

Université Paris-Dauphine

Advanced Convex Analysis

Paul Pegon

2024–2025

Contents

1	Convex sets	3
1.1	Affine sets, convex sets and cones	3
1.2	Geometric and topological properties of convex sets	6
1.3	The Hahn-Banach theorems and external representation theorems	8
1.4	The internal representation theorems of Minkowski and Krein-Milman	16
2	Convex functions	19
2.1	Definition and elementary properties	19
2.2	Regularity	23
2.3	Subdifferential	31
2.4	First order optimality conditions	36
2.5	Lower semicontinuous convex functions	38
2.6	Fenchel-Legendre conjugate	41
2.7	Convex duality	44

1 Convex sets

In this course, the setting will always be a vector space or a normed vector space $\mathcal{X}$, equipped with a norm $\|\cdot\|$, endowing $\mathcal{X}$ with a topology. We will denote by $B(x, r)$ and $B[x, r]$ respectively the open and closed ball of center x and radius r , and by $\mathcal{X}^*$ the dual space of $\mathcal{X}$, the space of continuous linear forms x^* equipped with the dual norm $\|x^*\|^* = \sup_{x \in \mathcal{X}: \|x\| \leq 1} f(x)$. We mimic the notation of the scalar product in a Hilbert space by writing $\langle x^*, x \rangle = x^*(x)$. The bilinear map $\langle \cdot, \cdot \rangle : \mathcal{X}^* \times \mathcal{X} \rightarrow \mathbb{R}$ is called the *duality pairing* between $\mathcal{X}^*$ and $\mathcal{X}$.

Many statements still hold in the more general setting of locally convex topological vector spaces.

1.1 Affine sets, convex sets and cones

Given two points $x, y \in \mathcal{X}$, consider the point

$$x_\theta = (1 - \theta)x + \theta y$$

with parameter $\theta \in \mathbb{R}$. When θ runs through $\mathbb{R}$, x_θ describes the *line* passing through x and y . Notice that $x_0 = x$ and $x_1 = y$: more generally when $\theta \in [0, 1]$ the point x_θ describes the *line segment* between x and y , denoted by $[x, y]$. When $\theta > 1$, x_θ is “after” y , and when $\theta < 0$ it is “before” x . When $x = 0$ and $\theta \geq 0$, x_θ describes the *ray* (or *half-line*) based at 0 and passing by y . Similarly, we define $[x, y) = \{(1 - \theta)x + \theta y : \theta \in [0, 1)\}$ and $(x, y), (x, y]$.

Definitions and examples

Definition 1.1 (Convex sets, affine sets and cones). A subset E of a vector space $\mathcal{X}$ is

- an affine set if for every $x, y \in E$ the line $L(x, y) = \{(1 - \theta)x + \theta y : \theta \in \mathbb{R}\}$ is included in E ,
- a convex set if for every $x, y \in E$ the line segment $[x, y] = \{(1 - \theta)x + \theta y : \theta \in [0, 1]\}$ is included in E ,
- a cone if for every $y \in E$ the ray $\mathbb{R}_+ y = \{\theta y : \theta \in \mathbb{R}_+\}$ is included in E ,
- a convex cone if it is both a convex set and a cone, or equivalently if $\lambda x + \mu y \in E$ for every $\lambda, \mu \geq 0$ and $x, y \in E$.

It is clear that an affine set is a convex set.

Example 1.2. • Open and closed balls $B(x, r)$ and $B[x, r]$ in a normed space are convex sets.

- Hyperplanes, i.e. subsets of the form $\{x \in \mathcal{X} : f(x) = \alpha\}$ where $f : \mathcal{X} \rightarrow \mathbb{R}$ is a linear form and $\alpha \in \mathbb{R}$ are affine sets.
- The sets S_d^+ (resp. S_d^{++}) of nonnegative semidefinite (resp. definite) symmetric matrices of size $d \times d$ are convex cones.
- The subset $\{f : a \leq f(x) \leq b \text{ for every } x \in \mathbb{R}\}$ of the vector space $\mathcal{F}(\mathbb{R}, \mathbb{R})$ of functions from $\mathbb{R}$ to $\mathbb{R}$, where $a \leq b$, is a convex set.

Combinations and hulls

Given n points $(x_1, \dots, x_n)$ and n reals $(\theta_1, \dots, \theta_n)$, a combination of the form

$$\sum_{i=1}^n \theta_i x_i$$

is called

- an *affine combination* of the x_i 's if $\sum_{i=1}^n \alpha_i = 1$,
- a *conic combination* of the x_i 's if $\alpha_i \geq 0$ for every i ,
- a *convex combination* of the x_i 's if $\sum_{i=1}^n \alpha_i = 1$ and $\alpha_i \geq 0$ for every i .

The notions of affine sets, convex sets and convex cones can be re-phrased with these combinations.

Lemma 1.3. *A subset $E \subseteq \mathcal{X}$ is (an affine set/a convex set/a convex cone) if and only if every (affine/convex/conic) combination of points in E is also in E .*

Proposition 1.4. 1. *The image or pre-image of an (affine set/convex set) by an affine map is an (affine set/convex set).*

2. *The image or pre-image of a convex cone by a linear map is a convex cone.*

3. *An intersection of (affine sets/convex sets/convex cones) is still (an affine set/a convex set/a convex cone).*

The stability by intersection allows us to define (affine/convex/conic) hulls.

Definition-Proposition 1.5. *For every $E \subseteq \mathcal{X}$, there exists a smallest (affine set/convex set/convex cone) containing E , which is called the (affine hull/convex hull/conic hull) of E , denoted by $\text{aff}(E)/\text{conv}(E)/\text{cone}(E)$. Moreover:*

$$\begin{aligned} \text{aff}(E) &= \bigcap \{C : C \supseteq E, C \text{ affine}\} = \{\text{affine combinations of points of } E\}, \\ \text{conv}(E) &= \bigcap \{C : C \supseteq E, C \text{ convex}\} = \{\text{convex combinations of points of } E\}, \\ \text{cone}(E) &= \bigcap \{C : C \supseteq E, C \text{ convex cone}\} = \{\text{conic combinations of points of } E\}. \end{aligned}$$

Operations on (convex) sets

We will need to dilate (convex) sets by setting

$$\lambda A := \{\lambda x : x \in A\}$$

for every $A \subseteq \mathcal{X}$ and $\lambda \in \mathbb{R}$. We will also need to “sum” convex sets, in the sense of Minkowski:

Definition 1.6 (Minkowski sum). If $A, B \subseteq \mathcal{X}$ we define the Minkowski sum of A and B as

$$A + B = \{x + y : x \in A, y \in B\},$$

and the Minkowski difference

$$A - B = \{x - y : x \in A, y \in B\} = A + (-B)$$

If $x \in \mathcal{X}$ we will write for short $x \pm B = \{x\} \pm B$ and $A \pm x = A \pm \{x\}$.

Proposition 1.7. Let $A, B \subseteq \mathcal{X}$ be two convex sets, and $s, t \in \mathbb{R}$. Then:

1. $sA \pm tB$ is convex,
2. $sA + tA = (s + t)A$ if $s, t \geq 0$.

Remark 1.8. The second item is not true if A is not convex: for example taking $A = [-2, -1] \cup [1, 2]$ and $s = t = 1$ we have $A + A = [-4, -2] \cup [-1, 1] \cup [2, 4]$ while $2A = [-4, -2] \cup [2, 4]$. It is also not true if A, B are convex but s, t are not nonnegative: take for example $A = [0, 1]$ and $s = t = -1$, so that $A + (-A) = [-1, 1] \neq \{0\} = 0A$. However, it is clear that we always have

$$(s + t)A \subseteq sA + tA,$$

A being any subset and s, t any reals.

Proof. For the first item, it is clear that sA is convex, thus it suffices to show that $A + B$ is convex. Take $z, z' \in A + B$, thus $z = x + y, z' = x' + y'$ where $x, x' \in A$ and $y, y' \in B$, and take $\theta \in [0, 1]$. We have

$$(1 - \theta)z + \theta z' = (1 - \theta)(x + y) + \theta(x' + y') = [(1 - \theta)x + \theta x'] + [(1 - \theta)y + \theta y'] \in A + B$$

because A and B are convex.

For the second item, we already remarked that $(s + t)A \subseteq sA + tA$, we only have to show the converse inclusion. Let $sx \in sA$ and $ty \in tA$. If $s = t = 0$ it is clear that $sx + ty = 0 \in (s + t)A$, thus we assume that $t + s > 0$. We have:

$$sx + ty = (s + t) \left(\frac{s}{s + t}x + \frac{t}{s + t}y \right) \in (s + t) \operatorname{conv}(A) = (s + t)A$$

because A is convex. □

1.2 Geometric and topological properties of convex sets

Lemma 1.9. *Let C be a nonempty convex subset of $\mathcal{X}$.*

1. *The closure of C is convex.*
2. *If $x_0 \in \text{int } C$ and $x \in C$ then $[x_0, x) \subseteq \text{int } C$. In particular if C has nonempty interior, it is the closure of its interior.*

Proof. Let C be a nonempty convex set. Let $x, y \in \overline{C}$ and $\theta \in [0, 1]$. There exists sequences (x_n) and (y_n) in C such that $x_n \rightarrow x$ and $y_n \rightarrow y$. Since C is convex,

$$\forall n, (1 - \theta)x_n + \theta y_n \in C,$$

and by passing to the limit in n we get

$$(1 - \theta)x + \theta y \in \overline{C},$$

which shows that $\overline{C}$ is convex.

Let $x \in C$ and $x_0 \in \text{int } C$, so that $B(x_0, r) \subseteq \text{int } C$ for a small $r > 0$. By convexity of C , for every $\theta \in [0, 1)$ the set

$$(1 - \theta)B(x_0, r) + \theta x$$

is a subset of C , but it is also the ball of center $x_\theta = (1 - \theta)x_0 + \theta x$ and radius $(1 - \theta)r$, thus $x_\theta \in \text{int } C$.

Let C of non-empty interior. To show that $\overline{\text{int } C} = \overline{C}$, it suffices to show that $C \subseteq \overline{\text{int } C}$. Fix $x_0 \in \text{int } C$ and let $x \in C$. Considering x_θ as above for $\theta \in [0, 1)$, $x_\theta \rightarrow x$ as $\theta \rightarrow 1$ and we get $x \in \overline{\text{int } C}$. $\square$

Definition 1.10 (Closed convex hull). If A is a set, we denote by $\overline{\text{conv}} A$ the smallest closed and convex set containing A , which is called the *called convex hull* of A .

Proposition 1.11. *For every subset A , $\overline{\text{conv}}(A) = \overline{\text{conv } A}$.*

Proof. Left as exercise. $\square$

In the 3-dimensional euclidean space $\mathbb{R}^3$, the disk $D = \{(x, y, 0) : x^2 + y^2 \leq 1\}$ is a convex set with empty interior. Nonetheless, we are tempted to say that “relative to D ”, the set $\{(x, y, 0) : x^2 + y^2 < 1\}$ is the interior of D , and $\{(x, y, 0) : x^2 + y^2 = 1\}$ is its boundary. This is formalized by the notion of *relative interior* and *relative boundary*.

Definition 1.12. Let C is a convex subset of $\mathcal{X}$.

- The *relative interior* of C is the interior of C as a subset of the affine set $\text{aff}(C)$, namely:

$$\text{ri } C = \{x : \exists r > 0, B(x, r) \cap \text{aff}(C) \subseteq C\}.$$

- The *relative boundary* of C is the boundary of C as a subset of the affine set $\text{aff}(C)$, namely:

$$\text{rbd } C = \overline{C} \setminus \text{ri } C.$$

Proposition 1.13. *Let C be a nonempty convex set. If $x_0 \in \text{ri } C, x \in C$ and $\theta \in [0, 1)$, then $(1 - \theta)x_0 + \theta x \in \text{ri } C$. In particular $\overline{\text{ri } C} = \overline{C}$.*

Proof. It is a variation of the proof of the fact that $\overline{\text{int } C} = \overline{C}$ whenever $\text{int } C \neq \emptyset$, left as an exercise. $\square$

Basics on affine space If $E \subseteq \mathcal{X}$ is a nonempty affine set, then for any $x_0 \in E$,

$$E - x_0$$

is a vector space, which does not depend on the choice of x_0 , it is called the *direction* of E and denoted by $\vec{E}$.

As a real vector space, it has a dimension $\dim \vec{E}$, equal to $+\infty$ if $\vec{E}$ is not finite-dimensional, and to the usual dimension in linear algebra when $\vec{E}$ is a finite-dimensional vector space. We thus set

$$\dim E = \dim \vec{E}$$

If $(x_0, \dots, x_d)$ are $d + 1$ points such that

$$\left(\sum_i \theta_i = 1 \quad \text{and} \quad \sum_i \theta_i x_i = 0 \right) \implies \forall i, \theta_i = 0$$

we say that they are affinely independent. If moreover $A = \text{aff}\{x_0, \dots, x_d\}$ we say that $(x_0, \dots, x_d)$ is an affine basis of A .

Definition 1.14 (Dimension of a convex set). If C is a nonempty convex set we define

$$\dim C = \dim(\text{aff } E).$$

A convex set of dimension 0 is a point.

Proposition 1.15. *Let A be an affine subspace with affine basis $(x_0, \dots, x_n)$. Then the relative interior of $\text{conv}(\{x_0, \dots, x_n\})$ is not empty and equal to:*

$$\text{ri conv}(\{x_0, \dots, x_n\}) = \left\{ \sum \theta_i x_i : \sum_i \theta_i = 1 \text{ and } \forall i, \theta_i > 0 \right\}.$$

Proof. Since $(x_0, \dots, x_n)$ is an affine basis of A , $(x_1 - x_0, \dots, x_n - x_0)$ is a vector basis of $\vec{A}$. Consider its dual base $(e_1^*, \dots, e_n^*)$ (basis of $\vec{A}^*$ such that $e_i^*(x_j - x_0) = \delta_{ij}$). Consider a point $x \in A$, which can thus be written as $\sum_{i=0}^n \theta_i x_i$ with $\sum_{i=0}^n \theta_i = 1$. We have $x - x_0 = \sum_{i=1}^n \theta_i (x_i - x_0)$ so that $\theta_i = e_i^*(x - x_0)$ for $1 \leq i \leq n$ and $\theta_0 = 1 - \sum_{i=1}^n e_i^*(x - x_0)$. As a consequence

$$\text{conv}\{x_0, \dots, x_n\} = \left\{ x \in A : \forall i \in \{1, \dots, n\}, e_i^*(x - x_0) \geq 0 \text{ and } \sum_{i=1}^n e_i^*(x - x_0) \leq 1 \right\},$$

and one can easily deduce its relative interior (i.e. interior in A)

$$\begin{aligned} \text{ri conv}\{x_0, \dots, x_n\} \\ = \left\{ x \in A : \forall i \in \{1, \dots, n\}, e_i^*(x - x_0) > 0 \text{ and } \sum_{i=1}^n e_i^*(x - x_0) < 1 \right\}. \end{aligned}$$

□

Definition 1.16 (Extreme points). A point $u \in C$ which cannot be written as convex combination $u = (1 - \theta)x + \theta y$ with $x, y \in C$ and $\theta \in (0, 1)$ is called an *extreme point*.

Definition 1.17 (Face). Let C be a nonempty convex subset of $\mathcal{X}$. A convex subset $F \subseteq C$ is called a *face* of C if

$$\forall x, y \in C, \quad (x, y) \cap F \neq \emptyset \implies [x, y] \subseteq F.$$

It is called a *proper face* if $F \neq C$.

Remark 1.18. • 0-dimensional faces are exactly the singletons $\{u\}$ where u is an extreme point of C .

- Faces of codimension 1, i.e. faces F such that $\text{aff } F$ is a hyperplane, are called *facets* of C .

Lemma 1.19. Let $C \subseteq \mathcal{X}^*$ be convex set and F be a face of C . A subset $G \subseteq F$ is a face of F if and only if it is a face of C . In particular

$$\text{ext } F = F \cap \text{ext } C.$$

Proof. Left as exercise. □

1.3 The Hahn-Banach theorems and external representation theorems

A fundamental result in convex analysis is the identification of closed convex sets with intersections of closed half-spaces, the proof of which will rely on the Hahn-Banach extension theorems.

The Hahn-Banach extension theorem

Definition 1.20 (Sublinear functions). A function $p : \mathcal{X} \rightarrow \mathbb{R}$ is called *sublinear* if it is

- *positively 1-homogeneous*: $p(\lambda x) = \lambda p(x)$ for every $x \in \mathcal{X}$ and $\lambda > 0$,
- *subadditive*: $p(x + y) \leq p(x) + p(y)$ for every $x, y \in \mathcal{X}$.

Lemma 1.21. Let $p : \mathcal{X} \rightarrow \mathbb{R}$ be a sublinear function. The following assertions are equivalent:

(i) p is continuous,

(ii) p is continuous at 0,

(iii) $\sup_{x \in \mathcal{X}: \|x\| \leq 1} p(x) = \sup_{x \in \mathcal{X} \setminus \{0\}} \frac{p(x)}{\|x\|} < +\infty$.

Proof. (i) $\implies$ (ii) is clear. Assume that p is continuous at 0. By definition, there exists $\delta > 0$ such that $\|x\| \leq \delta$ implies $|p(x)| \leq 1$. By homogeneity, for every $x \in \mathcal{X}$ such that $\|x\| \leq 1$,

$$p(x) = \frac{1}{\delta} p(\delta x) \leq \frac{1}{\delta} \times 1,$$

and (iii) follows. Assume now that $M := \sup_{x \in \mathcal{X}: \|x\| \leq 1} p(x) < +\infty$. Take $x_0 \in \mathcal{X}$. By subadditivity we have

$$p(x) \leq p(x - x_0) + p(x_0) \quad \text{and} \quad p(x_0) \leq p(x_0 - x) + p(x)$$

thus

$$|p(x) - p(x_0)| \leq \max\{|p(x - x_0)|, |p(x_0 - x)|\} \leq M \|x - x_0\| \xrightarrow{x \rightarrow x_0} 0,$$

which proves (i). $\square$

Norms or semi-norms (which satisfy all axioms of norms except for the separation axiom) are sublinear functions. Other examples arise as *gauges* of some convex sets, that we shall see later.

