decreasing (resp. strictly decreasing) if for all $x, y \in \mathbb{R}$, $x < y$ implies $f(x) \geq f(y)$ (resp. $f(x) > f(y)$).

You may find other documents in the English-speaking literature in which "increasing" mean our "strictly increasing", and "non-decreasing" mean our "increasing". We adopt here the convention of [AB06], which is close to the French terminology.

Indeterminate forms. We shall avoid as much as possible indeterminate forms, but in case they happen we adopt the convention $0 \times \infty = 0 \times (-\infty) = 0$.

Moreover we define by convention

$$\inf \emptyset = +\infty, \quad \sup \emptyset = -\infty. \tag{1}$$

Line segments. Given two points $x, y \in X$, we define the *closed line segment* joining x and y as $[x, y] = \{\, \theta x + (1 - \theta) y \mid 0 \leq \theta \leq 1 \,\}$. The *open line segment* joining x and y is $]x, y[\, = [x, y] \setminus \{x, y\}$. Note that $]x, x[\, = \emptyset$ and that if $x \neq y$, $]x, y[\, = \{\, \theta x + (1 - \theta) y \mid 0 \leq \theta \leq 1 \,\}$. The intervals $]x, y]$ and $[x, y[$ are defined similarly and are left to the reader.

Interior, closure. If $A \subseteq X$, we denote by $\operatorname{int}(A)$ (resp. $\overline{A}$) the interior (resp. closure) of A. The open ball centered at x with radius $r > 0$ is denoted by

$$B(x, r) \overset{\text{def.}}{=} \{\, y \in X \mid \|x - y\| < r \,\}.$$

Chapter 1

Extended-valued functions

In this chapter, we introduce a notion of extended-valued function, that is a function which may take the value $+\infty$ so as to encode constaints. This point of view will be particularly useful when handling convex functions. Then, we discuss the first fundamental question in optimization: how to ensure that an extended-valued function has a minimizer. It is the occasion to review several concepts and definitions that are also useful in the real-valued case and to see how they adapt to this framework.

1.1 Extended-valued functions, graphs and epigraphs

1.1.1 Extended-valued functions.

Assume that we are interested in minimizing a function $f\colon C \to \mathbb{R}$, where $C \subseteq X$ is a constraint set (or simply the subset of X where f is defined), that is, we want to solve the problem

$$\min_{x\in C} f(x). \tag{1.1}$$

It is sometimes convenient to encode the constraint set C in the function f and to pretend as if f were defined on the whole space X. To this end, we may extend f in the following way,

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } x \in C \\ +\infty & \text{otherwise.} \end{cases} \tag{1.2}$$

so that (1.1) is equivalent to

$$\min_{x\in X} \tilde{f}(x). \tag{1.3}$$

Example 1.1. *One short way to encode the constraint set in the program*

$$\min_{x\in[-1,1]} x^2$$

is to write it as

$$\min_{x\in\mathbb{R}} \left(x^2 + \chi_{[-1,1]}(x)\right),$$

where $\chi_{[-1,1]}$ *is the* indicator function *(also called characteristic function) of the set* $[-1,1]$, *defined as*

$$\forall x \in \mathbb{R}, \quad \chi_{[-1,1]}(x) = \begin{cases} 0 & \text{if } x \in [-1,1] \\ +\infty & \text{otherwise.} \end{cases} \tag{1.4}$$

Writing the objective a sum of functions enables to study those functions separately (e.g. compute their (sub)derivatives) and gather their properties when studying the sum.

The points $x \in X$ where $\tilde{f}(x) = +\infty$ are the points that we really do not want to obtain while solving the program (1.3). Similarly, we may encounter points x where $\tilde{f}(x) = -\infty$, that is, points that we should accept without further discussion in our minimization program. In practice, that is of little interest, since the problem is readily solved if such a point x exists, but the value $-\infty$ might appear in our intermediate calculations, and we should be aware of that.

From now on, we drop the tilde and we simply write f instead of $\tilde{f}$. We may thus encounter extended-valued functions $f\colon X \to [-\infty,+\infty]$ (*i.e.* $\mathbb{R}\cup\{\pm\infty\}$), $f\colon X \to]-\infty,+\infty]$ (*i.e.* $\mathbb{R}\cup\{+\infty\}$) or real-valued functions $f\colon C \to \mathbb{R}$. The reader should pay particular attention to the difference.

When referring to the set of points where $f(x) < +\infty$, we speak of *effective domain* (or simply *domain*) of f,

$$\operatorname{dom}(f) \stackrel{\text{def.}}{=} \{\, x \in X \mid f(x) < +\infty \,\}. \tag{1.5}$$

It may happen that $f(x) = -\infty$ for some $x \in \operatorname{dom}(f)$, so the effective domain should not be mistaken for the usual definition domain of a function, *i.e.* the set of points where it is well-defined and real-valued.

1.1.2 Proper functions

As we shall see in the following, functions which take the value $-\infty$ (especially convex ones) are quite specific, as well as functions which are identically equal to $+\infty$. It is therefore convenient to exclude such cases.

Definition 1.1 (Proper function)**.** *Let* $f\colon X \to [-\infty,+\infty]$ *be an extended-valued function. We say that* f *is proper if*

1. *there exists* $x \in X$ *such that* $f(x) < +\infty$ *(i.e.* $\operatorname{dom} f \neq \emptyset$*),*

2. *for all* $x \in X$, $f(x) > -\infty$.

The term "proper function" is standard in convex analysis. Unfortunately, "proper function" has also a different meaning in analysis, where one requires that the inverse image of a compact set is a compact set. Both notions are unrelated.

1.1.3 Epigraph of a function

A cornerstone of convex analysis (but not only) is the equivalence between (extended-valued) functions and some specific sets. We know that a function is characterized by its *graph*, that is the set

$$\mathrm{graph}(f) \overset{\text{def.}}{=} \{ (x,y) \in X \times \mathbb{R} \mid y = f(x) \} . \tag{1.6}$$

In convex analysis, a bigger role is played by the *epigraph* of f, *i.e.* the set of points above the graph,

$$\mathrm{epi}(f) \overset{\text{def.}}{=} \{ (x,r) \in X \times \mathbb{R} \mid r \geq f(x) \} . \tag{1.7}$$

The fundamental idea is that most operations on functions can be interpreted as operations on their epigraphs.

Questions.

1. *To which functions correspond the empty epigraph* $\mathrm{epi}(f) = \emptyset$ *and the full epigraph* $\mathrm{epi}(f) = X \times \mathbb{R}$*?*
2. *Characterize the subsets of* $X \times \mathbb{R}$ *which are epigraphs of extended valued functions.*
3. *The domain* $\mathrm{dom}(f)$ *can be recovered from the epigraph* $\mathrm{epi}(f)$ *by a simple operation. Which one?*

1.2 How to ensure the existence of a minimizer?

When dealing with an optimization problem like (1.1) of (1.3), it is wise to ensure that the problem has indeed a minimizer before trying to write optimality conditions or designing an optimization algorithm. Otherwise, we might be wasting our time!

1.2.1 The Weierstrass theorem

For that purpose, an important tool is the Weierstrass theorem. The reader has probably already met it.

Theorem 1.2 (Weierestrass)**.** *Let $C \subseteq X$ be a nonempty compact set, and let $f : C \to \mathbb{R}$ be a continuous function. Then, f has a minimizer on C, i.e.*

$$\exists x^* \in C,\ f(x^*) = \min_{x \in C} f(x).$$

The reader probably knows one or many proofs of that result. Let us examine a standard proof which uses sequences.

Proof. Since C is nonempty, $\ell \overset{\text{def.}}{=} \inf_{x \in C} f(x)$ exists in $\mathbb{R} \cup \{-\infty\}$.

Step 1. Considering a minimizing sequence. By definition of the infimum, there exists a sequence[1] $(x_n)_{n\in\mathbb{N}}$ in C such that

$$\lim_{n\to\infty} f(x_n) = \ell. \tag{1.8}$$

Step 2. Extracting a convergent subsequence. Since C is a compact subset of X, we may extract a subsequence[2] $(x_{\varphi(n)})_{n\in\mathbb{N}}$ which converges in C. In other words, there exists $x^* \in C$ such that

$$\lim_{n\to\infty} x_{\varphi(n)} = x^*. \tag{1.9}$$

Step 3. Passing to the limit. Since f is continuous, we may pass to the limit in the subsequence,

$$f(x^*) = \lim_{n\to\infty} f(x_{\varphi(n)}) = \ell. \tag{1.10}$$

Hence x^* is a minimizer of f on K. Incidentally, we note that $\ell > -\infty$.

□

The Weierstrass theorem is important, and in general it is the quickest way to ensure the existence of a minimizer. However, it does not apply to every case. For instance, how can we prove that the functions f_1, f_2 defined by

$$f_1(x) = \begin{cases} 1/x & \text{for } 0 < x < 1 \\ (x-1)^2 & \text{for } x \geq 1 \end{cases}, \tag{1.11}$$

$$f_2(x) = \lceil |x| \rceil = \min \{\, n \in \mathbb{N} \mid n \geq |x| \,\} \tag{1.12}$$

have a minimizer?

[1]Recall the definition of the infimum: for all $t > \ell$, there exists $x \in C$ such that $\ell \leq f(x) \leq t$. We may pick for instance $t_n = \ell + 1/2^n$ if $\ell \in \mathbb{R}$, or $t_n = -n$ if $\ell = -\infty$, to construct such a sequence $(x_n)_{n\in\mathbb{N}}$.

[2]A subsequence of $(x_n)_{n\in\mathbb{N}}$ is a sequence of the form $(x_{\varphi(n)})_{n\in\mathbb{N}}$, where $\varphi : \mathbb{N} \to \mathbb{N}$ is strictly increasing. We may also denote the subsequence by $(x_{k_n})_{n\in\mathbb{N}}$.

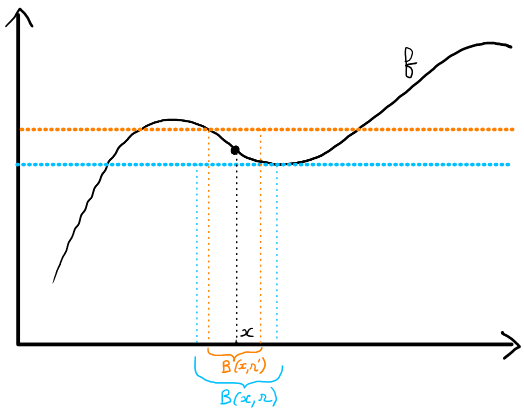

Figure 1.1: The continuity of a function is equivalent to (1.13) and (1.14)

1.2.2 Lower semi-continuity

Continuity is a condition which is a bit too strong to deal with all the practical optimization problems, especially if we use indicator functions in the objective to encode constraints.

Let us remind that the continuity of a function $f\colon X \to \mathbb{R}$ at some point $x \in X$ can be formulated as the following two conditions (see Figure 1.1),

$$\forall t' > f(x), \exists r' > 0, \forall y \in B(x, r'), \quad f(y) < t', \tag{1.13}$$

$$\forall t < f(x), \exists r > 0, \forall y \in B(x, r), \quad f(y) > t. \tag{1.14}$$

It turns out than only "half of the definition" of continuity, namely (1.14), is useful when proving the existence of a minimizer.

Definition 1.3 (Semi-continuity). *Let $f\colon X \to [-\infty, +\infty]$. We say that f is* lower semi-continuous *(l.s.c.) at $x \in X$ if*

$$\forall t < f(x), \exists r > 0, \forall y \in B(x, r), \quad f(y) > t. \tag{1.15}$$

Similarly, we say that f is upper semi-continuous *(u.s.c.) at $x \in X$ if*

$$\forall t' > f(x), \exists r' > 0, \forall y \in B(x, r'). \quad f(y) < t', \tag{1.16}$$

Upper semi-continuity is mostly useful when considering maximization problems. Since we focus here on minimization problems, we only consider lower semi-continuity in the following.

Examples of lower semi-continuous functions are given in Figure 1.2.

Just like continuity can be characterized using sequences, there is a sequential characterization of lower semi-continuity.

♡

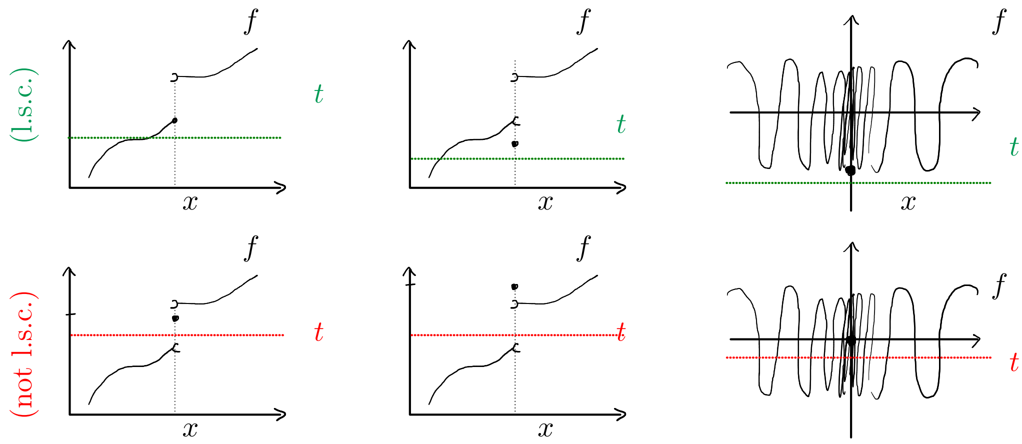

Figure 1.2: *Top row:* examples of lower semi-continuous (l.s.c.) functions. The last function is defined by $f(x) = \sin(1/x)$ for $x \neq 0$, $f(0) = -1$. *Bottom row:* examples of functions which are not l.s.c. The last function is defined by $f(x) = \sin(1/x)$ for $x \neq 0$, $f(0) = 0$.

Proposition 1.4 (Sequential characterization)**.** *The function $f\colon X \to [-\infty, +\infty]$ is lower semi-continuous if and only if for every sequence $(x_n)_{n\in\mathbb{N}} \in X^{\mathbb{N}}$ which converges towards x,*

$$\liminf_{n\to\infty} f(x_n) \geq f(x). \tag{1.17}$$

See Appendix B for a reminder on the limit inferior of a sequence.

Proof. ($\Leftarrow$) Let $(x_n)_{n\in\mathbb{N}}$ be a sequence converging to x, and let $t < f(x)$. By (1.15), there exists $r > 0$ such that $f(y) > t$ for all $y \in B(0, r)$. There exists $N \in \mathbb{N}$ such that for all $n \geq N$, $x_n \in B(x, r)$. Hence, for $n \geq N$,

$$\inf_{k\geq n} f(x_k) \geq t.$$

so that $\liminf_{n\to\infty} f(x_n) \geq t$. Since this is true for all $t < f(x)$, we let $t \to f(x)$ and we obtain (1.17).

($\Rightarrow$) Assume that (1.17) holds, and by contradiction, suppose that (1.15) does not hold. Then,

$$\exists t < f(x), \quad \forall r > 0,\ \exists y \in B(x, r), f(y) \leq t.$$

We choose $r = 2^{-n}$, for $n \in \mathbb{N}$, and we let y_n be a corresponding element in $B(x, 2^{-n})$. Then $\lim_{n\to\infty} y_n = x$, but

$$\liminf_{n\to\infty} f(y_n) \leq t < f(x),$$

which contradicts the assumption (1.17). □

In practice, we usually require the lower semi-continuity at every point of X, for which the following criterion holds.

