A SET OF LECTURE NOTES ON CONVEX OPTIMIZATION WITH SOME APPLICATIONS TO PROBABILITY THEORY INCOMPLETE DRAFT. MAY 06

CHRISTIAN LÉONARD

Contents

Preliminaries

This is an incomplete draft. In this version of the notes, I introduce basic tools of convex analysis and the saddle-point method. As an application of this powerful method, I derive the basic results of the optimal transport theory.
Other probabilistic applications in connection with large deviations are possible. I will try to write some of them later.
The reference list hasn't been worked at this stage. No reference to the literature is given, I only state references to precise results in the body of the proofs. Nevertheless, before completing these notes, I inform the reader that good references for convex analysis are the monographs of R. T. Rockafellar: [7], [6] and good references for the optimal transport are the books of S. Rachev and L. Rüschendorf [5] and of C. Villani [9].

The minimization problem. We are interested in the optimization problem

$$\text{minimize } h(x) \text{ subject to } x \in C \tag{0.1}$$

where h is a convex function and C a convex set. We wish to find the minimum value $\inf_{x\in C} h(x)$ together with the x_*'s in C for which $h(x_*)$ is minimal.

Definition 0.2 (Value and Minimizers)**.** *Such x_*'s are the solutions of the problem (0.1), they are called* minimizers *while* $\inf_{x\in C} h(x)$ *is called the* value *of the minimization problem.*

1. Convexity without topology

1.1. **Basic notions.** For convexity to be defined, one needs an addition and a scalar multiplication. This means that the underlying space X must be a vector space.

Definition 1.1 (Convex set)**.** *A subset C of the vector space X is* convex *if for all $x_0, x_1 \in C$, the segment $[x_0, x_1] := \{x_t = (1-t)x_0 + tx_1 \in X; 0 \le t \le 1\}$ is included in C. As a convention, the empty set is considered to be convex.*

Date: May 21, 2006.

Exercice 1.2. *Show that the image of a convex subset of X by a linear transformation from X to some vector space Y is a convex subset of Y.*

Definition 1.3 (Convex hull)**.** *The* convex hull *of a subset B of X is the smallest convex set which contains B. It is denoted* $\mathrm{cv}\, B$.

The following proposition states that the convex hull of any set exists.

Proposition 1.4. *Let $\{C_i; i \in I\}$ be a collection of convex subsets of X. Then its intersection $\cap_{i\in I} C_i$ is still a convex set.*
The convex hull of B is

$$\mathrm{cv}\, B = \bigcap \{C \subset X : C \textit{ convex}, B \subset C\}.$$

Proof. The last statement is a direct consequence of the first one.
Let us prove the first statement. Let x_0, x_1 stand in $\cap_{i\in I} C_i$. This means that for all $i \in I$, $x_0, x_1 \in C_i$. As C_i is convex, we have $[x_0, x_1] \subset C_i$, for all $i \in I$. That is $[x_0, x_1] \subset \cap_{i\in I} C_i$ which is the desired result. □

It will turn out to be comfortable to work with *extended real-valued* functions on X, that is $\mathbb{R}\cup\{-\infty, +\infty\} = [-\infty, +\infty]$-valued functions. For instance, considering for any $x \in X$,

$$f(x) = \begin{cases} h(x) & \text{if } x \in C \\ +\infty & \text{otherwise} \end{cases} \tag{1.5}$$

the minimization problem (0.1) is equivalent to

$$\text{minimize } f(x), \quad x \in X \tag{1.6}$$

Let us introduce the useful notions of effective domain and epigraph of a function.

Definition 1.7 (Effective domain)**.** *The* effective domain *of the extended real-valued function f is* $\mathrm{dom}\, f := \{x \in X; f(x) < +\infty\}$.

Clearly, with f given by (1.5), $\mathrm{dom}\, f$ is a subset of C.

Definition 1.8 (Epigraph)**.** *The* epigraph *of an extended real-valued function f is the subset of $X \times \mathbb{R}$ defined by* $\mathrm{epi}\, f := \{(x, \alpha) \in X \times \mathbb{R}; f(x) \le \alpha\}$.

Note that $(x, \alpha) \in \mathrm{epi}\, f$ implies that $x \in \mathrm{dom}\, f$. More precisely the projection $\{x \in X; (x, \alpha) \in \mathrm{epi}\, f$ for some $\alpha \in \mathbb{R}\}$ is exactly $\mathrm{dom}\, f$.
The epigraph of f is the set of all points in $X \times \mathbb{R}$ that lie above the graph of f (including this graph).
One usually defines a real valued function f (with $\mathrm{dom}\, f$ the entire space) to be convex if for any x_0, x_1 in X, the graph of f restricted to the segment $[x_0, x_1]$ lies below the chord joining $(x_0, f(x_0))$ to $(x_1, f(x_1))$. In other words, f is a convex function if for all $x_0, x_1 \in X$ and all $0 \le t \le 1$, $f((1-t)x_0 + tx_1) \le (1-t)f(x_0) + tf(x_1)$. The following definition extends this notion to extended real-valued functions.

Definition 1.9 (Convex function)**.** *The extended real-valued function f on X is* convex *if its epigraph is a convex set.*

Conventions 1.10. From now on, all functions will be supposed to be $[-\infty, +\infty]$-valued. The function which is identically $+\infty$ is convex since $\mathrm{epi}\,(+\infty) = \emptyset$ is a convex set.

Proposition 1.11. *Let us take an extended real-valued function $f : X \to [-\infty, \infty]$.*

(a) If f is a convex function, $\mathrm{dom}\, f$ *is a convex set.*

(b) f is a convex function if and only if for all $x_0, x_1 \in \operatorname{dom} f$ and all $0 \le t \le 1$,

$$f((1-t)x_0 + tx_1) \le (1-t)f(x_0) + tf(x_1).$$

Proof. The easy proof of (b) is left as an exercice.
Let us prove (a). As $\operatorname{dom} f = \{x \in X : \exists \alpha \in \mathbb{R}, (x, \alpha) \in \operatorname{epi} f\}$, it is the canonical projection of the epigraph of f onto X. Therefore, $\operatorname{dom} f$ is the image of the convex set $\operatorname{epi} f$ by a linear transformation. Hence, it is a convex set. □

Examples 1.12. Here are typical convex functions.

(1) On $X = \mathbb{R}$, $f(x) = |x|^p$ with $p \ge 1$, $f(x) = e^{ax}$ with $a \in \mathbb{R}$, or for any $a \ge 0$,
$$f(x) = \begin{cases} x \log x & \text{if } x > 0 \\ a & \text{if } x = 0 \\ +\infty & \text{if } x < 0 \end{cases}$$
(2) Any twice differentiable real-valued function f on $\mathbb{R}^d$ with a nonnegative Hessian matrix $D^2 f(x) = (\partial_i \partial_j f(x))_{1 \le i,j \le d}$ i.e. $\langle \xi, D^2 f(x) \xi \rangle \ge 0$ for all $x, \xi \in \mathbb{R}^d$.
(3) Any affine function $f(x) = \langle u, x \rangle + \alpha$ where u is a linear form on X and α is real.
(4) Any function $f(x) = g(\langle u, x \rangle)$ where u is a linear form on X and g is an extended real-valued convex function on $\mathbb{R}$.
(5) More generally, $f(x) = g(Ax)$ where A is a linear operator and g is a convex function.
(6) Any function $f(x) = g(\|x\|)$ where $\|x\|$ is a norm on X and g is an extended real-valued increasing convex function on $[0, \infty)$. Typically, $\|x\|^p$ with $p \ge 1$.
(7) Beware: the composition of two convex functions may not be convex. For instance $f(x) = |x|$ and $g(y) = -y$ are convex but $g[f(x)] = -|x|$ isn't. Another counterexample is by $g \circ f$ with $f(x) = x^2$ and $g(y) = e^{-y}$.
(8) Let $\varphi(t, x)$ be a function on $T \times X$ such that for each $t \in T$, $x \mapsto \varphi(t, x)$ is a $(-\infty, \infty]$-valued convex function. For any nonnegative measure $\mu(dt)$ on T, $x \mapsto \int_T \varphi(t, x)\, \mu(dt)$ is a convex function, provided that this integral is meaningful. In particular, any nonnegative linear combination of convex functions is convex.
(9) If you are used to abstract integration and probability theory, you already know the indicator of a set: $\mathbf{1}_C(x) = \begin{cases} 1 & \text{if } x \in C \\ 0 & \text{otherwise} \end{cases}$ which is not convex. Convex analysis requires another indicator which is defined by
$$\zeta_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise} \end{cases}$$
Of course, $\operatorname{epi} \zeta_C = C \times [0, \infty)$ and C is a convex set if and only if ζ_C is a convex function. We also have formally $\mathbf{1}_C = e^{-\zeta_C}$.

Definition 1.13 (Strictly convex function)**.** *A function f on X is* strictly convex *if* $\operatorname{dom} f$ *is a convex set and $f((1-t)x_0 + tx_1) < (1-t)f(x_0) + tf(x_1)$ for all distinct $x_0, x_1 \in X$ and all $0 < t < 1$.*

Of course, a strictly convex function is convex. Strict convexity is very useful to derive uniqueness results.

Proposition 1.14 (Uniqueness of the minimizer)**.** *Let f be a stricly convex function, then it admits at most one minimizer.*

Proof. Suppose that x_0 and x_1 are two distinct minimizers of f. They are in $\operatorname{dom} f$, $x_{0.5} = (x_0 + x_1)/2$ is also in $\operatorname{dom} f$ and $\inf f \le f(x_{0.5}) < (f(x_0) + f(x_1))/2 = \inf f$, which is a contradiction. □

Definition 1.15 (Convex envelope)**.** *Let f be any extended real-valued function on X. Its* convex envelope *is the greatest convex function which is less than f. It is denoted* $\mathrm{cv}\, f$.

The following proposition shows that this definition is meaningful.

Proposition 1.16. *Let $\{f_i; i \in I\}$ be a collection of convex functions. Then $\sup_i f_i$ is also convex.*
The convex envelope of a function f is

$$\mathrm{cv}\, f = \sup\{g : g \text{ convex extended real-valued function on } X \text{ such that } g \le f\}.$$

Proof. The last statement is a direct consequence of the first one.
To prove the first statement, note that the epigraph of a supremum is the intersection of the epigraphs (see Lemma 1.17 below) and the intersection of convex sets is a convex set (Proposition 1.4). □

Lemma 1.17. *Let $\{f_i; i \in I\}$ be a collection of extended real-valued functions. Then*

$$\bigcap_{i\in I} \mathrm{epi}\, f_i = \mathrm{epi}\, \sup_{i\in I} f_i.$$

Proof. For any $x \in X$, the section $(\{x\} \times \mathbb{R}) \cap (\cap_i \mathrm{epi}\, f_i)$ is $\cap_i[(\{x\} \times \mathbb{R}) \cap \mathrm{epi}\, f_i] = \{x\} \times \cap_i [f_i(x), +\infty) = \{x\} \times [\sup_i f_i(x), +\infty) = (\{x\} \times \mathbb{R}) \cap \mathrm{epi}\, \sup_i f_i$. □

Notations 1.18. We write $\inf f$ for short instead of $\inf_{x\in X} f(x)$.
The set of all the (global) minimizers of f is denoted $\mathrm{argmin}\, f$.

The next result relates the global minimization problems of f and $\mathrm{cv}\, f$.

Proposition 1.19. *Let g be a convex function, then* $\mathrm{argmin}\, g$ *is a convex set.*
Let f be any function on X. Then, $\inf f = \inf \mathrm{cv}\, f$ *and* $\mathrm{argmin}\, f \subset \mathrm{cv}\,\mathrm{argmin}\, f \subset \mathrm{argmin}\, \mathrm{cv}\, f$.

Proof. Let g be a convex function and $x_o, x_1 \in \mathrm{argmin}\, g$. Then, for all $0 \le t \le 1$, $\inf g \le g((1-t)x_o + tx_1) \le (1-t)g(x_o) + tg(x_1) = (1-t)\inf g + t \inf g = \inf g$. This proves that $[x_o, x_1] \subset \mathrm{argmin}\, g$ and means that $\mathrm{argmin}\, g$ is a convex set.
Let f be any function. As the constant function $\inf f$ is convex and below f, we have $\inf f \le \mathrm{cv}\, f \le f$. This yields the identity $\inf f = \inf \mathrm{cv}\, f$ together with the inclusion $\mathrm{argmin}\, f \subset \mathrm{argmin}\, \mathrm{cv}\, f$. Now, taking the convex hulls on both sides of this inclusion, one gets: $\mathrm{cv}\,\mathrm{argmin}\, f \subset \mathrm{cv}\,\mathrm{argmin}\, \mathrm{cv}\, f = \mathrm{argmin}\, \mathrm{cv}\, f$ since $\mathrm{argmin}\, \mathrm{cv}\, f$ is a convex set (first part of this proposition with $g = \mathrm{cv}\, f$). □

Examples 1.20. We give some examples on $X = \mathbb{R}$ to illustrate this proposition.

1. Let $f(x) = 1/(1+x^2), x \in \mathbb{R}$. Its convex envelope is $\mathrm{cv}\, f \equiv 0$, $\mathrm{argmin}\, f$ is empty and $\mathrm{argmin}\, \mathrm{cv}\, f = \mathbb{R}$. Note that $\emptyset = \mathrm{cv}\,\mathrm{argmin}\, f \subsetneq \mathrm{argmin}\, \mathrm{cv}\, f = \mathbb{R}$.
2. Let $f(x) = x^4 - x^2, x \in \mathbb{R}$. We have $\mathrm{argmin}\, f = \{-1/\sqrt{2}, 1/\sqrt{2}\}$ and $\inf f = -1/4$. The convex envelope of f is $\mathrm{cv}\, f(x) = \begin{cases} \inf f, & \text{if } x \in [-1/\sqrt{2}, 1/\sqrt{2}]; \\ f(x), & \text{otherwise.} \end{cases}$ Therefore, $\mathrm{argmin}\, \mathrm{cv}\, f = \mathrm{cv}\,\mathrm{argmin}\, f = [-1/\sqrt{2}, 1/\sqrt{2}]$ and $\inf \mathrm{cv}\, f = \inf f = -1/4$.
3. Let $f(x) = \begin{cases} +\infty, & \text{if } x < 0 \\ -x^2, & \text{if } x \ge 0 \end{cases}$ Then, $\mathrm{cv}\, f(x) = \begin{cases} +\infty, & \text{if } x < 0 \\ 0, & \text{if } x = 0. \\ -\infty, & \text{if } x > 0 \end{cases}$ Then, $\inf \mathrm{cv}\, f = \inf f = -\infty$ and $\mathrm{argmin}\, f = \emptyset \subsetneq \mathrm{argmin}\, \mathrm{cv}\, f = (0, +\infty)$.
4. Let $f(x) = \begin{cases} |x|, & \text{if } x \ne 0 \\ 1, & \text{if } x = 0 \end{cases}$ Then, $\mathrm{cv}\, f(x) = |x|$ for all $x \in \mathbb{R}$ and $\emptyset = \mathrm{cv}\,\mathrm{argmin}\, f \subsetneq \mathrm{argmin}\, \mathrm{cv}\, f = \{0\}$.

(5) Let $f(x) = \begin{cases} x^2, & \text{if } x \le 1 \\ x, & \text{if } x \ge 1 \end{cases}$. Then $\operatorname{cv} f(x) = \begin{cases} x^2, & \text{if } x \le 1/2 \\ x - 1/4 & \text{if } x \ge 1/2 \end{cases}$, so that $f \neq \operatorname{cv} f$, but $\operatorname{argmin} f = \operatorname{cv} \operatorname{argmin} f = \operatorname{argmin} \operatorname{cv} f = \{0\}$.

1.2. **Subdifferentiability.** Working with convex functions usually requires less regularity than the standard differentiability. Nevertheless, the notion of affine approximations of a convex function is crucial. For this to be defined, one needs a vector space U of linear functions on X. *At this stage, it is not necessary to consider a topological structure on X.* It is not required either that U is the space of *all* linear forms on X. We introduce a

Definition 1.21 (Algebraic dual pairing)**.** *Let U and X be two vector spaces. An* algebraic dual pairing *of U and X is a bilinear form $\langle u, x\rangle, u \in U, x \in X$.*

Note that this notion gives a symmetric role to U and X : X acts linearly on U and U acts linearly on X.
Also note that it is not supposed that this pairing separates U or X.
An *affine function* on X is given by

$$x \in X \mapsto \langle u, x\rangle + \alpha \in \mathbb{R}$$

with $u \in U$ and $\alpha \in \mathbb{R}$ and an affine function on U is given by

$$u \in U \mapsto \langle u, x\rangle + \xi \in \mathbb{R}$$

with $x \in X$ and $\xi \in \mathbb{R}$.

Definition 1.22 (Subgradient, subdifferential)**.** *Let f be an extended real-valued function on X (possibly nonconvex). A linear form $u \in U$ is a* subgradient *of f at $x_o \in X$ if*

$$f(x) \ge f(x_o) + \langle u, x - x_o\rangle, \forall x \in X$$

The set of all these subgradients is called the subdifferential *of f at x_o and is denoted $\partial f(x_o)$.*

In contrast with the usual notion of differential which is local and requires regularity, this notion is a global one and f may not even be continuous at x_o for $\partial f(x_o)$ to be nonempty.
Note that this definition depends on the underlying dual pairing $\langle U, X\rangle$. In particular, $\partial f(x_o)$ is a subset of U.
Clearly, if $f(x_o) = +\infty$, $\partial f(x_o)$ is empty unless f is identically $+\infty$ (in which case $\partial f(x) = U$ for all x). On the other hand, if $f(x_o) = -\infty$, $\partial f(x_o) = U$.

Exercice 1.23. *Show that for all $x \in X$, $\partial f(x)$ is a convex set.*

Subgradients are well designed for minimization. Indeed, playing with the definitions, we get

Proposition 1.24. *Let f be any extended real-valued function on X. The point x_* is a global minimizer of f if and only if*

$$0 \in \partial f(x_*).$$

1.3. **Convex conjugation.** For all $u \in \partial f(x_o)$, $x \mapsto f(x_o) + \langle u, x - x_o\rangle$ is an affine function which is tangent to f at x_o. More, the graph of f lies above the graph of the tangent line on the whole space.
Let us rewrite the equation of this tangent line. For all $x \in X$,

$$x \mapsto f(x_o) + \langle u, x - x_o\rangle = \langle u, x\rangle - \beta$$

with $\beta = \langle u, x_o\rangle - f(x_o)$. Since this tangent line is the highest one below f with slope u, the additive constant $-\beta$ is the greatest $\alpha \in \mathbb{R}$ such that $f(x) \ge \langle u, x\rangle + \alpha$ for all $x \in X$.

Hence, $-\beta = \inf_{x\in X}\{f(x) - \langle u, x\rangle\}$ or equivalently $\beta = \sup_{x\in X}\{\langle u, x\rangle - f(x)\}$. Looking at β as a function of the slope u, one introduces the following definition.

Definition 1.25 (Convex conjugate)**.** *Let f be an extended real function on X. We define the* convex conjugate *of f with respect to the algebraic dual pairing $\langle U, X\rangle$, by*

$$f^*(u) = \sup_{x\in X}\{\langle u, x\rangle - f(x)\}, \quad u \in U$$

Examples 1.26*.* As an exercice, prove these results.

(1) $X = U = \mathbb{R}$, $f(x) = |x|^p/p$ with $p \geq 1$, $f^*(u) = |u|^q/q$ where $1/p + 1/q = 1$.

(2) $X = U = \mathbb{R}$, $f(x) = e^x - 1$, $f^*(u) = \begin{cases} u\log u - u + 1 & \text{if } u > 0 \\ 1 & \text{if } u = 0 \\ +\infty & \text{if } u < 0 \end{cases}$

Proposition 1.27. *The function f^* is an extended real-valued convex function on X.*

Proof. As the supremum of convex (affine) functions, f^* is convex (Proposition 1.16), even if f is not convex. □

As $\beta = f^*(u)$ and $\beta = \langle u, x_o\rangle - f(x_o)$, we have just proved that for all $u \in U$ and $x \in X$ such that $u \in \partial f(x)$, we have $f^*(u) = \langle u, x\rangle - f(x)$. The converse of this statement will be proved at Proposition 1.37.
A direct consequence of the definition of f^* is the following inequality.

Proposition 1.28 (Young's inequality)**.** *For all $x \in X$ and $u \in U$,*

$$f^*(u) \geq \langle u, x\rangle - f(x).$$

Remark 1.29*.* To emphasize the u, x-symmetry, one is tempted to rewrite Young's inequality as $\langle u, x\rangle \leq f^*(u) + f(x)$. This is true when it is meaningful. But troubles occur when the right hand side is $+\infty - \infty$ or $-\infty + \infty$.

Exercice 1.30.

- *Show that if $f \leq g$, then $g^* \leq f^*$.*
- *Let $g(x) = f(\lambda x)$, $x \in X$, with $\lambda \neq 0$. Show that $g^*(u) = f^*(u/\lambda)$, $u \in U$.*
- *Let $g = \lambda f$, with $\lambda > 0$. Show that $g^*(u) = \lambda f^*(u/\lambda)$, $u \in U$.*

One can take advantage of the symmetric role played by X and U. The convex conjugate (with respect to the pairing $\langle U, X\rangle$) of any function g on U, is

$$g^*(x) = \sup_{u\in U}\{\langle u, x\rangle - g(u)\}, \quad x \in X$$

We will have to consider subgradients of f^* and more generally subgradients of functions g on U with respect to the dual pairing $\langle U, X\rangle$. They are defined by

$$\partial g(u_o) := \{x \in X : g(u) \geq g(u_o) + \langle x, u - u_o\rangle, \forall u \in U\}$$

Note that $\partial g(u)$ is a subset of X.
One can iterate convex conjugation and consider $f^{**} = (f^*)^*$ which is defined by

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

Proposition 1.31. *Let f be any function on X.*

(a) f^{**} is a convex function and $f^{**} \leq f$.

(b) Denoting $f^{*n} = (f^{*(n-1)})^*$ the n-th iterate for the convex conjugation, we have:

$$\begin{cases} f^{*n} = f^*, & \text{if } n \text{ is odd;} \\ f^{*n} = f^{**}, & \text{if } n \text{ is even, } n \geq 2. \end{cases}$$

Proof. Let us prove (a). As a convex conjugate, f^{**} is a convex function. For all $x \in X$, we have

$$\begin{aligned} f^{**}(x) &= \sup_u \{\langle u, x\rangle - f^*(u)\} \\ &= \sup_u \{\langle u, x\rangle - \sup_y (\langle u, y\rangle - f(y))\} \\ &= \sup_u \inf_y \{\langle u, x - y\rangle + f(y)\} \\ &\leq f(x) \end{aligned}$$

where the last inequality is obtained by choosing $y = x$.
To prove (b), it is enough to show that $f^{***} = f^*$. For all $u \in U$, $f^{***}(u) = \sup_x \{\langle u, x\rangle - f^{**}(x)\} \geq \sup_x \{\langle u, x\rangle - f(x)\} = f^*(u)$, where the inequality follows from $f^{**} \leq f$. Therefore, $f^{***} \geq f^*$. But we also know by (a) that $f^{***} = (f^*)^{**} \leq f^*$, so that $f^{***} = f^*$. □

Reversing conjugation is often useful.

