

Monge-Kantorovich Problem

Warning

These notes have been written mainly to help me in the preparation of a seminar at NTNU, focused on the results contained in Chapter 2 of [19]. I tried to select the results to keep the seminar as self contained as possible, and I put here most of the tools needed for these results. As a byproduct, these notes could help people who attended my (long and not really clear) seminar to fill the gaps I left in the hurry of the seminar.

Notice that not all the material here comes from [19]. I also used [20] and, here and there, added different definitions or proofs to keep the consistency of the notations. Anyway, no proof, example or remark is original: I predated many well-written books and papers to collect the results I needed. Take a look at the Bibliography, if you want to know where my arguments come from.

In the future, I do not plan to update these pages, other than to correct small mistakes. This does not mean I claim these notes to be perfect: many more results could be added and many arguments could be improved. For sure, I would like to add a proof of Kantorovich duality as it is presented in Chapter 1 of [19] and few results contained in papers I was not able to access (e.g. [13]). But I do not know if it will ever happen.

A note to the reader: many preliminaries are piled up in the first Section before a correct statement of the main results could be given. Anyway, not all the preliminaries are needed to read (or to prove) each result. Namely,

- Theorem 2.1 requires only the preliminary results on the tightness of a set of measures;
- Theorem 2.3 and Theorem 2.5 require a few basic concepts from convex analysis to be stated and most of the preliminary results to be proved (but you can safely skip the section about c -cyclically monotonicity, c -convexity and c -concavity);
- Theorem 2.7 is the only result which needs generalized concepts of convexity, concavity and monotonicity (depending on the cost c).

Finally, a thanks to SAS and Trenitalia: without long and boring periods to spend waiting in airports and train stations during my travel to Bari, I would have never found the time to write most (30+ pages) of these notes in L^AT_EX.

1 Preliminaries

1.1 Statement of the problem

Let X, Y be Polish spaces (i.e. complete, separable metric spaces). We will denote with $\mathcal{P}(X)$ (resp. $\mathcal{P}(Y)$) the space of Borel probability measures on X (resp. Y).

For any couple of measures $(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y)$, we will define the set of measures with marginals μ and ν the following set:

$$\Pi(\mu, \nu) \doteq \left\{ \pi \in \mathcal{P}(X \times Y) ; \mu = \text{Proj}_\#^X \pi, \nu = \text{Proj}_\#^Y \pi \right\}, \quad (1)$$

where $\text{Proj}^X : X \times Y \rightarrow X$ and $\text{Proj}^Y : X \times Y \rightarrow Y$ are the canonical projections on the components. While a general definition of push-forward measures $\text{Proj}_\#^X \pi, \text{Proj}_\#^Y \pi$ will be given in Definition 1.10, at this point note that $\mu = \text{Proj}_\#^X \pi$ and $\nu = \text{Proj}_\#^Y \pi$ simply mean the following

$\pi(A \times Y) = \mu(A)$ and $\pi(X \times B) = \nu(B)$ for any $A \subset X, B \subset Y$ measurable sets.

Equivalently, $\pi \in \Pi(\mu, \nu)$ if and only if for any $(\varphi, \psi) \in \mathbf{L}^1(\mu) \times \mathbf{L}^1(\nu)$ we have

$$\int_{X \times Y} [\varphi(x) + \psi(y)] d\pi(x, y) = \int_X \varphi(x) d\mu(x) + \int_Y \psi(y) d\nu(y). \quad (2)$$

We are now ready to state Kantorovich optimal transportation problem. Let $c : X \times Y \rightarrow [0, +\infty]$ be a lower semicontinuous map. We want to find

$$\inf_{\pi \in \Pi(\mu, \nu)} I(\pi) \doteq \inf_{\pi \in \Pi(\mu, \nu)} \int_{X \times Y} c(x, y) d\pi(x, y). \quad (3)$$

Any minimizer of this problem will be called *optimal transportation plan* for c and possibly denoted as $\mathcal{T}_c(\mu, \nu)$.

As showed by Kantorovich duality theorem, this problem is equivalent to the following. Given c as above, set

$$\Phi_c \doteq \{(\varphi, \psi) \in \mathbf{L}^1(\mu) \times \mathbf{L}^1(\nu) ; \varphi(x) + \psi(y) \leq c(x, y) \text{ } \mu\text{-a.e. } x \in X, \nu\text{-a.e. } y \in Y\}.$$

Then, finding $\mathcal{T}_c(\mu, \nu)$ is the same as solving the problem

$$\sup_{(\varphi, \psi) \in \Phi_c} J(\varphi, \psi) \doteq \sup_{(\varphi, \psi) \in \Phi_c} \int_X \varphi(x) d\mu(x) + \int_Y \psi(y) d\nu(y). \quad (4)$$

Remark 1.1. Notice that we have assumed $c \geq 0$. All the results in the following could be immediately extended to the more general case of a cost satisfying a lower bound as the following one

$$c(x, y) \geq a(x) + b(y) \quad \forall (x, y) \in X \times Y, \quad (5)$$

for suitable upper semicontinuous functions $a \in \mathbf{L}^1(\mu), b \in \mathbf{L}^1(\nu)$. Indeed, we can always define

$$\bar{c}(x, y) \doteq c(x, y) - a(x) - b(y) \geq 0, \quad \Lambda \doteq \int_X a d\mu + \int_Y b d\nu < \infty, \quad (6)$$

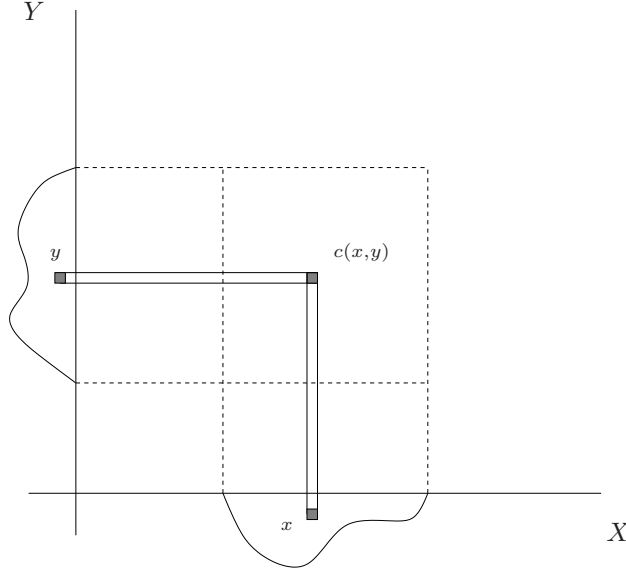


Figure 1: Transportation Problem

and, for all $(\varphi, \psi) \in \mathbf{L}^1(\mu) \times \mathbf{L}^1(\nu)$,

$$\overline{\varphi} \doteq \varphi - a, \quad \overline{\psi} \doteq \psi - b. \quad (7)$$

With these definitions, we can prove easily that:

c real valued & lower semicontinuous $\Rightarrow \bar{c}$ real valued & lower semicontinuous,

$$\begin{aligned} \varphi \in \mathbf{L}^1(\mu) &\Leftrightarrow \overline{\varphi} \in \mathbf{L}^1(\mu) & \psi \in \mathbf{L}^1(\nu) &\Leftrightarrow \overline{\psi} \in \mathbf{L}^1(\nu), \\ (\varphi, \psi) \in \Phi_c &\Leftrightarrow (\overline{\varphi}, \overline{\psi}) \in \Phi_{\bar{c}}, \end{aligned}$$

$$\forall \pi \in \Pi(\mu, \nu) \quad \int_{X \times Y} \bar{c} d\pi = \int_{X \times Y} c d\pi - \Lambda,$$

$$\forall (\varphi, \psi) \in \mathbf{L}^1(\mu) \times \mathbf{L}^1(\nu) \quad \int_X \overline{\varphi} d\mu + \int_Y \overline{\psi} d\nu = \int_X \varphi d\mu + \int_Y \psi d\nu - \Lambda.$$

Therefore, to any solution of the Kantorovich transportation problem (in both its formulations) with nonnegative cost there corresponds a solution of the Kantorovich transportation problem with cost bounded below as in (5).

1.2 Differentiability of Lipschitz continuous functions

Definition 1.2. Let $A \subset \mathbb{R}^n$.

- (i) A function $f: A \rightarrow \mathbb{R}^m$ is called Lipschitz continuous if there exists a constant C such that

$$|f(x) - f(y)| \leq C|x - y|, \quad (8)$$

for all $x, y \in A$. The smallest constant C such that (8) holds is denoted by $Lip(f)$. When $C = Lip(f)$ we will also say that f is C -Lipschitz continuous.

(ii) A function $f: A \rightarrow \mathbb{R}^m$ is called locally Lipschitz continuous if, for each compact $K \subset A$, there exists a constant C_K such that

$$|f(x) - f(y)| \leq C_K |x - y|, \quad (9)$$

for all $x, y \in K$.

Definition 1.3. A function $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ is differentiable at $x \in \mathbb{R}^n$ if there exists a linear mapping

$$L: \mathbb{R}^n \rightarrow \mathbb{R}^m,$$

such that

$$\lim_{y \rightarrow x} \frac{|f(y) - f(x) - L(y - x)|}{|x - y|} = 0. \quad (10)$$

Remark 1.4. If such a linear mapping L exists, it is unique. We call it the derivative of f at x and denote it with $Df(x)$.

Theorem 1.5. (Rademacher's Theorem) Let $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a locally Lipschitz continuous function. Then f is differentiable in $\mathcal{L}^n$ -a.e. $x \in \mathbb{R}^n$.

For a proof of Rademacher's Theorem, please refer to [7] (precisely, Theorem 2 at page 81)

1.3 Hahn–Banach theorems

Definition 1.6. Let X be a vector space over a field $\mathbb{F}$ and $\ell: X \rightarrow \mathbb{F}$ a function. ℓ is said to be a linear functional if it is linear, i.e. if for all $x, y \in X$ and for all $\lambda \in \mathbb{F}$

$$\ell(x + y) = \ell(x) + \ell(y), \quad \ell(\lambda x) = \lambda \ell(x) \quad (11)$$

Theorem 1.7. (Hahn–Banach Theorem) Let X be a vector space over $\mathbb{R}$ and $p: X \rightarrow \mathbb{R}$ such that

(i) p is positively homogeneous, i.e. for all $x \in X$ and for all $a > 0$

$$p(ax) = a p(x), \quad (12)$$

(ii) p is subadditive, i.e. for all $x, y \in X$

$$p(x + y) \leq p(x) + p(y). \quad (13)$$

If $Y \subset X$ is a vector subspace and $\ell: Y \rightarrow \mathbb{R}$ is a linear functional over Y such that for all $y \in Y$

$$\ell(y) \leq p(y), \quad (14)$$

then there exists a linear functional $\Lambda: X \rightarrow \mathbb{R}$ over X such that $\Lambda|_Y = \ell$ and for all $x \in X$

$$\Lambda(x) \leq p(x). \quad (15)$$

Theorem 1.8. (Hahn–Banach Separation Theorem) Let X be a normed vector space on $\mathbb{R}$, with dual X^* , and let $A, B \subset X$ be nonempty, convex and disjoint sets.

(i) If A is open, then there exist $\xi \in X^* \setminus \{0\}$ and $c \in \mathbb{R}$ such that for all $a \in A, b \in B$

$$\langle \xi, a \rangle \leq c \leq \langle \xi, b \rangle. \quad (16)$$

(ii) If A is closed and B compact, then there exist $\xi \in X^* \setminus \{0\}$ and $c \in \mathbb{R}$ such that for all $a \in A, b \in B$

$$\langle \xi, a \rangle < c < \langle \xi, b \rangle. \quad (17)$$

You can find a proof for Theorems 1.7–1.8 in many books, e.g. in [6] or [11].

1.4 Ingredients from measure theory

In the following, let $\mathcal{P}(X)$ be the set of Borel probability measures on X .

Definition 1.9. Let X be a metric space. Given a probability measure $\mu \in \mathcal{P}(X)$ and a set $S \subset X$, we say that μ is concentrated in S if $\mu(X \setminus S) = 0$. We also define the support of μ as the smallest (w.r.t. inclusion) closed set $\text{Supp}(\mu)$ on which μ is concentrated.

Definition 1.10. Let $(X, \Sigma_1), (Y, \Sigma_2)$ be measure spaces. Given a probability measure $\mu \in \mathcal{P}(X)$ and a μ -measurable map $f: X \rightarrow Y$, we define a probability measure in $\mathcal{P}(Y)$ as follows. For all $A \in \Sigma_2$ we set

$$f_{\#}\mu(A) \doteq \mu(f^{-1}(A)) = \mu(\{x \in X ; f(x) \in A\}). \quad (18)$$

We call $f_{\#}\mu$ the push-forward measure of μ through f . Measurability of f implies that $f_{\#}\mu$ is indeed a measure on (Y, Σ_2) .

Remark 1.11. Given $g: Y \rightarrow \mathbb{R}$ summable w.r.t. $f_{\#}\mu$, then $g \circ f$ is summable w.r.t. μ and we have the equality

$$\int_Y g(y) d(f_{\#}\mu)(y) = \int_X g(f(x)) d\mu(x).$$

Notice that this characterizes the probability measure $f_{\#}\mu$: if ν is a probability measure on Y such that for all $g \in \mathbf{L}^1(\nu)$ we have

$$\int_Y g(y) d\nu(y) = \int_X g(f(x)) d\mu(x),$$

then $\nu = f_{\#}\mu$. Indeed, for all $E \subset Y$ measurable, $\chi_E \in \mathbf{L}^1(\nu)$. Therefore, $\nu(E) = \int_Y \chi_E d\nu = \int_X \chi_E \circ f d\mu = \int_X \chi_{f^{-1}(E)} d\mu = \mu(f^{-1}(E))$.

Definition 1.12. Let X be a metric space. A subset $S \subset \mathcal{P}(X)$ is said to be tight if $\forall \varepsilon > 0$ there exists a compact set $K_\varepsilon \subset X$ such that

$$\mu(X \setminus K_\varepsilon) \leq \varepsilon \quad \forall \mu \in S. \quad (19)$$

Proposition 1.13. Let X, Y be Polish spaces and let $A \subset \mathcal{P}(X), B \subset \mathcal{P}(Y)$ be tight subsets of $\mathcal{P}(X)$ and $\mathcal{P}(Y)$ respectively. Then the set

$$\Pi(A, B) \doteq \{\pi \in \mathcal{P}(X \times Y) ; \pi \in \Pi(\mu, \nu) \text{ for some } \mu \in A, \nu \in B\}$$

is tight in $\mathcal{P}(X \times Y)$.

Proof. Let $\mu \in A$, $\nu \in B$ and $\varepsilon > 0$. By tightness of A, B , there exist compact sets $K_\varepsilon \subset X$ and $L_\varepsilon \subset Y$, independent by the choice of μ, ν , such that

$$\mu(X \setminus K_\varepsilon) \leq \frac{\varepsilon}{2}, \quad \nu(Y \setminus L_\varepsilon) \leq \frac{\varepsilon}{2}.$$

But then for all $\pi \in \Pi(A, B)$, $\pi \in \Pi(\mu, \nu)$ for suitable μ, ν and

$$\begin{aligned} \pi((X \times Y) \setminus (K_\varepsilon \times L_\varepsilon)) &\leq \pi((X \setminus K_\varepsilon) \times Y) + \pi(X \times (Y \setminus L_\varepsilon)) \\ &= \mu(X \setminus K_\varepsilon) + \nu(Y \setminus L_\varepsilon) \leq \varepsilon. \end{aligned}$$

This concludes the proof. $\diamond$

Theorem 1.14. (Prokhorov's Theorem) *Let X be a Polish space. A subset $S \subset \mathcal{P}(X)$ is relatively (sequentially) compact in $\mathcal{P}(X)$ if and only if S is tight.*

For a proof of this Theorem, please refer to [4].

Remark 1.15. In Prokhorov's Theorem above, relatively compact in $\mathcal{P}(X)$ means that the closure w.r.t. the topology $\sigma(\mathcal{P}(X), C_b(X))$ is compact in $\mathcal{P}(X)$.

Corollary 1.16. *Let X be a Polish space and $\mu \in \mathcal{P}(X)$. Then $\{\mu\}$ is a tight set in $\mathcal{P}(X)$.*

Remark 1.17. Actually there is no need to use Theorem 1.14 to prove tightness of $\{\mu\}$. Indeed, in Polish spaces, every probability measure μ is regular (see again [4] for a proof), i.e. for any Borel set A ,

$$\begin{aligned} \mu(A) &= \sup\{\mu(K) ; K \text{ compact}, K \subset A\} \\ &= \inf\{\mu(O) ; O \text{ open}, A \subset O\}. \end{aligned}$$

Therefore, since $\mu(X) = 1 < \infty$, we can always find a compact K whose measure is as close as we want to $\mu(X)$, and tightness follows.

Definition 1.18. *Let (X, Σ) be a measure space and let μ a positive measure on it. A family $\mathcal{F}$ in $\mathbf{L}^1(X, \mu)$ is said to be equiintegrable if the following conditions hold:*

(i) *for any $\varepsilon > 0$ there exists a μ -measurable set A with $\mu(A) < \infty$ such that*

$$\int_{X \setminus A} |f| d\mu < \varepsilon \quad \forall f \in \mathcal{F}, \quad (20)$$

(ii) *for any $\varepsilon > 0$ there exists $\delta > 0$ such that for any μ -measurable set E , $\mu(E) < \delta$ implies*

$$\int_E |f| d\mu < \varepsilon \quad \forall f \in \mathcal{F}. \quad (21)$$

Remark 1.19. Notice that, for finite measures (e.g. for probability measures), the first condition is always trivially satisfied by the choice $A = X$. Therefore, equiintegrability for a family $\mathcal{F} \subset \mathbf{L}^1(X, \mu)$, with μ finite, reduces to condition (ii) only.

Theorem 1.20. (Dunford–Pettis Theorem) *Let μ be a positive σ -finite measure on the measure space (X, Σ) and $\mathcal{F}$ be a bounded family in $\mathbf{L}^1(X, \mu)$. Then $\mathcal{F}$ is weakly (sequentially) compact if and only if $\mathcal{F}$ is equiintegrable.*

A rigorous proof of Theorem 1.20 would require some additional results from measure theory, therefore we only refer to [2] for a proof (precisely, Theorem 1.38 at page 18).

1.5 Ingredients from convex analysis

In this subsection, we will recall few results in convex analysis that we need in the following.

Definition 1.21. Let X be a vector space. A function $\varphi: X \rightarrow \mathbb{R} \cup \{+\infty\}$ is said to be a *convex function* if $\forall x, y \in X$ and $\forall t \in [0, 1]$ we have

$$\varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y). \quad (22)$$

Moreover, φ is said to be *strictly convex* if the equality in (22) implies either $x = y$ or $t \in \{0, 1\}$. Finally, φ is said to be *proper* if $\exists x \in X$ such that $\varphi(x) < \infty$. Equivalently, setting

$$\text{Dom}(\varphi) \doteq \{x \in X ; \varphi(x) < \infty\}, \quad (23)$$

φ is proper if $\text{Dom}(\varphi) \neq \emptyset$.

Remark 1.22. Whenever φ is convex, $\text{Dom}(\varphi)$ is a convex set. In particular, when $X = \mathbb{R}^n$, since any convex set C has ∂C with Hausdorff dimension $n-1$ ¹, $\partial(\text{Dom}(\varphi))$ is a set with finite $\mathcal{H}^{n-1}$ measure.

