https://0aax.github.io/convex-eval https://github.com/0aax/convex-eval https://0aax.github.io/convex-eval/docs

# Convex Eval

Audrey Xie

February 9, 2026

# Chapter 1

# Convex Sets

## 1.1 Generalities

### 1.1.1 Definitions and first examples

**Definition 1.1.1.** The set $C \subset \mathbb{R}^n$ is said to be convex if $\alpha x + (1-\alpha)x'$ is in $C$ whenever $x$ and $x'$ are in $C$, and $\alpha \in ]0,1[$ (or equivalently $\alpha \in [0,1]$).

### 1.1.2 Convexity preserving operations on sets

**Proposition 1.1.1.** *Let $\{C_j\}_{j \in J}$ be an arbitrary family of convex sets. Then*

$$C := \bigcap \{C_j : \; j \in J\}$$

*is convex.*

*Proof.* Immediate from the very Definition 1.1.1. □

**Proposition 1.1.2.** *For $i = 1, \ldots, k$, let $C_i \subset \mathbb{R}^{n_i}$ be convex sets. Then $C_1 \times \cdots \times C_k$ is a convex set of $\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$.*

*Proof.* Straightforward. □

**Proposition 1.1.3.** *Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be an affine mapping and $C$ a convex set of $\mathbb{R}^n$. The image $A(C)$ of $C$ under $A$ is convex in $\mathbb{R}^m$.*

*If $D$ is a convex set of $\mathbb{R}^m$, the inverse image*

$$A^{-1}(D) := \{x \in \mathbb{R}^n : \; A(x) \in D\}$$

*is convex in $\mathbb{R}^n$.*

*Proof.* For $x$ and $x'$ in $\mathbb{R}^n$, the image under $A$ of the segment $[x, x']$ is clearly the segment $[A(x), A(x')] \subset \mathbb{R}^m$. This proves the first claim, but also the second: indeed, if $x$ and $x'$ are such that $A(x)$ and $A(x')$ are both in the convex set $D$, then every point of the segment $[x, x']$ has its image in $[A(x), A(x')] \subset D$. □

**Proposition 1.1.4.** *If $C$ is convex, so are its interior* $\operatorname{int} C$ *and its closure* $\overline{C}$.

*Proof.* For given different $x$ and $x'$, and $\alpha \in ]0,1[$, we set $x'' = \alpha x + (1-\alpha)x' \in ]x, x'[$.

Take first $x$ and $x'$ in $\operatorname{int} C$. Choosing $\delta > 0$ such that $B(x', \delta) \subset C$, we show that $B(x'', (1-\alpha)\delta) \subset C$. As often in convex analysis, it is probably best to draw a picture. The ratio $\|x'' - x\| / \|x' - x\|$ being precisely $1 - \alpha$, Fig. 1.2.3 clearly shows that $B(x'', (1-\alpha)\delta)$ is just the set $\alpha x + (1-\alpha)B(x', \delta)$, obtained from segments with endpoints in $\operatorname{int} C$: $x'' \in \operatorname{int} C$.

Now, take $x$ and $x'$ in $\operatorname{cl} C$: we select in $C$ two sequences $\{x_k\}$ and $\{x'_k\}$ converging to $x$ and $x'$ respectively. Then, $\alpha x_k + (1-\alpha)x'_k$ is in $C$ and converges to $\alpha x + (1-\alpha)x'$, which is therefore in $\operatorname{cl} C$. $\square$

### 1.1.3 Convex combinations and convex hulls

**Definition 1.1.2.** A *convex combination* of elements $x_1, \ldots, x_k$ in $\mathbb{R}^n$ is an element of the form

$$\sum_{i=1}^{k} \alpha_i x_i$$

where

$$\sum_{i=1}^{k} \alpha_i = 1 \quad \text{and} \quad \alpha_i \geq 0 \text{ for } i = 1, \ldots, k.$$

**Proposition 1.1.5.** *A set $C \subset \mathbb{R}^n$ is convex if and only if it contains every convex combination of its elements.*

*Proof.* The condition is sufficient: convex combinations of two elements just make up the segment joining them. To prove necessity, take $x_1, \ldots, x_k$ in $C$ and $\alpha = (\alpha_1, \ldots, \alpha_k) \in \Delta_k$. One at least of the $\alpha_i$'s is positive, say $\alpha_1 > 0$. Then form

$$y_2 := \frac{\alpha_1}{\alpha_1 + \alpha_2} x_1 + \frac{\alpha_2}{\alpha_1 + \alpha_2} x_2 \qquad \left[ = \frac{1}{\alpha_1 + \alpha_2}(\alpha_1 x_1 + \alpha_2 x_2) \right]$$

which is in $C$ by Definition 1.1.1 itself. Therefore,

$$y_3 := \frac{\alpha_1 + \alpha_2}{\alpha_1 + \alpha_2 + \alpha_3} y_2 + \frac{\alpha_3}{\alpha_1 + \alpha_2 + \alpha_3} x_3 \qquad \left[ = \frac{1}{\sum_{i=1}^{3} \alpha_i} \sum_{i=1}^{3} \alpha_i x_i \right]$$

is in $C$ for the same reason; and so on until

$$y_k := \frac{\alpha_1 + \cdots + \alpha_{k-1}}{1} y_{k-1} + \frac{\alpha_k}{1} x_k \qquad \left[ = \tfrac{1}{1} \sum_{i=1}^{k} \alpha_i x_i \right].$$

$\square$

**Proposition 1.1.6.** *The convex hull can also be described as the set of all convex combinations:*

$$\operatorname{co} S := \bigcap \{C : C \text{ is convex and contains } S\} = \Big\{x \in \mathbb{R}^n : \text{for some } k \in \mathbb{N}_*, \text{ there exist } x_1, \dots, x_k \in S \text{ and } \alpha = \tag{1.3.2}$$

*Proof.* Call $T$ the set described in the rightmost side of (1.3.2). Clearly, $T \supset S$. Also, if $C$ is convex and contains $S$, then it contains all convex combinations of elements

For this, take two points $x$ and $y$ in $T$, characterized respectively by $(x_1, \alpha_1), \dots,$ $(x_k, \alpha_k)$ and by $(y_1, \beta_1), \dots, (y_\ell, \beta_\ell)$; take also $\lambda \in [0, 1]$. Then $\lambda x + (1-\lambda)y$ is a certain combination of $k + \ell$ elements of $S$; this combination is convex because its coefficients $\lambda\alpha_i$ and $(1-\lambda)\beta_j$ are nonnegative, and their sum is

$$\lambda \sum_{i=1}^{k} \alpha_i + (1-\lambda) \sum_{j=1}^{\ell} \beta_j = \lambda + 1 - \lambda = 1.$$

$\square$

**Theorem 1.1.1.** *(C. Carathéodory) Any $x \in \operatorname{co} S \subset \mathbb{R}^n$ can be represented as a convex combination of $n+1$ elements of $S$.*

*Proof.* Take an arbitrary convex combination $x = \sum_{i=1}^{k} \alpha_i x_i$, with $k > n+1$. We will show that one of the $x_i$'s can be assigned a 0-coefficient without changing $x$. For this, assume that all coefficients $\alpha_i$ are positive (otherwise we are done).

The $k > n+1$ elements $x_i$ are certainly affinely dependent: (1.3.1) tells us that we can find $\delta_1, \dots, \delta_k$, not all zero, such that

$$\sum_{i=1}^{k} \delta_i x_i = 0 \quad \text{and} \quad \sum_{i=1}^{k} \delta_i = 0.$$

There is at least one positive $\delta_i$; and we can set $\alpha_i' := \alpha_i - t^* \delta_i$ for $i = 1, \dots, k$, where

$$t^* := \max\{t \geq 0 : \ \alpha_i - t\delta_i \geq 0 \text{ for } i = 1, \dots, k\} = \min_{\delta_j > 0} \frac{\alpha_j}{\delta_j}.$$

Clearly enough,

$$\alpha_i' \geq 0 \quad \text{for } i = 1, \dots, k \qquad [\text{automatic if } \delta_i \leq 0, \text{ by construction of } t^* \text{ if } \delta_i > 0]$$

$$\sum_{i=1}^{k} \alpha_i' = \sum_{i=1}^{k} \alpha_i - t^* \sum_{i=1}^{k} \delta_i = 1;$$

$$\sum_{i=1}^{k} \alpha_i' x_i = x - t^* \sum_{i=1}^{k} \delta_i x_i = x;$$

and there exists $i_0$ such that $\alpha'_{i_0} = 0$. [by construction of $t^*$]

In other words, we have expressed $x$ as a convex combination of $k-1$ among the $x_i$'s; our claim is proved.

Now, if $k-1 = n+1$, the proof is finished. If not, we can apply the above construction to the convex combination $x = \sum_{i=1}^{k-1} \alpha'_i x_i$ and so on. The process can be continued until there remain only $n+1$ elements (which may be affinely independent). $\square$

**Theorem 1.1.2.** *(W. Fenchel and L. Bunt) If $S \subset \mathbb{R}^n$ has no more than $n$ connected components (in particular, if $S$ is connected), then any $x \in \operatorname{co} S$ can be expressed as a convex combination of $n$ elements of $S$.*

### 1.1.4 Closed convex sets and hulls

**Definition 1.1.3.** The *closed convex hull* of a nonempty set $S \subset \mathbb{R}^n$ is the intersection of all closed convex sets containing $S$. It will be denoted by $\overline{\operatorname{co}}\, S$.

**Proposition 1.1.7.** *The closed convex hull $\overline{\operatorname{co}} S$ of Definition 1.4.1 is the closure* $\operatorname{cl}(\operatorname{co} S)$ *of the convex hull of $S$.*

*Proof.* Because $\operatorname{cl}(\operatorname{co} S)$ is a closed convex set containing $S$, it contains $\overline{\operatorname{co}} S$ as well. On the other hand, take a closed convex set $C$ containing $S$; being convex, $C$ contains $\operatorname{co} S$; being closed, it contains also the closure of $\operatorname{co} S$. Since $C$ was arbitrary, we conclude $\bigcap C \supset \operatorname{cl} \operatorname{co} S$. $\square$

**Theorem 1.1.3.** *If $S$ is bounded [resp. compact], then $\operatorname{co} S$ is bounded [resp. compact].*

*Proof.* Let $x = \sum_{i=1}^{n+1} \alpha_i x_i \in \operatorname{co} S$. If $S$ is bounded, say by $M$, we can write

$$\|x\| \leq \sum_{i=1}^{n+1} \alpha_i \|x_i\| \leq M \sum_{i=1}^{n+1} \alpha_i = M.$$

Now take a sequence $\{x^k\} \subset \operatorname{co} S$. For each $k$ we can choose

$$x_1^k, \ldots, x_{n+1}^k \in S \qquad \text{and} \qquad \alpha^k = (\alpha_1^k, \ldots, \alpha_{n+1}^k) \in \Delta_{n+1}$$

such that $x^k = \sum_{i=1}^{n+1} \alpha_i^k x_i^k$. Note that $\Delta_{n+1}$ is compact. If $S$ is compact, we can extract a subsequence as many times as necessary (not more than $n+2$ times) so that $\{\alpha^k\}$ and each $\{x_i^k\}$ converge: we end up with an index set $K \subset \mathbb{N}$ such that, when $k \to +\infty$,

$$\{x_i^k\}_{k \in K} \to x_i \in S \qquad \text{and} \qquad \{\alpha^k\}_{k \in K} \to \alpha \in \Delta_{n+1}.$$

Passing to the limit for $k \in K$, we see that $\{x^k\}_{k \in K}$ converges to a point $x$, which can be expressed as a convex combination of points of $S$: $x \in \operatorname{co} S$, whose compactness is thus established. $\square$

**Definition 1.1.4.** A *conical combination* of elements $x_1, \dots, x_k$ is an element of the form $\sum_{i=1}^k \alpha_i x_i$, where the coefficients $\alpha_i$ are nonnegative.

The set of all conical combinations from a given nonempty $S \subset \mathbb{R}^n$ is the *conical hull* of $S$. It is denoted by $\operatorname{cone} S$.

**Definition 1.1.5.** The *closed conical hull* (or rather closed convex conical hull) of a nonempty set $S \subset \mathbb{R}^n$ is

$$\overline{\operatorname{cone} S} := \operatorname{cl}\big(\operatorname{cone} S\big) = \overline{\Big\{\sum_{i=1}^{k} \alpha_i x_i : \ \alpha_i \geq 0, \ x_i \in S \ \text{ for } i = 1, \dots, k\Big\}}.$$

**Proposition 1.1.8.** *Let $S$ be a nonempty compact set such that $0 \notin \operatorname{co} S$. Then*

$$\overline{\operatorname{cone} S} = \mathbb{R}^+(\operatorname{co} S) \quad [= \operatorname{cone} S].$$

*Proof.* The set $C := \operatorname{co} S$ is compact and does not contain the origin; we prove that $\mathbb{R}^+ C$ is closed. Let $\{t_k x_k\} \subset \mathbb{R}^+ C$ converge to $y$; extracting a subsequence if necessary, we may suppose $x_k \to x \in C$; note: $x \neq 0$. We write

$$t_k \frac{x_k}{\|x_k\|} \longrightarrow \frac{y}{\|x\|},$$

which implies $t_k \to \|y\|/\|x\| =: t \geq 0$. Then, $t_k x_k \to tx = y$, which is thus in $\mathbb{R}^+ C$. $\square$

## 1.2 Convex Sets attached to a Convex Set

### 1.2.1 Relative interior

**Definition 1.2.1.** The *relative interior* $\operatorname{ri} C$ (or $\operatorname{relint} C$) of a convex set $C \subset \mathbb{R}^n$ is the interior of $C$ for the topology relative to the affine hull of $C$. In other words: $x \in \operatorname{ri} C$ if and only if

$$x \in \operatorname{aff} C \quad \text{and} \quad \exists \delta > 0 \text{ such that } (\operatorname{aff} C) \cap B(x, \delta) \subset C.$$

The *dimension* of a convex set $C$ is the dimension of its affine hull, that is to say the dimension of the subspace parallel to $\operatorname{aff} C$.

**Theorem 1.2.1.** *If $C \neq \varnothing$, then $\operatorname{ri} C \neq \varnothing$. In fact,* $\dim(\operatorname{ri} C) = \dim C$.

*Proof.* Let $k := 1 + \dim C$. Since $\operatorname{aff} C$ has dimension $k-1$, $C$ contains $k$ elements affinely independent $x_1, \dots, x_k$. Call $\Delta := \operatorname{co}\{x_1, \dots, x_k\}$ the simplex that they generate; see fig. 2.1.1; $\operatorname{aff} \Delta = \operatorname{aff} C$ because $\Delta \subset C$ and $\dim \Delta = k - 1$. The proof will be finished if we show that $\Delta$ has nonempty relative interior.

Take $\bar{x} := 1/k \sum_{i=1}^k x_i$ (the "center" of $\Delta$) and describe $\operatorname{aff} \Delta$ by points of the form

$$\bar{x} + y = \bar{x} + \sum_{i=1}^{k} \alpha_i(y) x_i = \sum_{i=1}^{k} \Big[\tfrac{1}{k} + \alpha_i(y)\Big] x_i,$$

where $\alpha(y) = (\alpha_1(y), \ldots, \alpha_k(y)) \in \mathbb{R}^k$ solves

$$\sum_{i=1}^{k} \alpha_i x_i = y, \qquad \sum_{i=1}^{k} \alpha_i = 0.$$

Because this system has a unique solution, the mapping $y \mapsto \alpha(y)$ is (linear and) continuous: we can find $\delta > 0$ such that $\|y\| \leq \delta$ implies

$$|\alpha_i(y)| \leq 1/k \quad \text{for } i = 1, \ldots, k,$$

hence $\bar{x} + y \in \Delta$.

In other words, $\bar{x} \in \operatorname{ri} \Delta \subset \operatorname{ri} C$.

It follows in particular $\dim \operatorname{ri} C = \dim \Delta = \dim C$. □

**Lemma 1.2.1.** *Let $x \in \operatorname{cl} C$ and $x' \in \operatorname{ri} C$. Then the half-open segment*

$$]x, x'] = \{\alpha x + (1-\alpha)x' : \ 0 \leq \alpha < 1\}$$

*is contained in* $\operatorname{ri} C$*.*

*Proof.* Take $x'' = \alpha x + (1-\alpha)x'$, with $1 > \alpha \geq 0$. To avoid writing "$\operatorname{ri} C$" every time, we assume without loss of generality that $\operatorname{aff} C = \mathbb{R}^n$.

Since $x \in \operatorname{cl} C$, for all $\varepsilon > 0$, $x \in C + B(0, \varepsilon)$ and we can write

$$B(x'', \varepsilon) = \alpha x + (1-\alpha)x' + B(0, \varepsilon) = \alpha C + (1-\alpha)x' + (1+\alpha)B(0, \varepsilon) = \alpha C + (1-\alpha)\{x' + B\big(0, \tfrac{1+\alpha}{1-\alpha}\varepsilon\big)\}.$$

Since $x' \in \operatorname{int} C$, we can choose $\varepsilon$ so small that $x' + B\big(0, \frac{1+\alpha}{1-\alpha}\varepsilon\big) \subset C$. Then we have

$$B(x'', \varepsilon) \subset \alpha C + (1-\alpha)C = C$$

(where the last equality is just the definition of a convex set). □

**Proposition 1.2.1.** *The three convex sets* $\operatorname{ri} C$*,* $C$ *and* $\operatorname{cl} C$ *have the same affine hull (and hence the same dimension), the same relative interior and the same closure (and hence the same relative boundary).*

*Proof.* The case of the affine hull was already seen in Theorem 2.1.3. For the others, the key result is Lemma 2.1.6 (as well as for most other properties involving closures and relative interiors). We illustrate it by restricting our proof to one of the properties, say: $\operatorname{ri} C$ and $C$ have the same closure.

Thus, we have to prove that $\operatorname{cl} C \subset \operatorname{cl}(\operatorname{ri} C)$. Let $x \in \operatorname{cl} C$ and take $x' \in \operatorname{ri} C$ (it is possible by virtue of Theorem 2.1.3). Because $[x, x'] \subset \operatorname{ri} C$ (Lemma 2.1.6), we do have that $x$ is a limit of points in $\operatorname{ri} C$ (and even a "radial" limit); hence $x$ is in the closure of $\operatorname{ri} C$. □

**Proposition 1.2.2.** *Let the two convex sets $C_1$ and $C_2$ satisfy* $\operatorname{ri} C_1 \cap \operatorname{ri} C_2 \neq \varnothing$*. Then*

$$\operatorname{ri}(C_1 \cap C_2) = \operatorname{ri} C_1 \cap \operatorname{ri} C_2 \tag{1.1}$$

$$\operatorname{cl}(C_1 \cap C_2) = \operatorname{cl} C_1 \cap \operatorname{cl} C_2. \tag{1.2}$$

*Proof.* First we show that $\operatorname{cl} C_1 \cap \operatorname{cl} C_2 \subset \operatorname{cl}(C_1 \cap C_2)$ (the converse inclusion is always true). Given $x \in \operatorname{cl} C_1 \cap \operatorname{cl} C_2$, we pick $x'$ in the nonempty $\operatorname{ri} C_1 \cap \operatorname{ri} C_2$. From Lemma 2.1.6 applied to $C_1$ and to $C_2$,

$$]x, x'[ \subset \operatorname{ri} C_1 \cap \operatorname{ri} C_2.$$

Taking the closure of both sides, we conclude

$$x \in \operatorname{cl}(\operatorname{ri} C_1 \cap \operatorname{ri} C_2) \subset \operatorname{cl}(C_1 \cap C_2),$$

which proves (1.2) because $x$ was arbitrary; the above inclusion is actually an equality.

Now, we have just seen that the two convex sets $\operatorname{ri} C_1 \cap \operatorname{ri} C_2$ and $C_1 \cap C_2$ have the same closure. According to Remark 2.1.9, they have the same relative interior:

$$\operatorname{ri}(C_1 \cap C_2) = \operatorname{ri}(\operatorname{ri} C_1 \cap \operatorname{ri} C_2) \subset \operatorname{ri} C_1 \cap \operatorname{ri} C_2.$$

It remains to prove the converse inclusion, so let $y \in \operatorname{ri} C_1 \cap \operatorname{ri} C_2$. If we take $x' \in C_1$ [resp. $C_2$], the segment $[x', y]$ is in $\operatorname{aff} C_1$ [resp. $\operatorname{aff} C_2$] and, by definition of the relative interior, this segment can be stretched beyond $y$ and yet stay in $C_1$ [resp. $C_2$] (see Fig. 2.1.3). Take in particular $x' \in \operatorname{ri}(C_1 \cap C_2)$, $x' \neq y$ (if such an $x'$ does not exist, we are done). The above stretching singles out an $x \in C_1 \cap C_2$ such that $y \in ]x, x'[$:

$$y = \alpha x + (1 - \alpha)x' \quad \text{for some } \alpha \in ]0, 1[.$$

Then Lemma 2.1.6 applied to $C_1 \cap C_2$ tells us that $y \in \operatorname{ri}(C_1 \cap C_2)$. □

**Proposition 1.2.3.** *For $i = 1, \dots, k$, let $C_i \subset \mathbb{R}^{n_i}$ be convex sets. Then*

$$\operatorname{ri}(C_1 \times \cdots \times C_k) = (\operatorname{ri} C_1) \times \cdots \times (\operatorname{ri} C_k).$$

*Proof.* It suffices to apply Definition 2.1.1 alone, observing that

$$\operatorname{aff}(C_1 \times \cdots \times C_k) = (\operatorname{aff} C_1) \times \cdots \times (\operatorname{aff} C_k).$$

□

**Proposition 1.2.4.** *Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be an affine mapping and $C$ a convex set of $\mathbb{R}^n$. Then*

$$\operatorname{ri}[A(C)] = A(\operatorname{ri} C). \tag{2.1.3}$$

*If $D$ is a convex set of $\mathbb{R}^m$ satisfying $A^{-1}(\operatorname{ri} D) \neq \varnothing$, then*

$$\operatorname{ri}\big[A^{-1}(D)\big] = A^{-1}(\operatorname{ri} D). \tag{2.1.4}$$

*Proof.* First, note that the continuity of $A$ implies $A(\operatorname{cl} S) \subset \operatorname{cl}[A(S)]$ for any $S \subset \mathbb{R}^n$. Apply this result to $\operatorname{ri} C$, whose closure is $\operatorname{cl} C$ (Proposition 2.1.8), and use the monotonicity of the closure operation:

$$A(C) \subset A(\operatorname{cl} C) = A[\operatorname{cl}(\operatorname{ri} C)] \subset \operatorname{cl}[A(\operatorname{ri} C)] \subset \operatorname{cl}[A(C)];$$

the closed set $\mathrm{cl}[A(\mathrm{ri}\,C)]$ is therefore $\mathrm{cl}[A(C)]$. Because $A(\mathrm{ri}\,C)$ and $A(C)$ have the same closure, they have the same relative interior (Remark 2.1.9):

$$\mathrm{ri}\ A(C) = \mathrm{ri}[A(\mathrm{ri}\,C)] \subset A(\mathrm{ri}\,C).$$

To prove the converse inclusion, let $w = A(y) \in A(\mathrm{ri}\,C)$, with $y \in \mathrm{ri}\,C$. We choose $z' = A(x') \in \mathrm{ri}\,A(C)$, with $x' \in C$ (we assume $z' \neq w$, hence $x' \neq y$).

Using in $C$ the same stretching mechanism as in Fig. 2.1.3, we single out $x \in C$ such that $y \in ]x, x'[$, to which corresponds $z = A(x) \in A(C)$. By affinity, $A(y) \in ]A(x), A(x')[ = ]z, z'[$. Thus, $z$ and $z'$ fulfill the conditions of Lemma 2.1.6 applied to the convex set $A(C)$: $w \in \mathrm{ri}\,A(C)$, and (2.1.3) is proved.

The proof of (2.1.4) uses the same technique. $\square$

### 1.2.2 The asymptotic cone

**Proposition 1.2.5.** *The closed convex cone $C_\infty(x)$ does not depend on $x \in C$.*

*Proof.* See Theorem I.2.3.1 and the pantographic Figure I.2.3.1. Take two different points $x_1$ and $x_2$ in $C$; it suffices to prove one inclusion, say $C_\infty(x_1) \subset C_\infty(x_2)$. Let $d \in C_\infty(x_1)$ and $t > 0$, we have to prove $x_2 + td \in C$. With $\varepsilon \in ]0, 1[$, consider the point

$$\bar{x}_\varepsilon := x_1 + td + (1 - \varepsilon)(x_2 - x_1).$$

Writing it as

$$\bar{x}_\varepsilon = \varepsilon\big(x_1 + \tfrac{t}{\varepsilon}d\big) + (1 - \varepsilon)x_2,$$

we see that $\bar{x}_\varepsilon \in C$ (use the definitions of $C_\infty(x_1)$ and of a convex set). On the other hand,

$$x_2 + td = \lim_{\varepsilon \downarrow 0} \bar{x}_\varepsilon \in \overline{C}.$$

$\square$

**Definition 1.2.2.** The asymptotic cone, or recession cone of the closed convex set $C$ is the closed convex cone $C_\infty$ defined by (2.2.1) or (2.2.2), in which Proposition 2.2.1 is exploited.

**Proposition 1.2.6.** *A closed convex set $C$ is compact if and only if $C_\infty = \{0\}$.*

*Proof.* If $C$ is bounded, it is clear that $C_\infty$ cannot contain any nonzero direction.

Conversely, let $\{x_k\} \subset C$ be such that $\|x_k\| \to +\infty$ (we assume $x_k \neq 0$). The sequence $\{d_k := x_k/\|x_k\|\}$ is bounded, extract a convergent subsequence: $d = \lim_{k \in K} d_k$ with $K \subset \mathbb{N}$ ($\|d\| = 1$). Now, given $x \in C$ and $t > 0$, take $k$ so large that $\|x_k\| \geq t$. Then, we see that

$$x + td = \lim_{k \in K} \Big[(1 - \tfrac{t}{\|x_k\|})x + \tfrac{t}{\|x_k\|}x_k\Big]$$

is in the closed convex set $C$, hence $d \in C_\infty$. $\square$

**Proposition 1.2.7.** *Proposition 2.2.5*

- If $\{C_j\}_{j \in J}$ is a family of closed convex sets having a point in common, then
$$\left(\bigcap_{j \in J} C_j\right)_\infty = \bigcap_{j \in J} (C_j)_\infty.$$
- If, for $j = 1, \ldots, m$, $C_j$ are closed convex sets in $\mathbb{R}^{n_j}$, then
$$(C_1 \times \cdots \times C_m)_\infty = (C_1)_\infty \times \cdots \times (C_m)_\infty.$$
- Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be an affine mapping. If $C$ is closed convex in $\mathbb{R}^n$ and $A(C)$ is closed, then
$$A(C_\infty) \subset [A(C)]_\infty.$$
- If $D$ is closed convex in $\mathbb{R}^m$ with nonempty inverse image, then
$$\left[\ A^{-1}(D)\ \right]_\infty = A^{-1}(D_\infty).$$

### 1.2.3 Extreme points

**Definition 1.2.3.** We say that $x \in C$ is an *extreme point* of $C$ if there are no two different points $x_1$ and $x_2$ in $C$ such that $x = \frac{1}{2}(x_1 + x_2)$.

**Proposition 1.2.8.** *If $C$ is compact, then* $\operatorname{ext} C \neq \varnothing$.

*Proof.* Because $C$ is compact, there is $\bar{x} \in C$ maximizing the continuous function $x \mapsto \|x\|^2$. We claim that $\bar{x}$ is extremal. In fact, suppose that there are $x_1$ and $x_2$ in $C$ with $\bar{x} = \frac{1}{2}(x_1 + x_2)$. Then, with $x_1 \neq x_2$ and using (2.3.1), we obtain the contradiction
$$\|\bar{x}\|^2 = \left\|\tfrac{1}{2}(x_1 + x_2)\right\|^2 < \tfrac{1}{2}\left(\|x_1\|^2 + \|x_2\|^2\right) \leq \tfrac{1}{2}\left(\|\bar{x}\|^2 + \|\bar{x}\|^2\right) = \|\bar{x}\|^2.$$
□

**Theorem 1.2.2.** *(H. Minkowski) Let $C$ be compact, convex in $\mathbb{R}^n$. Then $C$ is the convex hull of its extreme points: $C = \operatorname{co}(\operatorname{ext} C)$.*

**Definition 1.2.4.** A nonempty convex subset $F \subset C$ is a *face* of $C$ if it satisfies the following property: every segment of $C$, having in its relative interior an element of $F$, is entirely contained in $F$. In other words,
$$\left.\begin{array}{r}(x_1, x_2) \in C \times C \\ \exists \alpha \in ]0,1[:\ \alpha x_1 + (1-\alpha)x_2 \in F\end{array}\right\} \Longrightarrow\ [x_1, x_2] \subset F. \tag{2.3.2}$$

**Proposition 1.2.9.** *Let $F$ be a face of $C$. Then any extreme point of $F$ is an extreme point of $C$.*

*Proof.* Take $x \in F \subset C$ and assume that $x$ is not an extreme point of $C$: there are different $x_1, x_2$ in $C$ and $\alpha \in ]0,1[$ such that $x = \alpha x_1 + (1-\alpha)x_2 \in F$. From the very definition (2.3.2) of a face, this implies that $x_1$ and $x_2$ are in $F$: $x$ cannot be an extreme point of $F$. □ □

### 1.2.4 Exposed faces

**Definition 1.2.5** (Supporting Hyperplane)**.** An affine hyperplane $H_{s,r}$ is said to *support* the set $C$ when $C$ is entirely contained in one of the two closed half-spaces delimited by $H_{s,r}$: say

$$\langle s, y\rangle \leq r \qquad \text{for all } y \in C. \tag{2.4.1}$$

It is said to support $C$ at $x \in C$ when, in addition, $x \in H_{s,r}$; (2.4.1) holds, as well as

$$\langle s, x\rangle = r.$$

**Definition 1.2.6.** The set $F \subset C$ is an *exposed face* of $C$ if there is a supporting hyperplane $H_{s,r}$ of $C$ such that $F = C \cap H_{s,r}$.

An *exposed point*, or *vertex*, is a 0-dimensional exposed face, i.e. a point $x \in C$ at which there is a supporting hyperplane $H_{s,r}$ of $C$ such that $H_{s,r} \cap C$ reduces to $\{x\}$.

**Proposition 1.2.10.** *Proposition 2.4.3 An exposed face is a face.*

*Proof.* Let $F$ be an exposed face, with its associated support $H_{s,r}$. Take $x_1$ and $x_2$ in $C$:

$$\langle s, x_i\rangle \leq r \quad \text{for } i = 1, 2; \tag{1.3}$$

take also $\alpha \in ]0, 1[$ such that $\alpha x_1 + (1-\alpha)x_2 \in F \subset H_{s,r}$:

$$\langle s, \alpha x_1 + (1-\alpha)x_2\rangle = r.$$

Suppose that one of the relations (1.3) holds as strict inequality. By convex combination, we obtain $(0 < \alpha < 1!)$

$$\langle s, \alpha x_1 + (1-\alpha)x_2\rangle < r,$$

a contradiction. □

**Proposition 1.2.11.** *Let $C$ be convex and compact. For $s \in \mathbb{R}^n$, there holds*

$$\max_{x\in C}\langle s, x\rangle = \max_{x\in \operatorname{ext} C}\langle s, x\rangle.$$

Furthermore, the solution-set of the first problem is the convex hull of the solution-set of the second:

$$\operatorname{Argmax}_{x\in C}\langle s, x\rangle = \operatorname{co}\{\operatorname{Argmax}_{x\in \operatorname{ext} C}\langle s, x\rangle\}.$$

*Proof.* Because $C$ is compact, $\langle s, \cdot\rangle$ attains its maximum on $F_C(s)$. The latter set is convex and compact, and as such is the convex hull of its extreme points (Minkowski's Theorem 2.3.4); these extreme points are also extreme in $C$ (Proposition 2.3.7 and Remark 2.4.4). □

## 1.3 Projection Onto Closed Convex Set

### 1.3.1 The projection operator

**Theorem 1.3.1.** *A point $y_x \in C$ is the projection $p_C(x)$ if and only if*

$$\langle x - y_x,\ y - y_x\rangle \leq 0 \qquad \text{for all } y \in C. \tag{3.1.3}$$

*Proof.* Call $y_x$ the solution of (3.1.1); take $y$ arbitrary in $C$, so that $y_x + \alpha(y - y_x) \in C$ for any $\alpha \in ]0,1[$. Then we can write with the notation (3.1.2)

$$f_x(y_x) \leq f_x\big(y_x + \alpha(y - y_x)\big) = \tfrac{1}{2}\|y_x - x + \alpha(y - y_x)\|^2.$$

Developing the square, we obtain after simplification

$$0 \leq \alpha\langle y_x - x,\ y - y_x\rangle + \tfrac{1}{2}\alpha^2\|y - y_x\|^2.$$

Divide by $\alpha$ ($> 0$) and let $\alpha \downarrow 0$ to obtain (3.1.3).