Theorem 1.22 (Hahn-Banach). *Let $\mathcal{X}$ be a normed vector space and $p : \mathcal{X} \rightarrow \mathbb{R}$ a continuous sublinear function on $\mathcal{X}$. If E is a linear subspace of $\mathcal{X}$ and if f is a linear form on E satisfying $f \leq p$, then f can be extended into a continuous linear form $\hat{f} \in \mathcal{X}^*$ which still satisfies $\hat{f} \leq p$.*

Remark 1.23. We shall prove this under the extra assumption that $\mathcal{X}$ is separable, although it holds in general thanks to Zorn's lemma, and in an even more general setting ($\mathcal{X}$ being just a vector space, without continuity assumption on p).

Proof (continuous and separable case). Let $(u_i)_{i \in \mathbb{N}}$ a countable dense subset of $\mathcal{X}$ and set $E_0 = E$, $E_i = \text{span } E \cup \{v_k : k \leq i\}$ for $i \geq 1$. We have $E_0 \subseteq E_1 \subseteq \dots$. We shall extend f by induction to the E_i 's, starting from E_0 .

From E_{i-1} to E_i . We assume to have constructed a linear extension $f_{i-1} : E_{i-1} \rightarrow \mathbb{R}$ such that $f_i \leq p$ and wish to extend it to some $f_i : E_i \rightarrow \mathbb{R}$. If $u_i \in E_{i-1}$ then $E_i = E_{i-1}$ and there is nothing to do. If not, we shall define f_i by setting:

$$\forall x \in E_{i-1}, \forall t \in \mathbb{R}, \quad f_i(x + tu_i) = f_{i-1}(x) + t\alpha,$$

1 Convex sets

for some well-chosen $\alpha \in \mathbb{R}$. We want $f_i \leq p$, that is:

$$\begin{aligned} & \forall x \in E_{i-1}, \forall t \in \mathbb{R}, \quad f_{i-1}(x) + t\alpha \leq p(x + tu_i) \\ \iff & \forall x \in E_{i-1}, \forall t \in \mathbb{R}^*, \quad f_{i-1}(x/|t|) + \frac{t}{|t|}\alpha \leq p\left(\frac{x}{|t|} + \frac{t}{|t|}u_i\right) \\ \iff & \forall x \in E_{i-1}, \forall t \in \mathbb{R}^*, \quad f_{i-1}(x) \pm \alpha \leq p(x \pm u_i) \\ \iff & \underbrace{\sup_{x \in E_{i-1}} f_{i-1}(x) - p(x - u_i)}_{=m} \leq \alpha \leq \underbrace{\inf_{y \in E_{i-1}} p(y - u_i) - f_{i-1}(y)}_{=M} \end{aligned}$$

where we have used the homogeneity for the first equivalence. Thus, such a real α exists if and only if $m \leq M$ (since $m > -\infty$ and $M < +\infty$), i.e.

$$\begin{aligned} & \forall (x, y) \in E_{i-1}, \quad f_{i-1}(x) - p(x - u_i) \leq p(y + u_i) - f_{i-1}(y) \\ \iff & \forall (x, y) \in E_{i-1}, \quad f_{i-1}(x + y) \leq p(y + u_i) + p(x - u_i) \\ \iff & \forall (x, y) \in E_{i-1}, \quad f_{i-1}(x + y) \leq p(y + x), \end{aligned}$$

where we have used the subadditivity of p in the last implication. Since $f_{i-1} \leq p$ by assumption, we deduce the existence of a real α such that $f_i \leq p$.

Extension to $\mathcal{X}$. Since the f_j is an extension of f_i for $i \leq j$ we can extend f to $\tilde{E} = \bigcup_i E_i$ by setting (without ambiguity) $\tilde{f}(x) = f_i(x)$ if $x \in E_i$. We still have $\tilde{f} \leq p$. We now wish to extend it to $\mathcal{X}$ by density of $\tilde{E}$ (since it contains the dense family $(u_i)_{i \in \mathbb{N}}$). Notice that for every $x \in \tilde{E}$,

$$\tilde{f}(x) \leq p(x) \quad \text{and} \quad \tilde{f}(x) = -\tilde{f}(-x) \geq -p(-x)$$

thus

$$|\tilde{f}(x)| \leq \max\{|p(x)|, |p(-x)|\}.$$

By [Lemma 1.21](#) and by continuity of p ,

$$\sup_{x \in \tilde{E}: \|x\| \leq 1} |\tilde{f}(x)| \leq \sup_{x \in \mathcal{X}: \|x\| \leq 1} p(x) < +\infty.$$

This shows that $\tilde{f}$ is a continuous linear form on $\tilde{E}$, thus it can be extended by continuity to a linear form $\hat{f}$ on its closure $\mathcal{X}$. By continuity of p , we still have $\hat{f} \leq p$. $\square$

Corollary 1.24. *If f is a continuous linear form on a subset $\mathcal{Y} \subseteq \mathcal{X}$, then it can be extended into a continuous linear form x^* on the whole space $\mathcal{X}$ with same norm $\|f\| = \|x^*\|$.*

The Hahn-Banach separation theorems

Definition 1.25 (Linear separation). Two sets $A, B \subseteq \mathcal{X}$ are (linearly) separated if there exists $x^* \in \mathcal{X}^* \setminus \{0\}$ such that

$$\sup_{x \in A} \langle x^*, x \rangle \leq \inf_{x \in B} \langle x^*, x \rangle$$

and they are *strictly separated* if the inequality is strict.

We shall prove that a disjoint pair (convex - open convex) is separated, and a pair (closed convex - compact convex) is strictly separated. We start with a fundamental lemma on the gauge of a convex neighborhood of the origin.

Definition 1.26 (Gauge of a convex set). The *gauge* of a convex set $C \subseteq \mathcal{X}$ is the function $g_C : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ defined by

$$g_C(x) = \inf\{r > 0 : x \in rC\}.$$

Lemma 1.27. *Let C be a non-empty convex set such that $0 \in C$. We have:*

- (i) g_C is sublinear,
- (ii) $\{g_C < 1\} \subseteq C \subseteq \{g_C \leq 1\}$,
- (iii) if $0 \in \text{int } C$ then g_C is Lipschitz.

Proof. We start with the sublinearity. Clearly $g_C(0) = 0$ because $0 \in C$. Let $x, y \in \mathcal{X}$, $\lambda > 0$. One has

$$\begin{aligned} g_C(\lambda x) &= \inf\{r > 0 : \lambda x \in rC\} \\ &= \inf\{\lambda r > 0 : \lambda x \in \lambda rC\} = \lambda \inf\{r > 0 : x \in rC\} = \lambda g_C(x). \end{aligned}$$

Besides, for every $s > g_C(x)$ and $t > g_C(y)$ we have $x \in sC$ and $y \in tC$ thus by [Proposition 1.7](#),

$$(x + y) \in sC + tC = (s + t)C \implies g_C(x + y) \leq s + t.$$

Sending $s \rightarrow g_C(x)$ and $t \rightarrow g_C(y)$ yields $g_C(x + y) \leq g_C(x) + g_C(y)$.

Now let us show continuity of g_C . Since C is a neighborhood of 0, it contains a small ball $B[0, \delta]$ with $\delta > 0$, thus for every x such that $\|x\| \leq 1$,

$$x \in B[0, 1] = 1/\delta B[0, \delta] \subseteq \frac{1}{\delta}C \implies g_C(x) \leq \frac{1}{\delta},$$

which concludes by [Lemma 1.21](#).

If $g_C(x) < 1$, we can find $r < 1$ such that $x \in rC$ thus x is a convex combination of 0, which belongs to C , and an element of C , thus $x \in C$. $\square$

Theorem 1.28 (Hahn-Banach separation theorem). *Let A, B be two nonempty and disjoint convex subsets of $\mathcal{X}$, one of which is open. Then A and B can be separated*

Proof. The proof relies on the Hahn-Banach extension theorem. The first step is to establish the following: if C is a nonempty open convex set and $x \notin C$ then C and $\{x\}$ can be separated. Notice that C and $\{x\}$ can be separated if and only if $C - x_0$ and $\{x - x_0\}$ can be separated, for any given $x_0 \in \mathcal{X}$. Taking any $x_0 \in C$, the set $C - x_0$ contains 0. Up to renaming, we may thus assume that C contains 0, thus it is

a neighborhood of 0. By [Lemma 1.27](#), g_C is a continuous sublinear function on $\mathcal{X}$. We define a linear form ℓ on $\mathbb{R}x$ by setting

$$\forall t \in \mathbb{R}, \quad \ell(tx) = t,$$

so that $\ell(x) = 1$. Since $x \notin C$, by [Lemma 1.27](#) we know that $g_C(x) \geq 1 = \ell(x)$. As a consequence $\ell \leq g_C$ of $\mathbb{R}x$ and ℓ is continuous (as a linear form on a finite-dimensional space). We may apply the Hahn-Banach extension theorem [Theorem 1.22](#) to extend ℓ to some $x^* \in \mathcal{X}^*$ which satisfies $\langle x^*, y \rangle \leq g_C(y)$ for every $y \in \mathcal{X}$. By [Lemma 1.27](#), $g_C \leq 1$ on C thus

$$\sup_{y \in C} \langle x^*, y \rangle \leq \sup_{y \in C} g_C(y) \leq 1 = \ell(x) = \langle x^*, x \rangle,$$

which implies that C and x are separated.

We now tackle the separation of A and B , assuming for example that A is open. We consider the set

$$C = A - B$$

which by [Proposition 1.7](#) is a convex set. Notice that C is open, because

$$C = \bigcup_{b \in B} (A - b)$$

and all the $A - b$ are open. Moreover $0 \notin C$ for otherwise we would have some $(a, b) \in A \times B$ such that $0 = a - b$, i.e. $a = b \in A \cap B$, which cannot happen because A and B are disjoint. By what we have seen above, C and 0 are separated by some $x^* \in \mathcal{X}^* \setminus \{0\}$:

$$\begin{aligned} & \forall x \in C, \langle x^*, x \rangle \leq \langle x^*, 0 \rangle = 0 \\ \iff & \forall a \in A, \forall b \in B, \langle x^*, a \rangle \leq \langle x^*, b \rangle \\ \iff & \sup_{a \in A} \langle x^*, a \rangle \leq \inf_{b \in B} \langle x^*, b \rangle \end{aligned} \quad \square$$

As a corollary, we may state the strict separation theorem of Hahn-Banach.

Corollary 1.29 (Hahn-Banach strict separation theorem). *Let C, D be two nonempty convex subsets of $\mathcal{X}$.*

1. *If $\text{dist}(C, D) > 0$ then C and D are strictly separated.*
2. *In particular, if C is closed, D is compact and $C \cap D = \emptyset$, then they are strictly separated.*

Proof. Let C, D as in the first item and set $\varepsilon = d(C, D) > 0$. By definition of the distance,

$$(C + B(0, \varepsilon)) \cap D = \emptyset.$$

By the Hahn-Banach separation theorem ([Theorem 1.28](#)), $C + B(0, \varepsilon)$ and D are separated by some $x^* \in \mathcal{X}^*$. Notice that

$$\begin{aligned} \sup_{x \in C + B(0, \varepsilon)} \langle x^*, x \rangle &= \sup_{c \in C, u \in B(0, \varepsilon)} \langle x^*, c + u \rangle \\ &= \sup_{c \in C} \langle x^*, c \rangle + \sup_{u \in B(0, \varepsilon)} \langle x^*, u \rangle = \sup_{c \in C} \langle x^*, c \rangle + \varepsilon \|x^*\|^*, \end{aligned}$$

and thus

$$\sup_{c \in C} \langle x^*, c \rangle < \sup_{x \in C+B(0, \varepsilon)} \langle x^*, x \rangle \leq \inf_{d \in D} \langle x^*, d \rangle,$$

so that C and D are strictly separated.

For the second item, it suffices to show that a closed set C and a compact set D which are disjoint are at positive distance. If on the contrary $d(A, B) = 0$, then one can find sequences $(a_n) \in A, (b_n) \in B$ such that $d(a_n, b_n) \rightarrow 0$. Since A is compact, (a_n) admits a converging subsequence $(a_{n_k})_k$, of limit $a \in A$. By the triangle inequality, $d(b_{n_k}, a) \rightarrow 0$ so that b_{n_k} converges to a as well. Since B is closed, $a = \lim_{n_k} b_{n_k} \in B$ thus $a \in A \cap B$: a contradiction. $\square$

Remark 1.30 (Separation of weak topology). As a corollary, we deduce that the weak topology on $\mathcal{X}$ (the coarsest topology which makes all element $x^* \in \mathcal{X}^*$ continuous) is separated. Indeed take $x \neq y$ in $\mathcal{X}$. Since singletons are compact and convex, they are strictly separated, i.e. there exists $x^* \in \mathcal{X}^*$ such that

$$\langle x^*, x \rangle < \langle x^*, y \rangle,$$

thus we have found two disjoint neighborhoods of x and x^* for the weak-* topology: respectively $\{\langle x^*, \cdot \rangle < \alpha\}$ and $\{\langle x^*, \cdot \rangle > \alpha\}$ where $\alpha = \frac{1}{2}(\langle x^*, x \rangle + \langle x^*, y \rangle)$.

External representation of closed convex sets

We shall characterize a closed convex set by the half-spaces that contain it. We start by a characterization of a closed half-space. Recall that a half space is by definition a subset of the form

$$\{x : \ell(x) \leq \alpha\}$$

where $\alpha \in \mathbb{R}$ and ℓ a linear form (not necessarily continuous) on $\mathcal{X}$.

Lemma 1.31 (Closed half-spaces). *We consider a linear form $\ell : \mathcal{X} \rightarrow \mathbb{R}$ and $\alpha \in \mathbb{R}$. The following assertions are equivalent:*

- (i) *the half-space $\{x : \ell(x) \leq \alpha\}$ is closed,*
- (ii) *the hyperplane $\{x : \ell(x) = \alpha\}$ is closed,*
- (iii) *the linear form ℓ is continuous.*

Proof. It is clear that the third item implies the two first ones since the half-space is the reciprocal image by the continuous map ℓ of the closed set $(-\infty, \alpha]$ and the hyperplane of the closed set $\{\alpha\}$.

Now, if the half-space is closed, its complement is open, and in particular there exists a ball $B(x_0, \varepsilon) \subseteq \{\ell > \alpha\}$. Thus for $y \in B(0, 1)$,

$$x = x_0 + \varepsilon y \in B(x_0, \varepsilon) \implies \ell(x_0 + \varepsilon y) > \alpha \implies \ell(y) \geq \frac{\alpha - \ell(x_0)}{\varepsilon},$$

and by changing y into $-y$ yields

$$\sup_{B(0,1)} |\ell(y)| \leq \frac{|\alpha - \ell(x_0)|}{\varepsilon},$$

thus ℓ is continuous. If the hyperplane is closed, its complement is open and there also exists a ball $B(x_0, \varepsilon) \subseteq \{\ell > \alpha\}$: we conclude in the same way. $\square$

To give a precise description of a closed convex set in terms of half-space, we will use its *support function*.

Definition 1.32 (Support function). The *support function* $\sigma_A : \mathcal{X}^* \rightarrow \mathbb{R} \cup \{+\infty\}$ associated with a subset $A \subseteq \mathcal{X}$ is defined by

$$\sigma_A(x^*) = \sup\{\langle x^*, x \rangle : x \in A\}.$$

Proposition 1.33. *The support function associated with a set A is sublinear.*

Proof. Left as an exercise. $\square$

Definition 1.34. Let C be a convex set. If $x \in C$ and $x^* \in \mathcal{X}^*$ are such that

$$\sigma_C(x^*) = \langle x^*, x \rangle$$

then the hyperplane $\{\langle x^*, \cdot \rangle = \sigma_C(x^*)\}$ is called a *support hyperplane* of C at x .

Remark 1.35. Notice that such a hyperplane intersects C “tangentially” at x , in the sense that

$$C \subseteq \{\langle x^*, \cdot \rangle \leq \sigma_C(x^*)\} \text{ and } x \in C \cap \{\langle x^*, \cdot \rangle = \sigma_C(x^*)\},$$

Corollary 1.36 (External representation of closed convex sets). *Any nonempty closed convex set $C \subseteq \mathcal{X}$ is the intersection of all closed half-spaces that contain C , more precisely:*

$$C = \bigcap_{x^* \in \mathcal{X}^*} \{x : \langle x^*, x \rangle \leq \sigma_C(x^*)\}.$$

Consequently, the closed convex hull of a nonempty subset A can be written as:

$$\overline{\text{conv}} A = \bigcap_{x^* \in \mathcal{X}^*} \{x : \langle x^*, x \rangle \leq \sigma_A(x^*)\}.$$

Proof. By definition of σ_C , the intersection of all closed half-spaces that contain C is precisely

$$L = \bigcap_{x^* \in \mathcal{X}^*} \{x : \langle x^*, x \rangle \leq \sigma_C(x^*)\}.$$

Obviously $C \subseteq L$, so it suffices to show the inclusion $L \subseteq C$, or equivalently $\mathcal{X} \setminus C \subseteq \mathcal{X} \setminus L$. Let $x \notin C$. Since C is closed, we may apply the Hahn-Banach strict separation theorem (Corollary 1.29) to find some $x^* \in \mathcal{X}^*$ such that

$$\sigma_C(x^*) = \sup_{y \in C} \langle x^*, y \rangle < \langle x^*, x \rangle.$$

Thus $x \in \mathcal{X} \setminus \{y : \langle x^*, y \rangle \leq \sigma_C(x^*)\} \subseteq \mathcal{X} \setminus L$. $\square$

Definition 1.37 (Normal cone). Let C be a convex set. We define the (outer) normal cone of C at $x \in C$ by

$$\mathbf{N}_x C = \{x^* \in \mathcal{X}^* : \forall y \in C, \langle x^*, y - x \rangle \leq 0\}.$$

It is a closed convex cone. By definition, we set $\mathbf{N}_x C = \emptyset$ if $x \notin C$.

