Chapter 6

The Optimal Transportation Problem

Let X, Y be separable metric spaces such that any Borel probability measure in X, Y is tight (5.1.9), i.e. Radon spaces, according to Definition 5.1.4, and let $c : X \times Y \to [0, +\infty]$ be a Borel cost function. Given $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$ the optimal transport problem, in Monge's formulation, is given by

$$\inf \left\{ \int_X c(x, \boldsymbol{t}(x))\, d\mu(x) : \ \boldsymbol{t}_{\#}\mu = \nu \right\}. \tag{6.0.1}$$

This problem can be ill posed because sometimes there is no transport map $\boldsymbol{t}$ such that $\boldsymbol{t}_{\#}\mu = \nu$ (this happens for instance when μ is a Dirac mass and ν is not a Dirac mass). Kantorovich's formulation

$$\min \left\{ \int_{X \times Y} c(x, y)\, d\gamma(x, y) : \ \gamma \in \Gamma(\mu, \nu) \right\} \tag{6.0.2}$$

circumvents this problem (as $\mu \times \nu \in \Gamma(\mu, \nu)$). The existence of an optimal transport plan, when c is l.s.c., is provided by (5.1.15) and by the tightness of $\Gamma(\mu, \nu)$ (this property is equivalent to the tightness of μ, ν, a property always guaranteed in Radon spaces).

The problem (6.0.2) is truly a weak formulation of (6.0.1) in the following sense: if c is bounded and continuous, and if μ has no atom, then the "min" in (6.0.2) is equal to the "inf" in (6.0.1), see [81], [9]. This result can also be extended to unbounded cost functions, under the assumption (6.1.8), see [128].

In some special situations one can directly show the existence of optimal transport maps without any assumption on the cost function (besides positivity and lower semicontinuity).

Theorem 6.0.1 (Birkhoff theorem). *Let C be the convex set of all doubly stochastic $N\times N$ matrices, i.e. those matrices M whose entries M_{ij} satisfy*

$$\sum_{i=1}^{N} M_{ij}=\frac{1}{N}\quad \forall j=1,\ldots,N,\qquad \sum_{j=1}^{N} M_{ij}=\frac{1}{N}\quad \forall i=1,\ldots,N.$$

Then, the extreme points of C are permutation matrices, i.e. those matrices of the form

$$M_{ij}=\frac{1}{N}\delta_{i\sigma(j)}\quad \text{for some permutation } \sigma \text{ of } \{1,\ldots,N\}.$$

In particular, if μ (resp. ν) can be represented as the sum of N Dirac masses in distinct points x_i (resp. distinct points y_j) with weight $1/N$, then the minimum in (6.0.2) *is always provided by a transport map.*

Proof. For a proof the first statement see, for instance, the simple argument at the end of the introduction of [146].
The convex set $\Gamma(\mu,\nu)$ can be canonically identified with C, writing $\boldsymbol{\mu}_{ij}=\boldsymbol{\mu}(\{x_i\}\times\{y_j\})$, and transport maps correspond to permutation matrices. Since the energy functional is linear on $\Gamma(\mu,\nu)$, the minimum is surely attained on a extreme point of $\Gamma(\mu,\nu)$ and therefore on a transport map. □

Another special occasion occurs when $X=Y=\mathbb{R}$. In this case we can use the distribution function

$$F_\mu(t):=\mu\left((-\infty,t)\right)\qquad t\in\mathbb{R}$$

to characterize optimal transport maps and to give a simple formula for the minimum value in (6.0.2). We need to define also an inverse of F_μ, by the formula (notice that a priori F_μ need not be continuous or strictly increasing)

$$F_\mu^{-1}(s):=\sup\{x\in\mathbb{R}:\ F_\mu(x)\le s\}\qquad s\in[0,1].$$

Theorem 6.0.2 (Optimal transportation in $\mathbb{R}$). *Let μ, $\nu\in\mathscr{P}_p(\mathbb{R})$ and let $c(x,y)=h(x-y)$, with $h\ge 0$ convex and with p growth.*

(i) *If μ has no atom, i.e. F_μ is continuous, then $F_\nu^{-1}\circ F_\mu$ is an optimal transport map. It is the unique optimal transport map if h is strictly convex.*

(ii) *We have*

$$\min\left\{\int_{\mathbb{R}^2} c(x,y)\,d\boldsymbol{\gamma}:\ \boldsymbol{\gamma}\in\Gamma(\mu,\nu)\right\}=\int_0^1 c\left(F_\mu^{-1}(s),F_\nu^{-1}(s)\right)\,ds. \tag{6.0.3}$$

Proof. For the proof of the first statement see for instance [146], [82].
(ii) In this proof we use the following two elementary properties of the distribution function when μ has no atom: first, $F_{\mu\#}\mu=\chi_{(0,1)}\mathscr{L}^1$ (this fact can be checked in an elementary way on intervals and we omit the argument), second

$F_\mu^{-1} \circ F_\mu(x) = x$ for μ-a.e. x. The second property simply follows by the observation that the (maximal) open intervals in which F_μ is constant correspond, by the very definition of F_μ, to intervals where μ has no mass. Using statement (i) we have then

$$\begin{aligned}\int_{\mathbb{R}} c\left(x, F_\nu^{-1} \circ F_\mu(x)\right) d\mu(x) &= \int_{\mathbb{R}} c\left(F_\mu^{-1} \circ F_\mu(x), F_\nu^{-1} \circ F_\mu(x)\right) d\mu(x) \\ &= \int_0^1 c\left(F_\mu^{-1}(s), F_\nu^{-1}(s)\right) ds,\end{aligned}$$

in the case when μ has no atom. The general case can be achieved through a simple approximation. □

6.1 Optimality conditions

In this section we discuss the optimality conditions in the variational problem (6.0.2), assuming always that $c : X \times Y \to [0, +\infty]$ is a proper l.s.c. function.

Theorem 6.1.1 (Duality formula). *The minimum of the Kantorovich problem* (6.0.2) *is equal to*

$$\sup \left\{ \int_X \varphi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y) \right\} \tag{6.1.1}$$

where the supremum runs among all pairs $(\varphi, \psi) \in C_b^0(X) \times C_b^0(Y)$ *such that* $\varphi(x) + \psi(y) \le c(x, y)$.

Proof. This identity is well-known if c is bounded and continuous, see for instance [104, 129, 146]. A possible strategy is to show first that the support of any optimal plan is a c-monotone set, according to Definition 6.1.3 below, and than use this fact to build a maximizing pair (we will give this construction in Theorem 6.1.4 below, under more general assumptions on c).