Conversely, suppose that $y_x \in C$ satisfies (3.1.3). If $y_x = x$, then $y_x$ certainly solves (3.1.1). If not, write for arbitrary $y \in C$:

$$0 \geq \langle x - y_x,\ y - y_x\rangle = (x - y_x,\ y - x + x - y_x)$$

$$= \|x - y_x\|^2 + \langle x - y_x,\ y - x\rangle \geq \|x - y_x\|^2 - \|x - y\|\,\|x - y_x\|,$$

where the Cauchy-Schwarz inequality is used. Divide by $\|x - y_x\| > 0$ to see that $y_x$ solves (3.1.1). □

**Proposition 1.3.1.** *For all $(x_1, x_2) \in \mathbb{R}^n \times \mathbb{R}^n$, there holds*

$$\|p_C(x_1) - p_C(x_2)\|^2 \leq \langle p_C(x_1) - p_C(x_2),\ x_1 - x_2\rangle.$$

*Proof.* Write (3.1.3) with $x = x_1$, $y = p_C(x_2) \in C$:

$$\langle p_C(x_2) - p_C(x_1),\ x_1 - p_C(x_1)\rangle \leq 0;$$

likewise,

$$\langle p_C(x_1) - p_C(x_2),\ x_2 - p_C(x_2)\rangle \leq 0,$$

and conclude by addition

$$\langle p_C(x_1) - p_C(x_2),\ x_2 - x_1 + p_C(x_1) - p_C(x_2)\rangle \leq 0.$$

□

### 1.3.2 Projection onto a closed convex cone

**Definition 1.3.1.** Let $K$ be a convex cone, as defined in Example 1.1.4. The *polar cone* of $K$ (called negative polar cone by some authors) is:

$$K^\circ := \{s \in \mathbb{R}^n : \langle s, x\rangle \leq 0 \text{ for all } x \in K\}.$$

**Proposition 1.3.2.** *Let $K$ be a closed convex cone. Then $y_x = p_K(x)$ if and only if*

$$y_x \in K, \qquad x - y_x \in K^\circ, \qquad \langle x - y_x, y_x\rangle = 0. \tag{3.2.1}$$

*Proof.* We know from Theorem 3.1.1 that $y_x = p_K(x)$ satisfies

$$\langle x - y_x, y - y_x\rangle \leq 0 \qquad \text{for all } y \in K. \tag{3.2.2}$$

Taking $y = \alpha y_x$, with arbitrary $\alpha \geq 0$, this inequality implies

$$(\alpha - 1)\langle x - y_x, y_x\rangle \leq 0 \qquad \text{for all } \alpha \geq 0.$$

Since $\alpha - 1$ can have either sign, this implies $\langle x - y_x, y_x\rangle = 0$ and (3.2.2) becomes

$$\langle y, x - y_x\rangle \leq 0 \quad \text{for all } y \in K, \qquad \text{i.e.} \quad x - y_x \in K^\circ.$$

Conversely, let $y_x$ satisfy (3.2.1). For arbitrary $y \in K$, use the notation (3.1.2):

$$f_x(y) = \tfrac{1}{2}\|x - y_x + y_x - y\|^2 \geq f_x(y_x) + \langle x - y_x, y_x - y\rangle;$$

but (3.2.1) shows that

$$\langle x - y_x, y_x - y\rangle = -\langle x - y_x, y\rangle \geq 0,$$

hence $f_x(y) \geq f_x(y_x)$: $y_x$ solves (3.1.1). □

**Theorem 1.3.2** (J.-J. Moreau)**.** *Let $K$ be a closed convex cone. For the three elements $x, x_1$ and $x_2$ in $\mathbb{R}^n$, the properties below are equivalent:*

*(i)* $x = x_1 + x_2$ *with* $x_1 \in K$, $x_2 \in K^\circ$ *and* $\langle x_1, x_2\rangle = 0$;

*(ii)* $x_1 = \mathrm{p}_K(x)$ *and* $x_2 = \mathrm{p}_{K^\circ}(x)$.

*Proof.* Straightforward, from (3.2.3) and the characterization (3.2.1) of $x_1 = \mathrm{p}_K(x)$. □

## 1.4 Separation and Applications

### 1.4.1 Separation between convex sets

**Theorem 1.4.1.** *Let $C \subset \mathbb{R}^n$ be nonempty closed convex, and let $x \notin C$. Then there exists $s \in \mathbb{R}^n$ such that*

$$\langle s, x\rangle > \sup_{y \in C}\langle s, y\rangle. \tag{4.1.1}$$

*Proof.* Set $s := x - p_C(x) \neq 0$. We write (3.1.3) as

$$0 \geq (s, y - x + s) = (s, y) - (s, x) + \|s\|^2.$$

Thus we have

$$(s, x) - \|s\|^2 \geq (s, y) \qquad \text{for all } y \in C,$$

and our $s$ is a convenient answer for (4.1.1). $\square$

**Corollary 1.4.1** (Strict Separation of Convex Sets)**.** *Let $C_1$, $C_2$ be two nonempty closed convex sets with $C_1 \cap C_2 = \varnothing$. If $C_2$ is bounded, there exists $s \in \mathbb{R}^n$ such that*

$$\sup_{y \in C_1} \langle s, y \rangle \; < \; \min_{y \in C_2} \langle s, y \rangle. \tag{4.1.2}$$

PROOF. The set $C_1 - C_2$ is convex (Proposition 1.2.4) and closed (because $C_2$ is compact). To say that $C_1$ and $C_2$ are disjoint is to say that $0 \notin C_1 - C_2$, so we have by Theorem 4.1.1 an $s \in \mathbb{R}^n$ separating $\{0\}$ from $C_1 - C_2$:

$$\sup\{\langle s, y \rangle : y \in C_1 - C_2\} < \langle s, 0 \rangle = 0.$$

This means:

$$0 \; > \; \sup_{y_1 \in C_1} \langle s, y_1 \rangle + \sup_{y_2 \in C_2} \langle s, -y_2 \rangle = \sup_{y_1 \in C_1} \langle s, y_1 \rangle - \inf_{y_2 \in C_2} \langle s, y_2 \rangle.$$

Because $C_2$ is bounded, the last infimum (is a min and) is finite and can be moved to the left-hand side. $\square$

**Theorem 1.4.2.** *(Proper Separation of Convex Sets) If the two nonempty convex sets $C_1$ and $C_2$ satisfy $\operatorname{ri} C_1 \cap (\operatorname{ri} C_2) = \varnothing$, they can be properly separated.*

### 1.4.2 First consequences of the separation properties

**Lemma 1.4.1.** *Let $x \in \partial C$, where $C \neq \varnothing$ is convex in $\mathbb{R}^n$ (naturally $C \neq \mathbb{R}^n$). There exists a hyperplane supporting $C$ at $x$.*

*Proof.* Because $C$, $\operatorname{cl} C$ and their complements have the same boundary (remember Remark 2.1.9), a sequence $\{x_k\}$ can be found such that

$$x_k \notin C \quad \text{for } k = 1, 2, \ldots \quad \text{and} \quad \lim_{k \to +\infty} x_k = x.$$

For each $k$ we have by Theorem 4.1.1 some $s_k$ with $\|s_k\| = 1$ such that

$$\langle s_k, \; x_k - y \rangle > 0 \quad \text{for all } y \in C \subset \operatorname{cl} C.$$

Extract a subsequence if necessary so that $s_k \to s$ (note: $s \neq 0$) and pass to the limit to obtain

$$\langle s, \; x - y \rangle \geq 0 \quad \text{for all } y \in C,$$

which is the required result $\langle s, x \rangle = r \geq \langle s, y \rangle$ for all $y \in C$. $\square$

**Proposition 1.4.1.** *Let $S \subset \mathbb{R}^n$ and $C := \operatorname{co} S$. Any $x \in C \cap \operatorname{bd} C$ can be represented as a convex combination of $n$ elements of $S$.*

*Proof.* Because $x \in \operatorname{bd} C$, there is a hyperplane $H_{s,r}$ supporting $C$ at $x$: for some $s \neq 0$ and $r \in \mathbb{R}$,

$$\langle s, x\rangle - r = 0 \qquad \text{and} \qquad \langle s, y\rangle - r \leq 0 \ \text{ for all } y \in C. \tag{4.2.1}$$

On the other hand, Carathéodory's Theorem 1.3.6 implies the existence of points $x_1, \dots, x_{n+1}$ in $S$ and convex multipliers $\alpha_1, \dots, \alpha_{n+1}$ such that $x = \sum_{i=1}^{n+1} \alpha_i x_i$; and each $\alpha_i$ can be assumed positive (otherwise the proof is finished).

Setting successively $y = x_i$ in (4.2.1), we obtain by convex combination

$$0 = \langle s, x\rangle - r = \sum_{i=1}^{n+1} \alpha_i \big(\langle s, x_i\rangle - r\big) \leq 0,$$

so each $\langle s, x_i\rangle - r$ is actually 0. Each $x_i$ is therefore not only in $S$, but also in $H_{s,r}$, a set whose dimension is $n-1$. It follows that our starting $x$, which is in $\operatorname{co}\{x_1, \dots, x_{n+1}\}$, can be described as the convex hull of only $n$ among these $x_i$'s. ☐

**Theorem 1.4.3.** *Let $\varnothing \neq C \subseteq \mathbb{R}^n$ be convex. The set $C^*$ defined above is the closure of $C$.*

*Proof.* By construction, $C^* \supseteq \operatorname{cl} C$. Conversely, take $x \notin \operatorname{cl} C$; we can separate $x$ and $\operatorname{cl} C$: there exists $s_0 \neq 0$ such that

$$\langle s_0, x\rangle > \sup_{y \in C} \langle s_0, y\rangle =: r_0.$$

Then $(s_0, r_0) \in \Sigma_C$; but $x \notin H_{s_0, r_0}$, hence $x \notin C^*$. ☐

**Corollary 1.4.2.** *The data $(s_j, r_j) \in \mathbb{R}^n \times \mathbb{R}$ for $j$ in an arbitrary index set $J$ is equivalent to the data of a closed convex set $C$ via the relation*

$$C = \bigcap_{j \in J} \{x \in \mathbb{R}^n : \langle s_j, x\rangle \leq r_j\}.$$

*Proof.* If $C$ is given, define $\{(s_j, r_j)\}_J := \Sigma_C$ as in (4.2.2). If $\{(s_j, r_j)\}_J$ is given, the intersection of the corresponding half-spaces is a closed convex set. Note here that we can define at the same time the whole of $\mathbb{R}^n$ and the empty sets as two extreme cases. ☐

**Definition 1.4.1** (Polyhedral Sets)**.** A *closed convex polyhedron* is an intersection of finitely many half-spaces. Take $(s_1, r_1), \dots, (s_m, r_m)$ in $\mathbb{R}^n \times \mathbb{R}$, with $s_i \neq 0$ for $i = 1, \dots, m$; then define

$$P := \{x \in \mathbb{R}^n : \langle s_j, x\rangle \leq r_j \ \text{ for } j = 1, \dots, m\},$$

or in matrix notations (assuming the dot-product for $\langle \cdot, \cdot \rangle$),

$$P = \{x \in \mathbb{R}^n : Ax \leq b\},$$

if $A$ is the matrix whose rows are $s_j$ and $b \in \mathbb{R}^m$ has coordinates $r_j$.

A *closed convex polyhedral cone* is the special case where $b = 0$.

Proposition 4.2.7 Let $K$ be a convex cone with polar $K^\circ$; then, the polar $K^{\circ\circ}$ of $K^\circ$ is the closure of $K$.

*Proof.* We exploit Remark 4.1.2: due to its conical character ($\alpha x \in K$ if $\alpha \in K$ and $\alpha > 0$), $\operatorname{cl} K$ has a very special support function:

$$\sigma_{\operatorname{cl} K}(s) = \begin{cases} \langle s, 0 \rangle = 0 & \text{if } \langle s, x \rangle \leq 0 \text{ for all } x \in \operatorname{cl} K, \\ +\infty & \text{otherwise.} \end{cases}$$

In other words, $\sigma_{\operatorname{cl} K}$ is 0 on $K^\circ$, $+\infty$ elsewhere. Thus, the characterization

$$x \in \operatorname{cl} K \iff \langle \cdot, x \rangle \leq \sigma_{\operatorname{cl} K}(\cdot)$$

becomes

$$x \in \operatorname{cl} K \iff \begin{cases} \langle s, x \rangle \leq 0 & \text{for all } s \in K^\circ \\ (\langle s, x \rangle \text{ arbitrary}) & \text{for } s \notin K^\circ, \end{cases}$$

in which we recognize the definition of $K^{\circ\circ}$. ☐

### 1.4.3 The lemma of Minkowski and Farkas

**Lemma 1.4.2** (Farkas I)**.** *Let $b, s_1, \ldots, s_m$ be given in $\mathbb{R}^n$. The set*

$$\{x \in \mathbb{R}^n : \ \langle s_j, x \rangle \leq 0 \quad \text{for } j = 1, \ldots, m\} \tag{4.3.1}$$

*is contained in the set*

$$\{x \in \mathbb{R}^n : \ \langle b, x \rangle \leq 0\} \tag{4.3.2}$$

*if and only if (see Definition 1.4.5 of a conical hull)*

$$b \in \operatorname{cone}\{s_1, \ldots, s_m\}. \tag{4.3.3}$$

**Lemma 1.4.3** (Farkas II)**.** *Let $b, s_1, \ldots, s_m$ be given in $\mathbb{R}^n$. Then exactly one of the following statements is true.*

$P :=$ (**??**) *has a solution $\alpha \in \mathbb{R}^n$.*

$Q :=$

$$\begin{cases} \langle b, x \rangle > 0, \\ \langle s_j, x \rangle \leq 0 \quad \text{for } j = 1, \ldots, m \end{cases}$$

*has a solution $x \in \mathbb{R}^n$.*

**Lemma 1.4.4** (Farkas III)**.** *Let* $s_1, \ldots, s_m$ *be given in* $\mathbb{R}^n$*. Then the convex cone*

$$K := \operatorname{cone}\{s_1, \ldots, s_m\} = \Big\{ \sum_{j=1}^{m} \alpha_j s_j : \ \alpha_j \geq 0 \textit{ for } j = 1, \ldots, m \Big\}$$

*is closed.*

*Proof.* It is quite similar to that of Carathéodory's Theorem 1.3.6. First, the proof is easy if the $s_j$'s are linearly independent: then, the convergence of

$$x^k = \sum_{j=1}^{m} \alpha_j^k s_j \quad \text{for } k \to \infty \tag{1.4}$$

is equivalent to the convergence of each $\{\alpha_j^k\}$ to some $\alpha_j$, which must be nonnegative if each $\alpha_j^k$ in (1.4) is nonnegative.

Suppose, on the contrary, that the system $\sum_{j=1}^m \beta_j s_j = 0$ has a nonzero solution $\beta \in \mathbb{R}^m$ and assume $\beta_j < 0$ for some $j$ (change $\beta$ to $-\beta$ if necessary). As in the proof of Theorem 1.3.6, write each $x \in K$ as

$$x = \sum_{j=1}^{m} \alpha_j s_j = \sum_{j=1}^{m} [\alpha_j + t^*(x)\beta_j] s_j = \sum_{j \neq i(x)} \alpha'_j s_j,$$

where

$$i(x) \in \operatorname{Argmin}_{\beta_j < 0} \frac{-\alpha_j}{\beta_j}, \qquad t^*(x) := \frac{-\alpha_{i(x)}}{\beta_{i(x)}},$$

so that each $\alpha'_j = \alpha_j + t^*(x)\beta_j$ is nonnegative. Letting $x$ vary in $K$, we thus construct a decomposition

$$K = \bigcup \{K_i : \ i = 1, \ldots, m\},$$

where $K_i$ is the conical hull of the $m-1$ generators $s_j, \ j \neq i$.

Now, if there is some $i$ such that the generators of $K_i$ are linearly dependent, we repeat the argument for a further decomposition of this $K_i$. After finitely many such operations, we end up with a decomposition of $K$ as a finite union of polyhedral convex cones, each having linearly independent generators. All these cones are therefore closed (first part of the proof), so $K$ is closed as well. $\square$

**Theorem 1.4.4** (Generalized Farkas)**.** *Let be given* $(b, r)$ *and* $(s_j, \rho_j)$ *in* $\mathbb{R}^n \times \mathbb{R}$*, where* $j$ *varies in an (arbitrary) index set* $J$*. Suppose that the system of inequalities*

$$\langle s_j, x \rangle \leq \rho_j \qquad \textit{for all } j \in J \tag{4.3.6}$$

*has a solution* $x \in \mathbb{R}^n$ *(the system is consistent). Then the following two properties are equivalent:*

*(i)* $\langle b, x \rangle \leq r$ *for all* $x$ *satisfying (4.3.6);*

*(ii)* $(b, r)$ *is in the closed convex conical hull of* $S := \{(0,1)\} \cup \{(s_j, \rho_j)\}_{j \in J}$.

*(ii)* ⇒ *(i).* Let first $(b, r)$ be in $K := \operatorname{cone} S$. In other words, there exists a finite set $\{1, \dots, m\} \subset J$ and nonnegative $\alpha_0, \alpha_1, \dots, \alpha_m$ such that (we adopt the convention $\sum \emptyset = 0$)

$$b = \sum_{j=1}^{m} \alpha_j s_j \qquad \text{and} \qquad r = \alpha_0 + \sum_{j=1}^{m} \alpha_j \rho_j.$$

For each $x$ satisfying (4.3.6) we can write

$$\langle b, x \rangle \leq r - \alpha_0 \leq r. \tag{4.3.7}$$

If, now, $(b, r)$ is in the closure of $K$, pass to the limit in (4.3.7) to establish the required conclusion (i) for all $(b, r)$ described by (ii).

[(i) ⇒ (ii)] If $(b, r) \notin \operatorname{cl} K$, separate $(b, r)$ from $\operatorname{cl} K$: equipping $\mathbb{R}^n \times \mathbb{R}$ with the scalar product

$$\langle\langle (b, r), (d, t) \rangle\rangle := \langle b, d \rangle + rt,$$

there exists $(d, -t) \in \mathbb{R}^n \times \mathbb{R}$ such that

$$\sup_{(s,\rho) \in K} [\langle s, d \rangle - \rho t] < \langle b, d \rangle - rt. \tag{4.3.8}$$

It follows first that the left-hand supremum is a finite number $\kappa$. Then the conical character of $K$ implies $\kappa \leq 0$, because $\alpha \kappa \leq \kappa$ for all $\alpha > 0$; actually $\kappa = 0$ because $(0, 0) \in K$. In summary, we have singled out $(d, t) \in \mathbb{R}^n \times \mathbb{R}$ such that

$$t \geq 0 \qquad [\text{take } (0, 1) \in K]$$

$$(*) \qquad \langle s_j, d \rangle - \rho_j t \leq 0 \;\; \text{for all } j \in J \qquad [\text{take } (s_j, \rho_j) \in K]$$

$$(**) \qquad \langle b, d \rangle - rt > 0. \qquad [\text{don't forget (4.3.8)}]$$

Now consider two cases:

- If $t > 0$, divide $(*)$ and $(**)$ by $t$ to exhibit the point $x = d/t$ violating (i).
- If $t = 0$, take $x_0$ satisfying (4.3.6). Observe from $(*)$ that, for all $\alpha > 0$, the point $x(\alpha) = x_0 + \alpha d$ satisfies (4.3.6) as well. Yet, let $\alpha \to +\infty$ in

  $$\langle b, x(\alpha) \rangle = \langle b, x_0 \rangle + \alpha \langle b, d \rangle$$

  to realize from $(**)$ that $x(\alpha)$ violates (i) if $\alpha$ is large enough.

Thus we have proved in both cases that ”not (ii) ⇒ not (i)”. □

## 1.5 Conical approximations of convex sets

### 1.5.1 Convenient definitions of tangent cones

**Definition 1.5.1.** Let $S \subset \mathbb{R}^n$ be nonempty. We say that $d \in \mathbb{R}^n$ is a direction *tangent* to $S$ at $x \in S$ when there exists a sequence $\{x_k\} \subset S$ and a sequence $\{t_k\}$ such that, when $k \to +\infty$,

$$x_k \to x, \qquad t_k \downarrow 0, \qquad \frac{x_k - x}{t_k} \to d. \tag{5.1.1}$$

The set of all such directions is called the *tangent cone* (also called the contingent cone, or Bouligand's cone) to $S$ at $x \in S$, denoted by $T_S(x)$.

**Proposition 1.5.1.** *A direction $d$ is tangent to $S$ at $x \in S$ if and only if*

$$\exists (d_k) \to d,\ \exists (t_k) \downarrow 0 \quad \textit{such that} \quad x + t_k d_k \in S \textit{ for all } k.$$

**Proposition 1.5.2.** *The tangent cone is closed.*

*Proof.* Let $\{d_\ell\} \subset T_S(x)$ be converging to $d$; for each $\ell$ take sequences $\{x_{\ell,k}\}_k$ and $\{t_{\ell,k}\}_k$ associated with $d_\ell$ in the sense of Definition 5.1.1. Fix $\ell > 0$: we can find $k_\ell$ such that

$$\left\| \frac{x_{\ell,k_\ell} - x}{t_{\ell,k_\ell}} - d_\ell \right\| \leq \frac{1}{\ell}.$$

Letting $\ell \to \infty$, we then obtain the sequences $\{x_\ell, t_\ell\}_\ell$ and $\{t_\ell, k_\ell\}_\ell$ which define $d$ as an element of $T_S(x)$. ☐

### 1.5.2 The tangent cones and normal cones to a convex set

**Proposition 1.5.3.** *The tangent cone to $C$ at $x$ is the closure of the cone generated by $C - \{x\}$:*

$$\begin{aligned} T_C(x) &= \overline{\text{cone}}(C - x) = \overline{\text{cl}}\mathbb{R}^+(C - x) \\ &= \overline{\{d \in \mathbb{R}^n : \ d = \alpha(y - x),\ y \in C,\ \alpha \geq 0\}}. \end{aligned} \tag{5.2.1}$$

*Proof.* We have just said that $C - \{x\} \subset T_C(x)$. Because $T_C(x)$ is a closed cone (Proposition 5.1.3), it immediately follows that $\overline{\mathbb{R}^+}(C - x) \subset T_C(x)$. Conversely, for $d \in T_C(x)$, take $\{x_k\}$ and $\{t_k\}$ as in the definition (5.1.1): the point $(x_k - x)/t_k$ is in $\mathbb{R}^+(C - x)$, hence its limit $d$ is in the closure of this latter set. ☐

**Definition 1.5.2.** The direction $s \in \mathbb{R}^n$ is said *normal* to $C$ at $x \in C$ when

$$\langle s, y - x \rangle \leq 0 \quad \text{for all } y \in C. \tag{5.2.2}$$

The set of all such directions is called *normal cone* to $C$ at $x$, denoted by $N_C(x)$.

**Proposition 1.5.4.** *The normal cone is the polar of the tangent cone.*

*Proof.* If $\langle s, d\rangle \le 0$ for all $d \in C - x$, the same holds for all $d \in \mathbb{R}^+(C - x)$, as well as for all $d$ in the closure $T_C(x)$ of the latter. Thus, $N_C(x) \subset [T_C(x)]^\circ$.

Conversely, take $s$ arbitrary in $[T_C(x)]^\circ$. The relation $\langle s, d\rangle \le 0$, which holds for all $d \in T_C(x)$, a fortiori holds for all $d \in C - x \subset T_C(x)$; this is just (5.2.2). □

**Corollary 1.5.1.** *The tangent cone is the polar of the normal cone:*

$$T_C(x) = \{d \in \mathbb{R}^n : \ \langle s, d\rangle \le 0 \text{ for all } s \in N_C(x)\}.$$

### 1.5.3 Some properties of tangent and normal cones

**Proposition 1.5.5.** *5.3.1 Here, the $C$'s are nonempty closed convex sets.*

*[(i)]*

1. *For $x \in C_1 \cap C_2$, there holds*

   $$T_{C_1 \cap C_2}(x) \subset T_{C_1}(x) \cap T_{C_2}(x) \qquad \text{and} \qquad N_{C_1 \cap C_2}(x) \supset N_{C_1}(x) + N_{C_2}(x).$$

2. *With $C_i \subset \mathbb{R}^{n_i}$, $i = 1, 2$ and $(x_1, x_2) \in C_1 \times C_2$,*

   $$T_{C_1 \times C_2}(x_1, x_2) = T_{C_1}(x_1) \times T_{C_2}(x_2), \qquad N_{C_1 \times C_2}(x_1, x_2) = N_{C_1}(x_1) \times N_{C_2}(x_2).$$

3. *With an affine mapping $A(x) = y_0 + A_0 x$ ($A_0$ linear) and $x \in C$,*

   $$T_{A(C)}[A(x)] = \mathrm{cl}[A_0 T_C(x)] \qquad \text{and} \qquad N_{A(C)}[A(x)] = A_0^{-*}[N_C(x)].$$

4. *In particular (start from (ii), (iii) and proceed as when proving (1.2.2)):*

   $$T_{C_1+C_2}(x_1+x_2) = \mathrm{cl}[T_{C_1}(x_1)+T_{C_2}(x_2)], \qquad N_{C_1+C_2}(x_1+x_2) = N_{C_1}(x_1) \cap N_{C_2}(x_2).$$

**Proposition 1.5.6.** *For $x \in C$ and $s \in \mathbb{R}^n$, the following properties are equivalent:*

*(i)* $s \in N_C(x)$;

*(ii)* $x$ *is in the exposed face* $F_C(s)$: $\langle s, x\rangle = \max_{y \in C}\langle s, y\rangle$;

*(iii)* $x = p_C(x + s)$.

*Proof.* Nothing really new: everything comes from the definitions of normal cones, supporting hyperplanes, exposed faces, and the characteristic property (3.1.3) of the projection operator. □

**Proposition 1.5.7.** *For given $x \in C$ and $d \in \mathbb{R}^n$, there holds*

$$\lim_{t \downarrow 0} \frac{\mathrm{P}_C(x + td) - x}{t} = \mathrm{P}_{T_C(x)}(d). \qquad (5.3.3)$$

*Proof.* HINT. Start from the characterization (3.1.3) of a projection, to observe that the difference quotient $[\mathrm{P}_C(x+td)-x]/t$ is the projection of $d$ onto $(C-x)/t$. Then let $t \downarrow 0$; the result comes as well with the help of (5.1.4) and Remark 5.2.2. □

# Chapter 2

# Convex functions

## 2.1 Basic Definitions and Examples

### 2.1.1 The definitions of a Convex Function

**Definition 2.1.1.** Let $C$ be a nonempty convex set in $\mathbb{R}^n$. A function $f : C \to \mathbb{R}$ is said to be *convex on* $C$ when, for all pairs $(x, x') \in C \times C$ and all $\alpha \in [0, 1]$, there holds

$$f(\alpha x + (1-\alpha)x') \le \alpha f(x) + (1-\alpha) f(x'). \tag{1.1.1}$$

**Proposition 2.1.1.** $ConvexEval.FCA_chap_B.FCA_HUL_{112}Thefunctionfisstronglyconvexon Cwithmodulusci$ $\frac{1}{2}c\|\cdot\|^2$ *is convex on* $C$.

*Proof.* Use direct calculations in the definition (1.1.1) of convexity applied to the function $f - \frac{1}{2}c\|\cdot\|^2$, namely:

$$f(\alpha x + (1-\alpha)x') - \tfrac{1}{2}c\|\alpha x + (1-\alpha)x'\|^2 \le$$

$$\le \alpha f(x) + (1-\alpha) f(x') - \tfrac{1}{2}c\big[\alpha\|x\|^2 + (1-\alpha)\|x'\|^2\big].$$

□

**Definition 2.1.2.** (The Set $\operatorname{Conv}\mathbb{R}^n$) A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, not identically $+\infty$, is said to be convex when, for all $(x, x') \in \mathbb{R}^n \times \mathbb{R}^n$ and all $\alpha \in [0, 1]$, there holds

$$f\big(\alpha x + (1-\alpha)x'\big) \le \alpha f(x) + (1-\alpha) f(x'),$$

considered as an inequality in $\mathbb{R} \cup \{+\infty\}$.

The class of such functions is denoted by $\operatorname{Conv}\mathbb{R}^n$.

**Definition 2.1.3.** (Domain of a Function) The *domain* (or also effective domain) of $f \in \operatorname{Conv}\mathbb{R}^n$ is the nonempty set

$$\operatorname{dom} f := \{x \in \mathbb{R}^n : f(x) < +\infty\}.$$

**Definition 2.1.4.** (Epigraph of a Function) Given $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, not identically equal to $+\infty$, the *epigraph* of $f$ is the nonempty set

$$\operatorname{epi} f := \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : r \geq f(x)\}.$$

Its *strict epigraph* $\operatorname{epi}_s f$ is defined likewise, with "$\geq$" replaced by "$>$" (beware that the word "strict" here has nothing to do with strict convexity).

**Proposition 2.1.2.** *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be not identically equal to $+\infty$. The three properties below are equivalent:*

*(i) $f$ is convex in the sense of Definition 1.1.3;*

*(ii) its epigraph is a convex set in $\mathbb{R}^n \times \mathbb{R}$;*

*(iii) its strict epigraph is a convex set in $\mathbb{R}^n \times \mathbb{R}$.*

*Proof.* Left as an exercise. □

**Theorem 2.1.1.** *(Inequality of Jensen) Let $f \in \operatorname{Conv} \mathbb{R}^n$. Then, for all collections $\{x_1, \ldots, x_k\}$ of points in dom $f$ and all $\alpha = (\alpha_1, \ldots, \alpha_k)$ in the unit simplex of $\mathbb{R}^k$, there holds (inequality of Jensen in summation form)*

$$f\Big(\sum_{i=1}^{k} \alpha_i x_i\Big) \leq \sum_{i=1}^{k} \alpha_i f(x_i).$$

*Proof.* The $k$ points $(x_i, f(x_i)) \in \mathbb{R}^n \times \mathbb{R}$ are clearly in epi $f$, a convex set. Their convex combination

$$\sum_i \alpha_i (x_i, f(x_i)) = \Big(\sum_i \alpha_i x_i, \sum_i \alpha_i f(x_i)\Big)$$

is also in epi $f$ (Proposition A.1.3.3). This is just the claimed inequality. □

**Proposition 2.1.3.** *Let $f \in$ Conv $\mathbb{R}^n$. The relative interior of $\operatorname{epi} f$ is the union over $x \in \operatorname{ri} \operatorname{dom} f$ of the open half-lines with bottom endpoints at $f(x)$:*

$$\operatorname{ri} \operatorname{epi} f = \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : x \in \operatorname{ri} \operatorname{dom} f,\ r > f(x)\}.$$

*Proof.* Since $\operatorname{dom} f$ is the image of $\operatorname{epi} f$ under the linear mapping "projection onto $\mathbb{R}^n$", Proposition A.2.1.12 tells us that

$$\operatorname{ri} \operatorname{dom} f \text{ is the projection onto } \mathbb{R}^n \text{ of } \operatorname{ri} \operatorname{epi} f. \tag{1.1.5}$$

Now take $x$ arbitrary in $\operatorname{ri} \operatorname{dom} f$. The subset of $\operatorname{ri} \operatorname{epi} f$ that is projected onto $x$ is just $((\{x\} \times \mathbb{R}) \cap \operatorname{ri} \operatorname{epi} f)$, which in turn is $\operatorname{ri}((\{x\} \times \mathbb{R}) \cap \operatorname{epi} f)$ (use Proposition A.2.1.10). This latter set is clearly $\{x\} \times (f(x), +\infty)$.

In summary, we have proved that, for $x \in \operatorname{ri} \operatorname{dom} f$, $(x, r) \in \operatorname{ri} \operatorname{epi} f$ if and only if $r > f(x)$. Together with (1.1.5), this proves our claim. □

### 2.1.2 Special convex functions: Affinity and Closedness

**Proposition 2.1.4.** *Any* $f \in \operatorname{Conv} \mathbb{R}^n$ *is minorized by some affine function. More precisely: for any* $x_0 \in \operatorname{ri} \operatorname{dom} f$*, there is* $s$ *in the subspace parallel to* $\operatorname{aff} \operatorname{dom} f$ *such that*

$$f(x) \geq f(x_0) + \langle s,\, x - x_0 \rangle \quad \textit{for all } x \in \mathbb{R}^n.$$

In other words, the affine function can be forced to coincide with $f$ at $x_0$.

*Proof.* We know that $\operatorname{dom} f$ is the image of $\operatorname{epi} f$ under the linear mapping "projection onto $\mathbb{R}^n$". Look again at the definition of an affine hull (§A.1.3) to realize that

$$\operatorname{aff} \operatorname{epi} f = (\operatorname{aff} \operatorname{dom} f) \times \mathbb{R}\,.$$

Denote by $V$ the linear subspace parallel to $\operatorname{aff} \operatorname{dom} f$, so that $\operatorname{aff} \operatorname{dom} f = \{x_0\} + V$ with $x_0$ arbitrary in $\operatorname{dom} f$; then we have

$$\operatorname{aff} \operatorname{epi} f = \{x_0 + V\} \times \mathbb{R}. \tag{2.1}$$

We equip $V \times \mathbb{R}$ and $\mathbb{R}^n \times \mathbb{R}$ with the scalar product of product-spaces. With $x_0 \in \operatorname{ri} \operatorname{dom} f$, Proposition 1.1.9 tells us that $(x_0, f(x_0)) \in \operatorname{rbd} \operatorname{epi} f$ and we can take a nontrivial hyperplane supporting $\operatorname{epi} f$ at $(x_0, f(x_0))$: using Remark A.4.2.2 and (2.1), there are $s = s_V \in V$ and $\alpha \in \mathbb{R}$, not both zero, such that

$$\langle s, x \rangle + \alpha r \leq \langle s, x_0 \rangle + \alpha f(x_0) \tag{2.2}$$

for all $(x, r)$ with $f(x) \leq r$. Note: this implies $\alpha \leq 0$ (let $r \to +\infty$!).