Proposition 1.5 (Global characterization of semi-continuity)**.** *Let $f\colon X \to [-\infty, +\infty]$. The following assertions are equivalent.*

1. *f is lower semi-continuous at x for every $x \in X$,*
2. *the level set*
$$\{f \leq t\} \stackrel{\text{def.}}{=} \{\, x \in X \mid f(x) \leq t \,\}, \tag{1.18}$$
is closed.
3. epi(f) *is closed,*

Proof. $(i) \Rightarrow (iii)$. If $\operatorname{epi}(f) = \emptyset$, it is closed and there is nothing to prove. Otherwise, let $((x_n, r_n))_{n\in\mathbb{N}}$ be a sequence in $\operatorname{epi}(f)$, and assume that $(x_n, r_n) \to (x, r) \in X \times \mathbb{R}$ for $n \to +\infty$. We want to prove that $(x, r) \in \operatorname{epi}(f)$. By definition of the epigraph,

$$\forall n \in \mathbb{N}, \quad r_n \geq f(x_n).$$

Passing to the limit inferior and using (1.17)

$$r \geq \liminf_{n\to\infty} f(x_n) \geq f(x),$$

we obtain that $(x, r) \in \operatorname{epi}(f)$. Hence $\operatorname{epi}(f)$ is closed.

$(iii) \Rightarrow (ii)$. Let $t \in \mathbb{R}$, we note that

$$\operatorname{epi}(f) \cap X \times \{t\} = \{f \leq t\} \times \{t\},$$

so that $\{f \leq t\} \times \{t\}$ is closed, being the intersection of two closed sets. Since $\{f \leq t\}$ is homeomorphic to $\{f \leq t\} \times \{t\}$, we deduce that it is closed as well.

$(ii) \Rightarrow (i)$. Let $t < f(x)$, by assumption the level set $\{f \leq t\}$ is closed. Equivalently, its complement $\{f > t\}$ is open. Since $x \in \{f > t\}$, there exists $r > 0$ such that $B(x, r) \subseteq \{f > t\}$, which is exactly (1.15).

□

It is instructive to apply the global characterization of lower semi-continuity to indicator functions.

Example 1.2 (Closed constraint set and lower semi-continuity)**.** *Let $C \subseteq X$, and $f = \chi_C$ the indicator function of C, i.e.*

$$\chi_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ +\infty & \text{otherwise.} \end{cases} \tag{1.19}$$

Then f is lower semi-continuous if and only if C is closed.

In this particular case, we note that the lower semi-continuity is almost necessary, as it provides some form of closedness of the constraint set. Examples such as

$$\min_{x\in\mathbb{R}} \left(\frac{1}{2}x^2 + \chi_{]0,+\infty[}(x) \right) \tag{1.20}$$

show that without the closedness of the constraint set, optimization problems may have no solution. We may (in the end) remove the compactness assumption, but we should always assume some kind of closedness in our problem: in the proof of Theorem 1.2, our minimizing subsequence should converge inside the set of feasible points.

Questions.

1. *Compare Proposition 1.5 with the global characterization of continuity.*

2. *Give an example of an (extended-valued) lower semi-continuous function whose* domain *is not closed.*

3. *Let $f\colon X \to [-\infty, +\infty]$ be an (extended-valued) function, and let $S \overset{\text{def.}}{=} \overline{\mathrm{epi}(f)}$ be the closure of its epigraph. Prove that S is the epigraph of the largest lower semi-continuous function h such that $h \leq f$. Such a function is called the lower semi-continuous enveloppe of f.*

4. *Give a "functional definition" (that is, involving only functions) of the lower semi-continuous enveloppe of f. Hint: the supremum of a family of l.s.c. functions is l.s.c.*

1.2.3 Coercive functions

An additional observation is that, in Step 2 of the proof of Theorem 1.2, if we want to be able to extract a converging subsequence, we need to prevent it from "escaping towards infinity". It is possible if the functions takes large values at infinity.

Definition 1.6. *Let $f\colon X \to [-\infty, +\infty]$ be an extended-valued function. We say that f is coercive if*

$$\lim_{\|x\|\to\infty} f(x) = +\infty. \tag{1.21}$$

Questions.

1. *Check that the functions f_1, f_2 defined in (1.11) (and possibly extended by $+\infty$ as in (1.2)) are coercive.*

2. *Prove that f is coercive if and only if, for every $t \in \mathbb{R}$, its level set*

$$\{f \leq t\} \overset{\text{def.}}{=} \{\, x \in X \mid f(x) \leq t \,\}, \tag{1.22}$$

is bounded.

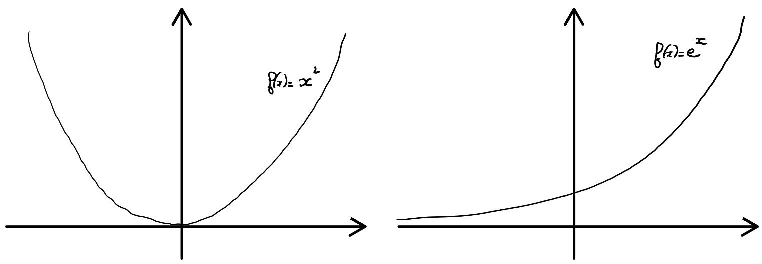

Figure 1.3: The function $x \mapsto x^2$ is coercive, whereas the function $x \mapsto \exp(x)$ is not coercive (incidentally, the second one has no minimizer)

1.2.4 Main existence result

The combination of coercivity and lower semi-continuity allows to extend the Weierstrass theorem.

Theorem 1.7. *Let $f \colon X \to]-\infty, +\infty]$ be a proper, lower semi-continuous and coercive function. Then f has a minimizer, i.e.*

$$\exists x^* \in X,\ f(x^*) = \min_{x \in X} f(x).$$

Proof. Let $\ell = \inf_X f$. Since there exists $x \in \operatorname{dom} f$, we have $\ell \in \mathbb{R} \cup \{-\infty\}$.

Step 1. Considering a minimizing sequence. Let $(x_n)_{n\in\mathbb{N}}$ be a minimizing sequence in X, *i.e.*

$$\lim_{n\to\infty} f(x_n) = \ell. \tag{1.23}$$

Step 2. Extracting a convergent subsequence. Let $t > \ell$. There is an index $N \in \mathbb{N}$ such that for all $n \geq N$, $x_n \in \{f \leq t\}$. Since f is coercive, $\{f \leq t\}$ is bounded, *i.e.* there exists $R > 0$ such that $\{f \leq t\} \subseteq B(0, R)$. Moreover, by the lower semi-continuity of f, the set $\{f \leq t\}$ is closed; it is therefore compact. We may thus extract a subsequence $(x_{\varphi(n)})_{n\in\mathbb{N}}$ which converges towards some $x^* \in \{f \leq t\}$.

Step 3. Passing to the limit inferior. Since f is lower semi-continuous, we have

$$f(x^*) \leq \liminf_{n\to\infty} f(x_{\varphi(n)}) = \ell. \tag{1.24}$$

Hence x^* is a minimizer of f on X. Since f is proper, we obtain that $\ell > -\infty$. □

Remark 1.1. *The assumption that f is proper in Theorem 1.7 may be removed, but we should be ready to accept points x^* such that $f(x^*) = \pm\infty$, which is rarely interesting.*

Remark 1.2. *The idea of proof of Theorem 1.7 is the same as in the proof of Theorem 1.2: pick a minimizing sequence, extract a convergent subsequence and prove that the limit point is a minimizer. That line of argument is common in variational analysis and is often called the* direct method in the calculus of variations.

Questions.

1. *The above formulation uses extended valued functions on X. Prove that every proper, lower semi-continuous real-valued function attains its infimum on every non-empty compact set $C \subseteq X$.*

2. *Prove the existence of minimizers for* (1.11).

1.2.5 Some remarks on the infinite-dimensional case (♠)

Sequential definitions. We have used in the previous paragraphs the sequential definitions of properties (compactness, continuity. . .). It is a good exercise to write down these results using the standard definitions (with coverings, etc.). In particular, in infinite dimension (with a general locally convex topology), these properties are not always equivalent to their sequential counterpart!

Coercivity. In Definition 1.6, we have defined coercivity using the norm, but in infinite dimension or in metric spaces, it is more relevant for the direct method to require that for all $t \in \mathbb{R}$, the closure of the level set, $\overline{\{f \leq t\}}$, is compact (see [DM93] for more detail, in particular the use of sequential definitions). In $X = \mathbb{R}^p$, both definitions are equivalent (check this as an exercise).

Choice of topology. In $X = \mathbb{R}^p$ all (vector-space) topologies are equivalent, but in the infinite-dimensional setting, one of the main concerns when applying the direct method is finding a suitable topology. A too strong topology does not yield the required compactness on the level sets of f, but, in a too weak topology, f may not be lower semi-continuous. In reflexive Banach spaces (in particular in Hilbert spaces) or in spaces that are the dual of a vector space, the Banach-Alaoglu-Bourbaki theorem often provides the desired compactness in the weak (or weak-*) topology, for sets that are bounded in norm. See for instance [Bre11].

1.3 How to prove the uniqueness of the minimizer?

A standard tool to prove the uniqueness of a minimizer (if it exists), is the strict convexity of f. We study in further detail in the next chapter, but let us briefly mention the definition we need.

Definition 1.8. *A proper function $f : X \to]-\infty, +\infty]$ is strictly convex if*

$$\forall x, y \in \operatorname{dom}(f), x \neq y, \forall \theta \in]0,1[, \quad f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta) f(y). \tag{1.25}$$

Proposition 1.9. *Let $f : X \to]-\infty, \infty]$ be a proper, strictly convex function. Then f has at most one minimizer.*

Proof. By contradiction, assume that there exist $x_1, x_2 \in X$ such that $f(x_1) = f(x_2) = \min_X f$.

Then

$$f\left(\frac{1}{2}x_1 + \frac{1}{2}x_2\right) < \frac{1}{2}f(x_1) + \frac{1}{2}f(x_2) = \min_X f, \tag{1.26}$$

which contradicts the definition of the minimimum. As a result, there is a most one minimizer. □

Remark 1.3. *Instead of considering a* proper *extended valued function, one could simply consider a real-valued function on a convex set $C \subseteq X$.*

If the functional f is not strictly convex (or not even convex), one should proceed with a case-specific study, for instance by exploiting the optimality conditions.

Chapter 2

Convex sets

Contents

2.1 Definiton and elementary examples 21

2.2 Operations that preserve convexity 22

2.2.1 Intersection 22

2.2.2 Cartesian product 23

2.2.3 Affine maps 23

2.2.4 Sum of convex sets 23

2.2.5 Interior, closure 24

2.3 Convex hulls 24

2.4 Relative interior, extreme points, and faces . . . 26

2.4.1 Relative interior of a convex set 26

2.4.2 Extreme points 27

2.5 Projection onto a convex set and consequences 29

2.5.1 The projection theorem 29

2.5.2 The separating hyperplane theorem 30

It is sometimes more convenient, instead of studying directly an extended valued function, to study its epigraph or its level sets. That is all the more true when dealing with convex functions, whose properties directly follow from the convexity properties of their epigraphs. This chapter summarizes many interesting properties of convex sets.

2.1 Definiton and elementary examples

Let $x, y \in X$. The (closed) line segment joining x to y is

$$[x, y] \stackrel{\text{def.}}{=} \{ (1-\theta)x + \theta y \mid 0 \le \theta \le 1 \} . \tag{2.1}$$

The quantity $(1-\theta)x+\theta y$ is called a convex combination of x and y (or sometimes a barycenter of x and y).

The open line segment[1] joining x to y is

$$]x,y[\overset{\text{def.}}{=} [x,y] \setminus \{x,y\} = \begin{cases} \emptyset & \text{if } x = y, \\ \{\, (1-\theta)x+\theta y \mid 0<\theta<1 \,\} & \text{otherwise.} \end{cases} \tag{2.2}$$

Definition 2.1. *Let $C \subseteq X$. We say that C is convex if for all $x, y \in C$, $[x,y] \subseteq C$.*

Example 2.2. *The following sets are convex.*

- *The empty set $\emptyset$.*
- *Linear spaces, affine spaces,*
- *The unit (or probability) simplex*

$$\Delta_p = \left\{ x \in \mathbb{R}^p \mid \sum_{j=1}^{p} x_j = 1 \text{ and } \forall i \in [\![1\,;p]\!], x_i \geq 0 \right\}. \tag{2.3}$$

- *Balls (open or closed) for any norm N on X,*

$$\{\, y \in X \mid N(y-x) \leq r \,\},\ \{\, y \in X \mid N(y-x) \leq r \,\}. \tag{2.4}$$

- *The set of positive semidefinite matrices,*

$$\mathbb{S}_p^+(\mathbb{R}) \overset{\text{def.}}{=} \left\{ M \in \mathbb{R}^p \times p \mid M^\top = M \text{ and } \forall x \in \mathbb{R}^p, x^\top M x \geq 0 \right\}. \tag{2.5}$$

Remark 2.3. *By induction it is possible to consider convex combinations of more than two points.The set C is convex if and only if*

$$\forall n \in \mathbb{N}^*,\ \forall x_1,\ \ldots, x_n \in C,\ \forall(\theta_1, \ldots, \theta_n) \in \Delta_n,\quad \sum_{i=1}^{n} \theta_i x_i \in C. \tag{2.6}$$

2.2 Operations that preserve convexity

2.2.1 Intersection

Any intersection (possibly infinite) of convex sets is convex.

Proposition 2.4. *Let $\{C_i\}_{i\in I}$ be a family of convex sets in X. Then $\bigcap_{i\in I} C_i$ is convex.*

Proof. If $\bigcap_{i\in I} C_i = \emptyset$, there is nothing to prove. Otherwise, let $x, y \in \bigcap_{i\in I} C_i$. For each $i \in I$, both x and y belong to C_i, hence $[x,y] \subseteq C_i$. Since this holds for every $i \in I$, we have $[x,y] \subseteq \bigcap_{i\in I} C_i$, hence the convexity of $\bigcap_{i\in I} C_i$. □

[1] Sometimes in the English-speaking literature, the open line segment is dented by (x, y).

2.2.2 Cartesian product

Proposition 2.5. *Let $C_1, \dots, C_k$ be convex subsets of the vector spaces $X_1, \dots, X_k$ respectively. Then $C_1 \times \dots \times C_k$ is a convex subset of $X_1 \times \dots \times X_k$.*

Example 2.6. *The cylinder*

$$C = \left\{ (x_1, \dots, x_p) \in X \mid x_1^2 + \dots + x_{p-1}^2 \leq 1, \quad |x_p| \leq 1 \right\}$$

is convex.

2.2.3 Affine maps

An affine map $f\colon \mathbb{R}^p \to \mathbb{R}^n$ is a map of the form $x \mapsto (Ax + b)$, where $A \in \mathbb{R}^{n \times p}$ is a matrix and $b \in \mathbb{R}^n$ is a vector.

Proposition 2.7. *Let $f\colon \mathbb{R}^p \to \mathbb{R}^n$ be an affine map. Then*

1. *For all convex set $C \subseteq \mathbb{R}^p$, the image $f(C)$ is convex.*
2. *For all convex set $D \subseteq \mathbb{R}^n$, the preimage $f^{-1}(D) = \{ x \in \mathbb{R}^p \mid f(x) \in D \}$ is convex.*

Proof. Let $f\colon x \mapsto Ax + b$ be an affine map, with $A \in \mathbb{R}^{n \times p}$, $b \in \mathbb{R}^n$. We assume that C and D are nonempty, otherwise there is nothing to prove.

Let $y_1, y_2 \in f(C)$. There exist $x_1, x_2 \in \mathbb{R}^p$ such that $y_i = f(x_i)$, $1 \leq i \leq 2$. For any $\theta \in [0, 1]$,

$$\begin{aligned}(1-\theta)y_1 + \theta y_2 &= (1-\theta)(Ax_1 + b) + \theta(Ax_2 + b) \\ &= A((1-\theta)x_1 + \theta x_2) + b = f\left((1-\theta)x_1 + \theta x_2\right)\end{aligned}$$

which is in $f(C)$ since C is convex.

Now, let $x_1, x_2 \in f^{-1}(D)$, and let $\theta \in [0, 1]$. Then by the linearity of A,

$$f\left((1-\theta)x_1 + \theta x_2\right) = (1-\theta)f(x_1) + \theta f(x_2) \in D$$

since D is convex. Hence $((1-\theta)x_1 + \theta x_2) \in f^{-1}(D)$. $\square$

Example 2.8. *As important particular cases, we note that if C is convex,*

- *the translate $a + C \stackrel{\text{def.}}{=} \{ a + x \mid x \in C \}$, where $a \in X$, is convex,*
- *the dilation $\alpha C \stackrel{\text{def.}}{=} \{ \alpha x \mid x \in C \}$, where $\alpha \in \mathbb{R}$.*

2.2.4 Sum of convex sets

For $C_1, C_2 \subseteq X$, one may define their (Minkowski) sum:

$$C_1 + C_2 \stackrel{\text{def.}}{=} \{ x_1 + x_2 \mid x_1 \in C_1, x_2 \in C_2 \}. \tag{2.7}$$

If C_1 or C_2 is empty, then so is $C_1 + C_2$.