In the general case it suffices to approximate c from below by an increasing sequence of bounded continuous functions c_h, defined for instance by (compare with (5.1.4))

$$c_h(x, y) := \inf_{(x', y') \in X \times Y} \left\{ c(x', y') \wedge h + h d_X(x, x') + h d_Y(y, y') \right\},$$

noticing that a simple compactness argument gives

$$\min \left\{ \int_{X \times Y} c_h\, d\boldsymbol{\gamma} : \ \boldsymbol{\gamma} \in \Gamma(\mu, \nu) \right\} \quad \uparrow \quad \min \left\{ \int_{X \times Y} c\, d\boldsymbol{\gamma} : \ \boldsymbol{\gamma} \in \Gamma(\mu, \nu) \right\}$$

and that any pair (φ, ψ) such that $\varphi + \psi \le c_h$ is admissible in (6.1.1). □

We recall briefly the definitions of c-transform, c-concavity and c-monotonicity, referring to the papers [68], [82] and to the book [129] for a more detailed analysis.

Definition 6.1.2 (c-transform, c-concavity). (1) *For $u : X \to \overline{\mathbb{R}}$, the* c-transform *$u^c : Y \to \overline{\mathbb{R}}$ is defined by*

$$u^c(y) := \inf_{x \in X} c(x, y) - u(x)$$

with the convention that the sum is $+\infty$ whenever $c(x, y) = +\infty$ and $u(x) = +\infty$. Analogously, for $v : Y \to \overline{\mathbb{R}}$, the c-transform *$v^c : X \to \overline{\mathbb{R}}$ is defined by*

$$v^c(x) := \inf_{y \in Y} c(x, y) - v(y)$$

with the same convention when an indetermination of the sum is present.
(2) *We say that $u : X \to \overline{\mathbb{R}}$ is* c-concave *if $u = v^c$ for some v; equivalently, u is c-concave if there is some family $\{(y_i, t_i)\}_{i \in I} \subset Y \times \overline{\mathbb{R}}$ such that*

$$u(x) = \inf_{i \in I} c(x, y_i) + t_i \qquad \forall x \in X. \tag{6.1.2}$$

An analogous definition can be given for functions $v : Y \to \overline{\mathbb{R}}$.

It is not hard to show that $u^{cc} \geq u$ and that equality holds if and only if u is c-concave. Analogously, $v^{cc} \geq v$ and equality holds if and only if v is c-concave.
Let us also introduce the concept of c-monotonicity.

Definition 6.1.3 (c-monotonicity). *We say that $\Gamma \subset X \times Y$ is* c-monotone *if*

$$\sum_{i=1}^{n} c(x_i, y_{\sigma(i)}) \geq \sum_{i=1}^{n} c(x_i, y_i)$$

whenever $(x_1, y_1), \ldots, (x_n, y_n) \in \Gamma$ and σ is a permutation of $\{1, \ldots, n\}$.

With these definitions we can prove the following result concerning necessary and sufficient optimality conditions and the existence of maximizing pairs (φ, ψ) in (6.1.1). The proof is taken from [16], see also [146], [82], [129] for similar results. Notice also that conditions (6.1.3) and (6.1.4) do not apply to the cost functions considered in [79, 80, 101], in a infinite-dimensional framework.

Theorem 6.1.4 (Necessary and sufficient optimality conditions).
(Necessity) If $\boldsymbol{\gamma} \in \Gamma(\mu, \nu)$ is optimal and $\int_{X \times Y} c \, d\boldsymbol{\gamma} < +\infty$, then $\boldsymbol{\gamma}$ is concentrated on a c-monotone Borel subset of $X \times Y$. Moreover, if c is continuous, then $\operatorname{supp} \boldsymbol{\gamma}$ *is c-monotone.*
(Sufficiency) Assume that c is real-valued, $\boldsymbol{\gamma} \in \Gamma(\mu, \nu)$ is concentrated on a c-monotone Borel subset of $X \times Y$, and

$$\mu\left(\left\{x \in X : \int_Y c(x, y) \, d\nu(y) < +\infty\right\}\right) > 0, \tag{6.1.3}$$

$$\nu\left(\left\{y \in Y : \int_X c(x,y)\, d\mu(x) < +\infty\right\}\right) > 0. \tag{6.1.4}$$

Then $\boldsymbol{\gamma}$ *is optimal,* $\int_{X\times Y} c\, d\boldsymbol{\gamma} < +\infty$ *and there exists a maximizing pair* (φ, ψ) *in* (6.1.1) *with* φ *c-concave and* $\psi = \varphi^c$.

Proof. Let (φ_n, ψ_n) be a maximizing sequence in (6.1.1) and let $c_n = c - \varphi_n - \psi_n$. Since

$$\int_{X\times Y} c_n\, d\gamma = \int_{X\times Y} c\, d\boldsymbol{\gamma} - \int_X \varphi_n\, d\mu - \int_Y \psi_n\, d\nu \to 0$$

and $c_n \geq 0$ we can find a subsequence $c_{n(k)}$ and a Borel set Γ on which $\boldsymbol{\gamma}$ is concentrated and c is finite, such that $c_{n(k)} \to 0$ on Γ. If $\{(x_i, y_i)\}_{1\leq i\leq p} \subset \Gamma$ and σ is a permutation of $\{1, \ldots, p\}$ we get

$$\begin{aligned}\sum_{i=1}^p c(x_i, y_{\sigma(i)}) &\geq \sum_{i=1}^p \varphi_{n(k)}(x_i) + \psi_{n(k)}(y_{\sigma(i)}) \\ &= \sum_{i=1}^p \varphi_{n(k)}(x_i) + \psi_{n(k)}(y_i) = \sum_{i=1}^p c(x_i, y_i) - c_{n(k)}(x_i, y_i)\end{aligned}$$

for any k. Letting $k \to \infty$ the c-monotonicity of Γ follows.

Now we show the converse implication, assuming that (6.1.3) and (6.1.4) hold. We denote by Γ a Borel and c-monotone set on which $\boldsymbol{\gamma}$ is concentrated; without loss of generality we can assume that $\Gamma = \cup_k \Gamma_k$ with Γ_k compact and $c|_{\Gamma_k}$ continuous. We choose continuous functions c_l such that $c_l \uparrow c$ and split the proof in several steps.

Step 1. There exists a c-concave Borel function $\varphi : X \to [-\infty, +\infty)$ such that $\varphi(x) > -\infty$ for μ-a.e. $x \in X$ and

$$\varphi(x') \leq \varphi(x) + c(x', y) - c(x, y) \qquad \forall x' \in X,\ (x, y) \in \Gamma. \tag{6.1.5}$$

To this aim, we use the explicit construction given in the generalized Rockafellar theorem in [134], setting

$$\begin{aligned}\varphi(x) := \inf\{&c(x, y_p) - c(x_p, y_p) + c(x_p, y_{p-1}) - c(x_{p-1}, y_{p-1}) \\ &+ \cdots + c(x_1, y_0) - c(x_0, y_0)\}\end{aligned}$$

where $(x_0, y_0) \in \Gamma_1$ is fixed and the infimum runs among all integers p and collections $\{(x_i, y_i)\}_{1\leq i\leq p} \subset \Gamma$.