Remark 1.38 (Closedness). Notice that

$$\mathbf{N}_x C = \bigcap_{y \in C} \{x^* \in \mathcal{X}^* : \langle x^*, y - x \rangle \leq 0\},$$

thus it is an intersection of pre-images of closed sets $(-\infty, 0]$ by evaluation maps $\langle \cdot, y - x \rangle$, so that the normal cone is closed (even weakly-* closed).

Proposition 1.39. *Let C be a convex set and $x \in C$. One has*

1. $x^* \in \mathbf{N}_x(C)$ if and only if $\{\langle x^*, \cdot \rangle = \sigma_C(x^*)\}$ is a support hyperplane of C at x , i.e.

$$x^* \in \mathbf{N}_x C \iff \sigma_C(x^*) = \langle x^*, x \rangle.$$

2. The intersection of C with a support hyperplane at x is a face of C , i.e.

$$x^* \in \mathbf{N}_x C \implies C \cap \{\langle x^*, \cdot \rangle = \sigma_C(x^*)\} \text{ is a face of } C.$$

Proof. For the first item, it follows from the following chain of equivalence:

$$\begin{aligned} x^* \in \mathbf{N}_x C &\iff \forall y \in C, \quad \langle x^*, y \rangle \leq \langle x^*, x \rangle \\ &\iff \sigma_C(x^*) \doteq \sup_{y \in C} \langle x^*, y \rangle \leq \langle x^*, x \rangle \\ &\iff \sigma_C(x^*) = \langle x^*, x \rangle. \end{aligned}$$

The last equivalence is due to the fact that $x \in C$ and thus $\sigma_C(x^*) \geq \langle x^*, x \rangle$.

For the second item, suppose $x^* \in \mathbf{N}_x C$. Let $u, v \in C$ such that there exists $\theta \in (0, 1)$ satisfying $(1 - \theta)u + \theta v \in \{\langle x^*, \cdot \rangle = \langle x^*, x \rangle\}$. By definition of the normal cone at x , we have

$$\langle x^*, u \rangle \leq \langle x^*, x \rangle \text{ and } \langle x^*, v \rangle \leq \langle x^*, x \rangle$$

thus

$$\begin{aligned} \langle x^*, x \rangle &= \langle x^*, (1 - \theta)u + \theta v \rangle = (1 - \theta)\langle x^*, u \rangle + \theta\langle x^*, v \rangle \\ &\leq (1 - \theta)\langle x^*, x \rangle + \theta\langle x^*, x \rangle = \langle x^*, x \rangle. \end{aligned}$$

Therefore, the inequalities above are actually equalities:

$$\langle x^*, u \rangle = \langle x^*, x \rangle \text{ and } \langle x^*, v \rangle = \langle x^*, x \rangle.$$

It means that $u, v \in C \cap \{\langle x^*, \cdot \rangle = \langle x^*, x \rangle\}$, and thus $C \cap \{\langle x^*, \cdot \rangle = \langle x^*, x \rangle\}$ is a face of C . $\square$

The existence of a nonzero element in the normal cone at a (relative) boundary point of a non-empty convex set C , and the existence of a proper face, are not always guaranteed, but we may provide some sufficient conditions.

Proposition 1.40. *Let C be a closed convex set. If $\text{ri} C \neq \emptyset$ and $x \in \text{rbd} C$ then $\mathbf{N}_x C \neq \{0\}$.*

Proof. Assume that there exists some $x_0 \in \text{ri} C$, and take $x \in \text{rbd} C$. We know that $x \in \text{aff}(C) \setminus \text{ri} C$, and that $\text{ri} C$ is an open subset of $\text{aff}(C)$. By the Hahn-Banach separation theorem, there exists a continuous linear form $\ell : \text{aff}(C) \rightarrow \mathbb{R}$ which separates $\text{ri} C$ and x in the ambient space $\text{aff}(C)$. By the Hahn-Banach extension theorem, ℓ may be extension to an element $x^* \in \mathcal{X}^*$, which separates $\text{ri} C$ and x in the whole space $\mathcal{X}^*$:

$$\sup_{y \in C} \langle x^*, y \rangle = \sup_{y \in C} \ell(y) < \ell(x) = \langle x^*, x \rangle,$$

which implies that $x^* \in \mathbf{N}_x C$. □

Proposition 1.41. *If C is a nonempty compact convex set and $x^* \in \mathcal{X}^* \setminus \{0\}$, there is $x \in \text{bd} C$ such that $x^* \in \mathbf{N}_x C$, thus $\mathbf{N}_x C \neq \{0\}$ and $F_{x^*} = C \cap \{\langle x^*, \cdot \rangle = \sigma_C(x^*)\}$ is a face of C containing x .*

Proof. Consider any $x^* \in \mathcal{X}^*$. Since x^* is continuous and C is compact, x^* attains its maximum at some $x \in C$:

$$\sigma_C(x^*) = \sup_{y \in C} \langle x^*, y \rangle = \langle x^*, x \rangle.$$

By [Proposition 1.39](#), we get the result. □

Proposition 1.42. *If C is a closed convex set in $\mathbb{R}^d$, then for every $x \in C$,*

$$\begin{cases} \mathbf{N}_x C = \{0\} & \text{if } x \in \text{int } C \\ \mathbf{N}_x C \neq \{0\} & \text{if } x \in \text{bd } C. \end{cases}$$

Proof. Seen as an exercise. □

1.4 The internal representation theorems of Minkowski and Krein-Milman

The aim of this section is to give a description of compact convex sets which is not external but rather internal, using their extreme points.

The finite-dimensional case

We have seen that the convex hull of a set E is obtained by forming every convex combination $\sum_{i=1}^n \theta_i x_i$ of points $x_i \in E$, where n is a priori arbitrary. However, it is not difficult to see that in $\mathbb{R}^2$, the convex hull of E can be obtained by considering at most $n = 3$ points: every point $y \in \text{conv}(E)$ lies in some (perhaps flat) triangle with vertices belonging to E . This is a general fact that holds in $\mathbb{R}^d$, where $d + 1$ points suffice.

Theorem 1.43 (Caratheodory's Theorem). *Let $E \subseteq \mathbb{R}^d$ be nonempty.*

1. *Any nonzero element of $\text{cone}(E)$ can be written as a conic combination of at most d (linearly independent) vectors of E .*
2. *Any element of $\text{conv}(E)$ can be written as a convex combination of at most $d + 1$ (affinely independent) points of E .*

Proof. We start with the first item. Take $x \in \text{cone}(E)$ such that $x \neq 0$ and consider $n \in \mathbb{N}^*$ the smallest integer such that x can be written as a conic combination of n points of E , as $x = \sum_{i=1}^n \theta_i x_i$. Necessarily, $\theta_i > 0$ for every i . Let us show that the x_i 's are linearly independent (thus there are at most d). Arguing by contradiction, we assume that they are not: $\sum_i \lambda_i x_i = 0$ for some λ_i 's which are not all zero. Up to changing all λ_i 's by $-\lambda_i$ we may assume that for some j , $\lambda_j > 0$. Then we have

$$x = \sum_{i=1}^n (\theta_i - \varepsilon \lambda_i) x_i, \quad \text{where } \varepsilon = \min\{\theta_i / \lambda_i : \lambda_i > 0\}.$$

By construction, $\theta_i - \varepsilon \lambda_i \geq 0$ for every i , and moreover for an index j which attains the minimum, that is $\varepsilon = \theta_j / \lambda_j$, $\lambda_j > 0$, we have $\theta_j - \varepsilon \lambda_j = 0$. We can thus remove x_j in the conic combination, to find x as a conic combination of $n - 1$ vectors: a contradiction.

We move to the second item. We consider $F = \{(x, 1) : x \in E\} \subseteq \mathbb{R}^{d+1}$ and notice the equivalence:

$$x \in \text{conv}(E) \iff (x, 1) \in \text{cone}(F).$$

Taking $x \in \text{conv}(E)$ and applying the previous item to $(x, 1) \in \text{cone}(F)$ allows us to write $(x, 1)$ as a conic combination

$$(x, 1) = \sum_{i=1}^n \theta_i (x_i, 1)$$

where $n \leq d + 1$. Extracting the last coordinate in the previous equality yields $1 = \sum_{i=1}^n \theta_i$, thus the d first coordinates give $x = \sum_{i=1}^n \theta_i x_i$ as a convex combination of at most $d + 1$ elements. Moreover, it is not difficult to see that the (x_i) are affinely independent in $\mathbb{R}^d$ since the vectors $(x_i, 1)$ are linearly independent in $\mathbb{R}^{d+1}$. $\square$

Theorem 1.44 (Minkowski Theorem). *If C a compact convex subset of a finite-dimensional space, then it is equal to the convex hull of its extreme points.*

Proof. The proof is the object of an (nontrivial) exercise. $\square$

Corollary 1.45 (Minkowski-Carathéodory). *Let C be a compact convex subset of $\mathbb{R}^d$. Any point $x \in C$ is a convex combination of at most $d+1$ (affinely independent) extreme points of C .*

The infinite-dimensional case

Theorem 1.46 (Krein-Milman Theorem). *If C be a nonempty compact convex set, then C is the closed convex hull of its extreme points.*

Proof. First of all, let us show that $\text{ext } C$ is not empty. We shall use Zorn's Lemma. We know by [Proposition 1.41](#) that C admits at least a compact face. Moreover, if $(F_i)_i$ is an ordered (for the order $\supseteq$) family of compact faces of C , it admits a minorant: $F = \bigcap_i F_i$, which is non-empty face. Thus the family of compact faces of C equipped with $\supseteq$ is an inductive set and by Zorn's lemma it admits a minimal element G . If G is not a singleton, it admits two distinct points $x \neq y$. By the Hahn-Banach separation theorem there exists $x^* \in \mathcal{X}^*$ which separates them, for example

$$\langle x^*, x \rangle < \langle x^*, y \rangle \leq \sigma_G(x^*).$$

By [Proposition 1.41](#), there exists $z \in \text{bd } G$ such that $G' = G \cap \{\langle x^*, \cdot \rangle = \sigma_G(x^*)\}$ is a compact face of G , thus a face of C by [Lemma 1.19](#), such that $x \notin G'$: a contradiction to the minimality of G . Thus G is a singleton, consisting of an extreme point as noticed in [Remark 1.18](#).

We set $A = \overline{\text{conv}}(\text{ext } C)$. Let us show that $A = C$. It is clear that $\overline{\text{conv}}(\text{ext } C) \subseteq C$ so it suffices to establish the inclusion $C \subseteq A$. By contradiction, assume that there exists $x \in C \setminus A$. By the Hahn-Banach separation theorem we can (strictly) separate the compact set $\{x\}$ and the closed set A by a linear form $x^* \in \mathcal{X}^*$:

$$\sigma_A(x^*) = \sup_{y \in A} \langle x^*, y \rangle < \langle x^*, x \rangle \leq \sup_{y \in C} \langle x^*, y \rangle = \sigma_C(x^*). \quad (1.4.1)$$

By [Proposition 1.41](#), the set $F = C \cap \{\langle x^*, \cdot \rangle = \sigma_C(x^*)\}$ is a compact face of C . By what we have proved before, it admits an extreme point z , and by [Lemma 1.19](#) $z \in \text{ext } C$. But by [\(1.4.1\)](#), $F \cap A = \emptyset$ thus $z \notin A = \overline{\text{conv}}(\text{ext } C)$: a contradiction with $z \in \text{ext } C$. $\square$

2 Convex functions

2.1 Definition and elementary properties

Let $\mathcal{X}$ be a real vector space. We shall define the notion of convex function for functions that may take a real or $+\infty$ value, and are defined on the whole space $\mathcal{X}$. The connection with the usual definition of real-valued convex functions on convex sets is clarified later.

Recall that by convention $0 \times (+\infty) = 0$, and by definition $a < +\infty$ for every $a \in \mathbb{R}$.

Definition 2.1 (Domain and epigraph). If $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ we define its *domain* and *epigraph* as

$$\text{dom } f = \{x \in \mathcal{X} : f(x) < +\infty\}.$$

and

$$\text{epi } f = \{(x, t) \in \mathcal{X} \times \mathbb{R} : f(x) \leq t\}.$$

We say that f is proper whenever $\text{dom } f \neq \emptyset$.

Remark 2.2. Notice that $\text{epi } f \subseteq \text{dom } f \times \mathbb{R}$ and $\pi_{\mathcal{X}}(\text{epi } f) = \text{dom } f$.

Definition 2.3. A function $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ is said

- *convex* if

$$\forall x, y \in \mathcal{X}, \forall \theta \in [0, 1], \quad f((1 - \theta)x + \theta y) \leq (1 - \theta)f(x) + \theta f(y),$$

- *strictly convex* if

$$\forall x \neq y \in \mathcal{X}, \forall \theta \in (0, 1), \quad f((1 - \theta)x + \theta y) < (1 - \theta)f(x) + \theta f(y),$$

- λ -*strongly convex* for some $\lambda > 0$ if

$$\forall x, y \in \mathcal{X}, \forall \theta \in [0, 1], \quad f((1 - \theta)x + \theta y) \leq (1 - \theta)f(x) + \theta f(y) - \lambda \theta(1 - \theta)\|y - x\|^2.$$

Remark 2.4 (Concave function). One can provide analogous definitions for concavity, which will correspond to convexity of $-f$. Namely: f is concave (resp strictly concave, λ -strongly concave) if $-f$ is convex (resp. strictly convex, λ -strongly convex).

Remark 2.5. By a direct induction argument, f is convex if and only if for every $(x_1, \dots, x_N) \in \mathcal{X}^N$ and all $(\theta_1, \dots, \theta_N) \in \mathbb{R}_+^N$ such that $\sum_i \theta_i = 1$,

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

2 Convex functions

Remark 2.6 (Extension and restriction of a convex function). The domain $C = \text{dom } f$ of a convex function f is necessarily a convex set, and if we restrict f to its domain, it becomes a real-valued function $g = f|_{\text{dom } f} : C \rightarrow \mathbb{R}$ that satisfies

$$\forall x, y \in C, \forall t \in [0, 1], \quad f((1-t)x + ty) \leq (1-t)f(x) + tf(y). \quad (2.1.1)$$

This corresponds probably to the definition of convex function the reader already knows.

Conversely, if $f : C \rightarrow \mathbb{R}$ is a real-valued function such that C is convex and (2.1.1) is satisfied, then its extension $\hat{f} : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ defined by

$$\hat{f}(x) = \begin{cases} f(x) & \text{if } x \in C \\ +\infty & \text{if not} \end{cases}$$

is a convex function in the sense of **Definition 2.3**.

Remark 2.7 (Modulus of convexity). Let $\omega : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a nondecreasing function such that¹ $\omega(0) = 0$ and $\lim_{t \rightarrow 0} \omega(t) = 0$ and $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$.

- We say that f is ω -convex (or has *convexity modulus* ω) if

$$\forall x, y \in \mathcal{X}, \forall \theta \in [0, 1], \quad f((1-\theta)x + \theta y) + \theta(1-\theta)\omega(\|y-x\|) \leq (1-\theta)f(x) + \theta f(y).$$

- We say that f is ω -semi convex if

$$\forall x, y \in \mathcal{X}, \forall \theta \in [0, 1], \quad f((1-\theta)x + \theta y) \leq (1-\theta)f(x) + \theta f(y) + \theta(1-\theta)\omega(\|y-x\|).$$

- If f is convex, we say that it is ω -smooth if

$$\forall x, y \in \mathcal{X}, \forall \theta \in [0, 1], \quad f((1-\theta)x + \theta y) + \theta(1-\theta)\omega(\|y-x\|) \geq (1-\theta)f(x) + \theta f(y).$$

Notice that ω -convexity for $\omega \equiv 0$ corresponds to usual convexity, and ω -convexity for $\omega(t) = \frac{\lambda}{2}t^2$ corresponds to λ -strong convexity. The notion of ω -smoothness will be clarified later, but we can already say that if f is convex and ω -smooth (for some nice enough ω), ω will induce some modulus of continuity for ∇f .

Finally, these notions can be transposed to concavity: f is ω -concave/semi concave if $-f$ is ω -convex/semi convex, and a concave function f is σ -smooth if the convex function $-f$ is σ -smooth.

Definition 2.8 (Indicator function). If $A \subseteq \mathcal{X}$, we define its *indicator function* as

$$\chi_A(x) = \begin{cases} 0 & \text{if } x \in A \\ +\infty & \text{otherwise.} \end{cases}$$

Lemma 2.9. *If $A \subseteq \mathcal{X}$ then:*

$$\chi_A \text{ is a convex function} \iff A \text{ is a convex set.}$$

¹We call such a function a *modulus*

2 Convex functions

Proof. Assume that χ_A is convex. If $x, y \in A$ and $\theta \in [0, 1]$, then

$$\chi_A((1 - \theta)x + \theta y) \leq (1 - \theta)\chi_A(x) + \theta\chi_A(y) = (1 - \theta)0 + \theta 0 = 0,$$

thus $\chi_A((1 - \theta)x + \theta y) = 0$ and $(1 - \theta)x + \theta y \in A$.