Proposition 2.9. *Let $C_1, C_2 \subseteq X$ be two convex sets. Then $C_1 + C_2$ is convex.*

Proof. This follows from Proposition 2.5 and Proposition 2.7. But you may also write a direct proof as an exercise! $\square$

2.2.5 Interior, closure

Convexity is compatible with topological properties. We refer to Appendix A.1 for a reminder on the definitions of the interior and closure of a set.

Proposition 2.10. *Let $C \subseteq X$. Then both its interior* $\operatorname{int}(C)$ *and its closure* $\overline{C}$ *are convex.*

Proof. Interior. If $\operatorname{int}(C) = \emptyset$, there is nothing to prove. Otherwise, let $y, z \in \operatorname{int}(C)$, and let $x = (1-\theta)y + \theta z$, with $\theta \in [0, 1]$ (we may even assume that $0 < \theta < 1$, otherwise $x = y$ or $x = z$ and thus $x \in \operatorname{int}(C)$). There exists some radius $r > 0$ such that $B(y, r) \subseteq C$. Applying the convexity of C to every line segment between z and any element of $B(y, r)$, we note that

$$(1-\theta)B(y, r) + \theta z \subseteq C.$$

But $(1-\theta)B(y, r) + \theta z = B(x, (1-\theta)r)$, hence $x \in \operatorname{int}(C)$, and $\operatorname{int}(C)$ is convex.

Closure. A point $x \in X$ is in $\overline{C}$ if and only if there are points of C arbitrarily close to x. In other words,

$$\overline{C} = \bigcap_{\varepsilon > 0} (C + B(0, \varepsilon)) \,. \tag{2.8}$$

Thus, it follows from Proposition 2.9 and Proposition 2.4 that $\overline{C}$ is convex. □

While the closure operation is quite natural with convex sets, taking the interior is not always interesting, as one frequently works with convex sets that have smaller dimension than the ambient space. The notion of *relative interior* (see Section 2.4.1) is often more relevant.

2.3 Convex hulls

The convex hull of a subset A of X is the smallest convex subset of X which contains A. That set can be described both from the outside and from the inside.

Proposition 2.11 (Convex hull of a set)**.** *Let $A \subseteq X$. The convex hull of A is the convex set given by*

$$\operatorname{conv} A = \bigcap_{\substack{A \subset C, \\ C \text{ convex}}} C \tag{2.9}$$

$$= \left\{ \sum_{k=1}^{n} \theta_k x_k \mid n \in \mathbb{N}^*, \ \theta \in \Delta_n, \ x_1, \dots, x_n \in A \right\} . \tag{2.10}$$

The reader may check as an exercise that both representation describe the same set, and that is indeed the smallest convex which contains A.

Interestingly, it is not necessary to take n arbitrary large in (2.10).

Theorem 2.12 (Carathéodory)**.** *Let $p = \dim X$ and $A \subseteq X$. Then every point of* $\operatorname{conv} A$ *is a convex combination of at most $p + 1$ points of A.*

The proof relies on an induction argument, see for instance [HUL93, Thm. III.1.3.6].

Sometimes it is more convenient to work with a convex set that is also closed. The closed convex hull of A is the smallest *closed convex* set which contains A.

Proposition 2.13 (Closed convex hull of a set)**.** *Let $A \subseteq X$. The closed convex hull of A is the convex set given by*

$$\overline{\operatorname{conv} A = \bigcap_{\substack{A \subset C, \\ C \text{ closed convex}}} C} \tag{2.11}$$

$$= \left\{ \sum_{k=1}^{n} \theta_k x_k \mid n \in \mathbb{N}^*,\ \theta \in \Delta_n,\ x_1, \dots, x_n \in A \right\}. \tag{2.12}$$

Remark 2.14. *Many notions in mathematics have this kind of inner and outer representation. The linear span of a set is the smallest vector space which contains A,*

$$\operatorname{Vect} A = \bigcap_{\substack{A \subset C, \\ F \text{ vector space}}} F \tag{2.13}$$

$$= \left\{ \sum_{k=1}^{n} \lambda_k x_k \mid n \in \mathbb{N}^*,\ \lambda \in \mathbb{R}^n,\ x_1, \dots, x_n \in A \right\}. \tag{2.14}$$

The affine hull of a set is the smallest affine space which contains A,

$$\operatorname{Aff} A = \bigcap_{\substack{A \subset F, \\ F \text{ affine space}}} F \tag{2.15}$$

$$= \left\{ \sum_{k=1}^{n} \lambda_k x_k \mid n \in \mathbb{N}^*,\ \lambda \in \mathbb{R}^n,\ \sum_{k=1}^{n} \lambda_k = 1, x_1, \dots, x_n \in A \right\}. \tag{2.16}$$

The convex conic hull is the smallest convex cone which contains A,

$$\operatorname{cone} A = \bigcap_{\substack{A \subset C, \\ C \text{ convex cone}}} C \tag{2.17}$$

$$= \left\{ \sum_{k=1}^{n} \lambda_k x_k \mid n \in \mathbb{N}^*,\ \lambda \in \mathbb{R}_+^n,\ x_1, \dots, x_n \in A \right\}. \tag{2.18}$$

By relying on the notion of affine hull of a set, it is possible to define the *dimension of a convex set.*

Definition 2.15 (Dimension of a convex set)**.** *Let $C \subseteq X$ be a convex set. We define its dimension* $\dim C$ *as the dimension of its affine hull* $\operatorname{Aff} C$*.*

2.4 Relative interior, extreme points, and faces

2.4.1 Relative interior of a convex set

Consider the line segment $[a,b]$ for $a,b \in \mathbb{R}^2$, $a \neq b$. Obviously, its topological interior is empty. Nevertheless, we intuitively feel that the points $]a,b[$ are "more in the interior" of $[a,b]$ than the point a and b. Similarly, the square $[0,1]^2 \times \{0\}$ has empty interior in $\mathbb{R}^3$, but it is tempting to say that the points in $]0,1[^2 \times \{0\}$ are more in the interior than the edges of the square.

The notion of relative interior formalizes this idea by considering the interior when regarding the convex set as a subset of its affine hull.

Definition 2.16 (Relative interior)**.** *Let $C \subseteq X$ be a convex set, and $x \in C$. We say that x is in the* relative interior *of C if there exists $r > 0$ such that*

$$(\mathrm{Aff}\, C) \cap B(x,r) \subseteq C. \tag{2.19}$$

We denote by $\mathrm{rint}\, C$ *the collection of all such points.*

One important property of the points in the relative interior of C is that it is always possible to "step back" while looking in the direction of another point of C.

Proposition 2.17. *Let $C \subseteq X$ be a convex set and $x \in C$. The following assertions are equivalent.*

1. *x is in the relative interior of C,*
2. *For all $y \in C \setminus \{x\}$, there exists $z \in C$ such that $x \in]y,z[$.*

Proposition 2.18. *If $C \subseteq X$ is a nonempty convex set, then* $\mathrm{rint}\, C \neq \emptyset$*. Moreover* $\dim(\mathrm{rint}\, C) = \dim C$*.*

Proof. If C is a singleton $\{x\}$, the relative interior is reduced to $\{x\}$, and the result holds. We assume now that C has more than two points. We know that

$$\mathrm{Aff}\, C = \left\{ \sum_{i \in I} \lambda_i c_i \mid I \subseteq \mathbb{N} \text{ finite},\ \lambda \in \mathbb{R}^I,\ \sum_{i \in I} \lambda_i = 1, \{c_i\}_{i \in I} \subseteq C \right\}. \tag{2.20}$$

We fix some $c_0 \in C$ and we rewrite it as

$$\begin{aligned} &\mathrm{Aff}\, C = c_0 + V \\ &\text{where } V = \left\{ \sum_{i \in I} \lambda_i (c_i - c_0) \mid I \subseteq \mathbb{N} \text{ finite},\ \lambda \in \mathbb{R}^I,\ \sum_{i \in I} \lambda_i = 1, \{c_i\}_{i \in I} \subseteq C \right\} \\ &\qquad = \left\{ \sum_{i \in I} \lambda_i (c_i - c_0) \mid I \subseteq \mathbb{N} \text{ finite},\ \lambda \in \mathbb{R}^I,\ \{c_i\}_{i \in I} \subseteq C \right\}. \end{aligned}$$

In the last line, we have removed the constraint $\sum_{i \in I} \lambda_i = 1$, since it is always possible to add c_0 to the collection $\{c_i\}_{i \in I}$, with a weight such that $\sum_{i \in I} \lambda_i = 1$.

Therefore, the collection $\{c - c_0\}_{c\in C}$ spans V. By standard results in linear algebra, we may extract a basis $\{c_i - c_0\}_{1\leq i\leq m}$ of V, where $m = \dim V =: \dim C$. Then, any element $x \in \operatorname{Aff} C$ can be written as

$$x = c_0 + \sum_{i=1}^{m} \mu_i (c_i - c_0)$$

with $\mu \in \mathbb{R}^m$. Moreover, the μ_i's are the coefficient of $x - c_0$ in the basis $\{c_i - c_0\}_{1\leq i\leq m}$, hence they depend continuously on x.

Setting $x_* = \frac{1}{m+1}\left(c_0 + \sum_{i=1}^m c_i\right)$ so that $x_* \in C$, we deduce that there exists $r > 0$ such that for all $x \in B(x_*, r) \cap (\operatorname{Aff} C)$,

$$\forall i \in \{1, \ldots, m\}, \quad \left|\mu_i - \frac{1}{m+1}\right| < \frac{1}{m(m+1)}.$$

Then $\mu_i > 0$ for $1 \leq i \leq m$, and setting

$$\mu_0 \stackrel{\text{def.}}{=} 1 - \sum_{i=1}^{m} \mu_i = 1 - \frac{m}{m+1} - \sum_{i=1}^{m}\left(\mu_i - \frac{1}{m+1}\right) > \frac{1}{m+1} - \frac{m}{m(m+1)} > 0.$$

Thus $\sum_{i=0}^m \mu_i c_i$ is a convex combination of the c_i's and it belongs to C. We have proved that $B(x_*, r) \cap (\operatorname{Aff} C) \subseteq C$, hence the claimed result. □

Proposition 2.19 (Relative interior and affine maps (♠))**.** *Let $f\colon \mathbb{R}^n \to \mathbb{R}^p$ be an affine map.*

- *If $C \subseteq \mathbb{R}^n$ is a convex set, then $\operatorname{rint}(f(C)) = f(\operatorname{rint} C)$.*
- *If $D \subseteq \mathbb{R}^p$ is a convex set and $f^{-1}(\operatorname{rint} D) \neq \emptyset$, then $\operatorname{rint}(f^{-1}(D)) = f^{-1}(\operatorname{rint} D)$.*

We refer to [HUL93, Prop. III.2.1.12] for a proof.

Similarly to the relative interior, it is possible to define the notion of relative boundary,

$$\operatorname{rbd}(C) \stackrel{\text{def.}}{=} \overline{C} \setminus (\operatorname{rint} C). \tag{2.21}$$

2.4.2 Extreme points

While the elements of the relative interior of a convex set C are those that lie on open line segments in any suitable direction (see Proposition 2.17), sthe points which do not belong to any open line segment in C...

Definition 2.20. *Let $C \subseteq X$ be a convex set, and $x \in C$. We say that x is an extreme point of C is C does not belong to any open line segment in C.*

An alternative characterization of x being an extreme point is the following implication:

$$\forall y, z \in C, \ (x \in [y, z] \Longrightarrow x = y \text{ or } x = z.) \tag{2.22}$$

Proposition 2.21. *Let C be a nonempty* compact *convex set. Then C has at least one extreme point.*

Proof. We consider the function $f\colon X \to \mathbb{R}$, defined by

$$\forall x \in X, \quad f(x) = \frac{1}{2} \|x\|_2^2 .$$

It is continuous on the nonempty compact set C, hence it has a maximizer $x \in C$. The key point is that f is strictly convex (see Section 3.5.1), but we provide a self-contained argument here. If $x = (1-\theta)y + \theta z$ with $y, z \in C$ and $\theta \in [0,1]$,

$$\begin{aligned} \|x\|^2 = \|\theta z + (1-\theta)y\|^2 &= \theta^2 \|z\|^2 + (1-\theta)^2 \|y\|^2 + \underbrace{2\theta(1-\theta)\langle z, y\rangle}_{\theta(1-\theta)\left(\|z\|^2+\|y\|^2-\|z-y\|^2\right)} \\ &= \theta \|z\|^2 + (1-\theta)\|y\|^2 - \theta(1-\theta)\|z-y\|^2 . \end{aligned}$$

By assumption, $\|z\|^2 \leq \|x\|^2$ and $\|y\|^2 \leq \|x\|^2$. As a result,

$$0 \leq -\theta(1-\theta)\|z-y\|^2$$

which implies that $\theta \in \{0,1\}$ or $z = y$, hence in any case $x \in \{y, z\}$. The point x is thus an extreme point of C. □

Remark 2.22. *The assumption that C is compact is important. Can you find examples of unbounded or non-closed convex sets which do not have extreme points?*

The extreme points of a compact convex set contain all the necessary information to recover the whole convex set.

Theorem 2.23 (Minkowski)**.** *Let C be a nonempty compact convex set. Then $C = \operatorname{conv}(\operatorname{extr} C)$.*

Together with Carathéodory's theorem Theorem 2.12, this implies that each point of C is a convex combination of at most $n+1$ extreme points of C, where $n \overset{\text{def.}}{=} \dim C$.

A convenient way to prove that a point $x \in C$ is an extreme point of C is to prove that it is an *exposed point*, that is x is the unique maximizer (or minimizer) of some linear form on C.

Proposition 2.24 (Exposed points are extreme)**.** *Let $C \subset X$ be a convex set and let $x \in C$. If $\{x\} = \operatorname{argmax}_{c \in C} \langle c, v\rangle$ for some $v \in \mathbb{R}^p \setminus \{0\}$, then c is an extreme point of C.*

Proof. Assume that $x = (1-\theta)y + \theta z$ for some $y, z \in C$, $\theta \in]0,1[$. Then

$$\langle x, v\rangle = (1-\theta)\langle y, v\rangle + \theta\langle z, v\rangle \leq (1-\theta)\left(\max_{c\in C}\langle c, v\rangle\right) + \theta\left(\max_{c\in C}\langle c, v\rangle\right) = \max_{c\in C}\langle c, v\rangle ,$$

and the inequality is strict unless $\langle y, v\rangle = \langle z, v\rangle = \max_{c\in C}\langle c, v\rangle$, that is $y = z = x$. □

2.5 Projection onto a convex set and consequences

2.5.1 The projection theorem

One of the most important theorems on convex sets is the existence of a unique projection for the Euclidean ℓ^2 norm. In our finite-dimensional setting, it might not be very surprising, but it becomes a fundamental tool in infinite dimension (Hilbert spaces).

Theorem 2.25. *Let $C \subseteq X$ be a nonempty closed convex set. Then, for every $x \in X$, there exists a unique point $p \in C$, such that*

$$\|x-p\|_2^2 = \min_{c \in C} \|x-c\|_2^2 . \tag{2.23}$$

Moreover, p is the unique point in C such that

$$\forall c \in C, \quad \langle x-p,\, c-p\rangle \leq 0. \tag{2.24}$$

Proof. Existence. The extended-valued function

$$f\colon c \mapsto \|x-c\|_2^2 + \chi_C(c) = \begin{cases} \|x-c\| & \text{if } c \in C \\ +\infty & \text{otherwise} \end{cases} \tag{2.25}$$

is lower semi-continuous (l.s.c.), as the sum of l.s.c. functions. Moreover, it is finite at some point since $C \neq \emptyset$, and it is coercive: for all c with $\|c\| \geq \|x\|$

$$f(c) \geq (\|c\| - \|x\|)^2 \xrightarrow{\|c\| \to +\infty} +\infty.$$

By Theorem 1.7, there exists a minimizer $p \in X$, and necessarily $p \in C$.