Because of our choice of $s$ (in $V$) and $x_0$ (in $\operatorname{ri} \operatorname{dom} f$), we can take $\delta > 0$ so small that $x_0 + \delta s \in \operatorname{dom} f$, for which (2.2) gives

$$\delta \|s\|^2 \leq \alpha \big[ f(x_0) - f(x_0 + \delta s) \big] < +\infty;$$

this shows $\alpha \neq 0$ (otherwise, both $s$ and $\alpha$ would be zero). Without loss of generality, we can assume $\alpha = -1$; then (2.2) gives our affine function. $\square$

**Proposition 2.1.5.** *For* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$*, the following three properties are equivalent:*

*(i)* $f$ *is lower semi-continuous on* $\mathbb{R}^n$*;*

*(ii)* $\operatorname{epi} f$ *is a closed set in* $\mathbb{R}^n \times \mathbb{R}$*;*

*(iii) the sublevel-sets* $S_r(f)$ *are closed (possibly empty) for all* $r \in \mathbb{R}$*.*

*(i)* $\Rightarrow$ *(ii).* Let $(y_k, r_k)_k$ be a sequence of $\operatorname{epi} f$ converging to $(x, r)$ for $k \to +\infty$. Since $f(y_k) \leq r_k$ for all $k$, the l.s.c. relation (1.2.3) readily gives

$$r = \lim r_k \geq \liminf f(y_k) \geq \liminf_{y \to x} f(y) \geq f(x),$$

i.e. $(x, r) \in \operatorname{epi} f$.

[(ii) $\Rightarrow$ (iii)] Construct the sublevel-sets $S_r(f)$ as in Remark 1.1.7: the closed sets $\operatorname{epi} f$ and $\mathbb{R}^n \times \{r\}$ have a closed intersection.

[(iii) $\Rightarrow$ (i)] Suppose $f$ is not lower semi-continuous at some $x$: there is a (sub)sequence $(y_k)$ converging to $x$ such that $f(y_k)$ converges to $\rho < f(x) \leq +\infty$. Pick $r \in [\rho, f(x))$: for $k$ large enough, $f(y_k) \leq r < f(x)$; hence $S_r(f)$ contains the tail of $(y_k)$ but not its limit $x$. Consequently, this $S_r(f)$ is not closed. □

**Definition 2.1.5.** (Closed Functions) The function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to be closed if it is lower semi-continuous everywhere, or if its epigraph is closed, or if its sublevel-sets are closed.

**Definition 2.1.6** (Closure of a Function)**.** The *closure* (or lower semi-continuous hull) of a function $f$ is the function $\operatorname{cl} f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ defined by:

$$\operatorname{cl} f(x) := \liminf_{y \to x} f(y) \qquad \text{for all } x \in \mathbb{R}^n, \tag{1.2.4}$$

or equivalently by

$$\operatorname{epi}(\operatorname{cl} f) := \operatorname{cl}\big(\operatorname{epi} f\big). \tag{1.2.5}$$

**Proposition 2.1.6.** *Let $f \in \operatorname{Conv} \mathbb{R}^n$ and $x' \in \operatorname{ri} \operatorname{dom} f$. There holds (in $\mathbb{R} \cup \{+\infty\}$)*

$$\operatorname{cl} f(x) = \lim_{t \downarrow 0} f\big(x + t(x' - x)\big) \quad \textit{for all } x \in \mathbb{R}^n.$$

*Proof.* Since $x_t := x + t(x' - x) \to x$ when $t \downarrow 0$, we certainly have

$$(\operatorname{cl} f)(x) \leq \liminf_{t \downarrow 0} f\big(x + t(x' - x)\big).$$

We will prove the converse inequality by showing that

$$\limsup_{t \downarrow 0} f\big(x + t(x' - x)\big) \leq r \quad \text{for all } r \geq (\operatorname{cl} f)(x).$$

(Non-existence of such an $r$ means that $\operatorname{cl} f(x) = +\infty$, the proof is finished.)

Thus let $(x, r) \in \operatorname{epi}(\operatorname{cl} f) = \operatorname{cl}(\operatorname{epi} f)$. Pick $r' > f(x')$, hence $(x', r') \in \operatorname{epi} f$ (Proposition 1.1.9). Applying Lemma A.2.1.6 to the convex set $\operatorname{epi} f$, we see that

$$t(x', r') + (1 - t)(x, r) \in \operatorname{epi} f \subset \operatorname{epi} f \quad \text{for all } t \in ]0, 1].$$

This just means

$$f\big(x + t(x' - x)\big) \leq tr' + (1 - t)r \quad \text{for all } t \in ]0, 1]$$

and our required inequality follows by letting $t \downarrow 0$. □

**Proposition 2.1.7.** *1.2.6 For $f \in$ Conv $\mathbb{R}^n$, there holds*

$$\operatorname{cl} f \in \textit{Conv } \mathbb{R}^n; \tag{1.2.7}$$

$$\operatorname{cl} f \textit{ and } f \textit{ coincide on the relative interior of dom } f. \tag{1.2.8}$$

*Proof.* We already know from Proposition A.1.2.6 that epi $\mathrm{cl}\, f = \mathrm{cl}\, \mathrm{epi}\, f$ is a convex set; also $\mathrm{cl}\, f \leq f \not\equiv +\infty$; finally, Proposition 1.2.1 guarantees in the relation of definition (1.2.4) that $\mathrm{cl}\, f(x) > -\infty$ for all $x$: (1.2.7) does hold.

On the other hand, suppose $x \in \mathrm{ri}\ \mathrm{dom}\ f$. Then the one-dimensional function $\varphi(t) = f(x + td)$ is continuous at $t = 0$ (Theorem 0.6.2); it follows from Proposition 1.2.5 that $\mathrm{cl}\, f$ coincides with $f$ on ri dom $f$; besides, $\mathrm{cl}\, f(x)$ is obviously equal to $f(x) = +\infty$ for all $x \notin \mathrm{cl}\ \mathrm{dom}\ f$. Altogether, (1.2.8) is true. □

Notation 1.2.7 (The Set $\overline{\mathrm{Conv}}\mathbb{R}^n$) The set of closed convex functions on $\mathbb{R}^n$ is denoted by $\overline{\mathrm{Conv}}\mathbb{R}^n$. □

**Proposition 2.1.8.** *The closure of $f \in$ Conv $\mathbb{R}^n$ is the supremum of all affine functions minorizing $f$:*

$$\mathit{cl}\ f(x) = \sup_{(s,b)\in\mathbb{R}^n\times\mathbb{R}} \{\langle s, x\rangle - b : \langle s, y\rangle - b \leq f(y) \textit{ for all } y \in \mathbb{R}^n\}. \qquad (1.2.9)$$

*Proof.* A closed half-space containing epi $f$ is characterized by a nonzero vector $(s, \alpha) \in \mathbb{R}^n \times \mathbb{R}$ and a real number $b$ such that

$$\langle s, x\rangle + \alpha r \leq b \qquad \text{for all } (x, r) \in \mathrm{epi}\ f \qquad (1.2.10)$$

(we equip the graph-space $\mathbb{R}^n \times \mathbb{R}$ with the scalar product of a product-space). Let us denote by $\Sigma \subset \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}$ the index-set of such triples $\sigma = (s, \alpha, b)$, with corresponding half-space

$$H_\sigma^- := \{(x, r) : \langle s, x\rangle + \alpha r \leq b\}. \qquad (1.2.11)$$

In other words, $\mathrm{epi}(\overline{cl}\ f) = \overline{\mathrm{epi}\, f} = \bigcap_{\sigma\in\Sigma} H_\sigma^-$.

Because of the particular nature of an epigraph, (1.2.10) implies $\alpha \leq 0$ (let $r \to +\infty$) and, by positive homogeneity, the values $\alpha = 0$ and $\alpha = -1$ suffice: $\Sigma$ can be partitioned in

$$\Sigma_1 := \{(s, -1, b) : (1.2.10) \text{ holds with } \alpha = -1\}$$

and

$$\Sigma_0 := \{(s, 0, b) : (1.2.10) \text{ holds with } \alpha = 0\}.$$

Indeed, $\Sigma_1$ corresponds to affine functions minorizing $f$ (Proposition 1.2.1 tells us that $\Sigma_1 \neq \varnothing$) and $\Sigma_0$ to closed half-spaces of $\mathbb{R}^n$ containing dom $f$ (note that $\Sigma_0 = \varnothing$ if $\mathrm{dom}\, f = \mathbb{R}^n$).

We have to prove that, even when $\Sigma_0 \neq \varnothing$, intersecting the half-spaces $H_\sigma^-$ over $\Sigma$ or over $\Sigma_1$ produces the same set, namely epi $f$. For this we take arbitrary $\sigma_0 = (s_0, 0, b_0) \in \Sigma_0$ and $\sigma_1 = (s_1, -1, b_1) \in \Sigma_1$, we set

$$\sigma(t) := (s_1 + ts_0, -1, b_1 + tb_0) \in \Sigma_1 \quad \text{for all } t \geq 0,$$

and we prove (see Fig. 1.2.1)

$$H_{\sigma_0}^- \cap H_{\sigma_1}^- = \bigcap_{t\geq 0} H_{\sigma(t)}^- =: H^- .$$

Fig. 1.2.1. Closing a convex epigraph

It results directly from the definition (1.2.11) that an $(x,r)$ in $H_{\sigma_0}^- \cap H_{\sigma_1}^-$ satisfies

$$\langle s_1 + ts_0, x\rangle - (b_1 + tb_0) \leq r \quad \text{for all } t \geq 0, \tag{1.2.12}$$

i.e. $(x,r) \in H^-$. Conversely, take $(x,r) \in H^-$. Set $t = 0$ in (1.2.12) to see that $(x,r) \in H_{\sigma_1}^-$. Also, divide by $t > 0$ and let $t \to +\infty$ to see that $(x,r) \in H_{\sigma_0}^-$. The proof is complete. $\square$

### 2.1.3 First examples

**Theorem 2.1.2.** *Let $C$ be a nonempty subset of $\mathbb{R}^n \times \mathbb{R}$ satisfying* (**??**)*, and let its lower-bound function $\ell_C$ be defined by* (**??**)*.*

*[(i)]*

1. *If $C$ is convex, then $\ell_C \in \operatorname{Conv}\mathbb{R}^n$;*
2. *If $C$ is closed convex, then $\ell_C \in \overline{\operatorname{Conv}}\mathbb{R}^n$.*

*Proof.* We use the analytical definition (1.1.1). Take arbitrary $\varepsilon > 0$, $\alpha \in ]0,1[$ and $(x_i, r_i) \in C$ such that $r_i \leq \ell_C(x_i) + \varepsilon$ for $i = 1, 2$.

When $C$ is convex, $(\alpha x_1 + (1-\alpha)x_2, \alpha r_1 + (1-\alpha)r_2) \in C$, hence

$$\ell_C(\alpha x_1 + (1-\alpha)x_2) \leq \alpha r_1 + (1-\alpha)r_2 \leq \alpha\ell_C(x_1) + (1-\alpha)\ell_C(x_2) + \varepsilon.$$

The convexity of $\ell_C$ follows, since $\varepsilon > 0$ was arbitrary; (i) is proved.

Now take a sequence $(x_k, \rho_k)_k \subset \operatorname{epi}\ell_C$ converging to $(x, \rho)$; we have to prove $\ell_C(x) \leq \rho$ (cf. Proposition 1.2.2). By definition of $\ell_C(x_k)$, we can select, for each positive integer $k$, a real number $r_k$ such that $(x_k, r_k) \in C$ and

$$\ell_C(x_k) \leq r_k \leq \ell_C(x_k) + \tfrac{1}{k} \leq \rho_k + \tfrac{1}{k}. \tag{2.3}$$

We deduce first that $(r_k)$ is bounded from above. Also, when $\ell_C$ is convex, Proposition 1.2.1 implies the existence of an affine function minorizing $\ell_C$: $(r_k)$ is bounded from below.

Extracting a subsequence if necessary, we may assume $r_k \to r$. When $C$ is closed, $(x,r) \in C$, hence $\ell_C(x) \leq r$; but pass to the limit in (2.3) to see that $r \leq \rho$; the proof is complete. $\square$

## 2.2 Functional Operations Preserving Convexity

### 2.2.1 Operations preserving closedness

**Proposition 2.2.1.** *Let $f_1, \dots, f_m$ be in Conv $\mathbb{R}^n$ [resp. in $\overline{\text{Conv}}\,\mathbb{R}^n$], let $t_1, \dots, t_m$ be positive numbers, and assume that there is a point where all the $f_j$'s are finite. Then the function $f := \sum_{j=1}^m t_j f_j$ is in Conv $\mathbb{R}^n$ [resp. in $\overline{\text{Conv}}\,\mathbb{R}^n$].*

*Proof.* The convexity of $f$ is readily proved from the relation of definition (1.1.1). As for its closedness, start from

$$\liminf_{y\to x} t_j f_j(y) = t_j \liminf_{y\to x} f_j(y) \geq t_j f_j(x)$$

(valid for $t_j > 0$ and $f_j$ closed); then note that the lim inf of a sum is not smaller than the sum of lim inf's. □

**Proposition 2.2.2.** *Let $\{f_j\}_{j\in J}$ be an arbitrary family of convex [resp. closed convex] functions. If there exists $x_0$ such that $\sup_j f_j(x_0) < +\infty$, then their pointwise supremum $f := \sup\{f_j : j \in J\}$ is in* $\operatorname{Conv}\mathbb{R}^n$ *[resp. in* $\operatorname{Conv}_{\text{cl}}\mathbb{R}^n$*].*

*Proof.* The key property is that a supremum of functions corresponds to an intersection of epigraphs: $\operatorname{epi} f = \bigcap_{j\in J} \operatorname{epi} f_j$, which conserves convexity and closedness. The only needed restriction is nonemptiness of this intersection. □

**Proposition 2.2.3.** *Let $f \in \operatorname{Conv}\mathbb{R}^n$ [resp. $\overline{\operatorname{Conv}}\mathbb{R}^n$] and let $A$ be an affine mapping from $\mathbb{R}^m$ to $\mathbb{R}^n$ such that $\operatorname{Im} A \cap \operatorname{dom} f \neq \varnothing$. Then the function*

$$f \circ A : \mathbb{R}^m \supseteq x \mapsto (f \circ A)(x) = f(A(x))$$

*is in $\operatorname{Conv}\mathbb{R}^m$ [resp. $\overline{\operatorname{Conv}}\mathbb{R}^m$].*

*Proof.* Clearly $(f \circ A)(x) > -\infty$ for all $x$; besides, there exists by assumption $y = A(x) \in \mathbb{R}^n$ such that $f(y) < +\infty$. To check convexity, it suffices to plug the relation

$$A(\alpha x + (1-\alpha)x') = \alpha A(x) + (1-\alpha)A(x')$$

into the analytical definition (1.1.1) of convexity. As for closedness, it comes readily from the continuity of $A$ when $f$ is itself closed. □

**Proposition 2.2.4.** *Let $f \in \operatorname{Conv}\mathbb{R}^n$ [resp. $\overline{\operatorname{Conv}}\mathbb{R}^n$] and let $g \in \operatorname{Conv}\mathbb{R}$ [resp. $\overline{\operatorname{Conv}}\mathbb{R}$] be increasing. Assume that there is $x_0 \in \mathbb{R}^n$ such that $f(x_0) \in \operatorname{dom} g$, and set $g(+\infty) := +\infty$. Then the composite function $g \circ f : x \mapsto g(f(x))$ is in $\operatorname{Conv}\mathbb{R}^n$ [resp. in $\overline{\operatorname{Conv}}\mathbb{R}^n$].*

*Proof.* It suffices to check the inequalities of definition: (1.1.1) for convexity, (1.2.3) for closedness. □

### 2.2.2 Dilations and perspectives of a functions

**Proposition 2.2.5.** *If* $f \in \operatorname{Conv}\mathbb{R}^n$*, its perspective* $\tilde{f}$ *is in* $\operatorname{Conv}\mathbb{R}^{n+1}$.

*Proof.* Here also, it is better to look at $\tilde{f}$ with "geometric glasses":

$$\operatorname{epi}\tilde{f} = \{(u,x,r) \in \mathbb{R}_+^\times \times \mathbb{R}^n \times \mathbb{R} : \ f(x/u) \le r/u\} = \{u(1,x',r') : \ u > 0, (x',r') \in \operatorname{epi} f\}$$

$$= \bigcup_{u>0} u(\{1\} \times \operatorname{epi} f) = \mathbb{R}_+^\times(\{1\} \times \operatorname{epi} f)$$

and $\operatorname{epi}\tilde{f}$ is therefore a convex cone. $\square$

**Proposition 2.2.6.** *Let* $f \in \operatorname{Conv}\mathbb{R}^n$ *and let* $x' \in \operatorname{ri}$ *dom* $f$*. Then the closure* $\overline{f}$ *of its perspective is given as follows:*

$$(cl\ \tilde{f})(u,x) = \begin{cases} uf(x/u) & \text{if } u > 0,\\ \lim_{\alpha\downarrow 0} \alpha f(x' - x + \frac{x}{\alpha}) & \text{if } u = 0,\\ +\infty & \text{if } u < 0. \end{cases}$$

*Proof.* Suppose first $u < 0$. For any $x$, it is clear that $(u,x)$ is outside cl dom $\tilde{f}$ and, in view of (1.2.8), cl $\tilde{f}(u,x) = +\infty$.

Now let $u \ge 0$. Using (2.2.1), the assumption on $x'$ and the results of §A.2.1, we see that $(1,x') \in$ ri dom $\tilde{f}$, so Proposition 1.2.5 allows us to write

$$(\mathrm{cl}\ \tilde{f})(u,x) = \lim_{\alpha\downarrow 0} \tilde{f}\big((u,x) + \alpha[(1,x') - (u,x)]\big)$$

$$= \lim_{\alpha\downarrow 0} [u + \alpha(1-u)]\, f\Big(\frac{x + \alpha(x'-x)}{u + \alpha(1-u)}\Big).$$

If $u = 1$, this reads cl $\tilde{f}(1,x) =$ cl $f(x) = f(x)$ (because $f$ is closed); if $u = 0$, we just obtain our claimed relation. $\square$

### 2.2.3 Infimal convolution

**Definition 2.2.1.** Let $f_1$ and $f_2$ be two functions from $\mathbb{R}^n$ to $\mathbb{R} \cup \{+\infty\}$. Their *infimal convolution* is the function from $\mathbb{R}^n$ to $\mathbb{R} \cup \{\pm\infty\}$ defined by

$$\begin{aligned}(f_1 \overset{+}{\vee} f_2)(x) &:= \inf\{f_1(x_1) + f_2(x_2) : \ x_1 + x_2 = x\} \\ &= \inf_{y \in \mathbb{R}^n} [\, f_1(y) + f_2(x-y)\,].\end{aligned} \tag{2.4}$$

We will also call "infimal convolution" the *operation* expressed by (2.4). It is called exact at $x = \bar{x}_1 + \bar{x}_2$ when the infimum is attained at $(\bar{x}_1, \bar{x}_2)$, not necessarily unique. $\square$

**Proposition 2.2.7.** *Let the functions $f_1$ and $f_2$ be in* $\operatorname{Conv}\mathbb{R}^n$. *Suppose that they have a common affine minorant: for some $(s, b) \in \mathbb{R}^n \times \mathbb{R}$,*

$$f_j(x) \geq \langle s, x\rangle - b \qquad \textit{for } j = 1, 2 \textit{ and all } x \in \mathbb{R}^n.$$

*Then their infimal convolution is also in* $\operatorname{Conv}\mathbb{R}^n$.

*Proof.* For arbitrary $x \in \mathbb{R}^n$ and $x_1, x_2$ such that $x_1 + x_2 = x$, we have by assumption

$$f_1(x_1) + f_2(x_2) \geq \langle s, x\rangle - 2b > -\infty,$$

and this inequality extends to the infimal value $(f_1 \overset{+}{\vee} f_2)(x)$.

On the other hand, it suffices to choose particular values $x_j \in \operatorname{dom} f_j$, $j = 1, 2$, to obtain the point $x_1 + x_2 \in \operatorname{dom}(f_1 \overset{+}{\vee} f_2)$. Finally, the convexity of $f_1 \overset{+}{\vee} f_2$ results from the convexity of a lower-bound function, as seen in §1.3(g). ☐

### 2.2.4 Image of a function under a linear mapping

**Definition 2.2.2.** (Image Function) Let $A : \mathbb{R}^m \to \mathbb{R}^n$ be a linear operator and let $g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$. The image of $g$ under $A$ is the function $Ag : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ defined by

$$(Ag)(x) := \inf\{\, g(y) : \ Ay = x \,\} \tag{2.4.2}$$

(here as always, $\inf \varnothing = +\infty$).

**Theorem 2.2.1.** *Let $g$ of Definition 2.4.1 be in* $\operatorname{Conv}\mathbb{R}^m$. *Assume also that, for all $x \in \mathbb{R}^n$, $g$ is bounded below on the inverse image $A^{-1}(x) = \{y \in \mathbb{R}^m : Ay = x\}$. Then $Ag \in$* $\operatorname{Conv}\mathbb{R}^n$.

*Proof.* By assumption, $Ag$ is nowhere $-\infty$; also, $(Ag)(x) < +\infty$ whenever $x = Ay$, with $y \in \operatorname{dom} g$. Now consider the extended operator

$$A' : \mathbb{R}^m \times \mathbb{R} \rightrightarrows (y, r) \mapsto A'(y, r) := (Ay, r) \in \mathbb{R}^n \times \mathbb{R}.$$

The set $A'(\operatorname{epi} g) =: C$ is convex in $\mathbb{R}^n \times \mathbb{R}$; let us compute its lower-bound function (1.3.5): for given $x \in \mathbb{R}^n$,

$$\inf_r\{r : (x, r) \in C\} = \inf_{y,r}\{r : Ay = x \text{ and } g(y) \leq r\} = \inf_y\{g(y) : Ay = x\} = (Ag)(x),$$

and this proves the convexity of $Ag = \ell_C$. ☐

**Corollary 2.2.1.** *Let (2.4.1) have the following form: $U = \mathbb{R}^p$; $\varphi \in$ Conv $\mathbb{R}^p$; $X = \mathbb{R}^n$ is equipped with the canonical basis; the mapping $c$ has its components $c_j \in$ Conv $\mathbb{R}^p$ for $j = 1, \ldots, n$. Suppose also that the optimal value is $> -\infty$ for all $x \in \mathbb{R}^n$, and that*

$$\textit{dom } \varphi \cap \textit{dom } c_1 \cap \cdots \cap \textit{dom } c_n \neq \varnothing. \tag{2.4.4}$$

*Then the value function*

$$v_{\varphi,c}(x) := \inf\{\varphi(u) : c_j(u) \le x_j,\ \textit{for } j = 1, \dots, n\}$$

*lies in Conv* $\mathbb{R}^n$.

*Proof.* Note first that we have assumed $v_{\varphi,c}(x) > -\infty$ for all $x$. Take $u_0$ in the set (2.4.4) and set $M := \max_j c_j(u_0)$; then take $x_0 := (M, \dots, M) \in \mathbb{R}^n$, so that $v_{\varphi,c}(x_0) \le \varphi(u_0) < +\infty$. Knowing that $v_{\varphi,c}$ is an image-function, we just have to prove the convexity of the set (2.4.3); but this in turn comes immediately from the convexity of each $c_j$. ☐

**Definition 2.2.3.** (Marginal Function) Let $g \in \mathrm{Conv}(\mathbb{R}^n \times \mathbb{R}^m)$. Then

$$\mathbb{R}^n \ni x \mapsto \gamma(x) := \inf\{\, g(x, y) :\ y \in \mathbb{R}^m \,\}$$

is the marginal function of $g$.

**Corollary 2.2.2.** *With the above notation, suppose that $g$ is bounded below on the set $\{x\} \times \mathbb{R}^m$, for all $x \in \mathbb{R}^n$. Then the marginal function $\gamma$ lies in* Conv $\mathbb{R}^n$.

*Proof.* The marginal function $\gamma$ is the image of $g$ under the the linear operator $A$ projecting each $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ onto $x \in \mathbb{R}^n$: $A(x, y) = x$. ☐

### 2.2.5 Convex hull and closed convex hull of a function

**Proposition 2.2.8.** *Let $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, not identically $+\infty$, be minorized by some affine function: for some $(s, b) \in \mathbb{R}^n \times \mathbb{R}$,*

$$g(x) \ge \langle s, x\rangle - b \quad \textit{for all } x \in \mathbb{R}^n. \tag{2.5.1}$$

Then, the following three functions $f_1, f_2$ and $f_3$ are convex and coincide on $\mathbb{R}^n$:

$$\begin{aligned} f_1(x) &:= \inf\{r : (x, r) \in \mathrm{co\ epi}\, g\}, \\ f_2(x) &:= \sup\{h(x) : h \in \mathrm{Conv}\,\mathbb{R}^n,\ h \le g\}, \\ f_3(x) &:= \inf\Big\{\sum_{j=1}^{k} \alpha_j g(x_j) : k = 1, 2, \dots \\ &\qquad\qquad \alpha \in \Delta_k,\ x_j \in \mathrm{dom}\, g,\ \sum_{j=1}^{k} \alpha_j x_j = x\Big\}. \end{aligned} \tag{2.5.2}$$

*Proof.* We denote by $\Gamma$ the family of convex functions minorizing $g$. By assumption, $\Gamma \ne \varnothing$; then the convexity of $f_1$ results from §1.3(g).

$[f_2 \le f_1]$ Consider the epigraph of any $h \in \Gamma$: its lower-bound function $\ell_{\mathrm{epi}\, h}$ is $h$ itself; besides, it contains epi $g$, and co(epi $g$) as well (see Proposition A.1.3.4). In a word, there holds $h = \ell_{\mathrm{epi}\, h} \le \ell_{\mathrm{co\ epi}\, g} = f_1$ and we conclude $f_2 \le f_1$ since $h$ was arbitrary in $\Gamma$.

$[f_3 \leq f_2]$ We have to prove $f_3 \in \Gamma$, and the result will follow by definition of $f_2$; clearly $f_3 \leq g$ (take $\alpha \in \Delta_1$!), so it suffices to establish $f_3 \leq \operatorname{Conv} \mathbb{R}^n$. First, with $(s, b)$ of (2.5.1) and all $x$, $\{x_j\}$ and $\{\alpha_j\}$ as described by (2.5.2),

$$\sum_{j=1}^{k} \alpha_j g(x_j) \geq \sum_{j=1}^{k} \alpha_j (\langle s, x_j \rangle - b) = \langle s, x \rangle - b;$$

hence $f_3$ is minorized by the affine function $\langle s, \cdot \rangle - b$. Now, take two points $(x, r)$ and $(x', r')$ in the strict epigraph of $f_3$. By definition of $f_3$, there are $k$, $\{\alpha_j\}$, $\{x_j\}$ as described in (2.5.2), and likewise $k'$, $\{\alpha'_j\}$, $\{x'_j\}$, such that $\sum_{j=1}^{k} \alpha_j g(x_j) < r$ and likewise $\sum_{j=1}^{k'} \alpha'_j g(x'_j) < r'$.

For arbitrary $t \in ]0, 1[$, we obtain by convex combination

$$\sum_{j=1}^{k} t\alpha_j g(x_j) + \sum_{j=1}^{k'} (1-t)\alpha'_j g(x'_j) < tr + (1-t)r'.$$

Observe that

$$\sum_{j=1}^{k} t\alpha_j x_j + \sum_{j=1}^{k'} (1-t)\alpha'_j x'_j = tx + (1-t)x',$$

i.e. we have in the lefthand side a convex decomposition of $tx + (1-t)x'$ in $k + k'$ elements; therefore, by definition of $f_3$:

$$f_3(tx + (1-t)x') \leq \sum_{j=1}^{k} t\alpha_j g(x_j) + \sum_{j=1}^{k'} (1-t)\alpha'_j g(x'_j)$$

and we have proved that $\operatorname{epi}_s f_3$ is a convex set: $f_3$ is convex.

Let $x \in \mathbb{R}^n$ and take an arbitrary convex decomposition $x = \sum_{j=1}^{k} \alpha_j x_j$, with $\alpha_j$ and $x_j$ as described in (2.5.2). Since $(x_j, g(x_j)) \in \operatorname{epi} g$ for $j = 1, \ldots, k$,

$$\left( x, \sum_{j=1}^{k} \alpha_j g(x_j) \right) \in \operatorname{co} \operatorname{epi} g$$

and this implies $f_1(x) \leq \sum_{j=1}^{k} \alpha_j g(x_j)$ by definition of $f_1$. Because the decomposition of $x$ was arbitrary within (2.5.2), this implies $f_1(x) \leq f_3(x)$. $\square$

**Proposition 2.2.9.** *Let $g$ satisfy the hypotheses of Proposition 2.5.1. Then the three functions below*

$$\bar{f}_1(x) := \inf\{r : (x, r) \in \overline{\operatorname{epi} g}\}, \qquad \bar{f}_2(x) := \sup\{h(x) : \ h \in \operatorname{Conv} \mathbb{R}^n, \ h \leq g\},$$

$$\bar{f}_3(x) := \sup\{\langle s, x \rangle - b : \ \langle s, y \rangle - b \leq g(y) \text{ for all } y \in \mathbb{R}^n\}$$

*are closed, convex, and coincide on $\mathbb{R}^n$ with the closure of the function constructed in Proposition 2.5.1.*

**Definition 2.2.4** (Convex Hulls of a Function)**.** Let $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, not identically $+\infty$, be minorized by an affine function. The common function $f_1 = f_2 = f_3$ of Proposition 2.5.1 is called the *convex hull* of $g$, denoted by $\operatorname{co} g$. The *closed convex hull* of $g$ is any of the functions described by Proposition 2.5.2; it is denoted by $\overline{\operatorname{co}}\, g$ or $\operatorname{cl}\operatorname{co} g$.

**Proposition 2.2.10.** *Let $g_1, \dots, g_m$ be in $\operatorname{Conv}\mathbb{R}^n$, all minorized by the same affine function. Then the convex hull of their infimum is the function*

$$\mathbb{R}^n \ni x \mapsto [\operatorname{co}(\min_j g_j)](x) = \inf\Big\{ \sum_{j=1}^m \alpha_j g_j(x_j) \ : \ \alpha \in \Delta_m,\ x_j \in \operatorname{dom} g_j,\ \sum_{j=1}^m \alpha_j x_j = x \Big\}. \tag{2.5.3}$$

*Proof.* Apply Example A.1.3.5 to the convex sets $C_j = \operatorname{epi} g_j$. □

## 2.3 Local and Global Behavior of a Convex Function

### 2.3.1 Continuity properties

**Lemma 2.3.1.** *Let $f \in \operatorname{Conv}\mathbb{R}^n$ and suppose there are $x_0,\ \delta,\ m$ and $M$ such that*

$$m \le f(x) \le M \quad \text{for all } x \in B(x_0, 2\delta).$$

*Then $f$ is Lipschitzian on $B(x_0, \delta)$; more precisely: for all $y$ and $y'$ in $B(x_0, \delta)$,*

$$|f(y) - f(y')| \le \frac{M - m}{\delta}\|y - y'\|. \tag{2.5}$$

*Proof.* Look at Fig. 3.1.1: with two different $y$ and $y'$ in $B(x_0, \delta)$, take

$$y'' := y' + \delta\,\frac{y' - y}{\|y' - y\|} \in B(x_0, 2\delta);$$

by construction, $y'$ lies on the segment $[y, y'']$, namely

$$y' = \frac{\|y' - y\|}{\delta + \|y' - y\|}\, y'' + \frac{\delta}{\delta + \|y' - y\|}\, y.$$

Applying the convexity of $f$ and using the postulated bounds, we obtain

$$f(y') - f(y) \le \frac{\|y' - y\|}{\delta + \|y' - y\|}\big[f(y'') - f(y)\big] \le \frac{1}{\delta}\|y' - y\|(M - m).$$

Then, it suffices to exchange $y$ and $y'$ to prove (3.1.1). □

**Theorem 2.3.1.** *With $f \in \operatorname{Conv}\mathbb{R}^n$, let $S$ be a convex compact subset of* ri *dom $f$. Then there exists $L = L(S) \ge 0$ such that*

$$|f(x) - f(x')| \le L\|x - x'\| \quad \text{for all } x \text{ and } x' \text{ in } S. \tag{3.1.2}$$

*Preliminaries.* First of all, our statement ignores $x$-values outside the affine hull of the convex set dom $f$. Instead of $\mathbb{R}^n$, it can be formulated in $\mathbb{R}^d$, where $d$ is the dimension of dom $f$; alternatively, we may assume ri dom $f = \int$ dom $f$, which will simplify the writing.