It can be easily checked that

$$\varphi = \lim_{p\to\infty} \lim_{m\to\infty} \lim_{l\to\infty} \varphi_{p,m,l},$$

where

$$\begin{aligned}\varphi_{p,m,l}(x) := \inf\{&c_l(x, y_p) - c(x_p, y_p) + c_l(x_p, y_{p-1}) - c(x_{p-1}, y_{p-1}) \\ &+ \cdots + c_l(x_1, y_0) - c(x_0, y_0)\}\end{aligned}$$

and the infimum is made among all collections $\{(x_i, y_i)\}_{1\le i\le p} \subset \Gamma_m$. As all functions $\varphi_{p,m,l}$ are upper semicontinuous we obtain that φ is a Borel function.

Arguing as in [134] it is straightforward to check that $\varphi(x_0) = 0$ and that (6.1.5) holds. Choosing $x' = x_0$ we obtain that $\varphi > -\infty$ on $\pi_X(\Gamma)$ (here we use the assumption that c is real-valued). But since $\boldsymbol{\gamma}$ is concentrated on Γ the Borel set $\pi_X(\Gamma)$ has full measure with respect to $\mu = \pi_{X\#}\boldsymbol{\gamma}$, hence $\varphi \in \mathbb{R}$ μ-a.e.

Step 2. Now we show that $\psi := \varphi^c$ is ν-measurable, real-valued ν-a.e. and that

$$\varphi + \psi = c \quad \text{on } \Gamma. \tag{6.1.6}$$

It suffices to study ψ on $\pi_Y(\Gamma)$: indeed, as γ is concentrated on Γ, the Borel set $\pi_Y(\Gamma)$ has full measure with respect to $\nu = \pi_{Y\#}\gamma$. For $y \in \pi_Y(\Gamma)$ we notice that (6.1.5) gives

$$\psi(y) = c(x,y) - \varphi(x) \in \mathbb{R} \qquad \forall x \in \Gamma_y := \{x : \ (x,y) \in \Gamma\}.$$

In order to show that ψ is ν-measurable we use the disintegration $\boldsymbol{\gamma} = \gamma_y \times \nu$ of $\boldsymbol{\gamma}$ with respect to y and notice that the probability measure γ_y is concentrated on Γ_y for ν-a.e. y, therefore

$$\psi(y) = \int_X c(x,y) - \varphi(x)\, d\gamma_y(x) \qquad \text{for } \nu\text{-a.e. } y.$$

Since $y \mapsto \gamma_y$ is a Borel measure-valued map we obtain that ψ is ν-measurable.

Step 3. We show that φ^+ and ψ^+ are integrable with respect to μ and ν respectively (here we use (6.1.3) and (6.1.4)). By (6.1.3) we can choose x in such a way that $\int_Y c(x,y)\, d\nu(y)$ is finite and $\varphi(x) \in \mathbb{R}$, so that by integrating on Y the inequality $\psi^+ \le c(x,\cdot) + \varphi^-(x)$ we obtain that $\psi^+ \in L^1(Y,\nu)$. The argument for φ^+ uses (6.1.4) and is similar.

Step 4. Conclusion. The semi-integrability of φ and ψ gives the null-Lagrangian identity

$$\int_{X\times Y} (\varphi + \psi)\, d\tilde{\boldsymbol{\gamma}} = \int_X \varphi\, d\mu + \int_Y \psi\, d\nu \in \mathbb{R} \cup \{-\infty\} \qquad \forall \tilde{\boldsymbol{\gamma}} \in \Gamma(\mu,\nu),$$

so that choosing $\tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma}$ we obtain from (6.1.6) that $\int_{X\times Y} c\, d\boldsymbol{\gamma} < +\infty$ and $\varphi \in L^1(X,\mu)$, $\psi \in L^1(Y,\nu)$. Moreover, for any $\tilde{\gamma} \in \Gamma(\mu,\nu)$ we get

$$\begin{aligned}\int_{X\times Y} c\, d\tilde{\boldsymbol{\gamma}} \quad &\ge \quad \int_{X\times Y} (\varphi+\psi)\, d\tilde{\boldsymbol{\gamma}} = \int_X \varphi\, d\mu + \int_Y \psi\, d\nu \\ &= \quad \int_{X\times Y} (\varphi+\psi)\, d\boldsymbol{\gamma} = \int_\Gamma (\varphi+\psi)\, d\boldsymbol{\gamma} = \int_{X\times Y} c\, d\boldsymbol{\gamma}.\end{aligned}$$

This chain of inequalities gives that γ is optimal and, at the same time, that (φ,ψ) is optimal in (6.1.1). $\square$

We say that a Borel function $\varphi \in L^1(X,\mu)$ is a *maximal Kantorovich potential* if (φ, φ^c) is a maximizing pair in (6.1.1). In many applications it is useful to write the optimality conditions using a maximal Kantorovich potential, instead of the cyclical monotonicity.

Theorem 6.1.5. *Let $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$, assume that* (6.1.3) *and* (6.1.4) *hold, that c is real-valued and that the* sup *in* (6.1.1) *is finite. Then there exists a maximizing pair (φ, φ^c) in* (6.1.1) *and if $\boldsymbol{\gamma} \in \Gamma(\mu,\nu)$ is optimal then*

$$\varphi(x) + \varphi^c(y) = c(x,y) \qquad \boldsymbol{\gamma}\text{-a.e. in } X \times Y. \tag{6.1.7}$$

Moreover, if there exists a Borel potential $\varphi \in L^1(X,\mu)$ such that (6.1.7) *holds, then $\boldsymbol{\gamma}$ is optimal.*

Proof. The existence of a maximizing pair is a direct consequence of the sufficiency part of the previous theorem, choosing an optimal $\boldsymbol{\gamma}$ and (by the necessity part of the statement) a c-monotone set on which $\boldsymbol{\gamma}$ is concentrated.

If $\boldsymbol{\gamma}$ is optimal then

$$\int_{X\times Y} (c - \varphi - \varphi^c)\, d\boldsymbol{\gamma} = \int_{X\times Y} c\, d\boldsymbol{\gamma} - \int_X \varphi\, d\mu - \int_Y \varphi^c\, d\nu = 0.$$

As the integrand is nonnegative, it must vanish $\boldsymbol{\gamma}$-a.e. The converse implication is analogous. □

Remark 6.1.6. The assumptions (6.1.3), (6.1.4) are implied by

$$\int_{X\times Y} c(x,y)\, d\mu \times \nu(x,y) < +\infty. \tag{6.1.8}$$

In turn, (6.1.8) is implied by the condition

$$c(x,y) \le a(x) + b(y) \quad \text{with} \quad a \in L^1(\mu),\ b \in L^1(\nu).$$

6.2 Optimal transport maps and their regularity

In this section we go back to the original Monge problem (6.0.1), finding natural conditions on c and μ ensuring the existence of optimal transport maps. The first results in this direction, in Euclidean spaces and with the quadratic cost function $|x-y|^2$, have been estabilished in [35, 36, 100]; the case of a Riemannian manifold is considered in [112].