Uniqueness. The function f is strictly convex (see Section 3.5.1), hence the minimizer is unique.

Characterization. The point $p \in C$ is a minimizer if and only if, for every $c \in C$, every $\theta \in [0,1]$, the inequality $f(p) \leq f(\theta c + (1-\theta)p)$ holds. But

$$\begin{aligned} f(p) - f(\theta c + (1-\theta)p) &= \|x-p\|^2 - \|x - (\theta c + (1-\theta)p)\|^2 \\ &= \langle x - p + (x - (\theta c + (1-\theta)p))\,,\, \theta(c-p)\rangle \\ &= \langle 2(x-p) + \theta(p-c),\, \theta(c-p)\rangle \\ &= 2\theta\,\langle x-p,\, c-p\rangle - \theta^2\,\|c-p\|^2 . \end{aligned} \tag{2.26}$$

If p is a minimizer, the left hand side is non-positive. We divide (2.26) by $\theta \neq 0$, and, letting $\theta \to 0^+$, we obtain $0 \geq 2\,\langle x-p,\, c-p\rangle$.

Conversely, if (2.24) holds, we take $\theta = 1$ in (2.26), and we obtain

$$f(p) - f(c) = 2\,\langle x-p,\, c-p\rangle - \|c-p\|^2 \leq 0,$$

which implies that p is a minimizer. ☐

The point p is called the projection of x onto C.

Remark 2.26. *If C is not closed, the point p might not exist. For instance, consider $C = \,]1, +\infty[$ and $x = 0$.*

If C is closed, but not convex, the point p exists, but it might not be unique. For instance, let $C = \left\{ x \in \mathbb{R}^2 \mid \|x\| = 1 \right\}$, and $x = (0,0)$.

2.5.2 The separating hyperplane theorem

A hyperplane is a set of the form

$$H = \{ x \in X \mid \langle x, y \rangle = \alpha \},$$

where $y \neq 0$ and $\alpha \in \mathbb{R}^p$.

An open half space is a set of the form

$$D = \{ x \in X \mid \langle x, y \rangle > \alpha \} \text{ or } D = \{ x \in X \mid \langle x, y \rangle < \alpha \}.$$

with the same conditions on $y$and α. A closed half space is defined similarly, replacing the symbol $<$ with $\leq$ and $>$ with $\geq$.

We say that the hyperplane H *separates a point* $x_0 \in X$ *from the set* $C \subseteq X$ if one of the corresponding open half-spaces contains x_0 and the other one contains C.

In other words, H separates x_0 from C if

- $\langle x_0, y \rangle > \alpha$,
- $\langle c, y \rangle < \alpha$ for every $c \in C$

(or symmetrically exchanging the symbols $>$ and $<$).

Theorem 2.27. *Let $C \subseteq X$ be a nonempty, closed convex set, and $x_0 \in X \setminus C$.*

Then, there exists a hyperplane which separates x_0 and C.

Proof. Let p denote the projection of x_0 onto C. We know that

$$\begin{aligned} \forall c \in C, \quad & \langle x_0 - p, c - p \rangle \leq 0, \\ \text{or equivalently} \quad & \langle c, x_0 - p \rangle \leq \langle p, x_0 - p \rangle . \end{aligned}$$

Setting $y \stackrel{\text{def.}}{=} x_0 - p \neq 0$, we obtain $\langle c, y \rangle \leq \langle p, y \rangle$ for all $c \in C$.

Moreover,

$$\langle x_0, y \rangle = \underbrace{\langle x_0 - p, y \rangle}_{= \|x_0 - p\|^2 > 0} + \langle p, y \rangle > \langle p, y \rangle .$$

As a result, setting $\alpha = \frac{1}{2} (\langle p, y \rangle + \langle x_0, y \rangle)$, we have

$$\forall c \in C, \quad \langle c, y \rangle \leq \langle p, y \rangle < \alpha < \langle x_0, y \rangle , \tag{2.27}$$

which yields the claimed result. □

Corollary 2.28. *Every closed convex set is the intersection of all the closed half-spaces that contain it.*

Proof. If $C = X$, it is not contained in any half-space. The intersection over the empty collection being equal to X, the claim holds. If $C = \emptyset$, the result also holds (take any two half-spaces with empty intersection).

If $\emptyset \subsetneq C \subsetneq X$, let F be the the intersection of all the half-spaces that contain C. Obviously, $C \subseteq F$. Assume by contradiction that there is some $x_0 \in F \setminus C$. By Theorem 2.27, there exists a hyperplane which separates x_0 and C, for instance $C \subseteq \{ x \in X \mid \langle x, y \rangle < \alpha \}$ and $\langle x_0, y \rangle > \alpha$.

As a result, $x_0 \notin \{ x \in X \mid \langle x, y \rangle \leq \alpha \} \supseteq F$, which contradicts the assumption. Hence $F = C$. □

Chapter 3

Convex (extended-valued) functions

Contents

3.1 Definitions **31**
3.1.1 Geometric definition 31
3.1.2 Definition with inequalities 33
3.2 Some consequences of the definition **34**
3.2.1 Level sets of a convex function 34
3.2.2 Convex functions which take the value $-\infty$ 34
3.3 Operations that preserve convexity **34**
3.4 Convex functions and continuity **35**
3.5 Strictly convex functions **36**
3.5.1 Definition 36
3.5.2 Properties 36
3.6 Strongly convex functions **37**
3.6.1 Definition and characterization 37
3.6.2 Why are strongly convex functions useful? 37

Convex functions play an important role in optimization. One reason is that at each at each point, the slope of the function gives a rough idea of where the set of minimizers is. In this chapter, we define the notion of convex extended valued functions, and we summarize their elementary properties.

3.1 Definitions

3.1.1 Geometric definition

The definition of convexity is a simple geometric criterion which can be applied to real-valued functions as well as to extended-valued functions.

Definition 3.1. *Let $f \colon X \to [-\infty, +\infty]$. We say that f is* convex *if its epigraph,*

$$\mathrm{epi}(f) \overset{\text{def.}}{=} \{ (x, r) \in X \times \mathbb{R} \mid r \geq f(x) \} , \tag{3.1}$$

is a convex set.

There are several common convex functions.

Affine functions If $f(x) = \langle p, x \rangle - \alpha$ for some $p \in X$ and $\alpha \in \mathbb{R}$, then

$$\mathrm{epi}(f) = \left\{ (x, r) \in X \times \mathbb{R} \mid \left\langle \begin{pmatrix} p \\ -1 \end{pmatrix}, \begin{pmatrix} x \\ r \end{pmatrix} \right\rangle \leq \alpha \right\}$$

is a half space. Therefore f is convex.

Squared Euclidean norm If $f(x) = \frac{1}{2} \|x\|^2$, its epigraph is the region above a paraboloid. Therefore it is convex.

Constantly infinity If $f(x) = +\infty$ for all x, then $\mathrm{epi}(f) = \emptyset$. If $f(x) = -\infty$ for all x, $\mathrm{epi}(f) = X \times \mathbb{R}$. In both cases, f is convex.

Indicator function If $f(x) = \chi_C(x)$ (as in (1.19)), then $\mathrm{epi}(f) = C \times \mathbb{R}_+$, and f is convex if and only if C is convex.

Remark 3.1. *The reader has probably already encountered a slightly different definition which is used for real-valued functions. Consider a set $C \subseteq X$ and a mapping $f \colon C \to \mathbb{R}$. The function f is said to be convex if its epigraph*

$$\mathrm{epi}(f) \overset{\text{def.}}{=} \{ (x, r) \in C \times \mathbb{R} \mid r \geq f(x) \} , \tag{3.2}$$

is convex. Note that in that case, C must be convex.

Moreover, if we extend f into $\tilde{f} \colon X \to]-\infty, +\infty]$ as in (1.2), the epigraphs of f and $\tilde{f}$ coincide, i.e.

$$\mathrm{epi}(\tilde{f}) = \{ (x, r) \in C \times \mathbb{R} \mid r \geq f(x) \} = \mathrm{epi}(f),$$

so that $\tilde{f}$ is convex (in the sense of Definition 3.1) if and only if f is convex (in the classical sense).

As a result, in the following, we go back and forth between extended-valued functions defined on X, and real-valued functions defined on a (possibly smaller) set $C \subseteq X$.

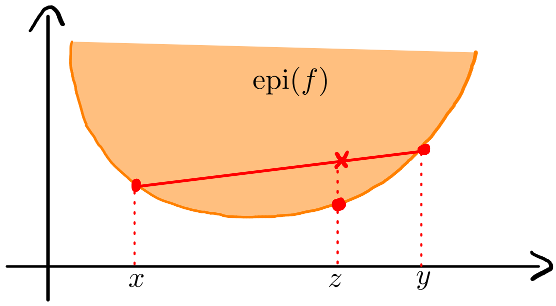

Figure 3.1: The convexity of the epigraph of f is equivalent to (3.3) (where $z \stackrel{\text{def.}}{=} \theta x + (1-\theta)y$). The image of the barycenter is below the barycenter of the images.

3.1.2 Definition with inequalities

Another definition of convexity, perhaps more common, involves convex combinations and inequalities. We recall that the effective domain is $\mathrm{dom}(f) = \{ x \in X \mid f(x) < +\infty \}$.

Proposition 3.2. *An extended-valued function $f\colon X \to [-\infty, +\infty]$ is convex if and only if*

$$\forall x, y \in \mathrm{dom}(f),\ \forall \theta \in]0,1[, \quad f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y). \tag{3.3}$$

Checking the equivalence stated in Proposition 3.2 is left to the reader (see Figure 3.1).

Additionally, we may reformulate the convexity criterion with $n \geq 2$ points. By induction, one may check that f is convex if and only if for all $x_1, \dots, x_n \in \mathrm{dom}(f)$ and all $\theta_1, \dots, \theta_n > 0$ such that $\sum_{i=1}^n \theta_i = 1$,

$$f\left(\sum_{i=1}^n \theta_i x_i\right) \leq \sum_{i=1}^n \theta_i f(x_i). \tag{3.4}$$

Remark 3.2. *There are several variants of (3.3). The main difficulty is to avoid indeterminate forms. See for instance [ET99, Def. I.2.1] or [Roc97, Th. 4.1, Th. 4.2] for alternative formulations.*

Remark 3.3 (Real-valued functions on a subset)**.** *In view of Remark 3.1, we note that a function $f\colon C \to \mathbb{R}$ is convex if and only if C is convex and*

$$\forall x, y \in C,\ \forall \theta \in]0,1[, \quad f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y). \tag{3.5}$$

3.2 Some consequences of the definition

3.2.1 Level sets of a convex function

Proposition 3.3. *If $f\colon X \to [-\infty, +\infty]$ is convex, then, for all $t \in \mathbb{R}$, the level set*

$$\{f \leq t\} \stackrel{\text{def.}}{=} \{\, x \in X \mid f(x) \leq t \,\} \tag{3.6}$$

is convex.

Proof. We note that

$$\mathrm{epi}(f) \cap (X \times \{t\}) = \{\, (x,t) \in X \times \mathbb{R} \mid f(x) \leq t \,\} = \{f \leq t\} \times \{t\}.$$

Since the left-hand side is convex (intersection of convex sets), we deduce that $\{f \leq t\} \times \{t\}$ is convex, hence so is its projection $\{f \leq t\}$. □

The converse is not true! There are non-convex functions whose level sets are all convex. A function with convex level sets is called a quasi-convex function.

3.2.2 Convex functions which take the value $-\infty$

Convex functions which take the value $-\infty$ may arise in some calculations (if we are not cautious), but they are not very interesting. For instance, in dimension one, it is possible to check that if $g\colon \mathbb{R} \to [-\infty, +\infty]$ is convex and takes the value $-\infty$, then there exists $-\infty \leq a < b \leq +\infty$ such that

$$g(t) = \begin{cases} -\infty & \text{for } t \in \,]a,b[\\ +\infty & \text{for } t \in \mathbb{R} \setminus [a,b]. \end{cases} \tag{3.7}$$

or

$$g(t) = \begin{cases} -\infty & \text{for } t = a \\ +\infty & \text{for } t \in \mathbb{R} \setminus \{a\}. \end{cases} \tag{3.8}$$

As a result, there are at most two points where g may take a finite value! Similarly, in dimension $p \geq 2$, f can be finite only on the boundary of the convex set $\{f \leq -\infty\}$.

Another difficulty which arises with functions taking values $-\infty$ is that in sums or convex combinations, we might face undetermined forms $-\infty + \infty$. Therefore, we usually work with convex *proper* functions as much as possible.

3.3 Operations that preserve convexity

Very often, it is easier to check that a function is convex by relying on some well known convex functions instead of applying the definition of convexity.

Proposition 3.4. *The following functions are convex.*

1. $f_1 + f_2$, *where* $f_1, f_2 \colon X \to \mathbb{R} \cup \{+\infty\}$ *are convex,*

2. αf, *where* $\alpha > 0$ *and* $f \colon X \to [-\infty, \infty]$ *is convex,*

3. $\sup_{i \in I} f_i$, *where* $\{f_i\}_{i \in I}$ *is any collection (possibly uncountable) of convex functions* $f_i \colon X \to [-\infty, \infty]$,

4. $x_1 \mapsto (\inf_{x_2} f(x_1, x_2))$, *where* $f \colon X_1 \times X_2 \to [-\infty, \infty]$ *is convex.*

Some other sufficient conditions, which rely on the derivatives, are given in Chapter 4.

Proof. 1. One may check that (3.3) holds on $\operatorname{dom}(f_1 + f_2) = \operatorname{dom}(f_1) \cap \operatorname{dom}(f_2)$.

2. Observe that $\operatorname{epi}(\alpha f) = u_\alpha(\operatorname{epi}(f)$, where u_α is the linear map $(x, r) \mapsto (x, \alpha r)$.
3. Note that $(x, r) \in \operatorname{epi}\left(\sup_{i \in I} f_i\right)$ if and only if, for all $i \in I$, $(x, r) \in \operatorname{epi}(f_i)$. Therefore $\operatorname{epi}\left(\sup_{i \in I} f_i\right) = \bigcap_{i \in I} \operatorname{epi}(f_i)$ is convex.
4. Let $g \colon x_1 \mapsto \inf_{x_2} f(x_1, x_2)$ and $(x_1, r) \in X_1 \times \mathbb{R}$. We note that $(x_1, r) \in \operatorname{epi} g$ if and only if for all $\varepsilon > 0$, there exists $x_2 \in X_2$ such that $r \geq f(x_1, x_2) - \varepsilon$. In other words iff $(x_1, r) \in \operatorname{Proj}_1(\operatorname{epi}(f - \varepsilon))$, where $\operatorname{Proj}_1 \colon X_1 \times X_2 \to X_1$ is the projection onto the first component, $\operatorname{Proj}_1(x_1, x_2) = x_1$.

   As a result,
   $$\operatorname{epi} g = \bigcap_{\varepsilon > 0} \operatorname{Proj}_1(\operatorname{epi}(f - \varepsilon))$$
   Since each function $f - \varepsilon$ is convex, its epigragph is convex and so is its projection (see Proposition 2.7). Their intersection is thus convex Proposition 2.4, hence g is convex.

□

3.4 Convex functions and continuity

Convexity have some regulatity, at least in the interior of their domain.

Proposition 3.5. *Let* $f \colon X \to \mathbb{R} \cup \{+\infty\}$ *be convex. Then* f *is continuous on* $\operatorname{int}(\operatorname{dom} f)$.

Remark 3.6. *In fact, the proof shows that* f *is locally Lipschitz, i.e. on every compact subset of* $\operatorname{int}(\operatorname{dom} f)$, *there is some* $L > 0$ *such that* f *is* L*-Lipchitz.*

It may happen that $\operatorname{int}(\operatorname{dom} f) = \emptyset$, in which case the function cannot be continuous at any point of $\operatorname{dom} f$. However, it might be interesting to apply the above result to $f|_V$, the restriction of f to $V \overset{\text{def.}}{=} \operatorname{Aff}(\operatorname{dom} f)$, the affine hull of $\operatorname{dom} f$. The interior of $\operatorname{dom} f|_V$ is then the relative interior of $\operatorname{dom} f$, $\operatorname{rint}(\operatorname{dom} f)$.