Make this assumption and let $x_0 \in S$. We will prove that there are $\delta = \delta(x_0) > 0$ and $L = L(x_0, \delta)$ such that the ball $B(x_0, \delta)$ is included in $\int$ dom $f$ and

$$|f(y) - f(y')| \leq L\|y - y'\| \quad \text{for all } y \text{ and } y' \text{ in } B(x_0, \delta). \tag{3.1.3}$$

If this holds for all $x_0 \in S$, the corresponding balls $B(x_0, \delta)$ will provide a covering of the compact set $S$, from which we will extract a finite covering $(x_1, \delta_1, L_1), \ldots, (x_k, \delta_k, L_k)$. With these balls, we will divide an arbitrary segment $[x, x']$ of the convex set $S$ into finitely many subsegments, of endpoints $y_0 := x, \ldots, y_i, \ldots, y_\ell := x'$. Ordering properly the $y_i$'s, we will have $\|x - x'\| = \sum_{i=1}^{\ell} \|y_i - y_{i-1}\|$; furthermore, $f$ will be Lipschitzian on each $[y_{i-1}, y_i]$ with the common constant $L := \max\{L_1, \ldots, L_k\}$. The required Lipschitz property (3.1.2) will follow.

[Main Step] To establish (3.1.3), we use Lemma 3.1.1, which requires boundedness of $f$ in the neighborhood of $x_0$. For this, we construct as in the proof of Theorem A.2.1.3 (see Fig. A.2.1.1) a simplex $\Delta = \mathrm{co}\{v_0, \ldots, v_n\} \cap \mathrm{dom}\, f$ having $x_0$ in its interior: we can take $\delta > 0$ such that $B(x_0, 2\delta) \subset \Delta$.

Then any $y \in B(x_0, 2\delta)$ can be written: $y = \sum_{i=0}^{n} \alpha_i v_i$ with $\alpha \in \Delta_{n+1}$, so that the convexity of $f$ gives

$$f(y) \leq \sum_{i=0}^{n} \alpha_i f(v_i) \leq \max\{f(v_0), \ldots, f(v_n)\} =: M.$$

On the other hand, Proposition 1.2.1 tells us that $f$ is bounded from below, say by $m$, on this very same $B(x_0, 2\delta)$. Our claim is proved: we have singled out $\delta > 0$ such that $m \leq f(y) \leq M$ for all $y \in B(x_0, 2\delta)$. $\square$

**Theorem 2.3.2.** *Let the convex functions $f_k : \mathbb{R}^n \to \mathbb{R}$ converge pointwise for $k \to +\infty$ to $f : \mathbb{R}^n \to \mathbb{R}$. Then $f$ is convex and, for each compact set $S$, the convergence of $f_k$ to $f$ is uniform on $S$.*

*Proof.* Convexity of $f$ is trivial: pass to the limit in the definition (1.1.1) itself. For uniformity, we want to use Lemma 3.1.1, so we need to bound $f_k$ on $S$ independently of $k$; thus, let $r > 0$ be such that $S \subset B(0, r)$.

[Step 1] First the function $g := \sup_k f_k$ is convex, and $g(x) < +\infty$ for all $x$ because the convergent sequence $(f_k(x))_k$ is certainly bounded. Hence, $g$ is continuous and therefore bounded, say by $M$, on the compact set $B(0, 2r)$:

$$f_k(x) \leq g(x) \leq M \qquad \text{for all } k \text{ and all } x \in B(0, 2r).$$

Second, the convergent sequence $(f_k(0))_k$ is bounded from below:

$$\mu \leq f_k(0) \qquad \text{for all } k.$$

Then, for $x \in B(0, 2r)$ and all $k$, use convexity on $[-x, x] \subset B(0, 2r)$:

$$2\mu \leq 2f_k(0) \leq f_k(x) + f_k(-x) \leq f_k(x) + M,$$

i.e. the $f_k$'s are bounded from below, independently of $k$. Thus, we are within the conditions of Lemma 3.1.1: there is some $L$ (independent of $k$) such that

$$|f_k(y) - f_k(y')| \leq L\|y - y'\| \quad \text{for all } k \text{ and all } y, y' \in B(0, r). \tag{3.1.5}$$

Naturally, the same Lipschitz property is transmitted to the limiting function $f$.

[Step 2] Now fix $\varepsilon > 0$. Cover $S$ by the balls $B(x, \varepsilon)$ for $x$ describing $S$, and extract a finite covering $S \subset B(x_1, \varepsilon) \cup \cdots \cup B(x_m, \varepsilon)$. With $x$ arbitrary in $S$, take an $x_i$ such that $x \in B(x_i, \varepsilon)$. There is $k_{i,\varepsilon}$ such that, for all $k \geq k_{i,\varepsilon}$,

$$|f_k(x) - f(x)| \leq |f_k(x) - f_k(x_i)| + |f_k(x_i) - f(x_i)| + |f(x_i) - f(x)| \leq (2L + 1)\varepsilon$$

where we have also used (3.1.5), knowing that $x$ and $x_i$ are in $S \subset B(0, r)$. The above inequality is then valid uniformly in $x$, providing that

$$k \geq \max\{k_{1,\varepsilon}, \ldots, k_{m,\varepsilon}\} =: k_\varepsilon.$$

□

### 2.3.2 Behavior at infinity

**Proposition 2.3.1.** *For $f \in \overline{\mathrm{Conv}}\mathbb{R}^n$, the asymptotic cone of* $\operatorname{epi} f$ *is the epigraph of the function $f'_\infty \in \overline{\mathrm{Conv}}\mathbb{R}^n$ defined by*

$$d \mapsto f'_\infty(d) := \sup_{t>0} \frac{f(x_0 + td) - f(x_0)}{t} = \lim_{t \to +\infty} \frac{f(x_0 + td) - f(x_0)}{t}, \tag{3.2.2}$$

*where $x_0$ is arbitrary in* $\operatorname{dom} f$.

*Proof.* Since $(x_0, f(x_0)) \in \operatorname{epi} f$, (3.2.1) tells us that $(d, \rho) \in (\operatorname{epi} f)_\infty$ if and only if $f(x_0 + td) \leq f(x_0) + t\rho$ for all $t > 0$, which means

$$\sup_{t>0} \frac{f(x_0 + td) - f(x_0)}{t} \leq \rho. \tag{3.2.3}$$

In other words, $(\operatorname{epi} f)_\infty$ is the epigraph of the function whose value at $d$ is the left hand side of (3.2.3); and this is true no matter how $x_0$ has been chosen in $\operatorname{dom} f$. The rest follows from the fact that the difference quotient in (3.2.3) is closed convex in $d$, and increasing in $t$ (the function $t \mapsto f(x_0 + td)$ is convex and enjoys the property of increasing slopes, namely Proposition 0.6.1). □ □

**Definition 2.3.1.** (Asymptotic function) The function $f'_\infty$ of Proposition 3.2.1 is called the *asymptotic function*, or recession function, of $f$.

**Proposition 2.3.2.** *Let $f \in \operatorname{Conv} \mathbb{R}^n$. All the nonempty sublevel-sets of $f$ have the same asymptotic cone, which is the sublevel-set of $f^\infty$ at the level 0:*

$$\forall r \in \mathbb{R} \text{ with } S_r(f) \neq \varnothing, \qquad [S_r(f)]_\infty = \{d \in \mathbb{R}^n : \; f^\infty(d) \leq 0\}.$$

*In particular, the following statements are equivalent:*

(i) There is r for which $S_r(f)$ is nonempty and compact;
(ii) all the sublevel-sets of $f$ are compact;
(iii) $f_\circ^\infty(d) > 0$ for all nonzero $d \in \mathbb{R}^n$.

*Proof.* By definition (A.2.2.1), a direction $d$ is in the asymptotic cone of the nonempty sublevel-set $S_r(f)$ if and only if

$$x \in S_r(f) \qquad \Longrightarrow \qquad [x + td \in S_r(f) \text{ for all } t > 0],$$

which can also be written — see (1.1.4):

$$(x, r) \in \operatorname{epi} f \qquad \Longrightarrow \qquad (x + td, r + t \times 0) \in \operatorname{epi} f \;\text{ for all } t > 0;$$

and this in turn just means that $(d, 0) \in (\operatorname{epi} f)_\infty = \operatorname{epi} f_\circ^\infty$. We have proved the first part of the theorem.

A particular case is when the sublevel-set $S_0(f_\circ^\infty)$ is reduced to the singleton $\{0\}$, which exactly means (iii). This is therefore equivalent to $[S_r(f)]_\infty = \{0\}$ for all $r \in \mathbb{R}$ with $S_r(f) \neq \emptyset$, which means that $S_r(f)$ is compact (Proposition A.2.2.3). The equivalence between (i), (ii) and (iii) is proved. □

**Definition 2.3.2.** (Coercivity) The functions $f \in \operatorname{Conv} \mathbb{R}^n$ satisfying (i), (ii) or (iii) are called 0-coercive. Equivalently, the 0-coercive functions are those that "increase at infinity":

$$f(x) \to +\infty \qquad \text{whenever} \qquad \|x\| \to +\infty,$$

and closed convex 0-coercive functions achieve their minimum over $\mathbb{R}^n$.

An important particular case is when $f'_\infty(d) = +\infty$ for all $d \neq 0$, i.e. when $f'_\infty = \iota_{\{0\}}$. It can be seen that this means precisely

$$\frac{f(x)}{\|x\|} \to +\infty \qquad \text{whenever} \qquad \|x\| \to +\infty.$$

(to establish this equivalence, extract a cluster point of $(x_k/\|x_k\|)$, and use the lower semi-continuity of $f'_\infty$). In words: at infinity, $f$ increases to infinity faster than any affine function; such functions are called 1-coercive, or sometimes just coercive.

**Proposition 2.3.3.** *A function $f \in \operatorname{Conv} \mathbb{R}^n$ is Lipschitzian on the whole of $\mathbb{R}^n$ if and only if $f'_\infty$ is finite on the whole of $\mathbb{R}^n$. The best Lipschitz constant for $f$ is then*

$$\sup\{f'_\infty(d) : \|d\| = 1\}. \tag{3.2.4}$$

*Proof.* When the (convex) function $f'_\infty$ is finite-valued, it is continuous (§3.1) and therefore bounded on the compact unit sphere:

$$\sup\{f'_\infty(d) : \|d\| = 1\} =: L < +\infty,$$

which implies by positive homogeneity

$$f'_\infty(d) \leq L\|d\| \quad \text{for all } d \in \mathbb{R}^n.$$

Now use the definition (3.2.2) of $f'_\infty$:

$$f(x+d) - f(x) \leq L\|d\| \quad \text{for all } x \in \operatorname{dom} f \text{ and } d \in \mathbb{R}^n;$$

thus, $\operatorname{dom} f$ is the whole space ($f(x+d) < +\infty$ for all $d$) and we do obtain that $L$ is a global Lipschitz constant for $f$.

Conversely, let $f$ have a global Lipschitz constant $L$. Pick $x_0 \in \operatorname{dom} f$ and plug the inequality

$$f(x_0 + td) - f(x_0) \leq Lt\|d\| \quad \text{for all } t > 0 \text{ and } d \in \mathbb{R}^n$$

into the definition (3.2.2) of $f'_\infty$ to obtain $f'_\infty(d) \leq L\|d\|$ for all $d \in \mathbb{R}^n$.

It follows that $f'_\infty$ is finite everywhere, and the value (3.2.4) does not exceed $L$. □

**Proposition 2.3.4** (3.2.8)**.**

1. *Let $f_1, \ldots, f_m$ be $m$ functions of Conv $\mathbb{R}^n$, and $t_1, \ldots, t_m$ be positive numbers. Assume that there is $x_0$ at which each $f_j$ is finite. Then,*

$$\text{for } f := \sum_{j=1}^m t_j f_j, \qquad \text{we have } f'_\infty = \sum_{j=1}^m t_j (f_j)'_\infty.$$

2. *Let $\{f_j\}_{j \in J}$ be a family of functions in Conv $\mathbb{R}^n$. Assume that there is $x_0$ at which $\sup_{j \in J} f_j(x_0) < +\infty$. Then,*

$$\text{for } f := \sup_{j \in J} f_j, \qquad \text{we have } f'_\infty = \sup_{j \in J} (f_j)'_\infty.$$

3. *Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be affine with linear part $A_0$, and let $f \in$ Conv $\mathbb{R}^m$. Assume that $A(\mathbb{R}^n) \cap$ dom $f \neq \emptyset$. Then $(f \circ A)'_\infty = f'_\infty \circ A_0$.*

## 2.4 First- and Second-order Differentiation

### 2.4.1 Differentiable convex functions

**Theorem 2.4.1.** *Let $f$ be a function differentiable on an open set $\Omega \subset \mathbb{R}^n$, and let $C$ be a convex subset of $\Omega$. Then*

1. *$f$ is convex on $C$ if and only if*

$$f(x) \geq f(x_0) + \langle \nabla f(x_0),\, x - x_0 \rangle \quad \textit{for all } (x_0, x) \in C \times C; \tag{4.1.1}$$

2. *$f$ is strictly convex on $C$ if and only if strict inequality holds in (4.1.1) whenever $x \neq x_0$;*

3. *$f$ is strongly convex with modulus $c$ on $C$ if and only if, for all $(x_0, x) \in C \times C$,*

$$f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \tfrac{1}{2} c \|x - x_0\|^2. \tag{4.1.2}$$

*(i).* Let $f$ be convex on $C$: for arbitrary $(x_0, x) \in C \times C$ and $\alpha \in ]0, 1[$, we have from the definition (1.1.1) of convexity

$$f(\alpha x + (1 - \alpha)x_0) - f(x_0) \leq \alpha[f(x) - f(x_0)].$$

Divide by $\alpha$ and let $\alpha \downarrow 0$: observing that $\alpha x + (1 - \alpha)x_0 = x_0 + \alpha(x - x_0)$, the lefthand side tends to $\langle \nabla f(x_0), x - x_0 \rangle$ and (4.1.1) is established.

Conversely, take $x_1$ and $x_2$ in $C$, $\alpha \in ]0, 1[$ and set $x_0 := \alpha x_1 + (1-\alpha)x_2 \in C$. By assumption,

$$f(x_i) \geq f(x_0) + \langle \nabla f(x_0), x_i - x_0 \rangle \quad \text{for } i = 1, 2 \tag{4.1.3}$$

and we obtain by convex combination

$$\alpha f(x_1) + (1 - \alpha) f(x_2) \geq f(x_0) + \langle \nabla f(x_0), \alpha x_1 + (1 - \alpha)x_2 - x_0 \rangle$$

which, after simplification, is just the relation of definition (1.1.1).

[(ii)] If $f$ is strictly convex, we have for $x_0 \neq x$ in $C$ and $\alpha \in ]0, 1[$,

$$f(x_0 + \alpha(x - x_0)) - f(x_0) < \alpha[f(x) - f(x_0)];$$

but $f$ is in particular convex and we can use (i):

$$\langle \nabla f(x_0), \alpha(x - x_0) \rangle \leq f(x_0 + \alpha(x - x_0)) - f(x_0),$$

so the required strict inequality follows.

For the converse, proceed as for (i), starting from strict inequalities in (4.1.3).

[(iii)] Using Proposition 1.1.2, just apply (i) to the function $f - \frac{1}{2} c \| \cdot \|^2$, which is of course differentiable. $\square$

**Definition 2.4.1.** Let $C \subset \mathbb{R}^n$ be convex. The mapping $F : C \to \mathbb{R}^n$ is said to be *monotone* [resp. strictly monotone, resp. strongly monotone with modulus $c > 0$] on $C$ when, for all $x$ and $x'$ in $C$,

$$\langle F(x) - F(x'),\, x - x' \rangle \geq 0$$

[resp. $\langle F(x) - F(x'),\, x - x' \rangle > 0$ whenever $x \neq x'$, resp. $\langle F(x) - F(x'),\, x - x' \rangle \geq c\|x - x'\|^2$].

**Theorem 2.4.2.** *Let $f$ be a function differentiable on an open set $\Omega \subset \mathbb{R}^n$, and let $C$ be a convex subset of $\Omega$. Then, $f$ is convex [resp. strictly convex, resp. strongly convex with modulus $c$] on $C$ if and only if its gradient $\nabla f$ is monotone [resp. strictly monotone, resp. strongly monotone with modulus $c$] on $C$.*

*Proof.* We combine the "convex $\Leftrightarrow$ monotone" and "strongly convex $\Leftrightarrow$ strongly monotone" cases by accepting the value $c = 0$ in the relevant relations such as (4.1.2).

Thus, let $f$ be [strongly] convex on $C$: in view of Theorem 4.1.1, we can write for arbitrary $x_0$ and $x$ in $C$:

$$f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \tfrac{1}{2} c \|x - x_0\|^2$$

$$f(x_0) \geq f(x) + \langle \nabla f(x), x_0 - x \rangle + \tfrac{1}{2} c \|x_0 - x\|^2,$$

and mere addition shows that $\nabla f$ is [strongly] monotone.

Conversely, let $(x_0, x_1)$ be a pair of elements in $C$. Consider the univariate function $t \mapsto \varphi(t) := f(x_t)$, where $x_t := x_0 + t(x_1 - x_0)$; for $t$ in an open interval containing $[0, 1]$, $x_t \in \Omega$ and $\varphi$ is well-defined and differentiable; its derivative at $t$ is $\varphi'(t) = \langle \nabla f(x_t), x_1 - x_0 \rangle$. Thus, we have for all $0 \leq t' < t \leq 1$

$$\varphi'(t) - \varphi'(t') = \langle \nabla f(x_t) - \nabla f(x_{t'}),\, x_1 - x_0 \rangle = \frac{1}{t - t'} \langle \nabla f(x_t) - \nabla f(x_{t'}),\, x_t - x_{t'} \rangle \tag{4.1.4}$$

and the monotonicity relation for $\nabla f$ shows that $\varphi'$ is increasing, $\varphi$ is therefore convex (Corollary 0.6.5).

For strong convexity, set $t' = 0$ in (4.1.4) and use the strong monotonicity relation for $\nabla f$:

$$\varphi'(t) - \varphi'(0) \geq \frac{1}{t} c \|x_t - x_0\|^2 = tc \|x_1 - x_0\|^2. \tag{4.1.5}$$

Because the differentiable convex function $\varphi$ is the integral of its derivative, we can write

$$\varphi(1) - \varphi(0) - \varphi'(0) = \int_0^1 [\varphi'(t) - \varphi'(0)]\, dt \geq \tfrac{1}{2} c \|x_1 - x_0\|^2$$

which, by definition of $\varphi$, is just (4.1.2) (the coefficient $1/2$ is $\int_0^1 t\, dt$!).

The same technique proves the "strictly monotone $\Leftrightarrow$ strictly convex" case; then, (4.1.5) becomes a strict inequality — with $c = 0$ — and remains so after integration. $\square$

## 2.4.2 Nondifferentiable convex functions

**Proposition 2.4.1** (Proposition 4.2.1)**.** *For $f \in \operatorname{Conv} \mathbb{R}^n$ and $x \in \operatorname{int} dom\, f$, the three statements below are equivalent:*

1. *The function*
$$\mathbb{R}^n \ni d \mapsto \lim_{t \downarrow 0} \frac{f(x+td)-f(x)}{t}$$
*is linear in d;*

2. *for some basis of* $\mathbb{R}^n$ *in which* $x = (\xi^1, \ldots, \xi^n)$, *the partial derivatives* $\frac{\partial f}{\partial \xi^i}(x)$ *exist at* $x$, *for* $i = 1, \ldots, n$;

3. $f$ *is differentiable at* $x$.

*Proof.* First of all, remember from Theorem 0.6.3 that the one-dimensional function $t \mapsto f(x+td)$ has half-derivatives at 0: the limits considered in (i) exist for all $d$. We will denote by $\{b_1, \ldots, b_n\}$ the basis postulated in (ii), so that $x = \sum_{i=1}^n \xi^i b_i$.

Denote by $d \mapsto \ell(d)$ the function defined in (i); taking $d = \pm b_i$, realize that, when (i) holds,
$$\lim_{\tau \downarrow 0} \frac{f(x+\tau b_i)-f(x)}{-\tau} = \ell(-b_i) = -\ell(b_i) = -\lim_{t \downarrow 0} \frac{f(x+tb_i)-f(x)}{t}.$$

This means that the two half-derivatives at $t = 0$ of the function $t \mapsto f(x + tb_i)$ coincide: the partial derivative of $f$ at $x$ along $b_i$ exists, (ii) holds. That (iii) implies (i) is clear: when $f$ is differentiable at $x$,
$$\lim_{t \downarrow 0} \frac{f(x+td)-f(x)}{t} = \langle \nabla f(x), d \rangle.$$

We do not really complete the proof here, because everything follows in a straightforward way from subsequent chapters. More precisely, [(ii) $\Rightarrow$ (i)] is Proposition C.1.1.6, which states that the function $\ell$ is linear on the space generated by the $b_i$'s, whenever it is linear along each $b_i$. Finally [(i) $\Rightarrow$ (iii)] results from Lemma D.2.1.1 and the proof goes as follows. One of the possible definitions of (iii) is:
$$\lim_{t \downarrow 0,\ d' \to d} \frac{f(x+td')-f(x)}{t} \quad \text{is linear in } d.$$

Because $f$ is locally Lipschitzian, the above limit exists whenever it exists for fixed $d' = d$—i.e. the expression in (i). □

**Theorem 2.4.3.** *Let* $f \in \operatorname{Conv} \mathbb{R}^n$. *The subset of* $\operatorname{int} \operatorname{dom} f$ *where* $f$ *fails to be differentiable is of zero (Lebesgue) measure.*

### 2.4.3 Second-order differentiation

**Theorem 2.4.4.** *Let* $f$ *be twice differentiable on an open convex set* $\Omega \subset \mathbb{R}^n$. *Then* *(i) $f$ is convex on $\Omega$ if and only if $\nabla^2 f(x_0)$ is positive semi-definite for all $x_0 \in \Omega$;*

*(ii) if $\nabla^2 f(x_0)$ is positive definite for all $x_0 \in \Omega$, then $f$ is strictly convex on $\Omega$;*

*(iii) $f$ is strongly convex with modulus $c$ on $\Omega$ if and only if the smallest eigenvalue of $\nabla^2 f(\cdot)$ is minorized by $c$ on $\Omega$: for all $x_0 \in \Omega$ and all $d \in \mathbb{R}^n$,*

$$\langle \nabla^2 f(x_0)d, d\rangle \geq c\|d\|^2.$$

*Proof.* For given $x_0 \in \Omega$, $d \in \mathbb{R}^n$ and $t \in \mathbb{R}$ such that $x_0 + td \in \Omega$, we will set

$$x_t := x_0 + td \quad \text{and} \quad \varphi(t) := f(x_t) = f(x + td),$$

so that $\varphi''(t) = \langle \nabla^2 f(x_t)d, d\rangle$.

[(i)] Assume $f$ is convex on $\Omega$; let $(x_0, d)$ be arbitrary in $\Omega \times \mathbb{R}^n$, with $d \neq 0$: $\varphi$ is then convex on the open interval $I := \{t \in \mathbb{R} : x_0 + td \in \Omega\}$. It follows

$$0 \leq \varphi''(t) = \langle \nabla^2 f(x_t)d, d\rangle \quad \text{for all } t \in I \ni 0$$

and $\nabla^2 f(x_0)$ is positive semi-definite.

Conversely, take an arbitrary $[x_0, x_1] \subset \Omega$, set $d := x_1 - x_0$ and assume $\nabla^2 f(x_t)$ positive semi-definite, i.e. $\varphi''(t) \geq 0$, for $t \in [0, 1]$. Then Theorem 0.6.6 tells us that $\varphi$ is convex on $[0, 1]$, i.e. $f$ is convex on $[x_0, x_1]$. The result follows since $x_0$ and $x_1$ were arbitrary in $\Omega$.

[(ii)] To establish the strict convexity of $f$ on $\Omega$, we prove that $\nabla f$ is strictly monotone on $\Omega$: Theorem 4.1.4 will apply. As above, take an arbitrary $[x_0, x_1] \subset \Omega$, $x_1 \neq x_0$, $d := x_1 - x_0$, and apply the mean-value theorem to the function $\varphi'$, differentiable on $[0, 1]$, for some $\tau \in ]0, 1[$,

$$\varphi'(1) - \varphi'(0) = \varphi''(\tau) = \langle \nabla^2 f(x_\tau)d, d\rangle > 0$$

and the result follows since

$$\varphi'(1) - \varphi'(0) = \langle \nabla f(x_1) - \nabla f(x_0), x_1 - x_0\rangle.$$

[(iii)] Using Proposition 1.1.2, apply (i) to the function $f - \frac{1}{2}c\| \cdot \|^2$, whose Hessian operator is $\nabla^2 f - cI_n$ and has the eigenvalues $\lambda - c$, with $\lambda$ describing the eigenvalues of $\nabla^2 f$. $\square$

# Chapter 3

# Sublinearity and Support Functions

## 3.1 Sublinear Functions

### 3.1.1 Definitions and first properties

**Definition 3.1.1.** A function $\sigma : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to be *sublinear* if it is convex and positively homogeneous (of degree 1): $\sigma \in \operatorname{Conv} \mathbb{R}^n$ and

$$\sigma(tx) = t\sigma(x) \qquad \text{for all } x \in \mathbb{R}^n \text{ and } t > 0. \tag{1.1.1}$$

**Proposition 3.1.1.** *A function $\sigma : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is sublinear if and only if its epigraph* $\operatorname{epi} \sigma$ *is a nonempty convex cone in $\mathbb{R}^n \times \mathbb{R}$.*

*Proof.* We know that $\sigma$ is a convex function if and only if $\operatorname{epi} \sigma$ is a nonempty convex set in $\mathbb{R}^n \times \mathbb{R}$ (Proposition B.1.1.6). Therefore, we just have to prove the equivalence between positive homogeneity and $\operatorname{epi} \sigma$ being a cone.

Let $\sigma$ be positively homogeneous. For $(x, r) \in \operatorname{epi} \sigma$, the relation $\sigma(x) \leq r$ gives

$$\sigma(tx) = t\sigma(x) \leq tr \qquad \text{for all } t > 0,$$

so $\operatorname{epi} \sigma$ is a cone. Conversely, if $\operatorname{epi} \sigma$ is a cone in $\mathbb{R}^n \times \mathbb{R}$, the property $(x, \sigma(x)) \in \operatorname{epi} \sigma$ implies $(tx, t\sigma(x)) \in \operatorname{epi} \sigma$, i.e.

$$\sigma(tx) \leq t\sigma(x) \qquad \text{for all } t > 0.$$

From Remark 1.1.2, this is just positive homogeneity. □

**Proposition 3.1.2.** *A function $\sigma : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, not identically equal to $+\infty$, is sublinear if and only if one of the following two properties holds:*

$$\sigma(t_1 x_1 + t_2 x_2) \leq t_1 \sigma(x_1) + t_2 \sigma(x_2) \qquad \text{for all } x_1, x_2 \in \mathbb{R}^n \text{ and } t_1, t_2 > 0, \tag{1.1.4}$$

*or*

$$\sigma \textit{ is positively homogeneous and subadditive.} \tag{1.1.5}$$

*sublinearity $\implies$ (1.1.4).* For $x_1, x_2 \in \mathbb{R}^n$ and $t_1, t_2 > 0$, set $t := t_1 + t_2 > 0$; we have

$$\begin{aligned}\sigma(t_1 x_1 + t_2 x_2) &= \sigma\big(t\big[\tfrac{t_1}{t}x_1 + \tfrac{t_2}{t}x_2\big]\big)\\ &= t\sigma\big(\tfrac{t_1}{t}x_1 + \tfrac{t_2}{t}x_2\big) \qquad \text{[positive homogeneity]}\\ &\le t\big[\tfrac{t_1}{t}\sigma(x_1) + \tfrac{t_2}{t}\sigma(x_2)\big] \qquad \text{[convexity]},\end{aligned}$$

and (1.1.4) is proved.

[(1.1.4) $\implies$ (1.1.5)] A function satisfying (1.1.4) is obviously subadditive (take $t_1 = t_2 = 1$) and satisfies (take $x_1 = x_2 = x$, $t_1 = t_2 = 1/2t$)

$$\sigma(tx) = t\sigma(x),$$

i.e. it is positively homogeneous.

[(1.1.5)$\Rightarrow$ sublinearity] Take $t_1, t_2 > 0$ with $t_1 + t_2 = 1$ and apply successively subadditivity and positive homogeneity:

$$\sigma(t_1 x_1 + t_2 x_2) \le \sigma(t_1 x_1) + \sigma(t_2 x_2) = t_1\sigma(x_1) + t_2\sigma(x_2),$$

hence $\sigma$ is convex. $\square$

**Corollary 3.1.1.** *If $\sigma$ is sublinear, then*

$$\sigma(x) + \sigma(-x) \ge 0 \quad \textit{for all } x \in \mathbb{R}^n. \tag{1.1.6}$$

*Proof.* Take $x_2 = -x_1$ in (1.1.3) and remember that $\sigma(0) \ge 0$. $\square$

**Proposition 3.1.3.** *Let $\sigma$ be sublinear and suppose that there exist $x_1, \dots, x_m$ in dom $\sigma$ such that*

$$\sigma(x_j) + \sigma(-x_j) = 0 \quad \textit{for } j = 1, \dots, m. \tag{1.1.7}$$

*Then $\sigma$ is linear on the subspace spanned by $x_1, \dots, x_m$.*

*Proof.* With $x_1, \dots, x_m$ as stated, each $-x_j$ is in dom $\sigma$. Let $x := \sum_{j=1}^m t_j x_j$ be an arbitrary linear combination of $x_1, \dots, x_m$; we must prove that $\sigma(x) = \sum_{j=1}^m t_j \sigma(x_j)$. Set

$$J_1 := \{j : t_j > 0\}, \qquad J_2 := \{j : t_j < 0\},$$

and obtain (as usual, $\sum_\emptyset = 0$):

$$\begin{aligned}
\sigma(x) &= \sigma\Big(\sum_{j\in J_1} t_j x_j + \sum_{j\in J_2}(-t_j)(-x_j)\Big) && (3.1)\\
&\le \sum_{j\in J_1} t_j\sigma(x_j) + \sum_{j\in J_2}(-t_j)\sigma(-x_j) && \text{[from (1.1.4)]}\\
&= \sum_{j\in J_1} t_j\sigma(x_j) + \sum_{j\in J_2} t_j\sigma(x_j) = \sum_{j=1}^{m} t_j\sigma(x_j) && \text{[from (1.1.7)]}\\
&= -\sum_{j\in J_1} t_j\sigma(-x_j) - \sum_{j\in J_2}(-t_j)\sigma(x_j) && \text{[from (1.1.7)]}\\
&\le -\sigma\big(-\sum_{j=1}^{m} t_j x_j\big) && \text{[from (1.1.4)]}\\
&= -\sigma(-x) \le \sigma(x). && \text{[from (1.1.6)]}
\end{aligned}$$

In summary, we have proved $\sigma(x) \le \sum_{j=1}^{m} t_j\sigma(x_j) \le -\sigma(-x) \le \sigma(x)$. □

**Proposition 3.1.4.** *Let $\sigma$ be sublinear. If $x \in U$, i.e. if*

$$\sigma(x) + \sigma(-x) = 0,$$

*then there holds*

$$\sigma(x+y) = \sigma(x) + \sigma(y) \qquad \text{for all } y \in \mathbb{R}^n.$$

*Proof.* In view of subadditivity, we just have to prove "≥" in (1.1.10). Start from the identity $y = x + y - x$; apply successively subadditivity and (1.1.9) to obtain

$$\sigma(y) \le \sigma(x+y) + \sigma(-x) = \sigma(x+y) - \sigma(x).$$

□

### 3.1.2 Some examples

**Definition 3.1.2.** (Gauge) Let $C$ be a closed convex set containing the origin. The function $\gamma_C$ defined by

$$\gamma_C(x) := \inf\{\lambda > 0 : x \in \lambda C\} \tag{1.2.2}$$

is called the *gauge* of $C$. As usual, we set $\gamma_C(x) := +\infty$ if $x \notin \lambda C$ for no $\lambda > 0$.

**Theorem 3.1.1.** *Let $C$ be a closed convex set containing the origin. Then*

*(i) its gauge $\gamma_C$ is a nonnegative closed sublinear function;*

*(ii) $\gamma_C$ is finite everywhere if and only if 0 lies in the interior of $C$;*

*(iii)* $C_\infty$ *being the asymptotic cone of* $C$*,*

$$\{x \in \mathbb{R}^n : \gamma_C(x) \leq r\} = rC \quad \textit{for all } r > 0, \qquad \{x \in \mathbb{R}^n : \gamma_C(x) = 0\} = C_\infty .$$

*(i) and (iii).* Nonnegativity and positive homogeneity are obvious from the definition of $\gamma_C$; also, $\gamma_C(0) = 0$ because $0 \in C$. We prove convexity via a geometric interpretation of (1.2.2). Let

$$K_C := \operatorname{cone}(C \times \{1\}) = \{(\lambda c, \lambda) \in \mathbb{R}^n \times \mathbb{R} : \ c \in C, \ \lambda \geq 0\}$$

be the convex conical hull of $C \times \{1\} \subset \mathbb{R}^n \times \mathbb{R}$. It is convex (beware that $K_C$ need not be closed) and $\gamma_C$ is clearly given by

$$\gamma_C(x) = \inf\{\lambda : \ (x, \lambda) \in K_C\}.$$

Thus, $\gamma_C$ is the lower-bound function of §B.1.3(g), constructed on the convex set $K_C$; this establishes the convexity of $\gamma_C$, hence its sublinearity.