We also mention the next equivalent definitions of a convex function. You can prove them as an exercise or look for them in [12]–[17].

Proposition 1.23. Let $\varphi: X \rightarrow \mathbb{R} \cup \{+\infty\}$. Then, the following fact are equivalent:

- (i) φ is convex,
- (ii) the epigraph of φ , defined as follows

$$\text{epi}(\varphi) \doteq \{(x, \lambda) \in X \times \mathbb{R} ; \varphi(x) \leq \lambda\}, \quad (24)$$

is a convex set.

Moreover, if $X = \mathbb{R}$, both conditions are also equivalent to

- (iii) for all $x_o \in \mathbb{R}$ the map

$$x \mapsto \frac{\varphi(x) - \varphi(x_o)}{x - x_o},$$

is nondecreasing for all $x \neq x_o$.

Proposition 1.24. Let X be a normed space and $\varphi: X \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex map.

- (i) The following facts are equivalent:

- a. there exist $\mathcal{O} \subset X$, open and non-empty, and $\alpha \in \mathbb{R}$ such that

$$\varphi(x) \leq \alpha \quad \forall x \in \mathcal{O},$$

¹This should be a consequence of a result contained in [8], but I haven't tried to obtain it in detail.

b. φ is proper, $\text{Int}(\text{Dom}(\varphi)) \neq \emptyset$ and φ is continuous in $\text{Int}(\text{Dom}(\varphi))$.

(ii) If $x_o \in X$, $r_o > 0$, $m, M \in \mathbb{R}$ are such that

$$m \leq \varphi(x) \leq M \quad \forall x \in B(x_o, r_o),$$

then φ is Lipschitz continuous on $B(x_o, r)$, for all $0 < r < r_o$, with Lipschitz constant $\frac{M - m}{r_o - r}$.

Proof. (i) b. $\Rightarrow$ a. is clear. Let $x \in \text{Int}(\text{Dom}(\varphi))$ and $\alpha = 2\varphi(x)$. By continuity there exists $\mathcal{O} \subset \text{Int}(\text{Dom}(\varphi))$, open neighborhood of x , such that $\varphi \leq \alpha$ on $\mathcal{O}$. As for a. $\Rightarrow$ b. we want to prove that f is continuous at $x \in \text{Int}(\text{Dom}(\varphi))$. Fixed $v \in \mathcal{O}$, let $t > 0$ such that $x - tv \in \text{Int}(\text{Dom}(\varphi))$. Now consider the map

$$y \mapsto h(y) \doteq \frac{(t+1)y - x}{t} + v.$$

Notice that h is smooth and $h(x) = v$. Then $\mathcal{N} = h^{-1}(\mathcal{O})$ is a neighborhood of x . Let $y \in \mathcal{N}$. Hence,

$$y = \frac{t}{t+1} h(y) + \frac{1}{t+1} (x - tv)$$

and

$$\varphi(y) \leq \frac{t}{t+1} \alpha + \frac{1}{t+1} \varphi(x - tv) \leq \alpha + \varphi(x - tv).$$

Therefore, $\varphi \leq \beta \doteq \alpha + \varphi(x - tv)$ on $\mathcal{N}$. We claim that this is enough to have φ continuous in x .

Indeed, for any $\varepsilon > 0$ we want to find U neighborhood of x such that $|\varphi(\xi) - \varphi(x)| < \varepsilon$ whenever $\xi \in U$. Let V a symmetric neighborhood of 0 such that $x + V \subset \mathcal{N}$ and let $\eta \in \varepsilon' V$, $\varepsilon' = \frac{\varepsilon}{\beta}$. Then:

- $\frac{\eta}{\varepsilon'} \in V$, we can write

$$x + \eta = (1 - \varepsilon')x + \varepsilon' \left(x + \frac{\eta}{\varepsilon'} \right),$$

and there holds

$$\varphi(x + \eta) \leq (1 - \varepsilon')\varphi(x) + \varepsilon' \varphi \left(x + \frac{\eta}{\varepsilon'} \right) \leq \varphi(x) + \varepsilon' \beta = \varphi(x) + \varepsilon.$$

- $-\frac{\eta}{\varepsilon'} \in V$, we can write

$$x = \frac{\varepsilon'}{1 + \varepsilon'} \left(x - \frac{\eta}{\varepsilon'} \right) + \frac{1}{1 + \varepsilon'} (x + \eta),$$

and

$$\varphi(x) \leq \frac{\varepsilon'}{1 + \varepsilon'} \varphi \left(x - \frac{\eta}{\varepsilon'} \right) + \frac{1}{1 + \varepsilon'} \varphi(x + \eta),$$

which, finally, implies

$$\varphi(x + \eta) \geq \varphi(x) - \varepsilon' \varphi \left(x - \frac{\eta}{\varepsilon'} \right) \geq \varphi(x) - \varepsilon' \beta = \varphi(x) - \varepsilon.$$

These inequalities on $U = x + V$ imply the continuity of φ in x and the proof of (i) is concluded.

(ii) Let $x, y \in B(x_o, r)$ with $r < r_o$ and let $z \doteq y + \frac{r_o - r}{\|y - x\|_X} (y - x)$. Since $z \in B(x_o, r_o)$, we have $\varphi(z) \leq M$. Moreover,

$$y = \frac{\|y - x\|_X}{r_o - r + \|y - x\|_X} z + \frac{r_o - r}{r_o - r + \|y - x\|_X} x,$$

and hence

$$\begin{aligned} \varphi(y) - \varphi(x) &\leq \frac{\|y - x\|_X}{r_o - r + \|y - x\|_X} \varphi(z) + \left(\frac{r_o - r}{r_o - r + \|y - x\|_X} - 1 \right) \varphi(x) \\ &\leq \frac{\|y - x\|_X}{r_o - r + \|y - x\|_X} (M - m) \\ &\leq \frac{M - m}{r_o - r} \|y - x\|_X. \end{aligned}$$

Inverting the role of x, y you obtain that φ is Lipschitz on $B(x_o, r)$ and (ii) is proved. $\diamond$

Proposition 1.25. *Let X be a normed space and let $\varphi: X \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper and convex function. Then φ is locally Lipschitz continuous in $\text{Int}(\text{Dom}(\varphi))$.*

Proof. Let us start proving continuity of φ in $\text{Int}(\text{Dom}(\varphi))$. If $\bar{x} \in \text{Int}(\text{Dom}(\varphi))$, then we can put $\bar{x}$ in the interior of a simplex $\mathcal{S}$ and $\varphi(x)$ is bounded by the maximum value attained by φ on the vertices of $\mathcal{S}$.² Therefore, continuity follows by Proposition 1.24-(i) with $\mathcal{O} = \mathcal{S}$.

As for local Lipschitz continuity, we simply need to find also a lower bound to φ in a neighborhood of $\bar{x} \in \text{Int}(\text{Dom}(\varphi))$, e.g. on a simplex around $\bar{x}$, and then apply 1.24-(ii) to a ball centered in $\bar{x}$ and contained in the simplex.

A possible solution is to prove the existence of an affine function ℓ which satisfies $\varphi(x) \geq \ell(x)$ for all $x \in X$, by using Theorem 1.8 (Hahn–Banach separation theorem). Once this has been done, we can simply take the minimum of ℓ in the simplex to finish the proof. Anyway, a much shorter and smarter proof has been suggested to me by X. Raynaud. We already have a simplex $\mathcal{S}$ around $\bar{x}$ such that $\varphi \leq M$ on $\mathcal{S}$. Let us consider $B(\bar{x}, r_o) \subset \mathcal{S}$ and prove that φ is also bounded below in $B(\bar{x}, r_o)$.

Indeed, for all $y \in B(\bar{x}, r_o)$ there exists $z \in B(\bar{x}, r_o)$ such that $\bar{x} = \frac{y}{2} + \frac{z}{2}$.

Hence, $\varphi(\bar{x}) \leq \frac{\varphi(y)}{2} + \frac{\varphi(z)}{2}$ and

$$\varphi(y) \geq 2\varphi(\bar{x}) - \varphi(z) \geq m \doteq 2\varphi(\bar{x}) - M.$$

This concludes the proof. $\diamond$

²This result follows from the elementary remark that any convex function φ attains its maximum over a segment in one of the extremal points of the segment.

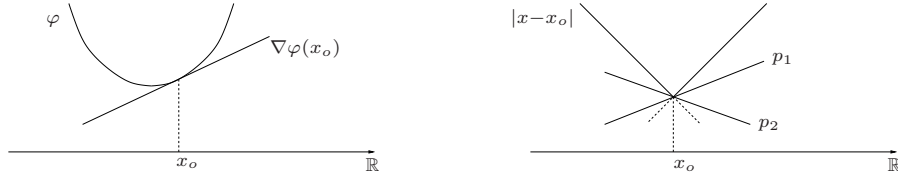


Figure 2: From the gradient to the subdifferential.

Remark 1.26. From the previous Proposition 1.25 and from Theorem 1.5 (Rademacher's Theorem), it follows that for $\mathcal{L}^n$ -a.e. x in $\text{Int}(\text{Dom}(\varphi))$ there exists $\nabla\varphi(x)$ and $\nabla\varphi$ is locally bounded.

Remark 1.27. If φ is differentiable in a point x , then we have

$$\varphi(z) \geq \varphi(x) + \nabla\varphi(x) \cdot (z - x) \quad \forall z \in \mathbb{R}^n.$$

From a geometrical point of view, this means that the graph of φ is always above its tangent hyperplane in x .

Motivated by Remark 1.27, we define the subdifferential of a convex function as the collection of all the hyperplanes such that the graph of φ is above them.

Definition 1.28. Let X be a topological vector space with dual X^* , $\varphi: X \rightarrow \mathbb{R} \cup \{+\infty\}$ be convex and $x \in X$. We define $\partial\varphi(x)$ as the set

$$\partial\varphi(x) \doteq \{\xi \in X^* ; \varphi(z) \geq \varphi(x) + \langle \xi, z - x \rangle \quad \forall z \in X\}. \quad (25)$$

The set $\partial\varphi(x)$ is called subdifferential of φ in x , and its elements are called subgradients. We also call subdifferential of φ the map $x \mapsto \partial\varphi(x)$, seen as a function between X and 2^{X^*} .

Remark 1.29. It follows immediately that

$$\bar{x} \text{ is a minimum for } \varphi \quad \Leftrightarrow \quad 0 \in \partial\varphi(\bar{x}).$$

Proposition 1.30. Let X be a topological vector space endowed with a norm $\|\cdot\|_X$. Then for all $x \in \text{Int}(\text{Dom}(\varphi))$ there holds $\partial\varphi(x) \neq \emptyset$.

Proof. Let $\bar{x} \in \text{Int}(\text{Dom}(\varphi))$. By Proposition 1.25, we know that φ is continuous in $\bar{x}$. Fix then $d \in X$ such that $\|d\|_X = 1$ and define on $\text{span}\{d\}$ the linear functional

$$\langle \ell_d, kd \rangle \doteq k \lim_{t \rightarrow 0^+} \frac{\varphi(\bar{x} + td) - \varphi(\bar{x})}{t} = k \inf_{t > 0} \frac{\varphi(\bar{x} + td) - \varphi(\bar{x})}{t}$$

for all $k \in \mathbb{R}$. This is a well defined linear functional due to convexity of φ . Notice also that $p: X \rightarrow \mathbb{R}$, defined as follows

$$p(v) \doteq \lim_{t \rightarrow 0^+} \frac{\varphi(\bar{x} + tv) - \varphi(\bar{x})}{t} = \inf_{t > 0} \frac{\varphi(\bar{x} + tv) - \varphi(\bar{x})}{t},$$

is subadditive and positively homogeneous. Indeed,

a. For $\lambda \geq 0$,

$$p(\lambda v) = \lim_{t \rightarrow 0^+} \frac{\varphi(x + t\lambda v) - \varphi(x)}{t} = \lambda \lim_{s \rightarrow 0^+} \frac{\varphi(x + sv) - \varphi(x)}{s} = \lambda p(v),$$

where we have set $s = \lambda t$.

b. For all v, w

$$\begin{aligned} p(v + w) &= \lim_{t \rightarrow 0^+} \frac{\varphi(x + tv + tw) - \varphi(x)}{t} \\ &= \lim_{t \rightarrow 0^+} \frac{\varphi\left(\frac{x + 2tv}{2} + \frac{x + 2tw}{2}\right) - \varphi(x)}{t} \\ &\leq \lim_{t \rightarrow 0^+} \frac{\varphi(x + 2tv) + \varphi(x + 2tw) - 2\varphi(x)}{2t} \\ &= \lim_{t \rightarrow 0^+} \frac{\varphi(x + 2tv) - \varphi(x)}{2t} + \lim_{t \rightarrow 0^+} \frac{\varphi(x + 2tw) - \varphi(x)}{2t} \\ &= \lim_{s \rightarrow 0^+} \frac{\varphi(x + sv) - \varphi(x)}{s} + \lim_{s \rightarrow 0^+} \frac{\varphi(x + sw) - \varphi(x)}{s} \\ &= p(v) + p(w), \end{aligned}$$

where we have used the convexity and set $s = 2t$ ³.

Therefore, by Theorem 1.7 (Hahn-Banach theorem), there exists a linear functional $\xi_d \in X^*$ such that $\langle \xi_d, kd \rangle = \langle \ell_d, kd \rangle$ and for all $v \in X$

$$\langle \xi_d, v \rangle \leq p(v) = \inf_{t > 0} \frac{\varphi(\bar{x} + tv) - \varphi(\bar{x})}{t}. \quad (26)$$

This means⁴ $\xi_d \in \partial\varphi(\bar{x})$. In particular, $\partial\varphi(\bar{x}) \neq \emptyset$. $\diamond$

Remark 1.31. Note that Proposition 1.30 holds *in any point*, not almost everywhere! Nevertheless, it is not that useful if $\text{Dom}(\varphi)$ has no interior points. In fact, it could be proved that $\partial\varphi(x)$ is non-empty if $x \in \mathcal{D} \subset \text{Dom}(\varphi)$, for a suitable dense subset $\mathcal{D}$, but such a proof would require much more sophisticated tools from convex analysis (e.g. approximated subdifferentials) which are far beyond the scope of these notes. Refer to [12] or [17] for extensive treatments of the topic.

Proposition 1.32. Let $\varphi: X \rightarrow \mathbb{R} \cup \{+\infty\}$ be convex and proper. Then

- (i) φ is (Gateaux) differentiable in $x \in X \Rightarrow \partial\varphi(x) = \{\nabla\varphi(x)\}$,
- (ii) φ continuous in $\bar{x} \in X$ and $\partial\varphi(\bar{x}) = \{\bar{\xi}\} \Rightarrow \varphi$ is (Gateaux) differentiable in $\bar{x} \in X$ and $\nabla\varphi(\bar{x}) = \bar{\xi}$.

³We have also used the fact that, since x is in the interior of the domain, $x + 2tv, x + 2tw$ belong to the interior of the domain as well (for t small enough). But this was implicitly used in the whole proof, since the definition of ℓ_d .

⁴Indeed, for all $z \in X$, choosing $t = 1$ and $v = z - \bar{x}$ in (26), we have $\varphi(z) \geq \varphi(\bar{x}) + \langle \xi_d, z - \bar{x} \rangle$.

Remark 1.33. Notice that the continuity of φ in x cannot be dropped in (ii). Indeed, it is possible to have a convex set A with empty interior such that in a point $\bar{x} \in A$ the normal cone reduces to $\{0\}$. Setting the *indicator function* of A as the following map

$$I_A(x) \doteq \begin{cases} 0 & \text{if } x \in A \\ +\infty & \text{if } x \notin A \end{cases}$$

we would have $\partial I_A(\bar{x}) = \{0\}$, but I_A is not differentiable in $\bar{x}$. Notice that I_A is convex if and only if A is convex and $\text{Dom}(I_A) = A$. Here the key assumption is $\text{Int}(A) = \text{Int}(\text{Dom}(I_A)) = \emptyset$. Indeed, one could prove that $\text{Int}(\text{Dom}(\varphi)) \neq \emptyset$ implies $\partial\varphi(\bar{x})$ not a singleton whenever φ is discontinuous at $\bar{x}$ (i.e. at boundary points of the domain, the subdifferential is either empty or unbounded). In particular, if $\text{Int}(\text{Dom}(\varphi)) \neq \emptyset$, then one has the nicer result

$$\varphi \text{ differentiable in } x \quad \Leftrightarrow \quad \partial\varphi \text{ is a singleton.}$$

We refer to [12] for a set A with the required properties and for a precise proof of the statement about subdifferentials in discontinuity points.

Proof. (i) First of all, notice that φ convex implies for all $v \in X$

$$t \mapsto \frac{\varphi(x + tv) - \varphi(x)}{t}$$

is an increasing function. Therefore, for any $t > 0$ we have

$$\frac{\varphi(x + tv) - \varphi(x)}{t} \geq \lim_{t \rightarrow 0^+} \frac{\varphi(x + tv) - \varphi(x)}{t} = \langle \nabla\varphi(x), v \rangle,$$

i.e.

$$\varphi(x + tv) \geq \varphi(x) + \langle \nabla\varphi(x), tv \rangle.$$

Since an identical estimate holds for $t < 0$, we conclude that $\nabla\varphi(x) \in \partial\varphi(x)$. If now $\xi \in \partial\varphi(x)$, for all $v \in X$ and $t > 0$

$$\varphi(x + tv) \geq \varphi(x) + \langle \xi, tv \rangle,$$

i.e.

$$\langle \xi, v \rangle \leq \inf_{t > 0} \frac{\varphi(x + tv) - \varphi(x)}{t} = \langle \nabla\varphi(x), v \rangle.$$

Since we can take in the previous inequality both v and $-v$, for any $v \in X$, this actually implies $\xi = \nabla\varphi(x)$.

(ii) Fix $d \in X$ such that $\|d\|_X = 1$ and proceed exactly as in the proof of Proposition 1.30, we define on $\text{span}\{d\}$ the linear functional

$$\langle \ell_d, kd \rangle \doteq k \lim_{t \rightarrow 0^+} \frac{\varphi(\bar{x} + td) - \varphi(\bar{x})}{t} = k \inf_{t > 0} \frac{\varphi(\bar{x} + td) - \varphi(\bar{x})}{t}$$

for all $k \in \mathbb{R}$. By Hahn-Banach theorem (Theorem 1.7 in these notes), there exists a linear functional $\xi_d \in X^*$ such that $\langle \xi_d, kd \rangle = \langle \ell_d, kd \rangle$ and for all $v \in X$

$$\langle \xi_d, v \rangle \leq \lim_{t \rightarrow 0^+} \frac{\varphi(\bar{x} + tv) - \varphi(\bar{x})}{t}.$$

This means $\xi_d \in \partial\varphi(\bar{x})$, and hence, $\xi_d = \bar{\xi}$. Therefore, we have for all $v \in X$

$$\langle \bar{\xi}, v \rangle = \langle \ell_w, \|v\|_X w \rangle = \|v\|_X \lim_{t \rightarrow 0^+} \frac{\varphi(\bar{x} + tw) - \varphi(\bar{x})}{t} = \lim_{s \rightarrow 0^+} \frac{\varphi(\bar{x} + sv) - \varphi(\bar{x})}{s}$$

where we set $w = \frac{v}{\|v\|_X}$ and, in the last equality, $s = \frac{t}{\|v\|_X}$. Finally, this means that φ is Gateaux differentiable and $\nabla\varphi(x) = \bar{\xi}$. $\diamond$

Definition 1.34. Let X be a locally convex Hausdorff space, X^* be its dual space and $\varphi: X \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper function. Then, for any $\xi \in X^*$ we define the Fenchel-Legendre conjugate of φ in ξ as

$$\varphi^*(\xi) \doteq \sup_{x \in X} \{\langle \xi, x \rangle - \varphi(x)\}. \quad (27)$$

Proposition 1.35. Let $\varphi: X \rightarrow \mathbb{R} \cup \{+\infty\}$ proper. Then $\varphi^*: X^* \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex and lower semicontinuous. Moreover, if φ is convex, then φ^* is also proper.