Proposition 1.32. *For any functions f on X and g on U, we have*

$$\begin{cases} f^* &= g \\ f &= f^{**} \end{cases} \Leftrightarrow \begin{cases} f &= g^* \\ g &= g^{**} \end{cases}$$

Proof. Suppose that $f^* = g$ and $f = f^{**}$. Then, $f = f^{**} = g^*$ and $g = f^* = g^{**}$. The converse follows the same line. □

Proposition 1.33 (Geometric characterization of a convex biconjugate)**.** *Let f be any function on X. Its convex biconjugate f^{**} is the supremum of all the affine functions (with respect to the pairing $\langle U, X\rangle$) which are less than f. In other words,*

$$f^{**}(x) = \sup\{h(x) : h \text{ affine such that } h \leq f\}, \quad x \in X$$

or equivalently

$$\operatorname{epi} f^{**} = \bigcap \{\operatorname{epi} h : h \text{ affine such that } \operatorname{epi} h \supset \operatorname{epi} f\}.$$

Proof. For all $x \in X$,

$$\begin{aligned} f^{**}(x) &= \sup_u \{\langle u, x\rangle - f^*(u)\} \\ &= \sup_u \sup_{\alpha : \alpha \geq f^*(u)} \{\langle u, x\rangle - \alpha\}. \end{aligned}$$

But,

$$\begin{aligned} \alpha \geq f^*(u) &\Leftrightarrow \alpha \geq \sup_y \{\langle u, y\rangle - f(y)\} \\ &\Leftrightarrow \alpha \geq \langle u, y\rangle - f(y), \forall y \in X \\ &\Leftrightarrow \langle u, y\rangle - \alpha \leq f(y), \forall y \in X \end{aligned}$$

and the first statement is proved.
The last statement is a rewriting based on Lemma 1.17. □

What happens when a convex function f achieves the value $-\infty$ at some point x_o? Usually, a degenerate behaviour occurs. For instance, suppose that f is defined on $\mathbb{R}$, and $f(0) = -\infty$. If $f(1)$ is finite (say), then one must have $f(x) = -\infty$ for all $0 \le x < 1$ and $f(x) = +\infty$ for all $x > 1$.

Proposition 1.34 (Pathologies). *Let's have a look at some degenerate situations.*

(a) *Let $f \equiv +\infty$. Then, $f^* \equiv -\infty$ and $f = f^{**}$.*
(b) *Suppose that $f(x_o) = -\infty$ for some $x_o \in X$. Then, $f^* \equiv +\infty$ and $f^{**} \equiv -\infty$.*
(c) *Suppose that $f^*(u_o) = -\infty$ for some $u_o \in U$. Then, $f \equiv +\infty$. In particular, $f^* \equiv -\infty$.*

Proof. These direct easy computations are left as an exercice. □

1.4. **Relations between subdifferentials and convex conjugates.** The main result of this section is stated at Theorem 1.38. It states that if $f = f^{**}$, then the set-valued function $\partial f : x \mapsto \partial f(x) \subset U$ is the inverse of the set-valued function $\partial f^* : u \mapsto \partial f^*(u) \subset X$.

Remark 1.35. It is very important to be remember that both notions of subgradient and convex conjugate are associated with a given algebraic dual pairing $\langle U, X \rangle$.

Proposition 1.36. *Let f be any function on X. For all $x \in X$ and $u \in U$,*

$$u \in \partial f(x) \Leftrightarrow f^*(u) = \langle u, x \rangle - f(x).$$

Proof. For any $x \in X$ and $u \in U$, we have:

$$\begin{aligned} u \in \partial f(x) &\Leftrightarrow f(x) + \langle u, y - x \rangle \le f(y), \forall y \in X \\ &\Leftrightarrow \langle u, y \rangle - f(y) \le \langle u, x \rangle - f(x), \forall y \in X \\ &\Leftrightarrow f^*(u) \le \langle u, x \rangle - f(x) \\ &\Leftrightarrow f^*(u) = \langle u, x \rangle - f(x) \end{aligned}$$

where the last equivalence follows from Young's inequality (Proposition 1.28). □

Proposition 1.37. *Let f be any function on X.*

(a) For all $x \in X$ such that $f(x) = f^{**}(x)$ we have $\partial f(x) = \partial f^{**}(x)$.
(b) For all $x \in X$ such that $\partial f(x) \neq \emptyset$, we have $f(x) = f^{**}(x)$ and $\partial f(x) = \partial f^{**}(x)$.
(c) If $f(x_o) > -\infty$, then $\partial f(x_o) \subset \operatorname{dom} f^*$.
(d) If $f(x_o) = -\infty$, then $f^* \equiv +\infty$. Hence $\partial f(x_o) = U$ and $\operatorname{dom} f^* = \emptyset$.

Proof. • Let us prove (a). For all $u \in U$ and $x \in X$, applying Proposition 1.36 with f^{**} instead of f, one obtains that $u \in \partial f^{**}(x) \Leftrightarrow f^{***}(u) = \langle u, x \rangle - f^{**}(x)$. But, $f^{***} = f^*$ (Proposition 1.31), and we have

$$u \in \partial f^{**}(x) \Leftrightarrow f^*(u) = \langle u, x \rangle - f^{**}(x).$$

The desired result follows from this together with Proposition 1.36 and the hypothesis: $f(x) = f^{**}(x)$.
• Let us prove (b). It is assumed that $\partial f(x) \neq \emptyset$. As we already know that $f^{**} \le f$, with Proposition 1.36 again we get $u \in \partial f(x) \Rightarrow f^*(u) \le \langle u, x \rangle - f^{**}(x)$. By Young's inequality, this last inequality is equivalent to the corresponding equality, so that

$$u \in \partial f(x) \Rightarrow f^*(u) = \langle u, x \rangle - f^{**}(x).$$

For $u \in U$ and $x \in X$ with $u \in \partial f(x)$, we have shown that $f^*(u) = \langle u, x \rangle - f^{**}(x) = f^*(u) = \langle u, x \rangle - f(x)$. As $\langle u, x \rangle$ is finite, we get the equality $f(x) = f^{**}(x)$. Statement

(b) now follows from (a).
• Let us prove (c). Note that if $\partial f(x_o)$ is empty, (c) holds trivially. Suppose now that $\partial f(x_o) \neq \emptyset$. It is worth discussing the cases where $f(x_o) = +\infty$ and $-\infty < f(x_o) < +\infty$.

Case where $f(x_o) = +\infty$, $\partial f(x_o) \neq \emptyset$: In this case, $f \equiv +\infty$ and it follows that: $f^* \equiv -\infty$, $f^{**} \equiv +\infty$, $\partial f(x_o) = \partial f^{**}(x_o) = U$ and $\mathrm{dom}\, f^* = U$. One sees that (c) holds in this situation.

Case where $-\infty < f(x_o) < +\infty$, $\partial f(x_o) \neq \emptyset$: We have $u \in \partial f(x_o) \Leftrightarrow f^*(u) = \langle u, x_o\rangle - f(x_o)$. As $\langle u, x_o\rangle$ and $f(x_o)$ are finite, so is $f^*(u)$; and (c) is satisfied.

• Let us get rid of the special case (d). As $f(x_o) = -\infty$, we have $f^* \equiv +\infty$ and $\partial f(x_o) = U$. □

Theorem 1.38. *For any function f on X, the following assertions hold.*

(a) For all $x \in X$, $u \in U$, $u \in \partial f(x) \Rightarrow x \in \partial f^*(u)$.
(b) For all $x \in X$, $u \in U$, if $f(x) = f^{**}(x)$ then $u \in \partial f(x) \Leftrightarrow x \in \partial f^*(u)$.
(c) If $f(0) = f^{**}(0)$, in particular if $\partial f(0) \neq \emptyset$, then $\partial f(0) = \mathrm{argmin}\, f^*$.

Of course, (b) implies that if $f = f^{**}$, then

$$u \in \partial f(x) \Leftrightarrow x \in \partial f^*(u), \quad \forall x \in X, u \in U.$$

Proof. Applying Proposition 1.36 with f and f^*, one sees that for all $x \in X$ and $u \in U$,

$$\begin{aligned} u \in \partial f(x) \quad &\Leftrightarrow \quad f^*(u) = \langle u, x\rangle - f(x) \\ x \in \partial f^*(u) \quad &\Leftrightarrow \quad f^{**}(x) = \langle u, x\rangle - f^*(u) \end{aligned}$$

Let us prove (a). By Proposition 1.37-b, we have $u \in \partial f(x) \Rightarrow \partial f(x) \neq \emptyset \Rightarrow f(x) = f^{**}(x)$. Therefore, $u \in \partial f(x) \Rightarrow \left\{ \begin{array}{l} f^*(u) = \langle u, x\rangle - f(x) \\ f(x) = f^{**}(x) \end{array} \right. \Rightarrow f^{**}(x) = \langle u, x\rangle - f^*(u) \Rightarrow x \in \partial f^*(u)$, which is (a).
Let us show that (b) holds. If $f(x) = f^{**}(x)$, then $u \in \partial f(x) \Leftrightarrow f^*(u) = \langle u, x\rangle - f(x) \Leftrightarrow f^{**}(x) = \langle u, x\rangle - f^*(u) \Leftrightarrow x \in \partial f^*(u)$.
Assertion (c) follows from (b), Proposition 1.37-b and Proposition 1.24. □

As appears with this theorem, it is worth knowing when $f = f^{**}$. This problem is solved at the next section in terms of a topological characterization.

1.5. **Similar tools for maximizing.** Maximization problems occur naturally when working with the saddle-point method (see Section 3). Note that maximizing g is equivalent to minimizing $-g$ since $\sup g = -\inf(-g)$ and $\mathrm{argmax}\, g = \mathrm{argmin}\, (-g)$ where $\mathrm{argmax}\, g$ is the set of all the maximizers of g : $\mathrm{argmax}\, g = \{x \in X : g(x) = \sup g\}$.
One says that a function g is *concave* if $-f$ is convex.
A maximum criterion similar to Proposition 1.24 will be useful. To state it, one needs the notion of *supergradient*.

Definition 1.39 (Supergradient, superdifferential)**.** *Let g be an extended real-valued function on X. A linear form $u \in U$ is a* supergradient *of g at $x_o \in X$ if*

$$g(x) \leq g(x_o) + \langle u, x - x_o\rangle, \forall x \in X$$

The set of all these supergradients is called the superdifferential *of g at x_o and is denoted $\widehat{\partial} g(x_o)$.*

Supergradients are well designed for maximization. Indeed, playing with the definitions, we get

Proposition 1.40. *Let g be any extended real-valued function on X. The point x_* is a global maximizer of g if and only if*

$$0 \in \widehat{\partial} g(x_*).$$

Of course, there is a relation between subgradients and supergradient: for all $x \in X$,

$$\partial(-g)(x) = -\widehat{\partial} g(x). \tag{1.41}$$

Now, let us introduce the notion of concave conjugate. For all $u \in \widehat{\partial} g(x_o)$, $x \mapsto g(x_o) + \langle u, x - x_o \rangle$ is an affine function which is tangent to g at x_o. More, the graph of g lies below the graph of the tangent line on the whole space.
Let us rewrite the equation of this tangent line. For all $x \in X$, $x \mapsto g(x_o) + \langle u, x - x_o \rangle = \langle u, x \rangle - \beta$ with $\beta = \langle u, x_o \rangle - g(x_o)$. Since this tangent line is the lowest one above g with slope u, the additive constant $-\beta$ is the least $\alpha \in \mathbb{R}$ such that $g(x) \le \langle u, x \rangle + \alpha$ for all $x \in X$. Hence, $-\beta = \sup_{x \in X}\{g(x) - \langle u, x \rangle\}$ or equivalently $\beta = \inf_{x \in X}\{\langle u, x \rangle - g(x)\}$. Looking at β as a function of the slope u, one introduces the following definition.

Definition 1.42 (Concave conjugate)**.** *Let g be an extended real function on X. We define the* concave conjugate *of g with respect to the algebraic dual pairing $\langle U, X \rangle$, by*

$$g^{\hat{*}}(u) = \inf_{x \in X}\{\langle u, x \rangle - g(x)\}, \quad u \in U$$

Concave and convex conjugates are related as follows. For all $u \in U$,

$$(-g)^*(u) = -g^{\hat{*}}(-u). \tag{1.43}$$

Using the relations (1.41) and (1.43), one can translate all the preceding convex results into concave ones. For instance, $g^{\hat{*}}$ is a concave function,

$$(-f)^{\hat{*}\hat{*}} = -f^{**}$$

and Theorem 1.38 becomes

Theorem 1.44. *For any function g on X, the following assertions hold.*

(a) *For all $x \in X$, $u \in U$, $u \in \widehat{\partial} g(x) \Rightarrow x \in \widehat{\partial} g^{\hat{*}}(u)$.*
(b) *For all $x \in X$, $u \in U$, if $g(x) = g^{\hat{*}\hat{*}}(x)$ then $u \in \widehat{\partial} g(x) \Leftrightarrow x \in \widehat{\partial} g^{\hat{*}}(u)$.*
(c) *If $g(0) = g^{\hat{*}\hat{*}}(0)$, in particular if $\widehat{\partial} g(0) \neq \emptyset$, then $\widehat{\partial} g(0) = \operatorname{argmax} g^{\hat{*}}$.*

Of course, (b) implies that if $g = g^{\hat{*}\hat{*}}$, then

$$u \in \widehat{\partial} g(x) \Leftrightarrow x \in \widehat{\partial} g^{\hat{*}}(u), \quad \forall x \in X, u \in U.$$

2. Convexity with a topology

Introducing a topological structure on X is useful to derive optimization results.

(i) While optimizing possibly nonconvex functions, topology may be used to prove that optimum values are attained: for instance, we feel at ease with a continuous function on a compact subset.
(ii) Usual optimization of nonconvex regular functions requires a notion of differentiability: local optimizers x_* satisfy $f'(x_*) = 0$. Talking about *local* properties refers to an underlying topology. In addition, to define a derivative one needs to consider *limiting* increment rates; this requires a topological structure.

As regards item(i), in the framework of global minimization lower semicontinuity of functions is the good notion of regularity to be considered. This is developped at Section 2.2 below. The main result of attainment of global minimum values is stated at Theorem 2.12.
As was already seen at Proposition 1.24 and Theorem 1.38-c, while considering *global minimization* problems, instead of local optimization, it is enough to work with a geometric notion of subgradients (without any reference to a topology). This is in contrast with (ii) above. Nevertheless, it will be needed to derive criteria for $f = f^{**}$ to be able to apply Theorem 1.38. Again, a useful tool will be the lower semicontinuity. This criterion in terms of lower semicontinuity is stated at Theorem 2.30.
The main results of this section are Theorem 2.12 about the existence of minimizers and Theorem 2.46 about the main properties the *Fenchel transform*: $f \to f^*$.

2.1. **Compactness results.** Let us recall basic compactness results. Let X be a topological space (possibly not a vector space). An open cover of a subset E of X is a collection of open sets $\{O_i; i \in I\}$ such that $E \subset \bigcup_{i\in I} O_i$. As a definition, a subset K of X is a *compact* set if, from any open cover of K, it is possible to extract a finite subcover.
A topological space X is said to be *Hausdorff* if for all distinct $x, y \in X$, there exist two open neighbourhoods $G_x \ni x$ and $G_y \ni y$ such that $G_x \cap G_y = \emptyset$.
A useful result to derive attainment results (see Theorem 2.12) is the following proposition.

Proposition 2.1. *Let X be a Hausdorff topological space.*
For any nonincreasing sequence of compact sets $(K_n)_{n\geq 1}$ such that $\bigcap_{n\geq 1} K_n$ is empty, there exists $N \geq 1$ such that K_n is empty for all $n \geq N$.
This implies that for any nonincreasing sequence of nonempty compact sets $(K_n)_{n\geq 1}$, we have $\bigcap_{n\geq 1} K_n \neq \emptyset$.

Proof. As X is supposed to be Hausdorff, by Lemma 2.2 below, each K_n is closed: its complement K_n^c is open.
Suppose that $\bigcap_n K_n$ is empty. Then $\{K_n^c; n \geq 1\}$ is an open subcover of X and a fortiori of the compact set K_1. One can extract a finite subcover $(K_n^c)_{1\leq n\leq N}$ such that $K_1 \subset \bigcup_{1\leq n\leq N} K_n^c$. But, $\bigcup_{1\leq n\leq N} K_n^c = K_N^c$ and $K_1^c \subset K_N^c$, so that $X = K_1 \bigcup K_1^c \subset K_N^c$. This implies that $K_N = \emptyset$. □

Lemma 2.2. *If X is Hausdorff, each compact set is closed.*

There exist non-Hausdorff spaces with non-closed compact sets. Indeed, let X be endowed with the coarsest topology: the open sets are X and $\emptyset$. Then, all subset E of X is trivially compact and each subset E which is different from X and $\emptyset$ is non-closed.

Proof. Let K be a compact subset of the Hausdorff space X and x_o be any point in the complement of K. We have to prove that there exists an open neighbourhood of x_o which doesn't intersect K.
As X is Hausdorff, for each $x \in K$, there exist two open sets $G_x \ni x$ and $O_x \ni x_o$ such that $O_x \cap G_x = \emptyset$. As K is compact and $K \subset \bigcup_{x\in K} G_x$, there exists a finite subset $\{x_i; i \in I\}$ of K such that $K \subset \bigcup_{i\in I} G_{x_i}$. Since $\bigcap_{i\in I} O_{x_i}$ and $\bigcup_{i\in I} G_{x_i}$ are disjoint sets, we have $K \cap (\bigcap_{i\in I} O_{x_i}) = \emptyset$. One concludes noting that as a finite intersection of open neighbourhoods, $\bigcap_{i\in I} O_{x_i}$ is still an open neighbourhood of x_o. □

We have recalled the proofs of these basic compactness results to emphasize the role of the Hausdorff assumption.

2.2. **Lower semicontinuity.** A good notion of regularity for a minimization problem is the lower semicontinuity. Indeed, it will be proved in a while at Theorem 2.12 that any lower semicontinuous function on a compact space attains its minimum value.

Definition 2.3 (Lower semicontinuity)**.** *Let X be a topological space (possibly not a vector space). An extended real-valued function f on X is* lower semicontinuous *if its epigraph is a closed subset of $X \times \mathbb{R}$.*

This is a definition of *global* lower semicontinuity. One says that f is lower semicontinuous at $x \in X$, if

$$f(x) \leq \sup_{V \in \mathcal{N}(x)} \inf_{y \in V} f(y) \tag{2.4}$$

where $\mathcal{N}(x)$ is the collection of all open neighbourhoods of x. In particular, if X is a metric space, f is lower semicontinuous at x if and only if for any sequence (x_n) with $\lim_{n\to\infty} x_n = x$, we have $f(x) \leq \liminf_{n\to\infty} f(x_n)$ in $[-\infty, +\infty]$.
Note that the converse inequality $f(x) \geq \sup_{V \in \mathcal{N}(x)} \inf_{y \in V} f(y)$ always holds, so that (2.4) is equivalent to the corresponding equality:

$$f(x) = \sup_{V \in \mathcal{N}(x)} \inf_{y \in V} f(y).$$

Exercice 2.5. *Show that the global lower semicontinuity is equivalent to the local lower semicontinuity at every point of X.*

A function f is said to be *upper semicontinuous* at x if $-f$ is lower semicontinuous at x. In other words, if $f(x) \geq \inf_{V \in \mathcal{N}(x)} \sup_{y \in V} f(y)$.

Exercice 2.6. *Show that f is continuous at x if and only if f is both upper and lower semicontinuous at x.*

Definition 2.7 (Level set)**.** *The* level sets *of a function f are the subsets of X of the form*

$$\{f \leq \alpha\} = \{x \in X : f(x) \leq \alpha\} \subset X$$

with $\alpha \in \mathbb{R}$.

Clearly, $\{f \leq \alpha\}$ is a nondecreasing collection of sets (as α increases).

Proposition 2.8. *Let f be a function on X. The following statements are equivalent.*

(a) f is lower semicontinuous.
(b) All the level sets of f are closed.

Proof. $(a) \Rightarrow (b)$. For all real α, $(X \times \{\alpha\}) \bigcap \operatorname{epi} f$ is closed. We have, $(X \times \{\alpha\}) \bigcap \operatorname{epi} f = \{(x, \beta) : x \in X, \beta \in \mathbb{R}, \beta = \alpha, f(x) \leq \beta\} = \{(x, \alpha) : x \in X, f(x) \leq \alpha\} = \{f \leq \alpha\} \times \{\alpha\}$. Hence, the level set $\{f \leq \alpha\}$ is the inverse image of the closed set $(X \times \{\alpha\}) \bigcap \operatorname{epi} f$ for the continuous mapping $x \in X \mapsto (x, \alpha) \in X \times \mathbb{R}$. Therefore, it is closed.
$(b) \Rightarrow (a)$. Let us prove that the complement of $\operatorname{epi} f$ is open. Let $(x_o, \alpha) \notin \operatorname{epi} f$. This means that $\alpha < f(x_o)$. Take $\beta < \infty$ such that $\alpha < \beta < f(x_o)$. Hence, x_o doesn't belong to the closed level set $\{f \leq \beta\}$ and there exists an open neighbourhood G of x_o which is disjoint from $\{f \leq \beta\}$. Finally, $G \times]-\infty, \beta[$ is an open neighbourhood of (x_o, α) which doesn't intersect $\operatorname{epi} f$. □

Definition 2.9 (Lower semicontinuous envelope)**.** *Let f be any extended real-valued function on X. Its* lower semicontinuous envelope *is the greatest lower semicontinuous function which is less than f. It is denoted* $\operatorname{ls} f$.

The following proposition shows that this definition is meaningful.

Proposition 2.10. *The following statements are true.*

(a) Let $\{f_i; i \in I\}$ be a collection of lower semicontinuous functions. Then $\sup_i f_i$ is also lower semicontinuous.

(b) The lower semicontinuous envelope of a function f is

$$\operatorname{ls} f = \sup\{h : h \text{ lower semicontinuous function on } X \text{ such that } h \le f\}.$$

(c) The epigraph of $\operatorname{ls} f$ is the topological closure of the epigraph of f :

$$\operatorname{epi} \operatorname{ls} f = \operatorname{cl} \operatorname{epi} f.$$

Proof. (a) By Lemma 1.17, we have $\operatorname{epi} \sup_i f_i = \cap_i \operatorname{epi} f_i$ which is closed as the intersection of a collection of closed sets.
(b) and (c) are direct consequences of (a). □

Proposition 2.11. *Let f be any function, its lower semicontinuous envelope is given for all $x \in X$ by*

$$\operatorname{ls} f(x) = \sup_{V \in \mathcal{N}(x)} \inf_{y \in V} f(y)$$

where $\mathcal{N}(x)$ is the collection of all open neighbourhoods of x.