Conversely, assume that A is a convex set. Let $x, y \in \mathcal{X}$ and $\theta \in [0, 1]$. We want to show that

$$\chi_A((1 - \theta)x + \theta y) \leq (1 - \theta)\chi_A(x) + \theta\chi_A(y).$$

If $\theta \in \{0, 1\}$ it holds trivially, and if $\theta \in (0, 1)$ and either $x \notin A$ or $y \notin A$ then the right-hand side is $+\infty$ and the inequality holds. If $\theta \in (0, 1)$, $x \in A$ and $y \in A$, then $(1 - \theta)x + \theta y \in A$ and both sides are 0: the inequality holds in any case. $\square$

Proposition 2.10 (Epigraph characterization of convexity). *Let $f : \mathcal{X} \rightarrow +\infty$. The following equivalence holds:*

$$f \text{ is a convex function} \iff \text{epi } f \text{ is a convex set.}$$

Proof. Assume that f is convex and take two point $(x, t), (y, s) \in \text{epi } f$. Let $\theta \in [0, 1]$. We have:

$$f((1 - \theta)x + \theta y) \leq (1 - \theta)f(x) + \theta f(y) \leq (1 - \theta)t + \theta s.$$

This implies that

$$(1 - \theta)(x, t) + \theta(y, s) = ((1 - \theta)x + \theta y, (1 - \theta)t + \theta s) \in \text{epi } f.$$

Conversely, assume that $\text{epi } f$ is convex. Take $x, y \in \mathcal{X}$ and $\theta \in [0, 1]$. We want to show that

$$f((1 - \theta)x + \theta y) \leq (1 - \theta)f(x) + \theta f(y). \quad (2.1.2)$$

We assume that $x \neq y$ and $\theta \in (0, 1)$, otherwise it holds trivially. If either $f(x) = +\infty$ or $f(y) = +\infty$ then the right-hand side of (2.1.2) is equal to $+\infty$ thus (2.1.2) holds. Otherwise, $x, y \in \text{dom } f$ and $(x, f(x)), (y, f(y)) \in \text{epi } f$. Consequently, by convexity of $\text{epi } f$ we get

$$((1 - \theta)x + \theta y, (1 - \theta)f(x) + \theta f(y)) \in \text{epi } f,$$

which is equivalent to $(1 - \theta)f(x) + \theta f(y) \in \mathbb{R}$ and

$$f((1 - \theta)x + \theta y) \leq (1 - \theta)f(x) + \theta f(y),$$

which is what we wanted to prove. $\square$

Remark 2.11. If f is convex, then its sublevel sets $\{x : f(x) \leq \ell\}$ (and $\{x : f(x) < \ell\}$) for every $\ell \in \mathbb{R}$ are convex. However, the converse is not true: take for example $f(x) = \sqrt{|x|}$ on $\mathbb{R}$.

Proposition 2.12 (Stability properties of convex functions). *(i) If $(f_i)_{i \in I}$ is a family of convex functions then $\sup_i f_i : x \mapsto \sup_{i \in I} f_i(x)$ is also convex.*

2 Convex functions

(ii) If $A : \mathcal{X} \rightarrow \mathcal{Y}$ is affine and f is convex on $\mathcal{Y}$ then $f \circ A$ is convex on $\mathcal{X}$.

(iii) Any conic combination of convex functions is convex².

(iv) A pointwise limit $f = \lim_{n \rightarrow +\infty} f_n$ of convex functions is also convex.

Proof. For the first item, one proof consists in writing the epigraph of $f = \sup_i f_i$ as

$$\text{epi } f = \bigcap_{i \in I} \text{epi } f_i$$

and use the fact that a general intersection of convex sets is convex. A direct proof is also possible: taking $x, y \in \mathcal{X}$ and $\theta \in [0, 1]$, we have for every $j \in I$

$$f_j((1 - \theta)x + \theta y) \leq (1 - \theta)f_j(x) + \theta f_j(y) \leq (1 - \theta)f(x) + \theta f(y),$$

and taking the supremum in j yields the result.

The second item is straightforward since $A((1 - \theta)x + \theta y) = (1 - \theta)Ax + \theta Ay$. The third item is also straightforward, looking for instance at the stability by positive multiples and by sum.

For the final item, simply pass to the limit $n \rightarrow +\infty$ in the inequality

$$f_n((1 - \theta)x + \theta y) \leq (1 - \theta)f_n(x) + \theta f_n(y),$$

for fixed $x, y \in \mathcal{X}$ and $\theta \in [0, 1]$. □

Example 2.13. 1. The indicator functions χ_C where C is convex, are convex.

2. Linear forms are convex.

3. Sublinear functions (in particular norms, gauges and support functions) are convex.

4. If f is convex and C is convex then so is the restriction $f|_C$ of f to C defined as $f + \chi_C$.

5. If A is a positive semidefinite square matrix then $x \mapsto \langle Ax, x \rangle$ is convex.

6. The composition of convex functions is not necessarily convex, e.g. $f(x) = x^2$ and $g(x) = -x$ are convex but $g \circ f$ is not convex.

7. However if $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex and $g : \mathbb{R} \cup \{+\infty\} \rightarrow \mathbb{R} \cup \{+\infty\}$ is a nondecreasing function such that $g(+\infty) = +\infty$ and g is convex on $\mathbb{R}$, then $g \circ f$ is convex.

²It means that the set of convex functions on $\mathcal{X}$ is a convex cone.

2.2 Regularity

Continuity properties

We recall that a function f valued in $\bar{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ is continuous at a point x_0 if

$$f(x_0) = \lim_{x \rightarrow x_0} f(x),$$

no matter if $f(x_0) \in \mathbb{R}$, $f(x_0) = +\infty$ or $f(x_0) = -\infty$.

Theorem 2.14. *Let $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex function and $x_0 \in \mathcal{X}$. The following assertions are equivalent:*

- (i) f is bounded from above on a neighborhood of x_0 ,
- (ii) $x_0 \in \text{dom } f$ and f is Lipschitz on a neighborhood of x_0 ,
- (iii) $x_0 \in \text{dom } f$ and f is continuous at x_0 .

In such a case $x_0 \in \text{int}(\text{dom } f)$.

Proof. It suffices to show that (i) implies (ii), the rest being trivial. Assume that $f \leq M$ with $M > 0$ on a ball $B[x_0, r_0]$. In particular f is finite on this ball and $x_0 \in \text{int dom } f$. Let $x \in B[x_0, r_0]$ and set

$$y = x_0 + r_0 \frac{x - x_0}{\|x - x_0\|}.$$

Notice that $x \in [x_0, y]$, more precisely

$$x = (1 - \theta)x_0 + \theta y \quad \text{where} \quad \theta = \frac{\|x - x_0\|}{r_0} \in [0, 1].$$

By convexity of f ,

$$f(x) - f(x_0) \leq (1 - \theta)f(x_0) + \theta f(y) - f(x_0) = \theta(f(y) - f(x_0)) \leq \theta(M - f(x_0)). \quad (2.2.1)$$

We do the same for the symmetric points of x and y w.r.t. x_0 , namely $x' = x_0 - (x - x_0)$ and $y' = x_0 - (y - x_0)$, noticing that $x' = (1 - \theta)x_0 + \theta y'$, so as to obtain

$$f(x') - f(x_0) \leq \theta(M - f(x_0)). \quad (2.2.2)$$

Since $x_0 = \frac{1}{2}(x + x')$ by convexity of f we know that

$$f(x_0) \leq \frac{1}{2}(f(x) + f(x')) \iff -(f(x) - f(x_0)) \leq f(x') - f(x_0).$$

Putting this together with (2.2.2) and (2.2.1) yields

$$|f(x) - f(x_0)| \leq \theta(M - f(x_0)) = \frac{\|x - x_0\|}{r_0}(M - f(x_0)).$$

2 Convex functions

This show that f is continuous at x_0 . Let us show that it is Lipschitz on every ball $B(x_0, r)$ for $r < r_0$. If $x \in B(x_0, r)$ then $B[x, r_0 - r] \subseteq B[x_0, r_0]$, thus f is bounded from above by M on $B[x, r_0 - r]$ and we can apply the above computation: if $y \in B[x, r_0 - r]$ then

$$|f(y) - f(x)| \leq \|y - x\| \frac{M - f(x)}{r_0 - r}.$$

But we know that $f(x) \geq -(M - f(x_0))$, thus

$$|f(y) - f(x)| \leq \|y - x\| \frac{2(M - f(x_0))}{r_0 - r},$$

which holds for every $x, y \in B(x_0, r)$ satisfying $\|x - y\| \leq r_0 - r$. To obtain it without the latter condition, we use the convexity of the ball $B(x_0, r)$: if $x, y \in B(x_0, r)$ we may divide the segment $[x, y]$ in consecutive segments $[x_i, x_{i+1}]$ (of the form $x_i = x + \theta_i(y - x)$ with (θ_i) positive and increasing) of length smaller than $r_0 - r$. Thus:

$$|f(y) - f(x)| \leq \sum_{i=1}^{n-1} |f(x_{i+1}) - f(x_i)| \leq L \sum_{i=1}^{n-1} \|x_{i+1} - x_i\| = L\|y - x\|$$

where $L = 2 \frac{M - f(x_0)}{r_0 - r}$. This concludes the proof. $\square$

We say that a function $f : A \rightarrow \mathbb{R}$ is locally Lipschitz on A for every $x_0 \in A$, f is Lipschitz on a neighborhood of x_0 in A .

Corollary 2.15. *Let $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$. The following assertions are equivalent:*

- (i) *there exists x_0 such that f is bounded from above on a neighborhood of x_0 ,*
- (ii) *$\text{int}(\text{dom } f) \neq \emptyset$ and f is locally Lipschitz on $\text{int}(\text{dom } f)$.*

Proof. Let show the implication (i) $\implies$ (ii), the converse being trivial. We assume the first item holds. By the previous proposition, $x_0 \in \text{int dom } f$ (thus the latter is not empty), and f is Lipschitz on a ball $B(x_0, r)$. In particular $|f| \leq M$ for some $M > 0$ on $B(x_0, r)$.

Take $y_0 \in \text{int dom } f$ and assume that $y_0 \neq x_0$. There is a small $\varepsilon > 0$ such that $z_0 = y_0 + \varepsilon(y_0 - x_0) \in \text{dom } f$, so that $y_0 \in (x_0, z_0)$, thus there exists $\theta \in (0, 1)$ such that $y_0 = (1 - \theta)x_0 + \theta z_0$. We know that

$$B(y_0, (1 - \theta)r) = (1 - \theta)B(x_0, r) + \theta z_0 \subseteq \text{dom } f,$$

thus for every $y \in B(y_0, (1 - \theta)r)$, y can be written as $y = (1 - \theta)x + \theta z_0$ for some $x \in B(x_0, r)$, and by convexity of f

$$f(y) \leq (1 - \theta)f(x) + \theta f(z_0) \leq (1 - \theta)M + \theta f(z_0).$$

We have proved that f is bounded from above on a ball around y_0 , and by the previous proposition it is Lipschitz on a neighborhood of y_0 . $\square$

2 Convex functions

Proposition 2.16. *If $\mathcal{X}$ is a finite-dimensional vector space and $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ is a convex function, then f is locally Lipschitz on $\text{ri}(\text{dom } f)$.*

Proof. First of all, let us replace $\mathcal{X}$ by $\text{aff}(\text{dom } f)$, so that $\text{ri}(\text{dom } f) = \text{int}(\text{dom } f)$. Let us consider points $(x_0, \dots, x_n)$ in $\text{dom } f$ which form an affine basis of $\mathcal{X}$. Since $\text{dom } f$ is convex, we know that $\text{conv}\{x_0, \dots, x_n\} \subseteq \text{dom } f$. By [Proposition 1.15](#) we know that the interior of $\text{conv}\{x_0, \dots, x_n\}$ is not empty, and moreover f is bounded from above on it. Indeed if $x = \sum_{i=0}^n \theta_i x_i$ then $f(x) \leq \sum_{i=0}^n \theta_i f(x_i) \leq n \max_{0 \leq i \leq n} \{f(x_i)\}$. By the previous corollary, f is locally Lipschitz on $\text{ri}(\text{dom } f)$. $\square$

Differentiability properties

In this section we shall define and study increasingly stronger notions of differentiability: directional, Gâteaux- and Fréchet-differentiability.

Definition 2.17. Let $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$, $x_0 \in \text{dom } f$ and $v \in \mathcal{X}$. When the following limits exist (in $\mathbb{R} \cup \{-\infty, +\infty\}$),

$$f'_-(x, v) = \lim_{t \rightarrow 0^-} \frac{f(x + tv) - f(x)}{t}, \quad f'_+(x, v) = \lim_{t \rightarrow 0^+} \frac{f(x + tv) - f(x)}{t},$$

we say that f is respectively *left- and right-differentiable at x_0 in the direction v* and we call them respectively the *left and right directional derivatives of f at x_0 in the direction v* .

Remark 2.18. Notice that for every x and v , $f'_+(x, v)$ exists if and only if $f'_-(x, -v)$ does, in which case $f'_+(x, v) = -f'_-(x, -v)$. This is why we shall concentrate on the right derivative f'_+ in the following.

Proposition 2.19 (Directional derivatives of convex functions). *Any convex function $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ is left- and right-differentiable at every point of its domain (with possibly infinite derivative), and moreover,*

(i) *for every $x \in \text{dom } f$, $f'_\pm(x, \cdot)$ is positively homogeneous, and it is sublinear (or equivalently, convex) if it is real-valued,*

(ii) *for every $x \in \text{dom } f$, $f'_\pm(x, \cdot)$ and $v \in \mathcal{X}$*

$$f'_+(x, v) = \inf_{t > 0} \frac{f(x + tv) - f(x)}{t} \quad \text{and} \quad f'_-(x, v) = \sup_{t < 0} \frac{f(x + tv) - f(x)}{t},$$

(iii) *$f'_-(x, v) \leq f'_+(x, v)$ for every $x \in \text{dom } f$ and $v \in \mathcal{X}$,*

(iv) *$f'_+(x, y - x) \leq f(y) - f(x)$ for every $x \in \text{dom } f$ and $y \in \mathcal{X}$,*

(v) *$f'_+(x, y - x) \leq f'_+(y, y - x)$ for every $x, y \in \text{dom } f$.*

2 Convex functions

Remark 2.20 (Convex functions are above tangents). Notice that as a consequence, we recover the fact that f is “above its tangents”:

$$\forall x \in \text{dom } f, \forall v \in \mathcal{X}, \forall t \in \mathbb{R}, \quad f(x + tv) \geq f(x) + tf'_+(x, v).$$

Proof. Let $x \in \text{dom } f$ and $v \in \mathcal{X}$. The map $\phi(t) = f(x + tv) - f(x)$ is convex on $\mathbb{R}_+$ and $\phi(0) = 0$ thus if $0 \leq s < t$, $s/t \in [0, 1)$ and

$$\phi(s) = \phi((s/t) \times t + (1 - (s/t)) \times 0) \leq (s/t)\phi(t) + (1 - (s/t))\phi(0) = s \frac{\phi(t)}{t},$$

so that $t \mapsto \phi(t)/t$ is nondecreasing on $(0, +\infty)$. We deduce the existence of the limit as $t \rightarrow 0^+$, i.e. of $f'_+(x, v)$, and the second item holds true. For the third item, we can assume that $-\infty < f'_-(x, v)$, i.e. $f'_+(x, -v) < +\infty$, and that $f'_+(x, v) < +\infty$, otherwise there is nothing to prove. By convexity of f , we check that for every $t > 0$

$$0 = \frac{f(x + 0) - f(x)}{t} \leq \frac{1}{2} \left(\frac{f(x + tv) - f(x)}{t} + \frac{f(x - tv) - f(x)}{t} \right),$$

and since both terms on the right have a limit in $(-\infty, +\infty)$ we may pass to the limit and get $0 \leq f'_+(x, v) + f'_+(x, -v)$. Thus both $f'_+(x, v)$ and $f'_+(x, -v)$ are finite and we get $f'_+(x, v) \geq -f'_+(x, -v) = f'_-(x, v)$, hence (iii) holds true.

Besides, taking $v = (y - x)$, we get

$$f'_+(x, y - x) = \lim_{t \rightarrow 0^+} \frac{\phi(t)}{t} \leq \frac{\phi(1)}{1} = f(y) - f(x),$$

thus (iv) holds true. Then interchanging the roles of x, y yields

$$f(y) - f(x) \leq -f'_+(y, x - y) = f'_-(y, y - x)$$

which together with (iii) and (iv) yields (v). □

Remark 2.21. • Notice that since $f'_+(x, v)$ is written as an infimum, it is $+\infty$ if and only if $f(x + tv) = +\infty$ for every $t > 0$, i.e. if the open half line of direction v and based at x does not intersect $\text{dom } f$. This may happen: consider for example $f = \chi_{[0,1]}$, $x = 1$ and $v = 1$.

- It is possible that $f'_+(x, v) = -\infty$. Consider for example $f(t) = t \log t$ for $t > 0$, $f(0) = 0$ and $f(t) = -\infty$ for $t < 0$, with $x = 0$ and $v = 1$.
- It is possible that $f'_-(x, v) < f'_+(x, v)$. Consider for example $f(t) = |t|$ at $t = 0$ and $v = 1$: $f'_-(x, v) = -f'_+(x, -1) = -1 < 1 = f'_+(x, v)$.

We now introduce a stronger notion of differentiability, that requires $f'_+(x_0, \cdot)$ to be linear.

Definition 2.22 (Gâteaux differentiability). Let $f : \mathcal{X} \rightarrow \mathbb{R} \cup -\infty, +\infty$ and $x_0 \in \text{dom } f$. We say that f is Gâteaux-differentiable at x_0 if

2 Convex functions

- for every $v \in \mathcal{X}$, the following limit exists in $\mathbb{R} \cup \{-\infty, +\infty\}$,

$$f'(x_0, v) = \lim_{t \not\rightarrow 0} \frac{f(x_0 + tv) - f(x_0)}{t}$$

- there exists a continuous linear form $p^* \in \mathcal{X}^*$ such that

$$\forall v \in \mathcal{X}, \quad f'(x_0, v) = \langle p^*, v \rangle.$$

This linear form p^* is then unique, called the Gâteaux differential of f at x_0 , and denoted by $Df(x_0)$ (or sometimes $f'(x_0)$ or $\nabla f(x_0)$).

Remark 2.23. The notion of Gâteaux-differentiability is stronger than directional differentiability, but is still generally quite weak. For example the function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by

$$f(x, y) = \begin{cases} 1 & \text{if } x \neq 0 \text{ and } y = x^2 \\ 0 & \text{otherwise,} \end{cases}$$

is Gâteaux-differentiable but not even continuous at 0!