Proof. Convexity and lower semicontinuity follows from the fact that φ^* is pointwise supremum of affine functions.

To prove that φ^* is proper, we can apply Theorem 1.8 (Hahn–Banach separation theorem) to find an affine function $\langle \xi, x \rangle + r$ which supports φ in $\bar{x} \in \text{Dom}(\varphi)$. Indeed, let $\bar{x} \in \text{Dom}(\varphi)$ and $\alpha < \varphi(\bar{x})$. By Theorem 1.8, we can strictly separate with an hyperplane the compact set $\{(\bar{x}, \alpha)\} \subset X \times \mathbb{R}$ and the closed convex set $\text{epi}(\varphi) \subset X \times \mathbb{R}$, where $\text{epi}(\varphi)$ is the epigraph of φ as defined in (24). This means that there exist $\xi \in X^* \setminus \{0\}$ and $r, c \in \mathbb{R}$ such that for all $(x, \lambda) \in \text{epi}(\varphi)$ there holds

$$\langle \xi, x \rangle + r\lambda > c > \langle \xi, \bar{x} \rangle + \alpha. \quad (28)$$

In particular, for $(\bar{x}, \varphi(\bar{x})) \in \text{epi}(\varphi)$, we get $r(\varphi(\bar{x}) - \alpha) > 0$ and, hence, $r > 0$. This ensures that the hyperplane is not vertical ⁵ and it allows to set $\xi^* = -\frac{\xi}{r}$, $q = \frac{c}{r}$ and, finally,

$$\ell(x) \doteq \langle \xi^*, x \rangle + q.$$

$\varphi(x) \geq \ell(x)$ immediately follows from (28) ⁶.

This implies that $\varphi^*(\xi^*) \leq \ell^*(\xi^*) = -q < \infty$ and, hence, φ^* is proper. $\diamond$

Remark 1.36. Directly from Definition 1.34, it follows that $\forall x \in X$ and $\xi \in X^*$

$$\langle \xi, x \rangle \leq \varphi(x) + \varphi^*(\xi). \quad (29)$$

Proposition 1.37. Let $\varphi, \psi: X \rightarrow \mathbb{R} \cup \{+\infty\}$. Then:

(i) $\varphi \leq \psi$ implies $\varphi^* \geq \psi^*$,

⁵In the opposite case, the proof would have been a bit longer: we would have had to prove that we could slightly change our hyperplane with a non-vertical one, still maintaining the separation property.

⁶ $\ell(\bar{x}) > \alpha$ as well, but we don't need this here.

- (ii) $\varphi^{**} \leq \varphi$,
- (iii) $\varphi^{***} = \varphi^*$.

Proof. (i) It immediately follows from

$$\langle \xi, x \rangle - \varphi(x) \geq \langle \xi, x \rangle - \psi(x).$$

(ii) Notice that, using the definition of Fenchel conjugate, we have for all $x \in X$ and $\xi \in X^*$

$$\langle \xi, x \rangle - \varphi^*(\xi) \leq \varphi(x).$$

Hence, passing to the supremum over all ξ , we can conclude that $\varphi^{**} \leq \varphi$.

(iii) From (i) and (ii), it follows that $\varphi^{***} \geq \varphi^*$. On the other hand, by definition of φ^{**} , we have for all $x \in X$ and $\xi \in X^*$

$$\langle \xi, x \rangle - \varphi^{**}(x) \leq \varphi^*(\xi).$$

Hence, passing to the supremum over all x , we can conclude that $\varphi^{***} \leq \varphi^*$. $\diamond$

Proposition 1.38. *Let X be a normed space and let $\varphi: X \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper map. Then the following facts are equivalent:*

- (i) φ is convex and lower semicontinuous,
- (ii) $\varphi \equiv \psi^*$ for some proper function ψ ,
- (iii) $\varphi^{**} \equiv \varphi$.

Since the proof is quite long and requires a few concepts we do not need otherwise, it is deferred to the Appendix.

Proposition 1.39. *Let X be a normed space and $\varphi: X \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex, lower semicontinuous and proper map. Then, for all $x \in X$ and $\xi \in X^*$, there holds*

$$\xi \in \partial\varphi(x) \Leftrightarrow \langle \xi, x \rangle = \varphi(x) + \varphi^*(\xi) \Leftrightarrow x \in \partial\varphi^*(\xi). \quad (30)$$

Proof. Fixed $x \in X$ and $\xi \in X^*$, we have

$$\begin{aligned} \xi \in \partial\varphi(x) &\Leftrightarrow \forall z \in X, \quad \varphi(z) \geq \varphi(x) + \langle \xi, z - x \rangle \\ &\Leftrightarrow \forall z \in X, \quad \langle \xi, x \rangle \geq \varphi(x) + \langle \xi, z \rangle - \varphi(z) \\ &\Leftrightarrow \langle \xi, x \rangle \geq \varphi(x) + \varphi^*(\xi). \end{aligned}$$

Since the opposite inequality always holds true, as noticed in (29), we have proved half of the result. Applying now the same argument to φ^* and recalling that Proposition 1.38 implies $\varphi(x) = \varphi^{**}(x)$, we have

$$x \in \partial\varphi^*(\xi) \Leftrightarrow \langle x, \xi \rangle = \varphi^*(\xi) + \varphi^{**}(x) \Leftrightarrow \langle x, \xi \rangle = \varphi^*(\xi) + \varphi(x),$$

which concludes the proof. $\diamond$

Remark 1.40. Notice that, even if here we do not enter in the details, there is a sort of dual correspondence between strict convexity of φ and differentiability of φ^* , as Figure 3 tries to underline.

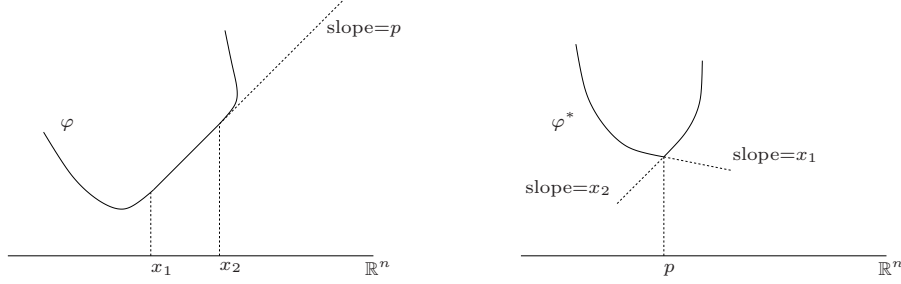


Figure 3: Loss of strict convexity in φ results in loss of regularity in φ^*

1.6 Monotone functions in $\mathbb{R}^n$

Definition 1.41. Let $u: \mathbb{R}^n \rightarrow 2^{\mathbb{R}^n}$ be a multifunction, i.e. u maps points $x \in \mathbb{R}^n$ to subsets $u(x) \subset \mathbb{R}^n$. Then:

- (i) for any fixed $B \subset \mathbb{R}^n$, u is univalued on B if $u(x)$ consists of at most one point for every $x \in B$,
- (ii) we define the domain of u as

$$\text{Dom}(u) \doteq \{x \in \mathbb{R}^n ; u(x) \neq \emptyset\} , \quad (31)$$

- (iii) we define the graph of u as

$$\text{Graph}(u) \doteq \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n ; y \in u(x)\} , \quad (32)$$

- (iv) u is monotone if

$$\langle y_1 - y_2, x_1 - x_2 \rangle \geq 0 \quad \forall x_i \in \mathbb{R}^n \quad \forall y_i \in u(x_i) \quad (i = 1, 2) , \quad (33)$$

- (v) for any given multifunctions u, v we say that

$$u \supset v ,$$

if $\text{Graph}(v) \subset \text{Graph}(u)$, i.e. if $v(x) \subset u(x)$ for all $x \in \mathbb{R}^n$,

- (v) u is maximal monotone if the following implication holds

$$v \supset u, v \text{ monotone} \Rightarrow v = u \quad (34)$$

in other words, u is maximal monotone when it is maximal w.r.t. the inclusion among the monotone functions.

Remark 1.42. For univalued functions in $\mathbb{R}$, condition (33) reduces to the usual monotonicity. Indeed, it is equivalent to

$$x_1 \leq x_2 \text{ and } y_1 \leq y_2 \quad \text{or} \quad x_1 \geq x_2 \text{ and } y_1 \geq y_2 .$$

Therefore, any monotone nondecreasing map in $\mathbb{R}$ is a monotone function in the sense of (33).

Example 1.43. An example of univalued monotone function on $\mathbb{R}^n$ is provided by the gradient $\nabla\varphi$ of a convex function. Whenever φ is differentiable at $x, y \in \mathbb{R}^n$ then

$$(\nabla\varphi(y) - \nabla\varphi(x)) \cdot (y - x) \geq 0. \quad (35)$$

Indeed, by Remark 1.27, for all x, y in which φ is differentiable there holds

$$\varphi(y) \geq \varphi(x) + \nabla\varphi(x) \cdot (y - x), \quad \varphi(x) \geq \varphi(y) + \nabla\varphi(y) \cdot (x - y).$$

It is now sufficient to sum up the two inequalities and cancel the terms with φ to obtain (35). Clearly, the same argument holds also for the subdifferential, whose definition was indeed modeled on the property in Remark 1.27. Anyway, we will come back later on the subdifferential and we will prove that it is in fact maximal monotone.

Remark 1.44. Let u be monotone. Then u is maximal monotone if and only if

$$\langle y - \eta, x - \xi \rangle \geq 0 \quad \forall (x, y) \in \text{Graph}(u) \quad \Rightarrow \quad (\xi, \eta) \in \text{Graph}(u). \quad (36)$$

Indeed, by contradiction if (36) does not hold for a fixed (ξ, η) , then we could define

$$v(x) \doteq \begin{cases} u(\xi) \cup \{\eta\} & \text{if } x = \xi, \\ u(x) & \text{otherwise.} \end{cases}$$

By properties of u and (36), v would be monotone and $v \supset u$, but $\eta \in v(\xi) \setminus u(\xi)$, violating maximal monotonicity of u .

On the other hand, take any $v: \mathbb{R}^n \rightarrow 2^{\mathbb{R}^n}$ monotone such that $v \supset u$ and fix $(\xi, \eta) \in \text{Graph}(v)$. By monotonicity, we would have

$$\langle y - \eta, x - \xi \rangle \geq 0 \quad \forall (x, y) \in \text{Graph}(v).$$

In particular, the same holds for all $(x, y) \in \text{Graph}(u) \subset \text{Graph}(v)$ and, by (36), $(\xi, \eta) \in \text{Graph}(u)$. Therefore, $\text{Graph}(v) \subset \text{Graph}(u)$, i.e. also $u \supset v$ and u is maximal.

We refer to [1] for a more complete development of the theory of maximal monotone functions and their properties. Here we only state a few results and their application to the multifunction $\partial\varphi$ for a convex map φ .

Proposition 1.45. *Let $u: \mathbb{R}^n \rightarrow 2^{\mathbb{R}^n}$ be maximal monotone. Then $\text{Graph}(u)$ is closed and $u(x)$ is a convex and closed (possibly empty) set for all $x \in \mathbb{R}^n$*

Proof. Given $(\xi_n, \eta_n) \in \text{Graph}(u)$ such that $(\xi_n, \eta_n) \rightarrow (\xi, \eta)$, we have by monotonicity

$$\langle \eta_n - y, \xi_n - x \rangle \geq 0 \quad \forall (x, y) \in \text{Graph}(u).$$

Passing to the limit, we have

$$\langle \eta - y, \xi - x \rangle \geq 0 \quad \forall (x, y) \in \text{Graph}(u),$$

and therefore, by (36), $(\xi, \eta) \in \text{Graph}(u)$. This proves that $\text{Graph}(u)$ is closed ⁷.

⁷Actually, to pass to the limit we would only need a sequence (ξ_n, η_n) that weakly converges. In other words, $\text{Graph}(u)$ is in fact weakly closed.

Now fix $x \in \mathbb{R}^n$. Set

$$S_\xi \doteq \{y \in \mathbb{R}^n ; \langle y - \eta, x - \xi \rangle \geq 0 \ \forall \eta \in u(\xi)\} ,$$

and

$$S \doteq \bigcap_{\xi \in \mathbb{R}^n} S_\xi .$$

Since all S_ξ are closed and convex, also S is closed and convex. We claim that $u(x) = S$. Indeed, by monotonicity, $u(x) \subset S$. On the other hand, $y \in S$ implies that (36) holds, and therefore $y \in u(x)$. $\diamond$

A remarkable property of maximal monotone functions is that there exists a one-to-one correspondence between graphs of maximal monotone functions and graphs of 1-Lipschitz continuous functions.

Proposition 1.46. *Define the Cayley transformation $\Psi: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n \times \mathbb{R}^n$ as follows:*

$$\Psi(x, y) \doteq \frac{1}{\sqrt{2}} (x + y, -x + y) . \quad (37)$$

Then the following facts hold:

- (i) *if u is a maximal monotone function, then $(u + \text{Id})^{-1}$ is defined on the whole $\mathbb{R}^n$ and $\Psi(\text{Graph}(u))$ is the graph of a 1-Lipschitz continuous function $\phi: \mathbb{R}^n \rightarrow \mathbb{R}^n$.⁸*
- (ii) *if $\phi: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a 1-Lipschitz continuous function, then $\Psi^{-1}(\text{Graph}(\phi))$ is the graph of a maximal monotone function on $\mathbb{R}^n$.*

In particular, u is maximal if and only if $(u + \text{Id})$ is surjective.

Another important property of maximal monotone functions is that the set of points x such that $u(x)$ is not univalued, is not too large.

Proposition 1.47. *Let u be a maximal monotone function and set*

$$\Xi_1(u) \doteq \{x \in \mathbb{R}^n ; \dim(u(x)) \geq 1\} ,$$

where $\dim(u(x))$ is the Hausdorff dimension of $u(x) \subset \mathbb{R}^n$. Then, $\dim(\Xi_1) \leq n - 1$, i.e. the Hausdorff dimension of Ξ_1 is at most $n - 1$. In particular, $\mathcal{H}^{n-1}(\Xi_1) < \infty$.

We do not give here a proof of the previous two results. You can find them in [1]. More precisely: Proposition 1.46 corresponds to Proposition 1.1 in [1]; Proposition 1.47 it is a particular case of Theorem 2.2 in [1] for $k = 1$.

Proposition 1.48. *Let φ be a proper convex function. Then $\partial\varphi$ is a maximal monotone function.*

Proof. Monotonicity follows from the definition of subdifferential. Indeed, for all $\xi_x \in \partial\varphi(x), \xi_y \in \partial\varphi(y)$ we have

$$\begin{aligned} \varphi(z) &\geq \varphi(x) + \langle \xi_x, z - x \rangle \\ \varphi(z) &\geq \varphi(y) + \langle \xi_y, z - y \rangle \end{aligned} \quad \forall z \in \mathbb{R}^n ,$$

⁸Namely, $z \mapsto \phi(z) \doteq z - \sqrt{2} (u + \text{Id})^{-1}(\sqrt{2}z)$, which can be proved to be univalued and 1-Lipschitz, but we do not need this explicit form in the following.

Choosing $z = y$ in the first, $z = x$ in the second and summing up, we have exactly

$$\langle \xi_x - \xi_y, x - y \rangle \geq 0.$$

As for being maximal, thanks to Proposition 1.46, we only need to prove that $\partial\varphi + \text{Id}$ is onto. Let $y \in \mathbb{R}^n$ and $\bar{x} \in \mathbb{R}^n$ be the unique minimizer of the coercive lower semicontinuous map

$$x \mapsto \varphi(x) + \frac{|x|^2}{2} - \langle y, x \rangle.$$

Then 0 is in the subdifferential at $\bar{x}$ of this map, i.e. for all $z \in \mathbb{R}^n$

$$\varphi(z) + \frac{|z|^2}{2} - \langle y, z \rangle \geq \varphi(\bar{x}) + \frac{|\bar{x}|^2}{2} - \langle y, \bar{x} \rangle,$$

or equivalently

$$\varphi(z) + \frac{|z|^2}{2} \geq \varphi(\bar{x}) + \frac{|\bar{x}|^2}{2} - \langle y, \bar{x} - z \rangle.$$

This in turn means $y \in \partial \left(\varphi(\bar{x}) + \frac{|\bar{x}|^2}{2} \right) = (\partial\varphi + \text{Id})(\bar{x})$. Therefore, $(\partial\varphi + \text{Id})$ is surjective. $\diamond$

Combining Proposition 1.48 with Propositions 1.45 and 1.47, one immediately obtains the following results.

Corollary 1.49. *Let φ be a proper convex function. Then $\text{Graph}(\partial\varphi)$ is a closed set.*

Corollary 1.50. *Let φ be a proper convex function. Then the set in which φ is not differentiable has finite $\mathcal{H}^{n-1}$ measure.*

1.7 Cyclical monotonicity, c -convexity and c -concavity

Definition 1.51. *Let X, Y be arbitrary spaces and $c: X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ be a function. A subset $\Gamma \subset X \times Y$ is said to be c -cyclically monotone if for all $n \in \mathbb{N}$ and for all families of points $(x_1, y_1), \dots, (x_n, y_n) \in \Gamma$ the following inequality holds:*

$$\sum_{i=1}^n c(x_i, y_i) \leq \sum_{i=1}^n c(x_i, y_{i+1}), \quad (38)$$

with the convention $y_{n+1} = y_1$. A c -cyclically monotone set Γ is said to be maximal if it is maximal w.r.t. the set inclusion.

Example 1.52. Let start with two trivial examples in $X = Y = \mathbb{R}^2$ and $c(x, y) = |x - y|^p$,⁹ just to learn how to handle Definition 1.51:

- $\Gamma_1 \doteq \{(0, 0), (1, 1), (2, 2)\}$ is a c -cyclically monotone set;
- $\Gamma_2 \doteq \{(0, 0), (1, 0), (0, 1), (1, 1)\}$ it is not a c -cyclically monotone set, indeed with $m = 2$ and $(1, 0), (0, 1)$, it fails to satisfy (38).

⁹A power $|x - y|^p$, $p > 1$, would work as well, but this is not really an interesting example anyway...

Remark 1.53. In particular, in the case $X = Y = \mathbb{R}^n$ and $c(x, y) = |x - y|^2$, a set $\Gamma \subset \mathbb{R}^n \times \mathbb{R}^n$ is c -cyclically monotone if for all $m \in \mathbb{N}$ and for all families $(x_1, y_1), \dots, (x_m, y_m) \in \Gamma$ there holds

$$\sum_{i=1}^m |x_i - y_i|^2 \leq \sum_{i=1}^m |x_i - y_{i+1}|^2, \quad (39)$$

with the convention $y_{m+1} = y_1$. Equivalently, we could require

$$\sum_{i=1}^m y_i \cdot (x_{i+1} - x_i) \leq 0, \quad (40)$$

with the convention $x_{m+1} = x_1$.