Proof. Let us denote $h(x) := \sup_{V \in \mathcal{N}(x)} \inf_{y \in V} f(y)$. We have $h \le f$ since for all $x \in X$ and $V \in \mathcal{N}(x)$, $\inf_{y \in V} f(y) \le f(x)$, which implies that $h(x) := \sup_{V \in \mathcal{N}(x)} \inf_{y \in V} f(y) \le f(x)$. More, h is lower semicontinuous since it satisfies (2.4). Indeed,

$$\begin{aligned}
\sup_{V \in \mathcal{N}(x)} \inf_{y \in V} h(y) &= \sup_{V \in \mathcal{N}(x)} \inf_{y \in V} \sup_{W \in \mathcal{N}(y)} \inf_{z \in W} f(z) \\
&\ge \sup_{V \in \mathcal{N}(x)} \inf_{y \in V} \inf_{z \in V} f(z) \\
&= \sup_{V \in \mathcal{N}(x)} \inf_{y \in V} f(y) \\
&= h(x)
\end{aligned}$$

where the inequality is obtained by choosing $W = V$.
Now, let h_o be a lower semicontinuous function such that $h \le h_o \le f$. As h_o is lower semicontinuous, we have $h_o(x) \le \sup_{V \in \mathcal{N}(x)} \inf_{y \in V} h_o(y)$. With, $h_o \le f$ this gives us $h_o(x) \le \sup_{V \in \mathcal{N}(x)} \inf_{y \in V} f(y) := h(x)$. This proves that $h = h_o$. Hence, h is the greatest lower semicontinuous minorant of f. □

Theorem 2.12 (Attainment of the minimum on a compact set)**.** *Let X be a Hausdorff topological space, f a lower semicontinuous function and K a compact subset of X. Then, there exists at least one $x_* \in K$ such that $f(x_*) = \inf_K f$.*

Proof. If $K \cap \operatorname{dom} f = \emptyset$, we have $\inf_K f = +\infty$ and this infimum is realized at each point of K.
If $K \cap \operatorname{dom} f \ne \emptyset$, we have $\inf_K f < +\infty$.
Suppose that $-\infty < \beta := \inf_K f$. As f is lower semicontinuous, its level sets are closed (Proposition 2.8). Since K is compact, $K_n := K \cap \{f \le \beta + 1/n\}$ is compact. Since $-\infty < \inf_K f < +\infty$, (K_n) is a nonincreasing sequence of nonempty compact sets. By Proposition 2.1, its limit $\bigcap_n K_n = \{x \in K; f(x) = \beta\}$ is also nonempty.
Suppose now that $\inf_K f = -\infty$. Considering the sets $K_n := K \cap \{f \le -n\}$, $n \ge 1$, we obtain a nonincreasing sequence of non-empty compact sets, since the level sets $\{f \le -n\}$ are closed (Proposition 2.8). As $\bigcap_n K_n = \{x \in K : f(x) = -\infty\}$, the result follows again by means of Proposition 2.1. □

Let us state a useful corollary of this theorem.

Definition 2.13. *A function f on a topological space X is said to be* inf-compact *if all its level sets are compact.*

Definition 2.14. *A sequence $(x_n)_{n\geq 1}$ in X is a* minimizing sequence *of a function f if $\lim_{n\to\infty} f(x_n) = \inf f$.*

Corollary 2.15. *Let X be a Hausdorff topological space and f an inf-compact function on X. Then f is lower semicontinuous and its infimum is attained.*
Any minimizing sequence admits accumulation points and these accumulation points are minimizers of f.

Proof. As the level sets of f are compact and X is Hausdorff, they are closed (Lemma 2.2). By Proposition 2.8, this proves that f is lower semicontinuous.
If $f \equiv +\infty$, its infimum is attained at every point. As f is inf-compact, this implies that X is compact and any sequence admits accumulation points.
Otherwise, $\inf f < +\infty$ and the non-empty level set $K = \{f \leq \beta\}$ is compact, with $\beta > \inf_X f$. Clearly, $\inf_X f = \inf_K f$. One concludes with Theorem 2.12.
Let (x_n) be a minimizing sequence. For n large enough, x_n is in the compact level set $\{f \leq \inf f + 1\}$. Hence there exist accumulation points. Let x_* be one of them: there exists a subsequence (x'_k) with $\lim_k x'_k = x_*$. As f is lower semicontinuous, we have $\inf f \leq f(x_*) \leq \liminf_k f(x'_k) = \lim_n f(x_n) = \inf f$. This proves that $f(x_*) = \inf f$: x_* is a minimizer. □

For a useful criterion of inf-compactness, see Proposition 2.43 below.

2.3. **Hahn-Banach Theorem.** We are going to give at Theorem 2.30 a characterization in terms of lower semicontinuity of the functions f which satisfy the identity $f = f^{**}$. The proof of this result relies upon a geometric form of Hahn-Banach Theorem which we are going to state.

Definition 2.16. *Let X be a vector space endowed with some topology. It is a* topological vector space *if both the addition $(x, y) \in X \times X \mapsto x+y \in X$ and the scalar multiplication $(x, \lambda) \in X \times \mathbb{R} \mapsto \lambda x \in X$ are continuous functions.*
A topological vector space X is locally convex *if $0 \in X$ possesses a fundamental system of convex neighbourhoods.*

Since the addition is continuous, in a locally convex topological vector space each vector possesses a fundamental system of convex neighbourhoods.
Let X be a topological vector space, u a continuous linear form on X and α a real number. The set

$$H = \{x \in X : \langle u, x\rangle = \alpha\}$$

is called a *closed affine hyperplane.* As u is continuous, it is clearly a closed set.
A closed affine hyperplane is said to *separate* two sets E and F if each of the *closed* half-spaces bounded by H : $\{x \in X : \langle u, x\rangle \leq \alpha\}$ and $\{x \in X : \langle u, x\rangle \geq \alpha\}$, contains one of them. If $\langle u, x\rangle = \alpha$ is the equation of H, this means

$$\langle u, x\rangle \leq \alpha, \forall x \in E \quad \text{and} \quad \langle u, x\rangle \geq \alpha, \forall x \in F.$$

Similarly, a closed affine hyperplane is said to *strictly separate* two sets E and F if each of the *open* half-spaces bounded by H : $\{x \in X : \langle u, x\rangle < \alpha\}$ and $\{x \in X : \langle u, x\rangle > \alpha\}$, contains one of them. If $\langle u, x\rangle = \alpha$ is the equation of H, this means

$$\langle u, x\rangle < \alpha, \forall x \in E \quad \text{and} \quad \langle u, x\rangle > \alpha, \forall x \in F.$$

The proofs of all the following results can be found for instance in the book of F. Trèves ([8], Chapter 18).

Theorem 2.17 (Hahn-Banach Theorem)**.** *Let X be a* topological vector space*, E an open non-empty convex set and F a non-empty affine subspace which does not intersect E. Then there exists a closed affine hyperplane H which contains F and does not intersect E.*

Let us give some important corollaries of this foundamental result.

Corollary 2.18. *Let X be a topological vector space, E an open non-empty convex set and F a non-empty convex set which does not intersect E. Then there exists a closed affine hyperplane H which separates E and F.*

Corollary 2.19. *Let X be a* locally convex *topological vector space, E a closed non-empty convex set and K a compact non-empty convex set which does not intersect E. Then there exists a closed affine hyperplane H which* stricly *separates E and K.*

As a consequence of this last result, we have

Corollary 2.20. *In a locally convex topological vector space, every closed convex set is the intersection of the closed half-spaces which contain it.*

Although Corollary 2.18 will not be used later, we have stated it to emphasize the role of the locally convex assumption in Corollaries 2.19 and 2.20.
Note that it is not assumed that X is Hausdorff to get these separation results.

2.4. **Closed envelopes.** Mixing the notions of convex hull and closure one obtains the

Definition 2.21 (Closed convex hull)**.** *Let E be any subset of X. Its* closed convex hull *is the smallest closed convex set which contains E. It is denoted* $\operatorname{clcv} E$.

Mixing the notions of convex and lower semicontinuous envelopes one obtains the

Definition 2.22 (Lower semicontinuous convex envelope)**.** *Let f be any extended real-valued function on X. Its* lower semicontinuous convex envelope *is the greatest convex lower semicontinuous function which is less than f. It is denoted* $\operatorname{lscv} f$.

The following proposition shows that this definition is meaningful.

Proposition 2.23. *The lower semicontinuous convex envelope of a function f is the pointwise supremum of all the convex and lower semicontinuous functions less than f :*

$$\operatorname{lscv} f = \sup\{h : h \textit{ convex lower semicontinuous function on } X \textit{ such that } h \le f\}.$$

Proof. Because of Propositions 1.16 and 2.10, the pointwise supremum of any collection of convex lower semicontinuous functions is convex and lower semicontinuous. It follows that $\sup\{h : h$ convex lower semicontinuous function on X such that $h \le f\}$ is convex and lower semicontinuous. □

Proposition 2.24. *Assuming that X is a locally convex topological vector space, the following statements are true.*

(a) The closure of a convex set is convex.
(b) The closed convex hull of E is the closure of its convex hull:

$$\operatorname{clcv} E = \operatorname{cl}(\operatorname{cv} E).$$

(c) The lower semicontinuous envelope of a convex function is convex.

Proof. Proof of (a). Let E be a convex subset of X, x_o, x_1 stand in $\mathrm{cl}\, E$ and $0 \le t \le 1$. We want to show that $x_t := (1-t)x_o + tx_1$ stands in $\mathrm{cl}\, E$.
Let W be any open neighbourhood of x_t. There exists a convex neighbourhood of the origin V such that $x_t + V \subset W$. As x_o, x_1 stand in $\mathrm{cl}\, E$, there exist $y_o \in x_o + V$ and $y_1 \in x_1 + V$ with y_o, y_1 in the convex set E. It follows that $y_t := (1-t)y_o + ty_1$ is also in E and $y_t \in (1-t)[x_o + V] + t[x_1 + V] \subset x_t + V \subset W$, since $(1-t)V + tV \subset V$ as V is convex. We have shown that for any neighbourhood W of x_t, there exists a point y_t in $W \cap E$. This means that x_t belongs to $\mathrm{cl}\, E$.
Let us prove (b). As $\mathrm{cv}\, E$ is the smallest convex set which contains E and $\mathrm{clcv}\, E$ is a convex set which contains E, we have: $\mathrm{cv}\, E \subset \mathrm{clcv}\, E$. Taking the closures, we get $E \subset \mathrm{cl}\,(\mathrm{cv}\, E) \subset \mathrm{clcv}\, E$. But, we have just proved at (a) that $\mathrm{cl}\,(\mathrm{cv}\, E)$ is closed and convex. As $\mathrm{clcv}\, E$ is the smallest closed convex set which contains E, we get $\mathrm{cl}\,(\mathrm{cv}\, E) = \mathrm{clcv}\, E$.
Let us prove (c). Let f be a convex function. Then, $\mathrm{epi}\, f$ is a convex set and $\mathrm{epi}\,\mathrm{ls}\, f = \mathrm{cl}\,\mathrm{epi}\, f$ (Proposition 2.10 (c)). Thanks to statement (a), $\mathrm{epi}\,\mathrm{ls}\, f$ is a convex set. □

Proposition 2.25. *Assuming that X is a locally convex topological vector space, the following statements are true.*

(a) The lower semicontinuous convex envelope of a function f is the lower semicontinuous envelope of its convex envelope:

$$\mathrm{lscv}\, f = \mathrm{ls}\,(\mathrm{cv}\, f).$$

(b) The epigraph of $\mathrm{lscv}\, f$ *is the closed convex hull of the epigraph of f :*

$$\mathrm{epi}\,\mathrm{lscv}\, f = \mathrm{clcv}\,\mathrm{epi}\, f.$$

Proof. Let us prove (a). As $\mathrm{cv}\, f$ is the greatest convex function below f and $\mathrm{lscv}\, f$ is a convex function below f, we have $\mathrm{lscv}\, f \le \mathrm{cv}\, f$. As $\mathrm{lscv}\, f$ is lower semicontinuous, taking the lower semicontinuous envelopes in both sides of this inequality implies that $\mathrm{lscv}\, f = \mathrm{ls}\,(\mathrm{lscv}\, f) \le \mathrm{ls}\,(\mathrm{cv}\, f) \le f$. Since $\mathrm{ls}\,(\mathrm{cv}\, f)$ is clearly lower semicontinuous. To obtain the desired identity: $\mathrm{lscv}\, f = \mathrm{ls}\,(\mathrm{cv}\, f)$, it remains to notice that $\mathrm{ls}\,(\mathrm{cv}\, f)$ is also a convex function. This holds because of Proposition 2.24 (c), assuming that X is a locally convex topological vector space.
Statement (b) is the epigraph version of statement (a), noting that the closure of an epigraph is still an epigraph (Proposition 2.10 (c)). □

2.5. **Convex conjugation and topology.** Now, let X be a vector space endowed with some topology. At this stage it is not necessary that X is a topological vector space. Let U be its topological dual space: $U = X'$. We consider the associated dual pairing $\langle u, x\rangle$: the action of the continuous linear form u on the vector x. This means that for all function f on X

$$\begin{aligned} f^*(u) &= \sup_{x\in X}\{\langle u,x\rangle - f(x)\}, \quad u \in X' \\ f^{**}(x) &= \sup_{u\in X'}\{\langle u,x\rangle - f^*(u)\}, \quad x \in X \end{aligned}$$

where X' is the topological dual space of X.

Proposition 2.26. *With these assumptions, the convex conjugate*

$$x \in X \mapsto g^*(x) = \sup_{u\in X'}\{\langle u,x\rangle - g(u)\} \in [-\infty, +\infty]$$

of any function g on U is convex lower semicontinuous and satisfies (with $f = g^*$)

$$(\exists x_o \in X, f(x_o) = -\infty) \Rightarrow f \equiv -\infty. \tag{2.27}$$

In particular, the biconjugate f^{**} *of any function* f *on* X *is convex lower semicontinuous and satisfies (2.27).*

Proof. As a convex conjugate, g^* is convex (Proposition 1.27). Since $U = X'$, g^* is the supremum of continuous (affine) functions. It follows from Proposition 2.10 that it is lower semicontinuous. The property (2.27) is the statement (c) of Proposition 1.34. $\square$

Hahn-Banach Theorem will allow us to prove the converse result at Theorem 2.30.

Definition 2.28 (Closed functions and $\Gamma(X)$)**.** *A function* f *on* X *is said to be* closed function *if*

$$\begin{cases} f \text{ is lower semicontinuous and } f(x) > -\infty, \forall x \in X \\ \text{or} \\ f \equiv -\infty \end{cases}$$

We denote $\Gamma(X)$ *the set of all* closed convex functions *on* X.

Of course, a closed convex function is also lower semicontinuous convex, since a function f is closed if and ony if it is lower semicontinuous and satisfies (2.27).
To make precise the difference between closed convex functions and lower semicontinuous convex functions, we state the following result.

Proposition 2.29. *Let* f *be a lower semicontinuous convex function. If there exists* $x_o \in X$ *such that* $-\infty < f(x_o) < +\infty$, *then* $f(x) > -\infty$ *for all* $x \in X$.
In particular, a lower semicontinuous convex function which admits one finite value is closed convex.

Proof. Suppose that there exists $x_1 \in X$ such that $f(x_1) = -\infty$. As f is convex, for all $0 < t \leq 1$, we have

$$f((1-t)x_0 + tx_1) \leq (1-t)f(x_0) + tf(x_1) = -\infty$$

But f is lower semicontinuous at x_o, and letting t tend to zero one gets: $f(x_o) \leq \liminf_{t\to 0} f((1-t)x_0 + tx_1) = -\infty$. Which contradicts $f(x_o) > -\infty$. $\square$

Theorem 2.30. *Let* X *be a locally convex topological vector space. For all function* f *on* X, *we have*

$$f = f^{**} \Leftrightarrow f \in \Gamma(X).$$

In particular, for all $f \in \Gamma(X)$, $g = f^* \Rightarrow f = g^*$.

Note that it is not assumed that X is Hausdorff.
Let us state the corresponding results with concave functions.

Definition 2.31 (Closed concave function)**.** *A function* f *is said to be* closed concave *if* $-f$ *is closed convex. We denote* $-\Gamma(X)$ *the set of all closed concave functions defined by* $f \in -\Gamma(X) \Leftrightarrow -f \in \Gamma(X)$.

A function f is closed concave if and only if it is concave, upper semicontinuous and satisfies: $(\exists x_o \in X, f(x_o) = +\infty) \Rightarrow f \equiv +\infty$.

Corollary 2.32. *Let* X *be a locally convex topological vector space. For all function* f *on* X, *we have*

$$f = f^{\hat{*}\hat{*}} \Leftrightarrow f \in -\Gamma(X).$$

In particular, for all $f \in -\Gamma(X)$, $g = f^{\hat{*}} \Rightarrow f = g^{\hat{*}}$.

Proof. Translate the results of Theorem 2.30, by means of (1.43): $(-f)^* = -f^{\hat{*}}(-\cdot)$. $\square$

Let us proceed with the proof of the theorem.

Proof of Theorem 2.30. The "$\Rightarrow$" part is already proved at Proposition 2.26. Let us prove the "$\Leftarrow$" part. Suppose that $f \in \Gamma(X)$.
If $f \equiv +\infty$, then $f = f^{**}$ is satisfied
If there exists $x_o \in X$ such that $f(x_o) = -\infty$, then it is assumed that $f \equiv -\infty$. Consequently, $f^{**} \equiv -\infty = f$.
Suppose now that $f(x) > -\infty$ for all $x \in X$ and that there exists at least one x such that $f(x) < +\infty$. In view of Proposition 1.33 applied with $U = X'$, we have to prove that f is the pointwise supremum of continuous affine functions. All we have to show is that for all $x_o \in X$ and all real α_o such that $\alpha_o < f(x_o)$, there exists a continous affine function h such that $\alpha_o < h(x_o) < f(x_o)$ and $h \leq f$.
As $\alpha_o < f(x_o)$, the point (x_o, α_o) is not in $\operatorname{epi} f$. But $\{(x_o, \alpha_o)\}$ is a convex compact set and $\operatorname{epi} f$ is a convex closed set since f is convex and lower semicontinuous. By the Corollary 2.19 of Hahn-Banach Theorem (note that X is assumed to be a locally convex topological vector space for this purpose), there exists $u_o \in X'$ and two real numbers β, γ such that

$$\langle u_o, x_o\rangle + \gamma\alpha_o < \beta < \langle u_o, x\rangle + \gamma\alpha, \quad \forall (x, \alpha) \in \operatorname{epi} f.$$

This implies that $\gamma \geq 0$. Indeed, suppose that $\gamma < 0$, letting α tend to $+\infty$ in $\beta < \langle u_o, x_1\rangle + \gamma\alpha$ for some x_1 in $\operatorname{dom} f$, we obtain $\beta < -\infty$ which is absurd.
More precisely,

$$x_o \in \operatorname{dom} f \Rightarrow \gamma > 0 \quad \text{or equivalently} \quad \gamma = 0 \Rightarrow x_o \notin \operatorname{dom} f. \tag{2.33}$$

Indeed, if $\gamma = 0$ one gets $\langle u_o, x_o\rangle < \beta < \langle u_o, x\rangle$ for all $x \in \operatorname{dom} f$. This clearly implies that $x_o \notin \operatorname{dom} f$, otherwise we would have $\langle u_o, x_o\rangle < \langle u_o, x_o\rangle$ which is absurd.
If $\gamma > 0$, one gets

$$\beta/\gamma - \langle u_o, x\rangle/\gamma < f(x), \quad \forall x \in X$$

and

$$\alpha_o < \beta/\gamma - \langle u_o, x_o\rangle/\gamma < f(x_o)$$

which is the desired result with $h(x) = \beta/\gamma - \langle u_o, x\rangle/\gamma$, $x \in X$.
If $\gamma = 0$, then denoting $h_o(x) = -\langle u_o, x\rangle + \beta$, $x \in X$, we have,

$$h_o(x) < 0 < h_o(x_o), \quad \forall x \in \operatorname{dom} f.$$

Choosing x_1 in $\operatorname{dom} f$, the associated γ_1 must be positive thanks to (2.33) and we have just proved that there exists a continuous affine function h_1 such that $h_1 < f$. Now, for all $c \geq 0$, we have $h_1(x) + ch_o(x) < f(x)$ for all $x \in X$, and choosing c large enough, we get $h_1(x_o) + ch_o(x_o) > \alpha_o$ and the desired results holds with $h = h_1 + ch_o$. □

To state the next result which is a corollary of this theorem, one needs a new definition.

Definition 2.34 (Closed convex envelope)**.** *Let f be any function on X. Its* closed convex envelope *is denoted* $\operatorname{clcv} f$ *and defined as the greatest function in $\Gamma(X)$ which is less than f.*

This definition is meaningful since $\Gamma(X)$ is stable under an arbitrary number of supremum operations.

Proposition 2.35. *Let f be any function on X. Then,*

$$\operatorname{clcv} f = \begin{cases} \operatorname{lscv} f, & \text{if for all } x \in X,\ \operatorname{lscv} f(x) > -\infty \\ -\infty, & \text{if there exists } x_o \in X \text{ such that } \operatorname{lscv} f(x_o) = -\infty \end{cases}$$

Proof. This is a direct consequence of Propositions 2.23 and 2.29. □

Corollary 2.36 (of Theorem 2.30). *Let X be a locally convex topological vector space. For any function f on X, we have*

$$f^{**} = \operatorname{clcv} f.$$

In particular, if f is bounded below by a continuous affine function, then

$$f^{**} = \operatorname{lscv} f$$

and

$$f^* = (\operatorname{cv} f)^* = (\operatorname{ls} f)^* = (\operatorname{lscv} f)^*.$$

Proof. As $f^{**} \leq f$ (see Proposition 1.31) and f^{**} is convex and lower semicontinuous, we also have:

$$f^{**} \leq \operatorname{lscv} f. \tag{2.37}$$

If there exists $x_o \in X$ such that $\operatorname{lscv} f(x_o) = -\infty$, then $f^{**}(x_o) \leq \operatorname{lscv} f(x_o) = -\infty$. This means that $f^{**}(x_o) = -\infty$. It follows by Proposition 1.34 that $f^{**} \equiv -\infty$.
Otherwise, by Theorem 2.30, we have $\operatorname{lscv} f = (\operatorname{lscv} f)^{**}$. With (2.37), this gives

$$f^{**} \leq \operatorname{lscv} f = (\operatorname{lscv} f)^{**} \leq f^{**}$$

where the last inequality follows from $\operatorname{lscv} f \leq f$. This proves the identity $f^{**} = \operatorname{lscv} f$.
Let f be bounded below by a continuous affine function $h : h \leq f$. As h is convex and lower semicontinuous, we have $h = \operatorname{lscv} h \leq \operatorname{lscv} f$ which implies that $\operatorname{lscv} f(x) > -\infty$ for all $x \in X$. We have just proved that in this situation: $f^{**} = \operatorname{lscv} f$.
As $\operatorname{lscv} f \leq \operatorname{cv} f \leq f$, we get: $f^* \leq (\operatorname{cv} f)^* \leq (\operatorname{lscv} f)^* = f^{***} = f^*$, where the last equality is obtained at Proposition 1.31. This proves that: $f^* = (\operatorname{cv} f)^* = (\operatorname{lscv} f)^*$. A similar proof works with $\operatorname{ls} f$ instead of $\operatorname{cv} f$. □

2.6. **Weak topologies.** Let $\langle X, U\rangle$ be an algebraic dual pairing.