Proposition 2.24 (Characterization of Gâteaux differentiability). *A convex function $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ is Gâteaux-differentiable at $x \in \text{dom } f$ if and only if $f'_+(x, \cdot)$ is a continuous and odd real-valued function, or equivalently, if and only if it is a continuous linear form, in which case $Df(x) = f'_+(x, \cdot)$.*

Proof. As a convex function, f is directionally differentiable at every point $x \in \text{dom } f$. Notice that

$$\begin{aligned} f'(x, v) &= \lim_{t \not\rightarrow 0} \frac{f(x_0 + tv) - f(x_0)}{t} \text{ exists} \\ \iff f'_+(x, v) &= f'_-(x, v) \\ \iff f'_+(x, -v) &= -f'_+(x, v), \end{aligned}$$

thus $f'(x, v)$ exists for every v if and only if $f'_+(x, \cdot)$ is odd. In such a case, $f'(x, \cdot)$ is sub-linear and homogeneous, thus linear: indeed for every v, w , $f'_+(x, v + w) = -f'_+(x, -v - w) \geq -(f'_+(x, -v) + f'_+(x, -w)) = f'_+(x, v) + f'_+(x, w)$ and f'_+ is not only subadditive but also superadditive. We deduce that f is Gâteaux-differentiable at x if and only if $f'_+(x, \cdot)$ is odd (or equivalently, linear), continuous and real-valued. Moreover $Df(x) = f'_+(x, \cdot)$ in that case. $\square$

Remark 2.25. Gâteaux differentiability implies directional differentiability, but the converse is not true. Take for example the function $f(x) = |x|$ at $x = 0$: it is directionally differentiable (because it is convex) but not Gâteaux-differentiable since $f'_+(x, 1) = 1 \neq -1 = f'_+(x, -1)$.

Proposition 2.26. *Let $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex function.*

2 Convex functions

1. If $x \in \text{int}(\text{dom } f)$ then $f'_+(x, \cdot)$ is a real-valued sublinear function.
2. If moreover f is continuous at x then $f'_+(x, \cdot)$ is a continuous sublinear function.
3. If moreover $f'_+(x, \cdot)$ is odd, then f is Gâteaux-differentiable at x .

Proof. Assume that $x \in \text{int}(\text{dom } f)$. Then for any fixed $v \in \mathcal{X}$ and for small $t > 0$, $x + tv \in \text{dom } f$, therefore by the characterization of $f'_+(x, v)$ as an infimum, it cannot take the value $+\infty$, which implies also that $f'_+(x, v) \geq f'_-(x, v) = -f'_+(x, -v) > -\infty$ for every v . Thus $f'_+(x, \cdot)$ is a real-valued sublinear function by [Proposition 2.19](#).

Assume moreover that f is continuous at x . Thus $|f(y) - f(x)| \leq M$ on $B(x, r)$ for some $M > 0$ and $r > 0$. Therefore if $v \in B(0, r)$, we have

$$f'_+(x, v) = \inf_{t>0} \frac{f(x + tv) - f(x)}{t} \leq f(x + v) - f(x) \leq M$$

and

$$f'_+(x, v) \geq -f'_+(x, -v) \geq -M.$$

Thus $f'_+(x, \cdot)$ is continuous as a sublinear function which is bounded on a neighborhood of 0.

If moreover $f'_+(x, \cdot)$ is odd, then by [Proposition 2.26](#), f is Gâteaux-differentiable at x . $\square$

Finally, we recall the stronger (and most usual) notion of differentiability, Fréchet differentiability.

Definition 2.27. A function $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ is Fréchet-differentiable at a point $x \in \text{dom } f$ if there exists a continuous linear form $p^* \in \mathcal{X}^*$ such that

$$f(x + h) = f(x) + \langle p^*, h \rangle + o_{h \rightarrow 0}(\|h\|).$$

Such p^* is then unique, called the *Fréchet differential of f at x* and denoted by $Df(x)$ (or sometimes $f'(x)$ or $\nabla f(x)$).

Remark 2.28. Of course

$$\text{Fréchet-differentiable} \implies \text{Gâteaux-differentiable} \implies \text{directional differentiable},$$

and if f is Fréchet-differentiable then the Fréchet and Gateau-differentials coincide. The converse implications are not true. We have already seen the example of $f(x) = |x|$ which is not Gâteaux-differentiable but convex thus directionally differentiable. The function $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by

$$g(x, y) = \begin{cases} \frac{|x|}{2} & \text{if } y = x^2 \\ 0 & \text{otherwise,} \end{cases}$$

2 Convex functions

is Gâteaux-differentiable with $f'_+(0) = 0$ and continuous at 0 but not Fréchet-differentiable, since $g(x, x^2) \sim |x|/2$ while $\|(x, x^2)\| \sim |x|$ thus $g(x, y) - (g(0) + \langle f'(0), (x, y) \rangle) \neq o(\|(x, y)\|)$ as $(x, y) \rightarrow 0$.

For convex functions neither, Gâteaux-differentiability does not imply Fréchet differentiability in general, but it is true in finite-dimensional vector spaces.

Proposition 2.29. *We assume that $\mathcal{X}$ has finite dimension. If $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex and Gâteaux-differentiable at $x \in \text{dom } f$ then it is Fréchet-differentiable at x .*

The proof is the object of an exercise.

Proof. First of all, the Gâteaux differentiability at x implies that $f'(x, \cdot)$ is finite, thus

$$x \in \text{core}(\text{dom } f) = \{y \in \text{dom } f : \forall v, \exists t > 0, [x - tv, x + tv] \subseteq \text{dom } f\},$$

which in finite dimension means that $x \in \text{int}(\text{dom } f)$.

Since f is convex in finite dimension, it is locally Lipschitz, thus there exists a closed ball $B[x_0, r]$ where f is L -Lipschitz. Notice that f is Fréchet-differentiable at x_0 if and only if the family of maps $f_t : B[0, 1] \rightarrow \mathbb{R}$, $t \leq r$ defined by

$$f_t(v) = \frac{f(x + tv) - f(x)}{t}$$

converges uniformly to $f'(x)$ as $t \rightarrow 0^+$. Let us show by a sequential criterion that $\|f_t - f'(x)\|_\infty \rightarrow 0$ as $t \rightarrow 0$. Consider a sequence (t_k) going to 0^+ . Since $B[0, 1]$ is compact, there exists $v_k \in B[0, 1]$ such that

$$\|f_{t_k} - f'(x)\|_\infty = \left| \frac{f(x + t_k v_k) - f(x)}{t_k} - \langle f'(x), v_k \rangle \right|,$$

and by compactness, up to subsequence $v_{t_k} \rightarrow v$ for some $v \in B[0, 1]$. We estimate:

$$\begin{aligned} & \|f_{t_k} - f'(x)\|_\infty \\ &= \left| \frac{f(x + t_k v_k) - f(x)}{t_k} - \langle f'(x), v_k \rangle \right| \\ &\leq \left| \frac{f(x + t_k v_k) - f(x)}{t_k} - \frac{f(x + t_k v) - f(x)}{t_k} \right| + \left| \langle f'(x), v \rangle - \frac{f(x + t_k v) - f(x)}{t_k} \right| \\ &\leq L\|v_k - v\| + \left| \langle f'(x), v \rangle - \frac{f(x + t_k v) - f(x)}{t_k} \right| \\ &\xrightarrow{k \rightarrow +\infty} 0. \end{aligned}$$

Since the sequence (t_k) was arbitrary, it shows that f_t converges to $f'(x)$ uniformly, thus that f is Fréchet-differentiable at x . $\square$

Proposition 2.30 (Characterizations of smooth convex functions). *Let $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper function with open domain. Assume that f is Gâteaux-differentiable on its domain. The following assertions are equivalent:*

2 Convex functions

(i) f is convex,

(ii) $\forall x, y \in \text{dom } f, \quad f(y) \geq f(x) + \langle f'(x), y - x \rangle,$

(iii) $\forall x, y \in \text{dom } f, \quad \langle f'(y) - f'(x), y - x \rangle \geq 0.$

If moreover f is twice differentiable on its domain, it is equivalent to

(iv) for every $x \in \text{dom } f$, $D^2f(x)$ is positive semi-definite, i.e. $\langle D^2f(x)v, v \rangle \geq 0$ for every v .

The proof is the object of an exercise.

Proof. We already know from **Proposition 2.19** that (i) implies (ii). If (ii) holds, for every $x, y \in \text{dom } f$ we have

$$f(y) \geq f(x) + \langle f'(x), y - x \rangle \quad \text{and} \quad f(x) \geq f(y) + \langle f'(y), x - y \rangle$$

and by summing the two we get (iii). Now assume that (iii) holds. For fixed $x, y \in \text{dom } f$, the function $g : t \mapsto f(x + t(y - x))$ is (Gâteaux-)differentiable at every $t \in [0, 1]$ and $g'(t) = \langle f'(x + t(y - x)), y - x \rangle$. By (iii) applied to $x_t = x + t(y - x)$ and $x_s = x + s(y - x)$ we see that $g'(t) - g'(s) \geq 0$ for $0 \leq s \leq t \leq 1$ thus g' is nondecreasing and thus convex on $[0, 1]$. We get $g(t) \leq tg(1) + (1 - t)g(0)$, i.e. $f((1 - t)x + ty) \leq (1 - t)f(x) + tf(y)$. $\square$

We may deduce similar characterizations for λ -strongly convex functions thanks to the following lemma.

Lemma 2.31. *A function $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ is λ -strongly convex for $\lambda > 0$ if and only if $f - \lambda \frac{\|\cdot\|^2}{2}$ is convex.*

Proof. Left as exercise. $\square$

Corollary 2.32. *Let $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper function with open domain. Assume that f is Gâteaux-differentiable on its domain, and let $\lambda > 0$. The following assertions are equivalent:*

(i) f is λ -strongly convex,

(ii) $\forall x, y \in \text{dom } f, \quad f(y) \geq f(x) + \langle f'(x), y - x \rangle + \lambda \frac{\|y - x\|^2}{2},$

(iii) $\forall x, y \in \text{dom } f, \quad \langle f'(y) - f'(x), y - x \rangle \geq \lambda \|y - x\|^2.$

If moreover f is twice differentiable on $\text{dom } f$, it is equivalent to

(iv) for every $x \in \text{dom } f$, $D^2f(x) \succeq \lambda I$, i.e. $\langle D^2f(x)v, v \rangle \geq \lambda \|v\|^2$ for every v .

2.3 Subdifferential

Definition 2.33 (Subderivative and subdifferential). Let $f : \mathcal{X} \rightarrow (-\infty, +\infty]$ be a function and $x \in \text{dom } f$. We say that $x^* \in \mathcal{X}^*$ is a *subderivative* (or *subgradient*) of f at x if,

$$\forall y \in \mathcal{X}, \quad f(y) \geq f(x) + \langle x^*, y - x \rangle,$$

and we denote by $\partial f(x)$ the set of all subderivatives of f at x . By definition we set $\partial f(x) = \emptyset$ if $x \notin \text{dom } f$. If $\partial f(x) \neq \emptyset$ we say that f is *subdifferentiable* at x .

Notice that a function f which is subdifferentiable at some point $x \in \text{dom } f$ never takes the value $-\infty$.

Remark 2.34 (Geometric interpretation). A linear form x_0^* is a subderivative of f at x_0 if and only if in the space $\mathcal{X} \times \mathbb{R}$ the affine closed hyperplane $\{(x, t) : f(x_0) + \langle x^*, x - x_0 \rangle = t\}$ is a support hyperplane of $\text{epi } f$ at $(x_0, f(x_0))$.

Proposition 2.35 (Elementary properties). *Let $f : \mathcal{X} \rightarrow (-\infty, +\infty]$. The following properties hold:*

(i) *for every x , $\partial f(x)$ is convex and closed*³,

(ii) *∂f is monotone, in the sense that*

$$\forall x^* \in \partial f(x), \forall y^* \in \partial f(y), \quad \langle y^* - x^*, y - x \rangle \geq 0,$$

(iii) *if $x_n \rightarrow x$, $\partial f(x_n) \ni x_n^* \rightarrow x^*$ and f is continuous at x then $x^* \in \partial f(x)$* ⁴.

Moreover, assuming that f is convex,

(iv) *if f is continuous at x then $\partial f(x)$ is bounded,*

(v) $\partial f(x) = \{x^* : f'_+(x, \cdot) \geq \langle x^*, \cdot \rangle\}$.

Proof. Left as exercise. □

Proposition 2.36 (Subdifferentiability implies convexity). *Let $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$. If $\text{dom } f$ is convex and f is subdifferentiable at every point of its domain, then f coincides on $\text{dom } f$ with the supremum of its continuous affine minorants, and is as such convex (and lower semi-continuous) function.*

Proof. We set

$$\bar{f} = \sup\{g : g \leq f \text{ and } g \text{ is continuous and affine}\}.$$

By definition, $\bar{f} \leq f$. Let us take $x \in \text{dom } f$ and pick some $x^* \in \partial f(x)$. By definition, $f \geq g$ where $g(y) = f(x) + \langle x^*, y - x \rangle$, and thus $\bar{f}(x) \geq g(x) = f(x)$, so that $\bar{f} = f$ on $\text{dom } f$. Since $\bar{f}$ is convex (and lower semi-continuous), then so is $f|_{\text{dom } f} = f$. □

³Even weakly-* closed, as we shall see.

⁴The result holds true also when f is lower semicontinuous and x_n^* converges weakly-* to x^* .

2 Convex functions

To study the subdifferentiability properties of convex functions, we shall rely on a functional separation theorem in the spirit of (and resulting from) the Hahn-Banach theorems.

Theorem 2.37 (Functional separation theorem). *Let $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex function and $g : \mathcal{X} \rightarrow \mathbb{R} \cup \{-\infty\}$ be a concave function such that $f \leq g$. We assume that f and g are both finite at x_0 and that f is continuous at x_0 . Then there exists a continuous and affine function $a : E \rightarrow \mathbb{R} \cup \{+\infty\}$ such that $g \leq a \leq f$ on $\mathcal{Y}$.*

Proof. Since f is continuous at x_0 , $\text{int epi } f \neq \emptyset$ and $x_0 \in \text{int dom } f$. Moreover, $\text{int epi } f$ is included in the strict epigraph of f , that is $S = \{(x, t) \in \mathcal{X} \times \mathbb{R}; f(x) < t\}$, and since $f \geq g$, the latter is disjoint from the hypograph of g , defined by $H = \{(x, t) \in \text{dom } g \times \mathbb{R}; g(x) \geq t\}$. Therefore, $\text{int epi } f$ and H are disjoint non-empty convex subsets of $\mathcal{X} \times \mathbb{R}$, the first one being open, and we may apply the Hahn-Banach separation theorem to find a non-trivial linear form on $\mathcal{X} \times \mathbb{R}$ which separates them. Since $\text{epi } f = \overline{\text{int epi } f}$ (because the interior is not empty), we can say that there exists $(x^*, \lambda) \in \mathcal{X}^* \times \mathbb{R} \setminus \{(0, 0)\}$ such that

$$\inf_{(x,t) \in \text{epi } f} \langle x^*, x \rangle + \lambda t \geq \sup_{(y,s) \in H} \langle x^*, y \rangle + \lambda s, \quad (2.3.1)$$

and in particular

$$\forall (x, t) \in \text{epi } f, \quad \langle x^*, x \rangle + \lambda t \geq \langle x^*, x_0 \rangle + \lambda g(x_0).$$

Letting $t \rightarrow +\infty$ for $x = x_0$ yields $\lambda \geq 0$. If $\lambda = 0$, we would have $\langle x^*, x - x_0 \rangle \geq 0$ for every $x \in \text{dom } f$, and since $\text{dom } f$ contain some ball $B(x_0, r)$ centered at x_0 it would imply $x^* = 0$, a contradiction with the fact that $(x^*, \lambda) \neq (0, 0)$. Consequently $\lambda > 0$, and renaming x^* to replace x^* by $-\lambda x^*$ in (2.3.1) yields

$$\forall x \in \text{dom } f, \forall y \in \text{dom}(-g), \quad \langle -x^*, x \rangle + f(x) \geq b \geq -\langle x^*, y \rangle + g(y)$$

where $b = \inf_{x \in \text{dom } f} \langle -x^*, x \rangle + f(x) \in \mathbb{R}$. The above inequality is also trivially true when $x \notin \text{dom } f$ or $y \notin \text{dom}(-g)$. Taking $x = y$ in this inequality yields:

$$f(x) \geq b + \langle x^*, x \rangle \geq g(x).$$

□

We obtain the following theorem as a corollary.