Theorem 1.54. (Rockafellar's Theorem) *A nonempty set $\Gamma \subset \mathbb{R}^n \times \mathbb{R}^n$ is cyclically monotone (w.r.t. $c(x, y) = |x - y|^2$) if and only if it is included in the graph of the subdifferential of a proper, lower semicontinuous and convex function $\varphi: \mathbb{R} \cup \{+\infty\}$. Moreover, every maximal cyclically monotone set coincides with such a graph.*

Proof. a. Let φ be a proper, lower semicontinuous and convex function. We prove that $\text{Graph}(\partial\varphi)$ is a cyclically monotone set. In particular, this implies that its subsets are cyclically monotone. Let $m \in \mathbb{N}$ and $(x_1, y_1), \dots, (x_m, y_m) \in \text{Graph}(\partial\varphi)$, i.e. for all $i = 1, \dots, m$ $y_i \in \partial\varphi(x_i)$. By Definition 25, this means that for all $z \in \mathbb{R}^n$ and for all i we have

$$\varphi(z) \geq \varphi(x_i) + y_i \cdot (z - x_i).$$

In particular, choosing $z = x_{i+1}$,

$$\begin{aligned} \varphi(x_2) &\geq \varphi(x_1) + y_1 \cdot (x_2 - x_1), \\ \varphi(x_3) &\geq \varphi(x_2) + y_2 \cdot (x_3 - x_2), \\ &\vdots \\ \varphi(x_1) &\geq \varphi(x_m) + y_m \cdot (x_1 - x_m). \end{aligned}$$

Adding these inequalities, we recover (40) ¹⁰.

b. Let Γ be a nonempty cyclically monotone subset of $\mathbb{R}^n \times \mathbb{R}^n$. Fixed $(x_o, y_o) \in \Gamma$, we define $\varphi: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ as

$$\varphi(x) \doteq \sup \{ y_m \cdot (x - x_m) + \dots + y_o \cdot (x_1 - x_o) ; \\ m \in \mathbb{N}, (x_1, y_1), \dots, (x_m, y_m) \in \Gamma \}. \quad (41)$$

Since φ is the supremum of affine functions, it is lower semicontinuous and convex. Moreover, since Γ is cyclically monotone, we have that $\varphi(x_o)$ is the

¹⁰Notice that this part of the proof uses exactly the same elementary idea used in Example 1.43 to prove that the graph of the gradient of a convex function is a monotone set (and also the graph of its subdifferential).

supremum of sums of the form in (40), i.e. is the supremum of nonpositive terms. Therefore, $\varphi(x_o) \leq 0$ ¹¹ and φ is proper¹².

c. To prove that $\Gamma \subset \text{Graph}(\partial\varphi)$, for φ defined above, we need to prove that for all $(\bar{x}, \bar{y}) \in \Gamma$ there holds

$$\forall z \in \mathbb{R}^n \quad \varphi(z) \geq \varphi(\bar{x}) + \bar{y} \cdot (z - \bar{x}).$$

Given $z \in \mathbb{R}^n$ and $\alpha < \varphi(\bar{x})$, there exist m and $(x_1, y_1), \dots, (x_m, y_m)$ in Γ such that

$$\alpha \leq y_m \cdot (\bar{x} - x_m) + \dots + y_o \cdot (x_1 - x_o).$$

Thus,

$$\alpha + y \cdot (z - \bar{x}) \leq y \cdot (z - \bar{x}) + y_m \cdot (\bar{x} - x_m) + \dots + y_o \cdot (x_1 - x_o),$$

where now the right hand side is one of the terms which would appear in the definition of $\varphi(z)$. Therefore,

$$\alpha + y \cdot (z - \bar{x}) \leq \varphi(z),$$

and we can conclude by arbitrariness of $\alpha < \varphi(\bar{x})$.

d. Concerning maximality, let Γ be a maximal cyclically monotone set. Then it is contained (by c.) in the $\text{Graph}(\partial\varphi)$ for a suitable φ proper, convex and lower semicontinuous. But then, since Γ is maximal, it have to coincide with it. $\diamond$

Definition 1.55. Let X, Y be sets and $c: X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ a function. Then:

(i) $\varphi: X \rightarrow \mathbb{R} \cup \{+\infty\}$ is said to be a proper c -convex function, if $\varphi \not\equiv +\infty$ and there exists $\zeta: Y \rightarrow \mathbb{R} \cup \{\pm\infty\}$ such that

$$\forall x \in X \quad \varphi(x) = \sup_{y \in Y} (\zeta(y) - c(x, y)), \quad (42)$$

with the convention that $(+\infty) - (+\infty) = -\infty$.

(ii) $\psi: Y \rightarrow \mathbb{R} \cup \{-\infty\}$ is said to be a proper c -concave function, if $\psi \not\equiv -\infty$ and there exists $\zeta: X \rightarrow \mathbb{R} \cup \{\pm\infty\}$ such that

$$\forall y \in Y \quad \psi(y) = \inf_{x \in X} (\zeta(x) + c(x, y)), \quad (43)$$

with the convention that $(-\infty) + (+\infty) = +\infty$.

Remark 1.56. Roughly speaking, a c -convex function is a function whose graph can be “caressed” from below with an object having the shape of $-c$ (this shape might of course vary when you change point). Figure 4 could help to visualize the situation.

¹¹Actually $\varphi(x_o) = 0$, since you can take $m = 1$ and $(x_1, y_1) = (x_o, y_o)$ among the terms in the supremum.

¹²This is the only part which requires the cyclically monotonicity of Γ .

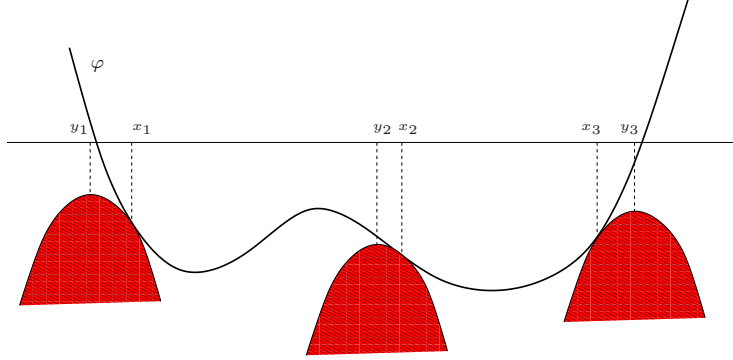


Figure 4: a c -convex φ in the case of a convex cost c (the “tools” supporting the graph of φ have the shape of $-c$) and a few points such that $y_i \in \partial_c \varphi(x_i)$.

Remark 1.57. Notice that, if $c(x, y) = -x \cdot y$, then c -convexity coincides with usual convexity plus lower semicontinuity (indeed, we would have φ c -convex if and only if $\varphi = (-\zeta)^*$ for some ζ and the equivalence would follow from Proposition 1.38).

Definition 1.58. Let X, Y be sets, $c: X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ be a function and $\varphi: X \rightarrow \mathbb{R} \cup \{+\infty\}$ be c -convex. Then:

(i) we define the c -transform of φ as follows

$$\varphi^c \doteq \begin{cases} Y & \rightarrow \mathbb{R} \cup \{-\infty\} \\ y & \mapsto \inf_{x \in X} (\varphi(x) + c(x, y)) \end{cases}$$

(ii) we define the c -subdifferential of φ as follows

$$\partial_c \varphi \doteq \{(x, y) \in X \times Y ; \varphi^c(y) - \varphi(x) = c(x, y)\} ,$$

(iii) we define, for all $x \in X$ the c -subdifferential in x of φ as follows

$$\partial_c \varphi(x) \doteq \{y \in Y ; (x, y) \in \partial_c \varphi\} ,$$

or, equivalently ¹³, the set of $y \in Y$ such that for all $z \in X$ there holds

$$\varphi(x) + c(x, y) \leq \varphi(z) + c(z, y) .$$

Remark 1.59. It is immediate to notice that φ^c is c -concave, whenever φ is c -convex ¹⁴. This behavior is different from the usual one with convex functions and their Fenchel-Legendre transforms. Namely, if $c(x, y) = -x \cdot y$ and φ convex, then one always has $\varphi^c = -\varphi^*$.

¹³Notice that $(x, y) \in \partial_c \varphi$ if and only if

$$\varphi(x) + c(x, y) = \varphi^c(y) = \inf_{z \in X} \varphi(z) + c(z, y) .$$

¹⁴Even more: the definition of c -concave functions can be restated as $\psi = \zeta^c$ for some ζ .

Definition 1.60. Let X, Y be sets, $c: X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ be a function and $\psi: Y \rightarrow \mathbb{R} \cup \{-\infty\}$ be c -concave. Then:

(i) we define the c -transform of ψ as follows

$$\psi^c \doteq \begin{cases} X & \rightarrow \mathbb{R} \cup \{+\infty\} \\ x & \mapsto \sup_{y \in Y} (\psi(y) - c(x, y)) \end{cases}$$

(ii) we define the c -superdifferential of ψ as follows

$$\partial^c \psi \doteq \{(x, y) \in X \times Y ; \psi(y) - \psi^c(x) = c(x, y)\} ,$$

(iii) we define, for all $y \in Y$ the c -superdifferential in y of ψ as follows

$$\partial^c \psi(y) \doteq \{x \in X ; (x, y) \in \partial^c \psi\} ,$$

or, equivalently¹⁵, the set of $x \in X$ such that for all $w \in Y$ there holds

$$\psi(y) - c(x, y) \geq \psi(w) + c(x, w) .$$

Remark 1.61. Notice that both the c -subdifferentials $\partial_c \varphi$ and the superdifferentials $\partial^c \psi$ are c -cyclically monotone sets.

Remark 1.62. We want to stress that, in both Definition 1.55, Definition 1.58 and Definition 1.60, we never assumed any measurability and c is any function defined on the product $X \times Y$. Hence, a priori, measurability of φ, ψ or φ^c, ψ^c is not guaranteed. Assuming c continuous, we would be able to say more: a c -convex (resp. c -concave) function is lower semicontinuous (resp. upper semicontinuous) and its c -subdifferential (resp. c -superdifferential) is a closed set.

Despite of the loss of symmetry between X and Y induced by using infima or suprema depending on the c -convexity or c -concavity, and of the differences with standard convexity (as underlined in Remark 1.59), previous definitions allow to recover some of the properties of Fenchel–Legendre transform for convex functions, as the following properties show.

Lemma 1.63. Let X, Y be sets and $c: X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$, $\varphi: X \rightarrow \mathbb{R} \cup \{+\infty\}$ and $\psi: Y \rightarrow \mathbb{R} \cup \{-\infty\}$ be functions. Then

$$\varphi^{ccc} \doteq ((\varphi^c)^c)^c = \varphi^c \quad \text{and} \quad \psi^{ccc} \doteq ((\psi^c)^c)^c = \psi^c .$$

Proof. We prove the equality only for φ . A completely similar argument can be used to prove the equality for ψ . Recall that for all $y \in Y$

$$\varphi^{ccc}(y) = \inf_{z \in X} \sup_{w \in Y} \inf_{x \in X} (\varphi(x) + c(x, w) - c(z, w) + c(z, y)) .$$

Hence,

¹⁵As in the previous case, notice that $(x, y) \in \partial^c \psi$ if and only if

$$\psi(y) - c(x, y) = \psi^c(x) = \sup_{w \in Y} \psi(w) - c(x, w) .$$

- with $w = y$ we obtain

$$\varphi^{ccc}(y) \geq \inf_{x \in X} (\varphi(x) + c(x, y)) = \varphi^c(y),$$

- with $x = z$ we obtain

$$\varphi^{ccc}(y) \leq \inf_{z \in X} \sup_{w \in Y} (\varphi(z) + c(z, y)) = \varphi^c(y).$$

This concludes the proof. $\diamond$

Proposition 1.64. *Let X, Y be sets and $c: X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$, $\varphi: X \rightarrow \mathbb{R} \cup \{+\infty\}$ and $\psi: Y \rightarrow \mathbb{R} \cup \{-\infty\}$ be functions. Then:*

- (i) φ is c -convex if and only if $\varphi = \varphi^{cc} \doteq (\varphi^c)^c$,
- (ii) ψ is c -concave if and only if $\psi = \psi^{cc} \doteq (\psi^c)^c$.

Proof. We prove the equivalence only for φ , being the proof for ψ entirely similar.

$\Leftarrow$ If $\varphi = \varphi^{cc}$, in particular φ is the c -transform of φ^c , therefore it is c -convex with the choice $\zeta = \varphi^c$.

$\Rightarrow$ If φ is c -convex, there exists ζ such that $\varphi = \zeta^c$. Hence, by Lemma 1.63, $\varphi^{cc} = \zeta^{ccc} = \zeta^c = \varphi$.

This concludes the proof. $\diamond$

2 Main Results

Theorem 2.1. *Let $(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y)$. Then there exists a minimizer π_* of the problem, i.e. an element in $\Pi(\mu, \nu)$ such that*

$$I(\pi_*) = \inf_{\Pi(\mu, \nu)} I(\pi).$$

Remark 2.2. For general measures μ, ν , one cannot expect any uniqueness result. Indeed, consider $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$ such that μ is concentrated in $\{(0, 0), (1, 1)\}$ and ν is concentrated in $\{(1, 0), (0, 1)\}$. In this situation it is clear that both the plans

$$\Pi_1: \begin{cases} (0, 0) \mapsto (1, 0) \\ (1, 1) \mapsto (0, 1) \end{cases} \quad \Pi_2: \begin{cases} (0, 0) \mapsto (0, 1) \\ (1, 1) \mapsto (1, 0) \end{cases}$$

are optimal for the quadratic distance cost $c(x, y) = |x - y|^2$. Figure 5 could help to visualize the situation.

Let now restrict our study to a particular situation. Let assume:

- (H1) we are in the Euclidean setting, i.e. $X = Y = \mathbb{R}^n$,
- (H2) the cost function is the square of the distance

$$c(x, y) = \frac{|x - y|^2}{2}, \tag{44}$$

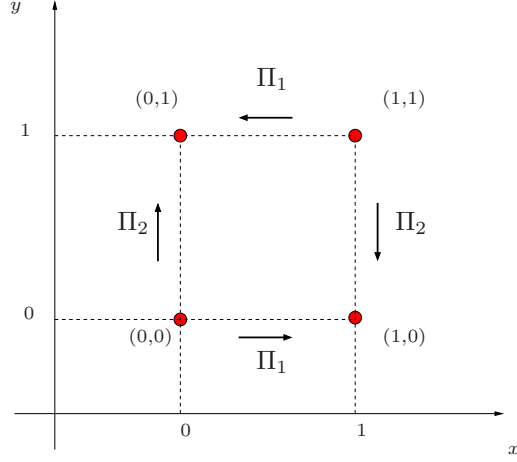


Figure 5: Two optimal transportation plans.

(H3) μ, ν are such that the following quantity

$$M_2 \doteq \int_{\mathbb{R}^n} \frac{|x|^2}{2} d\mu(x) + \int_{\mathbb{R}^n} \frac{|y|^2}{2} d\nu(y), \quad (45)$$

i.e. the sum of the second moments of μ, ν , is finite.

Then we can state the following Theorem.

Theorem 2.3. *Let assume (H1), (H2) and (H3). Then the following facts hold.*

- (i) **(Knott–Smith optimality criterion)** *A plan $\pi \in \Pi(\mu, \nu)$ is optimal if and only if there exists $\varphi: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ proper, convex and lower semicontinuous function such that*

$$\text{Supp}(\pi) \subset \text{Graph}(\partial\varphi), \quad (46)$$

or, equivalently, $y \in \partial\varphi(x)$ for π -a.e. $(x, y) \in X \times Y$. Moreover, the couple (φ, φ^*) solves the problem

$$\inf \left\{ \int_X \varphi(x) d\mu(x) + \int_Y \psi(y) d\nu(y) ; x \cdot y \leq \varphi(x) + \psi(y) \quad \forall (x, y) \right\}. \quad (47)$$

- (ii) **(Brenier’s Theorem)** *If μ does not give mass to sets with finite $\mathcal{H}^{n-1}$, then there exists a unique optimal transportation plan π and it is given by*

$$\pi = (\text{Id} \times \nabla\varphi)_\# \mu, \quad (48)$$

where $\nabla\varphi$ is the unique (read: uniquely determined μ -a.e.) gradient of a convex function φ such that $\nabla\varphi_\# \mu = \nu$.

- (iii) *As a consequence, if μ does not give mass to sets with finite $\mathcal{H}^{n-1}$, then $\nabla\varphi$ is the unique solution to Monge’s transportation problem, i.e.*

$$\int_{\mathbb{R}^n} |x - \nabla\varphi(x)|^2 d\mu(x) = \inf_{T_\# \mu = \nu} \int_{\mathbb{R}^n} |x - T(x)|^2 d\mu(x). \quad (49)$$

Remark 2.4. Notice that (φ, φ^*) does not solve the dual Monge-Kantorovich problem. It is the couple

$$\left(\frac{|x|^2}{2} - \varphi(x), \frac{|y|^2}{2} - \varphi^*(y) \right), \quad (50)$$

which provides a solution to the dual problem.

What about more general situations? A complete characterization, as in Theorem 2.3, is not known at this point. Anyway, one could prove the following results. First, without assuming finiteness of the second moments M_2 , the following Theorem holds true.

Theorem 2.5. *Let assume (H1) and (H2). Then the following facts hold.*

- (i) *If a plan $\pi \in \Pi(\mu, \nu)$ is optimal, then there exists $\varphi: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ proper, convex and lower semicontinuous function such that*

$$\text{Supp}(\pi) \subset \text{Graph}(\partial\varphi). \quad (51)$$

- (ii) **(Brenier's Theorem, refined version)** *If μ does not give mass to sets with finite $\mathcal{H}^{n-1}$, then there exists a unique $\varphi: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ proper, convex and lower semicontinuous function such that $\nabla\varphi_{\#}\mu = \nu$.*

Remark 2.6. Comparing Theorem 2.5 to Theorem 2.3, you can see that here there is no more a characterization of optimality, i.e. there is no proof that, for general measures μ, ν , transportation plans π concentrated on a cyclically monotone set are optimal. The equivalence holds when μ does not give mass to sets with finite $\mathcal{H}^{n-1}$ measure, and it is a consequence of the uniqueness in Brenier's Theorem. It is also true when you assume

$$\int_{X \times Y} |x - y|^2 d\mu(x) d\nu(y) < +\infty,$$

as proved by in [3]. But this is the closest result known at present.

Finally, for a general cost c in Polish spaces X, Y , one can prove the following Theorem.

Theorem 2.7. *Let X, Y be Polish spaces and $c: X \times Y \rightarrow [0, +\infty]$ be lower semicontinuous. Then the following facts hold.*

- (i) *Kantorovich duality theorem can be extended in the following way*

$$\begin{aligned} \inf_{\pi \in \Pi(\mu, \nu)} \int_{X \times Y} c d\pi &= \sup_{(\varphi, \psi) \in \Phi_c} \int_X \varphi d\mu + \int_Y \psi d\nu \\ &= \sup_{\varphi \in \mathbf{L}^1(\mu)} \int_Y \varphi^c d\nu - \int_X \varphi d\mu \\ &= \sup_{\psi \in \mathbf{L}^1(\nu)} \int_Y \psi d\nu - \int_X \psi^c d\mu \end{aligned} \quad (52)$$

where φ^c is the c -transform of φ , as in Definition 1.58, and ψ^c is the c -transform of ψ , as in Definition 1.60, and one might as well impose that φ is c -convex or ψ is c -concave in the last two suprema.