Definition 6.2.1 (Gaussian measures and Gaussian null sets). *Let X be a separable Banach space with dual X', and let $\mu \in \mathscr{P}(X)$. We say that μ is a nondegenerate Gaussian (probability) measure in X if for any $L \in X'$ the image measure $L_{\#}\mu \in \mathscr{P}(\mathbb{R})$ has a Gaussian distribution, i.e. there exist $m = m(L) \in \mathbb{R}$ and $\sigma = \sigma(L) > 0$ such that*

$$\mu\left(\{x \in X : \ a < L(x) < b\}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_a^b e^{-|t-m|^2/2\sigma^2}\, dt \quad \forall\, (a,b) \subset \mathbb{R}.$$

We say that $B \in \mathscr{B}(X)$ is a Gaussian null set if $\mu(B) = 0$ for any nondegenerate Gaussian measure μ in X.

We refer to [32] for the general theory of Gaussian measures. Here we use Gaussian measures only to define the σ-ideal of Gaussian null sets. Starting from Definition 6.2.1 and recalling (5.2.4), it is easy to check that if μ is a (nondegenerate) Gaussian measure in X and Y is another (separable) Banach space, then

$$\begin{aligned} &\pi_{\#}\mu \quad \text{is a (nondegenerate) Gaussian measure in } Y \\ &\text{for every continuous (surjective) linear map} \quad \pi : X \to Y. \end{aligned} \tag{6.2.1}$$

One can also check that in the case $X = \mathbb{R}^d$ nondegenerate Gaussian measures are absolutely continuous with respect to $\mathscr{L}^d$, with density given by

$$\frac{1}{\sqrt{(2\pi)^d \det A}} e^{-\frac{1}{2}\langle A^{-1}(x-m),(x-m)\rangle}$$

for some $m \in \mathbb{R}^d$ and some positive definite symmetric matrix A. Therefore Gaussian null sets coincide with $\mathscr{L}^d$-negligible sets. See also [59] for the equivalence between Gaussian null sets and null sets in the sense of Aronszajn, a concept that involves only the Lebesgue measure on the real line.

Definition 6.2.2 (Regular measures). *We say that $\mu \in \mathscr{P}(X)$ is regular if $\mu(B) = 0$ for any Gaussian null set B. We denote by $\mathscr{P}^r(X)$ the class of regular measures.*

By definition of Gaussian null sets, all Gaussian measures are regular. By the above remarks on Gaussian null sets, in the finite dimensional case $X = \mathbb{R}^d$ the class $\mathscr{P}^r(X)$ reduces to the standard family of measures absolutely continuous with respect to $\mathscr{L}^d$.

We recall the following classical infinite-dimensional version of Rademacher's theorem (see for instance Theorem 5.11.1 in [32]).

Theorem 6.2.3 (Differentiability of Lipschitz functions). *Let X be a separable Hilbert space and let $\phi : X \to \mathbb{R}$ be a locally Lipschitz function. Then the set of points where ϕ is not Gateaux differentiable is a Gaussian null set.*

Theorem 6.2.4 (Optimal transport maps in $\mathbb{R}^d$). *Assume that* $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$, $c(x,y) = h(x-y)$ *with* $h : \mathbb{R}^d \to [0,+\infty)$ *strictly convex, and the minimum in* (6.0.2) *finite.*
If μ, ν *satisfy* (6.1.3), (6.1.4), *and* $\mu \in \mathscr{P}^r(\mathbb{R}^d)$, *then the Kantorovich problem* (6.0.2) *has a unique solution* $\boldsymbol{\mu}$ *and this solution is induced by an* optimal transport, *i.e. there exists a Borel map* $\boldsymbol{r} : \mathbb{R}^d \to \mathbb{R}^d$ *such that the representation* (5.2.13) *holds. We have also*

$$\boldsymbol{r}(x) = x - (\partial h)^{-1}\left(\tilde{\nabla}\varphi(x)\right) \qquad \text{for } \mu\text{-a.e. } x, \tag{6.2.2}$$

for any c-concave and maximal Kantorovich potential φ *(recall that* $\tilde{\nabla}$ *stands for the approximate differential).*

Proof. By the necessity part in Theorem 6.1.4 we have the existence of an optimal plan, concentrated on a c-monotone subset of $\mathbb{R}^d \times \mathbb{R}^d$. By the sufficiency part we obtain the existence of a c-concave maximal Kantorovich potential φ. Theorem 6.1.5 gives that for μ-a.e. x there exists y such that $\varphi(x) + \varphi^c(y) = c(x,y)$. We have to show that y is unique and given by (6.2.2). To this aim, for any $R > 0$ we define

$$\varphi_R(x) := \inf_{z \in B_R(0)} c(x,z) - \varphi^c(z) \qquad x \in \mathbb{R}^d.$$

Notice that all functions φ_R are locally Lipschitz in $\mathbb{R}^d$ for R large enough (as soon as there is some z with $|z| < R$ and $\varphi^c(z) > -\infty$) and therefore differentiable $\mathscr{L}^d$-a.e. Moreover, the above mentioned existence of y for μ-a.e. x implies that the decreasing family of sets $\{\varphi < \varphi_R\}$ has a μ-negligible intersection, i.e. μ-a.e. x belongs to $\{\varphi = \varphi_R\}$ for R large enough.

It follows that for μ-a.e. x the following two conditions are satisfied: x is a point of density 1 of $\{\varphi = \varphi_R\}$ for some R (recall Remark 5.5.2 and φ_R is differentiable at x. By the very definition of approximate differential, φ is approximately differentiable at x and $\tilde{\nabla}\varphi(x) = \nabla\varphi_R(x)$. If $\varphi(x) + \varphi^c(y) = h(x-y)$, since $x' \mapsto h(x'-y) - \varphi(x')$ attains its minimum (equal to $\varphi^c(y)$) at x, by differentiation of both sides we get

$$\tilde{\nabla}\varphi(x) \in \partial h(x-y).$$

This immediately gives that y is unique and given by (6.2.2). □

In the following remark we point out some extensions of the previous existence result and we recall some cases when the approximate differential in (6.2.2) is indeed a classical differential.

Remark 6.2.5. a) Classical differential. As the proof shows, the approximate differential is actually a classical differential if ν has a bounded support. Under a technical condition on the level sets of h at infinity (this condition includes the model case $h(z) = |z|^p$, $p > 1$) the differential is still classical even when ν has an unbounded support, see [82].

b) More general initial measures. It has been shown in [82] that for $h \in C^{1,1}_{\rm loc}(\mathbb{R}^d)$ and ν with bounded support the same properties hold if μ satisfies the more general condition

$$\mu(B) = 0 \quad \text{whenever } B \in \mathscr{B}(\mathbb{R}^d) \text{ and } \mathscr{H}^{d-1}(B) < +\infty. \tag{6.2.3}$$

The proof is based on a refinement of Rademacher theorem, valid for convex or semi-convex functions, see for instance [4].

c) The case when h is not strictly convex. Here the difficulty arises from the fact that $(\partial h)^{-1}$ is not single-valued in general, so the first variation argument of the previous proofs does not produce anymore a unique y, for given x. This problem, even when $h(z) = \|z\|$ for some norm $\|\cdot\|$ in $\mathbb{R}^d$, is not yet completely understood, see the discussions in [14]. Only the case when $\|\cdot\|$ is the Euclidean norm (or, more generally, a C^2 and uniformly convex norm) has been settled (see [142], [74], [43], [143], [9], [16]). See also [14] for an existence result in the case when the norm $\|\cdot\|$ is crystalline (i.e. its unit sphere is contained in finitely many hyperplanes).