Theorem 2.38 (Non-emptiness of the subdifferential). *Let $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex function. If f is continuous at a point of its domain then it is continuous and sub-differentiable at every point $x_0 \in \text{int dom } f$. Moreover, for every such x_0 and every $v \in \mathcal{X}$, there exists $p^* \in \partial f(x_0)$ such that $f'_+(x_0, v) = \langle p^*, v \rangle$.*

Remark 2.39. As a consequence of this theorem, if f is continuous at x_0 , then for every v ,

$$f'_+(x_0, v) = \max_{p^* \in \partial f(x_0)} \langle p^*, v \rangle,$$

thus doing the identification $\mathcal{X} \subseteq \mathcal{X}^{**}$ we may write in particular:

$$f'_+(x_0, v) = \sigma_{\partial f(x_0)}(v),$$

2 Convex functions

Proof. We have already seen that since f is convex and continuous at some point of its domain, then $\text{int dom } f$ is not empty and f is continuous at every point $x_0 \in \text{int dom } f$. Consider such a point. By [Proposition 2.26](#), $f'_+(x_0, \cdot)$ is a real-valued sublinear function. Let $v \in \mathcal{X}$ and define the concave function $g : \mathcal{X} \rightarrow \mathbb{R} \cup \{-\infty\}$ by $g(x_0 + tv) = f(x_0) + tf'_+(x_0, v)$ on the affine line $L = x_0 + \mathbb{R}v$, and $g(x) = -\infty$ on $\mathcal{X} \setminus L$. By [Proposition 2.19](#) we know that $f \geq g$: indeed for $t \geq 0$ we have $f(x_0 + tv) \geq f(x_0) + tf'_+(x_0, v)$ while for $t < 0$, $f(x_0 + tv) \geq f(x_0) - tf'_+(x_0, -v) \geq f(x_0) + tf'_+(x_0, v)$. We are in position to apply [Theorem 2.37](#), so as to obtain an affine form $a : x \mapsto \langle x^*, x \rangle + b$ such that $g \leq a \leq f$. In particular at x_0 we get equality $g(x_0) = f(x_0) = \langle x^*, x_0 \rangle + b$, thus and $f \geq a$ rewrites as

$$\forall x \in \mathcal{X}, \quad f(x) \geq a(x_0) + (a(x) - a(x_0)) = f(x_0) + \langle x^*, x - x_0 \rangle,$$

which shows that $x^* \in \partial f(x_0)$. Besides, taking $x = x_0 + tv$ gives

$$\begin{aligned} \forall t \in \mathbb{R}, \quad g(x_0 + tv) - g(x_0) &\leq a(x) - a(x_0), \\ \iff \quad tf'_+(x_0, v) &\leq t\langle x^*, v \rangle. \end{aligned}$$

Taking $t = \pm 1$ gives $\langle x^*, v \rangle = f'_+(x, v)$. □

Corollary 2.40 (Characterization of Gâteaux-differentiability). *Let $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex function and $x_0 \in \text{dom } f$.*

1. *If f is Gâteaux-differentiable at x_0 then $\partial f(x_0)$ is the singleton $\{f'(x_0)\}$.*
2. *The converse is true when f is continuous at x_0 : f is Gâteaux-differentiable at x_0 if and only if $\partial f(x_0)$ is a singleton.*

Proof. We have seen that when f is Gâteaux-differentiable at a point x_0 , $f'_+(x_0, \cdot) = f'(x_0)$ and by [Proposition 2.19](#) it holds

$$f(x) \geq f(x_0) + f'(x_0, x - x_0) = f(x_0) + \langle f'(x_0), x - x_0 \rangle,$$

which means that $f'(x_0) \in \partial f(x_0)$. Conversely, take $p^* \in \partial f(x_0)$. For every $v \in \mathcal{X}$ and every $t \in \mathbb{R}$, we have

$$f(x_0 + tv) - f(x_0) \geq t\langle p^*, v \rangle,$$

Taking $t > 0$ then $t < 0$, dividing by t and taking the limit $t \rightarrow 0^\pm$ yields $f'(x_0) = p^*$, thus $\partial f(x_0) = \{f'(x_0)\}$.

Assume now that f is continuous at x_0 , and suppose that $\partial f(x_0) = \{p^*\}$. By [Theorem 2.38](#) we know that for every $v \in \mathcal{X}$, there exists a point $x^* \in \partial f(x_0)$ such that $f'_+(x_0, v) = \langle x^*, v \rangle$, but necessarily $x^* = p^*$ and thus $f'_+(x_0, \cdot) = p^*$: f is Gâteaux-differentiable at x_0 . □

Definition 2.41 (Adjoint operator). The adjoint of a continuous operator $A : \mathcal{X} \rightarrow \mathcal{Y}$ is the operator $A^* : \mathcal{Y}^* \rightarrow \mathcal{X}^*$ defined by

$$A^*(y^*) = (\mathcal{X} \ni x \mapsto \langle y^*, Ax \rangle) \in \mathcal{X}^*,$$

and is characterized by the property:

$$\forall y \in \mathcal{Y}^*, \forall x \in \mathcal{X}, \quad \langle A^*y^*, x \rangle = \langle y^*, Ax \rangle.$$

2 Convex functions

Proposition 2.42 (Inclusion properties of subdifferentials). *The following properties hold:*

- (i) $\partial(f + g)(x) \supseteq \partial f(x) + \partial g(x)$ for every functions $f, g : \mathcal{X} \rightarrow [-\infty, +\infty]$ and $x \in \mathcal{X}$,
- (ii) $\partial(f \circ A)(x) \supseteq A^* \partial f(Ax)$ for every function $f : \mathcal{Y} \rightarrow [-\infty, +\infty]$ and continuous linear map $A : \mathcal{X} \rightarrow \mathcal{Y}$,
- (iii) $\partial(\sup_i f_i)(x) \supseteq \overline{\text{conv}} \bigcup_{\text{argmax}_i f_i(x)} \partial f_i(x)$ for every family $(f_i)_{i \in I}$ of functions from $\mathcal{X} \rightarrow [-\infty, +\infty]$ and every $x \in \mathcal{X}$.

Remark 2.43. In the previous proposition, recall that $\emptyset + \Omega = \emptyset$ by definition.

The proof can be given as an exercise.

Proof. Let $p \in \partial f(x)$ and $q \in \partial g(x)$. Summing the two inequalities

$$f(y) \geq f(x) + \langle p, y - x \rangle \text{ and } g(y) \geq g(x) + \langle q, y - x \rangle$$

yields $(f + g)(y) \geq (f + g)(x) + \langle p + q, y - x \rangle$ for every y thus $p + q \in \partial(f + g)(x)$.

Let $p \in \partial f(Ax)$. For every y we have

$$f(y) \geq f(Ax) + \langle p, y - Ax \rangle.$$

and taking $y = A\tilde{x}$ yields

$$f(A\tilde{x}) \geq f(Ax) + \langle p, A\tilde{x} - Ax \rangle = f(Ax) + \langle A^*p, \tilde{x} - x \rangle,$$

which exactly means that $A^*p \in \partial f(Ax)$. Thus $A^* \partial f(Ax) \subseteq \partial(f \circ g)(x)$.

Finally, we set $f = \sup_i f_i$. Assume that j is such that $f_j(x) = \sup_i f_i(x)$, and $p \in \partial f_j(x)$. For every y we have

$$f(y) \geq f_j(y) \geq f_j(x) + \langle p, y - x \rangle = f(x) + \langle p, y - x \rangle$$

which implies that $\partial f_j(x) \subseteq \partial f(x)$. By convexity and closedness of $\partial f(x)$ we get (iii). $\square$

We shall now give conditions to ensure that these inclusions are actually equalities.

Theorem 2.44 (Qualification conditions – Moreau-Rockafellar). *Let $f, g : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ be convex functions and $A : \mathcal{Y} \rightarrow \mathcal{X}$ be a continuous linear operator. Then the following holds.*

- (i) *If there exists a point $x_0 \in \text{dom } f \cap \text{dom } g$ such that g is continuous at x_0 , then for every $x \in \mathcal{X}$,*

$$\partial f(x) + \partial g(x) = \partial(f + g)(x).$$

- (ii) *If there exists a point $y_0 \in \mathcal{Y}$ such that f is continuous at Ay_0 then for every $y \in \mathcal{Y}$,*

$$\partial(f \circ A)(y) = A^* \partial f(Ay).$$

2 Convex functions

Proof. We start with the first item. We assume that $f + g$ is subdifferentible at x and let $p \in \partial(f + g)(x)$. We have

$$\forall y, \quad f(y) + g(y) \geq f(x) + g(x) + \langle p, y - x \rangle.$$

In particular $f(x), g(x) \in \mathbb{R}$ because f, g are both finite at $y = x_0$. Thus the previous inequality may be rewritten as

$$\forall y, \quad \tilde{f}(y) := f(y) - f(x) - \langle p, y - x \rangle \geq g(x) - g(y) =: \tilde{g}(y).$$

We have $\tilde{g} \leq \tilde{f}$ with $\tilde{f}$ convex and $\tilde{g}$ concave, and by assumption there exists a point x_0 where both f, g take real value and g is continuous. Thus by the functional separation theorem [Theorem 2.37](#) there exists a continuous affine form $a : x \mapsto \langle q, x \rangle + \alpha$ such that $\tilde{g} \leq a \leq \tilde{f}$, i.e.

$$\forall y, \quad g(x) - g(y) \leq \langle q, y \rangle + \alpha \leq f(y) - f(x) - \langle p, y - x \rangle.$$

Taking $y = x$ gives $\langle q, x \rangle + \alpha = 0$ thus

$$\forall y, \quad g(x) - g(y) \leq \langle q, y - x \rangle \leq f(y) - f(x) - \langle p, y - x \rangle.$$

It implies that $-q \in \partial g(x)$ and $p + q \in \partial f(x)$. Thus $p = (p + q) + (-q) \in \partial f(x) + \partial g(x)$.

Let $y \in \mathcal{Y}$ and assume that $p \in \partial(f \circ A)(y)$. We have

$$\forall z \in \mathcal{Y}, \quad f(Az) \geq f(Ay) + \langle p, z - y \rangle.$$

Thus, if $w \in \ker A$ then $f(Ay) = f(A(y \pm w)) \geq f(Ay) + \langle p, w - y \rangle$, i.e. $\pm \langle p, w \rangle \leq 0$ and $w \in \ker p$. It means that $\langle p, \cdot \rangle$ is actually a function of A : there exists a linear form $\ell : \text{Img } A \rightarrow \mathbb{R}$ such that $\langle p, \cdot \rangle = \ell \circ A$. We define a concave function $g : \mathcal{X} \rightarrow \mathbb{R} \cup \{-\infty\}$ by setting $g(z) = \ell(z)$ if $z \in \text{Img } A$ and $g(z) = -\infty$ otherwise. We have $g(Ay) = \ell(Ay) = \langle p, y \rangle$ and

$$\forall x \in \mathcal{X}, \quad \tilde{f}(x) := f(x) - f(Ay) \geq g(x) - g(Ay) =: \tilde{g}(x).$$

We know that $\tilde{f}$ is convex, $\tilde{g}$ is concave, Ay_0 is a point belonging to both their domains and where $\tilde{f}$ is continuous. The functional separation theorem [Theorem 2.37](#) then guarantees the existence of a continuous affine form $a : x \mapsto \langle q, x \rangle + \alpha$ such that $\tilde{f} \geq a \geq \tilde{g}$. Since $\tilde{f} = \tilde{g} = 0$ at Ay we get that $a = x \mapsto \langle q, x - Ay \rangle$, and

$$\forall x \in \mathcal{X}, \quad f(x) - f(Ay) \geq \langle q, x - Ay \rangle \geq g(x) - g(Ay).$$

By the first inequality, $q \in \partial f(Ay)$. By the second equality, taking $x = Aw$ yields $\langle A^*q, w - y \rangle \geq \langle p, w \rangle$ for any w , thus $A^*q = p$ and $p \in A^*\partial f(Ay)$, which is what we wanted to prove. $\square$

Proposition 2.45 (Subdifferential of envelope – Dubovitskii-Milyutin). *Let $f_1, \dots, f_n$ be n convex functions defined on $\mathcal{X}$ and $x \in \mathcal{X}$. If all functions f_i are continuous at x then*

$$\partial(\sup_i f_i)(x) = \overline{\text{conv}} \left(\bigcup_{i \in \arg\max_i f_i(x)} \partial f_i(x) \right).$$

Proof. Admitted. $\square$

2.4 First order optimality conditions

Proposition 2.46 (Fermat's rule – fundamental). *Let $f : \mathcal{X} \rightarrow (-\infty, +\infty]$. The function f has a minimum at a point x if and only if $0 \in \partial f(x)$.*

Proof. The function f has a minimum at x if and only if

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

which exactly means that $0 \in \partial f(x)$. □

Example 2.47 (Fermat's problem). We consider the minimization problem

$$\inf_{x \in \mathbb{R}^d} \sum_{i=1}^n \lambda_i \|x - x_i\| =: F(x)$$

where $\alpha_1, \dots, \alpha_n \geq 0$, $x_1, \dots, x_n \in \mathbb{R}^d$ are distinct points and $\|\cdot\|$ denotes the Euclidean norm. Minimizers are called *Fermat points*.

The subdifferential of the norm is given by

$$\partial \|\cdot\|(x) = \begin{cases} \frac{x}{\|x\|} & \text{if } x \neq 0, \\ B[0, 1] & \text{if } x = 0, \end{cases}$$

and we can see in general that $\partial(f(\cdot - u))(x) = \partial f(x - u)$ thus setting $f_i = \|\cdot - x_i\|$ we have

$$\partial f_i(x) = \begin{cases} \frac{x - x_i}{\|x - x_i\|} & \text{if } x \neq x_i, \\ B[0, 1] & \text{if } x = x_i. \end{cases}$$

We can check that the qualification conditions are satisfied and we may apply [Theorem 2.44](#) to get

$$\partial F(x) = \sum_{i=1}^n \alpha_i \partial f_i(x) = \begin{cases} \sum_{i=1}^n \alpha_i \frac{x - x_i}{\|x - x_i\|} & \text{if } x \notin \{x_1, \dots, x_n\}, \\ B[0, \alpha_j] + \sum_{i \neq j} \alpha_i \frac{x - x_i}{\|x - x_i\|} & \text{if } x = x_j. \end{cases}$$

As a consequence x is a Fermat point if and only if

$$x \notin \{x_1, \dots, x_n\} \quad \text{and} \quad \sum_{i=1}^n \alpha_i \frac{x - x_i}{\|x - x_i\|} = 0, \tag{2.4.1}$$

or

$$\text{for some } j, \quad x = x_j \quad \text{and} \quad \left\| \sum_{i \neq j} \alpha_i \frac{x - x_i}{\|x - x_i\|} \right\| \leq \alpha_j. \tag{2.4.2}$$

2 Convex functions

Example 2.48 (LASSO problem). We consider the minimization problem

$$\inf_{x \in \mathbb{R}^d} \frac{1}{2} \|Ax - b\|_2^2 + \gamma \|x\|_1 =: G(x),$$

where $A \in M_{m,d}(\mathbb{R})$ is a $m \times d$ matrix, $b \in \mathbb{R}^m$ and $\gamma > 0$. It is a linear least-square problem with ℓ^1 penalization. Setting $f(x) = \frac{1}{2} \|Ax - b\|_2^2$ and $g(x) = \|x\|_1$, $a(x) = |x|$, we have

$$\partial f(x) = \{\nabla f(x)\} = \{A^T(Ax - b)\}, \quad (2.4.3)$$

$$\partial g(x) = \partial a(x_1) \times \dots \times \partial a(x_d). \quad (2.4.4)$$

Since f and g are continuous and convex, qualification constraints are satisfied and we may apply [Theorem 2.44](#) to get $\partial G = \partial f + \gamma \partial g$. As a consequence x is a minimizer of G if and only if

$$A^T(Ax - b) = -\gamma v \quad \text{where } v \text{ satisfies } v_i \begin{cases} = \text{sign}(x_i) & \text{if } x_i \neq 0, \\ \in [-1, 1] & \text{if } x_i = 0, \end{cases}$$

or equivalently, denoting A_i the i th column of A ,

$$\begin{cases} \langle A_i, Ax - b \rangle = \gamma \text{sign}(x_i) & \text{if } x_i \neq 0 \\ |\langle A_i, Ax - b \rangle| \leq \gamma & \text{if } x_i = 0. \end{cases}$$

Now, we are interested in the first order optimality conditions of convex problems with constraints, i.e.

$$\inf_{x \in C} f(x)$$

where C is a convex set and f a convex function. It may be re-phrased as an apparently unconstrained problem:

$$\inf_{x \in \mathcal{X}} f(x) + \chi_C(x).$$

Theorem 2.49 (Optimality condition for constrained convex problems). *Let $C \subseteq \mathcal{X}$ a convex subset and $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ a convex function. If*

$$(\text{int } C \cap \text{dom } f \neq \emptyset) \quad \text{or} \quad f \text{ is continuous at a point of } C,$$

then the following assertions are equivalent:

- (i) x is a minimizer of f on C ,
- (ii) $\mathbf{N}_x C \cap (-\partial f(x)) \neq \emptyset$,
- (iii) there exists $x^* \in \partial f(x)$ such that

$$\forall y \in C, \quad \langle x^*, y - x \rangle \geq 0.$$

For the proof, we shall use the following remark, which is easy to check.

2 Convex functions

Remark 2.50. Let C be a convex subset of $\mathcal{X}$. For every $x \in \mathcal{X}$,

$$\partial\chi_C(x) = \mathbf{N}_x C.$$

Moreover $\text{dom } \chi_C = C$ and $\text{int dom } C = \{x \in \text{dom } f : \chi_C \text{ is continuous at } x\}$.

Proof of Theorem 2.49. We have seen that x minimizes f on C if and only if it minimizes $F = f + \chi_C$, which is equivalent by Fermat's rule to $0 \in \partial(f + \chi_C)(x)$. Since $\text{dom } \chi_C = C$ and the set of continuity points of χ_C is $\text{int } C$, the two hypotheses correspond to the existence of a point $x_0 \in \text{dom } f \cap \text{dom } \chi_C$ where either χ_C is continuous or f is continuous. The qualification conditions of Theorem 2.44 are thus satisfied hence $\partial F(x) = \partial f(x) + \partial\chi_C(x) = \partial f(x) + \mathbf{N}_x C$. We thus get the equivalence of (i) and (ii), and (iii) is simply a rephrasing of (ii) recalling the definition of the normal cone. $\square$

Remark 2.51. From this theorem we may establish quite easily the Karush-Kuhn-Tucker (KKT) theorem.

2.5 Lower semicontinuous convex functions

A class of functions is particularly suited to minimization problems, and is fundamental in convex analysis: the class of lower semicontinuous (lsc) functions. It is a wider class than continuous functions, that allows jumps but only “downwards”.