Now we prove

$$\{x \in \mathbb{R}^n : \ \gamma_C(x) \leq 1\} = C. \tag{3.2}$$

This will imply the first part in (iii), thanks to positive homogeneity. Then the second part will follow because of (A.2.2.2): $C_\infty = \cap\{rC : \ r > 0\}$ and closedness of $\gamma_C$ will also result from (iii) via Proposition B.1.2.2.

So, to prove (1.2.3), observe first that $x \in C$ implies from (1.2.2) that certainly $\gamma_C(x) \leq 1$. Conversely, let $x$ be such that $\gamma_C(x) \leq 1$; we must prove that $x \in C$. For this we prove that $x_k := (1 - 1/k)x \in C$ for $k = 1, 2, \dots$ (and then, the desired property will come from the closedness of $C$). By positive homogeneity, $\gamma_C(x_k) = (1 - 1/k)\gamma_C(x) \leq 1$, so there is $\lambda_k \in [0, 1]$ such that $x_k = \lambda_k C$, or equivalently $x_k/\lambda_k \in C$. Because $C$ is convex and contains the origin, $\lambda_k(x_k/\lambda_k) + (1 - \lambda_k)0 = x_k$ is in $C$, which is what we want. $\square$

*(ii).* Assume $0 \in \operatorname{int} C$. There is $\varepsilon > 0$ such that for all $x \neq 0$, $x_\varepsilon := x/\|x\| \in C$; hence $\gamma_C(x_\varepsilon) \leq 1$ because of (1.2.3). We deduce by positive homogeneity

$$\gamma_C(x) = \frac{\|x\|}{\varepsilon}\, \gamma_C(x_\varepsilon) \leq \frac{\|x\|}{\varepsilon} \,;$$

this inequality actually holds for all $x \in \mathbb{R}^n$ ($\gamma_C(0) = 0$) and $\gamma_C$ is a finite function.

Conversely, suppose $\gamma_C$ is finite everywhere. By continuity (Theorem B.3.1.2), $\gamma_C$ has an upper bound $L > 0$ on the unit ball:

$$\|x\| \leq 1 \quad \Longrightarrow \quad \gamma_C(x) \leq L \quad \Longrightarrow \quad x \in LC,$$

where the last implication comes from (iii). In other words, $B(0, 1/L) \subset C$. $\square$

**Corollary 3.1.2.** *C is compact if and only if* $\gamma_C(x) > 0$ *for all* $x \neq 0$*.*

### 3.1.3 The convex cone of all closed sublinear functions

**Proposition 3.1.5.** *(i) If $\sigma_1$ and $\sigma_2$ are [closed] sublinear and $t_1, t_2$ are positive numbers, then $\sigma := t_1\sigma_1 + t_2\sigma_2$ is [closed] sublinear, if not identically $+\infty$.*

*(ii) If $\{\sigma_j\}_{j\in J}$ is a family of [closed] sublinear functions, then $\sigma := \sup_{j\in J} \sigma_j$ is [closed] sublinear, if not identically $+\infty$.*

*Proof.* Concerning convexity and closedness, everything is known from §B.2. Note in passing that a closed sublinear function is zero (hence finite) at zero. As for positive homogeneity, it is straightforward. □

**Proposition 3.1.6.** *Let $\{\sigma_j\}_{j\in J}$ be a family of sublinear functions all minorized by some linear function. Then*

*(i) $\sigma := \operatorname{co}\big(\inf_{j\in J} \sigma_j\big)$ is sublinear.*

*(ii) If $J = \{1, \ldots, m\}$ is a finite set, we obtain the infimal convolution*

$$\operatorname{co}\min\{\sigma_1, \ldots, \sigma_m\} = \sigma_1 \overset{+}{\vee} \cdots \overset{+}{\vee} \sigma_m.$$

*(i).* Once again, the only thing to prove for (i) is positive homogeneity. Actually, it suffices to multiply $x$ and each $x_j$ by $t > 0$ in a formula giving $\operatorname{co}\big(\inf_j \sigma_j\big)(x)$, say (B.2.5.3).

[(ii)] By definition, computing $\operatorname{co}\big(\min_j \sigma_j\big)(x)$ amounts to solving the minimization problem in the $m$ couples of variables $(x_j, \alpha_j) \in \operatorname{dom}\sigma_j \times \mathbb{R}$

$$\begin{aligned} &\inf \sum_{j=1}^{m} \alpha_j \sigma_j(x_j) \qquad \alpha_j \geq 0 \\ &\sum_{j=1}^{m} \alpha_j = 1, \ \sum_{j=1}^{m} \alpha_j x_j = x. \end{aligned} \tag{1.3.1}$$

In view of positive homogeneity, the variables $\alpha_j$ play no role by themselves: the relevant variables are actually the products $\alpha_j x_j$ and (1.3.1) can be written – denoting $\alpha_j x_j$ again by $x_j$:

$$\operatorname{co}\big(\min_j \sigma_j\big)(x) = \inf\Big\{\sum_{j=1}^{m} \sigma_j(x_j) : \sum_{j=1}^{m} x_j = x\Big\}.$$

We recognize the infimal convolution of the $\sigma_j$'s. □

**Theorem 3.1.2.** *For $\sigma_1$ and $\sigma_2$ in the set $\Phi$ of sublinear functions that are finite everywhere, define*

$$\Delta(\sigma_1, \sigma_2) := \max_{\|x\|\leq 1} |\sigma_1(x) - \sigma_2(x)|. \tag{1.3.2}$$

*Then $\Delta$ is a distance on $\Phi$.*

*Proof.* Clearly $\Delta(\sigma_1, \sigma_2) < +\infty$ and $\Delta(\sigma_1, \sigma_2) = \Delta(\sigma_2, \sigma_1)$. Now positive homogeneity of $\sigma_1$ and $\sigma_2$ gives for all $x \neq 0$

$$|\sigma_1(x) - \sigma_2(x)| = \|x\| \left| \sigma_1\left(\frac{x}{\|x\|}\right) - \sigma_2\left(\frac{x}{\|x\|}\right) \right| \leq \|x\| \max_{\|u\|=1} |\sigma_1(u) - \sigma_2(u)| \leq \|x\| \, \Delta(\sigma_1, \sigma_2).$$

In addition, $\sigma_1(0) = \sigma_2(0) = 0$, so

$$|\sigma_1(x) - \sigma_2(x)| \leq \|x\| \, \Delta(\sigma_1, \sigma_2) \qquad \text{for all } x \in \mathbb{R}^n$$

and $\Delta(\sigma_1, \sigma_2) = 0$ if and only if $\sigma_1 = \sigma_2$.

As for the triangle inequality, we have for arbitrary $\sigma_1, \sigma_2, \sigma_3$ in $\Phi$

$$|\sigma_1(x) - \sigma_3(x)| \leq |\sigma_1(x) - \sigma_2(x)| + |\sigma_2(x) - \sigma_3(x)| \qquad \text{for all } x \in \mathbb{R}^n,$$

so there holds

$$\Delta(\sigma_1, \sigma_3) \leq \max_{\|x\|=1} \big[ |\sigma_1(x) - \sigma_2(x)| + |\sigma_2(x) - \sigma_3(x)| \big] \leq \max_{\|x\|=1} |\sigma_1(x) - \sigma_2(x)| + \max_{\|x\|=1} |\sigma_2(x) - \sigma_3(x)|,$$

which is the required inequality. ☐

**Theorem 3.1.3.** *Let $(\sigma_k)$ be a sequence of finite sublinear functions and let $\sigma$ be a finite function. Then the following are equivalent when $k \to +\infty$:*

*(i)* $(\sigma_k)$ *converges pointwise to* $\sigma$*;*

*(ii)* $(\sigma_k)$ *converges to* $\sigma$ *uniformly on each compact set of* $\mathbb{R}^n$*;*

*(iii)* $\Delta(\sigma_k, \sigma) \to 0$.

*Proof.* First, the (finite) function $\sigma$ is of course sublinear whenever it is the pointwise limit of sublinear functions. The equivalence between (i) and (ii) comes from the general Theorem B.3.1.4 on the convergence of convex functions.

Now, (ii) clearly implies (iii). Conversely $\Delta(\sigma_k, \sigma) \to 0$ is the uniform convergence on the unit ball, hence on any ball of radius $L > 0$ (the maxmind in (1.3.2) is positively homogeneous), hence on any compact set. ☐

# 3.2 The Support Function of a Nonempty Set

## 3.2.1 Definitions, Interpretations

**Definition 3.2.1.** ( Support Function) Let $S$ be a nonempty set in $\mathbb{R}^n$. The function

$$\sigma_S : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$$

defined by

$$\mathbb{R}^n \ni x \mapsto \sigma_S(x) := \sup\{\langle s, x \rangle : \; s \in S\} \tag{2.1.1}$$

is called the *support function* of $S$.

**Proposition 3.2.1.** *A support function is closed and sublinear.*

*Proof.* This results from Proposition 1.3.1(ii) (a linear form is closed and convex!). Observe in particular that a support function is null (hence $< +\infty$) at the origin. $\square$

**Proposition 3.2.2.** *The support function of $S$ is finite everywhere if and only if $S$ is bounded.*

*Proof.* Let $S$ be bounded, say $S \subset B(0, L)$ for some $L > 0$. Then

$$\langle s, x\rangle \leq \|s\| \, \|x\| \leq L\|x\| \qquad \text{for all } s \in S,$$

which implies $\sigma_S(x) \leq L\|x\|$ for all $x \in \mathbb{R}^n$.

Conversely, finiteness of the convex $\sigma_S$ implies its continuity on the whole space (Theorem B.3.1.2), hence its local boundedness: for some $L$,

$$\langle s, x\rangle \leq \sigma_S(x) \leq L \qquad \text{for all } (s, x) \in S \times B(0, 1).$$

If $s \neq 0$, we can take $x = s/\|s\|$ in the above relation, which implies $\|s\| \leq L$. $\square$

**Definition 3.2.2.** (Breadth of a Set) The *breadth* of a nonempty set $S$ along $x \neq 0$ is

$$\sigma_S(x) + \sigma_S(-x) = \sup_{s \in S} \langle s, x\rangle - \inf_{s \in S} \langle s, x\rangle,$$

a number in $[0, +\infty]$. It is 0 if and only if $S$ lies entirely in some affine hyperplane orthogonal to $x$; such a hyperplane is expressed as

$$\{y \in \mathbb{R}^n : \ \langle y, x\rangle = \sigma_S(x)\},$$

which in particular contains $S$. The intersection of all these hyperplanes is just the affine hull of $S$.

### 3.2.2 Basic properties

**Proposition 3.2.3.** *For $S \subset \mathbb{R}^n$ nonempty, there holds $\sigma_S = \sigma_{\mathrm{cl}\, S} = \sigma_{\mathrm{co}\, S}$; whence*

$$\sigma_S = \sigma_{\overline{\mathrm{co}}\, S}. \tag{2.2.1}$$

*Proof.* The continuity [resp. linearity, hence convexity] of the function $\langle s, \cdot\rangle$, which is maximized over $S$, implies that $\sigma_S = \sigma_{\mathrm{cl}\, S}$ [resp. $\sigma_S = \sigma_{\mathrm{co}\, S}$]. Knowing that $\overline{\mathrm{co}}\, S = \mathrm{cl}\ \mathrm{co}\ S$ (Proposition A.1.4.2), (2.2.1) follows immediately. $\square$

**Theorem 3.2.1.** *For the nonempty $S \subset \mathbb{R}^n$ and its support function $\sigma_S$, there holds*

$$s \in \overline{\mathrm{co}} S \iff \big[\langle s, d\rangle \leq \sigma_S(d) \quad \textit{for all } d \in X\big], \tag{2.2.2}$$

*where the set $X$ can be indifferently taken as: the whole of $\mathbb{R}^n$, the unit ball $B(0, 1)$ or its boundary the unit sphere $\tilde{B}$, or* $\mathrm{dom}\, \sigma_S$.

*Proof.* First, the equivalence between all the choices for $X$ is clear enough; in particular due to positive homogeneity. Because "$\Rightarrow$" is Proposition 2.2.1, we have to prove "$\Leftarrow$" only, with $X = \mathbb{R}^n$ say.

So suppose that $s \notin \overline{\text{co}}S$. Then $\{s\}$ and $\overline{\text{co}}S$ can be strictly separated (Theorem A.4.1.1): there exists $d_0 \in \mathbb{R}^n$ such that

$$\langle s, d_0\rangle > \sup\{\langle s', d_0\rangle : s' \in \overline{\text{co}}S\} = \sigma_S(d_0),$$

where the last equality is (2.2.1). Our result is proved by contradiction. $\square$

**Theorem 3.2.2.** *Let $S$ be a nonempty closed convex set in $\mathbb{R}^n$. Then*

*(i) $s \in \operatorname{aff} S$ if and only if*

$$\langle s, d\rangle = \sigma_S(d) \quad \textit{for all } d \textit{ with } \sigma_S(d) + \sigma_S(-d) = 0; \tag{2.2.3}$$

*(ii) $s \in \operatorname{ri} S$ if and only if*

$$\langle s, d\rangle < \sigma_S(d) \quad \textit{for all } d \textit{ with } \sigma_S(d) + \sigma_S(-d) > 0; \tag{2.2.4}$$

*(iii) in particular, $s \in \operatorname{int} S$ if and only if*

$$\langle s, d\rangle < \sigma_S(d) \quad \textit{for all } d \neq 0. \tag{2.2.5}$$

*(i).* Let first $s \in S$. We have already seen in Definition 2.1.4 that

$$-\sigma_S(-d) \le \langle s, d\rangle \le \sigma_S(d) \qquad \text{for all } d \in \mathbb{R}^n.$$

If the breadth of $S$ along $d$ is zero, we obtain a pair of equalities: for such $d$, there holds

$$\langle s, d\rangle = \sigma_S(d),$$

an equality which extends by affine combination to any element $s \in \operatorname{aff} S$.

Conversely, let $s$ satisfy (2.2.3). A first case is when the only $d$ described in (2.2.3) is $d = 0$; as a consequence of our observations in Definition 2.1.4, there is no affine hyperplane containing $S$, i.e. $\operatorname{aff} S = \mathbb{R}^n$ and there is nothing to prove. Otherwise, there does exist a hyperplane $H$ containing $S$; it is defined by

$$H := \{p \in \mathbb{R}^n : \langle p, d_H\rangle = \sigma_S(d_H)\}, \tag{2.2.6}$$

for some $d_H \neq 0$. We proceed to prove $\langle s, \cdot\rangle \le \sigma_H$.

In fact, the breadth of $S$ along $d_H$ is certainly 0, hence $\langle s, d_H\rangle = \sigma_S(d_H)$ because of (2.2.3), while (2.2.6) shows that $\sigma_S(d_H) = \sigma_H(d_H)$. On the other hand, it is obvious that $\sigma_H(d) = +\infty$ if $d$ is not collinear to $d_H$. In summary, we have proved $\langle s, d\rangle \le \sigma_H(d)$ for all $d$, i.e. $s \in H$. We conclude that our $s$ is in any affine manifold containing $S$: $s \in \operatorname{aff} S$.

[(iii)] In view of positive homogeneity, we can normalize $d$ in (2.2.5). For $s \in \operatorname{int} S$, there exists $\varepsilon > 0$ such that $s + \varepsilon d \in S$ for all $d$ in the unit sphere $\widetilde{B}$. Then, from the very definition (2.1.1),

$$\sigma_S(d) \ge \langle s + \varepsilon d, d\rangle = \langle s, d\rangle + \varepsilon \quad \text{for all } d \in \widetilde{B}.$$

Conversely, let $s \in \mathbb{R}^n$ be such that

$$\sigma_S(d) - \langle s, d \rangle > 0 \quad \text{for all } d \in \widetilde{B},$$

which implies, because $\sigma_S$ is closed and the unit sphere is compact:

$$0 < \varepsilon := \inf\{\sigma_S(d) - \langle s, d \rangle : d \in \widetilde{B}\} \leq +\infty.$$

Thus

$$\langle s, d \rangle + \varepsilon \leq \sigma_S(d) \quad \text{for all } d \in \widetilde{B}.$$

Now take $u$ with $\|u\| < \varepsilon$. From the Cauchy-Schwarz inequality, we have for all $d \in \widetilde{B}$

$$\langle s + u, d \rangle = \langle s, d \rangle + \langle u, d \rangle \leq \langle s, d \rangle + \varepsilon \leq \sigma_S(d)$$

and this implies $s + u \in S$ because of Theorem 2.2.2: $s \in \operatorname{int} S$ and (iii) is proved.

[(ii)] Look at Fig. 2.2.2 again: decompose $\mathbb{R}^n = V \oplus U$, where $V$ is the subspace parallel to $\operatorname{aff} S$ and $U = V^\perp$. In the decomposition $d = d_V + d_U$, $\langle \cdot, d_U \rangle$ is constant over $S$, so $S$ has 0-breadth along $d_U$ and

$$\sigma_S(d) = \sup_{s \in S} \langle s, d_V + d_U \rangle = \langle s, d_U \rangle + \sup_{s \in S} \langle s, d_V \rangle$$

for any $s \in S$. With these notations, a direction described as in (2.2.4) is a $d$ such that

$$\sigma_S(d) + \sigma_S(-d) = \sigma_S(d_V) + \sigma_S(-d_V) > 0.$$

Then, (ii) is just (iii) written in the subspace $V$. $\square$

**Proposition 3.2.4.** *Let $S$ be a nonempty closed convex set in $\mathbb{R}^n$. Then $\overline{\operatorname{dom} \sigma_S}$ and the asymptotic cone $S_\infty$ of $S$ are mutually polar cones.*

*Proof.* Recall from §A.3.2 that, if $K_1$ and $K_2$ are two closed convex cones, then $K_1 \subset K_2$ if and only if $(K_1)^\circ \supset (K_2)^\circ$.

Let $p \in S_\infty$. Fix $s_0$ arbitrary in $S$ and use the fact that $S_\infty = \bigcap_{t>0} t(S - s_0)$ ( §A.2.2); for all $t > 0$, we can find $s_t \in S$ such that $p = t(s_t - s_0)$. Now, for $q \in \operatorname{dom} \sigma_S$, there holds

$$\langle p, q \rangle = t \langle s_t - s_0, q \rangle \leq t\big[\sigma_S(q) - \langle s_0, q \rangle\big] < +\infty$$

and letting $t \downarrow 0$ shows that $\langle p, q \rangle \leq 0$. In other words, $\operatorname{dom} \sigma_S \subset (S_\infty)^\circ$; then $\overline{\operatorname{dom} \sigma_S} \subset (S_\infty)^\circ$ since the latter is closed.

Conversely, let $q \in (\operatorname{dom} \sigma_S)^\circ$, which is a cone, hence $tq \in (\operatorname{dom} \sigma_S)^\circ$ for any $t > 0$. Thus, given $s_0 \in S$, we have for arbitrary $p \in \operatorname{dom} \sigma_S$

$$\langle s_0 + tq, p \rangle = \langle s_0, p \rangle + t \langle q, p \rangle \leq \langle s_0, p \rangle \leq \sigma_S(p),$$

so $s_0 + tq \in S$ by virtue of Theorem 2.2.2. In other words: $q \in (S - s_0)/t$ for all $t > 0$ and $q \in S_\infty$. $\square$

### 3.2.3 Examples

## 3.3 The Isomorphism Between Closed Convex Sets and Closed Sublinear Functions

### 3.3.1 The fundamental correspondence

**Theorem 3.3.1.** *Let $\sigma$ be a closed sublinear function; then there is a linear function minorizing $\sigma$. In fact, $\sigma$ is the supremum of the linear functions minorizing it. In other words, $\sigma$ is the support function of the nonempty closed convex set*

$$S_\sigma := \{s \in \mathbb{R}^n : \langle s, d\rangle \le \sigma(d) \text{ for all } d \in \mathbb{R}^n\}. \tag{3.1.1}$$

*Proof.* Being convex, $\sigma$ is minorized by some affine function (Proposition B.1.2.1): for some $(s, r) \in \mathbb{R}^n \times \mathbb{R}$,

$$\langle s, d\rangle - r \le \sigma(d) \qquad \text{for all } d \in \mathbb{R}^n. \tag{3.1.2}$$

Because $\sigma(0) = 0$, the above $r$ is nonnegative. Also, by positive homogeneity,

$$\langle s, d\rangle - \tfrac{1}{t} r \le \sigma(d) \qquad \text{for all } d \in \mathbb{R}^n \text{ and all } t > 0.$$

Letting $t \to +\infty$, we see that $\sigma$ is actually minorized by a linear function:

$$\langle s, d\rangle \le \sigma(d) \qquad \text{for all } d \in \mathbb{R}^n. \tag{3.1.3}$$

Now observe that the minorization (3.1.3) is sharper than (3.1.2): when expressing the closed convex $\sigma$ as the supremum of all the affine functions minorizing it (Proposition B.1.2.8), we can restrict ourselves to linear functions. In other words

$$\sigma(d) = \sup\{\langle s, d\rangle : \text{ the linear } \langle s, \cdot\rangle \text{ minorizes } \sigma\};$$

in the above index-set, we just recognize $S_\sigma$. □

**Corollary 3.3.1.** *For a nonempty closed convex set $S$ and a closed sublinear function $\sigma$, the following are equivalent:*

*(i) $\sigma$ is the support function of $S$.*

*(ii) $S = \{s : \langle s, d\rangle \le \sigma(d)$ for all $d \in X\}$, where the set $X$ can be indifferently taken as: the whole of $\mathbb{R}^n$, the unit ball $B(0,1)$ or its boundary, or* $\operatorname{dom} \sigma$.

*Proof.* The case $X = \mathbb{R}^n$ is just Theorem 3.1.1. The other cases are then clear. □

**Definition 3.3.1** (Direction Exposing a Face)**.** Let $C$ be a nonempty closed convex set, with support function $\sigma$. For given $d \ne 0$, the set

$$F_C(d) := \{x \in C : \langle x, d\rangle = \sigma(d)\}$$

is called the exposed face of $C$ associated with $d$, or the face exposed by $d$.

**Proposition 3.3.1.** *For $x$ in a nonempty closed convex set $C$, it holds*

$$x \in F_C(d) \iff d \in N_C(x).$$

**Proposition 3.3.2.** *For a nonempty closed convex set $C$, it holds*

$$\operatorname{bd} C = \bigcup \{F_C(d) : \ d \in X\}$$

*where $X$ can be indifferently taken as: $\mathbb{R}^n \setminus \{0\}$, the unit sphere $\widetilde{B}$, or $\operatorname{dom} \sigma_C \setminus \{0\}$.*

*Proof.* Observe from Definition 3.1.3 that the face exposed by $d \neq 0$ does not depend on $\|d\|$. This establishes the equivalence between the first two choices for $X$. As for the third choice, it is due to the fact that $F_C(d) = \varnothing$ if $d \notin \operatorname{dom} \sigma_C$.

Now, if $x$ is interior to $C$ and $d \neq 0$, then $x + \varepsilon d \in C$ and $x$ cannot be a maximizer of $\langle \cdot, d\rangle$; $x$ is not in the face exposed by $d$. Conversely, take $x$ on the boundary of $C$. Then $N_C(x)$ contains a nonzero vector $d$; by Proposition 3.1.4, $x \in F_C(d)$. □

### 3.3.2 Example: Norms and Their Duals, Polarity

**Proposition 3.3.3.** *Let $B$ and $B^*$ be defined by (3.2.1) and (3.2.2), where $\|\cdot\|$ is a norm on $\mathbb{R}^n$. The support function of $B$ and the gauge of $B^*$ are the same function $\|\cdot\|_*$ defined by*

$$\|s\|_* := \max\{\langle s, x\rangle : \|x\| \leq 1\}. \tag{3.2.3}$$

*Furthermore, $\|\cdot\|_*$ is a norm on $\mathbb{R}^n$. The support function of its unit ball $B^*$ and the gauge of its supported set $B$ are the same function $\|\cdot\|$: there holds*

$$\|x\| = \max\{\langle s, x\rangle : \|s\|_* \leq 1\}. \tag{3.2.4}$$

*Proof.* It is a particular case of the results 3.2.4 and 3.2.5 below. □

**Proposition 3.3.4.** *Let $C$ be a closed convex set containing the origin. Its gauge $\gamma_C$ is the support function of a closed convex set containing the origin, namely*

$$C^\circ := \{s \in \mathbb{R}^n : \langle s, d\rangle \leq 1 \textit{ for all } d \in C\}, \tag{3.2.8}$$

*which defines the polar (set) of $C$.*

*Proof.* We know that $\gamma_C$ (which, by Theorem 1.2.5(i), is closed, sublinear and nonnegative) is the support function of some closed convex set containing the origin, say $D$; from (3.1.1),

$$D = \{s \in \mathbb{R}^n : \ \langle s, d\rangle \leq r \ \text{ for all } (d, r) \in \operatorname{epi} \gamma_C\}.$$

As seen in (1.2.4), $\operatorname{epi} \gamma_C$ is the closed convex conical hull of $C \times \{1\}$; we can use positive homogeneity to write

$$D = \{s \in \mathbb{R}^n : \ \langle s, d\rangle \leq 1 \ \text{ for all } d \text{ such that } \gamma_C(d) \leq 1\}.$$

In view of Theorem 1.2.5(iii), the above index-set is just $C$; in other words, $D = C^\circ$. □

**Corollary 3.3.2.** *Let $C$ be a closed convex set containing the origin. Its support function $\sigma_C$ is the gauge of $C^\circ$.*

**Proposition 3.3.5.** *Let $C$ be a nonempty compact convex set having $0$ in its interior, so that $C^\circ$ enjoys the same properties. Then, for all $d$ and $s$ in $\mathbb{R}^n$, the following statements are equivalent (the notation (3.2.9) is used)*

*(i)* $H(s)$ *is a supporting hyperplane to* $C$ *at* $d$*;*

*(ii)* $H(d)$ *is a supporting hyperplane to* $C^\circ$ *at* $s$*;*

*(iii)* $d \in \operatorname{bd} C,\ s \in \operatorname{bd} C^\circ$ *and* $\langle s, d\rangle = 1$*;*

*(iv)* $d \in C,\ s \in C^\circ$ *and* $\langle s, d\rangle = 1$.

*Proof.* Left as an exercise; the assumptions are present to make sure that every nonzero vector in $\mathbb{R}^n$ does expose a face in each set. □

### 3.3.3 Calculus with support functions

**Theorem 3.3.2.** *Let $S_1$ and $S_2$ be nonempty closed convex sets; call $\sigma_1$ and $\sigma_2$ their support functions. Then*

$$S_1 \subset S_2 \iff \sigma_1(d) \le \sigma_2(d) \text{ for all } d \in \mathbb{R}^n.$$

*Proof.* Apply the equivalence stated in Corollary 3.1.2:

$$S_1 \subset S_2 \iff s \in S_2 \text{ for all } s \in S_1$$
$$\iff \sigma_2(d) \ge \langle s, d\rangle \text{ for all } s \in S_1 \text{ and all } d \in \mathbb{R}^n$$
$$\iff \sigma_2(d) \ge \sup_{s \in S_1} \langle s, d\rangle \text{ for all } d \in \mathbb{R}^n.$$

□

**Theorem 3.3.3.** *(i) Let $\sigma_1$ and $\sigma_2$ be the support functions of the nonempty closed convex sets $S_1$ and $S_2$. If $t_1$ and $t_2$ are positive, then*

$$t_1\sigma_1 + t_2\sigma_2 \text{ is the support function of } \overline{(t_1S_1 + t_2S_2)}.$$

*(ii) Let $\{\sigma_j\}_{j\in J}$ be the support functions of the family of nonempty closed convex sets $\{S_j\}_{j\in J}$. Then*

$$\sup_{j \in J} \sigma_j \text{ is the support function of } \overline{\text{co}}\{\bigcup_{j\in J} S_j :\ j \in J\}.$$

*(iii) Let $\{\sigma_j\}_{j\in J}$ be the support functions of the family of closed convex sets $\{S_j\}_{j\in J}$. If*

$$S := \bigcap_{j\in J} S_j \neq \varnothing,$$

*then*

$$\sigma_S = \overline{\text{co}}\{\inf \sigma_j :\ j \in J\}.$$

*(i).* Call $S$ the closed convex set $\operatorname{cl}(t_1S_1 + t_2S_2)$. By definition, its support function is

$$\sigma_S(d) = \sup\{\langle t_1s_1 + t_2s_2, d\rangle : \ s_1 \in S_1,\ s_2 \in S_2\}.$$

In the above expression, $s_1$ and $s_2$ run independently in their index sets $S_1$ and $S_2$, $t_1$ and $t_2$ are positive, so

$$\sigma_S(d) = t_1 \sup_{s\in S_1} \langle s, d\rangle + t_2 \sup_{s\in S_2} \langle s, d\rangle.$$

[(ii)] The support function of $S := \bigcup_{j\in J} S_j$ is

$$\sup_{s\in\bigcup_{j\in J} S_j} \langle s, d\rangle = \sup_{j\in J} \sup_{s_j\in S_j} \langle s_j, d\rangle = \sup_{j\in J} \sigma_j(d).$$

This implies (ii) since $\sigma_S = \sigma_{\overline{\mathrm{co}}\, S}$.

[(iii)] The set $S := \bigcap_j S_j$ being nonempty, it has a support function $\sigma_S$. Now, from Corollary 3.1.2,

$$\begin{aligned} s \in S &\iff s \in S_j \text{ for all } j \in J \\ &\iff \langle s, \cdot\rangle \le \sigma_j \text{ for all } j \in J \\ &\iff \langle s, \cdot\rangle \le \inf_{j\in J} \sigma_j \\ &\iff \langle s, \cdot\rangle \le \overline{\mathrm{co}}\big(\inf_{j\in J} \sigma_j\big), \end{aligned}$$

where the last equivalence comes directly from the Definition B.2.5.3 of a closed convex hull. Again Corollary 3.1.2 tells us that the closed sublinear function $\overline{\mathrm{co}}(\inf \sigma_j)$ is just the support function of $S$. □

**Proposition 3.3.6.** *Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be a linear operator, with adjoint $A^*$ (for some scalar product $\langle\cdot,\cdot\rangle$ in $\mathbb{R}^m$). For $S \subset \mathbb{R}^n$ nonempty, we have*

$$\sigma_{\mathrm{cl}\, A(S)}(y) = \sigma_S(A^*y) \qquad \textit{for all } y \in \mathbb{R}^m.$$

*Proof.* Just write the definitions

$$\sigma_{A(S)}(y) = \sup_{s\in S}\langle As, y\rangle = \sup_{s\in S}\langle s, A^*y\rangle$$

and use Proposition 2.2.1 to obtain the result. □

**Proposition 3.3.7.** *Let $A : \mathbb{R}^m \to \mathbb{R}^n$ be a linear operator, with adjoint $A^*$ (for some scalar product $\langle\cdot,\cdot\rangle$ in $\mathbb{R}^m$). Let $\sigma$ be the support function of a nonempty closed convex set $S \subset \mathbb{R}^m$. If $\sigma$ is minorized on the inverse image*

$$A^{-1}(d) = \{p \in \mathbb{R}^m : \ Ap = d\} \tag{3.3.3}$$

*of each $d \in \mathbb{R}^n$, then the support function of the set $(A^{-1})^*(S)$ is the closure of the image-function $A\sigma$.*

*Proof.* Our assumption is tailored to guarantee $A\sigma \in \operatorname{Conv} \mathbb{R}^n$ (Theorem B.2.4.2). The positive homogeneity of $A\sigma$ is clear: for $d \in \mathbb{R}^n$ and $t > 0$,

$$(A\sigma)(td) = \inf_{Ap=td} \sigma(p) = \inf_{A(p/t)=d} t\sigma(p/t) = t \inf_{Aq=d} \sigma(q) = t(A\sigma)(d).$$

Thus, the closed sublinear function $\operatorname{cl}(A\sigma)$ supports some set $S'$; by definition, $s \in S'$ if and only if

$$\langle s, d \rangle \le \inf\{\sigma(p) : \ Ap = d\} \qquad \text{for all } d \in \mathbb{R}^n;$$

but this just means

$$\langle s, Ap \rangle \le \sigma(p) \qquad \text{for all } p \in \mathbb{R}^m,$$

i.e. $A^* s \in S$, because $\langle s, Ap \rangle = \langle A^* s, p \rangle$. □

**Theorem 3.3.4.** *Let $S$ and $S'$ be two nonempty compact convex sets of $\mathbb{R}^n$. Then*

$$\Delta(\sigma_S, \sigma_{S'}) := \max_{\|d\| \le 1} |\sigma_S(d) - \sigma_{S'}(d)| = \Delta_H(S, S'). \tag{3.3.5}$$

*Proof.* As mentioned in §0.5.1, for all $r \ge 0$, the property

$$\max\{d_S(d) : \ d \in S'\} \le r \tag{3.3.6}$$

simply means $S' \subset S + B(0, r)$.