6.2.1 Approximate differentiability of the optimal transport map

In many applications it is useful to know that the optimal transport map is differentiable, at least in the approximate sense. The following theorem answers to this question and shows, adapting to a non-smooth setting an argument in [120], that the differential of the optimal transport map is diagonalizable and has non-negative eigenvalues. Notice that our assumption on the cost includes the model case $c(x, y) = |x - y|^p$, $p > 1$. In the proof of the theorem we will use a weak version of the second order Taylor expansion, but still sufficient to have a maximum principle.

Definition 6.2.6 (Approximate second order expansion). *Let $\Omega \subset \mathbb{R}^d$ be an open set and $\varphi : \Omega \to \mathbb{R}$. We say that φ has an approximate second order expansion at $x \in \Omega$ if*

$$\lim_{y \to x,\, y \in E} \frac{\varphi(y) - a - \langle b, y - x\rangle - \langle A(y - x), (y - x)\rangle}{|y - x|^2} = 0 \tag{6.2.4}$$

for some $a \in \mathbb{R}$, $b \in \mathbb{R}^d$ and some symmetric matrix A, with E having density 1 *at x.*

It is immediate to check that $a = \tilde{\varphi}(x)$, $b = \tilde{\nabla}\varphi(x)$ and that A is uniquely determined: we will denote it by $\tilde{\nabla}^2\varphi(x)$. Moreover, if φ has a minimum at x then $b = 0$ and $A \geq 0$.

Theorem 6.2.7 (Approximate differentiability of the transport map). *Assume that $\mu \in \mathscr{P}^r(\mathbb{R}^d)$, $\nu \in \mathscr{P}(\mathbb{R}^d)$ and let $c(x, y) = h(x - y)$ with $h : \mathbb{R}^d \to [0, +\infty)$ strictly convex with superlinear growth, $h \in C^1(\mathbb{R}^d) \cap C^2(\mathbb{R}^d \setminus \{0\})$, and $\nabla^2 h$ is positive definite in $\mathbb{R}^d \setminus \{0\}$. If the minimum in* (6.0.2) *is finite, then for μ-a.e. $x \in \mathbb{R}^d$*

the optimal transport map $\boldsymbol{r}$ is approximately differentiable at x and $\tilde\nabla \boldsymbol{r}(x)$ is diagonalizable with nonnegative eigenvalues.

Proof. Let φ be a maximal Kantorovich potential and let $N = \{\boldsymbol{r}(x) \neq x\}$. Clearly it suffices to show that the claimed properties are true μ-a.e. on N (as outside of N the approximate differential of $\boldsymbol{r}$ is the identity). We consider the countable family of triplets of balls (B, B', B'') centered at a rational point of $\mathbb{R}^d$, with $\overline{B} \subset B'$, $\overline{B}' \subset B''$ and with rational radii, the family of sets

$$N_{B,B',B''} := \{x \in B :\ \boldsymbol{r}(x) \in B'' \setminus B'\},$$

and the family of functions

$$\varphi_{B,B',B''}(x) := \min_{y \in B'' \setminus B'} h(x-y) - \varphi^c(y) \qquad x \in B.$$

Notice that $\varphi_{B,B',B''} = \varphi$ μ-a.e. on $N_{B,B',B''}$, as the minimum of $y \mapsto h(x-y) - \varphi^c(y)$ is achieved at $y = \boldsymbol{r}(x) \in B'' \setminus B'$ for μ-a.e. x.

Let $C = C(B, B', B'')$ be the Lipschitz constant $\mathrm{Lip}(\nabla h, B - (B'' \setminus B'))$ of ∇h in the set $B - (B'' \setminus B')$; it follows that all maps

$$x \mapsto h(x-y) - \varphi^c(y) - \frac{C}{2}|x|^2, \qquad y \in B'' \setminus B',$$

are concave in B, and therefore $\varphi_{B,B',B''} - C|x|^2/2$ is concave in B as well. By Alexandrov's differentiability theorem (see 5.5.4) we obtain that $\varphi_{B,B',B''}$ are twice differentiable and have a classical second order Taylor expansion for $\mathscr{L}^d$-a.e. $x \in B$.

Clearly the set N is contained in the union of all sets $N_{B,B',B''}$, therefore, by Remark 5.5.2, $\mathscr{L}^d$-a.e. $x \in N$ is a point of density 1 for one of the sets $N_{B,B',B''}$ and $\varphi_{B,B',B''}$ is twice differentiable at x. By Definition 6.2.6 we obtain that φ is twice differentiable in the approximate sense at x and (6.2.4) holds with $a = \varphi(x)$, $b = \tilde\nabla\varphi(x) = \nabla\varphi_{B,B',B''}$ and $A = \tilde\nabla^2\varphi(x) = \nabla^2\varphi_{B,B',B''}/2$. Since

$$\boldsymbol{r}(x) = x - (\partial h)^{-1}(\tilde\nabla\varphi(x)) = x - \nabla h^*(\tilde\nabla\varphi(x)),$$

we obtain that $\boldsymbol{r}$ is approximately differentiable μ-a.e. on N.

Since h has a superlinear growth at infinity, the gradient map $\nabla h : \mathbb{R}^d \to \mathbb{R}^d$ is a bijection and its inverse is ∇h^*, where h^* is the conjugate of h. Therefore ∇h^* is differentiable on $\mathbb{R}^d \setminus \{\nabla h(0)\}$.

Fix now a point x where the above properties hold and set $y = \boldsymbol{r}(x)$. Since $x' \mapsto h(x'-y) - \varphi(x')$ achieves its minimum, equal to $-\varphi^c(y)$, at x, we get

$$\nabla^2 h(x-y) \geq \tilde\nabla^2\varphi(x).$$

On the other hand, the identity $\nabla h(\nabla h^*(p)) = p$ gives

$$\nabla^2 h\left(\nabla h^*(p)\right) = \left[\nabla^2 h^*(p)\right]^{-1}.$$

Using the identity above with $p = \tilde{\nabla}\varphi(x) \neq \nabla h(0)$ we obtain

$$\left[\nabla^2 h^*(\tilde{\nabla}\varphi(x))\right]^{-1} \geq \tilde{\nabla}^2\varphi(x).$$

By Lemma 6.2.8 below with $A := \nabla^2 h^*(\tilde{\nabla}\varphi(x))$ and $B := -\tilde{\nabla}^2\varphi(x)$ we obtain that $\tilde{\nabla}\boldsymbol{r}(x) = \boldsymbol{i} + AB$ is diagonalizable and it has nonnegative eigenvalues. □

Again, under more restrictive assumptions (e.g. the supports of the two measures are compact and $\operatorname{dist}(\operatorname{supp}\mu, \operatorname{supp}\nu) > 0$) one can show that the optimal transport map $\boldsymbol{r}$ is μ-a.e. differentiable in a classical sense. As discussed in Section 5.5, approximate differentiability is however sufficient to establish an area formula and the rule for the computation of the density of $\boldsymbol{r}_\#(\rho\mathscr{L}^d)$.