Definition 2.52 (Lower semicontinuous functions). A function $f : \mathcal{X} \rightarrow (-\infty, +\infty]$ is lower semicontinuous (lsc) if one of the following equivalent assertions is satisfied:

- (i) the epigraph of f is closed,
- (ii) the large sublevel sets $\{x : f(x) \leq \ell\}$ are closed for every $\ell \in \mathbb{R}$,

Remark 2.53 (Upper semicontinuity). An upper semi-continuous function is a function f such that $-f$ is lsc, or equivalently

- (i) the large superlevel sets $\{x : f(x) \geq \ell\}$ are closed for every $\ell \in \mathbb{R}$,
- (ii) the hypograph of f is closed,

Proof of the equivalence. Notice that for a given ℓ , $\text{epi } f \cap (\mathcal{X} \times \{\ell\}) = \{x : f(x) \leq \ell\} \times \{\ell\}$, the latter being homeomorphic to $\{x : f(x) \leq \ell\}$. Thus (i) implies (ii).

Assume that (ii) holds true. If $\text{epi } f$ was not closed, there would exist $(x, t) \notin \text{epi } f$, i.e. $f(x) > t$ such that for every neighborhood V of x and every neighborhood W of t , $(V \times W) \cap \text{epi } f \neq \emptyset$. Fix $W = (t - \varepsilon, t + \varepsilon)$ where $\varepsilon > 0$ is small enough to ensure $t + \varepsilon < f(x)$ and consider the sub-level set $L = \{f \leq t + \varepsilon\}$. For any neighborhood V of x , we know that $(V \times (t - \varepsilon, t + \varepsilon)) \cap \text{epi } f \neq \emptyset$ thus there exists $y \in V, s \in (t - \varepsilon, t + \varepsilon)$ such that $f(y) \leq s \leq t + \varepsilon$ hence $V \cap L \neq \emptyset$. We have $x \notin L$ but every neighborhood of x intersects L : L is not closed. We have shown (i). $\square$

Example 2.54. 1. A subset $A \subseteq \mathcal{X}$ is closed if and only if χ_A is lsc.

2 Convex functions

2. A function f valued in $\mathbb{R}$ is continuous if and only if f and $-f$ are lsc.
3. The cumulative distribution function F of a probability measure on $\mathbb{R}$ is upper semicontinuous but not lower semicontinuous in general.

Proposition 2.55 (Sequential characterization of lsc). *In a metric space $\mathcal{X}$, a function $f : \mathcal{X} \rightarrow [-\infty, +\infty]$ is lsc if and only if for every x and every sequence $x_n \rightarrow x$,*

$$f(x) \leq \liminf_{n \rightarrow +\infty} f(x_n).$$

Remark 2.56. The property above is called *sequential lower semicontinuity*. In general topological spaces it is not equivalent to lower semicontinuity, but it will often be more useful for the purpose of Calculus of Variations.

Proof. Let us assume that f is lsc as in [Definition 2.52](#), and take a sequence $x_n \rightarrow x$. For every real value $\ell < f(x)$, since $\{f > \ell\}$ is an open set containing x , for n large enough we have $f(x_n) > \ell$. Taking the $\liminf_n$ then $\ell \rightarrow f(x)$ yields $\liminf_n f(x_n) \geq f(x)$.

We proceed by contraposition to prove the remaining implication. Assume that f is not lsc, thus there exists a level t such that $\{f > t\}$ is not open, i.e. there exists an x such that $f(x) > t$ and for every neighborhood V of x , $V \setminus \{f > t\} \neq \emptyset$. In particular for every $n \in \mathbb{N}^*$, there exists a point $x_n \in B(x, 1/n) \cap \{f \leq t\}$. By construction $x_n \rightarrow x$ and $f(x_n) \leq t$ for every n , which implies that $\liminf_n f(x_n) \leq t < f(x)$. $\square$

Proposition 2.57 (Stability properties of lsc functions). *1. A supremum of lsc functions is lsc.*

2. A conic combination of lsc functions is lsc.

Proof. If $(f_i)_{i \in I}$ is a family of lsc functions and $f = \sup_i f_i$, $\text{epi } f = \bigcap_i \text{epi } f_i$, and since an intersection of closed sets is closed then f is lsc.

If $(f_i)_{i \in I}$ is a finite family of lsc functions and $\alpha_i \geq 0$ for every i then defining $f = \sum_i \alpha_i f_i$, the set

$$\{f > \ell\} = \bigcup_{(\ell_i) : \sum_i \alpha_i \ell_i > \ell} \bigcap_{i \in I} \{f_i > \ell_i\},$$

is open as a union of finite intersection of open sets. This shows that f is lsc. $\square$

Remark 2.58 (Lsc envelope). By the previous proposition, given $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$, the function

$$\hat{f}(x) = \sup\{g(x) : g : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\} \text{ is lsc and } g \leq f\}$$

is lower semicontinuous and is by definition the largest lsc function that is smaller than f . It is called the *lower semicontinuous envelope* of f .

Definition 2.59 (Convex lsc functions). We denote by $\Gamma(\mathcal{X})$ the set of convex lsc functions $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ and $\Gamma_0(\mathcal{X})$ the subset of *proper* lsc convex functions.

2 Convex functions

Proposition 2.60 (Characterization of convex lsc functions). *The following assertions are equivalent:*

(i) $f \in \Gamma(\mathcal{X})$,

(ii) f is the supremum of its continuous and affine minorants, i.e.

$$f(x) = \sup\{\langle x^*, x \rangle + \alpha : x^* \in \mathcal{X}^*, \alpha \in \mathbb{R} \text{ satisfy } f \geq \langle x^*, \cdot \rangle + \alpha\},$$

(iii) f is a supremum of continuous affine functions.

Proof. We have seen that convexity and lower semicontinuity are stable by supremum, thus (iii) $\implies$ (i), and obviously (ii) $\implies$ (iii). It remains to show that (i) $\implies$ (ii). We define

$$\tilde{f}(x) = \sup\{\langle x^*, x \rangle + \alpha : x^* \in \mathcal{X}^*, \alpha \in \mathbb{R} \text{ satisfy } f \geq \langle x^*, \cdot \rangle + \alpha\}.$$

If f is identically $+\infty$ there is nothing to prove. If $\text{dom } f \neq \emptyset$ let us consider $x_0 \in \text{dom } f$. For every real $t < f(x_0)$, since f is lsc, the superlevel set $L(t) = \{f > t\}$ is an open set which contains x_0 , thus it contains some ball $B(x_0, r)$ for $r > 0$. We see that $f \geq g := t - \chi_{B(x_0, r)}$, where f is convex, g is concave. Since $x_0 \in \overline{\text{dom } f}$ we have $\text{dom } f \cap B(x_0, r)$ contains some point x'_0 : both functions are finite at x'_0 and g is locally constant thus continuous at x'_0 . We may apply the functional separation theorem [Theorem 2.37](#) to find a continuous affine function a such that $g \leq a \leq f$, and we have $\tilde{f}(x_0) \geq a(x_0) \geq g(x_0) = t$. Since this is true for every $t < f(x_0)$ we get $f(x_0) \leq \tilde{f}(x_0)$.

We have shown that f admits at least one continuous affine minorant a_0 , and that $f = \tilde{f}$ on $\overline{\text{dom } f}$. Now let $x_0 \in \mathcal{X} \setminus \overline{\text{dom } f}$. We may use the Hahn-Banach strict separation theorem [Corollary 1.29](#) to separate strictly the closed convex set $\overline{\text{dom } f}$ from the compact convex set $\{x_0\}$. There exists a nontrivial $x^* \in \mathcal{X}^*$ such that

$$\alpha := \sup_{x \in \text{dom } f} \langle x^*, x \rangle < \langle x^*, x_0 \rangle.$$

The affine function $a : x \mapsto \langle x^*, x \rangle - \alpha$ satisfies $a \leq 0$ on $\text{dom } f$ and $a(x_0) > 0$. Take $t \in \mathbb{R}$. For $\lambda > 0$ large enough, $b(x_0) = a_0(x_0) + \lambda a(x_0) > t$ where $b = a_0 + \lambda b$. In such a way, $b(x_0) > t$ but $b \leq a_0 \leq f$ on $\text{dom } f$, thus $b \leq f$. Consequently $\tilde{f}(x_0) \geq b(x_0) > t$, and since t was arbitrary we obtain $\tilde{f}(x_0) = +\infty = f(x_0)$. □

We have seen that convexity and lower semicontinuity are preserved by supremum, justifying the following definition.

Definition 2.61 (Convex lsc envelope). The *convex lsc envelope* of $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ is defined by

$$\overline{\text{conv}} f = \sup\{g : g \leq f \text{ and } g \in \Gamma(\mathcal{X})\}.$$

By definition it is the largest convex lsc function which is smaller than f .

Remark 2.62. By the previous proposition, it may be equivalently characterized as:

$$\overline{\text{conv}} f = \sup\{\langle x^*, x \rangle + \alpha : x^* \in \mathcal{X}^*, \alpha \in \mathbb{R} \text{ s.t. } f \geq \langle x^*, \cdot \rangle + \alpha\}.$$

2.6 Fenchel-Legendre conjugate

The *Fenchel-Legendre conjugate* (or Fenchel/Legendre/Legendre-Fenchel conjugate/transform, or convex conjugate) is to convex functions what the support function is to convex sets.

Definition 2.63 (Fenchel-Legendre conjugate). The Fenchel-Legendre conjugate of a function $f : \mathcal{X} \rightarrow [-\infty, +\infty]$ is the function $f^* : \mathcal{X}^* \rightarrow [-\infty, +\infty]$ defined by

$$f^*(x^*) = \sup_{x \in \mathcal{X}} \langle x^*, x \rangle - f(x).$$

Remark 2.64. If f is valued in $\mathbb{R} \cup \{+\infty\}$ and is proper ($\text{dom } f \neq \emptyset$) then f^* does not take the value $-\infty$. Indeed, there exists $x \in \text{dom } f$, implying $f^* \geq \langle \cdot, x \rangle - f(x) > -\infty$.

Example 2.65. 1. If A is a set then $\chi_A^* = \sigma_A$, the support function of A . In particular if $f = \chi_{B[0,1]}$ then $f^* = \|\cdot\|_*$.

2. If $f = \|\cdot\|$ then $f^* = \chi_{B^*}$ where B^* is the closed unit ball for $\|\cdot\|_*$. In particular if $f = |\cdot|$ on $\mathbb{R}$ then $f^* = \chi_{[-1,1]}$.

3. If $f = \langle p^*, \cdot \rangle$ then $f^* = \chi_{\{p^*\}}$.

4. If $f(x) = \frac{1}{p}|x|^p$ on $\mathbb{R}$ with $p \in (1, +\infty)$ and $\mathbb{R}$ is identified with its dual, then $f^*(y) = \frac{1}{q}|y|^q$ where $\frac{1}{p} + \frac{1}{q} = 1$.

5. If $f(x) = \phi(\|x\|)$ where $\phi : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$ is even and convex then $f^*(x^*) = \phi^*(\|x^*\|)$.

These examples may be treated as exercise.

Proposition 2.66 (Basic properties of the conjugate). *Let $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper. We have*

- (i) (Fenchel-Young inequality) $f(x) + f^*(x^*) \geq \langle x^*, x \rangle$ for every $x \in \mathcal{X}, x^* \in \mathcal{X}^*$,
- (ii) $f^*(0) = -\inf f$,
- (iii) f^* is convex and lsc,
- (iv) if $f \in \Gamma_0(\mathcal{X})$, then $f^* \in \Gamma_0(\mathcal{X}^*)$

Proof. The first two items are direct consequence of the definition of the conjugate. The third item comes from the fact that f^* is defined as a supremum of continuous affine functions, thus it is convex and lsc. Now if $f \in \Gamma_0(\mathcal{X})$, by [Remark 2.64](#) f^* is valued in $(-\infty, +\infty]$. Moreover, by [Proposition 2.60](#) it is equal to the supremum of its affine minorants, in particular it admits an affine minorant $f \geq \langle x^*, \cdot \rangle + \alpha$, thus $f^*(x^*) = \sup_{x \in E} \langle x^*, x \rangle - f(x) \leq -\alpha < +\infty$ so that f^* is indeed proper. $\square$

Example 2.67 (Hölder inequality). On $\mathbb{R}^d$, Hölder inequality

$$\langle x, y \rangle \leq \|x\|_p \|y\|_q$$

where $\frac{1}{p} + \frac{1}{q} = 1$, $p \in (1, +\infty)$, may be deduced from Fenchel-Young inequality with $f(x) = \frac{|x|^p}{p}$, making use of the homogeneity of Hölder inequality (Exercise).

2 Convex functions

Proposition 2.68 (Stability properties of the conjugate). 1. $f \leq g \implies f^* \geq g^*$,

2. if $\lambda > 0$, $(\lambda f)^* = \lambda f^*(\cdot/\lambda)$,

3. if $\lambda \neq 0$, $(f(\lambda \cdot))^* = f^*(\cdot/\lambda)$,

4. if $\alpha \in \mathbb{R}$, $(f + \alpha)^* = f^* - \alpha$,

5. if $v \in \mathcal{X}$, $f(\cdot - v)^* = f^* + \langle \cdot, v \rangle$,

6. $(\inf_{i \in I} f_i)^* = \sup_{i \in I} f_i^*$,

7. $(\sup_{i \in I} f_i)^* \leq \inf_{i \in I} f_i^*$.

Proof. Left as exercise. □

We shall want to compute the biconjugate, i.e. the conjugate of the conjugate, of a function f . Since f^* is a function defined on $\mathcal{X}^*$, f^{**} would normally be defined on $\mathcal{X}^{**}$, but we will not make this choice. We rather see $\mathcal{X}, \mathcal{X}^*$ as being related through the duality pairing $(\mathcal{X} \times \mathcal{X}^*) \ni (x, x^*) \mapsto \langle x^*, x \rangle$, and with a slight abuse when $g : \mathcal{X}^* \rightarrow [-\infty, +\infty]$ we denote by g^* the function from $\mathcal{X} \rightarrow [-\infty, +\infty]$ defined by

$$g^*(x) = \sup_{x^* \in \mathcal{X}^*} \langle x^*, x \rangle - g(x^*).$$

Remark 2.69. The basic properties and the stability properties seen above hold also for this (pre-)conjugate. In particular Fenchel-Young inequality is also true.

Notice that when $\mathcal{X}$ is a Hilbert space, $\mathcal{X}^*$ can be identified with $\mathcal{X}$ and the two transforms are the same.⁵

Theorem 2.70 (Fenchel-Moreau). *Let $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper function. Then f^{**} is the lower semicontinuous envelope of f , namely*

$$f^{**} = \overline{\text{conv}} f.$$

In particular $f^{**} \leq f$ and

$$f = f^{**} \iff f \in \Gamma_0(\mathcal{X}).$$

Proof. Let us show first that $f^{**} \leq f$. By Fenchel-Young's inequality, for every $x \in \mathcal{X}$ and $x^* \in \mathcal{X}^*$ we have $f(x) + f^*(x^*) \geq \langle x^*, x \rangle$, i.e. $\langle x^*, x \rangle - f^*(x^*) \leq f(x)$, thus

$$f^{**}(x) = \sup_{x^* \in \mathcal{X}^*} \langle x^*, x \rangle - f^*(x^*) \leq f(x).$$

If f^* is identically $+\infty$ then f^{**} is identically $-\infty$. If not, then by definition of f^{**} as a supremum, f^{**} never takes the value $-\infty$, and it is convex and lower semicontinuous as a supremum of continuous affine functions, thus $f^{**} \leq \overline{\text{conv}} f$. In any case we have $f^{**} \leq \overline{\text{conv}} f$.

⁵More generally, when $\mathcal{X}$ is reflexive, i.e. we can canonically identify $\mathcal{X} \approx \mathcal{X}^{**}$ then the pre-conjugate transform and the conjugate on the dual may be identified as well.

2 Convex functions

We shall prove the converse inequality now, using the fact that $\overline{\text{conv}} f$ is equal to the supremum of its continuous affine minorants. Let $a = \langle p^*, \cdot \rangle + \alpha$ be a continuous affine minorant of f . We have $a \leq f \implies a^* \geq f^* \implies a^{**} \leq f^{**}$. Let us compute a^{**} .

$$a^* = (\langle p^*, \cdot \rangle + \alpha)^* = \chi_{p^*} - \alpha,$$

and

$$\forall x \in \mathcal{X}, \quad a^{**}(x) = \sup_{x^* \in \mathcal{X}^*} \langle x^*, x \rangle - \chi_{p^*}(x^*) + \alpha = \langle p^*, x \rangle + \alpha = a(x).$$

We have shown that $f^{**} \geq a$ for every affine minorant of f , thus $f^{**} \geq \overline{\text{conv}} f$, and we actually have equality. $\square$

In line with the previous comments when we introduced the conjugate of a function defined on the dual space, as a function defined on the primal space, we define the subdifferential of a function $g : \mathcal{X}^* \rightarrow [-\infty, +\infty]$ as a subset of $\mathcal{X}$, namely for every $x^* \in \mathcal{X}^*$,

$$\partial g(x^*) = \{x \in \mathcal{X} : \forall y^* \in \mathcal{Y}^*, g(y^*) \geq g(x^*) + \langle y^* - x^*, x \rangle\}.$$

Theorem 2.71 (Subdifferential and conjugation). *Let $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper function, $x \in \mathcal{X}$ and $x^* \in \mathcal{X}^*$. The following conditions are equivalent,*

- (i) $x^* \in \partial f(x)$,
- (ii) $f(x) + f^*(x^*) = \langle x^*, x \rangle$.

If moreover $f \in \Gamma_0(\mathcal{X})$ then they are also equivalent with

- (iii) $x \in \partial f^*(x^*)$.

Proof. We know that

$$\begin{aligned} x^* \in \partial f(x) &\iff \forall y \in \mathcal{X}, f(y) \geq f(x) + \langle x^*, y - x \rangle \\ &\iff \forall y \in \mathcal{X}, \langle x^*, y \rangle - f(y) \leq \langle x^*, x \rangle - f(x) \\ &\iff f^*(x^*) \leq \langle x^*, x \rangle - f(x) \\ &\iff f(x) + f^*(x^*) \leq \langle x^*, x \rangle. \end{aligned}$$

Since the converse inequality is Fenchel-Young's inequality, we get the equivalence between the two first items.