Definition 2.38 (Weak topology). *The topology $\sigma(X, U)$ is the weakest topology on X such that for all $u \in U$, the linear form $x \in X \mapsto \langle u, x\rangle \in \mathbb{R}$ is continuous. It is called the* topology of X weakened by U.
It makes X a locally convex topological vector space: a fundamental system of neighbourhoods of $x_o \in X$ is $(\{x \in X : |\langle u, x - x_o\rangle| < \delta\}; u \in U, \delta > 0)$.

One can prove that `(voir Brézis, page 41, pour une preuve à écrire plus tard)`

$$(X, \sigma(X, U))' \simeq U.$$

This identity states that the topological dual space of X weakened by U is isomorphic to U. Rewriting Theorem 2.30 with this topology, we obtain the

Proposition 2.39. *Let $\langle X, U\rangle$ be an algebraic dual pairing. Then $f = f^{**}$ if and only if f is convex and $\sigma(X, U)$-closed.*

Weak topologies are interesting regards to compactness. Indeed, a weak topology is coarse in the sense that there are not many open sets, hence not many open covers and it is easier for a set to be compact than with a finer topology. As a consequence of Tychonov's theorem which states that the product of an arbitrary number of compact spaces is still compact with respect to the corresponding product topology, we have the following result.

Theorem 2.40 (Banach-Alaoglu). *Let X be a seminormed space. The unit ball of its dual space X' is compact for the $*$-weak topology $\sigma(X', X)$.*

Proof. See ([2], III.17, Corollaire 3). □

On the other hand, it is difficult for a function to be semicontinuous with respect to a coarse topology.

Remark 2.41. Weak topologies do not admit many semicontinuous functions but they admit many compact sets. On the other hand, it is easier for a function to be semicontinuous with respect to a stronger topology. But strenghtening the topology, one looses compact spaces!

Fortunately, *convex* functions are not too irregular.

Theorem 2.42. *Let X be a vector space with a topology τ which turns X into a locally convex topological vector space. Let X' be the topological dual space of (X,τ). Consider now the so-called weak topology on X : $\sigma(X,X')$. Then any* convex *function f on X is $\sigma(X,X')$-lower semicontinuous if and only if it is τ-lower semicontinuous.*

Proof. Since the topology $\sigma(X,X')$ is weaker than τ, if f is $\sigma(X,X')$-lower semicontinuous then it is τ-lower semicontinuous.
Suppose now that f is τ-lower semicontinuous. Its epigraph $\mathrm{epi}\, f$ is a convex τ-closed set. By Corollary 2.20 of Hahn-Banach theorem, $\mathrm{epi}\, f$ is the intersection of the τ-closed half-spaces which contain it. But, a τ-closed half-space is also $\sigma(X,X')$-closed, so that $\mathrm{epi}\, f$ is $\sigma(X,X')$-closed. This means that f is $\sigma(X,X')$-lower semicontinuous.
To see that a τ-closed half-space H is also $\sigma(X,X')$-closed, note that the generic equation of H is $x\in H \Leftrightarrow \langle u,x\rangle \le \alpha$ with $u\in X'$ and α a real number. □

The next result is a useful criterion for inf-compactness with respect to a weak topology.

Proposition 2.43. *Let X be a topological vector space and X' be its topological dual space. The pairing to be considered is $\langle X,X'\rangle$. If there exists a neighbourhood N of the origin in X such that $\sup_{x\in N} f(x) < +\infty$, then f^* is inf-compact on X' for the topology $\sigma(X',X)$.*

Proof. The set $N' = N\cap(-N)$ is still a neighbourhood of zero, with $\sup_{x\in N'} f(x) \le \sup_{x\in N} f(x) < +\infty$. Therefore, one can assume without restriction that $N=-N$.
Denoting $r := \sup_{x\in N} f(x)$ and $\zeta_N(x) = \begin{cases} 0 & \text{if } x\in N \\ +\infty & \text{if } x\notin N \end{cases}$, our assumption gives us $f\le r+\zeta_N$. Hence, $f^*\ge \zeta_N^* - r$ and for all $u\in X'$ and $\alpha\in\mathbb{R}$, $f^*(u)\le\alpha \Rightarrow \sup_{x\in N}\langle u,x\rangle = \zeta_N^*(u)\le \alpha+r$. As $N=-N$, one also obtains $\sup_{x\in N}|\langle u,x\rangle| \le \alpha + r$, which implies that the level set $\{f^*\le\alpha\}$ is an *equicontinuous* set of linear forms on X. It is therefore a relatively $\sigma(X',X)$-compact set (see [8], Proposition 32.8).
As f^* is $\sigma(X',X)$-lower semicontinuous on X (Proposition 2.26), $\{f^*\le\alpha\}$ is $\sigma(X',X)$-closed. One concludes that $\{f^*\le\alpha\}$ is $\sigma(X',X)$-compact. □

2.7. **Continuity of convex functions.**

We follow ([3], Chapter 1, Section 2.3). The main technical result for the study of continuous convex functions, is the following result.

Lemma 2.44. *Let f be a convex function on a locally convex topological vector space X. Let $x_o\in X$ satisfy $f(x_o) > -\infty$. If there exists an open neighbourhood V of x_o such that $\sup_{x\in V} f(x) < +\infty$, then f is continuous at x_o.*

Proof. Without restriction, V can be assumed to be convex. The set $W = [(V-x_o)\cap -(V-x_o)]$ is a symmetric open neighbourhood of 0. Let us take $0<t<1$. For all $x\in x_o+tW$, we have $x_o+(x-x_o)/t\in V$ and $x_o-(x-x_o)/t\in V$. As f is convex, it follows that

$$f(x)-f(x_o) \le t[f(x_o+(x-x_o)/t) - f(x_o)] \le t(\sup_V f - f(x_o)),$$

and

$$f(x_o) - f(x) \leq t[f(x_o - (x - x_o)/t) - f(x_o)] \leq t(\sup_V f - f(x_o)).$$

This gives $|f(x) - f(x_o)| \leq ta$ for all $x \in x_o + tW$ with $0 \leq a := \sup_V f - f(x_o) < \infty$. This completes the proof of the lemma. □

2.8. **Topologies on X and U.** It is sometimes useful to take advantage of the symmetric roles played by X and U in the algebraic setting of Section 1. It appears that it is worth giving U a topology, in order for instance to talk about lower semicontinuous functions on U to obtain a criterion for the identity $g = g^{**}$. If one wants to consider *simultaneously* characterizations of functions f on X such that $f = f^{**}$ (see Theorem 2.30) and functions g on U such that $g = g^{**}$, one has to impose that: $U = X'$ *and* $X = U'$.

Definition 2.45 (Topological dual pairing). *Let X and U be two vector spaces. They are* topologically paired *if*

- *Both X and U are locally convex topological vector spaces.*
- *The topological dual space X' of X is (isomorphic to) U.*
- *The topological dual space U' of U is (isomorphic to) X.*

The pairing is still denoted $\langle u, x\rangle$, $x \in X$, $u \in U$.

Note that X separates U and U separates X. Saying that X separates U means that for all distinct $u_1, u_2 \in U$, there exists $x \in X$, such that $\langle u_1, x\rangle \neq \langle u_2, x\rangle$.
This separation propoerty implies that X and U are Hausdorff spaces.
A typical example of topological dual pairing is as follows. We take X a Hausdorff locally convex topological vector space and $U = X'$ is endowed with the weak topology $\sigma(U, X)$ so that $U' = X$.
If X and U are topologically paired, they are Hausdorff locally convex topological vector spaces, so that all the preceding results hold.

Theorem 2.46. *Let $\langle X, U\rangle$ be a topological dual pairing. Let us recall that $\Gamma(X)$ is the set of all closed convex functions on X and $\Gamma(U)$ is the set of all closed convex functions on U.*
The Fenchel transform

$$f \in \Gamma(X) \mapsto f^* \in \Gamma(U)$$

induces a one-to-one correspondence between $\Gamma(X)$ and $\Gamma(U)$ such that

$$g = f^* \Leftrightarrow f = g^*, \quad \forall f \in \Gamma(X), g \in \Gamma(U).$$

More, for any $f \in \Gamma(X)$ we have

$$u \in \partial f(x) \Leftrightarrow x \in \partial f^*(u), \quad \forall x \in X, u \in U$$

and for any $g \in \Gamma(U)$ we have

$$x \in \partial g(u) \Leftrightarrow u \in \partial g^*(x), \quad \forall x \in X, u \in U.$$

Proof. We apply Theorem 2.30 with X and U to obtain $f \in \Gamma(X) \Leftrightarrow f = f^{**}$ and $g \in \Gamma(U) \Leftrightarrow g = g^{**}$.
The first result follows immediately from Proposition 1.32 and the last result from Theorem 1.38. □

One gets a similar result for concave functions.

Theorem 2.47. *Let $\langle X, U\rangle$ be a topological dual pairing. Let us recall that $-\Gamma(X)$ is the set of all closed concave functions on X and $-\Gamma(U)$ is the set of all closed concave functions on U.*
The concave Fenchel transform

$$f \in -\Gamma(X) \mapsto f^{\hat{*}} \in -\Gamma(U)$$

induces a one-to-one correspondence between $-\Gamma(X)$ and $-\Gamma(U)$ such that

$$g = f^{\hat{*}} \Leftrightarrow f = g^{\hat{*}}, \quad \forall f \in -\Gamma(X), g \in -\Gamma(U).$$

More, for any $f \in -\Gamma(X)$ we have

$$u \in \widehat{\partial} f(x) \Leftrightarrow x \in \widehat{\partial} f^{\hat{*}}(u), \quad \forall x \in X, u \in U$$

and for any $g \in -\Gamma(U)$ we have

$$x \in \widehat{\partial} g(u) \Leftrightarrow u \in \widehat{\partial} g^{\hat{*}}(x), \quad \forall x \in X, u \in U.$$

Proof. Translate the results of Theorem 2.46 by means of (1.41) and (1.43). □

3. The saddle-point method

3.1. **Primal and dual problems.** Let A be a set (which may not be a vector space) and $f : A \to [-\infty, +\infty]$ an extended real-valued function. We consider the following minimization problem

$$\text{minimize } f(a), \quad a \in A \tag{$\mathcal{P}$}$$

All the functions to be considered are supposed to be $[-\infty, +\infty]$-valued.
Let B be another set and K a function on $A \times B$ such that

$$f(a) = \sup_{b \in B} K(a, b), \quad a \in A. \tag{3.1}$$

Let us introduce the following maximization problem

$$\text{maximize } g(b), \quad b \in B \tag{$\mathcal{D}$}$$

where g is the function on B which is defined by

$$g(b) = \inf_{a \in A} K(a, b), \quad b \in B. \tag{3.2}$$

Vocabulary. The function f is the *objective function* of the *primal* minimization *problem* $(\mathcal{P})$. The function K is called the *Lagrangian.* The maximization problem $(\mathcal{D})$ is the *dual problem* and g is its objective function.
We denote $\inf(\mathcal{P}) = \inf_{a \in A} f(a)$ and $\sup(\mathcal{D}) = \sup_{b \in B} g(b)$ the values of the primal and dual problems.

Lemma 3.3. *We have*
(a) $g(b) \le K(a, b) \le f(a)$ for all $a \in A, b \in B$.
(b) $\sup(\mathcal{D}) \le \inf(\mathcal{P})$.

Proof. Statement (a) is immediate and (b) follows from it/ □

Definition 3.4 (Saddle-point)**.** *One says that $(\bar{a}, \bar{b}) \in A \times B$ is a* saddle-point *of the function K if*

$$K(\bar{a}, b) \le K(\bar{a}, \bar{b}) \le K(a, \bar{b}), \quad \forall a \in A, b \in B.$$

As an example, consider $K(a, b) = a^2 - b^2$ on $\mathbb{R}^2$ which admits $(0, 0)$ as a saddle-point.

Theorem 3.5 (Saddle-point theorem). *We assume that f and g are related to K by means of (3.1) and (3.2). The following statements are equivalent.*

(1) The point $(\bar{a}, \bar{b})$ is a saddle-point of the Lagrangian K
(2) $f(\bar{a}) \leq g(\bar{b})$
(3) The following three statements hold
  (a) we have the dual equality*: $\sup(\mathcal{D}) = \inf(\mathcal{P})$,*
  (b) $\bar{a}$ is a solution to the primal problem $(\mathcal{P})$ and
  (c) $\bar{b}$ is a solution to the dual problem $(\mathcal{D})$.

In this situation, one also gets

$$\sup(\mathcal{D}) = \inf(\mathcal{P}) = K(\bar{a}, \bar{b}) = f(\bar{a}) = g(\bar{b}). \tag{3.6}$$

Moreover, suppose that A and B are vector spaces and that we are given a couple of algebraic dual pairings $\langle A, P\rangle$ and $\langle B, Q\rangle$. Then, the point $(\bar{a}, \bar{b})$ is a saddle-point of K if and only if it satisfies

$$\begin{cases} \partial_a K(\bar{a}, \bar{b}) & \ni & 0 \\ \widehat{\partial}_b K(\bar{a}, \bar{b}) & \ni & 0 \end{cases} \tag{3.7}$$

where the subscript a or b indicates the unfixed variable.

Proof. We begin with a circular proof of (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3).

• (1) $\Rightarrow$ (2). Let $(\bar{a}, \bar{b})$ be a saddle-point of the Lagrangian K. Optimizing both sides of the saddle-point property: $K(\bar{a}, b) \leq K(\bar{a}, \bar{b}) \leq K(a, \bar{b})$, $\forall a \in A, b \in B$, one gets (2).

• (2) $\Rightarrow$ (3). Because of (2), we have $\inf(\mathcal{P}) \leq f(\bar{a}) \leq g(\bar{b}) \leq \sup(\mathcal{D})$. Thanks to Lemma 3.3-b, this yields the following sequence of equalities

$$\inf(\mathcal{P}) = f(\bar{a}) = g(\bar{b}) = \sup(\mathcal{D}) \tag{3.8}$$

which is clearly equivalent to (3).

• (3) $\Rightarrow$ (1). It is assumed that (3.8) holds. Together with Lemma 3.3-a, this gives $f(\bar{a}) = K(\bar{a}, \bar{b}) = g(\bar{b})$. But, by Lemma 3.3-a again , we have $K(\bar{a}, b) \leq f(\bar{a}), \forall b \in B$ and $g(\bar{b}) \leq K(a, \bar{b}), \forall a \in A$. Gathering these relations yields $K(\bar{a}, b) \leq K(\bar{a}, \bar{b}) \leq K(a, \bar{b}), \forall a \in A, b \in B : (\bar{a}, \bar{b})$ is a saddle-point.

The identity (3.6) follows from (3.8) and Lemma 3.3-a. The last statement is straightforward since (3.7) is simply a restatement of the saddle-point property in terms of subdifferentials and superdifferentials, see Propositions 1.24 and 1.40. □

The relations (3.7) are usually called the *Karush-Kuhn-Tucker relations.*

The interest of the saddle-point method. The main interests of this method:

- The dual equality $\inf(\mathcal{P}) = \sup(\mathcal{D})$ allows us to obtain the value of $(\mathcal{P})$ by computing the value of $(\mathcal{D})$. In general, the saddle-point method is interesting when the dual problem is easier to solve than the primal one.
- Suppose that you can compute a maximizer $\bar{b}$ of $(\mathcal{D})$. Then, the Karush-Kuhn-Tucker relations give us equations in a with $\bar{b}$ as a parameter, whose solutions are candidates to be solutions to the primal problem $(\mathcal{P})$.

3.2. **Perturbations.** Clearly, there are many Lagrangians K such that (3.1) holds for a given objective function f. Of course, different K's lead to different dual problems. Following Rockafellar's monograph [6], we are going to expose a general method for deriving Lagrangians. The main idea is to imbed the problem $(\mathcal{P})$ in a family of *perturbed* minimization problems $(\mathcal{P})_{q\in Q}$:

$$\textit{minimize } F(a,q), \quad a \in A \tag{$\mathcal{P}_q$}$$

where q runs through a vector space Q and $F: A\times Q \to [-\infty,+\infty]$ is a function such that

$$F(a,q=0) = f(a), \quad \forall a\in A \tag{3.9}$$

so that $(\mathcal{P}_{q=0})$ is $(\mathcal{P})$.
Let B be another vector space such that $\langle B,Q\rangle$ is a *topological* dual pairing. The Lagrangian associated with the perturbation F and the duality $\langle B,Q\rangle$ is defined by

$$K(a,b) \triangleq \inf_{q\in Q}\{\langle b,q\rangle + F(a,q)\}, \quad a\in A, b\in B. \tag{3.10}$$

In other words, for any $a\in A$, $b\mapsto K(a,b)$ is the concave conjugate of the function $q\mapsto -F(a,q)$ and as such it is a concave function: For all $a\in A, b\in B$

$$K_a(b) = (-F_a)^{\hat{*}}(b) \quad \text{or equivalently} \quad -K_a(-b) = F_a^*(b)$$

where the subscript $a\in A$ indicates the fixed variable.
Assuming that F is chosen such that $q\in Q\mapsto F(a,q)\in[-\infty,\infty]$ is a closed convex function for all $a\in A$, with Theorem 2.30 one can reverse the conjugate duality relation $-K_a(-b) = F_a^*(b)$ to obtain

$$F(a,q) = \sup_{b\in B}\{K(a,b) - \langle b,q\rangle\}, \forall a\in A, q\in Q. \tag{3.11}$$

In particular, with $q=0$ and (3.9) one recovers (3.1):

$$f(a) = \sup_{b\in B} K(a,b), \quad a\in A.$$

Let us think of K as a pivot: If K is convex in a then $-K$ is concave in a and convex in b. This suggests to introduce another vector space P such that $\langle P,A\rangle$ is a topological dual pairing and to introduce also the function

$$G(b,p) \triangleq \inf_{a\in A}\{K(a,b) - \langle a,p\rangle\}, b\in B, p\in P. \tag{3.12}$$

This formula is analogous to (3.11). Since

$$G(b,p) = \inf_{a,q}\{\langle b,q\rangle - \langle a,p\rangle + F(a,q)\}, b\in B, p\in P, \tag{3.13}$$

one sees that G is jointly closed concave, as a concave conjugate. Going on symetrically, one interprets G as the concave perturbation of the objective concave function

$$g(b) \triangleq G(b,0), b\in B$$

associated with the concave maximization problem

$$\textit{maximize } g(b), b\in B \tag{$\mathcal{D}$}$$

which is called the *dual problem* of the *primal problem* $(\mathcal{P})$. It is imbedded in the family of concave maximization problems $(\mathcal{D}_p)_{p\in P}$

$$\textit{maximize } G(b,p), b\in B. \tag{$\mathcal{D}_p$}$$

The *value function* of $(\mathcal{P}_q)_{q\in Q}$ is defined by

$$\varphi(q) \overset{\triangle}{=} \inf(\mathcal{P}_q) = \inf_{a\in A} F(a,q) \in [-\infty,+\infty], q \in Q.$$

Note that $\inf(\mathcal{P}) = \varphi(0)$.
It will be very useful that φ is a convex function. This is the reason why we are going to assume that the perturbation is chosen such that F is jointly convex on $A \times Q$.
Then, $(\mathcal{P}_q)_{q\in Q}$ is a family of *convex* minimization problems. In particular, because of (3.9), this requires that f *is a convex function.*

Lemma 3.14. *If F is jointly convex on $A \times Q$, then φ is convex.*

Proof. This follows from the fact that the epigraph of φ is "essentially" a linear (marginal) projection of the convex epigraph of F so that it is also convex.
Let us prove that $\operatorname{epi}\varphi$ is a convex set. For all $q \in Q$, $\alpha \in \mathbb{R}$,

$$\begin{aligned}
(q,\alpha) \in \operatorname{epi}\varphi \quad &\Leftrightarrow \quad \inf_a F(a,q) \le \alpha \\
&\Leftrightarrow \quad \forall\delta > 0, \exists a \in A, F(a,q) \le \alpha + \delta \\
&\Leftrightarrow \quad \forall\delta > 0, \exists a \in A, (a,q,\alpha+\delta) \in \operatorname{epi} F \\
&\Leftrightarrow \quad \forall\delta > 0, (q,\alpha+\delta) \in \operatorname{proj}(\operatorname{epi} F)
\end{aligned}$$

where $\operatorname{proj}(\operatorname{epi} F)$ is the linear canonical projection from $A \times Q \times \mathbb{R}$ onto $Q \times \mathbb{R}$ of the set $\operatorname{epi} F$. As $\operatorname{epi} F$ is convex and the projection is linear, $\operatorname{proj}(\operatorname{epi} F)$ is a convex set.
Let $q_0, q_1 \in Q$, $\alpha_0, \alpha_1 \in \mathbb{R}$ and for all $0 \le t \le 1$, define $q_t := (1-t)q_0 + tq_1$ and $\alpha_t := (1-t)\alpha_0 + t\alpha_1$. Suppose that (q_0,α_0) and (q_1,α_1) are in $\operatorname{epi}\varphi$. Then, for all $\delta > 0$, $(q_0, \alpha_0+\delta) \in \operatorname{proj}(\operatorname{epi} F)$ and $(q_1, \alpha_1+\delta) \in \operatorname{proj}(\operatorname{epi} F)$. As $\operatorname{proj}(\operatorname{epi} F)$ is convex, we also have that $(1-t)(q_0,\alpha_0+\delta) + t(q_1,\alpha_1+\delta) = (q_t,\alpha_t+\delta) \in \operatorname{proj}(\operatorname{epi} F)$, for all $0 \le t \le 1$. But this is equivalent to $(q_t,\alpha_t) \in \operatorname{epi}\varphi$, which is the desired result. $\square$

If F is jointly convex on $A \times Q$, for any $b \in B$, $a \mapsto K(a,b)$ is a convex function (same argument as for the convexity of φ based on Lemma 3.14). Therefore, K is a *convex-concave* function. We shall see that its *saddle-points* will play a central role.
Similarly, the value function of $(\mathcal{D}_p)_{p\in P}$ is defined by

$$\gamma(p) \overset{\triangle}{=} \sup_{b\in B} G(b,p), p \in P.$$

We have $\sup(\mathcal{D}) = \gamma(0)$.
As G is jointly concave, by Lemma 3.14, γ is a concave function.