3.5 Strictly convex functions

The notion of strict convexity provides a straightforward argument to prove that the minimizer of some function is unique (see Section 1.3).

3.5.1 Definition

As far as I know, there is no universal geometric definition for strictly convex functions: this concept is mostly relevant for real-valued functions, and most authors define it using inequalities. The main idea is that f should be convex and its graph should not contain any nonempty open line segment.

A more common definition of strict convexity is given using inequalities.

Proposition 3.7. *An extended-valued function $f\colon X \to [-\infty, +\infty]$ is strictly convex if and only if*

$$\forall x, y \in \mathrm{dom}(f),\ \ x \neq y, \forall \theta \in]0,1[\,,\ \ f\left(\theta x + (1-\theta)y\right) < \theta f(x) + (1-\theta) f(y). \tag{3.9}$$

The canonical example of strictly convex functions is $x \mapsto \|x\|^2$. Indeed, for all $\theta \in]0,1[$, $x, y \in X$,

$$\begin{aligned}\|\theta x + (1-\theta)y\|^2 &= \theta^2 \|x\|^2 + (1-\theta)^2 \|y\|^2 + \underbrace{2\theta(1-\theta)\langle x,\, y\rangle}_{\theta(1-\theta)\left(\|x\|^2+\|y\|^2-\|x-y\|^2\right)}\\ &= \theta \|x\|^2 + (1-\theta)\|y\|^2 - \theta(1-\theta)\|x-y\|^2\\ &< \theta \|x\|^2 + (1-\theta)\|y\|^2 \ \text{ for } x \neq y.\end{aligned}$$

3.5.2 Properties

Proposition 3.8. *The following functions are strictly convex.*

1. $f_1 + f_2$, *where* $f_1\colon X \to \mathbb{R} \cup \{+\infty\}$ *is convex and* $f_2\colon X \to \mathbb{R} \cup \{+\infty\}$ *is* strictly convex,
2. αf, *where* $\alpha > 0$ *and* $f\colon X \to [-\infty, \infty]$ *is strictly convex,*
3. $\max(f_1, f_2)$, *where* $f_1, f_2\colon X \to [-\infty, \infty]$ *are strictly convex functions.*

The proof is left to the reader.

If $(f_i)_{i\in I}$ is an *infinite* family of strictly convex functions, in general it is not true that $\sup_{i\in I} f_i$ is strictly convex (consider, *e.g.*, $f_n(x) = 1 + \frac{1}{2^n}x(x-1)$).

Several other sufficient conditions for strict convexity are given in Section 4.2.2.

3.6 Strongly convex functions

Strong convexity is an even stronger(!) property than strict convexity. It is helpful in optimization to derive convergence rates.

3.6.1 Definition and characterization

Definition 3.9. *Let $f\colon X \to [-\infty, +\infty]$ and $\lambda > 0$. We say that f is λ-strongly convex if $x \mapsto \left(f(x) - \frac{\lambda}{2}\|x\|^2\right)$ is convex.*

Remark 3.4. *Let f be a proper function, and write $f(x) = h(x) + \frac{\lambda}{2}\|x\|^2$ where h is convex. Since $x \mapsto \frac{\lambda}{2}\|x\|^2$ is strictly convex and h is convex, it follows that f is strictly convex.*

The class of λ-strongly convex functions is easy to handle since its characterization directly follows from the characterization of the convexity of functions of the form $h \overset{\text{def.}}{=} f - \frac{\lambda}{2}\|\cdot\|^2$.

Proposition 3.10 (Characterization using inequalities)**.** *Let $f\colon X \to [-\infty, +\infty]$ be an extended valued function and $\lambda > 0$. Then, f is λ-strongly convex if and only if for all $x, y \in \operatorname{dom}(f)$, all $\theta \in]0,1[$,*

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y) - \frac{\lambda}{2}\theta(1-\theta)\|x-y\|^2. \tag{3.10}$$

Proof. Apply the characterization of convexity to $h \overset{\text{def.}}{=} f - \frac{\lambda}{2}\|\cdot\|^2$ and recall that

$$\|\theta x + (1-\theta)y\|^2 = \theta\|x\|^2 + (1-\theta)\|y\|^2 - \theta(1-\theta)\|x-y\|^2$$

(see Section 3.5.1). □

3.6.2 Why are strongly convex functions useful?

The reader will see in lectures on algorithms how strong convexity can be exploited, but let us highlight an important property.

If $f\colon U \to \mathbb{R}$ is strongly convex, and $x^* \in U$ is the (unique) minimizer, then, by taking $x = x^*$ in (3.10), we obtain

$$f(x^*) \le f(\theta x^* + (1-\theta)y) \le \theta f(x^*) + (1-\theta)f(y) - \frac{\lambda}{2}\theta(1-\theta)\|x^* - y\|^2.$$

Dividing by $(1-\theta)$ and letting $\theta \to 1$, we get

$$\forall y \in U, \quad f(y) \ge f(x^*) + \frac{\lambda}{2}\|y - x^*\|^2. \tag{3.11}$$

In other words, if we have an upper bound on the gap in the objective values, *i.e.* if we know that our iterate $y = x^{(k)}$ satisfies

$$\left(f(x^{(k)}) - \min f\right) \le \alpha_k,$$

with (hopefully) $\alpha_k \to 0$, then we can deduce that our iterate is not too far from the minimizer, *i.e.*

$$\left\| x^{(k)} - x^* \right\|^2 \leq \frac{2}{\lambda} \alpha_k \to 0.$$

Chapter 4

Convexity, continuity and differentiability

Contents

4.1 Convexity on the real line 39

4.1.1 Convex functions 39

4.1.2 Strictly convex functions 44

4.2 Convexity in dimension $p \geq 2$ 45

4.2.1 Convex functions 45

4.2.2 Strictly convex functions in dimension $p \geq 2$ 47

4.2.3 Strongly convex functions 47

In this chapter, we focus on the characterizations of convex functions through their derivatives. Such characterizations turn out to be very useful in practice.

4.1 Convexity on the real line

We begin by studying convex functions in a one-dimensional setting. This section paves the way for the the proofs of the higher-dimensional setting and it contains most of the technical discussion. It may be skipped at first reading.

4.1.1 Convex functions

The convexity of a set is a one-dimensional property, in the sense that it only involves its intersection with lines. Similarly, the convexity of a function only involves its restriction to lines.

Proposition 4.1. *Let $f : X \to [-\infty, +\infty]$. Then f is convex if and only if for every $x, y \in \operatorname{dom}(f)$, the function $g : \mathbb{R} \to [-\infty, +\infty]$, defined by*

$$\forall t \in \mathbb{R}, \quad g(t) \stackrel{\text{def.}}{=} f(tx + (1-t)y), \tag{4.1}$$

is convex.

It is therefore particularly important to characterize univariate convex functions.

In view of Section 3.2.2, the only interesting case is when $g : \mathbb{R} \to]-\infty, +\infty]$, and by restricting our attention to $I \stackrel{\text{def.}}{=} \operatorname{dom}(g)$, we are led to study real-valued functions $g : I \to \mathbb{R}$, where $I \subseteq \mathbb{R}$ is an interval.

In the following, unless otherwise stated, $I \subseteq \mathbb{R}$ is an interval containing at least two points.

Increasing slope of the chords

We define the slope between two points $A = (a, a')$, and $B = (b, b')$ in $\mathbb{R}^2$ as

$$s(A, B) \stackrel{\text{def.}}{=} \frac{b' - a'}{b - a}. \tag{4.2}$$

The reader may check the following equivalence (see Figure 4.1).

Lemma 4.2. *Let $A = (a, a')$, $B = (b, b')$, $C = (c, c')$ be three points in $\mathbb{R}^2$ such that $a < b < c$. The following assertions are equivalent.*

(i) B *is below* $[A, C]$,

(ii) C *is above the line which joins* A *and* B,

(iii) A *is above the line which joins* B *and* C,

(iv) $s(A, B) \le s(A, C)$,

(v) $s(A, C) \le s(B, C)$.

For a typical point $x \in I$, we define the corresponding point $M_x \stackrel{\text{def.}}{=} (x, g(x)) \in \mathbb{R}^2$. Applying Lemma 4.2, it is possible to prove the following characterization of convexity using slopes.

Proposition 4.3 (Characterization with the slope of the chords)**.** *Let $g : I \to \mathbb{R}$ be a function. Then g is convex if and only if for all $a \in I$, the function*

$$x \mapsto s(M_a, M_x) = \frac{g(x) - g(a)}{x - a} \tag{4.3}$$

is increasing on $I \setminus \{a\}$.

Proof. We note that $a < b < c$ if and only if there is $\theta \in]0, 1[$ such that $b = \theta a + (1-\theta)c$. Moreover, M_b is below $[M_a, M_c]$ if and only if $g(\theta a + (1-\theta)c) \le \theta g(a) + (1-\theta)g(c)$. By Lemma 4.2 this is equivalent to $s(M_a, M_b) \le s(M_a, M_c)$. □

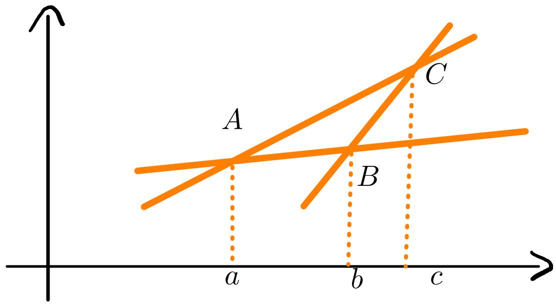

Figure 4.1: The relative positions of the lines (AB), (AC) and (BC) can be compared by examining their slopes.

Increasing slopes

Proposition 4.3 has interesting consequences on the differentiability of g. Let $x_0 \in \operatorname{int}(I)$. We say that g has a left derivative at x_0 if the limit

$$g'_\ell(x_0) \stackrel{\text{def.}}{=} \lim_{\substack{x \to x_0 \\ x < x_0}} \left(\frac{g(x) - g(x_0)}{x - x_0} \right) \tag{4.4}$$

exists and is finite. Similarly, we say that g has a right derivative at x_0 if the limit

$$g'_r(x_0) \stackrel{\text{def.}}{=} \lim_{\substack{x \to x_0 \\ x > x_0}} \left(\frac{g(x) - g(x_0)}{x - x_0} \right) \tag{4.5}$$

exists and is finite. The function g is differentiable at x_0 if and only if the left and right derivatives at x_0 exist and coincide.

It turns out that a convex function has left and right derivatives everywhere in the interior of its domain.

Proposition 4.4. *If $a, b \in \operatorname{int}(I)$ with $a < b$, and $g \colon I \to \mathbb{R}$ is convex, then $g'_r(a)$ and $g'_\ell(b)$ exist (in $\mathbb{R}$) and*

$$g'_r(a) \leq \frac{g(b) - g(a)}{b - a} \leq g'_\ell(b). \tag{4.6}$$

Remark 4.1. *If a (resp. b) lies on the boundary of I, i.e. $a = \inf I$ (resp. $b = \sup I$), the limit (4.5) for $x_0 = a$ still exists (resp. (4.4) for $x_0 = b$), but it could be $-\infty$ (resp. $+\infty$).*

Proof. Let $c \in]a, b[$. Using Lemma 4.2 (or Proposition 4.3), we get

$$\frac{g(c) - g(a)}{c - a} \leq \frac{g(b) - g(a)}{b - a} \leq \frac{g(b) - g(c)}{b - c}. \tag{4.7}$$

Since ratio on the left is increasing, its limit $g'_r(a)$ exists in $\mathbb{R}\cup\{-\infty\}$ as $c \to a$. But since $a \in \mathrm{int}(I)$, there exists $d < a$, and

$$-\infty < \frac{g(d)-g(a)}{d-a} \leq \frac{g(c)-g(a)}{c-a},$$

hence that limit must be finite.

A symmetric argument applies to the ratio on the right of (4.7) for $c \to b$. □

Theorem 4.5 (Characterization with derivatives)**.** *Let $g \colon I \to \mathbb{R}$, with $I \subseteq \mathbb{R}$ an* open *interval. The following assertions are equivalent.*

(i) g is convex,

(ii) g is continuous on I and there exists a countable set $S \subseteq I$ such that g is differentiable on $I \setminus S$, with g' is increasing on $I \setminus S$.

(iii) g is continuous on I and there exists a countable set $S \subseteq I$ such that g is differentiable on $I \setminus S$, with

$$\forall y \in I, \forall x \in I \setminus S, \quad f(y) \geq f(x) + f'(x)(y-x). \tag{4.8}$$

Proof (♠). (i)⇒ (ii). Assume that g is convex. Then g has left and right derivatives at every point of I, which implies that g is continuous on I. Moreover, for all $x \in I$, passing to the limit $h \to 0^+$ in the slopes inequality

$$\frac{g(x-h)-g(x)}{(-h)} \leq \frac{g(x+h)-g(x)}{h},$$

we note that $g'_\ell(x) \leq g'_r(x)$. Combining this observation with (4.6), we note that

$$\forall a, b \in I \text{ such that } a < b, \quad g'_\ell(a) \leq g'_r(a) \leq g'_\ell(b) \leq g'_r(b),$$

so that both g'_ℓ and g'_r are increasing functions, and at each point $x \in I$, the following assertions are equivalent:

- $g'_\ell(x) = g'_r(x)$,
- g'_ℓ is continuous at x,
- g'_r is continuous at x.

The discontinuity set of a increasing function is countable, hence g'_ℓ and g'_r are equal except on a countable set.

(ii) ⇒ (i). Assume by contradiction that g satisfies the properties of 2) but g is not convex. Then, by Proposition 4.3, there exists[1] $a < b < c$ such that

$$\frac{g(b)-g(a)}{b-a} > \frac{g(c)-g(b)}{c-b}.$$

But the mean value theorem[2] together with the fact that g' is increasing give

$$\frac{g(b)-g(a)}{b-a} \leq \sup_{x \in]a,b[\setminus S} g'(x) \leq \inf_{x \in]b,c[\setminus S} g'(x) \leq \frac{g(c)-g(b)}{c-b},$$

[1]This requires a careful check. First, negate the fact that the slopes are increasing for some a, so that the slopes corresponding to $x = b$ and $x = c$ with $b < c$ are not in the correct order. Then using Lemma 4.2, relabel a, b, and c to have $a < b < c$ and the claimed inequality.

[2]This is a refined version of the classical mean value theorem (*inégalité des accroissements finis* in French), which allows for g' to be undefined on a countable set, see for instance [Bou07, Thm. I.2.1].

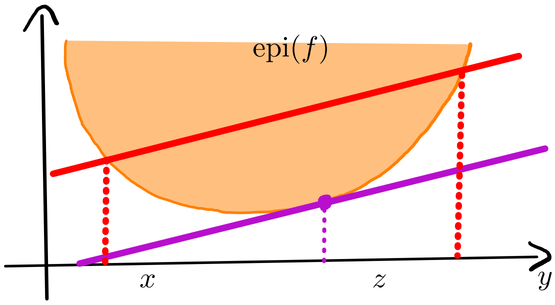

Figure 4.2: A convex function is below its chords (in between the two intersection points) and above its tangents.

which yields a contradiction. Hence g is convex.

(i) $\Rightarrow$ *(iii)*. Let $y \in I$ and $x \in I \setminus S$. If $y > x$, let $h > 0$ such that $x < x + h < y$. Using the chords inequality,

$$\frac{g(y) - g(x)}{y - x} \geq \frac{g(x+h) - g(x)}{h} \longrightarrow g'(x) \quad \text{as } h \to 0.$$

If $y < x$ let $h > 0$ such that $y < x - h < x$. Similarly,

$$\frac{g(y) - g(x)}{y - x} \leq \frac{g(x-h) - g(x)}{(-h)} \longrightarrow g'(x) \quad \text{as } h \to 0.$$

In both cases, we have obtained (4.8).