Now, the support function of $B(0, 1)$ is $\|\cdot\|$ — see (2.3.1). Calculus rules on support functions therefore tell us that (3.3.6) is also equivalent to

$$\sigma_{S'}(d) \le \sigma_S(d) + r\|d\| \qquad \text{for all } d \in \mathbb{R}^n,$$

which in turn can be written

$$\max_{\|d\| \le 1} \big[\sigma_{S'}(d) - \sigma_S(d)\big] \le r.$$

In summary, we have proved

$$\max_{d \in S'} d_S(d) = \max_{\|d\| \le 1} \big[\sigma_{S'}(d) - \sigma_S(d)\big]$$

and symmetrically

$$\max_{d \in S} d_{S'}(d) = \max_{\|d\| \le 1} \big[\sigma_S(d) - \sigma_{S'}(d)\big];$$

the result follows. □

**Proposition 3.3.8.** *A convex-compact-valued and locally bounded multifunction $F : \mathbb{R}^n \longrightarrow 2^{\mathbb{R}^n}$ is outer [resp. inner] semi-continuous at $x_0 \in \operatorname{int} \operatorname{dom} F$ if and only if its support function $x \mapsto \sigma_{F(x)}(d)$ is upper [resp. lower] semi-continuous at $x_0$ for all $d$ of norm 1.*

*Proof.* Calculus with support functions tells us that our definition (0.5.2) of outer semi-continuity is equivalent to

$$\forall \varepsilon > 0,\ \exists \delta > 0 :\ y \in B(x_0, \delta) \implies \sigma_{F(y)}(d) \leq \sigma_{F(x_0)}(d) + \varepsilon \|d\| \quad \text{for all } d \in \mathbb{R}^n$$

and division by $\|d\|$ shows that this is exactly upper semi-continuity of the support function for $\|d\| = 1$. Same proof for inner/lower semi-continuity. $\square$

**Corollary 3.3.3.** *Let $(S_k)$ be a sequence of nonempty convex compact sets and $S$ a nonempty convex compact set. When $k \to +\infty$, the following are equivalent*

*(i)* $S_k \to S$ *in the Hausdorff sense, i.e.* $\Delta_H(S_k, S) \to 0$*;*

*(ii)* $\sigma_{S_k} \to \sigma_S$ *pointwise;*

*(iii)* $\sigma_{S_k} \to \sigma_S$ *uniformly on each compact set of* $\mathbb{R}^n$*.*

### 3.3.4 Example: Support functions of closed convex polyhedra

# Chapter 4

# Subdifferentials of Finite Convex Functions

## 4.1 The Subdifferential: Definitions and Interpretations

### 4.1.1 First definition: Directional derivative

**Definition 4.1.1.** (Directional Derivative) The directional derivative of $f$ at $x$ in the direction $d$ is

$$f'(x,d) := \lim_{\{q(t):t\downarrow 0\}} = \inf\{q(t) : t > 0\}. \tag{1.1.2}$$

**Proposition 4.1.1.** *For fixed $x$, the function $f'(x,\cdot)$ is finite sublinear.*

*Proof.* Let $d_1, d_2 \in \mathbb{R}^n$, and positive $\alpha_1, \alpha_2$ with $\alpha_1+\alpha_2 = 1$. From the convexity of $f$:

$$f\big(x + t(\alpha_1 d_1 + \alpha_2 d_2)\big) - f(x) =$$
$$f\big(\alpha_1(x + td_1) + \alpha_2(x + td_2)\big) - \alpha_1 f(x) - \alpha_2 f(x) \leq$$
$$\leq \alpha_1\big[f(x + td_1) - f(x)\big] + \alpha_2\big[f(x + td_2) - f(x)\big].$$

for all $t$. Dividing by $t > 0$ and letting $t \downarrow 0$, we obtain

$$f'(x, \alpha_1 d_1 + \alpha_2 d_2) \leq \alpha_1 f'(x, d_1) + \alpha_2 f'(x, d_2)$$

which establishes the convexity of $f'$ with respect to $d$. Its positive homogeneity is clear: for $\lambda > 0$

$$f'(x, \lambda d) = \lim_{t\downarrow 0} \lambda \frac{f(x + \lambda td) - f(x)}{\lambda t} = \lambda \lim_{\tau\downarrow 0} \frac{f(x + \tau d) - f(x)}{\tau} = \lambda f'(x, d).$$

Finally suppose $\|d\| = 1$. As a finite convex function, $f$ is Lipschitz continuous around $x$ (Theorem B.3.1.2); in particular there exist $\varepsilon > 0$ and $L > 0$ such that

$$|f(x + td) - f(x)| \leq Lt \quad \text{for } 0 \leq t \leq \varepsilon.$$

Hence, $|f'(x, d)| \leq L$ and we conclude with positive homogeneity:

$$|f'(x, d)| \leq L\|d\| \qquad \text{for all } d \in \mathbb{R}^n. \tag{1.1.5}$$

□

**Definition 4.1.2** (Subdifferential I)**.** The subdifferential $\partial f(x)$ of $f$ at $x$ is the nonempty compact convex set of $\mathbb{R}^n$ whose support function is $f'(x, \cdot)$, i.e.

$$\partial f(x) := \{s \in \mathbb{R}^n : \langle s, d\rangle \leq f'(x, d) \text{ for all } d \in \mathbb{R}^n\}. \tag{1.1.6}$$

A vector $s \in \partial f(x)$ is called a *subgradient* of $f$ at $x$.

**Proposition 4.1.2.** *The finite sublinear function $d \mapsto \sigma(d) := f'(x, d)$ satisfies*

$$\sigma'(0, \delta) = f'(x, \delta) \qquad \textit{for all } \delta \in \mathbb{R}^n; \tag{1.1.8}$$

$$\sigma(\delta) = \sigma(0) + \sigma'(0, \delta) = \sigma'(0, \delta) \qquad \textit{for all } \delta \in \mathbb{R}^n; \tag{1.1.9}$$

$$\partial\sigma(0) = \partial f(x). \tag{1.1.10}$$

*Proof.* Because $\sigma$ is positively homogeneous and $\sigma(0) = 0$,

$$\frac{\sigma(t\delta) - \sigma(0)}{t} = \sigma(\delta) = f'(x, \delta) \qquad \text{for all } t > 0.$$

This implies immediately (1.1.8) and (1.1.9). Then (1.1.10) follows from uniqueness of the supported set. □

### 4.1.2 Second definition: Minorization by affine functions

**Definition 4.1.3.** (Subdifferential II) The subdifferential of $f$ at $x$ is the set of vectors $s \in \mathbb{R}^n$ satisfying

$$f(y) \geq f(x) + \langle s, y - x\rangle \qquad \text{for all } y \in \mathbb{R}^n. \tag{1.2.1}$$

**Theorem 4.1.1.** *The definitions 1.1.4 and 1.2.1 are equivalent.*

*Proof.* Let $s$ satisfy (1.1.6), i.e.

$$\langle s, d\rangle \leq f'(x, d) \qquad \text{for all } d \in \mathbb{R}^n. \tag{1.2.2}$$

The second equality in (1.1.2) makes it clear that (1.2.2) is equivalent to

$$\langle s, d\rangle \leq \frac{f(x + td) - f(x)}{t} \qquad \text{for all } d \in \mathbb{R}^n \text{ and } t > 0. \tag{1.2.3}$$

When $d$ describes $\mathbb{R}^n$ and $t$ describes $\mathbb{R}^*_+$, $y := x + td$ describes $\mathbb{R}^n$ and we realize that (1.2.3) is just (1.2.1). □

### 4.1.3 Geometric constructions and interpretations

**Proposition 4.1.3.** *(i) A vector $s \in \mathbb{R}^n$ is a subgradient of $f$ at $x$ if and only if $(s, -1) \in \mathbb{R}^n \times \mathbb{R}$ is normal to* epi $f$ *at $(x, f(x))$. In other words:*

$$N_{\operatorname{epi} f}(x, f(x)) = \{(\lambda s, -\lambda) : s \in \partial f(x),\ \lambda \geq 0\}.$$

*(ii) The tangent cone to the set* epi $f$ *at $(x, f(x))$ is the epigraph of the directional-derivative function $d \mapsto f'(x, d)$:*

$$T_{\operatorname{epi} f}(x, f(x)) = \{(d, r) : r \geq f'(x, d)\}.$$

*(i).* Apply Definition A.5.2.3 to see that $(s, -1) \in N_{\operatorname{epi} f}(x, f(x))$ means

$$\langle s, y - x\rangle + (-1)[r - f(x)] \leq 0 \qquad \text{for all } y \in \mathbb{R}^n \text{ and } r \geq f(y)$$

and the equivalence with (1.2.1) is clear. The formula follows since the set of normals forms a cone containing the origin.

[(ii)] The tangent cone to epi $f$ is the polar of the above normal cone, i.e. the set of $(d, r) \in \mathbb{R}^n \times \mathbb{R}$ such that

$$\langle \lambda s, d\rangle + (-\lambda)r \leq 0 \qquad \text{for all } s \in \partial f(x) \text{ and } \lambda \geq 0.$$

Barring the trivial case $\lambda = 0$, we divide by $\lambda > 0$ to obtain

$$r \geq \max\{\langle s, d\rangle : s \in \partial f(x)\} = f'(x, d).$$

□

**Lemma 4.1.1.** *For the convex function $f : \mathbb{R}^n \to \mathbb{R}$ and the sublevel-set (1.3.1), we have*

$$T_{S_{f(x)}}(x) \subset \{d :\ f'(x, d) \leq 0\}. \tag{1.3.2}$$

*Proof.* Take arbitrary $y \in S_{f(x)},\ t > 0$, and set $d := t(y - x)$. Then, using the second equality in (1.1.2),

$$0 \geq t[f(y) - f(x)] = \frac{f(x + d/t) - f(x)}{1/t} \geq f'(x, d).$$

So we have proved

$$\mathbb{R}^+[S_{f(x)} - x] \subset \{d :\ f'(x, d) \leq 0\} \tag{1.3.3}$$

(note: the case $d = 0$ is covered since $0 \in S_{f(x)} - x$).

Because $f'(\cdot, \cdot)$ is a closed function, the righthand set in (1.3.3) is closed. Knowing that $T_{S_{f(x)}}(x)$ is the closure of the lefthand side in (1.3.3) (Proposition A.5.2.1), we deduce the result by taking the closure of both sides in (1.3.3). □

**Proposition 4.1.4.** *Let $g : \mathbb{R}^n \to \mathbb{R}$ be convex and suppose that $g(x_0) < 0$ for some $x_0 \in \mathbb{R}^n$. Then*

$$\mathrm{cl}\{z : g(z) < 0\} = \{z : g(z) \leq 0\}, \tag{1.3.4}$$

$$\{z : g(z) < 0\} = \mathrm{int}\{z : g(z) \leq 0\}. \tag{1.3.5}$$

*It follows*

$$\mathrm{bd}\{z : g(z) \leq 0\} = \{z : g(z) = 0\}. \tag{1.3.6}$$

*Proof.* Because $g$ is (lower semi-) continuous, the inclusion "$\subset$" automatically holds in (1.3.4). Conversely, let $\bar{z}$ be arbitrary with $g(\bar{z}) \leq 0$ and, for $k > 0$, set

$$z_k := \tfrac{1}{k}x_0 + (1 - \tfrac{1}{k})\bar{z}.$$

By convexity of $g$, $g(z_k) < 0$, so (1.3.4) is established by letting $k \to +\infty$.

Now, take the interior of both sides in (1.3.4). The "int cl" on the left is actually an "int" (Proposition A.2.1.8), and this "int"-operation is useless because $g$ is (upper semi-) continuous: (1.3.5) is established. □

**Theorem 4.1.2.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and suppose $0 \notin \partial f(x)$. Then, $S_f(x)$ being the sublevel-set (1.3.1),*

$$T_{S_f(x)}(x) = \{d \in \mathbb{R}^n : \; f'(x,d) \leq 0\} \tag{1.3.7}$$

$$\mathrm{int}\left[T_{S_f(x)}(x)\right] = \{d \in \mathbb{R}^n : \; f'(x,d) < 0\} \neq \varnothing. \tag{1.3.8}$$

*Proof.* From the very definition (1.1.6), our assumption means that $f'(x,d) < 0$ for some $d$, and (1.1.2) then implies that $f(x+td) < f(x)$ for $t > 0$ small enough: our $d$ is of the form $(x + td - x)/t$ with $x + td \in S_f(x)$ and we have proved

$$\{d : \; f'(x,d) < 0\} \subset \mathbb{R}^+[S_f(x) - x] \subset T_{S_f(x)}(x). \tag{1.3.9}$$

Now, we can apply (1.3.4) with $g = f'(x, \cdot)$:

$$\mathrm{cl}\{d : \; f'(x,d) < 0\} = \{d : \; f'(x,d) \leq 0\},$$

so (1.3.7) is proved by closing the sets in (1.3.9) and using (1.3.2). Finally, take the interior of both sides in (1.3.7) and apply (1.3.5) with $g = f'(x, \cdot)$ to prove (1.3.8). □

**Theorem 4.1.3.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and suppose $0 \notin \partial f(x)$. Then a direction $d$ is normal to $Sf(x)$ at $x$ if and only if there is some $t \geq 0$ and some $s \in \partial f(x)$ such that $d = ts$:*

$$N_{Sf(x)}(x) = \mathbb{R}^+ \partial f(x).$$

*Proof.* Write (1.3.7) as

$$T_{Sf(x)}(x) = \{d \in \mathbb{R}^n : \langle s, d\rangle \leq 0 \text{ for all } s \in \partial f(x)\} = \{d \in \mathbb{R}^n : \langle \lambda s, d\rangle \leq 0 \text{ for all } \lambda \geq 0 \text{ and } s \in \partial f(x)\} = [\mathbb{R}^+ \partial$$

The result follows by taking the polar cone of both sides, and observing that the assumption implies closedness of $\mathbb{R}^+\partial f(x)$ (Proposition A.1.4.7):

$$N_{Sf(x)}(x) = \mathrm{cl}[\mathbb{R}^+\partial f(x)] = \mathbb{R}^+\partial f(x).$$

□

## 4.2 Local Properties of the Subdifferential

### 4.2.1 First-order developments

**Lemma 4.2.1.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and $x \in \mathbb{R}^n$. For any $\varepsilon > 0$, there exists $\delta > 0$ such that $\|h\| \leq \delta$ implies*

$$\big|f(x+h) - f(x) - f'(x,h)\big| \leq \varepsilon\|h\|. \tag{2.1.1}$$

*Proof.* Suppose for contradiction that there is $\varepsilon > 0$ and a sequence $(h_k)$ with $\|h_k\| =: t_k \leq 1/k$ such that

$$|f(x+h_k) - f(x) - f'(x,h_k)| > \varepsilon t_k \qquad \text{for } k = 1, 2, \ldots$$

Extracting a subsequence if necessary, assume that $h_k/t_k \to d$ for some $d$ of norm 1. Then take a local Lipschitz constant $L$ of $f$ (see Remark 1.1.3) and expand:

$$\begin{aligned}\varepsilon t_k < |f(x+h_k) - f(x) - f'(x,h_k)| &\leq |f(x+h_k) - f(x+t_k d)| + \\ &+ |f(x+t_k d) - f(x) - f'(x,t_k d)| + |f'(x,t_k d) - f'(x,h_k)| \\ &\leq 2L\|h_k - t_k d\| + |f(x+t_k d) - f(x) - t_k f'(x,d)|.\end{aligned}$$

Divide by $t_k > 0$ and pass to the limit to obtain the contradiction $\varepsilon \leq 0$. $\square$

**Corollary 4.2.1.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. At any $x$,*

$$f(x+h) = f(x) + \langle s, h\rangle + o(\|h\|)$$

*whenever one of the following equivalent properties holds:*

$$s \in \partial f(x)(h) \iff h \in N_{\partial f(x)}(s) \iff s = p_{\partial f(x)}(s+h).$$

**Corollary 4.2.2.** *If the convex $f$ is (Gâteaux) differentiable at $x$, its only subgradient at $x$ is its gradient $\nabla f(x)$. Conversely, if $\partial f(x)$ contains only one element $s$, then $f$ is (Fréchet) differentiable at $x$, with $\nabla f(x) = s$.*

**Proposition 4.2.1.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. For all $x$ and $d$ in $\mathbb{R}^n$, we have*

$$F_{\partial f(x)}(d) = \partial[f'(x,\cdot)](d).$$

*Proof.* If $s \in \partial f(x)$ then $f'(x,d') \geq \langle s, d'\rangle$ for all $d' \in \mathbb{R}^n$, simply because $f'(x,\cdot)$ is the support function of $\partial f(x)$. If, in addition, $\langle s, d\rangle = f'(x,d)$, we get

$$f'(x,d') \geq f'(x,d) + \langle s, d'-d\rangle \qquad \text{for all } d' \in \mathbb{R}^n \tag{2.1.4}$$

which proves the inclusion $F_{\partial f(x)}(d) \subset \partial[f'(x,\cdot)](d)$.

Conversely, let $s$ satisfy (2.1.4). Set $d'' := d' - d$ and deduce from subadditivity

$$f'(x,d) + f'(x,d'') \geq f'(x,d') \geq f'(x,d) + \langle s, d''\rangle \qquad \text{for all } d'' \in \mathbb{R}^n$$

which implies $f'(x,\cdot) \geq \langle s,\cdot\rangle$, hence $s \in \partial f(x)$. Also, putting $d' = 0$ in (2.1.4) shows that $\langle s, d\rangle \geq f'(x,d)$. Altogether, we have $s \in F_{\partial f(x)}(d)$. $\square$

**Definition 4.2.1.** A point $x$ at which $\partial f(x)$ has more than one element — i.e. at which $f$ is not differentiable — is called a *kink* (or corner-point) of $f$.

### 4.2.2 Minimality conditions

**Theorem 4.2.1.** *For $f : \mathbb{R}^n \to \mathbb{R}$ convex, the following three properties are equivalent:*

*[(i)]*

1. *$f$ is minimized at $x$ over $\mathbb{R}^n$, i.e., $f(y) \geq f(x)$ for all $y \in \mathbb{R}^n$;*
2. *$0 \in \partial f(x)$;*
3. *$f'(x, d) \geq 0$ for all $d \in \mathbb{R}^n$.*

*Proof.* The equivalence (i) $\Leftrightarrow$ (ii) [resp. (ii) $\Leftrightarrow$ (iii)] is obvious from (1.2.1) [resp. (1.1.6)]. $\square$

### 4.2.3 Mean-value theorems

**Lemma 4.2.2.** *The subdifferential of $\varphi$ defined by (2.3.1) is*

$$\partial\varphi(t) = \{\ s, y - x\ :\ s \in \partial f(x_t)\ \}$$

*or, more symbolically:*

$$\partial\varphi(t) = \langle \partial f(x_t),\, y - x\rangle.$$

*Proof.* In terms of right- and left-derivatives (see Theorem 0.6.3), we have

$$D_+\varphi(t) = \lim_{\tau \downarrow 0} \frac{f(x_t + \tau(y - x)) - f(x_t)}{\tau} = f'(x_t, y - x),$$

$$D_-\varphi(t) = \lim_{\tau \uparrow 0} \frac{f(x_t + \tau(y - x)) - f(x_t)}{\tau} = -f'(x_t, -(y - x));$$

so, knowing that

$$f'(x_t, y - x) = \max_{s \in \partial f(x_t)} \langle s, y - x\rangle,$$

$$-f'(x_t, -(y - x)) = \min_{s \in \partial f(x_t)} \langle s, y - x\rangle,$$

we obtain $\partial\varphi(t) := [D_-\varphi(t), D_+\varphi(t)] = \{\langle s, y - x\rangle : s \in \partial f(x)\}$. $\square$

**Theorem 4.2.2.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Given two points $x \neq y$ in $\mathbb{R}^n$, there exist $t \in ]0, 1[$ and $s \in \partial f(x_t)$ such that*

$$f(y) - f(x) = \langle s,\, y - x\rangle. \tag{2.3.2}$$

*In other words,*

$$f(y) - f(x) \in \bigcup_{t \in ]0,1[} \{\langle \partial f(x_t),\, y - x\rangle\}.$$

*Proof.* Start from the function $\varphi$ of (2.3.1) and, as usual in this context, consider the auxiliary function

$$\psi(t) := \varphi(t) - \varphi(0) - t[\varphi(1) - \varphi(0)],$$

which is clearly convex. Computing directional derivatives gives easily $\partial\psi(t) = \partial\varphi(t) - [\varphi(1) - \varphi(0)]$. Now $\psi$ is continuous on $[0, 1]$, it has been constructed so that $\psi(0) = \psi(1) = 0$, so it is minimal at some $t \in ]0, 1[$: at this $t$, $0 \in \partial\psi(t)$ (Theorem 2.2.1). In view of Lemma 2.3.1, this means that there is $s \in \partial f(x_t)$ such that

$$\langle s,\, y - x\rangle = \varphi(1) - \varphi(0) = f(y) - f(x).$$

$\square$

**Theorem 4.2.3.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. For $x, y \in \mathbb{R}^n$,*

$$f(y) - f(x) = \int_0^1 \langle \partial f(xt),\, y - x\rangle \, dt.$$

## 4.3 First Examples

## 4.4 Calculus Rules with Subdifferentials

### 4.4.1 Positive combinations of functions

**Theorem 4.4.1.** *Let $f_1, f_2$ be two convex functions from $\mathbb{R}^n$ to $\mathbb{R}$ and $t_1, t_2$ be positive. Then*

$$\partial(t_1 f_1 + t_2 f_2)(x) = t_1\partial f_1(x) + t_2\partial f_2(x) \qquad \textit{for all } x \in \mathbb{R}^n. \tag{4.1.1}$$

*Proof.* Apply Theorem C.3.3.2(i): $t_1\partial f_1(x) + t_2\partial f_2(x)$ is a compact convex set whose support function is

$$t_1 f_1'(x, \cdot) + t_2 f_2'(x, \cdot). \tag{4.1.2}$$

On the other hand, the support function of $\partial(t_1 f_1 + t_2 f_2)(x)$ is by definition the directional derivative $(t_1 f_1 + t_2 f_2)'(x, \cdot)$ which, from elementary calculus, is just (4.1.2). Therefore the two (compact convex) sets in (4.1.1) coincide, since they have the same support function. $\square$

### 4.4.2 Pre-composition with an affine mapping

**Theorem 4.4.2.** *Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be an affine mapping ( $Ax = A_0x + b$, with $A_0$ linear and $b \in \mathbb{R}^m$) and let $g$ be a finite convex function on $\mathbb{R}^m$. Then*

$$\partial(g \circ A)(x) = A_0^*\partial g(Ax) \qquad \textit{for all } x \in \mathbb{R}^n. \tag{4.2.1}$$

*Proof.* Form the difference quotient giving rise to $(g \circ A)'(x, d)$ and use the relation $A(x + td) = Ax + tA_0 d$ to obtain

$$(g \circ A)'(x, d) = g'(Ax, A_0 d) \qquad \text{for all } d \in \mathbb{R}^n.$$

From Proposition C.3.3.3, the righthand side in the above equality is the support function of the convex compact set $A_0^* \partial g(Ax)$. □

### 4.4.3 Post-composition with increasing convex function of several variables

**Theorem 4.4.3.** *Let $f$, $F$ and $g$ be defined as above. For all $x \in \mathbb{R}^n$,*

$$\partial(g \circ F)(x) = \Big\{ \sum_{i=1}^{m} \rho^i s_i : \ (\rho^1, \dots, \rho^m) \in \partial g(F(x)), \ s_i \in \partial f_i(x) \ \textit{ for } i = 1, \dots, m \Big\}. \tag{4.3.1}$$

*Preamble.* Our aim is to show the formula via support functions, hence we need to establish the convexity and compactness of the righthand side in (4.3.1) – call it $S$. Boundedness and closedness are easy, coming from the fact that a subdifferential (be it $\partial g$ or $\partial f_i$) is bounded and closed. As for convexity, pick two points in $S$ and form their convex combination

$$s = \alpha \sum_{i=1}^{m} \rho^i s_i + (1 - \alpha) \sum_{i=1}^{m} \rho'^i s'_i = \sum_{i=1}^{m} \big[ \alpha \rho^i s_i + (1 - \alpha) \rho'^i s'_i \big],$$

where $\alpha \in [0, 1]$. Remember that each $\rho^i$ and $\rho'^i$ is nonnegative and the above sum can be restricted to those terms such that $\rho''^i := \alpha \rho^i + (1 - \alpha) \rho'^i > 0$. Then we write each such term as

$$\rho''^i \big[ \frac{\alpha \rho^i}{\rho''^i} s_i + \frac{(1 - \alpha) \rho'^i}{\rho''^i} s'_i \big].$$

It suffices to observe that $\rho''^i \in \partial g(F(x))$, so the bracketed expression is in $\partial f_i(x)$; thus $s \in S$.

[Step 1] Now let us compute the support function $\sigma_S$ of $S$. For $d \in \mathbb{R}^n$, we denote by $F'(x, d) \in \mathbb{R}^m$ the vector whose components are $f'_i(x, d)$ and we proceed to prove

$$\sigma_S(d) = g'(F(x), F'(x, d)). \tag{4.3.2}$$

For any $s = \sum_{i=1}^{m} \rho^i s_i \in S$, we write $\langle s, d \rangle$ as

$$\sum_{i=1}^{m} \rho^i \langle s_i, d \rangle \le \sum_{i=1}^{m} \rho^i f'_i(x, d) \le g'(F(x), F'(x, d)); \tag{4.3.3}$$

the first inequality uses $\rho^i \ge 0$ and the definition of $f'_i(x, \cdot) = \sigma_{\partial f_i}(x)$; the second uses the definition $g'(F(x), \cdot) = \sigma_{\partial g(F(x))}$.

On the other hand, the compactness of $\partial g(F(x))$ implies the existence of an $m$-tuple $(\bar{\rho}^i) \in \partial g(F(x))$ such that

$$g'(F(x), F'(x,d)) = \sum_{i=1}^{m} \bar{\rho}^i f_i'(x,d),$$

and the compactness of each $\partial f_i(x)$ yields likewise an $\bar{s}_i \in \partial f_i(x)$ such that

$$f_i'(x,d) = \langle \bar{s}_i, d \rangle \quad \text{for } i = 1, \ldots, m.$$

Altogether, we have exhibited an $\bar{s} = \sum_{i=1}^{m} \bar{\rho}^i \bar{s}_i \in S$ such that equality holds in (4.3.3), so (4.3.2) is established.

[Step 2] It remains to prove that the support function (4.3.2) is really the directional derivative $(g \circ F)'(x,d)$. For $t > 0$, expand $F(x+td)$, use the fact that $g$ is locally Lipschitzian, and then expand $g(F(x+td))$:

$$g(F(x+td)) = g(F(x)+tF'(x,d)+o(t)) = g(F(x)+tF'(x,d))+o(t) = g(F(x))+tg'(F(x),F'(x,d))+o(t).$$

From there, it follows

$$(g \circ F)'(x,d) := \lim_{t \downarrow 0} \frac{g(F(x+td)) - g(F(x))}{t} = g'(F(x), F'(x,d)). \quad \square$$

$\square$

**Corollary 4.4.1.** *Let $f_1, \ldots, f_m$ be $m$ convex functions from $\mathbb{R}^n$ to $\mathbb{R}$ and define*

$$f := \max\{f_1, \ldots, f_m\}.$$

*Denoting by $I(x) := \{i : f_i(x) = f(x)\}$ the active index-set, we have*

$$\partial f(x) = \mathrm{co}\{\bigcup_{i \in I(x)} \partial f_i(x)\}. \tag{4.3.4}$$

*Proof.* Take $g(y) = \max\{y^1, \ldots, y^m\}$, whose subdifferential was computed in (3.7): $\{e_i\}$ denoting the canonical basis of $\mathbb{R}^m$,

$$\partial g(y) = \mathrm{co}\{e_i : i \text{ such that } y^i = g(y)\}.$$

Then, using the notation of Theorem 4.3.1, we write $\partial g(F(x))$ as

$$\left\{(\rho^1, \ldots, \rho^m) : \; \rho^i = 0 \text{ for } i \notin I(x), \; \rho^i \geq 0 \text{ for } i \in I(x), \; \sum_{i=1}^{m} \rho^i = 1\right\},$$

and (4.3.1) gives

$$\partial f(x) = \left\{\sum_{i \in I(x)} \rho^i \partial f_i(x) : \; \rho^i \geq 0 \text{ for } i \in I(x), \; \sum_{i \in I(x)} \rho^i = 1\right\}.$$

Remembering Example A.1.3.5, it suffices to recognize in the above expression the convex hull announced in (4.3.4). $\square$

### 4.4.4 Supremum of convex functions

**Lemma 4.4.1** (Lemma 4.4.1). *With the notation (4.4.1), (4.4.2),*

$$\partial f(x) \supset \operatorname{co}\{\partial f_j(x) : \ j \in J(x)\}. \tag{4.4.3}$$

*Proof.* Take $j \in J(x)$ and $s \in \partial f_j(x)$; from the definition (1.2.1) of the subdifferential,

$$f(y) \geq f_j(y) \geq f_j(x) + \langle s, y - x\rangle \qquad \text{for all } y \in \mathbb{R}^n,$$

so $\partial f(x)$ contains $\partial f_j(x)$. Being closed and convex, it also contains the closed convex hull appearing in (4.4.3). □

**Theorem 4.4.4.** *With the notation (4.4.1), (4.4.2), assume that $J$ is a compact set (in some metric space), on which the functions $j \mapsto f_j(x)$ are upper semi-continuous for each $x \in \mathbb{R}^n$. Then*

$$\partial f(x) = \operatorname{co}\{\cup \partial f_j(x) : \ j \in J(x)\}. \tag{4.4.4}$$

*Step 0.* Our assumptions make $J(x)$ nonempty and compact. Denote by $S$ the curly bracketed set in (4.4.4); because of (4.4.3), $S$ is bounded, let us check that it is closed. Take a sequence $(s_k) \subset S$ converging to $s$; to each $s_k$, we associate some $j_k \in J(x)$ such that $s_k \in \partial f_{j_k}(x)$, i.e.

$$f_{j_k}(y) \geq f_{j_k}(x) + \langle s_k, y - x\rangle \qquad \text{for all } y \in \mathbb{R}^n.$$

Let $k \to \infty$; extract a subsequence so that $j_k \to j \in J(x)$; we have $f_{j_k}(x) \equiv f(x) =: f_j(x)$; and by upper semi-continuity of the function $f_{(\cdot)}(y)$, we obtain

$$f_j(y) \geq \limsup f_{j_k}(y) \geq f_j(x) + \langle s, y - x\rangle \qquad \text{for all } y \in \mathbb{R}^n,$$

which shows $s \in \partial f_j(x) \subset S$. Altogether, $S$ is compact and its convex hull is also compact (Theorem A.1.4.3).

In view of Lemma 4.4.1, it suffices to prove the "$\subset$"-inclusion in (4.4.4); for this, we will establish the corresponding inequality between support functions which, in view of the calculus rule C.3.3.2(ii), is: for all $d \in \mathbb{R}^n$,

$$f'(x, d) \leq \sigma_S(d) = \sup\{f_j'(x, d) : \ j \in J(x)\}. \tag{4.4.5}$$

[Step 1] Let $\varepsilon > 0$; from the definition (1.1.2) of $f'(x, d)$,

$$\frac{f(x + td) - f(x)}{t} > f'(x, d) - \varepsilon \qquad \text{for all } t > 0. \tag{4.4.6}$$

For $t > 0$, set

$$J_t := \Big\{ j \in J : \ \frac{f_j(x + td) - f_j(x)}{t} \geq f'(x, d) - \varepsilon \Big\}.$$

The definition of $f(x+td)$ shows with (4.4.6) that $J_t$ is nonempty. Because $J$ is compact and $f_{(\cdot)}(x+td)$ is upper semi-continuous, $J_t$ is visibly compact. Observe that $J_t$ is a superlevel-set of the function

$$0 < t \mapsto \frac{f_j(x+td) - f_j(x)}{t} + \frac{f_j(x) - f(x)}{t},$$

which is nondecreasing: the first fraction is the slope of a convex function, and the second fraction has a nonpositive numerator. Thus, $J_{t_1} \subset J_{t_2}$ for $0 < t_1 \leq t_2$.