(ii) If c is real valued, then there exists $\Gamma \subset X \times Y$ measurable and c -cyclically monotone (additionally, Γ is closed, if c is continuous) such that for any $\pi \in \Pi(\mu, \nu)$ the following facts are equivalent:

- π is optimal;
- $\text{Supp}(\pi)$ is c -cyclically monotone;
- $\text{Supp}(\pi) \subset \Gamma$.

We will not prove Theorem 2.7. Refer to [20] for a complete proof, with slightly different notations.

Remark 2.8. Something similar to the first part of Theorem 2.7 will be recovered also in the setting of Theorem 2.3. Indeed, we will prove that a similar “restriction” can be performed for a slightly modified Kantorovich duality, obtained exploiting **(H1)**, **(H2)** and **(H3)**. Namely, we might consider only couples φ, φ^* with φ convex and lower semicontinuous (see the proof of Theorem 2.3 for more details).

Remark 2.9. The second part of Theorem 2.7, states the equivalence between optimality of a transportation plan π and c -cyclically monotonicity of $\text{Supp}(\pi)$. This is possible because we added the hypothesis of a real valued cost c . For a general cost the problem is still not completely solved¹⁶. The current state of the art is:

- the equivalence is **true** for c real valued and lower semicontinuous (i.e. Theorem 2.7);
- the equivalence is **true** for c continuous and possibly attaining infinite values (see [15]);
- the equivalence is **false** for a general c lower semicontinuous with possibly infinite values (see [3] for a counterexample).

Remark 2.10. As in Remark 1.1, the results in Theorem 2.7 still hold true for a cost which is only bounded below as follows

$$c(x, y) \geq a(x) + b(y) \quad \forall (x, y) \in X \times Y, \quad (53)$$

for suitable upper semicontinuous functions $a \in \mathbf{L}^1(\mu), b \in \mathbf{L}^1(\nu)$. Nevertheless, dealing with c -convexity and c -concavity, one has to slightly modify the argument from Remark 1.1. We still define

$$\bar{c}(x, y) \doteq c(x, y) - a(x) - b(y) \geq 0, \quad \Lambda \doteq \int_X a \, d\mu + \int_Y b \, d\nu < \infty, \quad (54)$$

but now, for all $\varphi: X \rightarrow \mathbb{R} \cup \{+\infty\}$ and $\psi: Y \rightarrow \mathbb{R} \cup \{-\infty\}$,

$$\bar{\varphi} \doteq \varphi + a, \quad \bar{\psi} \doteq \psi - b. \quad (55)$$

With these definitions, we can prove easily that

$$c \text{ real valued, lower semicontinuous} \Rightarrow \bar{c} \text{ real valued, lower semicontinuous,}$$

¹⁶A possible construction of an equivalent condition to optimality is sketched in [20] and it is in progress by Bianchini and Caravenna, but it requires a lot of measure theory and it is far beyond our goal here.

$$\begin{aligned}
\varphi \in \mathbf{L}^1(\mu) &\Leftrightarrow \bar{\varphi} \in \mathbf{L}^1(\mu) & \psi \in \mathbf{L}^1(\nu) &\Leftrightarrow \bar{\psi} \in \mathbf{L}^1(\nu), \\
(-\varphi, \psi) \in \Phi_c &\Leftrightarrow (-\bar{\varphi}, \bar{\psi}) \in \Phi_{\bar{c}}, \\
\forall \pi \in \Pi(\mu, \nu) & \int_{X \times Y} \bar{c} d\pi = \int_{X \times Y} c d\pi - \Lambda, \\
\forall (-\varphi, \psi) \in \mathbf{L}^1(\mu) \times \mathbf{L}^1(\nu) & \int_Y \bar{\psi} d\nu - \int_X \bar{\varphi} d\mu = \int_Y \psi d\nu - \int_X \varphi d\mu - \Lambda. \\
\varphi \text{ } c\text{-convex} &\Leftrightarrow \bar{\varphi} \text{ } \bar{c}\text{-convex}, & \psi \text{ } c\text{-concave} &\Leftrightarrow \bar{\psi} \text{ } \bar{c}\text{-concave}, \\
(\varphi, \psi) \text{ are } c\text{-conjugate} &\Leftrightarrow (\bar{\varphi}, \bar{\psi}) \text{ are } \bar{c}\text{-conjugate},
\end{aligned}$$

where (φ, ψ) are said to be c -conjugate if either $\psi = \varphi^c$ or $\varphi = \psi^c$. Finally a set $\Gamma \subset X \times Y$ is c -cyclically monotone if and only if it is $\bar{c}$ -cyclically monotone. Therefore, to any solution of the Kantorovich transportation problem with non-negative cost, corresponds a solution of the Kantorovich transportation problem with cost bounded below as in (53).

3 Proof of Main Results, I

Here we give a proof to Theorem 2.1.

Step 1

Since X, Y are Polish spaces, then we have seen (Corollary 1.16) that $\{\mu\}, \{\nu\}$ are tight sets. Therefore, by Proposition 1.13, $\Pi(\mu, \nu)$ is a tight set in $\mathcal{P}(X \times Y)$. By Prokhorov's Theorem (Theorem 1.14 in these notes), $\Pi(\mu, \nu)$ has compact closure in $\mathcal{P}(X \times Y)$ (where the closure is w.r.t. w -topology).

We claim that actually $\Pi(\mu, \nu)$ is weakly closed, and therefore it is weakly compact. Indeed, given a sequence $\pi_n \in \Pi(\mu, \nu)$, we have by (2)

$$\int_{X \times Y} [\varphi(x) + \psi(y)] d\pi_n(x, y) = \int_X \varphi(x) d\mu(x) + \int_Y \psi(y) d\nu(y),$$

for any $(\varphi, \psi) \in \mathbf{L}^1(\mu) \times \mathbf{L}^1(\nu)$. In particular it holds for any couple of functions in $\mathbf{C}_b(X) \times \mathbf{C}_b(Y)$. Therefore, if $\pi_n \xrightarrow{*} \pi_\infty$, we can pass to the limit in the left hand side and use the weak convergence to obtain

$$\int_{X \times Y} [\varphi(x) + \psi(y)] d\pi_\infty(x, y) = \int_X \varphi(x) d\mu(x) + \int_Y \psi(y) d\nu(y),$$

for all (φ, ψ) continuous and bounded. But in Polish spaces this is enough to conclude that the same property holds for any couple in $\mathbf{L}^1$, which is equivalent to $\pi_\infty \in \Pi(\mu, \nu)$. Hence, $\Pi(\mu, \nu)$ is weakly closed as claimed.

Step 2

Let $(\pi_k)_{k \in \mathbb{N}}$ be a minimizing sequence and let $\pi_* \in \Pi(\mu, \nu)$ a cluster point of the sequence (it exists due to compactness of $\Pi(\mu, \nu)$).

Since c is lower semicontinuous and nonnegative, we can find a sequence $\{c_\ell\}_{\ell \in \mathbb{N}}$ of nondecreasing bounded Lipschitz functions such that c is the point-wise supremum of the c_ℓ . Indeed, it is enough to consider the Moreau–Yosida approximation of c ,

$$\tilde{c}_\ell(x) \doteq \inf_{z \in X} \{c(z) + \ell d(x, z)\}, \quad (56)$$

and to make it bounded by setting $c_\ell \doteq \min\{\tilde{c}_\ell, \ell\}$.

Finally, using respectively the monotone convergence theorem, the fact that π_* is a cluster point, $c_\ell \leq c$ and π_k is a minimizing sequence, we can conclude

$$\begin{aligned} \int_{X \times Y} c(x, y) d\pi_*(x, y) &= \lim_{\ell} \int_{X \times Y} c_\ell(x, y) d\pi_*(x, y) \\ &\leq \lim_{\ell} \limsup_k \int_{X \times Y} c_\ell(x, y) d\pi_k(x, y) \\ &\leq \limsup_k \int_{X \times Y} c(x, y) d\pi_k(x, y) \\ &= \inf I(\pi). \quad \diamond \end{aligned}$$

4 Proof of Main Results, II

Here we prove Theorem 2.3 through various steps. The plan is the following:

1. to reformulate the problem (3)–(4) exploiting **(H1)**, **(H2)** and **(H3)** (Steps 1–2),
2. to give a rough idea of a possible strategy of proof (Step 3),
3. to prove two accessory results we need to make this rough idea more rigorous (Steps 4–5),
4. to prove Knott–Smith criterion (Steps 6–7),
5. to prove Brenier’s Theorem (Steps 8–10),
6. to conclude the whole proof with *(iii)* (Step 11).

Step 1

First of all notice that, under assumptions **(H1)**–**(H2)**–**(H3)**, we can rewrite (3) as follows.

$$\begin{aligned} \inf_{\pi \in \Pi(\mu, \nu)} I(\pi) &= \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{|x - y|^2}{2} d\pi(x, y) \\ &= M_2 - \sup_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} x \cdot y d\pi(x, y). \end{aligned} \quad (57)$$

On the other hand, we also have

$$\begin{aligned}
(\varphi, \psi) \in \Phi_c &\Leftrightarrow \varphi(x) + \psi(y) \leq \frac{|x-y|^2}{2} \quad \mu\text{-a.e. } x, \nu\text{-a.e. } y \\
&\Leftrightarrow x \cdot y \leq \underbrace{\left(\frac{|x|^2}{2} - \varphi(x) \right)}_{\tilde{\varphi}(x)} + \underbrace{\left(\frac{|y|^2}{2} - \psi(y) \right)}_{\tilde{\psi}(y)} \quad \mu\text{-a.e. } x, \nu\text{-a.e. } y
\end{aligned}$$

Therefore, there holds

$$\begin{aligned}
\sup_{(\varphi, \psi) \in \Phi_c} J(\varphi, \psi) &= \sup_{(\varphi, \psi) \in \Phi_c} M_2 - \int_{\mathbb{R}^n} \left(\frac{|x|^2}{2} - \varphi(x) \right) d\mu(x) \\
&\quad - \int_{\mathbb{R}^n} \left(\frac{|y|^2}{2} - \psi(y) \right) d\nu(y) \\
&= M_2 - \inf_{(\varphi, \psi) \in \tilde{\Phi}} J(\varphi, \psi),
\end{aligned} \tag{58}$$

where

$$\tilde{\Phi} \doteq \{(\varphi, \psi) \in \mathbf{L}^1(\mu) \times \mathbf{L}^1(\nu) ; x \cdot y \leq \varphi(x) + \psi(y) \quad \mu\text{-a.e. } x, \nu\text{-a.e. } y\}.$$

Finally, we can rewrite the Kantorovich duality theorem as follows

$$\sup_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} x \cdot y d\pi(x, y) = \inf_{(\varphi, \psi) \in \tilde{\Phi}} J(\varphi, \psi). \tag{59}$$

Step 2

We now prove that in the right hand side of (59), we can in fact consider the infimum over the smaller set

$$\{(\varphi, \psi) \in \mathbf{L}^1(\mu) \times \mathbf{L}^1(\nu) ; x \cdot y \leq \varphi(x) + \psi(y) \quad \forall x, y\} \tag{60}$$

Indeed, given a couple $(\varphi, \psi) \in \tilde{\Phi}$, let $\mathcal{N}_\varphi, \mathcal{N}_\psi \subset \mathbb{R}^n$ be such that

$$\mu(\mathcal{N}_\varphi) = 0 \quad \nu(\mathcal{N}_\psi) = 0,$$

$$x \cdot y \leq \varphi(x) + \psi(y) \quad \forall x \in \mathbb{R}^n \setminus \mathcal{N}_\varphi, \forall y \in \mathbb{R}^n \setminus \mathcal{N}_\psi.$$

Consider the functions (φ', ψ') defined as follows:

$$\varphi' \doteq \begin{cases} \varphi & \text{in } \mathbb{R}^n \setminus \mathcal{N}_\varphi \\ +\infty & \text{in } \mathcal{N}_\varphi \end{cases} \quad \psi' \doteq \begin{cases} \psi & \text{in } \mathbb{R}^n \setminus \mathcal{N}_\psi \\ +\infty & \text{in } \mathcal{N}_\psi \end{cases}$$

Since these new functions are obtained by the previous ones by changing the values only on negligible sets, we can deduce that

- $J(\varphi', \psi') = J(\varphi, \psi)$,
- $(\varphi', \psi') \in \tilde{\Phi}$,
- but now $x \cdot y \leq \varphi'(x) + \psi'(y)$ for all $x, y \in \mathbb{R}^n$.

Therefore, restricting our attention to this smaller set (60) in the infimum (59) does not make any difference.

Step 3

Thanks to Step 2, in the characterizing condition of $\tilde{\Phi}$

$$\psi(y) \geq x \cdot y - \varphi(x), \quad (61)$$

we can pass to the supremum for $x \in \mathbb{R}^n$, obtaining

$$\psi(y) \geq \sup_{x \in \mathbb{R}^n} \{x \cdot y - \varphi(x)\} = \varphi^*(y), \quad (62)$$

and, hence, $J(\varphi, \psi) \geq J(\varphi, \varphi^*)$.

Remark 4.1. Alternatively, one could try to use the essential supremum in the expression above, without passing through Step 2. Even if such a choice may sound reasonable at first, issues might arise quite soon. E.g. there could be need to pass from properties which hold for every x , ν -almost every y , to properties which hold ν -almost every y , for every x . In the former case, the ν -negligible set in which the property fails can well depend on x ; in the latter case, on the other hand, it has to be the same set for all x . Of course, in certain situations to find such a set in the latter case may become a nightmare!

Recalling Remark 1.36, there also holds for all $y \in \mathbb{R}^n$

$$\varphi(x) \geq x \cdot y - \varphi^*(y). \quad (63)$$

Passing to the supremum on y , we obtain

$$\varphi(x) \geq \sup_{y \in \mathbb{R}^n} \{x \cdot y - \varphi^*(y)\} = \varphi^{**}(x), \quad (64)$$

and, hence, $J(\varphi, \varphi^*) \geq J(\varphi^{**}, \varphi^*)$.

Inspired by the previous inequalities, one could think to proceed as follows:

- for any minimizing sequence (φ_k, ψ_k) to the Kantorovich optimization problem (59), to consider $(\varphi_k^{**}, \varphi_k^*)$, which attains a lower cost J and consists of more regular functions, since now it would consist of a convex and lower semicontinuous function and its Fenchel–Legendre transform,
- to ensure that $(\varphi_k^{**}, \varphi_k^*) \in \tilde{\Phi}$ and that the increased regularity leads to weak compactness of the sequence (in $\mathbf{L}^1$),
- to find a minimizer passing to the limit as $k \rightarrow \infty$ and to prove Theorem 2.3 using this limit function.

Unfortunately, even starting with a minimizing sequence $(\varphi_k, \varphi_k^*) \in \tilde{\Phi}$, with φ_k convex and lower semicontinuous, the weak compactness in $\mathbf{L}^1$ is far from guaranteed. Indeed, the sequence $(\varphi_k - k, \varphi_k^* + k)$ would still

- belong to $\tilde{\Phi}$,
- be a minimizing sequence for J , since $J(\varphi_k, \varphi_k^*) = J(\varphi_k - k, \varphi_k^* + k)$,
- be of the desired form, since $(\varphi_k - k)^* = \varphi_k^* + k$,

but it is not weakly compact in $\mathbf{L}^1$.

Therefore, to overcome these difficulties, we necessarily need to prove the following facts:

Claim 1. Given (φ_k, ψ_k) minimizing sequence for the Kantorovich Problem, then there exists $(a_k)_{k \in \mathbb{N}} \subset \mathbb{R}$ such that the sequence

$$(\overline{\varphi}_k, \overline{\psi}_k) \doteq (\varphi_k^{**} - a_k, \varphi_k^* + a_k)$$

satisfies:

- (i) $(\overline{\varphi}_k, \overline{\psi}_k)$ is a minimizing sequence for J ,
- (ii) $\forall x, y \in \mathbb{R}^n$

$$\overline{\varphi}_k(x) \geq -\frac{|x|^2}{2}, \quad \overline{\psi}_k(y) \geq -\frac{|y|^2}{2}, \quad (65)$$

- (iii)

$$\begin{aligned} \liminf_{k \rightarrow \infty} \inf_{x \in \mathbb{R}^n} \left(\overline{\varphi}_k(x) + \frac{|x|^2}{2} \right) &\leq \inf_{\tilde{\Phi}} J + M_2, \\ \liminf_{k \rightarrow \infty} \inf_{y \in \mathbb{R}^n} \left(\overline{\varphi}_k(y) + \frac{|y|^2}{2} \right) &\leq \inf_{\tilde{\Phi}} J + M_2. \end{aligned} \quad (66)$$

Claim 2. There exists φ proper, convex and lower semicontinuous on $\mathbb{R}^n$ such that $(\varphi, \varphi^*) \in \tilde{\Phi}$ and

$$\inf_{\tilde{\Phi}} J = J(\varphi, \varphi^*). \quad (67)$$

In other words:

- a. Claim 1 states that we can always modify our minimizing sequence so that it has the required form $(\varphi_k^{**}, \varphi_k^*)$ and it satisfies suitable additional bounds;
- b. Claim 2 states that such a modified sequence is weakly compact in $\mathbf{L}^1$ and the limit still has the same form.

In Steps 4–5 we will prove these claims. In Step 6–10 we will go back to the proof of the main result, following the approach sketched above.