(iii) $\Rightarrow$ *(ii)*. Let $x, y \in I \setminus S$ with $x < y$. By (4.8),

$$g(y) \geq g(x) + g'(x)(y - x)$$
$$g(x) \geq g(y) + g'(y)(x - y).$$

Summing both inequalities yield $0 \geq (g'(x) - g(y))(y - x)$, and since $x < y$, we obtain $g'(x) \leq g'(y)$. Hence g' is increasing. $\square$

In practice, we often know that the function g is smooth. In that case, Theorem 4.5 reformulates as

Proposition 4.6. *Let $I \subseteq \mathbb{R}$ be an* open *interval, and let $g \in \mathscr{C}^k(I)$, with $k \geq 1$. The following assertions are equivalent*

(i) *g is convex,*

(ii) *g' is increasing,*

(iii) *For all $x, y \in I$,*

$$g(y) \geq g(x) + g'(x)(y - x). \tag{4.9}$$

(iv) (if $k \geq 2$) *$g''(x) \geq 0$ for all $x \in I$.*

Remark 4.2. *Comparing* (3.5) *and* (4.9)*, we see that there are two competing geometric interpretations of convexity (see Figure 4.2):*

- *g is below its chords (between the two points of intersection, outside it is above the chord),*
- *g is above its tangent.*

4.1.2 Strictly convex functions

Again, strict convexity is a one-dimensional property. A function f is strictly convex if and only it is strictly convex on every line which intersects $\mathrm{dom}(f)$.

Here, we consider again a function $g\colon I \to \mathbb{R}$. The main results of Section 4.1.2 have a "strict" counterpart. We leave to the reader the task to check that the following results hold.

Lemma 4.7. *Let $A = (a, a')$, $B = (b, b')$, $C = (c, c')$ be three points in $\mathbb{R}^2$ such that $a < b < c$. The following assertions are equivalent.*

(i) *B is strictly below $[A, C]$,*

(ii) *C is strictly above the line which joins A and B,*

(iii) *A is strictly above the line which joins B and C,*

(iv) *$s(A, B) < s(A, C)$,*

(v) *$s(A, C) < s(B, C)$.*

Proposition 4.8 (Characterization with the slope of the chords)**.** *Let $g\colon I \to \mathbb{R}$ be a function. Then g is strictly convex if and only if for all $a \in I$, the function*

$$x \mapsto s(M_a, M_x) = \frac{g(x) - g(a)}{x - a} \tag{4.10}$$

is strictly increasing on $I \setminus \{a\}$.

Proposition 4.9. *If $a, b \in \mathrm{int}(I)$ with $a < b$, and $g\colon I \to \mathbb{R}$ is strictly convex, then $g'_r(a)$ and $g'_\ell(b)$ exist (in $\mathbb{R}$) and*

$$g'_r(a) < \frac{g(b) - g(a)}{b - a} < g'_\ell(b). \tag{4.11}$$

Theorem 4.10 (Characterization with derivatives)**.** *Let $g\colon I \to \mathbb{R}$, with $I \subseteq \mathbb{R}$ an* open *interval. The following assertions are equivalent.*

(i) *g is strictly convex,*

(ii) *g is continuous on I and there exists a countable set $S \subseteq I$ such that g is differentiable on $I \setminus S$, with g' is strictly increasing on $I \setminus S$.*

(iii) *g is continuous on I and there exists a countable set $S \subseteq I$ such that g is differentiable on $I \setminus S$, with*

$$\forall y \in I, \forall x \in I \setminus S \cup \{y\}, \quad f(y) > f(x) + f'(x)(y - x). \tag{4.12}$$

As, we often know beforehand that the function g is smooth, Theorem 4.10 reformulates as

Proposition 4.11. *Let $I \subseteq \mathbb{R}$ be an* open *interval, and let $g \in \mathscr{C}^k(I)$, with $k \geq 1$. The following assertions are equivalent*

(i) *g is strictly convex,*

(ii) *g' is strictly increasing,*

(iii) *For all $x, y \in I$, with $y \neq x$,*

$$g(y) > g(x) + g'(x)(y - x). \tag{4.13}$$

(iv) (if $k \geq 2$) *$g''(x) > 0$ for all x in a dense subset of I.*

Do not forget the dense subset in point 4. of Proposition 4.11! It is not true that any strictly convex function g satisfies $g''(x) > 0$ for all x, for the same reason that there exist strictly increasing functions whose derivative vanishes at isolated points

Proof. The equivalence of the first three point is left to the reader. We only focus on the fourth one.

(i) ⇒ (iv). Since g is convex, we know that $g''(x) \geq 0$ for all $x \in I$. Assume by contradiction that there is a nonempty open interval $J \subseteq I$ such that $g''(x) = 0$. Then g' is constant in J and g is affine in J, which contradicts the strict convexity of g. Hence $g''(x) > 0$ on a dense subset of I.

(iv) ⇒ (ii). By the continuity of g'' and by a density argument, $g''(x) \geq 0$ for all $x \in I$, hence g' is increasing. Now, let $x, y \in I$, $x < y$. By density there exists $z \in]x, y[$ such that $g''(x) > 0$, and for $h > 0$ small enough,

$$g'(z + h) \geq g'(z) + h/2 > g'(z).$$

Since g' is increasing, we obtain

$$g'(x) \leq g'(z) < g'(z + h) \leq g'(y)$$

hence g' is strictly increasing. □

4.2 Convexity in dimension $p \geq 2$

4.2.1 Convex functions

Now, we turn back to convex functions defined on $X = \mathbb{R}^p$, and we first consider smooth functions.

Smooth convex functions

From Proposition 4.1, we know that we can characterize convex functions on X using their restrictions on lines. The key point when exploiting the regularity of f is to relate the corresponding derivatives. Let U be an convex open set, $f \in \mathscr{C}^k(U)$ and $g(t) \overset{\text{def.}}{=} f(tx + (1-t)y)$. Then, for all $t \in \mathbb{R}$ such that $tx + (1-t)y \in U$,

$$g'(t) = \langle \nabla f\,(tx + (1-t)y)\,,\, x - y\rangle \tag{4.14}$$

$$g''(t) = \left\langle x - y,\, \nabla^2 f\,(tx + (1-t)y)\,(x-y)\right\rangle, \tag{4.15}$$

where $\nabla f(x) \in \mathbb{R}^p$ and $\nabla^2 f(x) \in \mathbb{R}^{p\times p}$ respectively denote[3] the gradient and the Hessian of f at x.

Applying Proposition 4.6 to all the restrictions of f to lines yields the important characterization.

Proposition 4.12. *Let $U \subseteq X$ be a nonempty* open *convex set, and let $f \in \mathscr{C}^k(U)$, with $k \geq 1$. The following assertions are equivalent*

(i) f is convex,

(ii) for all $x, y \in U$, $\langle \nabla f(x) - \nabla f(y),\, x - y\rangle \geq 0$,

(iii) for all $x, y \in U$, $f(y) \geq f(x) + \langle \nabla f(x),\, y - x\rangle$,

(iv) (for $k \geq 2$) for all $x \in U$, $\nabla^2 f(x) \succeq 0$.

Proof. (i) $\Rightarrow$ (ii,iii,iv). Assume that f is convex. Let $x, y \in U$. Then, the function $g : t \mapsto f(tx + (1-t)y)$ is convex, the set $[0,1]$ is contained in the open interval $\operatorname{dom} g$, and by Proposition 4.6,

- $0 \leq g'(1) - g'(0) = \langle \nabla f(y) - \nabla f(x),\, y - x\rangle$,
- $0 \leq g(1) - g(0) - g'(0) = f(y) - f(x) - \langle \nabla f(x),\, y - x\rangle$,
- $0 \leq g''(1) = \left\langle y - x,\, \nabla^2 f(x)(y-x)\right\rangle$.

That yields all the claimed items, except (iv). As the inequality $0 \leq \left\langle y - x,\, \nabla^2 f(x)(y-x)\right\rangle$ holds for all y in the open set U, it also holds in a small ball of radius $r > 0$ centered at x, hence

$$\forall u \in B(0,1),\; r^2 \left\langle u,\, \nabla^2 f(x) u\right\rangle \geq 0,$$

which implies that $\nabla^2 f(x) \succeq 0$.

(ii) or (iii) or (iv) $\Rightarrow$ (i). Conversely, let $x, y \in U$ and let $g : t \mapsto f(tx + (1-t)y)$. For all $t_1, t_2 \in \operatorname{dom}(g)$, setting $z_1 \overset{\text{def.}}{=} t_1 x + (1-t_1)y$ and $z_2 \overset{\text{def.}}{=} t_2 x + (1-t_2)y$, we note that

$$\begin{aligned}
(t_2 - t_1)\left(g'(t_2) - g'(t_1)\right) &= \langle \nabla f(z_2) - \nabla f(z_1),\, z_2 - z_1\rangle\,,\\
g(t_2) - \left(g(t_1) + g'(t_1)(t_2 - t_1)\right) &= f(z_2) - \left(f(z_1) + \langle \nabla f(z_1),\, z_2 - z_1\rangle\right),\\
g''(t) &= \left\langle x - y,\, \nabla^2 f\,(tx + (1-t)y)\,(x-y)\right\rangle
\end{aligned}$$

As a result, by Proposition 4.6, any of (ii), (iii) or (iv) implies that g is convex. Since this holds for any $x, y \in \operatorname{dom}(f)$, we deduce that f is convex. □

[3] The gradient represents the linear form $\mathrm{d}f(x)\colon \mathbb{R}^p \to \mathbb{R}$. It depends on the scalar product. The Hessian represents the bilinear form $\mathrm{d}^2 f(x)\colon \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$.

General convex functions ($\spadesuit$)

So far, in dimension p, we have assumed that f is smooth, *i.e.* $\mathscr{C}^k$ for $k \geq 1$. There are obviously nonsmooth convex functions (*e.g.*, the ℓ^1-norm, $\|x\|_1 = \sum_{i=1}^p |x_i|$), but, as in the one-dimensional case, a convex function cannot be too wild.

Theorem 4.13 (($\spadesuit$))**.** *Let $U \subseteq X$ be a nonempty* open *convex set, and let $f\colon U \to \mathbb{R}$.*

Then f is continuous on U and f is differentiable at x, for Lebesgue almost every $x \in U$.

Proof. The proof consists in observing that f must be Lipschitz on every compact subset of U, see the exercise sheet for the proof of that point.

The Lipschitz property implies both the continuity and the differentiability almost everywhere (this known as the Rademacher theorem), hence the result. □

4.2.2 Strictly convex functions in dimension $p \geq 2$

As for the case of convex functions, we may translate the results in dimension one to the multi-dimensional case.

Proposition 4.14. *Let $U \subseteq X$ be a nonempty* open *convex set, and let $f \in \mathscr{C}^k(U)$, with $k \geq 1$. The following assertions are equivalent*

(i) f is strictly convex,

(ii) for all $x, y \in U$ with $x \neq y$, $\langle \nabla f(x) - \nabla f(y), x - y \rangle > 0$,

(iii) for all $x, y \in U$ with $x \neq y$, $f(y) > f(x) + \langle \nabla f(x), y - x \rangle$,

A criterion involving the Hessian can be written but it is not practical (try to write it as an exercise). The following sufficient condition will be more convenient.

Proposition 4.15. *Let $U \subseteq X$ be a nonempty* open *convex set, and let $f \in \mathscr{C}^2(U)$. If*

$$\forall x \in U, \quad \nabla^2 f(x) \succ 0, \tag{4.16}$$

then f is strictly convex.

4.2.3 Strongly convex functions

Similarly, we may derive the equivalent of Proposition 4.12 for strongly convex functions (defined in Section 3.6)

Proposition 4.16. *Let $U \subseteq X$ be a nonempty* open *convex set, and let $f \in \mathscr{C}^k(U)$, with $k \geq 1$. The following assertions are equivalent* *(i)* f *is* λ*-strongly convex,*

(ii) for all $x, y \in U$*,* $\langle \nabla f(x) - \nabla f(y),\, x - y \rangle \geq \lambda \left\| x - y \right\|^2$*,*

(iii) for all $x, y \in U$*,* $f(y) \geq f(x) + \langle \nabla f(x),\, y - x \rangle + \frac{\lambda}{2} \left\| y - x \right\|^2$*,*

(iv) (for $k \geq 2$*) for all* $x \in U$*,* $\nabla^2 f(x) \succeq \lambda \mathrm{I}_p$*, where* $\mathrm{I}_p \in \mathbb{R}^{p \times p}$ *is the identity matrix.*

Chapter 5

Subgradients and conjugate functions

Contents

5.1 Convex enveloppe of a function 50

5.2 Subgradients of a function 52

5.2.1 Definition 52

5.2.2 Fermat's rule 53

5.2.3 Digression: the relative interior and polyhedral convex functions 53

5.2.4 Subdifferential and operations 54

5.3 Conjugate functions 55

5.3.1 The convex conjugate 55

5.3.2 The double convex conjugate 56

5.4 Fenchel-Rockafellar duality 57

In Chapter 4, we have seen that convex functions have some regularity (*e.g.* continuity on the interior of the domain, differentiability almost everywhere...). Nevertheless, they may well have "kinks" (points of non-differentiability) on some (small) set: think of the absolute value function $x \mapsto |x|$. Such points are very important in optimization, since they tend to attract minimizers when the function is perturbed. In fact, non-differentiability is one of the reasons of the success of sparsity-based methods in data sciences.

Therefore, it is important to be able to derive optimality conditions for such non-smooth problems. Incidentally, our working with extended-valued functions enables us to seamlessly tackle convex constrained optimization problems.

5.1 Convex enveloppe of a function

The cornerstone of this chapter is the following theorem in the theory of convex sets. We call a *closed half-space* a lower level set of a (non-zero) affine function, *i.e.* a set of the form

$$D = \{ x \in X \mid \langle p,\, x \rangle - \alpha \leq 0 \}, \tag{5.1}$$

for some $p \in X$, $\alpha \in \mathbb{R}$. We recall the following consequence of the separating hyperplane theorem, Corollary 2.28.

Theorem 5.1. *Let $C \subseteq X$ be a closed convex set. Then C is the intersection of all the closed half-spaces that contain C.*

Certainly, if $f\colon X \to [-\infty, +\infty]$ is convex lower semi-continuous, we may apply this theorem to $\mathrm{epi}(f) \subseteq X \times \mathbb{R}$. We obtain

Theorem 5.2. *Let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex, l.s.c. function. Then, for all $x \in X$,*

$$f(x) = \sup \{ L(x) \mid L\colon X \to \mathbb{R} \text{ is affine and } L \leq f \}. \tag{5.2}$$

Proof. The proof consists in applying Theorem 5.1 to $\mathrm{epi}(f)$ and write it as the intersection of half-spaces of the form

$$D = \left\{ (x,r) \in X \times \mathbb{R} \mid \left\langle \begin{pmatrix} p \\ s \end{pmatrix}, \begin{pmatrix} x \\ r \end{pmatrix} \right\rangle - \alpha \leq 0 \right\}. \tag{5.3}$$

There are two kinds of half-spaces in $X \times \mathbb{R}$,

- The oblique ones, *i.e.* such that $s \neq 0$.
- The vertical ones, *i.e.* such that $s = 0$.

We show that the oblique ones correpsond to some affine functions which are below f. The vertical ones, however, do not correspond to affine functions, and we need to be a bit more careful.