3.3. **The main abstract result.** The main abstract result of the theory is stated at Theorem 3.18 below.
We consider two topological pairings $\langle A, P\rangle$ and $\langle B, Q\rangle$. We have the following diagram

$$\begin{array}{ccccccc}
 & & \gamma(p) & & f(a) & & \\
 & \Big\langle & P & , & A & \Big\rangle & \\
G(b,p) & & & K(a,b) & & & F(a,q) \\
 & \Big\langle & B & , & Q & \Big\rangle & \\
 & & g(b) & & \varphi(q) & &
\end{array}$$

Because of (3.11) and (3.12) with $q = 0$ and $p = 0$ we obtain

$$\begin{aligned} f(a) &= \sup_{b \in B} K(a,b), a \in A && (3.15) \\ g(b) &= \inf_{a \in A} K(a,b), b \in B && (3.16) \end{aligned}$$

and the values of the optimization problems satisfy

$$\sup(\mathcal{D}) = \gamma(0) = \sup_b g(b) = \sup_b \inf_a K(a,b) \le \inf_a \sup_b K(a,b) = \inf_a f(a) = \varphi(0) = \inf(\mathcal{P}).$$

It appears that the *dual equality:* $\inf(\mathcal{P}) = \sup(\mathcal{D})$ holds if and only if $\sup_b \inf_a K(a,b) = \inf_a \sup_b K(a,b)$. If this occurs, it is said that K has a *saddle-value.*
The next lemma will be the keystone of the proof of Theorem 3.18.

Lemma 3.17. *We assume that $\langle P, A\rangle$ and $\langle B, Q\rangle$ are topological dual pairings.*

(a) *Without any additional assumptions, we have*

$$g = (-\varphi)^{\hat{*}}.$$

(b) *If F is jointly closed convex on $A \times Q$, we have*

$$f = (-\gamma)^{*}.$$

Proof. Statement (a) is a direct consequence of the definitions. Indeed, for all $b \in B$, $g(b) := \inf_a K(a,b) := \inf_{a,q}\{\langle b,q\rangle + F(a,q)\} := \inf_q\{\langle b,q\rangle + \varphi(q)\}$, which is the desired result.
Let us prove (b). Taking the convex conjugate of (3.13), one obtains

$$F^{**}(a,q) = \sup_{b,p}\{-\langle b,q\rangle + \langle a,p\rangle + G(b,p)\}.$$

As F is supposed to be jointly closed convex, by Theorem 2.30 we have $F = F^{**}$ so that for all a, q, $F(a,q) = \sup_{b,p}\{-\langle b,q\rangle + \langle a,p\rangle + G(b,p)\}$. In particular, with $q = 0$ we get $f(a) := F(a,0) = \sup_p\{\langle a,p\rangle + \sup_b G(b,p)\} := \sup_p\{\langle a,p\rangle + \gamma(p)\}$. This states that $f = (-\gamma)^*$. □

Theorem 3.18. *We assume that $\langle P, A\rangle$ and $\langle B, Q\rangle$ are topological dual pairings.*

(a) We have $\sup(\mathcal{D}) = \varphi^{**}(0)$.
Hence, the dual equality $\inf(\mathcal{P}) = \sup(\mathcal{D})$ *holds if and only if* $\varphi(0) = \varphi^{**}(0)$.

(b) In particular,

$$\left.\begin{array}{l} \bullet\ F \text{ is jointly convex} \\ \bullet\ \varphi \text{ is lower semicontinuous at } 0 \\ \bullet\ \sup(\mathcal{D}) > -\infty \end{array}\right\} \Rightarrow \inf(\mathcal{P}) = \sup(\mathcal{D})$$

(c) If the dual equality holds, then

$$\text{argmax } g = -\partial(\varphi)(0).$$

Let us assume in addition that F is jointly *convex on $A \times Q$ and $q \mapsto F(a,q)$ is a closed convex function for any $a \in A$. Of course, this holds in particular if F is jointly closed convex on $A \times Q$.*

(a') We have $\inf(\mathcal{P}) = \gamma^{\hat{*}\hat{*}}(0)$.
Hence, the dual equality $\inf(\mathcal{P}) = \sup(\mathcal{D})$ *holds if and only if* $\gamma(0) = \gamma^{\hat{*}\hat{*}}(0)$.

(b') In particular,

$$\left.\begin{array}{l}\bullet\ \gamma \textit{ is upper semicontinuous at } 0\\ \bullet\ \inf(\mathcal{P})<+\infty\end{array}\right\}\Rightarrow \inf(\mathcal{P})=\sup(\mathcal{D})$$

(c') If the dual equality holds, then

$$\operatorname{argmin} f = -\widehat{\partial}(\gamma)(0).$$

Proof. • Let us prove (a). Taking the conjugate of the identity of Lemma 3.17-a, one obtains: $-\varphi^{**} = (-\varphi)^{\hat{*}\hat{*}} = g^{\hat{*}}$. Hence, for all $q \in Q$, $\varphi^{**}(q) = \sup_b\{-\langle b, q\rangle + g(b)\}$. In particular, with $q = 0$, one gets $\varphi^{**}(0) = \sup_b g(b) = \sup(\mathcal{D})$. The dual equality is $\varphi(0) = \varphi^{**}(0)$.

• Let us prove (b). Suppose that F is jointly convex on $A \times Q$. By Lemma 3.14, φ is convex. As it is assumed that $\sup(\mathcal{D}) = \varphi^{**}(0) > -\infty$, we have $\operatorname{clcv}\varphi = \operatorname{lscv}\varphi$ (see Proposition 2.35). As φ is convex, we also have: $\operatorname{lscv}\varphi = \operatorname{ls}\varphi$. By Corollary 2.36, $\varphi^{**} = \operatorname{clcv}\varphi$. Therefore, $\varphi^{**} = \operatorname{ls}\varphi$ and in particular $\varphi^{**}(0) = \operatorname{ls}\varphi(0)$. With the help of the local property stated at the Proposition 2.11, we see that $\varphi(0) = \operatorname{ls}\varphi(0)$ if φ is lower semicontinuous at 0.

• Let us prove (c). The dual equality $\varphi(0) = \varphi^{**}(0)$ is equivalent to $-\varphi(0) = (-\varphi)^{\hat{*}\hat{*}}(0)$. The result follows from Theorem 1.44-c and Lemma 3.17-a.

• Let us prove (a'). Taking the conjugate of the identity of Lemma 3.17-b, one obtains: for all $p \in P$, $(-\gamma)^{**}(p) = f^*(p) = \sup_a\{\langle a, p\rangle - f(a)\}$. In particular, with $p = 0$, we get $\gamma^{\hat{*}\hat{*}}(0) = -(-\gamma)^{**}(0) = \inf_a f(a) = \inf(\mathcal{P})$ and the dual equality is $\gamma(0) = \gamma^{\hat{*}\hat{*}}(0)$.

• Let us prove (b'). As it is assumed that $-\inf(\mathcal{P}) = (-\gamma)^{**}(0) > -\infty$, we have $\operatorname{clcv}(-\gamma) = \operatorname{lscv}(-\gamma)$ (see Proposition 2.35). As $-\gamma$ is convex, we also have: $\operatorname{lscv}(-\gamma) = \operatorname{ls}(-\gamma)$. By Corollary 2.36, $(-\gamma)^{**} = \operatorname{clcv}(-\gamma)$. Therefore, $(-\gamma)^{**} = \operatorname{ls}(-\gamma)$ and in particular $(-\gamma)^{**}(0) = \operatorname{ls}(-\gamma)(0)$. With the help of the local property stated at the Proposition 2.11, we see that $(-\gamma)(0) = \operatorname{ls}(-\gamma)(0)$ if γ is upper semicontinuous at 0.

• Let us prove (c'). The dual equality is $(-\gamma)(0) = (-\gamma)^{**}(0)$. The result follows from Theorem 1.38-c and Lemma 3.17-b. ☐

As a remark, the following result shows that if one wants K to be convex-concave, F should be jointly convex.

Proposition 3.19. *We assume that $\langle P, A\rangle$ and $\langle B, Q\rangle$ are topological dual pairings. If for all $a \in A$, F_a is a closed convex function on Q, then K_b is convex on A for all $b \in B$ if and only if F is jointly convex on $A \times Q$.*

Proof. Let us prove the "⇒" part. Thanks to (3.11): $F(a, q) = \sup_b\{K_b(a) - \langle b, q\rangle\}$, one sees that F is convex as the supremum of convex functions.
Let us prove the "⇐" part. Thanks to the definition (3.10): $K_b(a) = \inf_q\{F(a, q)+\langle b, q\rangle\}$. As F is jointly convex, one concludes with Lemma 3.14 that K_b is convex. ☐

3.4. **Minimizing a convex function under convex constraints.** Let us consider the following minimization problem

$$\textit{minimize } h(a), \qquad \textit{subject to} \quad Ta \in C, \qquad a \in A \tag{$\mathcal{P}$}$$

where h is a convex $(-\infty, \infty]$-valued function on a vector space A, $T : A \to Q$ is a linear operator from A to another vector space Q and C is a convex subset of Q.
Defining

$$f(a) = h(a) + \zeta(a \mid Ta \in C), \quad a \in A$$

where ζ is the convex indicator function

$$\zeta(a \mid Ta \in C) = \begin{cases} 0 & \text{if } Ta \in C \\ +\infty & \text{if } Ta \notin C \end{cases},$$

so that $(\mathcal{P})$ is also

$$\textit{minimize } f(a), \quad a \in A$$

with f a convex function on A. The perturbation worth considering is

$$F(a,q) := h(a) + \zeta(a, q \mid Ta + q \in C), \quad a \in A, q \in Q.$$

Let B be another vector space topologically paired with Q. The corresponding Lagrangian is

$$\begin{aligned} K(a,b) &= \inf_{q \in Q} \{\langle b, q\rangle + F(a,q)\} \\ &= \inf_{q \in Q} \{\langle b, q\rangle + h(a) + \zeta(a, q \mid Ta + q \in C)\} \\ &= h(a) + \inf_{q \in Q} \{\langle b, q - Ta\rangle + \zeta(q \mid C)\} \\ &= h(a) - \langle b, Ta\rangle + \inf_{q \in C} \langle b, q\rangle \\ &= h(a) - [T^*b](a) + \inf_{q \in C} \langle b, q\rangle \end{aligned}$$

where T^* is the *algebraic* adjoint of T defined as a linear form on B, for all $b \in B$, by

$$[T^*b](a) := \langle b, Ta\rangle, \quad \forall a \in A.$$

Let us introduce another vector space P *topologically* paired with A and consider the function G on $B \times P$ defined for all $b \in B$ and $p \in P$ by

$$\begin{aligned} G(b,p) &:= \inf_a \{K(a,b) - \langle a, p\rangle\} \\ &= \inf_{q \in C} \langle b, q\rangle + \inf_a \{h(a) - [T^*b](a) - \langle a, p\rangle\} \\ &= \inf_{q \in C} \langle b, q\rangle - \sup_a \{[T^*b](a) + \langle a, p\rangle - h(a)\} \end{aligned}$$

We make the assumption that

$$T^*(B) \subset P, \tag{3.20}$$

so that one can write $[T^*b](a) = \langle T^*b, a\rangle_{P,A} = \langle b, Ta\rangle_{B,Q}$. It follows that the diagram

$$\begin{array}{ccccc} \Big\langle & P & , & A & \Big\rangle \\ & T^* \Big\uparrow & & \Big\downarrow T & \\ \Big\langle & B & , & Q & \Big\rangle \end{array}$$

is meaningful. It is now possible to rewrite

$$K(a,b) = \inf_{q \in C} \langle b, q\rangle + h(a) - \langle T^*b, a\rangle, \quad a \in A, b \in B. \tag{3.21}$$

and

$$G(b,p) = \inf_{q \in C} \langle b, q\rangle - h^*(T^*b + p).$$

With $p = 0$, we obtain the objective function

$$g(b) = \inf_{q \in C} \langle b, q\rangle - h^*(T^*b), \quad b \in B$$

of the associated dual problem

$$\textit{maximize} \inf_{q\in C}\langle b,q\rangle - h^*(T^*b), \quad b\in B. \tag{$\mathcal{D}$}$$

Note that ($\mathcal{D}$) is an *unconstrained* maximization problem. The value functions are

$$\varphi(q) = \inf_{a:Ta\in C-q} h(a), \quad q\in Q$$

and

$$\gamma(p) = \sup_b\{\inf_{q\in C}\langle b,q\rangle - h^*(T^*b+p)\}, \quad p\in P.$$

Assumptions 3.22. *Before stating the assumptions on h, C and T one has to describe the topologies.*
The topologies. We consider two topological dual pairings $\langle A,P\rangle$ and $\langle B,Q\rangle$, where

- *P is a Hausdorff locally convex topological vector space*
- *$A = P'$ is the topological dual space of P endowed with the weak topology $\sigma(A,P)$*
- *B is a Hausdorff locally convex topological vector space*
- *$Q = B'$ is the topological dual space of B endowed with the weak topology $\sigma(Q,B)$*

Note that A, B, P and Q are Hausdorff locally convex topological vector spaces.
The assumptions on h, C and T are

(A1) *h is a convex $\sigma(A,P)$-lower semicontinuous function.*
(A2) *h is bounded below: $\inf h > -\infty$.*
(A3) *C is a convex $\sigma(Q,B)$-closed set.*
(A4) *T is a linear operator from A to Q such that $T^*B\subset P$.*
(A5) *There exists an open neighbourhood N of zero in P such that*

$$\sup_{p\in N} h^*(p) < +\infty.$$

The next lemma will allow us to use the general results of Section 3.3.

Lemma 3.23. *Under these assumptions, the following assertions hold.*

(a) T is continuous
(b) h is a closed convex and inf-compact function on A.
(c) F is jointly closed convex on $A\times Q$.

Proof. Let us prove (a). To prove that T is continuous, one has to show that for any $b\in B$, $a\in A\mapsto\langle b,Ta\rangle\in\mathbb{R}$ is continuous. By (A4), we get $a\mapsto\langle b,Ta\rangle = \langle T^*b,a\rangle$ which is continuous since $T^*b\in P$.
Let us prove (b). By (A1) and (A2), h is a convex lower semicontinuous function such that $f(a) > -\infty, \forall a\in A$. Hence, it is closed convex. Thanks to (A5) and Proposition 2.43, h^{**} is inf-compact. But we also have $h = h^{**}$ by Theorem 2.30.
Let us prove (c). As T is linear continuous and C is closed convex, $\{(a,q); Ta+q\in C\}$ is closed convex in $A\times Q$. As h is closed convex on A, its epigraph is closed convex in $A\times\mathbb{R}$. It follows that $\operatorname{epi} F = (Q\times\operatorname{epi} h)\cap\{(a,q); Ta+q\in C\}$ is closed convex, which implies that F is convex and lower semicontinuous. As it is nowhere equal to $-\infty$ (since $\inf F\geq\inf h > -\infty$, by assumption (A2)), F is also a closed convex function. □

We are now ready to prove the primal attainment and the dual equality.
A minimizing sequence of ($\mathcal{P}$) is a sequence (a_n) such that $Ta_n\in C$ for all $n\geq 1$ and $\lim_{n\to\infty} h(a_n) = \inf(\mathcal{P})$.

Theorem 3.24 (Primal attainment)**.** *Under our assumptions, suppose that* $\inf(\mathcal{P}) < +\infty$. *Then the primal problem* $(\mathcal{P})$ *admits at least one solution and if* h *is stricly convex, this solution is unique.*
Any minimizing sequence admits $\sigma(A, P)$*-accumulation points. All these accumulation points are solutions to* $(\mathcal{P})$*. If* h *is strictly convex, any minimizing sequence* $\sigma(A, P)$*-converges to the unique solution of* $(\mathcal{P})$*.*

Proof. By Lemma 3.23, h is inf-compact and the constraint set $\{a \in A; Ta \in C\}$ is closed. All the statements of the theorem are direct consequences of Corollary 2.15 and Theorem 1.14.

□

Theorem 3.25 (Dual equality)**.** *Under our assumptions, the dual equality* $\inf(\mathcal{P}) = \sup(\mathcal{D})$ *holds. That is*

$$\inf\{h(a); a : Ta \in C\} = \sup_{b\in B}\{\inf_{q\in C}\langle b, q\rangle - h^*(T^*b)\} \in (-\infty, +\infty].$$

Proof. We consider separately the two situations where $\inf(\mathcal{P})$ is finite or infinite.
Case where $\inf(\mathcal{P}) < +\infty$. Thanks to Theorem 3.18-b', it is enough to prove that γ is upper semicontinuous at $p = 0$. We are going to prove that γ is continuous at $p = 0$. Indeed, for all $p \in P$,

$$-\gamma(p) = \inf_b\{h^*(T^*b + p) - \inf_{q\in Q}\langle b, q\rangle\} \le h^*(p)$$

where the inequality is obtained taking $b = 0$. But, by Assumption (A5), h^* is upper bounded on an open neighbourhood of 0, and so is the convex function $-\gamma$. Hence, $-\gamma$ is continuous at 0 by virtue of Lemma 2.44.

Case where $\inf(\mathcal{P}) = +\infty$. Note that $\sup(\mathcal{D}) \ge g(0) = -h^*(0) = \inf_a h(a) > -\infty$ by Assumption (A2), so that we can apply Theorem 3.18-b. It is enough to prove that

$$\mathrm{ls}\,\varphi(0) = +\infty$$

in the situation where $\varphi(0) = \inf(\mathcal{P}) = +\infty$. By Proposition 2.11, we have $\mathrm{ls}\,\varphi(0) = \sup_{U\in\mathcal{N}(0)} \inf\{h(a); a : Ta \in C + U\}$ where $\mathcal{N}(0)$ is the set of all the open neighbourhoods of $0 \in Q$. It follows that for all $U \in \mathcal{N}(0)$, there exists $a \in A$ such that $Ta \in C + U$ and $h(a) \le \mathrm{ls}\,\varphi(0)$. This implies that

$$T(\{h \le \mathrm{ls}\,\varphi(0)\}) \cap (C + U) \ne \emptyset, \quad \forall U \in \mathcal{N}(0). \tag{3.26}$$

On the other hand, $\inf(\mathcal{P}) = +\infty$ is equivalent to: $T(\mathrm{dom}\, h) \cap C = \emptyset$.
Now, we prove ad absurdum that $\mathrm{ls}\,\varphi(0) = +\infty$. Suppose that $\mathrm{ls}\,\varphi(0) < +\infty$. Because of $T(\mathrm{dom}\, h) \cap C = \emptyset$, we have a fortiori

$$T(\{h \le \mathrm{ls}\,\varphi(0)\}) \cap C = \emptyset.$$

As h is inf-compact and T is continuous (Lemma 3.23, (a) and (b)), $T(\{h \le \mathrm{ls}\,\varphi(0)\})$ is a compact subset of Q. It is also convex, since the level sets of a convex function are convex and the image of a convex set by a linear mapping is a convex set. But C is assumed to be closed and convex, so that by Hahn-Banach theorem (Corollary 2.19), C and $T(\{h \le \mathrm{ls}\,\varphi(0)\})$ are *strictly* separated. This contradicts (3.26), considering open neighbourhoods U of the origin in (3.26) which are open half-spaces. Consequently, $\mathrm{ls}\,\varphi(0) = +\infty$. □

Theorem 3.27 (Karush-Kuhn-Tucker relations)**.** *Under our assumptions, the following statements are equivalent.*

(1) The primal and dual problems are attained at $\bar{a}$ and $\bar{b}$ respectively
(2) The following KKT relations hold:
(a) $T\bar{a} \in C$
(b) $\langle \bar{b}, T\bar{a}\rangle \leq \langle \bar{b}, q\rangle$ for all $q \in C$
(c) $\bar{a} \in \partial h^*(T^*\bar{b})$.

Proof. This is a direct application of the KKT part of Theorem 3.5 with K defined by (3.21):

$$K(a,b) = \inf_{q\in C} \langle b, q\rangle + h(a) - \langle T^*b, a\rangle, a \in A, b \in B.$$

Note that as the dual equality holds by Theorem 3.25, (1) is the statement (3) of Theorem 3.5 with K as above. Hence, it remains to check that (2) is (3.7) with K as above.
We have

$$\begin{aligned} 0 \in \partial_a K(\bar{a}, \bar{b}) &\Leftrightarrow T^*\bar{b} \in \partial h(\bar{a}) \\ &\Leftrightarrow \bar{a} \in \partial h^*(T^*\bar{b}) \end{aligned}$$

where the last equivalence holds by Theorem 2.46 since h is closed convex by Assumptions (A1-2). We also have

$$\begin{aligned} 0 \in \widehat{\partial}_b K(\bar{a}, \bar{b}) &\Leftrightarrow T\bar{a} \in \widehat{\partial}\left[\inf_{q\in C}\langle \cdot, q\rangle\right](\bar{b}) \\ &\Leftrightarrow -T\bar{a} \in \partial\left[\sup_{q\in -C}\langle \cdot, q\rangle\right](\bar{b}) = \partial\zeta^*_{-C}(\bar{b}) \\ &\overset{(a)}{\Leftrightarrow} \bar{b} \in \partial\zeta_{-C}(-T\bar{a}) \\ &\Leftrightarrow \forall q \in Q, \zeta_{-C}(q) \geq \zeta_{-C}(-T\bar{a}) + \langle \bar{b}, q + T\bar{a}\rangle \\ &\Leftrightarrow \forall q \in -C, \zeta_{-C}(-T\bar{a}) + \langle \bar{b}, q + T\bar{a}\rangle \leq 0 \\ &\Leftrightarrow \begin{cases} -T\bar{a} \in -C \\ \langle \bar{b}, T\bar{a}\rangle \leq \langle \bar{b}, -q\rangle, \forall q \in -C. \end{cases} \end{aligned}$$

The equivalence (a) holds since C is a closed convex set so that the indicator function ζ_{-C} is a closed convex function and one can apply Theorem 2.46. This completes the proof of the theorem. □

This result is far from being the whole story. In practice, the dual attainment in B is not the rule and one has too work hard to obtain it in a larger space.

4. Optimal Transport

Let us consider two spaces $\mathcal{X}$ and $\mathcal{Y}$ equipped with σ-fields and c a measurable $[0,\infty)$-valued function on the product space $\mathcal{X} \times \mathcal{Y}$. We are given two probability measures μ on $\mathcal{X}$ and ν on $\mathcal{Y}$. The Monge-Kantorovich optimal transport cost of μ on ν for the cost function c is defined by

$$\inf_\pi \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\,\pi(dxdy) \in [0,\infty]$$

where the infimum is taken over all probability measures π on $\mathcal{X} \times \mathcal{Y}$ with first marginal $\pi_{\mathcal{X}}(dx) := \pi(dx \times \mathcal{Y}) = \mu(dx)$ and second marginal $\pi_{\mathcal{Y}}(dy) := \pi(\mathcal{X} \times dy) = \nu(dy)$. We denote the constraint set

$$P(\mu,\nu) = \{\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : \pi_{\mathcal{X}} = \mu, \pi_{\mathcal{Y}} = \nu\}.$$

In the special important case where $\mathcal{X} = \mathcal{Y}$ is equipped with a metric d, popular cost functions are $c(x,y) = d(x,y)^p$ where $p \geq 1$.
Denoting $\mathcal{P}(\mathcal{X})$ the set of all probability measures on $\mathcal{X}$, the Monge-Kantorovich problem is the following

$$\textit{minimize} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\,\pi(dxdy) \textit{ subject to } \pi \in P(\mu,\nu). \tag{MK-P}$$

Any minimizer $\pi \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$ of this problem is called an optimal transport plan of μ on ν for the cost function c. It is immediate to check that this is a *convex problem* since the constraint is convex and the objective function is linear.