Now let us prove the equivalence with the third item when $f \in \Gamma_0(\mathcal{X})$. We know from Fenchel-Moreau [Theorem 2.70](#) that $f^{**} = f$. Therefore

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

and re-doing the same computations as above, this is equivalent to $x \in \partial f^*(x^*)$. $\square$

2.7 Convex duality

Let us start with a very general problem in the form of a sum:

$$\text{minimize } \{f(x) + g(Ax) : x \in \mathcal{X}\}, \quad (\text{S})$$

where $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$, $g : \mathcal{Y} \rightarrow \mathbb{R} \cup \{+\infty\}$ and $A : \mathcal{X} \rightarrow \mathcal{Y}$ is a continuous linear operator.

How can we guess a dual problem for (S)? We use the fact that $g \geq g^{**}$:

$$\begin{aligned} \inf_{x \in \mathcal{X}} f(x) + g(Ax) &\geq \inf_{x \in \mathcal{X}} f(x) + \sup_{y^* \in \mathcal{X}^*} \langle y^*, Ax \rangle - g^*(y^*) \\ &= \inf_{x \in \mathcal{X}} \sup_{y^* \in \mathcal{X}^*} f(x) + \langle y^*, Ax \rangle - g^*(y^*) \end{aligned}$$

then we use the general fact that $\inf_{\mathcal{X}} \sup_{\mathcal{Y}} \geq \sup_{\mathcal{Y}} \inf_{\mathcal{X}}$

$$\begin{aligned} &\geq \sup_{y^* \in \mathcal{X}^*} \inf_{x \in \mathcal{X}} f(x) + \langle y^*, Ax \rangle - g^*(y^*) \\ &= \sup_{y^* \in \mathcal{X}^*} \left(\inf_{x \in \mathcal{X}} \langle A^* y^*, x \rangle + f(x) \right) - g^*(y^*) \\ &= \sup_{y^* \in \mathcal{X}^*} - \left(\sup_{x \in \mathcal{X}} \langle A^* (-y^*), x \rangle - f(x) \right) - g^*(y^*) \\ &= \sup_{y^* \in \mathcal{X}^*} -f^*(-A^* y^*) - g^*(y^*). \end{aligned}$$

Thus a candidate dual problem for (S) is:

$$\text{maximize } \{-f^*(-A^* y^*) - g^*(y^*) : y^* \in \mathcal{Y}^*\}. \quad (\text{S}^*)$$

The computations above show that *weak duality holds*, in the sense that

$$\inf (\text{S}) \geq \sup (\text{S}^*).$$

We shall establish this and also provide very general conditions to have equality above, i.e. *strong duality*.

Fenchel-Rockafellar duality theorem and applications

We consider again the problem in sum form

$$\text{minimize } \{f(x) + g(Ax) : x \in \mathcal{X}\}, \quad (\text{S})$$

and its candidate dual

$$\text{maximize } \{-f^*(-A^* y^*) - g^*(y^*) : y^* \in \mathcal{Y}^*\}. \quad (\text{S}^*)$$

2 Convex functions

Proposition 2.72 (Weak duality). *If $f \in \Gamma_0(\mathcal{X})$, $g \in \Gamma_0(\mathcal{Y})$ and $A : \mathcal{X} \rightarrow \mathcal{Y}$ a continuous linear operator then*

$$\inf (\text{S}) \geq \sup (\text{S}^*).$$

Moreover, if equality holds, then the following assertions are equivalent:

- (i) x is a minimizer of (S) and y^* a maximizer of (S*) ;
- (ii) $A^*y^* \in \partial f(x)$ and $-y^* \in \partial g(Ax)$;
- (iii) $x \in \partial f^*(A^*y)$ and $Ax \in \partial g^*(-y^*)$.

Proof. The inequality $\inf (\text{S}) \geq \sup (\text{S}^*)$ follows from the computations done at the very beginning of this chapter. Let us write it more efficiently, now that we know what is the candidate dual problem. By Fenchel-Young inequality, for every $x \in \mathcal{X}, x^* \in \mathcal{X}^*, y \in \mathcal{Y}, y \in \mathcal{Y}^*$:

$$\begin{aligned} f(x) + f^*(A^*y^*) &\geq \langle A^*y^*, x \rangle \\ g(Ax) + g^*(-y^*) &\geq \langle -y^*, Ax \rangle. \end{aligned}$$

Since $\langle A^*y^*, x \rangle = \langle y^*, Ax \rangle$, by summing the two and taking the infima in x, y^* we get

$$\inf \{f(x) + g(Ax) : x \in \mathcal{X}\} + \inf \{f^*(-A^*y^*) + g^*(y^*) : y^* \in \mathcal{Y}^*\} \geq 0, \quad (2.7.1)$$

which exactly means that $\inf (\text{S}) \geq \sup (\text{S}^*)$.

Assume that we have equality $\inf (\text{S}) = \sup (\text{S}^*)$. If x is a minimizer of (S) and y^* a maximizer of (S*), then we have equality in (2.7.1):

$$\begin{aligned} &f(x) + f^*(A^*y^*) + g(Ax) + g^*(-y^*) = 0 \\ \iff &\underbrace{f(x) + f^*(A^*y^*) - \langle A^*y^*, x \rangle}_{\geq 0} + \underbrace{g(Ax) + g^*(-y^*) - \langle -y^*, Ax \rangle}_{\geq 0} = 0 \\ \iff &f(x) + f^*(A^*y^*) = \langle A^*y^*, x \rangle \quad \text{and} \quad g(Ax) + g^*(-y^*) = \langle -y^*, Ax \rangle. \end{aligned}$$

By **Theorem 2.71**, the latter equalities are equivalent to $A^*y^* \in \partial f(x)$ and $-y^* \in \partial g(Ax)$, or equivalently $x \in \partial f^*(A^*y^*)$ and $Ax \in \partial g^*(-y^*)$. Thanks to the qualification condition, we can apply □

Theorem 2.73 (Fenchel-Rockafellar). *If the following qualification condition holds,*

$$\exists x_0 \in \text{dom } f \text{ s. t. } g \text{ is finite and continuous at } Ax_0,$$

then strong duality holds, and the dual problem admits a solution, i.e.

$$\inf (\text{S}) = \max (\text{S}^*).$$

Due to time constraints, we shall only give a proof in the case where the infimum of the primal problem is attained.

2 Convex functions

Proof assuming $\inf$ (S) is attained. In this case, we can give a proof using subdifferential calculus. Let us take x a minimizer of (S), i.e. $x \in \operatorname{argmin} f + g$, which is equivalent, by Fermat's rule, to $0 \in \partial(f + g \circ A)(x)$. The qualification assumption matches the qualification condition of [Theorem 2.44](#), so that

$$0 \in \partial(f + g \circ A)(x) = \partial f(x) + A^* \partial g(Ax),$$

thus there exists $y^* \in \partial g(Ax)$ such that $-A^* y^* \in \partial f(x)$. Using Fenchel-Young's equality case stated in [Theorem 2.71](#) we have

$$f(x) + f^*(-A^* y^*) = \langle -A^* y^*, x \rangle \quad \text{and} \quad g(Ax) + g^*(y^*) = \langle y^*, Ax \rangle$$

and summing the two yields

$$f(x) + g(Ax) + f^*(-A^* y^*) + g^*(y^*) = 0,$$

or equivalently

$$f(x) + g(Ax) = -f^*(-A^* y^*) - g^*(y^*).$$

Since by weak duality ([Proposition 2.72](#)) we have $\inf$ (S) $\geq$ $\sup$ (S*), it implies that we have equality and y^* is a maximizer of (S*). $\square$

Example 2.74 (Dual LASSO). We consider again the LASSO problem

$$\text{minimize} \quad \left\{ \frac{1}{2} \|Ax - b\|_2^2 + \lambda \|x\|_1 : x \in \mathbb{R}^d \right\}.$$

It is a problem of the form (S) with $f(x) = \lambda \|x\|_1$ and $g(y) = \frac{1}{2} \|y - b\|_2^2$. By [Example 2.65](#) and [Proposition 2.68](#) we obtain the Fenchel-Legendre transforms of f and g :

$$f^*(x^*) = \lambda \chi_{[-1,1]^d}(x^*/\lambda) = \chi_{[-\lambda, \lambda]^d}(x^*)$$

and

$$g^*(y^*) = \frac{\|y^*\|_2^2}{2} + \langle y^*, b \rangle = \frac{\|y^* + b\|_2^2}{2} - \frac{\|b\|_2^2}{2}.$$

Thus the dual problem reads:

$$\text{maximize} \quad \left\{ -\frac{1}{2} \|x^* + b\|_2^2 + \|b\|_2^2 : A^* x^* \in [-\lambda, \lambda]^d \right\}.$$

Since both functions f, g are continuous on $\mathcal{X}$ and $\mathcal{Y}$ respectively, we may apply then Fenchel-Rockafellar Theorem so that strong duality holds and we have existence to the dual problem:

$$\begin{aligned} & \inf \left\{ \frac{1}{2} \|Ax - b\|_2^2 + \lambda \|x\|_1 : x \in \mathbb{R}^d \right\} \\ &= \max \left\{ -\frac{1}{2} \|A^* x^* + b\|_2^2 + \|b\|_2^2 : x^* \in \mathbb{R}^d, A^* x^* \in [-\lambda, \lambda]^d \right\}. \end{aligned} \quad (2.7.2)$$

2 Convex functions

Example 2.75 (Optimal Transport). The optimal transport problem consists in the following. We are given two probability measures $\mu, \nu \in \mathcal{P}(K)$ on a compact subset $K \subseteq \mathbb{R}^d$, μ representing for example a density of sand (or people, particules, data points...) that we want to move to a target density ν representing the shape of the hole (or destination of people or particules, or other data points...). We are also given a cost function $c : K \times K \rightarrow \mathbb{R}$, where $c(x, y)$ represents the cost for moving a unit mass from position x to position y . Typically, $c(x, y) = \|y - x\|^p$ for some power $p \geq 1$. The optimal transport problem is the following,

$$\text{minimize } \left\{ \int_{X \times X} c(x, y) d\gamma(x, y) : \gamma \in \Pi(\mu, \nu) \right\}, \quad (2.7.3)$$

where $\Pi(\mu, \nu)$ is the set of probability measures γ on the product space $X \times X$ having μ, ν as first and second marginals, i.e. $\gamma(A \times X) = \mu(A)$ and $\gamma(X \times A) = \nu(A)$ for every Borel set A . In probabilistic terms, it means that if a random variable (X, Y) has law γ then $X \sim \mu$ and $Y \sim \nu$.

Can we cast the minimization problem (2.7.3) in the context of convex analysis, and derive a “dual” problem? To do this, we have to better understand the space of probability measures, and embed them in a normed vector space.

We already know that a positive finite measure $\mu \in \mathcal{M}^+(K)$ on a compact subset $K \subseteq \mathbb{R}^n$ induces a continuous linear form I_μ on $\mathcal{C}^0(K)$, by integration:

$$\forall \phi \in \mathcal{C}^0(K), \quad \langle I_\mu, \phi \rangle = \int_K \phi d\mu.$$

Since two positive measures μ, ν are equal if and only if $I_\mu = I_\nu$ we can identify a measure and its induced linear form on continuous functions: $\mu \simeq I_\mu \in \mathcal{C}^0(K)^*$.

A (finite) signed measure ω is defined as the difference of two finite positive measures, i.e. it can be written as $\omega = \omega_+ - \omega_-$ where $\omega_\pm \in \mathcal{M}^+(X)$. According to the identification $\mu \simeq I_\mu$ made above, signed measures can also be seen as belonging to $\mathcal{C}^0(X)^*$. Actually, it is possible to show that the set of finite signed measures, denoted by $\mathcal{M}(X)$, can be *exactly* identified with the dual of continuous functions⁶:

$$\mathcal{M}(X) \simeq \mathcal{C}^0(X)^*.$$

Hence the minimization problem (2.7.3) takes as unknown a measure $\gamma \in \mathcal{M}(X \times X)$ i.e. a dual object, and should be viewed as the dual problem of something: we want to find a primal (or “pre-dual”) problem⁷. We shall use Fenchel-Rockafellar Theorem, but first let us try to guess the pre-dual problem, in the spirit of the formal computations done at the beginning of the section (doing an inf – sup exchange). We have:

$$\begin{aligned} & \inf \left\{ \int_{X \times X} c(x, y) d\gamma(x, y) : \gamma \in \Pi(\mu, \nu) \right\} \\ &= \inf_{\gamma \in \mathcal{M}_+(X \times X)} \int_{X \times X} c d\gamma + \chi_{\gamma \text{ has marginals } (\mu, \nu)}. \end{aligned}$$

⁶This is due to a so-called Riesz representation theorem, stating that a continuous linear form on $\mathcal{C}^0(K)$ is of the form $I_{\omega_+} - I_{\omega_-}$.

⁷In Optimal Transport Theory, (2.7.3) is nonetheless called the primal problem.

2 Convex functions

Let us write the indicator function in the last term as a supremum of linear functions. Notice that for $\gamma \in \mathcal{M}(X \times X)$,

$$\chi_\gamma \text{ has marginals } (\mu, \nu) = \sup_{\phi, \psi \in \mathcal{C}^0(X)} \int_X \phi \, d\mu + \int_X \psi \, d\nu - \int_{X \times X} (\phi(x) + \psi(y)) \, d\gamma(x, y).$$

Indeed, if γ has marginals (μ, ν) then $\int_{X \times X} (\phi(x) + \psi(y)) \, d\gamma(x, y) = \int_{X \times X} \phi(x) \, d\gamma(x, y) + \int_{X \times X} \psi(y) \, d\gamma(x, y) = \int_X \phi \, d\mu + \int_X \psi \, d\nu$ and the right-hand side is zero. If on the contrary, the first marginal γ_1 of γ is not μ , it means that $\int_{X \times X} \phi(x) \, d\gamma(x, y) = \int_X \phi(x) \, d\gamma_1(x) \neq \int_X \phi(x) \, d\mu(x)$, then replacing ϕ by $\lambda\phi$ and making λ tend to $+\infty$ or $-\infty$ yields that the supremum on the right-hand side is $+\infty$.

Now we can pursue our formal computations, denoting $\phi \oplus \psi(x, y) = \phi(x) + \psi(y)$:

$$\begin{aligned} & \inf \left\{ \int_{X \times X} c(x, y) \, d\gamma(x, y) : \gamma \in \Pi(\mu, \nu) \right\} \\ &= \inf_{\gamma \in \mathcal{M}_+(X \times X)} \int_{X \times X} c \, d\gamma + \chi_{\gamma \text{ has marginals } (\mu, \nu)} \cdot \\ &= \inf_{\gamma \in \mathcal{M}_+(X \times X)} \sup_{\phi, \psi \in \mathcal{C}^0(X)} \int_{X \times X} (c - \phi \oplus \psi) \, d\gamma + \int_X \phi \, d\mu + \int_X \psi \, d\nu \\ &\geq \sup_{\phi, \psi \in \mathcal{C}^0(X)} \inf_{\gamma \in \mathcal{M}_+(X \times X)} \int_{X \times X} (c - \phi \oplus \psi) \, d\gamma + \int_X \phi \, d\mu + \int_X \psi \, d\nu \\ &= \sup_{\phi, \psi \in \mathcal{C}^0(X)} \left(\int_X \phi \, d\mu + \int_X \psi \, d\nu + \inf_{\gamma \in \mathcal{M}_+(X \times X)} \int_{X \times X} (c - \phi \oplus \psi) \, d\gamma \right). \end{aligned}$$

One can then check that the infimum in the last term is either 0 or $-\infty$:

$$\inf_{\gamma \in \mathcal{M}_+(X \times X)} \int_{X \times X} (c - \phi \oplus \psi) \, d\gamma = \begin{cases} 0 & \text{if } \phi \oplus \psi \leq c \text{ on } X \times X, \\ -\infty & \text{otherwise.} \end{cases}$$

Indeed if $\phi \oplus \psi \leq c$ then the best one can do is take $\gamma = 0$, while if $\phi(x) = \psi(y) > c(x, y)$ at a point (x, y) , taking $\gamma = m\delta_{(x, y)}$ for $m \rightarrow +\infty$ as competitor in the infimum yields $-\infty$. As a consequence,

$$\begin{aligned} & \inf \left\{ \int_{X \times X} c(x, y) \, d\gamma(x, y) : \gamma \in \Pi(\mu, \nu) \right\} \\ &= \sup_{\phi, \psi \in \mathcal{C}^0(X)} \left(\int_X \phi \, d\mu + \int_X \psi \, d\nu + \inf_{\gamma \in \mathcal{M}_+(X \times X)} \int_{X \times X} (c - \phi \oplus \psi) \, d\gamma \right) \\ &= \sup_{\phi, \psi \in \mathcal{C}^0(X)} \left\{ \int_X \phi \, d\mu + \int_X \psi \, d\nu : \phi \oplus \psi \leq c \right\}. \end{aligned}$$

The latter is the so-called “dual” problem of optimal transport, which is actually the primal or pre-dual problem.

2 Convex functions

To prove strong duality, we just have to apply Fenchel-Rockafellar to suitable functions f, g and operator A , namely:

$$A : \mathcal{C}^0(X) \times \mathcal{C}^0(X) \ni (\phi, \psi) \mapsto \phi \oplus \psi = \xi \in \mathcal{C}^0(X \times X), \quad (2.7.4)$$

$$g : \mathcal{C}^0(X) \times \mathcal{C}^0(X) \ni (\phi, \psi) \mapsto - \int \phi \, d\mu - \int \psi \, d\nu \quad (2.7.5)$$

$$f : \mathcal{C}^0(X \times X) \ni \xi \mapsto \chi_{\xi \leq c}. \quad (2.7.6)$$

We just have to compute A^* , the Fenchel-Legendre conjugates f^*, g^* and check that the qualification conditions are valid to apply Fenchel-Rockafellar's Theorem (Exercise).