Step 4

Proof of Claim 1. Let $(\varphi_k, \psi_k) \in \tilde{\Phi}$. Then, for all x, y we have $\varphi_k(x) \geq x \cdot y - \psi_k(y)$. Since $\psi_k \not\equiv +\infty$, then φ_k can be bounded below by some affine function, say $x \cdot y_o + b_o$ for some $y_o \in \mathbb{R}^n$ and $b_o \in \mathbb{R}$.¹⁷ Then

$$\varphi_k^*(y_o) = \sup_{x \in \mathbb{R}^n} \{x \cdot y_o - \varphi_k(x)\} \leq -b_o.$$

This means $\varphi_k^* \not\equiv +\infty$, i.e. φ_k^* proper. Similarly, being $\varphi_k \not\equiv +\infty$, φ_k^* is bounded below by a suitable affine function, and therefore

$$a_k \doteq \inf_{y \in \mathbb{R}^n} \left(\varphi_k^*(y) + \frac{|y|^2}{2} \right) \quad (68)$$

¹⁷It is enough to consider y_o such that $\psi_k(y_o) < \infty$ and $b_o = \psi_k(y_o)$.

is finite for all k . Let us prove that this choice of $(a_k)_{k \in \mathbb{N}}$ satisfies the Claim.

a. Set, as in Claim 1,

$$(\overline{\varphi}_k, \overline{\psi}_k) \doteq (\varphi_k^{**} + a_k, \varphi_k^* - a_k). \quad (69)$$

Note that $\overline{\varphi}_k = \overline{\psi}_k^*$. By construction, we immediately have

$$\inf_{y \in \mathbb{R}^n} \left(\overline{\psi}_k(y) + \frac{|y|^2}{2} \right) = 0.$$

Moreover,

$$\begin{aligned} \overline{\varphi}_k(x) + \frac{|x|^2}{2} &= \overline{\psi}_k^*(x) + \frac{|x|^2}{2} = \sup_{y \in \mathbb{R}^n} \left[x \cdot y - \overline{\psi}_k(y) + \frac{|x|^2}{2} \right] \\ &\leq \sup_{y \in \mathbb{R}^n} \left[-\frac{|y|^2}{2} - \overline{\psi}_k(y) \right] = -\inf_{y \in \mathbb{R}^n} \left[\frac{|y|^2}{2} + \overline{\psi}_k(y) \right] = 0. \end{aligned}$$

This proves (65).

b. Reasoning as in Step 3,

$$J(\overline{\varphi}_k, \overline{\psi}_k) = J(\varphi_k^{**}, \varphi_k^*) \leq J(\varphi_k, \psi_k) < +\infty.$$

In particular, it is finite $M_2 + J(\overline{\varphi}_k, \overline{\psi}_k)$, i.e.

$$\int_{\mathbb{R}^n} \left(\overline{\varphi}_k(x) + \frac{|x|^2}{2} \right) d\mu(x) + \int_{\mathbb{R}^n} \left(\overline{\psi}_k(y) + \frac{|y|^2}{2} \right) d\nu(y) < +\infty.$$

Since by *a.* both integrands are nonnegative, we deduce they are integrable. Moreover, recalling that μ, ν are assumed to have finite second moments, it means that $\overline{\varphi}_k \in \mathbf{L}^1(\mu)$ and $\overline{\psi}_k \in \mathbf{L}^1(\nu)$.

c. It is now quite easy to conclude. By construction, $(\overline{\varphi}_k, \overline{\psi}_k) \in \widetilde{\Phi}$ and it is a minimizing sequence for J . Moreover, there holds

$$\begin{aligned} J(\overline{\varphi}_k, \overline{\psi}_k) + M_2 &= \int_{\mathbb{R}^n} \left(\overline{\varphi}_k(x) + \frac{|x|^2}{2} \right) d\mu(x) + \int_{\mathbb{R}^n} \left(\overline{\psi}_k(y) + \frac{|y|^2}{2} \right) d\nu(y) \\ &\geq \inf_{x \in \mathbb{R}^n} \left(\overline{\varphi}_k(x) + \frac{|x|^2}{2} \right) + \inf_{y \in \mathbb{R}^n} \left(\overline{\psi}_k(y) + \frac{|y|^2}{2} \right). \end{aligned}$$

Since we have already shown that both these infima are nonnegative, we can conclude

$$\liminf_{k \rightarrow \infty} \inf_{x \in \mathbb{R}^n} \left(\overline{\varphi}_k(x) + \frac{|x|^2}{2} \right) \leq \liminf_{k \rightarrow \infty} J(\overline{\varphi}_k, \overline{\psi}_k) + M_2 = \inf_{\widetilde{\Phi}} J + M_2,$$

and similarly for the expression containing $\overline{\psi}_k$, i.e. we have proved (66) as well.

◇

Remark 4.2. Notice that exactly the same construction can be applied to any pair (φ, ψ) which minimizes J , to produce a new couple

$$(\varphi', \psi') = (\varphi^{**} + a', \varphi^* - a'),$$

for $a' \in \mathbb{R}$ as in (68), which still belongs to $\tilde{\Phi}$ and has the form (f, f^*) for some proper, convex and semicontinuous map f . Since

$$J(\varphi', \psi') \leq J(\varphi, \psi) = \inf_{\tilde{\Phi}} J,$$

the new couple minimizes J as well. This procedure is sometimes called *double convexification*.

Step 5

Proof of Claim 2.

a. Our first goal is to replace everything with nonnegative quantities. Let $(\varphi_k, \psi_k)_{k \in \mathbb{N}}$ be a minimizing sequence for J . Thanks to Claim 1, we can assume without loss of generality that

$$0 \leq \varphi_k(x) + \frac{|x|^2}{2}, \quad 0 \leq \psi_k(y) + \frac{|y|^2}{2}.$$

Notice that

$$(\varphi, \psi) \in \tilde{\Phi} \Leftrightarrow \left[\varphi(x) + \frac{|x|^2}{2} \right] + \left[\psi(y) + \frac{|y|^2}{2} \right] \geq \frac{|x|^2}{2} + \frac{|y|^2}{2} + x \cdot y = \frac{|x+y|^2}{2}$$

Being our sequence in $\tilde{\Phi}$, assume

$$\left[\varphi_k(x) + \frac{|x|^2}{2} \right] + \left[\psi_k(y) + \frac{|y|^2}{2} \right] \geq \frac{|x+y|^2}{2} \geq 0.$$

b. Now, we would like to prove some sort of weak compactness of our sequence in $\mathbf{L}^1$. In view of Theorem 1.20 (Dunford–Pettis Theorem), this requires some additional equiintegrability of the sequence. Hence, we start applying a truncation procedure. For each $\ell \in \mathbb{N}$, define $(\varphi_k^{(\ell)}, \psi_k^{(\ell)})$ as follows

$$\begin{aligned} \varphi_k^{(\ell)}(x) + \frac{|x|^2}{2} &\doteq \min \left\{ \varphi_k(x) + \frac{|x|^2}{2}, \ell \right\}, \\ \psi_k^{(\ell)}(y) + \frac{|y|^2}{2} &\doteq \min \left\{ \psi_k(y) + \frac{|y|^2}{2}, \ell \right\}. \end{aligned}$$

It is easy to verify that $\varphi_k^{(\ell)}$ satisfies

$$0 \leq \varphi_k^{(\ell)}(x) + \frac{|x|^2}{2} \leq \ell, \quad \varphi_k^{(1)} \leq \varphi_k^{(2)} \leq \dots \leq \varphi_k^{(\ell)} \leq \dots \quad (70)$$

and similar inequalities hold for $\psi_k^{(\ell)}$. Moreover,

$$J(\varphi_k^{(\ell)}, \psi_k^{(\ell)}) \leq J(\varphi_k, \psi_k), \quad (71)$$

and

$$\varphi_k^{(\ell)}(x) + \psi_k^{(\ell)}(y) \geq \min \left\{ \frac{|x+y|^2}{2}, \ell \right\} - \left(\frac{|x|^2}{2} + \frac{|y|^2}{2} \right). \quad (72)$$

c. We claim that, for any fixed $\ell \in \mathbb{N}$, the sequence $\varphi_k^{(\ell)}$ is weakly compact in $\mathbf{L}^1(\mu)$ (resp. the sequence $\psi_k^{(\ell)}$ is weakly compact in $\mathbf{L}^1(\nu)$). Indeed, we immediately see that, given $\varepsilon > 0$, if we choose $\delta = \frac{\varepsilon}{\ell}$, E with $\mu(E) < \delta$ and recalling (70),

$$\int_E \varphi_k^{(\ell)} + \frac{|x|^2}{2} d\mu(x) < \ell \mu(E) < \varepsilon,$$

so that $\left(\varphi_k^{(\ell)} + \frac{|x|^2}{2}\right)_{k \in \mathbb{N}}$ is equiintegrable, and hence weakly compact in $\mathbf{L}^1(\mu)$.

To prove weakly compactness of $\left(\varphi_k^{(\ell)}\right)_{k \in \mathbb{N}}$, it is enough to observe that it differs from a weakly compact sequence by a fixed element of $\mathbf{L}^1(\mu)$, namely $\frac{|x|^2}{2}$. Therefore, it is weakly compact as well.

d. Fixed $\ell \in \mathbb{N}$, up to a subsequence that we still label with k , we have that

$$\varphi_k^{(\ell)} \xrightarrow{k \rightarrow \infty} \varphi^{(\ell)} \quad \text{weakly in } \mathbf{L}^1(\mu).$$

Through a diagonal procedure, we can now extract a subsequence of indices $k \in \mathbb{N}$ such that the convergence above holds for all ℓ . Then, since weak convergence preserves the ordering, we can deduce from (70), (71) and (72) that

$$\varphi^{(1)} \leq \varphi^{(2)} \leq \dots \leq \varphi^{(\ell)} \leq \dots \quad \psi^{(1)} \leq \psi^{(2)} \leq \dots \leq \psi^{(\ell)} \leq \dots \quad (73)$$

$$J(\varphi^{(\ell)}, \psi^{(\ell)}) = \lim_{k \rightarrow \infty} J(\varphi_k^{(\ell)}, \psi_k^{(\ell)}) \leq \liminf_{k \rightarrow \infty} J(\varphi_k, \psi_k) = \inf_{\tilde{\Phi}} J, \quad (74)$$

$$\varphi^{(\ell)}(x) + \psi^{(\ell)}(y) \geq \min \left\{ \frac{|x+y|^2}{2}, \ell \right\} - \left(\frac{|x|^2}{2} + \frac{|y|^2}{2} \right). \quad (75)$$

e. At this point, we have constructed a sequence $(\varphi^{(\ell)}, \psi^{(\ell)})$ in $\mathbf{L}^1(\mu) \times \mathbf{L}^1(\nu)$, whose elements are:

- monotone nondecreasing, due to (73);
- uniformly bounded in $\mathbf{L}^1$, as a consequence of (74);
- bounded below by a fixed $\mathbf{L}^1$ function, due to (75).

Thus, we can apply the monotone convergence theorem and deduce the existence of $\mathbf{L}^1$ limits (φ, ψ) defined almost everywhere as

$$\varphi = \lim_{\ell \rightarrow \infty} \varphi^{(\ell)}, \quad \psi = \lim_{\ell \rightarrow \infty} \psi^{(\ell)},$$

and satisfying

$$J(\varphi, \psi) = \lim_{\ell \rightarrow \infty} J(\varphi^{(\ell)}, \psi^{(\ell)}) \leq \inf_{\tilde{\Phi}} J. \quad (76)$$

Notice that, passing to the limit as $\ell \rightarrow \infty$ in (75), we obtain $(\varphi, \psi) \in \tilde{\Phi}$. Therefore, from (76), it follows that we have found a minimizing pair.

f. Finally, it remains to prove that we can choose φ proper, convex and lower semicontinuous and $\psi = \varphi^*$. But this follows easily applying the double convexification procedure described in Remark 4.2. Therefore, Claim 2 is proved. $\diamond$

Step 6

We move back to the proof of Theorem 2.3, and prove that for any optimal plan π we have

$$\text{Supp}(\pi) \subset \text{Graph}(\partial\varphi).$$

Take any optimal transportation plan π (we know from Theorem 2.1 that there exists such a plan π). Let also φ be proper, convex and lower semicontinuous given by Claim 2. Then (φ, φ^*) is a minimizer and, recalling (2), it follows

$$\begin{aligned} \int_{\mathbb{R}^n \times \mathbb{R}^n} x \cdot y \, d\pi(x, y) &= \int_{\mathbb{R}^n} \varphi(x) \, d\mu(x) + \int_{\mathbb{R}^n} \varphi^*(y) \, d\nu(y) \\ &= \int_{\mathbb{R}^n \times \mathbb{R}^n} [\varphi(x) + \varphi^*(y)] \, d\pi(x, y). \end{aligned}$$

This is equivalent to say

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} [\varphi(x) + \varphi^*(y) - x \cdot y] \, d\pi(x, y) = 0,$$

which, in turn, means

$$\varphi(x) + \varphi^*(y) - x \cdot y = 0 \quad \pi - \text{a.e.}$$

Indeed, due to $(\varphi, \varphi^*) \in \tilde{\Phi}$, the integrand is nonnegative. Therefore, it has to vanish for π -a.e. (x, y) . This is equivalent, by (30), to say $y \in \partial\varphi(x)$ for π -a.e. (x, y) , i.e. $\pi(X \times Y \setminus \text{Graph}(\partial\varphi)) = 0$. Since, recalling Corollary 1.49, $\text{Graph}(\partial\varphi)$ is closed, the whole $\text{Supp}(\pi)$ is contained in $\text{Graph}(\partial\varphi)$ and (46) has been proved.

Step 7

To complete Knott–Smith optimality criterion, it remains to show that given φ as in (i) and $\pi \in \Pi(\mu, \nu)$ with support contained in the graph of $\partial\varphi$, this plan π is optimal. But this can be easily accomplished by using the previous arguments in reverse order. The hypothesis on $\text{Supp}(\pi)$ guarantees that $y \in \partial\varphi(x)$ for π -a.e. (x, y) . In turn, this implies by (30)

$$\varphi(x) + \varphi^*(y) - x \cdot y = 0 \quad \pi - \text{a.e.} \tag{77}$$

and, therefore,

$$\begin{aligned} \int_{\mathbb{R}^n \times \mathbb{R}^n} x \cdot y \, d\pi(x, y) &= \int_{\mathbb{R}^n \times \mathbb{R}^n} [\varphi(x) + \varphi^*(y)] \, d\pi(x, y) \\ &= \int_{\mathbb{R}^n} \varphi(x) \, d\mu(x) + \int_{\mathbb{R}^n} \varphi^*(y) \, d\nu(y) \\ &= \inf_{(\varphi, \psi) \in \tilde{\Phi}} J(\varphi, \psi) = \sup_{\sigma \in \Pi(\mu, \nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} x \cdot y \, d\sigma(x, y), \end{aligned}$$

where we used respectively (77), (2), optimality of (φ, φ^*) and Kantorovich duality theorem, as restated in (59).

This ensures that π is optimal, and therefore it concludes the proof of the first part of Theorem 2.3.

Step 8

Assume now that μ does not give mass to sets with finite $\mathcal{H}^{n-1}$ measure. We want to show that any optimal π has the required form. Let φ the proper, convex and lower semicontinuous given in (i).

Claim 3. $\varphi \in \mathbf{L}^1(\mu)$ implies φ finite μ -a.e. in $\mathbb{R}^n$.

Assume that Claim 3 has been proved (we will show it in Step 9). Then, it is equivalent to say that $\mu(\text{Dom}(\varphi)) = 1$.

On the other hand, as already noticed in Remark 1.22, $\partial\text{Dom}(\varphi)$ has finite $\mathcal{H}^{n-1}$ measure. Hence, $\mu(\text{Int}(\text{Dom}(\varphi))) = 1$.

Now, it follows from Corollary 1.50 that the set of $x \in \text{Int}(\text{Dom}(\varphi))$ in which φ is not differentiable has finite $\mathcal{H}^{n-1}$ measure. Therefore, it has to be

$$\mu(\{x \in \text{Int}(\text{Dom}(\varphi)) ; \varphi \text{ is diff. in } x\}) = 1,$$

and this means that $\partial\varphi(x) = \{\nabla\varphi(x)\}$ for μ -a.e. x .

Observe now that, since $\pi \in \Pi(\mu, \nu)$, i.e. π has marginals μ, ν ,

$$\begin{aligned} \pi(\{(x, y) ; \varphi \text{ is diff. in } x\}) &= \pi(\{x ; \varphi \text{ is diff. in } x\} \times Y) \\ &= \mu(\{x ; \varphi \text{ is diff. in } x\}) = 1, \end{aligned}$$

and we can conclude that $\partial\varphi(x) = \{\nabla\varphi(x)\}$ for π -a.e. (x, y) .

This is enough, because we already know by Knott–Smith criterion that π optimal implies $y \in \partial\varphi(x) = \{\nabla\varphi(x)\}$, i.e. $y = \nabla\varphi(x)$, for π -a.e. (x, y) .

Now we claim that this implies $\pi = (\text{Id} \times \nabla\varphi)_\# \mu$. Recalling Remark 1.11, we consider $f \in \mathbf{L}^1(\pi)$ and prove that

$$\int_{X \times Y} f \, d\pi = \int_X f \circ (\text{Id} \times \nabla\varphi) \, d\mu.$$

But, recalling (2), we have

$$\begin{aligned} \int_{X \times Y} f(x, y) \, d\pi(x, y) &= \int_{X \times Y} f(x, \nabla\varphi(x)) \, d\pi(x, y) \\ &= \int_X f(x, \nabla\varphi(x)) \, d\mu(x) \\ &= \int_X f \circ (\text{Id} \times \nabla\varphi)(x) \, d\mu(x). \end{aligned}$$

Similarly, we can prove $\nabla\varphi_\# \mu = \nu$. Indeed for all $(\phi, \psi) \in \mathbf{L}^1(\mu) \times \mathbf{L}^1(\nu)$ we have by (2),

$$\begin{aligned} \int_X \phi(x) \, d\mu(x) + \int_Y \psi(y) \, d\nu(y) &= \int_{X \times Y} (\phi(x) + \psi(y)) \, d\pi(x, y) \\ &= \int_{X \times Y} (\phi(x) + \psi(\nabla\varphi(x))) \, d\pi(x, y) \\ &= \int_X (\phi(x) + \psi(\nabla\varphi(x))) \, d\mu(x), \end{aligned}$$

which implies

$$\int_Y \psi(y) \, d\nu(y) = \int_X \psi(\nabla\varphi(x)) \, d\mu(x)$$

for all $\psi \in \mathbf{L}^1(\nu)$. This allows to conclude $\nu = \nabla\varphi_\# \mu$.

Step 9

Proof of Claim 3. This is a brief exercise in measure theory. Let

$$A_k \doteq \{x \in \mathbb{R}^n ; \varphi(x) > k\} .$$

Then we want to prove that

$$\mu \left(\bigcap_{k \in \mathbb{N}} A_k \right) = 0 .$$

First of all, notice that $\bigcap_k A_k$ is measurable and that, since $\mu(A_k) \leq \mu(\mathbb{R}^n) = 1$ for all k , $\lim_k \mu(A_k)$ is finite. Then $\mu(\bigcap_k A_k) = \lim_k \mu(A_k)$. On the other hand,

$$k \mu(A_k) = \int_{A_k} k \, d\mu \leq \int_{A_k} \varphi \, d\mu \leq \|\varphi\|_{\mathbf{L}^1} .$$

Therefore,

$$\mu \left(\bigcap_{k \in \mathbb{N}} A_k \right) = \lim_{k \rightarrow \infty} \mu(A_k) \leq \lim_{k \rightarrow \infty} \frac{\|\varphi\|_{\mathbf{L}^1}}{k} = 0 .$$

and this concludes the proof of Claim 3. $\diamond$

Step 10

We address now the uniqueness of the optimal transportation plan. In Step 8 we found a φ such that $\nabla \varphi_{\#} \mu = \nu$, $\pi = (\text{Id} \times \nabla \varphi)_{\#} \mu$ is an optimal plan and (φ, φ^*) is optimal for the dual problem. Let $\overline{\varphi}$ be another proper, convex and lower semicontinuous function such that $\nabla \overline{\varphi}_{\#} \mu = \nu$ and $(\overline{\varphi}, \overline{\varphi}^*)$ is optimal for the dual problem. This means

$$\int_{\mathbb{R}^n} \overline{\varphi}(x) \, d\mu(x) + \int_{\mathbb{R}^n} \overline{\varphi}^*(y) \, d\nu(y) = \int_{\mathbb{R}^n} \varphi(x) \, d\mu(x) + \int_{\mathbb{R}^n} \varphi^*(y) \, d\nu(y) .$$

Using (2) and the optimality of π , we can write

$$\begin{aligned} \int_{\mathbb{R}^n \times \mathbb{R}^n} [\overline{\varphi}(x) + \overline{\varphi}^*(y)] \, d\pi(x, y) &= \int_{\mathbb{R}^n \times \mathbb{R}^n} [\varphi(x) + \varphi^*(y)] \, d\pi(x, y) \\ &= \int_{\mathbb{R}^n \times \mathbb{R}^n} x \cdot y \, d\pi(x, y) . \end{aligned}$$

Since we know the explicit expression of π by (48), we can rewrite

$$\int_{\mathbb{R}^n} [\overline{\varphi}(x) + \overline{\varphi}^*(\nabla \varphi(x))] \, d\mu(x) = \int_{\mathbb{R}^n} x \cdot \nabla \varphi(x) \, d\mu(x) ,$$

or, equivalently,

$$\int_{\mathbb{R}^n} [\overline{\varphi}(x) + \overline{\varphi}^*(\nabla \varphi(x)) - x \cdot \nabla \varphi(x)] \, d\mu(x) = 0 .$$

We claim that the integrand is nonnegative. Indeed, $(\varphi, \overline{\varphi}^*)$ belongs to the set defined in (60). This means that

$$\overline{\varphi}(x) + \overline{\varphi}^*(y) - x \cdot y \geq 0 ,$$

for all x, y . Hence, it is enough to take $y = \nabla\varphi(x)$ in this inequality to prove our claim.