Let $x_0 \in X$, and $r_0 < f(x_0)$. As the point (x_0, r_0) does not belong to $\mathrm{epi}\, f$, there exists a hyperplane which separates them. In other words, there are some $(p,s) \in X \times \mathbb{R}$, some $\alpha \in \mathbb{R}$, such that

$$\left\langle \begin{pmatrix} p \\ s \end{pmatrix}, \begin{pmatrix} x_0 \\ r_0 \end{pmatrix} \right\rangle > \alpha, \tag{5.4}$$

$$\text{and } \forall (x,r) \in \mathrm{epi}\, f, \quad \left\langle \begin{pmatrix} p \\ s \end{pmatrix}, \begin{pmatrix} x \\ r \end{pmatrix} \right\rangle < \alpha. \tag{5.5}$$

First case: $x_0 \in \mathrm{dom}\, f$, that is $f(x_0) < +\infty$. We may as well write $r_0 = f(x_0) - \varepsilon$ for some $\varepsilon > 0$. Choosing $(x,r) = (x_0, f(x_0))$ in (5.5), we note that (5.4) implies

$$\langle p,\, x_0 \rangle + s f(x_0) < \alpha < \langle p,\, x_0 \rangle + s(f(x_0) - \varepsilon),$$

which implies that $s < 0$ (the hyperplane is oblique). Hence, dividing (5.5) by $(-s) > 0$, we get

$$\forall (x,r) \in \mathrm{epi}\, f, \quad \left\langle \begin{pmatrix} \tilde{p} \\ -1 \end{pmatrix}, \begin{pmatrix} x \\ r \end{pmatrix} \right\rangle < \tilde{\alpha} \tag{5.6}$$

for some $\tilde{p} \in X$, $\tilde{\alpha} \in \mathbb{R}$. In particular, for $r = f(x)$, we get $\tilde{L}(x) \overset{\text{def.}}{=} \langle \tilde{p}, x \rangle - \tilde{\alpha} < f(x)$, while (5.4) implies $\tilde{L}(x_0) > f(x_0) - \varepsilon$. The function $\tilde{L}$ is thus an affine minorant of f, and

$$\sup \{ L(x_0) \mid L \colon X \to \mathbb{R} \text{ is affine and } L \leq f \} \geq \tilde{L}(x_0) > f(x_0) - \varepsilon.$$

If $\Gamma(f)(x_0)$ denotes the left-hand side of the above inequality, letting $\varepsilon \to 0$, we obtain $\Gamma(f)(x_0) \geq f(x_0)$. The converse inequality is immediate, as $L(x_0) \leq f(x_0)$ for every admissible L.

Second case: $x_0 \notin \operatorname{dom} f$, that is, $f(x_0) = +\infty$. If $s < 0$, we proceed as before, and we build an affine minorant of f which takes a value larger than r_0 at x_0.

However, it is possible that $s = 0$ (vertical hyperplane), in which case we proceed slightly differently. We know that there exists at least some point in the domain of f, and from the first case above we deduce that there exists some affine minorant $\tilde{L}_1$ of f: for all $x \in X$, $\tilde{L}_1(x) \leq f(x)$. We set $\tilde{L}_2(x) \overset{\text{def.}}{=} \tilde{L}_1(x) + t\,(\langle p, x \rangle - \alpha)$ for some $t > 0$ chosen large enough so that

$$\tilde{L}_2(x_0) = \tilde{L}_1(x_0) + t\,\underbrace{(\langle p, x_0 \rangle - \alpha)}_{> 0 \text{ by } (5.4)} > r_0. \tag{5.7}$$

Next, for all $x \in \operatorname{dom} f$, we have from (5.5)

$$\tilde{L}_2(x) = \underbrace{\tilde{L}_1(x)}_{\leq f(x)} + t\,\underbrace{(\langle p, x \rangle - \alpha)}_{\leq 0 \text{ by } (5.5)} \leq f(x). \tag{5.8}$$

Hence $\tilde{L}_2$ is an affine minorant of f, and we deduce that $\Gamma(f)(x_0) \geq \tilde{L}_2(x) \geq r_0$. Letting $r_0 \to f(x_0) = +\infty$, we get the desired equality.

□

Remark 5.3. *The assumption that f is proper is important (or more precisely, that $f(x) > -\infty$ for every x). Consider for instance the function*

$$f(x) = \begin{cases} -\infty & \text{if } x = 0 \\ +\infty & \text{if } x \neq 0. \end{cases}$$

Then, the set on the right-hand side of (5.2) is empty, and $\sup \emptyset = -\infty$. Therefore (5.2) does not hold for $x \neq 0$.

If f is any function (not necessarily convex nor l.s.c.), the right-hand side of (5.2) defines a new function which is sometimes called (in French) *la Γ-régularisée de f*,

$$\Gamma(f)(x) \overset{\text{def.}}{=} \sup \{ L(x) \mid L \colon X \to \mathbb{R} \text{ is affine and } L \leq f \}. \tag{5.9}$$

If f has at least one affine minorant (*i.e.* if the r.h.s. of (5.2) or (5.9) is nonempty), the function $\Gamma(f)$ is the *greatest convex lower semi-continuous function below f*, also known as the closed convex hull of f, denoted $\overline{\operatorname{conv}} f$.

In the rest of this chapter we gather information of f by studying the "best" affine minorants L of f at some point x: what are their slopes (subgradient)? what are their intercept (minus the convex conjugate)?

5.2 Subgradients of a function

5.2.1 Definition

Definition 5.4. *Let $f\colon X \to [-\infty, +\infty]$ and let $x \in X$, $p \in X$.*

We say that p is a sugradient to f at x, and we write $p \in \partial f(x)$, if $f(x)$ is finite and

$$\forall y \in X,\ \ f(y) \geq f(x) + \langle p,\, y - x\rangle\,. \tag{5.10}$$

The collection of all subgradients $\partial f(x)$ is called the subdifferential of f at x.

Let us consider a few examples.

Gradient of a smooth convex function. Let f be a convex function which is finite and differentiable at some $x \in X$. By Proposition 4.12, we know that

$$\forall y \in \mathrm{dom}(f), \quad f(y) \geq f(x) + \langle \nabla f(x),\, y - x\rangle\,.$$

Since the inequality also holds for $y \notin \mathrm{dom}(f)$, we have (5.10), and $\nabla f(x) \in \partial f(x)$. In fact, it is the only subgradient (exercise!), and $\partial f(x) = \{\nabla f(x)\}$.

Absolute value function. In $X = \mathbb{R}$, let $f\colon x \mapsto |x|$. Observe that

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

Convex functions on the real line. More generally, if $I \subseteq \mathbb{R}$ is an interval and $f\colon I \to \mathbb{R}$ is convex, we know from (4.4) that it has left and right derivatives at every $x \in \mathrm{int}(I)$. Then,

$$\forall x \in \mathrm{int}(I),\ \ \partial f(x) = \begin{cases} \{f'(x)\} & \text{if } f \text{ is differentiable at } x \\ [f_l'(x), f_r'(x)] & \text{in general.} \end{cases}$$

Indicator of a convex set Let $C \subseteq X$ be a nonempty convex set. Then

$$\partial \chi_C(x) = \left\{\, s \in \mathbb{R}^p \mid \forall x' \in C,\ \left\langle s,\, x' - x\right\rangle \leq 0 \,\right\} = \mathcal{N}_C\,(x) \tag{5.11}$$

is the normal cone to C at x (see the constrained optimization lecture!).

Remark 5.5. *The notion of subgradient is global (defined by inequalities on all X), contrary to the classical notion of gradient. However, if the function is convex, it is sufficient to consider only local information.*

We leave the proof of the following observation to the reader.

Proposition 5.6. *For all $f\colon X \to [-\infty, +\infty]$, and $x \in X$, $\partial f(x)$ is a (possibly empty) closed convex set.*

5.2.2 Fermat's rule

The main reason we are interested in subgradients is the following theorem.

Theorem 5.7 (Fermat's rule)**.** *Let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ a proper function and $x \in X$. Then,*

$$(x \in \operatorname{argmin}_X f) \iff (0 \in \partial f(x)) . \tag{5.12}$$

Proof. It is almost a tautology!

$$\begin{aligned} x \in \operatorname{argmin}_X f &\iff \forall y \in X,\ f(y) \geq f(x) \\ &\iff \forall y \in X,\ f(y) \geq f(x) + \langle 0,\, y - x\rangle \\ &\iff 0 \in \partial f(x), \end{aligned}$$

since $f(x)$ must be finite. □

Fermat's rule holds for any (proper) function f, but it is mostly useful when dealing with convex functions, because we have theorems to compute the subdifferential.

5.2.3 Digression: the relative interior and polyhedral convex functions

In the next section, we describe the subdifferential of functions that result from basic operations (sum, composition…). The most precise formulas hold under some "qualification" conditions which involve two notions.

First, the relative interior of a convex set $C \subseteq X$ is its interior when seen as a subset of $\operatorname{Aff} C$ (the smallest affine space which contains C), see Section 2.4.1 for more detail. We denote it by $\operatorname{rint}(C)$. In other words,

$$\operatorname{rint}(C) \overset{\text{def.}}{=} \{ x \in C \mid \exists r > 0, (B(x,r) \cap \operatorname{Aff} C) \subseteq C \} \tag{5.13}$$

While the interior of a convex set might be empty, one may prove that the relative interior is never empty[1] if $C \neq \emptyset$.

Example 5.8. *The relative interior of $[0,1] \times \{0\} \subseteq \mathbb{R}^2$ is $]0,1[\times \{0\}$, but its interior is $\emptyset$!*

The relative interior of $0 \times \overline{B(0,r)} \subseteq \mathbb{R}^3$ is $0 \times B(0,r)$. Again, its interior is empty.

Second, we say that a convex function is *polyhedral* if its epigraph is polyhedral (*i.e.* the intersection of finitely many closed half spaces). A function is convex polyhedral if and only if it is of the form

$$f(x) = \max_{1 \leq i \leq n} (\langle p_i,\, x\rangle - \alpha_i) + \chi_C(x), \quad \text{where } C = \bigcap_{j=1}^{m} \{ x \in X \mid \langle x,\, q_j\rangle \leq \beta_j \} . \tag{5.14}$$

Example 5.9. *The ℓ^1 norm, $x \mapsto \|x\|_1 = \sum_{i=1}^{p} |x_i|$ is polyhedral.*

[1] In our finite-dimensional setting.

5.2.4 Subdifferential and operations

The result of most operations can be guessed by the known formulas for classical gradients. The scaling operation is the easiest case (prove the proposition below as an exercise!).

Proposition 5.10 (Scaling)**.** *Let $f\colon X \to [-\infty, +\infty]$, and $\alpha > 0$. Then, for all $x \in X$, $\partial(\alpha f)(x) = \alpha\,(\partial f(x))$.*

However, the subdifferential of a sum behaves as expected under some qualification condition.

♡ **Proposition 5.11** (Sum of convex functions)**.** *Let $f, g\colon X \to \mathbb{R} \cup \{+\infty\}$ be two proper functions, and $x \in X$. Then,*

$$\partial f(x) + \partial g(x) \subseteq \partial(f+g)(x). \tag{5.15}$$

If, moreover, f and g are proper convex, and

- $\operatorname{rint}(\operatorname{dom} f) \cap \operatorname{rint}(\operatorname{dom} g) \neq \emptyset$,
- *or g is polyhedral and* $\operatorname{rint}(\operatorname{dom} f) \cap \operatorname{dom} g \neq \emptyset$,

then

$$\partial f(x) + \partial g(x) = \partial(f+g)(x). \tag{5.16}$$

Proof. We only prove the first point. If $f(x)$ and $g(x)$ are finite, for any $p \in \partial f(x)$ and $q \in \partial g(x)$, then for all $y \in X$,

$$\begin{aligned} f(y) &\geq f(x) + \langle p,\, y - x\rangle\,, \\ g(y) &\geq g(x) + \langle q,\, y - x\rangle\,. \end{aligned}$$

Summing both inequalities, we obtain that $p + q \in \partial(f+g)(x)$.

For the second point, the interested reader may consult [Roc97, Th. 23.7]. □

The inclusion in (5.15) may be strict, even if f and g are convex. For instance, let

$$f(x) = -\sqrt{1-x^2} + \chi_{[-1,1]}(x), \quad g(x) = \chi_{[1,2]}(x).$$

Then $f(x) + g(x) = \chi_{\{1\}}(x)$, and it is possible to check that

$$\partial f(1) + \partial g(1) = \emptyset + \mathbb{R}_- = \emptyset \subsetneq \mathbb{R} = \partial(f+g)(x).$$

♡ **Proposition 5.12.** *Let $f\colon x \mapsto h(Ax)$, where $h\colon \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is proper and $A \in \mathbb{R}^{m \times p}$. Then,*

$$A^\top \partial h(Ax) \subseteq \partial f(x). \tag{5.17}$$

If, moreover, h is convex proper and

- $\mathrm{Im} A \cap \mathrm{rint}(\mathrm{dom}\, h) \neq \emptyset$,
- *or h is polyhedral and* $\mathrm{Im} A \cap \mathrm{dom}\, h \neq \emptyset$,

then

$$A^\top \partial h(Ax) = \partial f(x). \tag{5.18}$$

Proof. We only prove the first point. If $p \in (\partial h)(Ax)$, then for all $y \in X$,

$$h(Ay) \geq h(Ax) + \langle p,\, Ay - Ax\rangle = h(Ax) + \left\langle A^\top p,\, y - x\right\rangle.$$

Therefore, $A^\top p \in \partial f(x)$.

For the equality case, we refer to [Roc97, Thm. 23.9] □

5.3 Conjugate functions

5.3.1 The convex conjugate

Let us fix a slope $p \in X$. What is the "best" intercept for an affine minorant of f with slope p?

Let $L(x) = \langle p,\, x\rangle - \alpha$ be an affine minorant. We must have:

$$\forall x \in X,\ \langle p,\, x\rangle - \alpha \leq f(x) \iff \forall x \in X,\ \langle p,\, x\rangle - f(x) \leq \alpha,$$

so that the smallest possible value is $\alpha = \sup_{x\in X} (\langle p,\, x\rangle - f(x))$.

Remark 5.13. *The optimal value α is not necessarily reached (the supremum is not always a maximum). For instance, consider the function $f(x) = \sqrt{x^2+1}$ and the slope $p = 1$ corresponding to the function $L(x) = x$(the graph of L is an oblique asymptote to the graph of f, they never touch).*

Definition 5.14. *The convex conjugate (or Legendre-Fenchel conjugate) of $f\colon X \to [-\infty, +\infty]$ is the function $f^*\colon X \to [-\infty, \infty]$ defined by*

$$\forall p \in X,\ f^*(p) = \sup_{x\in X} \left(\langle p,\, x\rangle - f(x)\right). \tag{5.19}$$

It turns out that the function f^* has some interesting properties, even if f is not convex (check this as an exercise).

Proposition 5.15. *Let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a proper function. Then f^* is convex, l.s.c. and proper.*

Proposition 5.16 (Legendre-Fenchel inequality)**.** *Let f be proper. Then,*

$$\forall x, p \in X, \quad f(x) + f^*(p) \geq \langle p,\, x \rangle \,. \tag{5.20}$$

Moreover,

$$p \in \partial f(x) \Longleftrightarrow f(x) + f^*(p) = \langle p,\, x \rangle \,. \tag{5.21}$$

Proof. By definition of the convex conjugate $f^*(p) \geq \langle p,\, x \rangle - f(x)$. Since $f^*(p) > -\infty$ and $f(x) > -\infty$, we may add $f(x)$ and deduce $f(x) + f^*(p) \geq \langle p,\, x \rangle$.

Now, we characterize the equality case. Using that f is proper, we get

$$\begin{aligned} p \in \partial f(x) &\Longleftrightarrow \forall y \in \operatorname{dom}(f),\ f(y) \geq f(x) + \langle p,\, y - x \rangle \\ &\Longleftrightarrow \forall y \in \operatorname{dom}(f),\ \langle p,\, x \rangle - f(x) \geq \langle p,\, y \rangle - f(y), \\ &\Longleftrightarrow \langle p,\, x \rangle - f(x) = \max_{y \in \operatorname{dom}(f)} \left(\langle p,\, y \rangle - f(y)\right) \\ &\Longleftrightarrow \langle p,\, x \rangle - f(x) = f^*(p) \end{aligned}$$

□

5.3.2 The double convex conjugate

Let us recall the definition of the convex enveloppe and examine its connection with the conjugate function.

$$\begin{aligned} \Gamma(f)(x) &= \sup \{\, L(x) \mid L \text{ is affine and } L \leq f \,\} \\ &= \sup \{\, \langle p,\, x \rangle - \alpha \mid p \in X, \alpha \in \mathbb{R},\ \forall y \in X,\ \langle p,\, x \rangle - \alpha \leq f(x) \,\} \\ &= \sup \{\, \langle p,\, x \rangle - f^*(p) \mid p \in X \,\} \\ &= (f^*)^*(x) \end{aligned}$$

Writing f^{**} for $(f^*)^*$, we have obtained

Theorem 5.17. *If $f \colon X \to [-\infty, +\infty]$, then $f^{**} = \Gamma(f)$. In particular, if f is convex proper and lower semi-continuous, $f^{**} = f$.*

The double conjugate expresses a function f as a supremum of affine functions.