Remark 4.1. The objective function $\pi \mapsto \int c\,d\pi$ is convex but not strictly convex, since it is affine. Consequently, one may not expect uniqueness of the optimal plan in a general situation. Recall that uniqueness is the rule if the objective function is strictly convex (see Proposition 1.14).

4.1. **Primal attainment.** We first recall some useful results about compactness and probability measures built on a Borel σ-field.
A probability measure $\mu \in \mathcal{P}(\mathcal{X})$ is said to be *tight* if for any $\epsilon > 0$, there exists a compact set K_ϵ such that $\mu(\mathcal{X} \setminus K_\epsilon) \leq \epsilon$.
Similarly, a family $M \subset \mathcal{P}(\mathcal{X})$ of probability measures is said to be *uniformly tight* if for any $\epsilon > 0$, there exists a compact set K_ϵ such that $\sup_{\mu\in M} \mu(\mathcal{X} \setminus K_\epsilon) \leq \epsilon$.
On a metric space $\mathcal{X}$, if $M \subset \mathcal{P}(\mathcal{X})$ is uniformly tight, then it is relatively compact for the usual weak topology (see [1], Theorem 6.1).
Recall that a topological space is said to be *Polish* if it is a separable complete metric space. An important result is that any probability measure on a Polish space $\mathcal{X}$ is tight (see [4], Proposition II.7.3) and more generally, $M \subset \mathcal{P}(\mathcal{X})$ is uniformly tight if and only if it is relatively compact for the usual weak topology (see [1], Theorem 6.2).

Theorem 4.2 (Primal attainment). *We suppose that $\mathcal{X}$ and $\mathcal{Y}$ are Polish spaces and that the cost function c is a $[0,+\infty)$-valued lower semicontinuous function on $\mathcal{X}\times\mathcal{Y}$. If there exists $\pi_o \in P(\mu,\nu)$ such that $\int_{\mathcal{X}\times\mathcal{Y}} c\,d\pi_o < +\infty$, then the problem (MK-P) admits a minimizer.*
This is the case if there exist two $[0,\infty)$-valued measurable functions $c_\mathcal{X}$ on $\mathcal{X}$ and $c_\mathcal{Y}$ on $\mathcal{Y}$ such that $c \leq c_\mathcal{X} \oplus c_\mathcal{Y}$, $\int_\mathcal{X} c_\mathcal{X}\,d\mu < \infty$ and $\int_\mathcal{Y} c_\mathcal{Y}\,d\nu < \infty$ (take $\pi_o = \mu\otimes\nu$).

Proof. The sets of probability measures are equipped with their respective weak topologies.
Let us first prove that $P(\mu,\nu)$ is compact. As the mapping $\pi \in \mathcal{P}(\mathcal{X}\times\mathcal{Y}) \mapsto (\pi_\mathcal{X}, \pi_\mathcal{Y}) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ is continuous, $P(\mu,\nu)$ is closed.
But, μ and ν are tight, since they are built on Polish spaces. This means that for all $\epsilon > 0$, there exist two compact sets K and K' such that $\mu(\mathcal{X}\setminus K) \leq \epsilon$ and $\nu(\mathcal{Y}\setminus K') \leq \epsilon$.
Now, for all $\pi \in P(\mu,\nu)$, we have

$$\begin{aligned} \pi(\mathcal{X}\times\mathcal{Y} \setminus K\times K') &\leq \pi(\mathcal{X}\times\mathcal{Y}\setminus K\times\mathcal{Y}) + \pi(\mathcal{X}\times\mathcal{Y}\setminus\mathcal{X}\times K') \\ &= \mu(\mathcal{X}\setminus K) + \nu(\mathcal{Y}\setminus K') \leq 2\epsilon \end{aligned} \tag{4.3}$$

As $K \times K'$ is compact, this implies that $P(\mu,\nu)$ is uniformly tight and it follows that it is relatively compact. Therefore, $P(\mu,\nu)$ is compact. It is also assumed that it is non-empty.
To complete the proof, thanks to Theorem 2.12, it remains to show that $\pi \in \mathcal{P}(\mathcal{X}\times\mathcal{Y}) \mapsto \int_{\mathcal{X}\times\mathcal{Y}} c\,d\pi \in [0,+\infty]$ is lower semicontinuous. But, as c is assumed to be $[0,\infty)$-valued and lower semicontinuous, Lemma 4.4 below states that c is the limit of an increasing

sequence (c_n) of continuous bounded functions. By the monotone convergence theorem, we obtain that $\int c\,d\pi = \sup_n \int c_n\,d\pi$. It follows that $\pi \mapsto \int c\,d\pi$ is lower semicontinuous, being the supremum of continuous functions. $\square$

During this proof, we used the following lemma.

Lemma 4.4 (Moreau-Yosida approximation)**.** *On a metric space, any lower semicontinuous real valued function which is bounded below is the pointwise limit of an increasing sequence of Lipschitz continuous bounded functions.*

Proof. Let f be a lower semicontinuous function on X equipped with its metric d. As f is bounded below, one can assume without restriction that $f \geq 0$. It is enough to build a sequence (f_n) of (possibly not bounded) Lipschitz continuous functions such that (f_n) is increasing and $\lim_n f_n = f$ pointwise, since $\min(f_n, n)$ still does the same work.
A good sequence is

$$f_n(x) = \inf\{f(z) + nd(x,z); z \in X\}.$$

It is called the Moreau-Yosida approximation of f. Clearly (f_n) is an increasing sequence since $f(z) + md(x,z) \leq f(z) + nd(x,z)$ for all $x, z \in X$ and $m \leq n$.
As $x \mapsto nd(x,z)$ a Lipschitz function (with Lipschitz constant n) for each $z \in X$, f_n is also n-Lipschitz (even if f is irregular) by virtue of Lemma 4.7 below.
Now, let us prove that $\lim_n f_n(x) = f(x)$. For all $k \geq 1$, we define $g_k(x) = \inf\{f(z); z \in X, d(x,z) < 1/k\}$. As f is lower semicontinuous, by Proposition 2.11, we have

$$\lim_k g_k(x) = \operatorname{ls} f(x) = f(x). \tag{4.5}$$

On the other hand, for all k there exists N_k such that

$$g_k(x) \leq f_n(x), \quad \forall n \geq N_k. \tag{4.6}$$

This follows from

$$\begin{aligned}
f_n(x) &= \inf\{f(z) + nd(x,z); z \in X\} \\
&= \min\Big(\inf\{f(z) + nd(x,z); z : d(z,x) < 1/k\}, \inf\{f(z) + nd(x,z); z : d(z,x) \geq 1/k\}\Big) \\
&\geq \min\Big(\inf\{f(z); z : d(z,x) < 1/k\}, n/k + \inf\{f(z); z : d(z,x) \geq 1/k\}\Big) \\
&\geq \min\Big(g_k(x), n/k + \inf\{f(z); z \in X\}\Big) \\
&= g_k(x)
\end{aligned}$$

for all n such that $n/k \geq f(x) - \inf f$, so that $n/k + \inf f \geq f(x) \geq g_k(x)$.
Letting k tend to infinity in (4.6) and taking (4.5) into account together with the fact that $f_n(x)$ admits a limit as an increasing sequence, we obtain $f(x) \leq \limsup_n f_n(x) = \lim_n f_n(x)$. But the converse inequality: $\lim_n f_n(x) \leq f(x)$ follows directly from the definition of f_n which implies that $f_n \leq f$. We have proved that $\lim_n f_n(x) = f(x)$ and this completes the proof of the lemma. $\square$

During the proof of this lemma, we used the following general fact about Lipschitz functions.

Lemma 4.7. *Let $(f_i; i \in I)$ be a collection of K-Lipchitz functions (with respect to some metric d) for some constant $K \geq 0$. That is $|f_i(x) - f_i(y)| \leq Kd(x,y)$ for all $i \in I$ and all $x, y \in \mathcal{X}$. Then $\sup_{i \in I} f_i$ and $\inf_{i \in I} f_i$ are also K-Lipschitz functions.*

Proof. Let $\epsilon > 0$ and $x, y \in \mathcal{X}$. There exists $i_o \in I$ such that $\sup_i f_i(x) - \sup_i f(y) \leq f_{i_o}(x) + \epsilon - \sup_i f_i(y) \leq f_{i_o}(x) - f_{i_o}(y) + \epsilon \leq Kd(x, y) + \epsilon$. As this holds for all $\epsilon > 0$, we get $\sup_i f_i(x) - \sup_i f(y) \leq Kd(x, y)$. Inverting x and y leads us to the desired inequality. A similar proof works for $\inf_i f_i$. □

4.2. **An equivalent relaxed minimization problem.** We first relax the problem (MK-P). This will be the first step to obtain the dual equality for (MK-P) at Theorem 4.13. We assume that $\mathcal{X}$ and $\mathcal{Y}$ are topological spaces endowed with their Borel σ-fields. The space of continuous bounded functions on $\mathcal{X}$: $C_b(\mathcal{X})$, is equipped with the topology of uniform convergence and its topological space which is denoted $\mathcal{E}(\mathcal{X})$ is equipped with the $*$-weak topology $\sigma(\mathcal{E}(\mathcal{X}), C_b(\mathcal{X}))$. Similar notations hold for $\mathcal{Y}$ and $\mathcal{X} \times \mathcal{Y}$ instead of $\mathcal{X}$; similar topologies are also considered: $\sigma(\mathcal{E}(\mathcal{Y}), C_b(\mathcal{Y}))$ and $\sigma(\mathcal{E}(\mathcal{X} \times \mathcal{Y}), C_b(\mathcal{X} \times \mathcal{Y}))$. We are going to apply the results of Section 3 with

$$\begin{array}{ccc} \Big\langle \quad P = C_b(\mathcal{X} \times \mathcal{Y}) & , & A = \mathcal{E}(\mathcal{X} \times \mathcal{Y}) \quad \Big\rangle \\ T^* \ \uparrow & & \downarrow \ T \\ \Big\langle \ B = C_b(\mathcal{X}) \times C_b(\mathcal{Y}) & , & Q = \mathcal{E}(\mathcal{X}) \times \mathcal{E}(\mathcal{Y}) \ \Big\rangle \end{array}$$

where T is the marginal operator, defined for all $m \in \mathcal{E}(\mathcal{X} \times \mathcal{Y})$, by

$$Tm = (m_{\mathcal{X}}, m_{\mathcal{Y}}) \in \mathcal{E}(\mathcal{X}) \times \mathcal{E}(\mathcal{Y})$$

and

$$\begin{array}{rcl} \langle m_{\mathcal{X}}, \varphi \rangle & = & \langle \varphi \otimes 1, m \rangle, \quad \forall \varphi \in C_b(\mathcal{X}) \\ \langle m_{\mathcal{Y}}, \psi \rangle & = & \langle 1 \otimes \psi, m \rangle, \quad \forall \psi \in C_b(\mathcal{Y}) \end{array}$$

The relaxed version of the minimization problem (MK-P) is

$$\textit{minimize } \mathcal{C}(m) + \zeta_{\mathcal{E}_+}(m) \textit{ subject to } Tm = (\mu, \nu), \quad m \in \mathcal{E}(\mathcal{X} \times \mathcal{Y}) \tag{$\mathcal{P}$}$$

with

$$\mathcal{C}(m) := \sup\{\langle \tilde{c}, m \rangle; \tilde{c} \in C_b(\mathcal{X} \times \mathcal{Y}), \tilde{c} \leq c\}, \quad m \in \mathcal{E}(\mathcal{X} \times \mathcal{Y})$$

and $\mathcal{E}_+ := \{m \in \mathcal{E}(\mathcal{X} \times \mathcal{Y}); m \geq 0\}$ is the cone of the nonnegative elements of $\mathcal{E}(\mathcal{X} \times \mathcal{Y})$: $m \geq 0$ if and only if $\langle m, \theta \rangle \geq 0$ for all nonnegative $\theta \in C_b(\mathcal{X} \times \mathcal{Y})$.

Proposition 4.8. *Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are Polish spaces and that c is a lower semi-continuous $[0, +\infty]$-valued function on $\mathcal{X} \times \mathcal{Y}$. Then, the minimization problems (MK-P) and ($\mathcal{P}$) are equivalent.*

This means that they admit the same minimizers and values.

Proof. This result is a restatement of Lemmas 4.10 and 4.12 below. □

What a probability measure is. Before stating these lemmas, one needs to make precise what it is meant for an element ℓ of $\mathcal{E}(\mathcal{X})$ to be a probability measure. An element ℓ of $\mathcal{E}(\mathcal{X})$ is said to act as a probability measure if there exits a *unique* probability measure $\bar{\ell}$ on $\mathcal{X}$ equipped its Borel σ-field such that

$$\langle \varphi, \ell \rangle = \int_{\mathcal{X}} \varphi \, d\bar{\ell}, \quad \forall \varphi \in C_b(\mathcal{X}).$$

Any $\ell \in \mathcal{E}(\mathcal{X})$ acts linearly on $C_b(\mathcal{X})$ and for any sequence (φ_n) in $C_b(\mathcal{X})$ converging uniformly to zero, we have $\lim_n \langle \varphi_n, \ell \rangle = 0$ as ℓ is continuous with respect to the uniform topology. To be a probability measure, ℓ must of course be nonnegative: $\langle \varphi, \ell \rangle \geq 0$, for all nonnegative φ in $C_b(\mathcal{X})$, and have unit mass: $\langle 1, \ell \rangle = 1$. But there are such ℓ's in

$\mathcal{E}(\mathcal{X})$ which are *not* countably additive (but only finitely additive), and therefore are not measures. We have the following result. If $\mathcal{X}$ is a metric space, a nonnegative $\ell \in \mathcal{E}(\mathcal{X})$ with unit mass acts as a probability measure if and only if for any *decreasing* sequence (φ_n) in $C_b(\mathcal{X})$ such that $0 \le \varphi_n \le 1$ for all n and $\lim_n \varphi_n = 0$ *pointwise*, we have

$$\lim_{n\to\infty} \langle \varphi_n, \ell\rangle = 0. \tag{4.9}$$

This result is a generalized version of the extension result of Daniell's integrals. For a more general result with its proof, see ([4], Proposition II.7.2). The uniqueness of the extension follows from the fact that in a metric space, the Borel σ-field is generated by the continuous bounded functions.
We write shortly $\ell \in \mathcal{P}(\mathcal{X})$ to specify that ℓ acts as a probability measure on $\mathcal{X}$.

Lemma 4.10. *Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are Polish spaces. Then, any m in $\mathcal{E}_+$ such that $Tm = (\mu, \nu)$ with $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$, acts as a probability measure on $\mathcal{X} \times \mathcal{Y}$.*

Proof. Let m be any m in $\mathcal{E}_+$ such that $Tm = (\mu, \nu)$ with $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$. All we have to prove is that m satisfies the σ-additivity property (4.9). Indeed, m is in $\mathcal{E}_+$ (it is a nonnegative), and as its marginal projections have a unit mass, m has also a unit mass.
Let us prove that m satisfies (4.9). We have to check that for any sequence $(\theta_n)_{n\ge1}$ in $C_b(\mathcal{X}\times\mathcal{Y})$ such that $0 \le \theta_n \le 1$ for all n, which is decreasing and converging pointwise to zero: $\theta_n \downarrow 0$ as n tends to infinity, we have $\lim_n \langle \theta_n, m\rangle = 0$.
Since $\mathcal{X}$ and $\mathcal{Y}$ are Polish spaces, μ and ν are tight (see [4], Proposition II.7.3): for all $\epsilon > 0$, there exist K a compact subset of $\mathcal{X}$ and K' a compact subset of $\mathcal{Y}$ such that $\mu(\mathcal{X} \setminus K) \le \epsilon$ and $\nu(\mathcal{Y} \setminus K') \le \epsilon$. As a Polish space is completely regular, there also exist $\varphi_\epsilon \in C_b(\mathcal{X})$ and $\psi_\epsilon \in C_b(\mathcal{Y})$ both with a compact support such that $0 \le \varphi_\epsilon, \psi_\epsilon \le 1$, $\int_{\mathcal{X}}(1-\varphi_\epsilon)\,d\mu \le \epsilon$ and $\int_{\mathcal{Y}}(1-\psi_\epsilon)\,d\nu \le \epsilon$. But, as $1 - ab \le 2 - a - b$ for all $a, b \le 1$, taking the nonnegativity of m into account, we obtain

$$\begin{aligned}
\langle (1-\varphi_\epsilon\otimes\psi_\epsilon), m\rangle &\le \langle (1-\varphi_\epsilon\otimes 1), m\rangle + \langle (1-1\otimes\psi_\epsilon), m\rangle \\
&= \langle (1-\varphi_\epsilon), m_{\mathcal{X}}\rangle + \langle (1-\psi_\epsilon), m_{\mathcal{Y}}\rangle \\
&= \int_{\mathcal{X}} (1-\varphi_\epsilon)\,d\mu + \int_{\mathcal{Y}} (1-\psi_\epsilon)\,d\nu \\
&\le 2\epsilon.
\end{aligned}$$

Note that this argument mimicks (4.3).
Therefore, for all $\theta \in C_b(\mathcal{X}\times\mathcal{Y})$ with $0 \le \theta \le 1$, we have $0 \le \langle\theta, m\rangle \le 2\epsilon + \langle (\varphi_\epsilon\otimes\psi_\epsilon)\theta, m\rangle$. But, in restriction to any compact set, m is a measure. Indeed, let (θ_n) be a sequence in $C_b(\mathcal{X}\times\mathcal{Y})$ such that $0 \le \theta_n \le 1$ for all n, which is decreasing and converging pointwise to zero. Then, the sequence of continuous functions $((\varphi_\epsilon\otimes\psi_\epsilon)\theta_n)_{n\ge1}$ decreases pointwise to zero on the support of $\varphi_\epsilon\otimes\psi_\epsilon$ which is a compact set. Therefore, it converges *uniformly* (see Lemma 4.11 below) and we obtain that $\lim_n \langle (\varphi_\epsilon\otimes\psi_\epsilon)\theta_n, m\rangle = 0$ as m is continuous on $C_b(\mathcal{X}\times\mathcal{Y})$. Finally, for all $\epsilon > 0$, we have $0 \le \limsup_n \langle\theta_n, m\rangle \le 2\epsilon$. This completes the proof, since ϵ is arbitrary. □

In the above proof, the following lemma has been used.

Lemma 4.11. *Any decreasing sequence of $[0,\infty)$-valued upper semicontinuous functions on a Hausdorff compact set which converges pointwise to zero, also converges uniformly.*

Proof. Let (f_n) be a sequence of nonnegative upper semicontinuous functions on the compact set K such that for all $x \in K$, $f_n(x)$ decreases to zero as n tends to infinity. To work

with lower semicontinuous functions and epigraphs, let us consider $g_n = -f_n$. Then, for all n, $\mathrm{epi}\, g_n$ is a closed set of $K \times \mathbb{R}$. As, g_n increases to zero and g_1 attains its minimum value on the compact set K (see Theorem 2.12) we have $-\infty < \inf g_1 \le g_n(x)$ for all n and x.
For all $\epsilon > 0$, the sequence of compact subsets $\mathrm{epi}\, g_n \cap (K \times [\inf g_1, -\epsilon])$ decreases to the empty set: $\bigcap_n [\mathrm{epi}\, g_n \cap (K \times [\inf g_1, -\epsilon])] = \emptyset$. It follows from Proposition 2.1 that $\mathrm{epi}\, g_n \cap (K \times [\inf g_1, -\epsilon]) = \emptyset$ for all large enough n. But this means that $f_n < \epsilon$ for all large enough n. Since $0 \le f_n$, this completes the proof of the lemma. $\square$

Lemma 4.12. *Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are metric spaces and that c is a lower semicontinuous $[0, +\infty]$-valued function. Then, for all m in $\mathcal{E}(\mathcal{X} \times \mathcal{Y})$ acting as a probability measure on $\mathcal{X} \times \mathcal{Y}$, we have*

$$\mathcal{C}(m) = \int_{\mathcal{X} \times \mathcal{Y}} c(x, y)\, m(dxdy)$$

where the measure m in the right-hand side is the unique extension of $m \in \mathcal{E}(\mathcal{X} \times \mathcal{Y})$.

Note that a lower semicontinuous function f is Borel measurable since for all real α, $f^{-1}((-\infty, \alpha])$ is the level set $\{f \le \alpha\}$ which is closed (see Proposition 2.8). It follows that the integral $\int_{\mathcal{X} \times \mathcal{Y}} c(x, y)\, m(dxdy) \in [0, +\infty]$ is well-defined.