The following elementary lemma is also taken from [120].

Lemma 6.2.8. *Let A, B be symmetric matrices with A positive definite. If $-B \leq A^{-1}$ then $\boldsymbol{i} + AB$ is diagonalizable and has nonnegative eigenvalues.*

Proof. Let C be a positive definite symmetric matrix such that $C^2 = A$. Since

$$\boldsymbol{i} + AB = C(\boldsymbol{i} + CBC)C^{-1}$$

and since $\boldsymbol{i} + CBC$ is symmetric we obtain that $\boldsymbol{i} + AB$ is diagonalizable. In order to show that the eigenvalues are nonnegative we estimate:

$$\begin{aligned}\langle(\boldsymbol{i} + CBC)\xi, \xi\rangle &= |\xi|^2 + \langle C\xi, BC\xi\rangle \geq |\xi|^2 - \langle C\xi, A^{-1}C\xi\rangle \\ &= |\xi|^2 - \langle \xi, CA^{-1}C\xi\rangle = 0 \quad \square\end{aligned}$$

□

In the following theorem we establish, under more restrictive assumptions on $\boldsymbol{r}$ or h, some properties of the distributional derivative of $\boldsymbol{r}$ and the nonnegativity of the distributional divergence of $\boldsymbol{r}$ (or, better, of a canonical extension of $\boldsymbol{r}$ to the whole of $\mathbb{R}^d$: recall that $\boldsymbol{r}$ is a priori defined only μ-a.e.).

Theorem 6.2.9 (Distributional derivative of $\boldsymbol{r}$). *Let μ, $\nu \in \mathscr{P}^r(\mathbb{R}^d)$, with $\operatorname{supp}\nu$ bounded, let $c(x, y) = h(x - y)$ with $h : \mathbb{R}^d \to [0, +\infty)$ strictly convex and with superlinear growth and assume that the minimum in (6.0.2) is finite. Let $\boldsymbol{r}$ be the optimal transport map between μ and ν. Then*

- (i) *If $h \in C^2(\mathbb{R}^d)$ is locally uniformly convex then $\boldsymbol{r}$ has a canonical BV_{loc} extension to $\mathbb{R}^d$ satisfying $D \cdot \boldsymbol{r} \geq 0$.*
- (ii) *If $h \in C^2(\mathbb{R}^d \setminus \{0\})$ and $\nabla h(0) = 0$ we can find equi-bounded maps $\boldsymbol{r}_k \in BV_{\mathrm{loc}}(\mathbb{R}^d)$ satisfying $D \cdot \boldsymbol{r}_k \geq 0$ such that $\mu(\{\boldsymbol{r}_k \neq \boldsymbol{r}\}) \to 0$ as $k \to \infty$.*

Proof. (i) By the argument used in the proof of Theorem 6.2.4 we know that there exists a c-concave potential φ of the form

$$\varphi(x) = \inf_{y \in \operatorname{supp}\nu} h(x - y) - \psi(y) \tag{6.2.5a}$$

with $\psi = -\infty$ on $\mathbb{R}^d \setminus \operatorname{supp}\nu$, such that

$$\boldsymbol{r}(x) = x - (\nabla h)^{-1}(\tilde{\nabla}\varphi) \quad \mu\text{-a.e. in } \mathbb{R}^d. \tag{6.2.5b}$$

We take as an extension of $\boldsymbol{r}$ the right hand side in the previous identity (6.2.5b), for φ given by (6.2.5a). Notice that, on any ball B, all functions

$$x \mapsto h(x-y) - \psi(y) - C|x|^2 \quad \text{for } y \in \operatorname{supp}\nu, \quad \psi(y) > -\infty,$$

are concave for C large enough (depending on B and $\operatorname{supp}\nu$), so that $\varphi - C|x|^2$ is concave in B as well. This proves that φ is locally Lipschitz and locally BV in $\mathbb{R}^d$ and therefore, since the inverse of ∇h is locally Lipschitz in $\mathbb{R}^d$ as well (by the local uniform convexity assumption on h and the superlinear growth condition), also $\boldsymbol{r}$ is locally BV.

Let us show that $\boldsymbol{r}(x) \in \operatorname{supp}\nu$ and that $x' \mapsto \varphi(x') - h(x'-y)$ attains its maximum at x when $y = \boldsymbol{r}(x)$ for $\mathscr{L}^d$-a.e. $x \in \mathbb{R}^d$. Indeed, fix x where φ is differentiable and let $\bar{y} \in \operatorname{supp}\nu$ be a minimizer of $y \mapsto h(x-y) - \psi(y)$ (without loss of generality we can assume that ψ is upper semicontinuous: being $\operatorname{supp}\nu$ compact and $\psi(y) < +\infty$ for every $y \in X$, a minimizer exists). Then $\varphi(x') - h(x'-\bar{y})$ attains its maximum at x since (6.2.5a) yields

$$\varphi(x') - h(x'-\bar{y}) \le h(x'-\bar{y}) - \psi(\bar{y}) - h(x'-\bar{y}) = -\psi(\bar{y}) = \varphi(x) - h(x-\bar{y}),$$

and a differentiation yields $\bar{y} = \boldsymbol{r}(x)$.

It remains to show that $D\cdot\boldsymbol{r} \ge 0$. Since $\max_{\operatorname{supp}\nu} h(x-\cdot)$ is locally bounded we can find a strictly positive function $\rho \in L^1(\mathbb{R}^d)$ such that

$$\int_{\mathbb{R}^d} \max_{y \in \operatorname{supp}\nu} h(x-y)\rho(x)\,dx < +\infty. \tag{6.2.6}$$

Let $\bar{\mu} = \rho\mathscr{L}^d$, and notice that the minimality property above shows that the graph of $\boldsymbol{r}$ is (essentially, excluding points x where φ is not differentiable) c-monotone: indeed for any choice of differentiability points $x_1, \ldots, x_n$ of and for any permutation σ of $\{1, \ldots, n\}$ we have

$$\sum_{i=1}^{n} \varphi(x_{\sigma(i)}) - h(x_{\sigma(i)} - \boldsymbol{r}(x_i)) \le \sum_{i=1}^{n} \varphi(x_i) - h(x_i - \boldsymbol{r}(x_i)).$$

Removing from both sides $\sum_i \varphi(x_i)$ we obtain the c-monotonicity inequality.

Therefore, since by (6.2.6) the cost associated to $\boldsymbol{r}$ is finite, Theorem 6.1.4 gives that $\boldsymbol{r}$ is an optimal map between $\bar{\mu}$ and $\boldsymbol{r}_{\#}\bar{\mu}$.