[Step 2] By compactness, we deduce the existence of some $j^* \in \bigcap_{t>0} J_t$ (for each $\tau \in ]0,t]$, pick some $j_\tau \in J_\tau \subset J_t$; take a cluster point for $\tau \downarrow 0$: it is in $J_t$). We therefore have

$$f_{j^*}(x+td) - f(x) \geq t\big[f'(x,d) - \varepsilon\big] \qquad \text{for all } t > 0,$$

hence $j^* \in J(x)$ (continuity of the convex function $f_{j^*}$ for $t \downarrow 0$). In this inequality, we can replace $f(x)$ by $f_{j^*}(x)$, divide by $t$ and let $t \downarrow 0$ to obtain

$$\sigma_S(d) \geq f'_{j^*}(x,d) \geq f'(x,d) - \varepsilon.$$

Since $d \in \mathbb{R}^n$ and $\varepsilon > 0$ were arbitrary, (4.4.5) is established. $\square$

**Corollary 4.4.2.** *The notation and assumptions are those of Theorem 4.4.2. Assume also that each $f_j$ is differentiable; then*

$$\partial f(x) = \operatorname{co}\{\nabla f_j(x) : \ j \in J(x)\}.$$

**Corollary 4.4.3.** *For some compact set $Y \subset \mathbb{R}^p$, let $g : \mathbb{R}^n \times Y \to \mathbb{R}$ be a function satisfying the following properties:*

- *for each $x \in \mathbb{R}^n$, $g(x,\cdot)$ is upper semi-continuous;*
- *for each $y \in Y$, $g(\cdot,y)$ is convex and differentiable;*
- *the function $f := \sup_{y\in Y} g(\cdot,y)$ is finite-valued on $\mathbb{R}^n$;*
- *at some $x \in \mathbb{R}^n$, $g(x,\cdot)$ is maximized at a unique $y(x) \in Y$.*

*Then $f$ is differentiable at this $x$, and its gradient is*

$$\nabla f(x) = \nabla_x g\big(x, y(x)\big) \tag{4.4.8}$$

*(where $\nabla_x g(x,y)$ denotes the gradient of the function $g(\cdot,y)$ at $x$).*

### 4.4.5 Image of a function under a linear mapping

**Theorem 4.4.5.** *With the notation (4.5.1), (4.5.2), assume $A$ is surjective. Let $x$ be such that $Y(x)$ is nonempty. Then, for arbitrary $y \in Y(x)$,*

$$\partial(Ag)(x) = \{\, s \in \mathbb{R}^n : \ A^* s \in \partial g(y)\,\} = \big(A^*\big)^{-1}[\partial g(y)] \tag{4.5.3}$$

*(and this set is thus independent of the particular optimal $y$).*

*Proof.* By definition, $s \in \partial(Ag)(x)$ if and only if $(Ag)(x') \geq (Ag)(x) + \langle s, x' - x \rangle$ for all $x' \in \mathbb{R}^n$, which can be rewritten

$$(Ag)(x') \geq g(y) + \langle s, x' - Ay \rangle \qquad \text{for all } x' \in \mathbb{R}^n$$

where $y$ is arbitrary in $Y(x)$. Furthermore, because $A$ is surjective and by definition of $Ag$, this last relation is equivalent to

$$g(y') \geq g(y) + \langle s, Ay' - Ay \rangle = g(y) + \langle A^* s, y' - y \rangle \qquad \text{for all } y' \in \mathbb{R}^m$$

which means that $A^* s \in \partial g(y)$. □

**Corollary 4.4.4.** *Make the assumptions of Theorem 4.5.1. If $g$ is differentiable at some $y \in Y(x)$, then $Ag$ is differentiable at $x$.*

*Proof.* Surjectivity of $A$ is equivalent to injectivity of $A^*$: in (4.5.3), we have an equation in $s$: $A^* s = \nabla g(y)$, whose solution is unique, and is therefore $\nabla(Ag)(x)$. □

**Corollary 4.4.5.** *Suppose that the subdifferential of $g$ in (4.5.4) is associated with a scalar product $\langle\langle \cdot, \cdot \rangle\rangle$ preserving the structure of a product space:*

$$\langle\langle (x, y), (x', y') \rangle\rangle = \langle x, x' \rangle_n + \langle y, y' \rangle_m \qquad \textit{for } x, x' \in \mathbb{R}^n \textit{ and } y, y' \in \mathbb{R}^m.$$

*At a given $x \in \mathbb{R}^n$, take an arbitrary $y$ solving (4.5.4). Then*

$$\partial f(x) = \{ s \in \mathbb{R}^n : (s, 0) \in \partial_{(x,y)} g(x, y) \}.$$

*Proof.* With our notation, $A^* s = (s, 0)$ for all $s \in \mathbb{R}^n$. It suffices to apply Theorem 4.5.1 (the symbol $\partial_{(x,y)} g$ is used as a reminder that we are dealing with the subdifferential of $g$ with respect to the variable $(\cdot, \cdot) \in \mathbb{R}^n \times \mathbb{R}^m$). □

**Corollary 4.4.6.** *4.5.5 Let $f_1$ and $f_2 : \mathbb{R}^n \to \mathbb{R}$ be two convex functions minorized by a common affine function. For given $x$, let $(y_1, y_2)$ be such that the inf-convolution is exact at $x = y_1 + y_2$, i.e.: $(f_1 \overset{+}{\vee} f_2)(x) = f_1(y_1) + f_2(y_2)$. Then*

$$\partial(f_1 \overset{+}{\vee} f_2)(x) = \partial f_1(y_1) \cap \partial f_2(y_2). \tag{4.5.6}$$

*Proof.* First observe that $A^* s = (s, s)$. Also, apply Definition 1.2.1 to see that $(s_1, s_2) \in \partial g(y_1, y_2)$ if and only if $s_1 \in \partial f_1(y_1)$ and $s_2 \in \partial f_2(y_2)$. Then (4.5.6) is just the copy of (4.5.3) in the present context. □

## 4.5 Further Examples

### 4.5.1 Largest eigenvalue of a symmetric matrix

### 4.5.2 Nested optimization

### 4.5.3 Best approximation of a convex function on a compact interval

**Theorem 4.5.1.** *With the notations (5.3.1), (5.3.2), suppose $\varphi_0 \notin H$. A necessary and sufficient condition for $\bar{x} = (\bar{\xi}^1, \ldots, \bar{\xi}^n) \in \mathbb{R}^n$ to minimize $f$ of*

*(5.3.1) is that, for some positive integer $p \leq n+1$, there exist $p$ points $t_1, \dots, t_p$ in $T$, $p$ integers $\varepsilon_1, \dots, \varepsilon_p$ in $\{-1, +1\}$ and $p$ positive numbers $\alpha_1, \dots, \alpha_p$ such that*

$$\sum_{i=1}^{n} \xi_i^r \varphi_i(t_k) - \varphi_0(t_k) = \varepsilon_k f(\bar{x}) \qquad \text{for } k = 1, \dots, p,$$

$$\sum_{k=1}^{p} \alpha_k \varepsilon_k \varphi_i(t_k) = 0 \qquad \text{for } i = 1, \dots, n$$

*(or equivalently:* $\sum_{k=1}^{p} \alpha_k \varepsilon_k \psi(t_k) = 0 \quad \text{for all } \psi \in H$*).* □

## 4.6 Subdifferential as a multifunction

### 4.6.1 Monotonicity property of the subdifferential

**Proposition 4.6.1.** *The subdifferential mapping is monotone in the sense that, for all $x_1$ and $x_2$ in $\mathbb{R}^n$,*

$$\langle s_2 - s_1, x_2 - x_1 \rangle \geq 0 \quad \text{for all } s_i \in \partial f(x_i),\ i = 1, 2. \tag{6.1.1}$$

*Proof.* The subgradient inequalities

$$f(x_2) \geq f(x_1) + \langle s_1, x_2 - x_1 \rangle \quad \text{for all } s_1 \in \partial f(x_1)$$

$$f(x_1) \geq f(x_2) + \langle s_2, x_1 - x_2 \rangle \quad \text{for all } s_2 \in \partial f(x_2)$$

give the result simply by addition. □

**Theorem 4.6.1.** *A necessary and sufficient for a convex function $f : \mathbb{R}^n \to \mathbb{R}$ to be strongly convex with modulus $c > 0$ on a convex set $C$ is: for all $x_1, x_2$ in $C$,*

$$f(x_2) \geq f(x_1) + \langle s, x_2 - x_1 \rangle + \frac{c}{2}\|x_2 - x_1\|^2 \quad \text{for all } s \in \partial f(x_1), \tag{6.1.3}$$

*or equivalently*

$$\langle s_2 - s_1, x_2 - x_1 \rangle \geq c\|x_2 - x_1\|^2 \quad \text{for all } s_i \in \partial f(x_i),\ i = 1, 2. \tag{6.1.4}$$

*Proof.* For $x_1, x_2$ given in $C$ and $\alpha \in ]0, 1[$, we will use the notation

$$x^\alpha := \alpha x_2 + (1 - \alpha)x_1 = x_1 + \alpha(x_2 - x_1)$$

and we will prove (6.1.3) ⇒ (6.1.2) ⇒ (6.1.4) ⇒ (6.1.3).

[(6.1.3) ⇒ (6.1.2)] Write (6.1.3) with $x_1$ replaced by $x^\alpha \in C$: for $s \in \partial f(x^\alpha)$,

$$f(x_2) \geq f(x^\alpha) + \langle s, x_2 - x^\alpha \rangle + \frac{c}{2}\|x_2 - x^\alpha\|^2,$$

or equivalently

$$f(x_2) \geq f(x^\alpha) + (1-\alpha)\langle s, x_2 - x_1 \rangle + \frac{c}{2}(1-\alpha)^2 \|x_2 - x_1\|^2.$$

Likewise,

$$f(x_1) \geq f(x^\alpha) + \alpha \langle s, x_1 - x_2 \rangle + \frac{c}{2}\alpha^2 \|x_1 - x_2\|^2.$$

Multiply these last two inequalities by $\alpha$ and $(1-\alpha)$ respectively, and add to obtain

$$\alpha f(x_2) + (1-\alpha) f(x_1) \geq f(x^\alpha) + \frac{c}{2}\|x_2 - x_1\|^2 \left[\alpha(1-\alpha)^2 + (1-\alpha)\alpha^2\right].$$

Then realize after simplification that this is just (6.1.2).

[(6.1.2) $\Rightarrow$ (6.1.4)] Write (6.1.2) as

$$\frac{f(x^\alpha) - f(x_1)}{\alpha} + \frac{c}{2}(1-\alpha)\|x_2 - x_1\|^2 \leq f(x_2) - f(x_1)$$

and let $\alpha \downarrow 0$ to obtain $f'(x_1, x_2 - x_1) + \frac{c}{2}\|x_2 - x_1\|^2 \leq f(x_2) - f(x_1)$, which implies (6.1.3). Then, copying (6.1.3) with $x_1$ and $x_2$ interchanged and adding yields (6.1.4) directly.

[(6.1.4) $\Rightarrow$ (6.1.3)] Apply Theorem 2.3.4 to the one-dimensional convex function $\mathbb{R} \ni \alpha \mapsto \varphi(\alpha) := f(x^\alpha)$:

$$f(x_2) - f(x_1) = \varphi(1) - \varphi(0) = \int_0^1 \langle s^\alpha, x_2 - x_1 \rangle \, d\alpha \tag{6.1.5}$$

where $s^\alpha \in \partial f(x^\alpha)$ for $\alpha \in [0, 1]$. Then take $s_1$ arbitrary in $\partial f(x_1)$ and apply (6.1.4): $\langle s^\alpha - s_1, x^\alpha - x_1 \rangle \geq c\|x^\alpha - x_1\|^2$ i.e., using the value of $x^\alpha$,

$$\alpha \langle s^\alpha, x_2 - x_1 \rangle \geq \alpha \langle s_1, x_2 - x_1 \rangle + c\alpha^2 \|x_2 - x_1\|^2.$$

The result follows by using this inequality to minorize the integral in (6.1.5). □

**Proposition 4.6.2.** *A necessary and sufficient condition for a convex function $f : \mathbb{R}^n \to \mathbb{R}$ to be strictly convex on a convex set $C$ is: for all $x_1, x_2 \in C$ with $x_2 \neq x_1$,*

$$f(x_2) > f(x_1) + \langle s, x_2 - x_1 \rangle \quad \textit{for all } s \in \partial f(x_1)$$

*or equivalently*

$$\langle s_2 - s_1, x_2 - x_1 \rangle > 0 \quad \textit{for all } s_i \in \partial f(x_i),\ i = 1, 2.$$

*Proof.* Copy the proof of Theorem 6.1.2 with $c = 0$ and the relevant "$\geq$"-signs replaced by strict inequalities. The only delicate point is in the [(6.1.2) $\Rightarrow$ (6.1.4)]-stage: use monotonicity of the difference quotient. □

### 4.6.2 Continuity properties of the subdifferential

**Proposition 4.6.3.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. The graph of its subdifferential mapping is closed in $\mathbb{R}^n \times \mathbb{R}^n$.*

*Proof.* Let $(x_k, s_k)$ be a sequence in $\operatorname{gr} \partial f$ converging to $(x, s) \in \mathbb{R}^n \times \mathbb{R}^n$. We must prove that $(x, s) \in \operatorname{gr} \partial f$, which is easy. We have for all $k$

$$f(y) \geq f(x_k) + \langle s_k, y - x_k \rangle \quad \text{for all } y \in \mathbb{R}^n;$$

pass to the limit on $k$, using continuity of $f$ and of the scalar product. □

**Proposition 4.6.4.** *The mapping $\partial f$ is locally bounded, i.e. the image $\partial f(B)$ of a bounded set $B \subset \mathbb{R}^n$ is a bounded set in $\mathbb{R}^n$.*

*Proof.* For arbitrary $x$ in $B$ and $s \neq 0$ in $\partial f(x)$, the subgradient inequality implies in particular $f(x + s/\|s\|) \geq f(x) + \|s\|$. On the other hand, $f$ is Lipschitz-continuous on the bounded set $B + B(0, 1)$ (Theorem B.3.1.2). Hence $\|s\| \leq L$ for some $L$. □

**Theorem 4.6.2.** *The subdifferential mapping of a convex function $f : \mathbb{R}^n \to \mathbb{R}$ is outer semi-continuous at any $x \in \mathbb{R}^n$, i.e.*

$$\forall \varepsilon > 0,\ \exists \delta > 0 : \quad y \in B(x, \delta) \implies \partial f(y) \subset \partial f(x) + B(0, \varepsilon). \tag{6.2.1}$$

*Proof.* Assume for contradiction that, at some $x$, there are $\varepsilon > 0$ and a sequence $(x_k, s_k)_k$ with

$$x_k \to x \quad \text{for } k \to \infty \qquad \text{and} \qquad s_k \in \partial f(x_k),\ s_k \notin \partial f(x) + B(0, \varepsilon) \quad \text{for } k = 1, 2, \ldots \tag{6.2.2}$$

A subsequence of the bounded $(s_k)$ (Proposition 6.2.2) converges to $s \in \partial f(x)$ (Proposition 6.2.1). This contradicts (6.2.2), which implies $s \notin \partial f(x) + B(0, 1/2\varepsilon)$. □

**Corollary 4.6.1.** *For $f : \mathbb{R}^n \to \mathbb{R}$ convex, the function $f'(\cdot, d)$ is upper semi-continuous: at all $x \in \mathbb{R}^n$,*

$$f'(x, d) = \limsup_{y \to x} f'(y, d) \qquad \textit{for all } d \in \mathbb{R}^n.$$

*Proof.* Use Theorem 6.2.4, in conjunction with Proposition C.3.3.7. □

**Theorem 4.6.3.** *Let $(f_k)$ be a sequence of (finite) convex functions converging pointwise to $f : \mathbb{R}^n \to \mathbb{R}$ and let $(x_k)$ converge to $x \in \mathbb{R}^n$. For any $\varepsilon > 0$,*

$$\partial f_k(x_k) \subset \partial f(x) + B(0, \varepsilon) \qquad \textit{for } k \textit{ large enough.}$$

*Proof.* Let $\varepsilon > 0$ be given. Recall (Theorem B.3.1.4) that the pointwise convergence of $(f_k)$ to $f$ implies its uniform convergence on every compact set of $\mathbb{R}^n$.

First, we establish boundedness: for $s_k \neq 0$ arbitrary in $\partial f_k(x_k)$, we have

$$f_k(x_k + s_k/\|s_k\|) \geq f_k(x_k) + \|s_k\|.$$

The uniform convergence of $(f_k)$ to $f$ on $B(x, 2)$ implies for $k$ large enough

$$\|s_k\| \leq f(x_k + s_k/\|s_k\|) - f(x_k) + \varepsilon,$$

and the Lipschitz property of $f$ on $B(x, 2)$ ensures that $(s_k)$ is bounded.

Now suppose for contradiction that, for some infinite subsequence, there is some $s_k \in \partial f_k(x_k)$ which is not in $\partial f(x) + B(0, \varepsilon)$. Any cluster point of this $(s_k)$ — and there is at least one — is out of $\partial f(x) + B(0, 1/2\varepsilon)$. Yet, with $y$ arbitrary in $\mathbb{R}^n$, write

$$f_k(y) \geq f_k(x_k) + \langle s_k, y - x_k \rangle$$

and pass to the limit (on a further subsequence such that $s_k \to s$): pointwise [resp. uniform] convergence of $(f_k)$ to $f$ at $y$ [resp. around $x$], and continuity of the scalar product give $f(y) \geq f(x) + \langle s, y - x \rangle$. Because $y$ was arbitrary, we obtain the contradiction $s \in \partial f(x)$. $\square$

**Corollary 4.6.2.** *Let $(f_k)$ be a sequence of (finite) differentiable convex functions converging pointwise to the differentiable $f : \mathbb{R}^n \to \mathbb{R}$. Then $\nabla f_k$ converges to $\nabla f$ uniformly on every compact set of $\mathbb{R}^n$.*

*Proof.* Take $S$ compact; suppose for contradiction that there exists $\varepsilon > 0$ and a sequence $(x_k) \subset S$ such that

$$\|\nabla f_k(x_k) - \nabla f(x_k)\| > \varepsilon \quad \text{for } k = 1, 2, \ldots$$

Extracting a subsequence if necessary, we may suppose $x_k \to x \in S$; Theorem 6.2.7 assures that the sequences $(\nabla f_k(x_k))$ and $(\nabla f(x_k))$ both converge to $\nabla f(x)$, implying $0 \geq \varepsilon$. $\square$

### 4.6.3 Subdifferentials and limits of subgradients

**Theorem 4.6.4.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. With the notation (6.3.1), $\partial f(x) = \operatorname{co} \gamma f(x)$ for all $x \in \mathbb{R}^n$.*

**Proposition 4.6.5.** *Let $x$ and $d \neq 0$ be given in $\mathbb{R}^n$. For any sequence $(t_k, s_k, d_k) \subset \mathbb{R}_+^* \times \mathbb{R}^n \times \mathbb{R}^n$ satisfying*

$$t_k \downarrow 0, \quad s_k \in \partial f(x + t_k d_k), \quad d_k \to d, \qquad \text{for } k = 1, 2, \ldots$$

*and any cluster point $s$ of $(s_k)$, there holds*

$$s \in \partial f(x) \qquad \text{and} \qquad \langle s, d \rangle = f'(x, d).$$

*Proof.* The first property comes from the results in §6.2. For the second, use the monotonicity of $\partial f$:

$$0 \leq \langle s_k - s', x + t_k d_k - x \rangle = t_k \langle s_k - s', d_k \rangle \qquad \text{for all } s' \in \partial f(x).$$

Divide by $t_k > 0$ and pass to the limit to get $f'(x, d) \leq \langle s, d \rangle$. The converse inequality being trivial, the proof is complete. $\square$

# Chapter 5

# Conjugacy in Convex Analysis

## 5.1 The Convex Conjugate of a Function

### 5.1.1 Definition and first examples

**Definition 5.1.1.** The *conjugate* of a function $f$ satisfying (1.1.1) is the function $f^*$ defined by

$$\mathbb{R}^n \ni s \mapsto f^*(s) := \sup\{\langle s, x\rangle - f(x) : x \in \operatorname{dom} f\}. \tag{1.1.2}$$

For simplicity, we may also let $x$ run over the whole space instead of $\operatorname{dom} f$.

The mapping $f \mapsto f^*$ will often be called the *conjugacy* operation, or the Legendre–Fenchel transform.

**Theorem 5.1.1.** *For $f$ satisfying (1.1.1), the conjugate $f^*$ is a closed convex function: $f^* \in \mathrm{Conv}\, \mathbb{R}^n$.*

*Proof.* See Example B.2.1.3. □

### 5.1.2 Interpretations

**Proposition 5.1.1.** *There holds for all $x \in \mathbb{R}^n$*

$$f^*(s) = \sigma_{\operatorname{epi} f}(s, -1) = \sup\{\langle s, x\rangle - r : (x, r) \in \operatorname{epi} f\}. \tag{1.2.1}$$

It follows that the support function of $\operatorname{epi} f$ has the expression

$$\sigma_{\operatorname{epi} f}(s, -u) = \begin{cases} u\, f^*\!\left(\frac{1}{u} s\right) & \text{if } u > 0, \\ \sigma_{\operatorname{epi} f}(s, 0) = \sigma_{\operatorname{dom} f}(s) & \text{if } u = 0, \\ +\infty & \text{if } u < 0. \end{cases} \tag{1.2.2}$$

*Proof.* In (1.2.1), the right-most term can be written

$$\sup_x \sup_{r \geq f(x)} \big[\langle s, x\rangle - r\big] = \sup_x \big[\langle s, x\rangle - f(x)\big]$$

and the first equality is established. As for (1.2.2), the case $u < 0$ is trivial; when $u > 0$, use the positive homogeneity of support functions to get

$$\sigma_{\operatorname{epi} f}(s, -u) = u\sigma_{\operatorname{epi} f}\big(\tfrac{1}{u}s, -1\big) = uf^*\big(\tfrac{1}{u}s\big).$$

Finally, for $u = 0$, we have by definition

$$\sigma_{\operatorname{epi} f}(s, 0) = \sup\{\langle s, x\rangle : (x, r) \in \operatorname{epi} f \text{ for some } r \in \mathbb{R}\},$$

and we recognize $\sigma_{\operatorname{dom} f}(s)$. □

**Proposition 5.1.2.** *For $f \in \operatorname{Conv} \mathbb{R}^n$,*

$$\sigma_{\operatorname{dom} f}(s) = \sigma_{\operatorname{epi} f}(s, 0) = (f^*)^\infty(s) \quad \textit{for all } s \in \mathbb{R}^n. \tag{1.2.3}$$

*Proof.* Use direct calculations; or see Proposition B.2.2.2 and the calculations in Example B.3.2.3. □

### 5.1.3 First properties

**Proposition 5.1.3.** *1.3.1 The functions $f$, $f_j$ appearing below are assumed to satisfy (1.1.1).*

*(i) The conjugate of the function $g(x) := f(x) + r$ is $g^*(s) = f^*(s) - r$.*

*(ii) With $t > 0$, the conjugate of the function $g(x) := tf(x)$ is $g^*(s) = tf^*(s/t)$.*

*(iii) With $t \neq 0$, the conjugate of the function $g(x) := f(tx)$ is $g^*(s) = f^*(s/t)$.*

*(iv) More generally: if $A$ is an invertible linear operator, $(f \circ A)^* = f^* \circ (A^{-1})^*$.*

*(v) The conjugate of the function $g(x) := f(x - x_0)$ is $g^*(s) = f^*(s) + \langle s, x_0\rangle$.*

*(vi) The conjugate of the function $g(x) := f(x) + \langle s_0, x\rangle$ is $g^*(s) = f^*(s - s_0)$.*

*(vii) If $f_1 \leq f_2$, then $f_1^* \geq f_2^*$.*

*(viii) "Convexity" of the conjugation: if $\operatorname{dom} f_1 \cap \operatorname{dom} f_2 \neq \varnothing$ and $\alpha \in ]0, 1[$,*

$$[\alpha f_1 + (1 - \alpha)f_2]^* \leq \alpha f_1^* + (1 - \alpha)f_2^*;$$

*(ix) The Legendre-Fenchel transform preserves decomposition: with*

$$\mathbb{R}^n := \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_m} \ni x \mapsto f(x) := \sum_{j=1}^m f_j(x_j)$$

*and assuming that $\mathbb{R}^n$ has the scalar product of a product-space, there holds*

$$f^*(s_1, \ldots, s_m) = \sum_{j=1}^m f_j^*(s_j).$$

**Proposition 5.1.4.** *Let $f$ satisfy (1.1.1), let $H$ be a subspace of $\mathbb{R}^n$, and call $p_H$ the operator of orthogonal projection onto $H$. Suppose that there is a point in $H$ where $f$ is finite. Then $f + p_{iH}$ satisfies (1.1.1) and its conjugate is*

$$(f + p_{iH})^* = (f \circ p_H)^* \circ p_H. \tag{1.3.1}$$

*Proof.* When $y$ describes $\mathbb{R}^n$, $p_H y$ describes $H$ so we can write, knowing that $p_H$ is symmetric:

$$(f+p_{iH})^*(s) := \sup\{\langle s, x\rangle - f(x) : x \in H\} = \sup\{\langle s, p_H y\rangle - f(p_H y) : y \in \mathbb{R}^n\} = \sup\{\langle p_H s, y\rangle - f(p_H y) : y \in \mathbb{R}^n\}$$

□

**Proposition 5.1.5.** *For $f$ satisfying (1.1.1), let a subspace $V$ contain the subspace parallel to* $\operatorname{aff} \operatorname{dom} f$ *and set $U := V^\perp$. For any $z \in \operatorname{aff} \operatorname{dom} f$ and any $s \in \mathbb{R}^n$ decomposed as $s = s_U + s_V$, there holds*

$$f^*(s) = \langle s_U, z\rangle + f^*(s_V).$$

*Proof.* In (1.1.2), the variable $x$ can range through $z + V \supset \operatorname{aff} \operatorname{dom} f$:

$$\begin{aligned} f^*(s) &= \sup_{v \in V} [\langle s_U + s_V, z + v\rangle - f(z + v)] \\ &= \langle s_U, z\rangle + \sup_{v \in V} [\langle s_V, z + v\rangle - f(z + v)] \\ &= \langle s_U, z\rangle + f^*(s_V). \end{aligned}$$

□

**Theorem 5.1.2.** *For $f$ satisfying (1.1.1), the function $f^{**}$ of (1.3.2) is the pointwise supremum of all the affine functions on $\mathbb{R}^n$ majorized by $f$. In other words*

$$\operatorname{epi} f^{**} = \overline{\operatorname{co}}\big(\operatorname{epi} f\big). \tag{1.3.3}$$

*Proof.* Call $\Sigma \subset \mathbb{R}^n \times \mathbb{R}$ the set of pairs $(s, r)$ defining affine functions $\langle s, \cdot\rangle - r$ majorized by $f$:

$$\begin{aligned} (s, r) \in \Sigma &\iff f(x) \geq \langle s, x\rangle - r \quad \text{for all } x \in \mathbb{R}^n \\ &\iff r \geq \sup\{\, \langle s, x\rangle - f(x) : x \in \mathbb{R}^n \,\} \\ &\iff r \geq f^*(s) \qquad (\text{and } s \in \operatorname{dom} f^*!). \end{aligned}$$

Then we obtain, for $x \in \mathbb{R}^n$,

$$\begin{aligned} \sup_{(s,r) \in \Sigma} \{\langle s, x\rangle - r\} &= \sup\{\langle s, x\rangle - r : s \in \operatorname{dom} f^*,\ -r \leq -f^*(s)\} \\ &= \sup\{\langle s, x\rangle - f^*(s) : s \in \operatorname{dom} f^*\} = f^{**}(x). \end{aligned}$$

Geometrically, the epigraphs of the affine functions associated with $(s, r) \in \Sigma$ are the (non-vertical) closed half-spaces containing $\operatorname{epi} f$. From §B.2.5, the epigraph of their supremum is the closed convex hull of $\operatorname{epi} f$, and this proves (1.3.3). □

**Corollary 5.1.1.** *If $g$ is a function satisfying $\overline{\mathrm{co}}\, f \leq g \leq f$, then $g^* = f^*$. The function $f$ is equal to its biconjugate $f^{**}$ if and only if $f \in \mathrm{Conv}\,\mathbb{R}^n$.*

*Proof.* Immediate. ☐

**Definition 5.1.2.** A function $f$ satisfying (1.1.1) is said to be coercive [resp. 1-coercive] when

$$\lim_{\|x\|\to+\infty} f(x) = +\infty \qquad \Big[\text{resp.} \lim_{\|x\|\to+\infty} \frac{f(x)}{\|x\|} = +\infty\Big].$$

**Proposition 5.1.6.** *If $f$ satisfying (1.1.1) is 1-coercive, then $f^*(s) < +\infty$ for all $s \in \mathbb{R}^n$.*

*Proof.* Let $s$ be given. The 1-coercivity of $f$ implies the existence of a number $R$ such that $f(x) \geq \|s\|\,\|x\|$ (hence $\langle s, x\rangle - f(x) \leq 0$) whenever $\|x\| \geq R$. As a result, we have in (1.1.2)

$$\sup\{\langle s, x\rangle - \|s\|\,\|x\| : \|x\| \geq R\} \leq 0.$$

On the other hand, (1.1.1) implies an upper bound

$$\sup\{\langle s, x\rangle - f(x) : \|x\| \leq R\} \leq M.$$

Altogether, $f^*(s) \leq \max\{0, M\}$. ☐ ☐

**Proposition 5.1.7.** *For $f$ satisfying (1.1.1), the following holds:*

*(i) If $x_0 \in \mathrm{int}\, dom\ f$ then $f^* - \langle x_0, \cdot\rangle$ is 0-coercive;*

*(ii) in particular, if $f$ is finite over $\mathbb{R}^n$, then $f^*$ is 1-coercive.*

*Proof.* We know from (1.2.3) that $\sigma_{\mathrm{dom}\ f} = (f^*)^{\vee}_{\infty}$ so, using Theorem C.2.2.3(iii), $x_0 \in \mathrm{int}\,\mathrm{dom}\ f \subset \int(\mathrm{codom}\ f)$ implies $(f^*)^{\vee}_{\infty}(s) - \langle x_0, s\rangle > 0$ for all $s \neq 0$. By virtue of Proposition B.3.2.4, this means exactly that $f^* - \langle x_0, \cdot\rangle$ has compact sublevel-sets; (i) is proved.

Then, as demonstrated in Definition B.3.2.5, 0-coercivity of $f^* - \langle x_0, \cdot\rangle$ for all $x_0$ means 1-coercivity of $f^*$. ☐

### 5.1.4 Subdifferentials and extended-valued functions

**Theorem 5.1.3.** *For $f$ satisfying (1.1.1) and $\partial f$ defined by (1.4.1), $s \in \partial f(x)$ if and only if*

$$f^*(s) + f(x) - \langle s, x\rangle = 0 \quad (\textit{or}\ \leq 0). \tag{1.4.2}$$

*Proof.* To say that $s$ lies in the set (1.4.1) is to say that

$$\langle s, y\rangle - f(y) \leq \langle s, x\rangle - f(x) \quad \text{for all } y \in \mathrm{dom}\ f,$$

i.e. $f^*(s) \leq \langle s, x\rangle - f(x)$; but this is indeed an equality, in view of Fenchel's inequality (1.1.3). ☐

**Theorem 5.1.4.** *Let $f \in \operatorname{Conv} \mathbb{R}^n$. Then $\partial f(x) \neq \varnothing$ whenever $x \in \operatorname{ri} \operatorname{dom} f$.*

*Proof.* This is Proposition B.1.2.1. $\square$

**Proposition 5.1.8.** *For $f$ satisfying (1.1.1), the following properties hold:*

$$\partial f(x) \neq \varnothing \quad \Longrightarrow \quad (\overline{\operatorname{co}}\, f)(x) = f(x); \tag{1.4.3}$$

$$\overline{\operatorname{co}}\, f \leq g \leq f \quad \text{and} \quad g(x) = f(x) \quad \Longrightarrow \quad \partial g(x) = \partial f(x); \tag{1.4.4}$$

$$s \in \partial f(x) \quad \Longrightarrow \quad x \in \partial f^*(s). \tag{1.4.5}$$

*Proof.* Let $s$ be a subgradient of $f$ at $x$. From the definition (1.4.1) itself, the function $y \mapsto \ell_s(y) := f(x) + \langle s, y - x\rangle$ is affine and minorizes $f$, hence $\ell_s \leq \overline{\operatorname{co}}\, f \leq f$; because $\ell_s(x) = f(x)$, this implies (1.4.3).

Now, $s \in \partial f(x)$ if and only if (1.4.2) holds. From our assumption in (1.4.4), $f^* = g^* = (\overline{\operatorname{co}}\, f)^*$ (Corollary 1.3.6) and $g(x) = f(x)$. Therefore

$$s \in \partial f(x) \iff g^*(s) + g(x) - \langle s, x\rangle = 0,$$

which expresses exactly that $s \in \partial g(x)$; (1.4.4) is proved.