Proof. By Lemma 4.4, c is the limit of an increasing sequence (c_n) of continuous bounded functions. Let m be a nonnegative measure. By monotone convergence (Beppo-Levi theorem), we have $\int_{\mathcal{X} \times \mathcal{Y}} c\, dm = \sup_n \int_{\mathcal{X} \times \mathcal{Y}} c_n\, dm = \sup_n \langle c_n, m \rangle \le \mathcal{C}(m)$, where the inequality follows from $c_n \le c$ and the definition of $\mathcal{C}$.
On the other hand, we clearly have the converse inequality $\mathcal{C}(m) := \sup\{\int_{\mathcal{X} \times \mathcal{Y}} \tilde{c}\, dm; \tilde{c} \in C_b, \tilde{c} \le c\} \le \int_{\mathcal{X} \times \mathcal{Y}} c\, dm$. This completes the proof of the lemma. $\square$

4.3. **Dual equality.** We are now ready to apply Theorem 3.25 with

$$\begin{array}{ccccc} \Big\langle & P = C_b(\mathcal{X} \times \mathcal{Y}) & , & A = \mathcal{E}(\mathcal{X} \times \mathcal{Y}) & \Big\rangle \\ & T^* \ \uparrow & & \downarrow \ T & \\ \Big\langle & B = C_b(\mathcal{X}) \times C_b(\mathcal{Y}) & , & Q = \mathcal{E}(\mathcal{X}) \times \mathcal{E}(\mathcal{Y}) & \Big\rangle \end{array}$$

Keeping the notations of Section 3, the function $h(a)$ is $h(m) = \mathcal{C}(m) + \zeta_{\mathcal{E}_+}(m)$, the operator Ta is $Tm = (m_{\mathcal{X}}, m_{\mathcal{Y}})$ and the constraint set C is reduced to the point $\{(\mu, \nu)\}$. Since, $\mathcal{C}$ is the supremum linear continuous function, it is convex and lower semicontinuous. The cone $\mathcal{E}_+ = \bigcap_{\theta \in C_b; \theta \ge 0} \{m \in \mathcal{E}; \langle \theta, m \rangle \ge 0\}$ is convex closed, as the intersection of closed half-spaces. Therefore, h is convex and lower semicontinuous. This is assumption (A1) of Theorem 3.25. As, h is nonnegative, (A2) holds. As, C is a single point, (A3) holds.
Let us show (A4): $T^*(C_b(\mathcal{X}) \times C_b(\mathcal{Y})) \subset C_b(\mathcal{X} \times \mathcal{Y})$. By the very definition of T, we obtain for all $m \in \mathcal{E}(\mathcal{X} \times \mathcal{Y})$, $\varphi \in C_b(\mathcal{X})$ and $\psi \in C_b(\mathcal{Y})$,

$$\langle Tm, (\varphi, \psi) \rangle_{Q,B} = \langle m_{\mathcal{X}}, \varphi \rangle + \langle m_{\mathcal{Y}}, \psi \rangle = \langle m, \varphi \oplus \psi \rangle = \langle m, T^*(\varphi, \psi) \rangle.$$

Hence,

$$T^*(\varphi, \psi) = \varphi \oplus \psi \in C_b(\mathcal{X} \times \mathcal{Y})$$

where $\varphi \oplus \psi(x,y) := \varphi(x) + \psi(y)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$. This proves (A4).
Now, let us compute h^*. For any $\theta \in C_b(\mathcal{X} \times \mathcal{Y})$,

$$\begin{aligned} h^*(\theta) &= \sup_{m\in\mathcal{E}}\{\langle\theta,m\rangle - \mathcal{C}(m) - \zeta_{\mathcal{E}_+}(m)\} \\ &= \sup_{m\in\mathcal{E}_+}\{\langle\theta,m\rangle - \sup_{\tilde{c}\in C_b:\tilde{c}\le c}\langle\tilde{c},m\rangle\} \\ &= \sup_{m\in\mathcal{E}_+}\inf_{\tilde{c}\in C_b:\tilde{c}\le c}\langle\theta-\tilde{c},m\rangle \\ &= \begin{cases} 0 & \text{if } \forall m\in\mathcal{E}_+,\ \inf_{\tilde{c}\in C_b:\tilde{c}\le c}\langle\theta-\tilde{c},m\rangle \le 0 \\ +\infty & \text{if } \exists m\in\mathcal{E}_+,\ \inf_{\tilde{c}\in C_b:\tilde{c}\le c}\langle\theta-\tilde{c},m\rangle > 0 \end{cases} \\ &= \begin{cases} 0 & \text{if } \theta \le c \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

It appears that h^* is the convex indicator of $\{\theta \in C_b; \theta \le c\}$

$$h^*(\theta) = \zeta_{\{\theta\le c\}}$$

If $\inf c > 0$, $\theta = 0$ is in the topological interior of $\{\theta \in C_b; \theta \le c\}$ and (A5) holds. Otherwise, consider the new cost function $c_1 = c + 1$. Clearly, the relaxed problems $\mathcal{P}_c$ and $\mathcal{P}_{c_1}$ associated with c and c_1 admit the same minimizers and $\inf(\mathcal{P}_{c_1}) = \inf(\mathcal{P}_c) + 1$. We have checked all the requirements to apply Theorem 3.25, so that we have proved the dual equality $\inf\{h(a); a : Ta \in C\} = \sup_{b\in B}\{\inf_{q\in C}\langle b,q\rangle - h^*(T^*b)\} \in (-\infty,+\infty]$. Rewriting this result with the computed expressions for h, h^* and so on, and remembering that by Proposition 4.8: $\inf(\text{MK-P}) = \inf(\mathcal{P})$, we have proved the following theorem.

Theorem 4.13 (Kantorovich dual equality)**.** *Suppose that $\mathcal{X}$, $\mathcal{Y}$ are Polish spaces and c is a finite nonnegative lower semicontinuous function $\mathcal{X} \times \mathcal{Y}$. Then,*

$$\begin{aligned} &\inf\left\{\int_{\mathcal{X}\times\mathcal{Y}} c\,d\pi; \quad \pi \in P(\mu,\nu)\right\} \\ =\ &\sup\left\{\int_{\mathcal{X}}\varphi\,d\mu + \int_{\mathcal{Y}}\psi\,d\nu; \quad \varphi\in C_b(\mathcal{X}), \psi\in C_b(\mathcal{Y}) : \varphi\oplus\psi \le c\right\}. \end{aligned}$$

An interpretation of the Kantorovich duality. The effective cost for transporting a unit mass from x to y is $c(x,y)$.
I wish to transport the mass distribution $\mu \in \mathcal{P}(\mathcal{X})$ to $\nu \in \mathcal{P}(\mathcal{Y})$ and I ask a transport company to send me its price to do this job. It answers me that the price for taking a unit mass away from x is $\varphi(x)$ and for putting a unit mass down at y is $\psi(y)$. As these prices are such that $\varphi(x) + \psi(y) \le c(x,y)$ for all x, y, I find that this company is very competitive, I ask it to do the job and I pay $\int_{\mathcal{X}} \varphi(x)\,\mu(dx) + \int_{\mathcal{Y}} \psi(y)\,\nu(dy)$.
Since the transport company wants to maximize its income, it should have solved

$$\textit{maximize } (\varphi,\psi) \mapsto \int_{\mathcal{X}}\varphi(x)\,\mu(dx) + \int_{\mathcal{Y}}\psi(y)\,\nu(dy) \quad \textit{subject to} \quad \varphi\oplus\psi\le c \qquad (D)$$

On the other hand, the company also has to find the cheapest transport plan to minimize its expenditure. As the cost of a transport plan $\pi \in P(\mu,\nu)$ is $\int_{\mathcal{X}\times\mathcal{Y}} c\,d\pi$, the company should solve (MK-P) to find an optimal plan.
For any prices (φ,ψ) which are attractive in the sense that $\varphi\oplus\psi \le c$ and for any transport plan π from μ to ν, that is $\pi \in P(\mu,\nu)$, we have $\int_{\mathcal{X}}\varphi\,d\mu + \int_{\mathcal{Y}}\psi\,d\nu = \int_{\mathcal{X}\times\mathcal{Y}}\varphi\oplus\psi\,d\pi \le \int_{\mathcal{X}\times\mathcal{Y}} c\,d\pi$ which means: income$(\varphi,\psi) \le$ expenditure(π). On the other hand Kantorovich dual equality tells us that sup(D) = inf(MK-P) : the greatest possible income is equal to the lowest possible expenditure. In other words, unless the company acts optimally, it

looses money.
It will be proved below at Theorems 4.20 and 4.21, that under some assumptions there exist optimal plan and prices: π_* and (φ_*, ψ_*). Moreover it will be shown that although the prices are attractive: $\varphi_* \oplus \psi_* \le c$ everywhere, they are fair in the sense that $\varphi_* \oplus \psi_* = c$, π_*-almost everywhere.

4.4. **Dual attainment.** The dual attainment result is stated below at Theorem 4.20. Because of the monotonicity of

$$J(\varphi, \psi) := \int_{\mathcal{X}} \varphi \, d\mu + \int_{\mathcal{Y}} \psi \, d\nu$$

as a function of (φ, ψ), one can expect that the maximizers of J subject to $\varphi \oplus \psi \le c$ satisfy $\varphi \oplus \psi = c$ in some sense. As c is only supposed to be lower bounded and lower semicontinuous, one cannot expect that continuous bounded functions φ and ψ do the job. We are going to show that the dual attainment is achieved with $\varphi \in L_1(\mathcal{X}, \mu)$ and $\psi \in L_1(\mathcal{Y}, \nu)$. We define

$$\Phi_c := \{(\varphi, \psi) \in L_1(\mu) \times L_1(\nu) : \varphi \oplus \psi \le c\}.$$

More precisely, (φ, ψ) stands in Φ_c if there exist two Borel sets N_x and N_y such that $\mu(N_x) = 0$, $\nu(N_y) = 0$ and $\varphi(x) + \psi(y) \le c(x, y)$ for all $x \notin N_x$ and $y \notin N_y$. We say that this inequality holds (μ, ν)-almost everywhere to signify this.
In order to work at ease with negligible sets we consider the following versions of φ and ψ : $\tilde{\varphi}(x) = \varphi(x)$ if $x \notin N_x$, $\tilde{\varphi}(x) = -\infty$ if $x \in N_x$ and $\tilde{\psi}(y) = \psi(y)$ if $y \notin N_y$, $\tilde{\psi}(y) = -\infty$ if $y \in N_y$, so that $\tilde{\varphi} \oplus \tilde{\psi} \le c$ everywhere. From now on, the choice of these versions will be implicit and we drop the tilde.
The extended dual problem of interest is

$$\textit{maximize } J(\varphi, \psi) \quad \textit{subject to} \quad (\varphi, \psi) \in \Phi_c. \tag{MK-D}$$

The following proposition is a corollary of the Kantorovich dual equality. We denote $\Phi_c \cap C_b = \Phi_c \cap (C_b(\mathcal{X}) \times C_b(\mathcal{Y}))$.

Proposition 4.14. *Under the assumption of Theorem 4.13, we have*

$$\inf(\textit{MK-P}) := \inf_{\pi \in P(\mu,\nu)} \int_{\mathcal{X} \times \mathcal{Y}} c \, d\pi = \sup_{(\varphi,\psi) \in \Phi_c \cap C_b} J(\varphi, \psi) = \sup_{(\varphi,\psi) \in \Phi_c} J(\varphi, \psi) := \sup(\textit{MK-D}).$$

Proof. Let $(\varphi, \psi) \in \Phi_c$, then for all $\pi \in P(\mu, \nu)$ we have $\varphi \oplus \psi \le c$, π-almost everywhere. In fact, the undesirable set $(N_x \times \mathcal{Y}) \cup (\mathcal{X} \times N_y)$ is π-negligible since $0 \le \pi((N_x \times \mathcal{Y}) \cup (\mathcal{X} \times N_y)) \le \pi(N_x \times \mathcal{Y}) + \pi(\mathcal{X} \times N_y) = \mu(N_x) + \nu(N_y) = 0$. Hence, $J(\varphi, \psi) = \int \varphi \oplus \psi \, d\pi \le \int c \, d\pi$. Optimizing both sides of this inequality, one obtains

$$\sup_{(\varphi,\psi) \in \Phi_c} J(\varphi, \psi) \le \inf_{\pi \in P(\mu,\nu)} \int c \, d\pi.$$

This is the easy part of the dual equality: $\sup(\mathcal{D}) \le \inf(\mathcal{P})$, see Lemma 3.3-b.
The converse inequality follows from Theorem 4.13, since

$$\inf_{\pi \in P(\mu,\nu)} \int_{\mathcal{X} \times \mathcal{Y}} c \, d\pi = \sup_{(\varphi,\psi) \in \Phi_c \cap C_b} J(\varphi, \psi) \le \sup_{(\varphi,\psi) \in \Phi_c} J(\varphi, \psi).$$

□

c-conjugation. Let φ and ψ be functions on $\mathcal{X}$ and $\mathcal{Y}$ such that $\varphi(x)+\psi(y)\le c(x,y)$ for all x,y. The family of inequalities $\psi(y)\le c(x,y)-\varphi(x)$, for all x,y is equivalent to $\psi(y)\le \inf_x\{c(x,y)-\varphi(x)\}$ for all y. Therefore, the function

$$\varphi^c(y):=\inf_{x\in\mathcal{X}}\{c(x,y)-\varphi(x)\},\quad y\in\mathcal{Y}$$

satisfies $\varphi^c\ge\psi$ and $\varphi\oplus\varphi^c\le c$. As $J(\varphi,\psi)$ is an increasing function of its arguments φ and ψ, in view of maximizing J on Φ_c, the couple (φ,φ^c) is better than (φ,ψ). Performing this trick once again, we see that with

$$\psi^c(x):=\inf_{y\in\mathcal{Y}}\{c(x,y)-\psi(y)\},\quad x\in\mathcal{X},$$

the couple (φ^{cc},φ^c) is better than (φ,φ^c) and (φ,ψ). We have obtained the following result.

Lemma 4.15. *Let φ and ψ be $[-\infty,+\infty)$-valued functions on $\mathcal{X}$ and $\mathcal{Y}$. We assume that they are not identically $-\infty$ and $\varphi(x)+\psi(y)\le c(x,y)$ for all x,y. Then, φ^c and φ^{cc} are $[-\infty,+\infty)$-valued functions which are not identically $-\infty$. They also satisfy $\varphi^{cc}\ge\varphi,\varphi^c\ge\psi$ and $\varphi^{cc}(x)+\varphi^c(y)\le c(x,y)$ for all x,y.*

The operation $\varphi\to\varphi^c$ is very close to the concave conjugation defined at Definition 1.42, replacing formally $\langle x,y\rangle$ by $c(x,y)$. This is the reason why it is called c-conjugation.
Iterating the trick of Lemma 4.15 doesn't improve anything. Indeed, similarly to Proposition 1.31, we have $\varphi^{nc}=\varphi^{cc}$ if $n\ge 2$ is even and $\varphi^{nc}=\varphi^c$ if n is odd. To see this, it is enough to show that $\varphi^{ccc}=\varphi^c$. Let us show it. We have $\varphi^{ccc}=(\varphi^c)^{cc}\ge\varphi^c$ and the converse inequality holds since $\varphi^{cc}\ge\varphi$ implies that $\varphi^{ccc}(y)=(\varphi^{cc})^c(y)=\inf_x\{c(x,y)-\varphi^{cc}(x)\}\le\inf_x\{c(x,y)-\varphi(x)\}=\varphi^c(y)$, for all $y\in\mathcal{Y}$.
This indicates that a maximizer (φ_*,ψ_*) of the dual problem should satisfy $\varphi_*=(\psi_*)^c$ and $\psi_*=(\varphi_*)^c$ in some sense.
If c is a *continuous* cost function, it is easy to show that φ^c is measurable for all φ. Unfortunately, when c is only lower semicontinuous, this is not so easy to obtain. We shall restrict our attention to continuous φ's.

Lemma 4.16 (Measurability of φ^c)**.** *If c is a real valued bounded below continuous function, for any function φ, φ^c is a upper semicontinuous and therefore a* measurable *$[-\infty,+\infty)$-valued function.*
If c is a real valued bounded below lower semicontinuous function. For all upper semicontinuous real-valued function φ such that $\int_{\mathcal{X}}|\varphi|\,d\mu<\infty$, there exists a version $\tilde\varphi$ of φ ($\tilde\varphi$ is μ-almost everywhere equal to φ) such that $\tilde\varphi^c$ is a measurable *$[-\infty,+\infty)$-valued function.*

Note that when c is only lower semicontinuous, we do not prove the measurability of φ^c for all integrable φ : it is required that φ is upper semicontinuous. We are going to use Lemma 4.16 with φ continuous during the proof of Lemma 4.19 and with φ upper semicontinuous during the proof of Strassen's theorem (Proposition **??**).

Proof. If c is continuous, then φ^c is upper semicontinuous as the infimum of a collection of continuous functions. Hence, it is measurable.
Let us consider the general case where c is only lower semicontinuous. As μ is tight (it is a probability measure on a Polish space), there exists an increasing sequence (K_n) of compact subsets of $\mathcal{X}$ such that $\lim_n\mu(\mathcal{X}\setminus K_n)=0$. The set $N=\mathcal{X}\setminus(\cup_nK_n)$ is μ-negligible and we take for the version $\tilde\varphi$ of φ : $\tilde\varphi(x)=-\infty$ if $x\in N$ and $\tilde\varphi(x)=\varphi(x)$ otherwise. We have for all $y\in\mathcal{Y}$, $\tilde\varphi^c(y)=\inf_{x\in\mathcal{X}}\{c(x,y)-\tilde\varphi(x)\}=\inf_{x\in\cup_nK_n}\{c(x,y)-\varphi(x)\}=\lim_n\inf_{x\in K_n}\{c(x,y)-\varphi(x)\}$. In view of this result, it remains to show that φ^c is measurable

when $\mathcal{X}$ is a compact space.
Suppose that $\mathcal{X}$ is a compact space. Because of our assumptions on c, by Lemma 4.4 there exists an increasing sequence (c_n) of continuous functions such that c is the pointwise limit of (c_n). We have for all y, $\varphi^c(y) = \inf_x \sup_n f_n(x)$ with $f_n(x) = c_n(x,y) - \varphi(x)$. Since $c_n(\cdot, y)$ is continuous and φ is assumed to be upper semicontinuous, f_n is lower semicontinuous. As $\mathcal{X}$ is compact and (f_n) is increasing, thanks to Lemma 4.17 below, one can invert $\inf_x$ and $\sup_n$. This gives us $\varphi^c(y) = \sup_n \inf_x f_n(x) = \sup_n \inf_x \{c_n(x,y) - \varphi(x)\} = \lim_n \varphi^{c_n}(y)$. But c_n is continuous, so that φ^{c_n} is upper semicontinuous and a fortiori measurable. This completes the proof of the measurability of φ^c. □

During the proof of Lemma 4.16 we have used the following result.

Lemma 4.17. *Let (f_n) be an increasing sequence of lower semicontinuous functions on a compact Hausdorff space $\mathcal{X}$, then:* $\inf_{x\in\mathcal{X}} \sup_{n\geq 1} f_n(x) = \sup_{n\geq 1} \inf_{x\in\mathcal{X}} f_n(x)$.

Proof. As f_n is lower semicontinuous, by Proposition 2.11, we have for all x, $f_n(x) = \sup_{V\in\mathcal{N}(x)} \inf_{y\in V} f_n(y)$. It follows that

$$\sup_n f_n(x) = \sup_{V\in\mathcal{N}(x)} \sup_n \inf_{y\in V} f_n(y), \quad x \in \mathcal{X}. \tag{4.18}$$

By Theorem 2.12, for all n there exists $z_n \in \mathcal{X}$ such that $\inf_{y\in\mathcal{X}} f_n(y) = f_n(z_n)$. Let $(z_{n(k)})$ be a convergent subsequence with $\lim_k z_{n(k)} = z \in \mathcal{X}$. For any $V \in \mathcal{N}(z)$ and all large enough k, we have $z_{n(k)} \in V$ and

$$\inf_{y\in V} f_{n(k)}(y) = \inf_{y\in\mathcal{X}} f_{n(k)}(y).$$

As (f_n) is an increasing sequence, this also gives us

$$\sup_n \inf_{y\in V} f_n(y) = \sup_k \inf_{y\in V} f_{n(k)}(y) = \sup_k \inf_{y\in\mathcal{X}} f_{n(k)}(y) = \sup_n \inf_{y\in\mathcal{X}} f_n(y).$$

Taking the supremum over all $V \in \mathcal{N}(z)$, and making use of (4.18), one obtains

$$\sup_n f_n(z) = \sup_{V\in\mathcal{N}(z)} \sup_n \inf_{y\in V} f_n(y) = \sup_n \inf_{y\in\mathcal{X}} f_n(y).$$

It follows that $\inf_{x\in\mathcal{X}} \sup_n f_n(x) \leq \sup_n \inf_{x\in\mathcal{X}} f_n(x)$.
As the converse inequality holds trivially, this completes the proof of the lemma. □

To illustrate the importance of the assumption of lower semicontinuity in Lemma 4.17, we give an example. Let $\mathcal{X} = [0,1]$, (q_n) be an enumeration of the rational numbers in $[0,1]$ and $f_n(x) = \begin{cases} 0, & \text{if } x = q_k \text{ for some } k \geq n \\ 1, & \text{otherwise.} \end{cases}$. Then, (f_n) is an increasing sequence which converges pointwise to 1. It follows that, $1 = \inf_{x\in\mathcal{X}} \sup_{n\geq 1} f_n(x) > \sup_{n\geq 1} \inf_{x\in\mathcal{X}} f_n(x) = 0$. But, for all n the lower semicontinuous envelope of f_n is identically 0. This gives $\inf_{x\in\mathcal{X}} \sup_{n\geq 1} \mathrm{ls}\, f_n(x) = \sup_{n\geq 1} \inf_{x\in\mathcal{X}} \mathrm{ls}\, f_n(x) = 0$, in accordance with the lemma.

Lemma 4.19. *Let us assume that there exist two $[0,\infty)$-valued measurable functions $c_\mathcal{X}$ on $\mathcal{X}$ and $c_\mathcal{Y}$ on $\mathcal{Y}$ such that $c \leq c_\mathcal{X} \oplus c_\mathcal{Y}$, $\int_\mathcal{X} c_\mathcal{X}\, d\mu < \infty$ and $\int_\mathcal{Y} c_\mathcal{Y}\, d\nu < \infty$.*

- *(a) For all (φ,ψ) in $\Phi_c \cap C_b$ such that $J(\varphi,\psi) > -\infty$, there exists $(\bar\varphi,\bar\psi)$ in Φ_c such that $J(\bar\varphi,\bar\psi) \geq J(\varphi,\psi)$, $\bar\varphi \leq c_\mathcal{X}$ and $\bar\psi \leq c_\mathcal{Y}$.*
- *(b) If in addition c is continuous, for all (φ,ψ) in Φ_c such that $J(\varphi,\psi) > -\infty$, there exists $(\bar\varphi,\bar\psi)$ in Φ_c such that $J(\bar\varphi,\bar\psi) \geq J(\varphi,\psi)$, $\bar\varphi \leq c_\mathcal{X}$ and $\bar\psi \leq c_\mathcal{Y}$.*

In both cases, we can choose $\bar\varphi^c = \bar\psi$.

Proof. Let us first prove (a). As $J(\varphi,\psi) > -\infty$, there exists x_o such that $\varphi(x_o) > -\infty$. We have $\varphi^c(y) \le c(x_o, y) - \varphi(x_o)$, for all y. Consequently,

$$\begin{aligned} A &:= \sup_y \{\varphi^c(y) - c_{\mathcal{Y}}(y)\} \\ &\le \sup_y \{c_{\mathcal{X}}(x_o) + c_{\mathcal{Y}}(y) - \varphi(x_o) - c_{\mathcal{Y}}(y)\} \\ &= c_{\mathcal{X}}(x_o) - \varphi(x_o) \\ &< +\infty. \end{aligned}$$

We choose $(\bar\varphi, \bar\psi) = (\varphi + A, \varphi^c - A)$. Clearly:

- $J(\bar\varphi, \bar\psi) = J(\varphi, \varphi^c) \ge J(\varphi,\psi)$, by Lemma 4.15
- $\bar\varphi^c = \bar\psi$ so that $\bar\varphi \oplus \bar\psi \le c$.
- By construction, we also have $\bar\psi \le c_{\mathcal{Y}}$.