But then the integrand has to vanish μ -a.e. and, recalling (30), this is equivalent to say that $\nabla\varphi(x) \in \partial\overline{\varphi}(x)$ for μ -a.e. $x \in \mathbb{R}^n$. Since, as we observed in Remark 1.26, $\overline{\varphi}$ is differentiable μ -a.e., its subdifferential must reduce to contain a single element (the gradient of φ in the point) μ -a.e.

This is equivalent to say

$$\nabla\varphi(x) = \nabla\overline{\varphi}(x) \quad \mu - \text{a.e. } x \in \mathbb{R}^n,$$

i.e. that the gradient is uniquely determined almost everywhere (w.r.t. μ).

Remark 4.3. Notice that, in fact, we have shown not only the uniqueness of the optimal transportation plan $(\text{Id} \times \nabla\varphi)_\# \mu$, but also that there exists a unique gradient of a convex function $\nabla\varphi$ which pushes forward μ in ν .

Step 11

Finally (iii) follows easily from (ii). Indeed, to solve Monge optimal transportation problem means to find an optimal plan for the Kantorovich problem of the form

$$\pi = (\text{Id} \times T)_\# \mu,$$

for a suitable map T such that $T_\# \mu = \nu$. Hence, Brenier's theorem implies that $T = \nabla\varphi$ is a solution to the Monge problem and is unique.

Therefore, we have proved (iii) and we have finally completed the whole Theorem! $\diamond$

5 Restriction to the real line

In this Section we concentrate to the case $n = 1$: this allows to give a more explicit characterization of the optimal transportation plan than the general one given in Theorem 2.3.

First of all, there are the following one-to-one correspondences:

gradients of convex functions $\leftrightarrow$ monotone functions,

subdifferentials of convex functions $\leftrightarrow$ maximal monotone subsets of $\mathbb{R}^2$.

Let us consider now probability measures on $\mathbb{R}$. They enjoy a nice representation property: given any $\mu \in \mathcal{P}(\mathbb{R})$ we can define the *cumulative distribution function* associated to μ as

$$F_\mu(x) \doteq \int_{-\infty}^x d\mu = \mu([-\infty, x]). \quad (78)$$

F_μ is right continuous, nondecreasing and satisfies $F_\mu(-\infty) = 0$, $F_\mu(+\infty) = 1$. The opposite holds true as well: given any $\overline{F}: \mathbb{R} \rightarrow [0, 1]$ right continuous, nondecreasing and such that $\overline{F}(-\infty) = 0$, $\overline{F}(+\infty) = 1$, there exists a unique probability measure $\mu \in \mathcal{P}(\mathbb{R})$ such that $F_\mu = \overline{F}$. Indeed, $\mu([-\infty, x]) \doteq \overline{F}(x)$

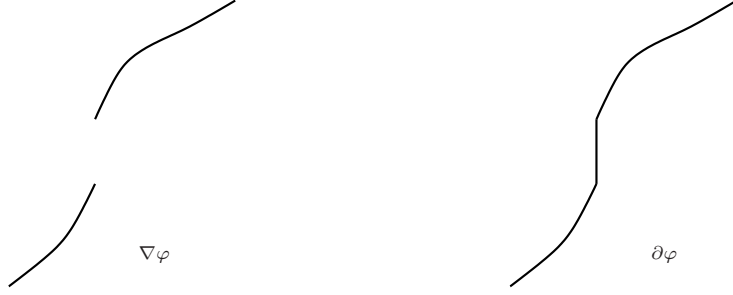


Figure 6: Gradients and subdifferential of convex functions in $\mathbb{R}$.

defines a unique Borel probability measure (because half-lines generates the whole Borel σ -algebra).

An interesting property of these cumulative distribution functions is that they admit *generalized inverses*, i.e. for all F we can define a (possibly discontinuous) map $F^{-1}: [0, 1] \rightarrow \mathbb{R}$ defined as follows

$$F^{-1}(t) \doteq \inf \{x \in \mathbb{R} ; F(x) > t\} . \quad (79)$$

Notice that F^{-1} is right continuous and it enjoys

$$\forall x \in \mathbb{R} \quad F^{-1}(F(x)) \geq x ,$$

$$\forall t \in [0, 1] \quad F(F^{-1}(t)) \geq t .$$

Indeed, whenever F is strictly monotone and continuous, there is no problem (F^{-1} is strictly monotone and continuous as well, and equality holds in both expressions).

If $F \equiv c$ on a given interval I , then F^{-1} has a jump in c between $\inf I = \lim_{\gamma \rightarrow c^-} F^{-1}(\gamma)$ and $\sup I = F^{-1}(c)$. In this case, $F(F^{-1}(c)) = F(\sup I) = c$, but for all $x' \in I$

$$F^{-1}(F(x')) = F^{-1}(c) = \sup I \geq x' .$$

On the other hand, if F has a jump in $\bar{x}$ between the values $c_1 < c_2$ (and $F(x) = c_2$ by right continuity), then $F^{-1}(\lambda c_1 + (1 - \lambda)c_2) = \bar{x}$ for all $\lambda \in [0, 1]$, i.e. F^{-1} is constant in $[c_1, c_2]$. In this case, $F^{-1}(F(\bar{x})) = F^{-1}(c_2) = \bar{x}$, but for all $\gamma \in [c_1, c_2]$

$$F(F^{-1}(\gamma)) = F(x) = c_2 \geq \gamma .$$

A similar representation can be given to measures in $\mathcal{P}(\mathbb{R}^2)$:

given $\pi \in \mathcal{P}(\mathbb{R}^2)$ and set

$$\mathcal{R}(\xi, \eta) \doteq \{(x, y) \in \mathbb{R}^2 ; x \leq \xi, y \leq \eta\} , \quad (80)$$

we define the *2d cumulative distribution function* associated to π as

$$H_\pi(x, y) \doteq \int_{\mathcal{R}(x, y)} d\pi . \quad (81)$$

viceversa, for any $\overline{H}: \mathbb{R}^2 \rightarrow \mathbb{R}$ such that $\overline{H}$ is nondecreasing and right continuous in both its variables, and its limits at $(-\infty, -\infty)$, $(+\infty, +\infty)$ are respectively 0 and 1, we find a unique measure $\pi \in \mathcal{P}(\mathbb{R}^2)$ such that $H_\pi = \overline{H}$ (because rectangles $\mathcal{R}$ generates the whole Borel σ -algebra in $\mathbb{R}^2$).

Finally, for probability measures on $\mathbb{R}^2$, there holds the following theorem due to Hoeffding and Fréchet (for more details on this theorem, check [16]).

Theorem 5.1. (Hoeffding–Fréchet Theorem) *Let $H: \mathbb{R}^2 \rightarrow \mathbb{R}$ be the 2d cumulative distribution of a probability measure π and let μ, ν probability measures on $\mathbb{R}$ with cumulative distributions F, G . Then $\pi \in \Pi(\mu, \nu)$ if and only if for all $(x, y) \in \mathbb{R}^2$*

$$F(x) + G(y) - 1 \leq H(x, y) \leq \min\{F(x), G(y)\}.$$

We can now state the characterization theorem in $\mathbb{R}$.

Theorem 5.2. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$ and let F, G be the cumulative distributions associated to them. Let $\pi \in \mathcal{P}(\mathbb{R}^2)$ be the probability measure associated to the 2-d cumulative distribution*

$$H(x, y) \doteq \min\{F(x), G(y)\}. \quad (82)$$

Then $\pi \in \Pi(\mu, \nu)$ and π is an optimal transportation plan for Kantorovich optimization problem with cost $c(x, y) = |x - y|^2$. Moreover, the optimal cost is given by

$$\mathcal{T}_c(\mu, \nu) \doteq \int_0^1 |F^{-1}(t) - G^{-1}(t)|^2 dt. \quad (83)$$

Proof. Notice that one could prove this Theorem without using Theorem 2.3. Nevertheless, we prefer to see it as a consequence of our main Theorem, to show the links between the general $\mathbb{R}^n$ case and this 1d setting.

a. First of all, due to the choice of H , $\pi \in \Pi(\mu, \nu)$ is a consequence of Theorem 5.1 (Hoeffding–Fréchet Theorem).

b. We claim that

$$\text{Supp}(\pi) \subset \{(x, y) \in \mathbb{R}^2 ; F(x^-) \leq G(y) \text{ and } G(y^-) \leq F(x)\}, \quad (84)$$

where $F(x^-), G(y^-)$ are the left limits of F, G in x, y respectively (by monotonicity these limits always exist).

Let (x, y) such that $F(x^-) > G(y)$, being the case $G(y^-) > F(x)$ similar (actually, the same up to a change of role between x and y). Fixed $\varepsilon \in \mathbb{R}$ such that $0 < 2\varepsilon < F(x^-) - G(y)$, we can exploit the right continuity and the monotonicity of G to find $\delta_1 > 0$ such that $G(y') < G(y) + \varepsilon$ for all $y' \in]y - \delta_1, y + \delta_1[$. Similarly, we can use the monotonicity of F and the definition of left limit to find $\delta_2 > 0$ such that $F(x') > F(x^-) - \varepsilon$ for all $x' \in]x - \delta_2, x + \delta_2[$. Setting $\delta = \min\{\delta_1, \delta_2\}$, we obtain

$$G(y') < G(y) + \varepsilon < F(x^-) - \varepsilon < F(x') \quad \forall (x', y') \in]x - \delta, x + \delta[\times]y - \delta, y + \delta[.$$

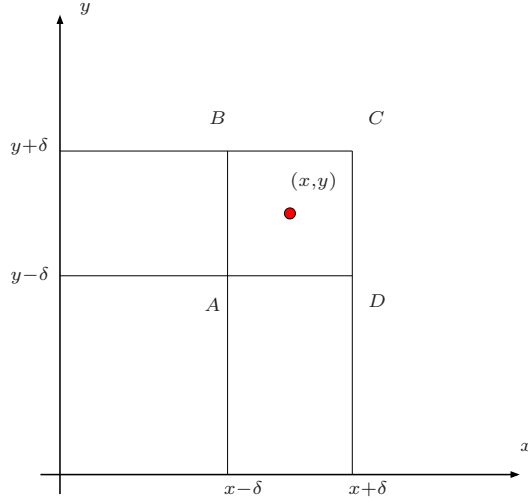


Figure 7: Rectangle with $H(x', y') = G(y')$.

So, $H(x', y') = \min\{F(x'), G(y')\} = G(y')$ and hence H does not depend on x' in a small square $\mathcal{S}$ around (x, y) . Now setting (as in Figure 7) $A = (x - \delta, y - \delta)$, $B = (x - \delta, y + \delta)$, $C = (x + \delta, y + \delta)$ and $D = (x + \delta, y - \delta)$, and using the definition of rectangles given in (80), we have

$$\begin{aligned}\pi(\mathcal{S}) &= \pi(\mathcal{R}_C) + \pi(\mathcal{R}_A) - \pi(\mathcal{R}_B) - \pi(\mathcal{R}_D) \\ &= H(C) + H(A) - H(B) - H(D) = 0,\end{aligned}$$

since H does not depend on x on $\mathcal{S}$. This proves that $(x, y) \notin \text{Supp}(\pi)$

c. Now we claim that $\text{Supp}(\pi)$ is a monotone subset of $\mathbb{R}^2$. Let $(x_1, y_1), (x_2, y_2)$ in $\text{Supp}(\pi)$ and assume $x_1 > x_2$, being the opposite case analogous. We have to prove $y_1 \geq y_2$. Since F is nondecreasing and (84) holds, we have

$$G(y_1) \geq F(x_1^-) \geq F(x_2) \geq G(y_2^-).$$

If $G(y_1) > G(y_2^-)$, we could conclude because also G is nondecreasing. Otherwise, all the quantities are equal, i.e.

$$G(y_1) = F(x_1^-) = F(x_2) = G(y_2^-).$$

Assume by contradiction that $y_2 > y_1$. This means that $F = G = \text{const.}$ on $[x_2, x_1[$ and $[y_1, y_2[$ respectively. Let us show that this would imply that $(x_1, y_1), (x_2, y_2) \notin \text{Supp}(\pi)$. Consider (x_2, y_2) , being the proof similar for (x_1, y_1) . Using monotonicity of F, G , their right continuity and the definition of left limit, one can find as in b. $\delta > 0$ small enough such that $H(x', y')$ does not depend on y' in the square $\mathcal{S}'$ with vertices (w, z) with $w \in \{x_2 - \delta, x_2 + \delta\}, z \in \{y_2 - \delta, y_2 + \delta\}$. This implies $\pi(\mathcal{S}') = 0$ and therefore $(x_2, y_2) \notin \text{Supp}(\pi)$.

d. Since $\text{Supp}(\pi)$ is a monotone subset of $\mathbb{R}^2$, i.e. the subdifferential of a convex function on $\mathbb{R}$, π has to be optimal by Theorem 2.3.

e. It remains to prove (83). We claim that

$$\pi = (F^{-1} \times G^{-1})_{\#} \mathcal{L}^1. \quad (85)$$

Of course, it is enough to prove it on rectangles $\mathcal{R}(x, y)$ for all x, y , i.e. for all x, y

$$\pi(\mathcal{R}(x, y)) = \mathcal{L}^1(\{t \in [0, 1] ; (F^{-1}(t), G^{-1}(t)) \in \mathcal{R}(x, y)\}).$$

Notice that the subset of $[0, 1]$ in the right hand side, is

$$\{t \in [0, 1] ; F^{-1}(t) \leq x\} \cap \{t \in [0, 1] ; G^{-1}(t) \leq y\},$$

and these sets coincides with either $[0, F(x)]$ or $[0, F(x))$ and with either $[0, G(y)]$ or $[0, G(y))$. In any case, the intersection will result an interval with endpoints 0 and $\min\{F(x), G(y)\}$, whose $\mathcal{L}^1$ measure is $\min\{F(x), G(y)\} = H(x, y)$. This proves our claim.

f. We can now conclude, since (85) implies that for any measurable function u on $\mathbb{R}^2$ we have

$$\int_{\mathbb{R}^2} u(x, y) d\pi(x, y) = \int_0^1 u(F^{-1}(t), G^{-1}(t)) dt.$$

Choosing $u = c$, we obtain (83). $\diamond$

Remark 5.3. In fact, the same π given in Theorem 5.2 is optimal not only for the quadratic distance, but for any cost $c(x, y)$ of the form $\gamma(x - y)$ with $\gamma: \mathbb{R} \rightarrow \mathbb{R}$ nonnegative, convex and symmetric. In this slightly more general case, the optimal cost takes the following expression

$$\mathcal{T}_\gamma(\mu, \nu) \doteq \int_0^1 \gamma(F^{-1}(t) - G^{-1}(t)) dt. \quad (86)$$

Remark 5.4. If additionally μ does not give mass to point, then $T = G^{-1} \circ F$ transports μ to ν and solves Monge problem. In particular, notice that we have

$$\int_{-\infty}^x d\mu = \int_{-\infty}^{T(x)} d\nu, \quad (87)$$

i.e. optimal plans are so called monotone rearrangements. Notice also that, whenever a jump occurs in G , we have an interval on which T is constant, i.e. there is a point in which we have to keep putting mass before moving on.

Remark 5.5. If $\mu, \nu \ll \mathcal{L}^1$ with densities f, g continuous and with $g > 0$, then the optimal plan for Monge problem is of class $\mathbf{C}^1$ and, deriving w.r.t. x in (87), we obtain that T satisfies

$$f(x) = g(T(x)) T'(x). \quad (88)$$

6 Proof of Main Results, III

Here we give a partial proof to Theorem 2.5. The missing details can be found in [13] and, partially, in [19].

First of all, we need the following Lemma.

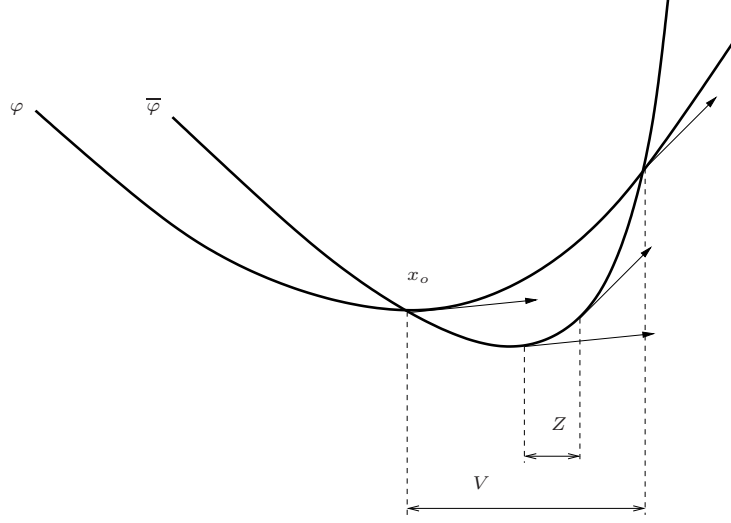


Figure 8: Representation of Aleksandrov's Lemma in a neighborhood of x_o .

Lemma 6.1. (Aleksandrov's Lemma) *Let φ and $\bar{\varphi}$ be two convex functions such that $\varphi(x_o) = \bar{\varphi}(x_o)$ but $\nabla\varphi(x_o) \neq \nabla\bar{\varphi}(x_o)$, for some $x_o \in \mathbb{R}^n$. Let*

$$V \doteq \{x \in \mathbb{R}^n ; \varphi(x) > \bar{\varphi}(x)\} \quad Z \doteq \nabla\bar{\varphi}^{-1}(\nabla\varphi(V)) .$$

Then $x_o \in \bar{V}$, $Z \subset V$ but Z lies a positive distance away from x_o .

A complete proof of this Lemma can be found in [13]. Here we only try to give an idea of the situation through Figure 8.

Finally, we can sketch a proof to Theorem 2.5.

Proof of Theorem 2.5 (sketch)

(i) One can prove that any optimal transportation plan π has support which is cyclically monotone (w.r.t. $c(x, y) = |x - y|^2$). We refer to [13] for a proof of this fact.

Then, the existence of a φ with the required properties follows from Theorem 1.54.

(ii) a. If φ in (i) is in $\mathbf{L}^1(\mu)$, we can simply repeat Step 8 from the proof of Theorem 2.3 to obtain that $\nabla\varphi$ transports μ in ν . Otherwise, one can still prove (see [13]) that

$$\pi(\{x \in \mathbb{R}^n ; \varphi \text{ is differentiable in } x\}) = 1$$

and then repeat the second part of Step 8, as in the other case. This complete the proof of the existence.

b. To prove uniqueness, assume by contradiction there exist two functions $\varphi, \bar{\varphi}$ which transport μ in ν and whose gradients do not coincide on $\text{Supp}(\mu)$. In

particular, let x_o such that $\nabla\varphi(x_o) \neq \nabla\bar{\varphi}(x_o)$ and assume, without loss of generality¹⁸, that $\varphi(x_o) = \bar{\varphi}(x_o)$.

Claim 1. In any neighborhood U of x_o , the set in which $\varphi = \bar{\varphi}$ has finite $\mathcal{H}^{n-1}$ measure.