Finally, we know that $f^{**} = \overline{\operatorname{co}}\, f \leq f$; so, when $s$ satisfies (1.4.2), we have

$$f^*(s) + f^{**}(x) - \langle s, x\rangle = f^*(s) + (\overline{\operatorname{co}}\, f)(x) - \langle s, x\rangle \leq 0,$$

which means $x \in \partial f^*(s)$: we have just proved (1.4.5). $\square$

**Corollary 5.1.2.** *If $f \in \operatorname{Conv} \mathbb{R}^n$, the following equivalences hold:*

$$f(x) + f^*(s) - \langle s, x\rangle = 0 \ (\text{or } \leq 0) \quad \iff \quad s \in \partial f(x) \quad \iff \quad x \in \partial f^*(s).$$

*Proof.* This is a rewriting of Theorem 1.4.1, taking into account (1.4.5) and the symmetric role played by $f$ and $f^*$ when $f \in \operatorname{Conv} \mathbb{R}^n$. $\square$

# 5.2 Conjugacy Rules on the Conjugacy Operation

## 5.2.1 Image of a function under a linear mapping

**Theorem 5.2.1.** *With the above notation, assume that $\operatorname{Im} A^* \cap \operatorname{dom} g^* \neq \varnothing$. Then $Ag$ satisfies (1.1.1) and its conjugate is*

$$(Ag)^* = g^* \circ A^*.$$

*Proof.* First, it is clear that $Ag \not\equiv +\infty$ (take $x = Ay$, with $y \in \operatorname{dom} g$). On the other hand, our assumption implies the existence of some $p_0 = A^* s_0$ such that $g^*(p_0) < +\infty$; with Fenchel's inequality (1.1.3), we have for all $y \in \mathbb{R}^m$:

$$g(y) \geq \langle A^* s_0, y\rangle_m - g^*(p_0) = \langle s_0, Ay\rangle_n - g^*(p_0).$$

For each $x \in \mathbb{R}^n$, take the infimum over those $y$ satisfying $Ay = x$: the affine function $\langle s_0, \cdot\rangle - g^*(p_0)$ minorizes $Ag$. Altogether, $Ag$ satisfies (1.1.1).

Then we have for $s \in \mathbb{R}^n$

$$(Ag)^*(s) = \sup_{x\in\mathbb{R}^n} [\langle s, x\rangle - \inf_{Ay=x} g(y)] = \sup_{x\in\mathbb{R}^n,\, Ay=x} [\langle s, x\rangle - g(y)] = \sup_{y\in\mathbb{R}^m} [\langle s, Ay\rangle - g(y)] = g^*(A^* s).$$

□

**Corollary 5.2.1.** *With $g : \mathbb{R}^n \times \mathbb{R}^p = \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ not identically $+\infty$, let $g^*$ be associated with a scalar product preserving the structure of $\mathbb{R}^m$ as a product space: $\langle\cdot,\cdot\rangle_m = \langle\cdot,\cdot\rangle_n + \langle\cdot,\cdot\rangle_p$. If there exists $s_0 \in \mathbb{R}^n$ such that $(s_0, 0) \in \operatorname{dom} g^*$, then the conjugate of $f$ defined by (2.1.2) is*

$$f^*(s) = g^*(s, 0) \qquad \textit{for all } s \in \mathbb{R}^n.$$

*Proof.* It suffices to observe that, $A$ being the projection defined above, there holds for all $y_1 = (x_1, z_1) \in \mathbb{R}^m$ and $x_2 \in \mathbb{R}^n$,

$$\langle Ay_1, x_2\rangle_n = \langle x_1, x_2\rangle_n = \langle x_1, x_2\rangle_n + \langle z_1, 0\rangle_p = \langle y_1, (x_2, 0)\rangle_m,$$

which defines the adjoint $A^*x = (x, 0)$ for all $x \in \mathbb{R}^n$. Then apply Theorem 2.1.1. □

**Corollary 5.2.2.** *Let $f_1$ and $f_2$ be two functions from $\mathbb{R}^n$ to $\mathbb{R} \cup \{+\infty\}$, not identically $+\infty$, and satisfying* $\operatorname{dom} f_1^* \cap \operatorname{dom} f_2^* \neq \varnothing$*. Then their inf-convolution satisfies (1.1.1), and $(f_1 \overset{+}{\vee} f_2)^* = f_1^* + f_2^*$.*

*Proof.* Equip $\mathbb{R}^n \times \mathbb{R}^n$ with the scalar product $\langle\cdot,\cdot\rangle + \langle\cdot,\cdot\rangle$. Using the above notation for $g$ and $A$, we have $g^*(s_1, s_2) = f_1^*(s_1) + f_2^*(s_2)$ (Proposition 1.3.1(ix)) and $A^*(s) = (s, s)$. Then apply the definitions. □

### 5.2.2 Pre-composition with an affine mapping

**Theorem 5.2.2.** *Take $g \in Conv\, \mathbb{R}^m$, $A_0$ linear from $\mathbb{R}^n$ to $\mathbb{R}^m$ and consider the affine operator $A(x) := A_0 x + y_0 \in \mathbb{R}^m$. Suppose that $A(\mathbb{R}^n) \cap dom\ g \neq \emptyset$. Then $g \circ A \in Conv\, \mathbb{R}^n$ and its conjugate is the closure of the convex function*

$$\mathbb{R}^n \ni s \mapsto \inf_p \{\, g^*(p) - \langle y_0, p\rangle_m : A_0^* p = s \,\}. \tag{2.2.1}$$

*Proof.* We start with the linear case ($y_0 = 0$): suppose that $h \in \operatorname{Conv} \mathbb{R}^n$ satisfies $\Im A_0 \cap \operatorname{dom} h \neq \emptyset$. Then Theorem 2.1.1 applied to $g := h^*$ and $A := A_0^*$ gives $(A_0^* h^*)^* = h \circ A_0$; conjugating both sides, we see that the conjugate of $h \circ A_0$ is the closure of the image-function $A_0^* h^*$.

In the affine case, consider the function $h := g(\cdot + y_0) \in \operatorname{Conv} \mathbb{R}^m$; its conjugate is given by Proposition 1.3.1(v): $h^* = g^* - \langle y_0, \cdot\rangle_m$. Furthermore, it is clear that

$$(g \circ A)(x) = g(A_0 x + y_0) = h(A_0 x) = (h \circ A_0)(x),$$

so (2.2.1) follows from the linear case. □

**Lemma 5.2.1.** *Let $g \in \operatorname{Conv} \mathbb{R}^m$ be such that $0 \in dom\ g$ and let $A_0$ be linear from $\mathbb{R}^n$ to $\mathbb{R}^m$. Make the following assumption:*

$$\Im A_0 \cap ri\ dom\ g \neq \emptyset \quad i.e. \quad 0 \in ri\ dom\ g - \Im A_0\ [= ri\ (dom\ g - \Im A_0)].$$

*Then $(g \circ A_0)^* = A_0^* g^*$; for every $s \in dom\ (g \circ A_0)^*$, the problem*

$$\inf_p \{\, g^*(p) : \ A_0^* p = s \,\} \tag{5.1}$$

*has at least one optimal solution $\bar{p}$ and there holds $(g \circ A_0)^*(s) = A_0^* g^*(s) = g^*(\bar{p})$.*

*Proof.* To prove $(g \circ A_0)^* = A_0^* g^*$, we have to prove that $A_0^* g^*$ is a closed function, i.e. that its sublevel-sets are closed (Definition B.1.2.3).

Thus, for given $r \in \mathbb{R}$, take a sequence $(s_k)$ converging to some $s$ and such that

$$(A_0^* g^*)(s_k) \leq r.$$

Take also $\delta_k \downarrow 0$; from the definition of the image-function, we can find $p_k \in \mathbb{R}^m$ such that

$$g^*(p_k) \leq r + \delta_k \quad \text{and} \quad A_0^* p_k = s_k.$$

Let $q_k$ be the orthogonal projection of $p_k$ onto the subspace $V := \operatorname{lin} \operatorname{dom} g - \Im A_0$. Since $V$ contains $\operatorname{lin} \operatorname{dom} g$, Proposition 1.3.4 (with $z = 0$) gives $g^*(p_k) = g^*(q_k)$. Furthermore, $V^\perp = (\operatorname{lin} \operatorname{dom} g)^\perp \cap \operatorname{Ker} A_0^*$; in particular, $q_k - p_k \in \operatorname{Ker} A_0^*$. In summary, we have singled out $q_k \in V$ such that

$$g^*(q_k) \leq r + \delta_k \quad \text{and} \quad A_0^* q_k = s_k \qquad \text{for all } k. \tag{2.2.3}$$

Suppose we can bound $q_k$. Extracting a subsequence if necessary, we will have $q_k \to \bar{q}$ and, passing to the limit, we will obtain (since $g^*$ is l.s.c)

$$g^*(\bar{q}) \leq \liminf g^*(q_k) \leq r \quad \text{and} \quad A_0^* \bar{q} = s.$$

The required closedness property $A_0^* g^*(\bar{q}) \leq r$ will follow by definition. Furthermore, this $\bar{q}$ will be a solution of (2.2.2) in the particular case $s_k \equiv s$ and $r = (A_0^* g^*)(s)$. In this case, $(q_k)$ will be actually a minimizing sequence of (2.2.2).

To prove boundedness of $q_k$, use the assumption: for some $\varepsilon > 0$, $B_m(0, \varepsilon) \cap V$ is included in $\operatorname{dom} g - \Im A$. Thus, for arbitrary $z \in B_m(0, \varepsilon) \cap V$, we can find $y \in \operatorname{dom} g$ and $x \in \mathbb{R}^n$ such that $z = y - A_0 x$. Then

$$\begin{aligned}
\langle q_k, z \rangle_m &= \langle q_k, y \rangle_m - \langle A_0^* q_k, x \rangle_n \\
&\leq g(y) + g^*(q_k) - \langle A_0^* q_k, x \rangle_n && \text{[Fenchel (1.1.3)]} \\
&\leq g(y) + r + \delta_k - \langle s_k, x \rangle_n. && \text{[(2.2.3)]}
\end{aligned}$$

We conclude that $\sup\{\langle q_k, z \rangle : k = 1, 2, \dots\}$ is bounded for any $z \in B_m(0, \varepsilon) \cap V$, which implies that $q_k$ is bounded; this is Proposition V.2.1.3 in the vector space $V$. $\square$

**Theorem 5.2.3.** *Take $g \in \operatorname{Conv} \mathbb{R}^m$, $A_0$ linear from $\mathbb{R}^n$ to $\mathbb{R}^m$ and consider the affine operator $A(x) := A_0 x + y_0 \in \mathbb{R}^m$. Make the following assumption:*

$$A(\mathbb{R}^n) \cap \operatorname{ri} \operatorname{dom} g \neq \varnothing. \tag{2.2.4}$$

*Then, for every $s \in \operatorname{dom}(g \circ A_0)^*$, the problem*

$$\min_p \{\, g^*(p) - \langle p, y_0 \rangle : \; A_0^* p = s \,\} \tag{2.2.5}$$

*has at least one optimal solution $\bar{p}$ and there holds $(g \circ A)^*(s) = g^*(\bar{p}) - \langle \bar{p}, y_0 \rangle$.*

*Proof.* By assumption, we can choose $\bar{x} \in \mathbb{R}^n$ such that $\bar{y} := A(\bar{x}) \in \operatorname{ri} \operatorname{dom} g$. Consider the function $\bar{g} := g(\bar{y} + \cdot) \in \operatorname{Conv} \mathbb{R}^m$. Observing that

$$(g \circ A)(x) = \bar{g}(A(x) - \bar{y}) = (\bar{g} \circ A_0)(x - \bar{x}),$$

we obtain from the calculus rule 1.3.1(v): $(g \circ A)^* = (\bar{g} \circ A_0)^* + \langle \cdot, \bar{x} \rangle$.

Then Lemma 2.2.2 allows the computation of this conjugate. We have 0 in the domain of $\bar{g}$, and even in its relative interior:

$$\operatorname{ri} \operatorname{dom} \bar{g} = \operatorname{ri} \operatorname{dom} g - \{\bar{y}\} \ni 0 \in \operatorname{Im} A_0.$$

We can therefore write: for all $s \in \operatorname{dom}(\bar{g} \circ A_0)^*[= \operatorname{dom}(g \circ A)^*]$,

$$(\bar{g} \circ A_0)^*(s) = \min_p \{\bar{g}^*(p) \; : \; A_0^* p = s\},$$

where the minimum is attained at some $\bar{p}$. Using again the calculus rule 1.3.1(v) and various relations from above, we have established

$$\begin{aligned}(g \circ A)^*(s) - \langle s, \bar{x} \rangle &= \min\{\bar{g}^*(p) - \langle p, A_0 \bar{x} + y_0 \rangle : \; A_0^* p = s\} \\ &= \min\{g^*(p) - \langle p, y_0 \rangle : \; A_0^* p = s\} - \langle s, \bar{x} \rangle.\end{aligned}$$

□

### 5.2.3 Sum of two functions

**Theorem 5.2.4.** *Let $g_1, g_2$ be in $\operatorname{Conv} \mathbb{R}^n$ and assume that $\operatorname{dom} g_1 \cap \operatorname{dom} g_2 \neq \varnothing$. The conjugate $(g_1 + g_2)^*$ of their sum is the closure of the convex function $g_1^* \overset{+}{\vee} g_2^*$.*

*Proof.* Call $f_i^* := g_i$, for $i = 1, 2$; apply Corollary 2.1.3: $(g_1^* \overset{+}{\vee} g_2^*)^* = g_1 + g_2$; then take the conjugate again. □

**Theorem 5.2.5.** *Let $g_1, g_2$ be in Conv $\mathbb{R}^n$ and assume that*

$$\begin{aligned}&\textit{the relative interiors of dom } g_1 \textit{ and dom } g_2 \textit{ intersect,} \\ &\textit{or equivalently: } 0 \in \textit{ri } (\textit{dom } g_1 - \textit{dom } g_2).\end{aligned} \tag{2.3.1}$$

*Then $(g_1 + g_2)^* = g_1^* \overset{+}{\vee} g_2^*$ and, for every $s \in$ dom $(g_1 + g_2)^*$, the problem*

$$\inf\{g_1^*(p) + g_2^*(q) \; : \; p + q = s\}$$

*has at least one optimal solution $(\bar{p}, \bar{q})$, which therefore satisfies*

$$g_1^*(\bar{p}) + g_2^*(\bar{q}) = (g_1^* \overset{+}{\vee} g_2^*)(s) = (g_1 + g_2)^*(s).$$

*Proof.* Define $g \in \text{Conv}\,(\mathbb{R}^n \times \mathbb{R}^n)$ by $g(x_1, x_2) := g_1(x_1) + g_2(x_2)$ and the linear operator $A : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ by $Ax := (x, x)$. Then $g \circ A = g_1 + g_2$, and we proceed to use Theorem 2.2.3. As seen in Proposition 1.3.1(ix), $g^*(p, q) = g_1^*(p) + g_2^*(q)$ and straightforward calculation shows that $A^*(p, q) = p + q$. Thus, if we can apply Theorem 2.2.3, we can write

$$(g_1+g_2)^*(s) = (g \circ A)^*(s) = (A^* g^*)(s) = \inf_{p,q}\{g_1^*(p)+g_2^*(q) \;:\; p+q = s\} = (g_1^* \overset{+}{\vee} g_2^*)(s)$$

and the above minimization problem does have an optimal solution. □

To check (2.2.4), note that $\text{dom}\; g = \text{dom}\; g_1 \times \text{dom}\; g_2$, and $\Im A$ is the diagonal set $\Delta := \{(s, s) : s \in \mathbb{R}^n\}$. We have

$$(x, x) \in \text{ri dom}\; g_1 \times \text{ri dom}\; g_2 = \text{ri}\;(\text{dom}\; g_1 \times \text{dom}\; g_2)$$

(Proposition A.2.1.11), and this just means that $\Im A = \Delta$ has a nonempty intersection with ri dom $g$.

Corollary 2.3.3 Take $f_1$ and $f_2$ in $\text{Conv}\,\mathbb{R}^n$, with $f_1$ 0-coercive and $f_2$ bounded from below. Then the inf-convolution problem (2.1.3) has a nonempty compact set of solutions; furthermore $f_1 \overset{+}{\vee} f_2 \in \text{Conv}\,\mathbb{R}^n$.

*Proof.* Letting $\mu$ denote a lower bound for $f_2$, we have $f_1(x_1) + f_2(x - x_1) \geq f_1(x_1) + \mu$ for all $x_1 \in \mathbb{R}^n$, and the first part of the claim follows.

For closedness of the infimal convolution, we set $g_i := f_i^*$, $i = 1, 2$; because of 0-coercivity of $f_1$, $0 \in \text{int}\,\text{dom}\; g_1$ (Remark 1.3.10), and $g_2(0) \leq -\mu$. Thus, we can apply Theorem 2.3.2 with the qualification assumption (2.3.Q.ij'). □

### 5.2.4 Infima and suprema

**Theorem 5.2.6.** *Let $\{f_j\}_{j \in J}$ be a collection of functions satisfying (1.1.1) and having a common affine minorant:* $\sup_{j \in J} f_j^*(s) < +\infty$ *for some $s \in \mathbb{R}^n$. Then their infimum $f := \inf_{j \in J} f_j$ satisfies (1.1.1), and its conjugate is the supremum of the $f_j^*$'s:*

$$(\inf_{j \in J} f_j)^* = \sup_{j \in J} f_j^*. \tag{2.4.1}$$

*Proof.* By definition, for all $s \in \mathbb{R}^n$

$$\begin{aligned} f^*(s) &= \sup_x \left[\langle s, x\rangle - \inf_{j \in J} f_j(x)\right] \\ &= \sup_x \sup_j \left[\langle s, x\rangle - f_j(x)\right] \\ &= \sup_j \sup_x \left[\langle s, x\rangle - f_j(x)\right] = \sup_{j \in J} f_j^*(s). \end{aligned}$$

□

**Theorem 5.2.7.** *Let $\{g_j\}_{j\in J}$ be a collection of functions in Conv $\mathbb{R}^n$. If their supremum $g := \sup_{j\in J} g_j$ is not identically $+\infty$, it is in Conv $\mathbb{R}^n$, and its conjugate is the closed convex hull of the $g_j^*$'s:*

$$\big(\sup_{j\in J} g_j\big)^* = \overline{co}\,\big(\inf_{j\in J} g_j^*\big). \tag{2.4.5}$$

*Proof.* Call $f_j := g_j^*$, hence $f_j^* = g_j$, and $g$ is nothing but the $f^*$ of (2.4.1). Taking the conjugate of both sides, the result follows from (1.3.4). ☐

### 5.2.5 Post-composition with an increasing convex function

**Theorem 5.2.8.** *With $f$ and $g$ defined as above, assume that $f(\mathbb{R}^n)\cap \operatorname{int}\operatorname{dom} g \neq \varnothing$. For all $s \in \operatorname{dom}(g\circ f)^*$, define the function $\psi_s \in \operatorname{Conv}\mathbb{R}$ by*

$$\mathbb{R} \ni \alpha \mapsto \psi_s(\alpha) := \begin{cases} \alpha f^*\big(\frac{1}{\alpha}s\big) + g^*(\alpha) & \text{if } \alpha > 0, \\ \sigma_{\operatorname{dom} f}(s) + g^*(0) & \text{if } \alpha = 0, \\ +\infty & \text{if } \alpha < 0. \end{cases}$$

*Then $(g\circ f)^*(s) = \min_{\alpha\in\mathbb{R}} \psi_s(\alpha)$.*

*Proof.* By definition,

$$-(g\circ f)^*(s) = \inf_x [g(f(x)) - \langle s, x\rangle] = \inf_{x,r}\{g(r) - \langle s,x\rangle : \ f(x) \le r\} = \inf_{x,r}[g(r) - \langle s, x\rangle + \iota_{\operatorname{epi} f}(x,r)].$$

[ $g$ is increasing]

We must compute the conjugate of the sum of the two functions $f_1(x,r) := g(r) - \langle s, x\rangle$ and $f_2 := \iota_{\operatorname{epi} f}$, at the obvious argument $(x,r)\in\mathbb{R}^n\times\mathbb{R}$. We have $\operatorname{dom} f_1 = \mathbb{R}^n \times \operatorname{dom} g$, so that $\operatorname{dom} f_1 = \mathbb{R}^n\times\operatorname{dom} g$; hence, by assumption:

$$\operatorname{int}\operatorname{dom} f_1 \cap \operatorname{dom} f_2 = (\mathbb{R}^n \times \operatorname{int}\operatorname{dom} g)\cap \operatorname{epi} f \neq \varnothing.$$

Theorem 2.3.2, more precisely Fenchel's duality theorem (2.3.2), can be applied with the qualification condition (2.3.Q.jj'):

$$(g\circ f)^*(s) = \min\{f_1^*(-p,\alpha) + f_2^*(p,-\alpha) : \ (p,\alpha)\in\mathbb{R}^n\times\mathbb{R}\}.$$

The computation of the above two conjugates is straightforward and gives

$$(g\circ f)^*(s) = \min_{p,\alpha}\big[g^*(\alpha) + \iota_{\{-s\}}(-p) + \sigma_{\operatorname{epi} f}(p,-\alpha)\big] = \min_\alpha \psi_s(\alpha),$$

where the second equality comes from (1.2.2). ☐

## 5.3 Various Examples

### 5.3.1 The Cramer transform

### 5.3.2 The conjugate of a convex partially quadratic function

**Proposition 5.3.1.** *The function $g$ of (3.2.1) has the conjugate*

$$g^*(s) = \begin{cases} \frac{1}{2}\langle s, (p_H \circ B \circ p_H)^- s\rangle & \text{if } s \in \operatorname{Im} B + H^\perp, \\ +\infty & \text{otherwise}, \end{cases} \tag{3.2.2}$$

*where $p_H$ is the operator of orthogonal projection onto $H$ and $(\cdot)^-$ is the Moore–Penrose pseudo-inverse.*

*Proof.* Set $f := \frac{1}{2}\langle B\cdot, \cdot\rangle$, so that $g = f + i_H$ and $g^* = (f \circ p_H)^* \circ p_H$ (Proposition 1.3.2). Knowing that $(f \circ p_H)(x) = \frac{1}{2}\langle (p_H \circ B \circ p_H)x, x\rangle$, we obtain from Example 1.1.4 the conjugate $g^*(s)$ under the form

$$(f \circ p_H)^*(p_H s) = \begin{cases} \frac{1}{2}\langle s, (p_H \circ B \circ p_H)^- s\rangle & \text{if } p_H s \in \operatorname{Im}(p_H \circ B \circ p_H), \\ +\infty & \text{otherwise.} \end{cases}$$

It could be checked directly that $\operatorname{Im}(p_H \circ B \circ p_H) + H^\perp = \operatorname{Im} B + H^\perp$. A simpler argument, however, is obtained via Theorem 2.3.2, which can be applied since $\operatorname{dom} f = \mathbb{R}^n$. Thus,

$$g^*(s) = (f^* \,\square\, i_{H^\perp})(s) = \min\{\tfrac{1}{2}\langle p, B^- p\rangle : \ p \in \operatorname{Im} B, \ s - p \in H^\perp\},$$

which shows that $\operatorname{dom} g^* = \operatorname{Im} B + H^\perp$. □

### 5.3.3 Polyhedral functions

**Proposition 5.3.2.** *At each $s \in \operatorname{co}\{s_1, \dots, s_k\} = \operatorname{dom} f^*$, the conjugate of $f$ has the value ( $\Delta_k$ is the unit simplex)*

$$f^*(s) = \min\Big\{\sum_{i=1}^k \alpha_i b_i : \ \alpha \in \Delta_k, \ \sum_{i=1}^k \alpha_i s_i = s\Big\}. \tag{3.3.2}$$

*Proof.* Set $g_i(s) := \iota_{\{s_i\}} + b_i$ and

$$g(s) := (\inf_i g_i)(s) = \begin{cases} b_i & \text{if } s = s_i \text{ for some } i = 1, \dots, k, \\ +\infty & \text{otherwise.} \end{cases}$$

Apply Proposition B.2.5.4 to see that $\overline{\operatorname{co}}\, g = f^*$ of (3.3.2). The rest follows from Theorem 2.4.1 or 2.4.4, with a notational flip of $f$ and $g$. □

## 5.4 Differentiability of a Conjugate Function

### 5.4.1 First-order differentiability

**Theorem 5.4.1.** *Let* $f \in \operatorname{Conv} \mathbb{R}^n$ *be strictly convex. Then* $\operatorname{int} dom\, f^* \neq \varnothing$ *and* $f^*$ *is continuously differentiable on* $\operatorname{int} dom\, f^*$*.*

*Proof.* For arbitrary $x_0 \in \operatorname{dom} f$ and nonzero $d \in \mathbb{R}^n$, consider Example 2.4.6. Strict convexity of $f$ implies that

$$0 < \frac{f(x_0 - td) - f(x_0)}{t} + \frac{f(x_0 + td) - f(x_0)}{t} \qquad \text{for all } t > 0,$$

and this inequality extends to the suprema: $0 < f'_\infty(-d) + f'_\infty(d)$. Remembering that $f'_\infty = \sigma_{\operatorname{dom} f^*}$ (Proposition 1.2.2), this means

$$\sigma_{\operatorname{dom} f^*}(d) + \sigma_{\operatorname{dom} f^*}(-d) > 0,$$

i.e. $\operatorname{dom} f^*$ has a positive breadth in every nonzero direction $d$: its interior is nonempty—Theorem C2.2.3(iii).

Now suppose that there is some $s \in \int \operatorname{dom} f^*$ such that $\partial f^*(s)$ contains two distinct points $x_1$ and $x_2$. Then $s \in \partial f(x_1) \cap \partial f(x_2)$; by convex combination of the relations

$$f^*(s) + f(x_i) = \langle s, x_i \rangle \qquad \text{for } i = 1, 2$$

we deduce, using Fenchel's inequality (1.1.3):

$$f^*(s) + \sum_{i=1}^{2} \alpha_i f(x_i) = \langle s, \sum_{i=1}^{2} \alpha_i x_i \rangle \leq f^*(s) + f\Big(\sum_{i=1}^{2} \alpha_i x_i\Big),$$

which implies that $f$ is affine on $[x_1, x_2]$, a contradiction. In other words, $\partial f^*$ is single-valued on $\int \operatorname{dom} f^*$, and this means that $f^*$ is continuously differentiable there (Remark 6.2.6). □

**Theorem 5.4.2.** *Let* $f \in \operatorname{Conv} \mathbb{R}^n$ *be differentiable on the set* $\Omega := \operatorname{int} \operatorname{dom} f$*. Then* $f^*$ *is strictly convex on each convex subset* $C \subset \nabla f(\Omega)$*.*

*Proof.* Let $C$ be a convex set as stated. Suppose that there are two distinct points $s_1$ and $s_2$ in $C$ such that $f^*$ is affine on the line-segment $[s_1, s_2]$. Then, setting $s := \frac{1}{2}(s_1 + s_2) \in C \subset \nabla f(\Omega)$, there is $x \in \Omega$ such that $\nabla f(x) = s$, i.e. $x \in \partial f^*(s)$. Using the affine character of $f^*$, we have

$$0 = f(x) + f^*(s) - \langle s, x \rangle = \tfrac{1}{2} \sum_{i=1}^{2} \big[ f(x) + f^*(s_i) - \langle s_i, x \rangle \big]$$

and, in view of Fenchel's inequality (1.1.3), this implies that each term in the bracket is 0: $x \in \partial f^*(s_1) \cap \partial f^*(s_2)$, i.e. $\partial f(x)$ contains the two points $s_1$ and $s_2$, a contradiction to the existence of $\nabla f(x)$. □

**Corollary 5.4.1** (Corollary 4.1.3)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be strictly convex, differentiable and 1-coercive. Then*

*(i) $f^*$ is also finite-valued on $\mathbb{R}^n$, strictly convex, differentiable and 1-coercive;*

*(ii) the continuous mapping $\nabla f$ is one-to-one from $\mathbb{R}^n$ onto $\mathbb{R}^n$, and its inverse is continuous;*

$$\textit{(iii)}\quad f^*(s) = \langle s, (\nabla f)^{-1}(s)\rangle - f\big((\nabla f)^{-1}(s)\big) \quad \textit{for all } s \in \mathbb{R}^n.$$

### 5.4.2 Lipschitz continuity of the gradient mapping

**Theorem 5.4.3.** *Assume that $f : \mathbb{R}^n \to \mathbb{R}$ is strongly convex with modulus $c > 0$ on $\mathbb{R}^n$: for all $(x_1, x_2) \in \mathbb{R}^n \times \mathbb{R}^n$ and $\alpha \in ]0, 1[$,*

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2) - \tfrac{1}{2}c\alpha(1-\alpha)\|x_1 - x_2\|^2. \quad (4.2.1)$$

*Then* $\operatorname{dom} f^* = \mathbb{R}^n$ *and $\nabla f^*$ is Lipschitzian with constant $1/c$ on $\mathbb{R}^n$:*

$$\|\nabla f^*(s_1) - \nabla f^*(s_2)\| \leq \tfrac{1}{c}\|s_1 - s_2\| \quad \textit{for all } (s_1, s_2) \in \mathbb{R}^n \times \mathbb{R}^n.$$

*Proof.* We use the various equivalent definitions of strong convexity (see Theorem D.6.1.2). Fix $x_0$ and $s_0 \in \partial f(x_0)$: for all $0 \neq d \in \mathbb{R}^n$ and $t \geq 0$

$$f(x_0 + td) \geq f(x_0) + t\langle s_0, d\rangle + \tfrac{1}{2}ct^2\|d\|^2,$$

hence $f'_\infty(d) = \sigma_{\operatorname{dom} f^*}(d) = +\infty$, i.e. $\operatorname{dom} f^* = \mathbb{R}^n$. Also, $f$ is in particular strictly convex, so we know from Theorem 4.1.1 that $f^*$ is differentiable (on $\mathbb{R}^n$). Finally, strong convexity of $f$ can also be written $(s_1 - s_2, x_1 - x_2) \geq c\|x_1 - x_2\|^2$, in which we have $s_i \in \partial f(x_i)$, i.e. $s_i = \nabla f^*(s_i)$, for $i = 1, 2$. The rest follows from the Cauchy–Schwarz inequality. □

**Theorem 5.4.4.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and have a gradient-mapping Lipschitzian with constant $L > 0$ on $\mathbb{R}^n$: for all $(x_1, x_2) \in \mathbb{R}^n \times \mathbb{R}^n$,*

$$\|\nabla f(x_1) - \nabla f(x_2)\| \leq L\|x_1 - x_2\|.$$

*Then $f^*$ is strongly convex with modulus $1/L$ on each convex subset $C \subset$* $\operatorname{dom} \partial f^*$. *In particular, there holds for all $(x_1, x_2) \in \mathbb{R}^n \times \mathbb{R}^n$*

$$\langle \nabla f(x_1) - \nabla f(x_2),\, x_1 - x_2\rangle \geq \tfrac{1}{L}\|\nabla f(x_1) - \nabla f(x_2)\|^2. \quad (4.2.2)$$

*Proof.* Let $s_1$ and $s_2$ be arbitrary in $\operatorname{dom} \partial f^* \subset \operatorname{dom} f^*$; take $s$ and $s'$ on the segment $[s_1, s_2]$. To establish the strong convexity of $f^*$, we need to minorize the remainder term $f^*(s') - f^*(s) - \langle x, s' - s\rangle$, with $x \in \partial f^*(s)$. For this, we minorize $f^*(s') = \sup_y\{\langle s', y\rangle - f(y)\}$, i.e. we majorize $f(y)$:

$$\begin{aligned} f(y) &= f(x) + \langle \nabla f(x),\, y - x\rangle + \int_0^1 \langle \nabla f(x + t(y-x)) - \nabla f(x),\, y - x\rangle dt \\ &\leq f(x) + \langle \nabla f(x),\, y - x\rangle + \tfrac{1}{2}L\|y - x\|^2 \\ &= -f^*(s) + \langle s, y\rangle + \tfrac{1}{2}L\|y - x\|^2 \end{aligned}$$

(we have used the property $\int_0^1 t\,dt = 1/2$, as well as $x \in \partial f^*(s)$, i.e. $\nabla f(x) = s$). In summary, we have

$$f^*(s') \geq f^*(s) + \sup_y \left[\, \langle s' - s, y\rangle - \tfrac{1}{2}L\|y - x\|^2 \right].$$

Observe that the last supremum is nothing but the value at $s'-s$ of the conjugate of $\frac{1}{2}L\|\cdot\|^2$. Using the calculus rule 1.3.1, we have therefore proved

$$f^*(s') \geq f^*(s) + \langle s' - s, x\rangle + \tfrac{1}{2L}\|s' - s\|^2 \tag{5.2}$$

for all $s, s'$ in $[s_1, s_2]$ and all $x \in \partial f^*(s)$. Replacing $s'$ in (4.2.3) by $s_1$ and by $s_2$, and setting $s = \alpha s_1 + (1-\alpha)s_2$, the strong convexity (4.2.1) for $f^*$ is established by convex combination.

On the other hand, replacing $(s, s')$ by $(s_1, s_2)$ in (4.2.3):

$$f^*(s_2) \geq f^*(s_1) + \langle s_2 - s_1, x_1\rangle + \tfrac{1}{2L}\|s_2 - s_1\|^2 \quad \text{for all } x_1 \in \partial f^*(s_1).$$

Then, replacing $(s, s')$ by $(s_2, s_1)$ and summing: $\langle x_1 - x_2, s_1 - s_2\rangle \geq \frac{1}{L}\|s_1 - s_2\|^2$. In view of the differentiability of $f$, this is just (4.2.2), which has to hold for all $(x_1, x_2)$ simply because $\Im \partial f^* = \operatorname{dom} \nabla f = \mathbb{R}^n$. $\square$