As $\bar\varphi \le \bar\varphi^{cc} = \bar\psi^c$, we have for all $x \in \mathcal{X}$,

$$\begin{aligned} \bar\varphi(x) - c_{\mathcal{X}}(x) &\le \bar\psi^c(x) - c_{\mathcal{X}}(x) \\ &= \inf_y \{c(x,y) - \bar\psi(y) - c_{\mathcal{X}}(x)\} \\ &\le \inf_y \{c_{\mathcal{Y}}(y) - \bar\psi(y)\} \\ &= -\sup_y \{\bar\psi(y) - c_{\mathcal{Y}}(y)\} \\ &= 0, \end{aligned}$$

that is $\bar\varphi \le c_{\mathcal{X}}$.
It remains to prove that $\bar\varphi \in L_1(\mu)$ and $\bar\psi \in L_1(\nu)$. This is clear for $\bar\varphi$, since it is continuous and bounded.
Let us show that $\bar\psi \in L_1(\nu)$. Thanks to Lemma 4.16, it is measurable. Let us consider the *nonpositive* function $\bar\psi - c_{\mathcal{Y}}$ so that the integral $\int_{\mathcal{Y}} (\bar\psi - c_{\mathcal{Y}})\, d\nu$ is meaningful in $[-\infty, 0]$
As $c_{\mathcal{X}}$ and $c_{\mathcal{Y}}$ are integrable and $J(\varphi,\psi) > -\infty$, we have

$$\begin{aligned} \int_{\mathcal{X}} (\bar\varphi - c_{\mathcal{X}})\, d\mu + \int_{\mathcal{Y}} (\bar\psi - c_{\mathcal{Y}})\, d\nu &= J(\bar\varphi, \bar\psi) - \int_{\mathcal{X}} c_{\mathcal{X}}\, d\mu - \int_{\mathcal{Y}} c_{\mathcal{Y}}\, d\nu \\ &\ge J(\varphi,\psi) - \int_{\mathcal{X}} c_{\mathcal{X}}\, d\mu - \int_{\mathcal{Y}} c_{\mathcal{Y}}\, d\nu \\ &> -\infty \end{aligned}$$

so that $\bar\psi - c_{\mathcal{Y}}$ is integrable and $\bar\psi \in L_1(\nu)$.
The proof of (b) follows exactly the same line. The only difference is that thanks to Lemma 4.16, for any φ, φ^c is measurable since c is assumed to be continuous. □

We are now ready to prove that a dual attainment result holds.

Theorem 4.20 (Dual attainment). *We assume that $\mathcal{X}$ and $\mathcal{Y}$ are Polish spaces and that the cost function c is $[0,+\infty)$-valued and lower semicontinuous. Let us take $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ and suppose that there exist two $[0,\infty)$-valued measurable functions $c_{\mathcal{X}}$ on $\mathcal{X}$ and $c_{\mathcal{Y}}$ on $\mathcal{Y}$ such that $c \le c_{\mathcal{X}} \oplus c_{\mathcal{Y}}$, $\int_{\mathcal{X}} c_{\mathcal{X}}\, d\mu < \infty$ and $\int_{\mathcal{Y}} c_{\mathcal{Y}}\, d\nu < \infty$.*
Then, the extended dual problem (MK-D) admits a solution (φ_*, ψ_*) in Φ_c.

Before giving the proof of this theorem, we derive the characterization of the optimal plans as an easy corollary.

Theorem 4.21 (Characterization of the optimal plans)**.** *Under the asumptions of Theorem 4.20, the primal and dual problems (MK-P) and (MK-D) both admit solutions. More, π_* is a solution of (MK-P) and (φ_*, ψ_*) is a solution of (MK-D) if and only if*

$$\begin{cases} \pi_* \in P(\mu,\nu), \ \int_{\mathcal{X}\times\mathcal{Y}} c\, d\pi_* < \infty, \\ \varphi_* \oplus \psi_* \leq c, \ \textit{everywhere and} \\ \varphi_* \oplus \psi_* = c, \ \pi_*\textit{-almost everywhere.} \end{cases} \tag{4.22}$$

This means that any optimal plan π_* is supported by $\{\varphi_* \oplus \psi_* = c\} \subset \mathcal{X} \times \mathcal{Y}$, where (φ_*, ψ_*) is any solution of (MK-D).

Proof of Theorem 4.21. The first statement is restatement of Theorem 4.2 and Theorem 4.20 which we admit for a while.
Let π_* and (φ_*, ψ_*) be solutions of (MK-P) and (MK-D). By Proposition 4.14, we have the dual equality: $\int_{\mathcal{X}} \varphi_*\, d\mu + \int_{\mathcal{Y}} \psi_*\, d\nu = \int_{\mathcal{X}\times\mathcal{Y}} c\, d\pi_*$. Therefore, $0 = \int_{\mathcal{X}\times\mathcal{Y}} (c - \varphi_* \oplus \psi_*)\, d\pi_*$. As $c - \varphi_* \oplus \psi_* \geq 0$, it follows that $c = \varphi_* \oplus \psi_*$, π_*-almost everywhere.
Conversely, let π_* and (φ_*, ψ_*) satisfy (4.22). Clearly,

$$\int_{\mathcal{X}} \varphi_*\, d\mu + \int_{\mathcal{Y}} \psi_*\, d\nu = \int_{\mathcal{X}\times\mathcal{Y}} c\, d\pi_*. \tag{4.23}$$

As $\int_{\mathcal{X}} \varphi\, d\mu + \int_{\mathcal{Y}} \psi\, d\nu \leq \int_{\mathcal{X}\times\mathcal{Y}} c\, d\pi$ for all $(\varphi,\psi) \in \Phi_c$ and all $\pi \in P(\mu,\nu)$, the equality (4.23) implies that π_* and (φ_*, ψ_*) respectively solve (MK-P) and (MK-D). □

Proof of Theorem 4.20. We need compactness. By means of a truncature procedure, we are going to be allowed to invoke Banach-Alaoglu theorem (see Theorem 2.40) in L_∞ : a bounded set in L_∞ is relatively $\sigma(L_\infty, L_1)$-compact.
Let $(\varphi_n, \psi_n)_{n\geq 1}$ be a maximizing sequence. By Proposition 4.14 one can take it in $\Phi_c \cap C_b$: For all n, $(\varphi_n, \psi_n) \in \Phi_c \cap C_b$ and $\lim_n J(\varphi_n, \psi_n) = \sup_{(\varphi,\psi)\in\Phi_c} J(\varphi,\psi)$.
As c is nonnegative, the dual equality stated at Proposition 4.14 implies that $\sup_{(\varphi,\psi)\in\Phi_c} J(\varphi,\psi) \geq 0$. It follows that one can take $J(\varphi_n, \psi_n) > -\infty$ for all n.
By Lemma 4.19, one can choose (φ_n, ψ_n) such that φ_n and ψ_n are measurable functions such that $\varphi_n \leq c_{\mathcal{X}}$ and $\psi_n \leq c_{\mathcal{Y}}$ for all n.
For all integer $k \geq 1$, let us put

$$\begin{aligned} f_n^k(x) &= \max(\varphi_n(x) - c_{\mathcal{X}}(x), -k), \ x \in \mathcal{X} \\ g_n^k(y) &= \max(\psi_n(y) - c_{\mathcal{Y}}(y), -k), \ y \in \mathcal{Y} \end{aligned}$$

This implies that

$$f_n^k(x) + g_n^k(y) \leq \max(c(x,y) - c_{\mathcal{X}}(x) - c_{\mathcal{Y}}(y), -k), \ x \in \mathcal{X}, y \in \mathcal{Y}. \tag{4.24}$$

We also have for all n and k

$$\begin{cases} -k \leq f_n^k \leq 0 \\ -k \leq g_n^k \leq 0 \end{cases}, \begin{cases} f_n^1 \geq f_n^2 \geq \cdots \geq f_n^k \geq \cdots \\ g_n^1 \geq g_n^2 \geq \cdots \geq g_n^k \geq \cdots \end{cases}, J(f_n^k, g_n^k) \geq J(\varphi_n, \psi_n) - J(c_{\mathcal{X}}, c_{\mathcal{Y}}). \tag{4.25}$$

For k fixed, the sequence $(f_n^k, g_n^k)_n$ is bounded in $L_\infty(\mu)\times L_\infty(\nu)$. Therefore, one can extract from it a $\sigma(L_\infty(\mu) \times L_\infty(\nu), L_1(\mu) \times L_1(\nu))$-convergent subsequence. By the diagonal subsequence trick, there exist two subsequences still denoted $(f_n^k)_n$ and $(g_n^k)_n$ such that $\lim_n f_n^k = f^k$ and $\lim_n g_n^k = g^k$ for all k.
As the $\sigma(L_\infty, L_1)$-convergence of a sequence implies its $\sigma(L_1, L_\infty)$-convergence (remember that μ and ν are bounded nonnegative measures so that $L_\infty \subset L_1$), we also have $\lim_n f_n^k = f^k$ and $\lim_n g_n^k = g^k$ for $\sigma(L_1, L_\infty)$, for all k. But by Theorem **??**, for all k there exists a sequence $(\breve{f}_n^k, \breve{g}_n^k)$ of convex combinations of (f_n^k, g_n^k) such that $\lim_n (\breve{f}_n^k, \breve{g}_n^k) = (f^k, g^k)$

where the limit is taken with respect to the strong topology of L_1. Once again extracting a subsequence, this limit also holds almost everywhere. The inequalities (4.24) and (4.25) are preserved by convex combinations (note that J is a *concave* function for the last inequality in (4.25)), so that they hold with $(\breve{f}^k_n, \breve{g}^k_n)$ instead of (f^k_n, g^k_n).
Writing (f^k_n, g^k_n) instead of $(\breve{f}^k_n, \breve{g}^k_n)$ not too overload notations, we have obtained the existence of (f^k_n, g^k_n) such that $\lim_n f^k_n = f^k$ and $\lim_n g^k_n = g^k$ strongly in L_1 and almost everywhere, for all k and such that (4.24) and (4.25) hold.
As pointwise convergence preserves the order, we also have

$$\begin{cases} f^1 \geq f^2 \geq \cdots \geq f^k \geq \cdots \\ g^1 \geq g^2 \geq \cdots \geq g^k \geq \cdots \end{cases}$$

(μ, ν)-almost everywhere.
Denoting the a.e-pointwise limits $f_* = \lim_k f^k = \inf_k f^k$ and $g_* = \lim_k g^k = \inf_k g^k$ and doing $\lim_k \lim_n$ in (4.24), we obtain

$$f_*(x) + g_*(y) \leq c(x, y) - c_{\mathcal{X}}(x) - c_{\mathcal{Y}}(y), \tag{4.26}$$

(μ, ν)-almost everywhere.
As J is continuous on L_1, it follows with (4.25) that for all k,

$$\begin{aligned} & J(f^k, g^k) = \lim_n J(f^k_n, g^k_n) \\ \geq \ & \limsup_n J(\varphi_n, \psi_n) - J(c_{\mathcal{X}}, c_{\mathcal{Y}}) = \sup_{(\varphi,\psi)\in\Phi_c} J(\varphi, \psi) - J(c_{\mathcal{X}}, c_{\mathcal{Y}}) > -\infty. \end{aligned} \tag{4.27}$$

Therefore, $\inf_k \int_{\mathcal{X}} f^k \, d\mu > -\infty$ and $\inf_k \int_{\mathcal{Y}} g^k \, d\nu > -\infty$ and by monotone convergence, we have $\int_{\mathcal{X}} f_* \, d\mu = \inf_k \int_{\mathcal{X}} f^k \, d\mu > -\infty$ and $\int_{\mathcal{Y}} g_* \, d\nu = \inf_k \int_{\mathcal{Y}} g^k \, d\nu > -\infty$. As f_* and g_* are nonpositive, this proves that f_* is in $L_1(\mu)$ and g_* is in $L_1(\nu)$.
Let us take $\varphi_* = f_* + c_{\mathcal{X}}$ and $\psi_* = g_* + c_{\mathcal{Y}}$. As $c_{\mathcal{X}}$ and $c_{\mathcal{Y}}$ are integrable, φ_* is in $L_1(\mu)$ and ψ_* is in $L_1(\nu)$. Noting that (4.26) is equivalent to $\varphi_* \oplus \psi_* \leq c$, (μ, ν)-almost everywhere, we have just proved that $(\varphi_*, \psi_*) \in \Phi_c$.
Finally, doing $\lim_k$ in (4.27), by monotone convergence, we have

$$J(\varphi_*, \psi_*) = J(f_*, g_*) + J(c_{\mathcal{X}}, c_{\mathcal{Y}}) \geq \sup_{(\varphi,\psi)\in\Phi_c} J(\varphi, \psi).$$

As $(\varphi_*, \psi_*) \in \Phi_c$, this implies that $J(\varphi_*, \psi_*) = \sup_{(\varphi,\psi)\in\Phi_c} J(\varphi, \psi)$ and completes the proof of the theorem. □

4.5. **Quadratic transport.** The quadratic transport corresponds to $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ and the quadratic cost function

$$c(x, y) = |x - y|^2/2$$

where $|x|$ is the usual Euclidean norm. It is assumed that the marginal measures μ and ν satisfy the following integrability condition

$$\int_{\mathbb{R}^d} |x|^2 \, \mu(dx) < \infty \quad \text{and} \quad \int_{\mathbb{R}^d} |y|^2 \, \nu(dy) < \infty. \tag{4.28}$$

Take any couple (φ, ψ) in Φ_c. Then, for all x, y in $\mathbb{R}^d$, $\varphi(x) + \psi(y) \leq |x - y|^2/2 = |x|^2/2 + |y|^2/2 - \langle x, y\rangle$, that is

$$\langle x, y\rangle \leq f(x) + g(y)$$

where

$$\begin{cases} f(x) &= |x|^2/2 - \varphi(x) \\ g(y) &= |y|^2/2 - \psi(y) \end{cases}$$

As $J(\varphi,\psi)=\int_{\mathbb{R}^d}|x|^2/2\,\mu(dx)+\int_{\mathbb{R}^d}|y|^2/2\,\nu(dy)-J(f,g)$, the dual problem in terms of (f,g) turns out to be

$$\textit{minimize } J(f,g) \quad \textit{subject to} \quad \langle x,y\rangle \le f(x)+g(y),\forall x,y$$

and it is easily seen that the improvement $J(\varphi,\psi)\le J(\varphi^{cc},\varphi^c)$ corresponds to

$$J(f^{**},f^*)\le J(f,g)$$

for all (f,g) in $\widetilde{\Phi}:=\{(f,g); f\in L_1(\mu), g\in L_1(\nu):\langle x,y\rangle\le f(x)+g(y),\forall x,y,(\mu,\nu)\text{-a.e.}\}$, where f^* and f^{**} are the convex conjugate and biconjugate of f.

Theorem 4.29 (Characterization of the optimal plans for the quadratic transport)**.** *The probability measure π_* on $\mathbb{R}^d\times\mathbb{R}^d$ is an optimal plan for the quadratic transport problem with marginals μ and ν satisfying (4.28) if and only if $\pi_*\in P(\mu,\nu)$ and there exists a closed convex function θ on $\mathbb{R}^d$ such that*

$$y\in\partial\theta(x) \textit{ for } \pi_*\textit{-almost every } (x,y).$$

Proof. By Theorem 4.20, there exists (f_o,g_o) in $\widetilde{\Phi}$ which solves the dual problem: $J(f_o,g_o)=\inf_{(f,g)\in\widetilde{\Phi}}J(f,g)$. With assumption (4.28), one can apply Lemma 4.19-b to assert that (f_o^{**},f_o^*) also solves the dual problem and is still in $\widetilde{\Phi}$. As a direct consequence of Theorem 4.21 we obtain with $\theta=f_o^{**}$ that $\pi_*\in P(\mu,\nu)$ is an optimal plan if and only if $\langle x,y\rangle\le\theta(x)+\theta^*(y)$ everywhere and $\langle x,y\rangle=\theta(x)+\theta^*(y)$ π_*-a.e. But the inequality is always true (Young's inequality) and the (Young's) equality is equivalent to $y\in\partial\theta(x)$. □

Remark 4.30. Clearly, this result still holds true when $(\mathcal{X},|\cdot|)$ is an Hilbert space.

One says that the closed convex function θ of Theorem 4.29 is a *Kantorovich potential* for the quadratic transport of μ on ν.
Consider the reverse transport problem of ν on μ. Denoting $R(x,y)=(y,x)$, $x,y\in\mathbb{R}^d$, as the quadratic cost is symmetric: $c=c\circ R$, it is immediate to see that π_* is an optimal plan for the direct problem if and only if $\pi_*\diamond R$ is an optimal plan for the reverse problem. As $y\in\partial\theta(x)$ is equivalent to $x\in\partial\theta^*(y)$ (see Proposition 1.36), one sees that $x\in\partial\theta^*(y)$ for π_*-almost every (x,y) or equivalently

$$y\in\partial\theta^*(x) \text{ for } \pi_*\diamond R\text{-almost every } (x,y).$$

In other words, θ is a Kantorovich potential for the direct quadratic transport problem if and only if its convex conjugate θ^* is a Kantorovich potential for the reverse problem.

4.6. **Kantorovich-Rubinstein theorem.** Let $\mathcal{X}=\mathcal{Y}$ be a Polish space. The cost function to be considered is $c(x,y)=d(x,y)$: a lower semicontinuous metric on $\mathcal{X}$ which may differ from the metric which turns $\mathcal{X}$ into a Polish space.
We denote φ^d and φ^{dd} the d-conjugate and d-biconjugate of φ.
In the sequel, the Lipschitz functions are to be considered with respect to the metric cost d and not with respect to the underlying metric on the Polish space $\mathcal{X}$. One writes that φ is d-Lipschitz(1) to specify that $|\varphi(x)-\varphi(y)|\le d(x,y)$ for all $x,y\in\mathcal{X}$.

Lemma 4.31. *For any function φ on $\mathcal{X}$,*

(a) φ^d is d-Lipschitz(1)
(b) $\varphi^{dd}=-\varphi^d$.
(c) If φ is continuous, then φ^d is measurable

Proof. (a) Since $y \mapsto d(x,y)$ is d-Lipschitz(1), by Lemma 4.7, $y \mapsto \varphi^d(y) = \inf_x\{d(x,y) - \varphi(x)\}$ is also d-Lipschitz(1).
(b) Hence for all x, y, $\varphi^d(y) - \varphi^d(x) \le d(x,y)$. But this implies that for all y, $-\varphi^d(x) \le d(x,y) - \varphi^d(y)$. Optimizing in y leads to $-\varphi^d(x) \le \varphi^{dd}(x)$.
On the other hand, $\varphi^{dd}(x) = \inf_y\{d(x,y) - \varphi^d(y)\} \le -\varphi^d(x)$ where the last inequality is obtained by taking $y = x$.
(c) is Lemma 4.16. □

Denote $\mathcal{P}_d := \{\mu \in \mathcal{P}_\mathcal{X}; \int_\mathcal{X} d(x_o, x)\, \mu(dx)\}$ where x_o is any fixed element in $\mathcal{X}$.
Let us denote

$$\|\varphi\|_{\mathrm{Lip}} = \sup_{x \neq y} \frac{|\varphi(y) - \varphi(x)|}{d(x,y)}.$$

It is the usual Lipschitz seminorm. Its dual norm gives for all μ, ν in $\mathcal{P}_d$

$$\|\mu - \nu\|^*_{\mathrm{Lip}} = \sup\left\{\int_\mathcal{X} \varphi(x)\,(\mu - \nu)(dx); \varphi \text{ measurable}, \|\varphi\|_{\mathrm{Lip}} \le 1\right\}.$$

As it is assumed that $\mu, \nu \in \mathcal{P}_d$, note that any measurable d-Lipschitz function is integrable with respect to μ and ν.
We also introduce a standard notation for the value of the transport problem

$$W_1(\mu,\nu) := \inf\left\{\int_{\mathcal{X}^2} d(x,y)\,\pi(dxdy); \pi \in P(\mu,\nu)\right\}, \quad \mu,\nu \in \mathcal{P}_d.$$

We are now ready to prove the main result of this section.

Theorem 4.32 (Kantorovich-Rubinstein)**.** *Let d be a lower semicontinuous metric on $\mathcal{X}$. Then, for all $\mu, \nu \in \mathcal{P}_d$*

$$\begin{aligned} W_1(\mu,\nu) &= \sup\left\{\int_\mathcal{X} \varphi(x)\,(\mu - \nu)(dx); \varphi \text{ measurable bounded}, \|\varphi\|_{\mathrm{Lip}} \le 1\right\} \\ &= \|\mu - \nu\|^*_{\mathrm{Lip}}. \end{aligned}$$

Proof. For all measurable d-Lipschitz(1) function φ and all π in $P(\mu,\nu)$, $\int_\mathcal{X} \varphi(x)\,(\mu - \nu)(dx) = \int_{\mathcal{X}^2} (\varphi(x) - \varphi(y))\,\pi(dxdy) \le \int_{\mathcal{X}^2} d(x,y)\,\pi(dxdy)$. Optimizing in φ and π one obtains $\|\mu - \nu\|^*_{\mathrm{Lip}} \le W_1(\mu,\nu)$.
With Kantorovich duality: $W_1(\mu,\nu) = \sup\left\{\int_\mathcal{X} \varphi\,d\mu + \int_\mathcal{X} \psi\,d\nu; (\varphi,\psi) \in \Phi_d \cap C_b\right\}$, and Lemma 4.31 we obtain that

$$\begin{aligned} W_1(\mu,\nu) &\le \sup\left\{\int_\mathcal{X} \varphi^{dd}\,d\mu + \int_\mathcal{X} \varphi^d\,d\nu; \varphi \text{ continuous bounded}\right\} \\ &\le \sup\left\{\int_\mathcal{X} \varphi\,d(\mu - \nu); \varphi \text{ measurable}, \|\varphi\|_{\mathrm{Lip}} \le 1\right\} \\ &= \|\mu - \nu\|^*_{\mathrm{Lip}}. \end{aligned}$$

This completes the proof of $W_1(\mu,\nu) = \|\mu - \nu\|^*_{\mathrm{Lip}}$. It remains to see that for any measurable d-Lipschitz(1) function φ and any $n \ge 1$, $\varphi_n := (-n) \vee \varphi \wedge n$ is bounded measurable d-Lipschitz(1) (see Lemma 4.7) and by the dominated convergence theorem: $\lim_n \int_\mathcal{X} \varphi_n\,d(\mu - \nu) = \int_\mathcal{X} \varphi\,d(\mu - \nu)$. □

References

[1] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 1968.

[2] N. Bourbaki. *Espaces vectoriels topologiques; Chapitres 1 à 5*. Masson, Paris, 1981.

[3] I. Ekeland and R. Témam. *Convex Analysis and Variational Problems*, volume 28 of *Classics in Applied Mathematics*. SIAM, 1999.

[4] J. Neveu. *Bases mathématiques du calcul des probabilités*. Masson, Paris, 1970.

[5] S. Rachev and L. Rüschendorf. *Mass Transportation Problems. Vol I : Theory, Vol. II : Applications*. Probability and its applications. Springer Verlag, New York, 1998.

[6] R.T. Rockafellar. *Conjugate Duality and Optimization*, volume 16 of *Regional Conferences Series in Applied Mathematics*. SIAM, Philadelphia, 1974.

[7] R.T. Rockafellar. *Convex Analysis*. Princeton landmarks in mathematics. Princeton University Press, Princeton, N.J., 1997. First published in the Princeton Mathematical Series in 1970.

[8] F. Trèves. *Topological Vector Spaces, Distributions and Kernels*, volume 25 of *Pure and Applied Mathematics*. Academic Press, 1967.

[9] C. Villani. *Topics in Optimal Transportation*. Graduate Studies in Mathematics 58. American Mathematical Society, Providence RI, 2003.

Modal-X, Université Paris 10. Bât. G, 200 av. de la République. 92001 Nanterre Cedex, France

CMAP, École Polytechnique. 91128 Palaiseau Cedex, France
E-mail address: `christian.leonard@polytechnique.fr`