A proof of this Claim, as well as a sort of implicit function theorem for convex map, can be found in [13].

Hence, any neighborhood U of x_o intersects either $\{\varphi > \bar{\varphi}\}$ or $\{\varphi < \bar{\varphi}\}$ in a set of nonzero measure μ . Let assume it is the former (otherwise, simply change φ and $\bar{\varphi}$ in the following). By Lemma 6.1, x_o lies a positive distance away from Z and, therefore, we have

$$\mu(Z) < \mu(V),$$

where we are using V, Z as defined in Lemma 6.1. We claim now that this leads to a contradiction. Indeed,

$$\begin{aligned} \nabla\bar{\varphi}_\# \mu(\nabla\varphi(V)) &= \mu(\nabla\bar{\varphi}^{-1}(\nabla\varphi(V))) = \mu(Z) \\ &< \mu(V) \leq \mu(\nabla\varphi^{-1}(\nabla\varphi(V))) = \nabla\varphi_\# \mu(\nabla\varphi(V)) \end{aligned}$$

and, therefore, $\nabla\bar{\varphi}_\# \mu \neq \nabla\varphi_\# \mu$, against our assumptions. $\diamond$

¹⁸Recall that, for all constants $c \in \mathbb{R}$ and all φ proper, convex and lower semicontinuous, $\partial\varphi = \partial(\varphi + c)$, i.e. φ is determined by its subdifferential up to an additive constant.

Appendix

A.1 Comments about Prokhorov's Theorem (by X. Raynaud)

Reading Villani's book [19]. and lecture notes [20], you could find two different versions of Prokhorov's Theorem. The first version is the one that can be found in Wikipedia and in [19] (and coincides with our Theorem 1.14).

Theorem A.1. (Prokhorov's Theorem, first version) *Let X be a Polish space. A subset S of $\mathcal{P}(X)$ is relatively compact in $\mathcal{P}(X)$ if and only if S is tight, that is, for all $\varepsilon > 0$, there exists a compact K_ε such that*

$$\mu(X \setminus K_\varepsilon) \leq \varepsilon$$

for all $\mu \in S$.

The second version (with a proof) comes from [20] (p.278):

Theorem A.2. (Prokhorov's Theorem, second version) *Let X be a Polish space and S a set of positive measures. Then, S is relatively compact for the topology $\sigma(\mathcal{M}(X), \mathbf{C}_b(X))$ if and only if*

- (i) S is bounded, i.e., there exists $C > 0$ such that $\|\mu\|_{\mathcal{M}(X)} \leq C$ for all $\mu \in S$,
- (ii) S is tight.

We can consider **two weak topologies** in the space of Radon measure $\mathcal{M}(X)$, namely $\sigma(\mathcal{M}(X), \mathbf{C}_0(X))$ and $\sigma(\mathcal{M}(X), \mathbf{C}_b(X))$. A sequence μ_n converges to μ in $\sigma(\mathcal{M}(X), \mathbf{C}_0(X))$, and respectively in $\sigma(\mathcal{M}(X), \mathbf{C}_b(X))$, if

$$\int_X \phi d\mu_n \rightarrow \int_X \phi d\mu \quad (89)$$

for all $\phi \in \mathbf{C}_c(X)$, respectively $\phi \in \mathbf{C}_b(X)$. Here, $\mathbf{C}_c(X)$ and $\mathbf{C}_b(X)$ denote respectively the space of continuous functions with compact support and the space of bounded continuous function. The space $\mathbf{C}_0(X)$ is the closure of $\mathbf{C}_c(X)$ in $\mathbf{L}^\infty(X)$.

The topology $\sigma(\mathcal{M}(X), \mathbf{C}_b(X))$ is **strictly stronger** than $\sigma(\mathcal{M}(X), \mathbf{C}_0(X))$ (and a sequence that converges in $\sigma(\mathcal{M}(X), \mathbf{C}_b(X))$ automatically converges also in $\sigma(\mathcal{M}(X), \mathbf{C}_c(X))$).

The space of probability measure $P(X)$ is a **closed** subset of $\mathcal{M}(X)$ with respect to $\sigma(\mathcal{M}(X), \mathbf{C}_b(X))$ but not with respect $\sigma(\mathcal{M}(X), \mathbf{C}_0(X))$. Let us prove this statement. We consider a sequence μ_n in $P(X)$ that converges to μ in $\sigma(\mathcal{M}(X), \mathbf{C}_b(X))$. We want to prove that $\mu \in P(X)$. Assume that we have proven that μ is positive¹⁹, then it only remains to prove that $\mu(X) = 1$. This follows from (89) after taking $\phi = 1$ since $\int_X d\mu_n = \mu_n(X) = 1$ for all n and $\int_X d\mu_n \rightarrow \int_X d\mu = \mu(X)$. Hence, $P(X)$ is closed with respect

¹⁹This actually requires a little bit of work. A very rough sketch of the proof: Let A be a set such that $\mu(A) < 0$. The regularity of μ implies that A can be approached in measure by a compact K and by an open O . If X is LCH, the conclusion follows from Uryshon's lemma which gives the existence of a continuous function ϕ such that $1_K \leq \phi \leq 1_O$. If X is a Polish space, one construct a continuous function which approaches the characteristic function 1_K of K by using the metric (The function $x \mapsto d(x, K)$ is continuous and $d(x, K) = 0$ if and only if $x \in K$).

to $\sigma(\mathcal{M}(X), \mathbf{C}_b(X))$. Now consider the sequence of probability measure δ_n (Dirac mass located at $x = n$). One can check that $\delta_n \rightarrow 0$ with respect to $\sigma(\mathcal{M}(X), \mathbf{C}_0(X))$ but 0 is not a probability measure and therefore $P(X)$ is not closed with respect to $\sigma(\mathcal{M}(X), \mathbf{C}_0(X))$. Note that the same sequence is not convergent (and none of its subsequence either) in $\sigma(\mathcal{M}(X), \mathbf{C}_b(X))$.

In the case of a **locally compact Hausdorff** (LCH) space, if we have a **tight** family (or sequence), then the **two topologies coincide**. Let us prove that. Let μ_n be a tight sequence that converges to μ in $\sigma(\mathcal{M}(X), \mathbf{C}_0(X))$ and ϕ a given function in $\mathbf{C}_b(X)$. Since the sequence is tight, for all $\varepsilon > 0$, there exists a compact K_ε such that $\mu_n(X \setminus K_\varepsilon) < \varepsilon$ for all n and $\mu(X \setminus K_\varepsilon) < \varepsilon$. By Urysohn's lemma (which can be applied because X is LCH), there exists ψ with compact support such that $0 \leq \psi \leq 1$ and $\psi|_{K_\varepsilon} = 1$. Then,

$$\begin{aligned} \left| \int_X \phi d\mu_n - \int_X \phi d\mu \right| &\leq \left| \int_X (1 - \psi)\phi d\mu_n \right| + \left| \int_X (1 - \psi)\phi d\mu \right| \\ &\quad + \left| \int_X \psi\phi d\mu_n - \int_X \psi\phi d\mu \right| \\ &\leq \|(1 - \psi)\phi\|_{\mathbf{L}^\infty} (\mu_n(X \setminus K_\varepsilon) + \mu(X \setminus K_\varepsilon)) \\ &\quad + \left| \int_X \psi\phi d\mu_n - \int_X \psi\phi d\mu \right| \\ &\leq 2\|\phi\|_{\mathbf{L}^\infty} \varepsilon + \left| \int_X \psi\phi d\mu_n - \int_X \psi\phi d\mu \right|. \end{aligned}$$

Since $\psi\phi$ has compact support and $\mu_n \rightarrow \mu$ in $\sigma(\mathcal{M}(X), \mathbf{C}_0(X))$, the second term in the inequality above tend to zero. The first one can be made as small as wanted. Hence, $\int_X \phi d\mu_n \rightarrow \int_X \phi d\mu$ and we have proved that $\mu_n \rightarrow \mu$ in $\sigma(\mathcal{M}(X), \mathbf{C}_b(X))$.

We can now prove one implication in the Prokhorov's theorem (tightness implies compactness) for a LCH space. By the Riesz representation theorem (which holds for LCH space), we have $\mathbf{C}_0(X)^* = \mathcal{M}(X)$. From Banach–Alaouglou theorem, any bounded set in $\mathcal{M}(X)$ is weakly-* compact so that any bounded sequence μ_n has a convergent subsequence μ_{n_i} in $\sigma(\mathcal{M}(X), \mathbf{C}_0(X))$. Since, for tight sets, convergence in $\sigma(\mathcal{M}(X), \mathbf{C}_0(X))$ implies convergence in $\sigma(\mathcal{M}(X), \mathbf{C}_b(X))$, the subsequence μ_{n_i} also converges in $\sigma(\mathcal{M}(X), \mathbf{C}_b(X))$ and we have proved the second version of Prokhorov's theorem. The first version now follows from the second as we have shown that $P(X)$ is closed with respect to the $\sigma(\mathcal{M}(X), \mathbf{C}_b(X))$ topology.

For **Polish spaces**, there are cases (which have applications! For example $C([0, 1], \mathbb{R}^n)$ for which $\mathbf{C}_0(X) = \{0\}$). Then, the $\sigma(\mathcal{M}(X), \mathbf{C}_0(X))$ topology is just the trivial topology (which consists of the two open sets $\{\emptyset, \mathcal{M}(X)\}$)! It means in practice that we will always consider the $\sigma(\mathcal{M}(X), \mathbf{C}_0(X))$ topology when working with Polish spaces. The proof of Prokhorov is a little bit more involved. One still uses the Riesz representation theorem and Banach–Alaouglou but repeatedly on a sequence of compact subset of X , see Villani's lecture note.

A.2 Proof of Proposition 1.38

Let $\varphi: X \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex map. We have seen in Proposition 1.23 that this is equivalent to say $\text{epi}(\varphi)$ is a convex set. Now we add a couple of

ingredients.

Proposition A.3. *Let X a topological vector space and $\varphi: X \rightarrow \mathbb{R} \cup \{+\infty\}$ be a map (not necessarily convex). Then φ is lower semicontinuous if and only if $\text{epi}(\varphi)$ is a closed set in $X \times \mathbb{R}$.*

Definition A.4. *A set $E \subset X \times \mathbb{R}$ is an epigraph if*

$$(x, \alpha) \in E \quad \Rightarrow \quad (x, \beta) \in E \quad \text{for all } \beta \geq \alpha.$$

If this is the case, we will have $E = \text{epi}(\phi)$ where, for all $x \in X$,

$$\phi(x) \doteq \inf \{ \alpha \in \mathbb{R} ; (x, \alpha) \in E \}.$$

Definition A.5. *Let $\varphi: X \rightarrow \mathbb{R} \cup \{+\infty\}$ be a map. We define:*

(i) *the lower semicontinuous regularization of φ as the map $\text{sc}[\varphi]$ defined by*

$$\text{epi}(\text{sc}[\varphi]) = \overline{\text{epi}(\varphi)},$$

(ii) *the convex lower semicontinuous regularization of φ as the map $\text{csc}[\varphi]$ defined by*

$$\text{epi}(\text{csc}[\varphi]) = \overline{\text{conv}(\text{epi}(\varphi))},$$

where $\text{conv}(A)$ is the convex hull of the set A .

Remark A.6. Since $\text{epi}(\varphi)$ is an epigraph in the sense of Definition A.4, also its closure and the closure of its convex hull are epigraph, and therefore the regularizations are well defined. Moreover, by Proposition A.3 and Proposition 1.23, they are respectively lower semicontinuous and convex lower semicontinuous as their names suggest. Finally, since the closure of a set A (respectively its convex hull) is the smallest closed (resp. convex) set containing A , $\text{sc}[\varphi]$ (resp. $\text{csc}[\varphi]$) is the largest lower semicontinuous (resp. convex and lower semicontinuous) function smaller than φ .

Lemma A.7. *Let X be a normed space and $\varphi: X \rightarrow \mathbb{R} \cup \{+\infty\}$ be a map. If there exists $\ell(x)$ affine such that $\varphi(x) \geq \ell(x)$ for all $x \in X$, then $\text{csc}[\varphi] = \varphi^{**}$*

Proof. If φ is not proper, then $\varphi = \text{csc}[\varphi] = \varphi^{**} \equiv \infty$ and we have concluded. Otherwise, it is clear that $\text{csc}[\varphi] \geq \varphi^{**}$, since φ^{**} is convex and lower semicontinuous. Moreover, applying twice Proposition 1.37-(i), for all $x \in X$

$$\varphi^{**}(x) \geq \ell^{**}(x) = \ell(x),$$

and hence we have $\varphi^{**}(x) > -\infty$. Notice that the last equality above follows directly from the definition of Fenchel conjugate for all affine functions.

By contradiction, assume now that there exists $\bar{x} \in X$ such that $\text{csc}[\varphi](\bar{x}) > \varphi^{**}(\bar{x})$. Assume also $\text{csc}[\varphi](\bar{x}) < \infty$. Then, using Theorem 1.8 (Hahn–Banach separation theorem), we can strictly separate the compact set $\{(\bar{x}, \varphi^{**}(\bar{x}))\}$ and the closed convex set $C \doteq \text{epi}(\text{csc}[\varphi])$. This means that there exists $\xi \in X^* \setminus \{0\}$ and $r, c \in \mathbb{R}$ such that for all $(x, \alpha) \in C$

$$\langle \xi, x \rangle + r \alpha > c > \langle \xi, \bar{x} \rangle + r \varphi^{**}(\bar{x}).$$

In particular, considering $\alpha = \text{csc}[\varphi](x)$ we can prove that $r > 0$ and then we have for all $x \in X$

$$\langle \xi', \bar{x} \rangle + \varphi^{**}(\bar{x}) < \langle \xi', x \rangle + \text{csc}[\varphi](x) \leq \langle \xi', x \rangle + \varphi(x),$$

where we set $\xi' = \frac{\xi}{r}$. This is equivalent to

$$\begin{aligned} \varphi^{**}(\bar{x}) &< \langle -\xi', \bar{x} \rangle - \sup_{x \in X} \{ \langle -\xi', x \rangle - \varphi(x) \} \\ &= \langle -\xi', \bar{x} \rangle - \varphi^*(-\xi'), \end{aligned}$$

which is a contradiction with the definition of φ^{**} .

It remains to deal with the case of $\varphi^{**}(\bar{x}) < \text{csc}[\varphi](\bar{x}) = \infty$. In this case, the separation still holds true, but we cannot prove $r > 0$, i.e. the hyperplane could be vertical. If $r \neq 0$, we can proceed as above to obtain a contradiction. On the other hand, if $r = 0$, the separation could be written as: there exists $\xi \in X^* \setminus \{0\}$ and $c \in \mathbb{R}$ such that

$$\langle \xi, x \rangle - c < 0 \quad \forall x \in \text{Dom}(\varphi), \quad \langle \xi, \bar{x} \rangle - c > 0.$$

Hence, for all $x \in X$ and $t > 0$

$$\varphi(x) > \ell(x) + t (\langle \xi, x \rangle - c),$$

since the inequality holds true (trivially) also outside $\text{Dom}(\varphi)$. Conjugating twice and evaluating the expression in $x = \bar{x}$, we obtain for all $t > 0$

$$\varphi^{**}(\bar{x}) \geq \ell(\bar{x}) + t (\langle \xi, \bar{x} \rangle - c).$$

Therefore, it has to be $\varphi^{**}(\bar{x}) = +\infty = \text{csc}[\varphi](\bar{x})$, a contradiction. $\diamond$

We are now ready to prove the last result we needed from convex analysis.

Proof of Proposition 1.38. It is clear that (iii) implies (ii) and that (ii) implies (i). It remains to be proved that (i) implies (iii). So, let φ be proper, convex and lower semicontinuous. Then $\varphi = \text{csc}[\varphi]$.

Let x such that $\varphi(x) < \infty$. Then, for all $\alpha < \varphi(x)$ we can use Hahn–Banach Theorem in $X \times \mathbb{R}$ to strictly separate the compact set $\{(x, \alpha)\}$ from the convex and closed set $\text{epi}(\varphi)$ by means of an hyperplane. Since this hyperplane provides an affine lower bound to φ , we can apply Lemma A.7 to prove that in fact $\varphi = \text{csc}[\varphi] = \varphi^{**}$. This concludes the proof. $\diamond$

Bibliography

- [1] G. Alberti & L. Ambrosio, *A geometrical approach to monotone functions in R^n* . Math. Z. **230** (1999), no. 2, 259–316.
- [2] L. Ambrosio, N. Fusco & D. Pallara, Functions of bounded variation and free discontinuity problems. *Oxford Mathematical Monographs*. The Clarendon Press, Oxford University Press, New York, 2000.
- [3] L. Ambrosio & A. Pratelli, Existence and stability results in the L^1 theory of optimal transportation. *Optimal transportation and applications (Martina Franca, 2001)*. Lecture Notes in Math., 1813. Springer, Berlin, 2003.
- [4] P. Billingsley, Convergence of probability measures. Second edition. *Wiley Series in Probability and Statistics: Probability and Statistics*. Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1999.
- [5] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*. Comm. Pure Appl. Math. **44** (1991), no. 4, 375–417.
- [6] H. Brezis, Analyse fonctionnelle. (French) *Theorie et applications. Collection Mathematiques Appliquees pour la Maitrise*. Masson, Paris, 1983.
- [7] L. Evans & R. F. Gariepy, Measure theory and fine properties of functions. *Studies in Advanced Mathematics*. CRC Press, Boca Raton, FL, 1992.
- [8] H. Federer, *Curvature measures*. Trans. Amer. Math. Soc. **93** (1959), 418–491.
- [9] L. V. Kantorovich, *On the Transfer of Masses*. Dokl. Akad. Nauk. SSSR **37** (1942), 227–229.
- [10] L. V. Kantorovich, *On a Problem of Monge*. Uspekhi Mat. Nauk. **3** (1948), 225–226.
- [11] P. D. Lax, Functional analysis. *Pure and Applied Mathematics*. Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 2002.
- [12] R. Lucchetti, Convexity and well-posed problems. *CMS Books in Mathematics/Ouvrages de Mathematiques de la SMC*, 22. Springer, New York, 2006.
- [13] R. J. McCann, *Existence and uniqueness of monotone measure-preserving maps*. Duke Math. J. **80** (1995), no. 2, 309–323.

- [14] G. Monge, *Memoire sur la Theorie des Déblais et des Remblais*. Histoire de l'Académie des Sciences de Paris, 1781.
- [15] A. Pratelli, *On the sufficiency of c -cyclical monotonicity for optimality of transport plans*. Math. Z., to appear (2007). Available at <http://cvgmt.sns.it/cgi/get.cgi/people/pratelli/>
- [16] S. T. Rachev & L. Rschendorf, Mass transportation problems. Vol. I. Theory. & Vol. II. Applications. *Probability and its Applications*. Springer-Verlag, New York, 1998.
- [17] R. T. Rockafellar, Convex analysis. *Princeton Mathematical Series*, 28. Princeton University Press, Princeton, NJ, 1970.
- [18] C. S. Smith & M. Knott, *Note on the optimal transportation of distributions*. J. Optim. Theory Appl. **52** (1987), no. 2, 323–329.
- [19] C. Villani, Topics in optimal transportation. *Graduate Studies in Mathematics*, 58. American Mathematical Society, Providence, RI, 2003.
- [20] C. Villani, Optimal Transport, Old and New. *Lecture notes from a course in Saint-Flour, expanded version*. Available at <http://www.umpa.ens-lyon.fr/~cvillani/surveys.html>