This optimality property of the extended map $\boldsymbol{r}$ shows that it suffices to prove that $D\cdot\boldsymbol{r} \ge 0$ only when $\operatorname{supp}\nu$ is made by finitely many points: the general case can be achieved by approximation, using the fact that optimality relative to $\bar{\mu}$ is stable in the limit and yields $L^p(\bar{\mu})$ convergence of the maps (see Lemma 5.4.1) and then, up to subsequences, $\mathscr{L}^d$-a.e. convergence, due to the fact that $\rho > 0$ $\mathscr{L}^d$-a.e. Under the assumption that $\operatorname{supp}\nu$ is finite the function $\boldsymbol{r}$ takes only finitely many values $\{y_1, \ldots, y_m\}$ and the distributional divergence is given by

$$D \cdot \boldsymbol{r} = \langle \boldsymbol{r}^+ - \boldsymbol{r}^-, \boldsymbol{n} \rangle \chi_S \mathscr{H}^{d-1},$$

where $\boldsymbol{r}^\pm$ are the approximate one sided limits on the approximate jump set S of $\boldsymbol{r}$ and $\boldsymbol{n}$ is the approximate normal to the jump set. For a given Borel choice of $\boldsymbol{n}$, let us consider the sets

$$S_{ij} := \Big\{ x \in S : \boldsymbol{r}^-(x) = y_i,\ \boldsymbol{r}^+(x) = y_j \Big\} \quad 1 \le i, j \le m,\ i \ne j, \quad S = \bigcup_{i \ne j} S_{ij}.$$

Since each neighborhood of $x \in S_{ij}$ contains points $x^\pm$ such that $\boldsymbol{r}(x^\pm) = \boldsymbol{r}^\pm(x)$ is the unique minimizer of $y \mapsto h(x^\pm - y) - \psi(y)$ in $\{y_1, \cdots, y_m\}$, S_{ij} is contained in ∂E_{ij}, with

$$E_{ij} := \big\{ x \in \mathbb{R}^d :\ h(x - y_i) - \psi(y_i) < h(x - y_j) - \psi(y_j) \big\} \qquad 1 \le i \ne j \le m$$

and the classical inner normal to E_{ij} is parallel (with the same direction) to the nonvanishing vector $\nabla h(x - y_j) - \nabla h(x - y_i)$. Therefore it suffices to check the inequality

$$\langle y_i - y_j, \nabla h(x - y_j) - \nabla h(x - y_i) \rangle \ge 0.$$

This is a direct consequence of the monotonicity of ∇h:

$$\langle (x - y_j) - (x - y_i), \nabla h(x - y_j) - \nabla h(x - y_i) \rangle \ge 0.$$

(ii) Let $h_k \ge h$ be in $C^2(\mathbb{R}^d)$ and locally uniformly convex, with the property that for any $z \in \mathbb{R}^d$ we have $h_k(z) = h(z)$ and $(\nabla h_k)^{-1}(z) = (\nabla h)^{-1}(z)$ for k large enough (the proof of the existence of this approximation, a regularization of h near the origin, is left to the reader) and let φ, ψ as in the proof of (i). We define

$$\varphi_k(x) := \inf_{y \in \operatorname{supp}\nu} h_k(x - y) - \psi(y)$$

so that $\varphi_k \ge \varphi$. Since the infimum in the problem defining φ is attained (by $y = \boldsymbol{r}(x)$) for μ-a.e. x, it follows that $\varphi_k(x) = \varphi(x)$ for μ-a.e. x for k large enough (precisely, such that $h_k(x - \boldsymbol{r}(x)) = h(x - \boldsymbol{r}(x))$, so that $\mu(\{\varphi_k \ne \varphi\}) \to 0$ as $k \to \infty$. Setting

$$\boldsymbol{r}_k := \boldsymbol{i} - (\nabla h_k)^{-1}(\tilde{\nabla}\varphi_k)$$

we know, by the c-monotonicity argument seen in the proof of statement (i), that $\boldsymbol{r}_k$ are optimal transport maps relative to the costs $h_k(x - y)$, that $\boldsymbol{r}_k \in \operatorname{supp}\nu$ μ-a.e. and that $D \cdot \boldsymbol{r}_k \ge 0$. Since the approximate differentials coincide at points of density 1 of the coincidence set we have $\mu(\{\tilde{\nabla}\varphi_k \ne \tilde{\nabla}\varphi\}) \to 0$ as $k \to \infty$ and therefore $\mu(\{\boldsymbol{r}_k \ne \boldsymbol{r}\}) \to 0$ as $h \to \infty$. □

6.2.2 The infinite dimensional case

In the infinite dimensional case we consider for simplicity only the case when $c(x,y) = |x-y|^p/p$, $p > 1$; when ν has a bounded support we are still able to recover, by the same argument used in the finite dimensional case, a differential characterization of the optimal transport map.

We denote by $\mathscr{P}^r_p(X)$ the intersection of $\mathscr{P}_p(X)$ (see (5.1.22)) with $\mathscr{P}^r(X)$.

Theorem 6.2.10 (Optimal transport maps in Hilbert spaces). *Assume that X is a separable Hilbert space, let $\mu \in \mathscr{P}^r_p(X)$, $\nu \in \mathscr{P}_p(X)$ and let $c(x,y) = |x-y|^p/p$ for $p \in (1,+\infty)$, $q^{-1} + p^{-1} = 1$. Then the Kantorovich problem* (6.0.2) *has a unique solution $\boldsymbol{\mu}$ and this solution is induced by an* optimal transport, *i.e. there exists a Borel map $\boldsymbol{r} \in L^p(X,\mu;X)$ such that the representation* (5.2.13) *holds. If ν has a bounded support we have also*

$$\boldsymbol{r}(x) = x - |\nabla\varphi(x)|^{q-2}\nabla\varphi(x) \qquad \text{for } \mu\text{-a.e. } x, \tag{6.2.7}$$

for some locally Lipschitz, c-concave and maximal Kantorovich potential φ (here $\nabla\varphi$ denotes the Gateaux differential of φ).

Proof. Let us assume first that $\operatorname{supp}\nu$ is bounded. We first define a canonical Kantorovich potential, taking into account the boundedness assumption on $\operatorname{supp}\nu$, as follows. Let ϕ be any maximal Kantorovich potential and define

$$\varphi(x) := \inf_{y\in\operatorname{supp}\nu} c(x,y) - \phi^c(y) \qquad x \in X. \tag{6.2.8}$$

Notice that the optimality conditions on ϕ ensure that for μ-a.e. x the infimum above is attained. By construction φ is a locally Lipschitz function and it is still a maximal Kantorovich potential. Indeed, $\varphi = \phi$ μ-a.e. and since φ is the c-transform of the function ψ equal to ϕ^c on $\operatorname{supp}\nu$ and equal to $-\infty$ otherwise we have $\varphi^c = (\psi^c)^c \geq \psi = \phi^c$ on $\operatorname{supp}\nu$.