Corollary 5.18. *Let f be a proper, convex, l.s.c. function. Then $p \in \partial f(x)$ if and only if $x \in \partial f^*(p)$.*

Proof. Using the equality case in the Legendre-Fenchel inequality twice,

$$\begin{aligned} p \in \partial f(x) &\Longleftrightarrow f(x) + f^*(p) = \langle p,\, x \rangle \\ &\Longleftrightarrow f^{**}(x) + f^*(p) = \langle p,\, x \rangle \\ &\Longleftrightarrow x \in \partial f^*(p). \end{aligned}$$

□

5.4 Fenchel-Rockafellar duality

In this section we consider some convex proper, lower semi-continuous functions $f: X \to \mathbb{R} \cup \{+\infty\}$ and $g: Y \to \mathbb{R} \cup \{+\infty\}$ together with a linear map $K: X \to Y$.

We are interested in the variational problem

$$\min_{x\in X} f(x) + g(Kx) \tag{5.22}$$

Typically $g \circ K$ is some loss function, and f is a regularization term, added to avoid overfitting. For instance, if f is the ℓ^1 norm and g is a square fidelity term, (5.22) is the LASSO. Alternatively, (5.22) may encode some constrained optimization problem, with $g = \chi_C$ the indicator function of the constraint set and f the objective.

Provided some qualification conditions hold (see for the middle inequality below), a point x^* is a minimizer[2] if and only if

$$0 \in \partial\left(f + g \circ K\right)(x^*) = \partial f(x^*) + K^\top \partial g(Kx^*). \tag{5.23}$$

In other words, x^* is optimal iff there exists some $z \in Y$ such that

$$z \in \partial g(Kx^*), \quad \text{ and } \; -K^\top z \in \partial f(x^*). \tag{5.24}$$

It turns out that z is an optimal solution to some optimization problem too. That result is known as duality.

For now, let us discuss the general situation (without qualification conditions).

Proposition 5.19 (Weak duality)**.** *Let $f: X \to \mathbb{R} \cup \{+\infty\}$ and $g: Y \to \mathbb{R}\cup\{+\infty\}$ be convex proper, lower semi-continuous functions and $K: X \to Y$ a linear map. Then the following inequality holds*

$$\inf_{x\in X} \left(f(x) + g(Kx)\right) \geq \sup_{z\in Y} \left(-f^*(K^\top z) - g^*(-z)\right). \tag{5.25}$$

Proof. For all $x \in X$, $z \in Y$, the Fenchel-Young inequality yield,

$$f(x) \geq \langle x, K^* z\rangle - f^*(K^\top z) \tag{5.26}$$

$$g(Kx) \geq \langle Kx, -z\rangle - g^*(-z). \tag{5.27}$$

Summing both equalities yields

$$f(x) + g(Kx) \geq -f^*(K^\top z) - g^*(-z),$$

and we obtained the claimed inequality by taking the infimum over x and the supremum over z. □

[2] Note that our assumptions do not guarantee the existence of x^*, but that is a different topic.

The problem

$$\sup_{z\in Y}\left(-f^*(K^\top z) - g^*(-z)\right). \tag{5.28}$$

is known as the *Fenchel-Rockafellar dual* to (5.22). It is a maximization problem for a concave objective (which is equivalent to a convex minimization problem by considering minus the objective). On the other hand, our original problem (5.22) is usually referred to as the *primal problem*.

It is interesting to note that, if equality holds in (5.25) (which is called *strong duality*) and if x^* and z^* are respectively optimal for the left and right hand sides of (5.25), then equality must hold in (5.26) and (5.27). In other words, we must have the following extremality relations,

$$\begin{cases} K^\top z^* \in \partial f(x^*) \\ \quad\;\; z^* \in \partial g(Kx^*). \end{cases} \tag{5.29}$$

Conversely, if the extremality relations (5.29) hold, then (5.25) is an equality and x^* and z^* are respectively optimal for the primal and dual problem.

Theorem 5.20 (Strong duality)**.** *Under the assumptions of Proposition 5.19, if there exists some $x_0 \in X$ such that $x_0 \in \operatorname{rint}\operatorname{dom} f$ and $Kx_0 \in \operatorname{rint}(\operatorname{dom} g)$, then strong duality holds*

$$\inf_{x\in X}\left(f(x)+g(Kx)\right) = \sup_{z\in Y}\left(-f^*(K^\top z) - g^*(-z)\right). \tag{5.30}$$

and there exists a solution to the dual problem (5.28)*.*

Moreover, given any pair $(x^*, z^*) \in X \times Y$, the following assertions are equivalent:

- *x^* is a solution to* (5.22) *and z^* is a solution to* (5.28)*,*
- *x^* and z^* satisfy the extremality conditions* (5.29)*.*

See for instance [Roc97, Cor. 31.2.1].

Appendix A

Reminder on topology

We recall here some basic notions of topology. Yet, we cannot summarize here all the useful results that are useful in optimization. The reader who is looking for a pedagogical and deep introduction to topology may consult [CCM97].

A.1 Interior, closure

Let $A \subseteq X$, and $x \in A$. As usual, we denote by

$$B(x,r) \stackrel{\text{def.}}{=} \{\, y \in X \mid \|x-y\| < r \,\}\,.$$

the open ball centered at x and with radius $r > 0$.

Given the point x, three cases may arise (see Figure A.1):

1. there exists $r > 0$ such that $B(x,r) \subseteq A$; we say that x lies *in the interior* of A, and we write $x \in \operatorname{int}(A)$.
2. for all $r > 0$, $B(x,r) \cap A \neq \emptyset$ and $(B(x,r) \setminus A) \neq \emptyset$; we say that x lies *on the boundary* of A, we write $x \in \partial A$.
3. there exists $r > 0$ such that $B(x,r) \subseteq (X \setminus A)$; we say that x lies *in the exterior* of A (which is nothing but the interior of $X \setminus A$).

The closure of A is defined as

$$\overline{A} \stackrel{\text{def.}}{=} \operatorname{int}(A) \cup (\partial A)\,.$$

We say that a set A is *closed* if $\overline{A} = A$. If A is the complement of a closed set, we say that A is open. A set is open if and only if it is equal to its interior (hence if and only if it is a union of open balls).

Proposition A.1 (The closure as the set of all possible limits)**.** *Let $A \subseteq X$, and $a \in X$. Then $a \in \overline{A}$ if and only if there is a sequence $(a_n)_{n\in\mathbb{N}}$ of elements in A such that $a = \lim_{n\to\infty} a_n$.*

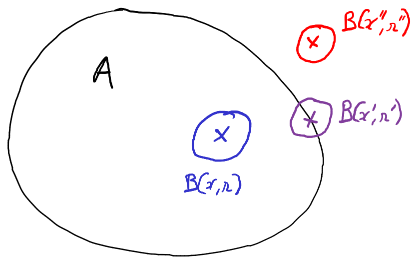

Figure A.1: The point x is in the interior of A, x' is on the boundary of A, and x'' is in the exterior of A.

A.2 Continuity

Let $X_1 = \mathbb{R}^{p_1}$ and $X_2 = \mathbb{R}^{p_2}$. We say that a function is continuous at $x \in X_1$ if for every $\varepsilon > 0$, there exists $r > 0$ such that $f(B(x,r)) \subseteq B(x,\varepsilon)$.

Proposition A.2 (Sequential characterization of continuity)**.** *The function f is continuous at $x \in X_1$ if and only if for every sequence $(x_n)_{n\in\mathbb{N}}$ of elements of X_1 which converges to some $x \in X_1$,*

$$f(x) = \lim_{n\to\infty} f(x_n).$$

Proposition A.3. *The following assertions are equivalent:*

1. *the function f is continuous on X_1 (i.e. at every point of X_1),*
2. *the inverse image of every open subset of X_2 is an open subset of X_1,*
3. *the inverse image of every closed subset of X_2 is an open subset of X_1.*

Remark A.4. *Similar statements hold for functions $f\colon A_1 \to A_2$ where A_i is a subset of X_i ($i \in \{1,2\}$) by considering the relative topology of A_i, i.e. in all the definitions of interior, boundary ..., the balls $B(x,r)$ should be replaced with $(B(x,r) \cap A_i)$.*

Appendix B

Lower and upper limits of sequences and functions.

B.1 Lower and upper limits of sequences

Let $(u_n)_{n\in\mathbb{N}}$ a sequence of *real* numbers. In general, there is no reason for $\lim_{n\to\infty} u_n$ to exist (*e.g.* take $u_n = (-1)^n$). If $(u_n)_{n\in\mathbb{N}}$ is bounded, we know by the Bolzano-Weierestrass theorem that we can extract a subsequence $(u_{\varphi(n)})_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} u_{\varphi(n)} = \ell$ for some $\ell \in \mathbb{R}$.

But can we choose the number ℓ? For instance, for $u_n = (-1)^n$ we could have $\ell = 1$ or -1 depending on the subsequence) ? And what if $(u_n)_{n\in\mathbb{N}}$ is not bounded?

The notion of *limit inferior* and *limit superior* address those issues.

Definition B.1. *Let $u = (u_n)_{n\in\mathbb{N}}$ be a sequence of* real numbers*. The limit inferior of u is*

$$\liminf_{n\to\infty} u_n \overset{\text{def.}}{=} \lim_{n\to\infty}\left(\inf_{k\geq n} u_k\right). \tag{B.1}$$

The limit inferior of u is

$$\limsup_{n\to\infty} u_n \overset{\text{def.}}{=} \lim_{n\to\infty}\left(\sup_{k\geq n} u_k\right). \tag{B.2}$$

Remark B.2. *The sequence $\left(\inf_{k\geq n} u_k\right)_{n\in\mathbb{N}}$ is increasing, hence by the monotone limt theorem, the limit in (B.1) always exists in $[-\infty, +\infty]$. Similarly, the sequence $\left(\sup_{k\geq n} u_k\right)_{n\in\mathbb{N}}$ is decreasing, so that the limit in (B.2) always exists.*

Remark B.3. *The limit inferior (resp. superior) is the lowest (highest) value $\ell \in [-\infty, +\infty]$ such that we may extract a subsequence which converges to ℓ.*

Example B.4. *Using the definition, we note*

- *For* $u_n = (-1)^n$, $\liminf_{n\to\infty}(-1)^n = -1$, *whereas* $\limsup_{n\to\infty}(-1)^n = 1$.
- *For* $u_n = n$, $\liminf_{n\to\infty} u_n = \limsup_{n\to\infty} u_n = +\infty$.
- *For* $u_n = -n$, $\liminf_{n\to\infty} u_n = \limsup_{n\to\infty} u_n = -\infty$.
- *For* $u_n = \sin(n)e^n$, $\liminf_{n\to\infty} u_n = -\infty$, $\limsup_{n\to\infty} u_n = +\infty$.

Proposition B.5. *Let* $u = (u_n)_{n\in\mathbb{N}}$ *be a sequence of* real numbers. *If* $\lim_{n\to\infty} u_n$ *exists, then* $\liminf_{n\to\infty} u_n = \limsup_{n\to\infty} u_n = \lim_{n\to\infty} u_n$.

Proposition B.6. *Let $u = (u_n)_{n\in\mathbb{N}}$ be a sequence of* real numbers. *Then*

- $\liminf_{n\to\infty}(-u_n) = -\limsup_{n\to\infty}(u_n)$
- $\liminf_{n\to\infty}(\alpha u_n) = \alpha\left(\liminf_{n\to\infty} u_n\right)$ *for all* $\alpha > 0$,
- $\limsup_{n\to\infty}(\alpha u_n) = \alpha\left(\limsup_{n\to\infty} u_n\right)$ *for all* $\alpha > 0$,
- $\liminf_{n\to\infty} u_n \leq \limsup_{n\to\infty} v_n$,
- *for all sequence* $(v_n)_{n\in\mathbb{N}}$ *in* $\mathbb{R}$,

$$\left(\liminf_{n\to\infty} u_n\right) + \left(\liminf_{n\to\infty} v_n\right) \leq \liminf_{n\to\infty}(u_n + v_n)$$

provided the left-hand side is not the inderterminate form $\infty - \infty$ *(or* $-\infty + \infty$*), and similarly*

$$\limsup_{n\to\infty}(u_n + v_n) \leq \left(\limsup_{n\to\infty} u_n\right) + \left(\limsup_{n\to\infty} v_n\right)$$

provided the right-hand side is not the inderterminate form $\infty - \infty$ *(or* $-\infty + \infty$*)*

Appendix C

References

There are many interesting references on convex and variational analysis.

The most authoritative treatise on convex analysis is perhaps the monograph by R.T. Rockafellar [Roc97], which contains a surprisingly large amount of results in a moderate number of pages, while being easy to read.

The two volume book [HUL93] by J.-B. Hiriart-Urruty and C. Lemaréchal covers a lot of material as well, with a lot of details and figures. It also deals with algorithmic aspect of convex optimization (which are not covered in [Roc97]). The reference [BV04] is also a comprehensive and pedagogical reference, and is illustrated with modern problems from signal processing or statistics. It is available on the website of its authors[1]. For the specificic case of univariate convex functions, the reader may consult the elegant study in [Bou07]; the corresponding section (Section 4.1.1) is directly inspired from it.

All the above-mentioned references are in finite dimension. The reader who is interested in infinite-dimensional analysis may consult [ET99, Roc89] and, in French, [ET74, Mor67].

[1] https://web.stanford.edu/~boyd/cvxbook/

Index

Coercive (function), 16
Convex hull, 24

Effective domain, 10
Epigraph, 11
Exposed point, 28
Extended valued function, 9
Extreme point, 27

Half space, 30
Hyperplane, 30

Lower semi-continuous (function), 13

Proper function, 10

Relative boundary, 27

Separating hyperplane, 30
Strictly convex (function), 18

Bibliography

[AB06] Charalambos D. Aliprantis and Kim C. Border. *Infinite dimensional analysis: a hitchhiker's guide.* Springer, Berlin ; New York, 3rd [rev. and enl.] ed edition, 2006. OCLC: ocm69983226.

[Bou07] N. Bourbaki. *Fonctions d'une variable réelle: Théorie élémentaire.* Éléments de Mathématique. Springer-Verlag, Berlin Heidelberg, 2007.

[Bre11] Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations.* Springer New York, New York, NY, 2011.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[CCM97] Gilles Christol, Anne Cot, and Charles-Michel Marle. *Topologie.* Mathématiques pour le 2e cycle. Ellipses, Paris, 1997.

[DM93] Gianni Dal Maso. *An Introduction to* Γ*-convergence.* Number v. 8 in Progress in nonlinear differential equations and their applications. Birkhäuser, Boston, MA, 1993.

[ET74] I. Ekeland and R. Temam. *Analyse convexe et problèmes variationnels.* Etudes mathématiques. Dunod, 1974.

[ET99] I. Ekeland and Roger Temam. *Convex analysis and variational problems.* Number 28 in Classics in applied mathematics. Society for Industrial and Applied Mathematics, Philadelphia, 1999.

[HUL93] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex Analysis and Minimization Algorithms I*, volume 305 of *Grundlehren der mathematischen Wissenschaften.* Springer Berlin Heidelberg, Berlin, Heidelberg, 1993.

[Mor67] Jean Jacques Moreau. Fonctionnelles convexes. *Séminaire Jean Leray*, (2):1–108, 1966-1967.

[Roc89] Ralph Tyrrell Rockafellar. *Conjugate duality and optimization*. Number 16 in Regional conference series in applied mathematics. Soc. for Industrial and Applied Mathematics, Philadelphia, Pa, 2. print edition, 1989. OCLC: 258450647.

[Roc97] R. Tyrrell Rockafellar. *Convex analysis*. Princeton landmarks in mathematics and physics. Princeton Univ. Press, Princeton, NJ, 1997. OCLC: 247646720.