As in the proof of Theorem 6.2.4 it can be shown that for μ-a.e. x there is only one y such that $\varphi(x) + \varphi^c(y) = c(x,y)$, and that y is given by (6.2.7); the only difference is that we have to consider Theorem 6.2.3 instead of the classical Rademacher theorem.

In the general case when $\operatorname{supp}\nu$ is possibly unbounded we can still prove existence and uniqueness of an optimal transport map as follows. Let $\boldsymbol{\gamma} \in \Gamma_o(\mu,\nu)$, let $\boldsymbol{\gamma}_n = \chi_{B_n}(y)\boldsymbol{\gamma}$ where $B_n := B_n(0)$ is the centered open ball of radius n, and let μ_n, ν_n be the marginals of $\boldsymbol{\gamma}_n$ (in particular $\nu_n = \chi_{B_n}\nu$ and μ_n is absolutely continuous with respect to μ, therefore still regular). By Theorem 6.1.5 we know that $\operatorname{supp}\boldsymbol{\gamma}$ is $|\cdot|^p$-monotone, and therefore $\operatorname{supp}\boldsymbol{\gamma}_n$ is $|\cdot|^p$-monotone as well. By applying Theorem 6.1.5 again and the first part of the present proof, we obtain that $\boldsymbol{\gamma}_n$ is an optimal plan, induced by a unique transport map $\boldsymbol{r}_n$. The inequality

$$(\boldsymbol{i}\times\boldsymbol{r}_n)_\#\mu_n = \boldsymbol{\gamma}_n \leq \boldsymbol{\gamma}_m = (\boldsymbol{i}\times\boldsymbol{r}_m)_\#\mu_m$$

immediately gives (for instance by disintegration of both sides with respect to x)

$$\boldsymbol{r}_n = \boldsymbol{r}_m \quad \mu_n\text{-a.e. whenever } n < m.$$

Therefore the map $\boldsymbol{r}$ such that $\boldsymbol{r} = \boldsymbol{r}_n$ μ_n-a.e. for any n is well defined, and passing to the limit as $n \to \infty$ in the identity $\boldsymbol{\gamma}_n = (\boldsymbol{i} \times \boldsymbol{r})_{\#}\mu_n$ we obtain $\boldsymbol{\gamma} = (\boldsymbol{i} \times \boldsymbol{r})_{\#}\mu$. This proves that $\boldsymbol{r}$ is an optimal transport map, and that any optimal plan is induced by an optimal transport map.

If there were two different optimal transport maps $\boldsymbol{r}$, $\boldsymbol{r}'$, then we could build an optimal transport plan

$$\boldsymbol{\gamma} := \frac{1}{2}\int_X \delta_{\boldsymbol{r}(x)} + \delta_{\boldsymbol{r}'(x)}\, d\mu(x)$$

which is not induced by any transport map. This contradiction proves the uniqueness of $\boldsymbol{r}$. □

Remark 6.2.11 (Essential injectivity of the transport map). Notice also that if ν is regular as well, under the assumption of Theorem 6.2.4 or Theorem 6.2.10, then the optimal transport map $\boldsymbol{r}$ between μ and ν is *μ-essentially injective* (i.e. its restriction to a set with full μ-measure is injective). This follows by the fact that, denoting by $\boldsymbol{s}$ the optimal transport map between ν and μ, the uniqueness of optimal plans gives $(\boldsymbol{i} \times \boldsymbol{r})_{\#}\mu = [(\boldsymbol{s} \times \boldsymbol{i})_{\#}\nu]^{-1}$, which leads to $\boldsymbol{s} \circ \boldsymbol{r} = \boldsymbol{i}$ μ-a.e. and to the essential injectivity of $\boldsymbol{r}$.

In the case when $p = 2$ and $\mu, \nu \in \mathscr{P}_2^r(\mathbb{R}^d)$ we can actually prove *strict monotonicity* of the optimal transport map.

Proposition 6.2.12 (Strict monotonicity of $\boldsymbol{r}$). *Let $\mu, \nu \in \mathscr{P}_2^r(\mathbb{R}^d)$, and let $\boldsymbol{r}$ be the unique optimal transport map relative to the cost $c(x,y) = |x-y|^2/2$. Then $\nabla\boldsymbol{r} > 0$ μ-a.e. and there exists a μ-negligible set $N \subset \mathbb{R}^d$ such that*

$$\langle \boldsymbol{r}(x_1) - \boldsymbol{r}(x_2), x_1 - x_2\rangle > 0 \qquad \forall x_1,\, x_2 \in \mathbb{R}^d \setminus N. \tag{6.2.9}$$

Proof. Let φ be a c-concave maximal Kantorovich potential. The c-concavity of φ and its construction ensure that $\varphi < +\infty$ globally, that $\varphi > -\infty$ μ-a.e. and that $\varphi - |x|^2/2$ is concave. In particular, denoting by C the interior of the convex hull of $\{\varphi \in \mathbb{R}\}$, we have that φ is finite on C and μ is concentrated on C. We have also that the optimal transport map $\boldsymbol{r}$ can be represented as $\nabla\phi$ with $\phi = |x|^2/2 - \varphi$ convex. Recalling that, by Alexandrov's theorem 5.5.4 convex functions are twice differentiable $\mathscr{L}^d$-a.e. in the classical sense, we can apply Lemma 5.5.3 to obtain that $\nabla\boldsymbol{r} > 0$ μ-a.e. in C, due to the fact that $\boldsymbol{r}_{\#}\mu \ll \mathscr{L}^d$.

Let now N be the μ-negligible set of points $x \in C$ where either ϕ is not twice differentiable or $\nabla^2\phi$ has some zero eigenvalue. The monotonicity inequality then gives (with $x_t = (1-t)x + ty$)

$$\langle \nabla\phi(y) - \nabla\phi(x), y - x\rangle \geq \lim_{t\downarrow 0} \frac{1}{t^2}\langle \nabla\phi(x_t) - \nabla\psi(x), x_t - x\rangle > 0$$

for any $x,\, y \in C \setminus N$. □

6.2.3 The quadratic case $p = 2$

In the case of $c(x, y) := \frac{1}{2}|x - y|^2$ in a Hilbert space X, the theory developed in the previous sections presents some more interesting features and stronger links with classical convex analysis.

Here we quote the most relevant aspects.

- *A function $u : X \to \bar{\mathbb{R}}$ is c-concave iff $u - \frac{1}{2}|\cdot|^2$ is u.s.c. and concave, i.e. $\tilde{u}(x) := \frac{1}{2}|x|^2 - u(x)$ is l.s.c. and convex.*
  For, from the representation of (6.1.2) we get
  $$u(x) - \frac{1}{2}|x|^2 = \inf_{i \in I} t_i + \frac{1}{2}|y_i|^2 - \langle x, y_i \rangle.$$
  This means that $u(x) - |x|^2/2$ is the infimum of a family of linear continuous functional on X.

- *If $v = u^c$ is the c-transform of u then $\tilde{v} = \tilde{u}^*$, the Legendre-Fenchel-Moreau conjugate functional defined as*
  $$\tilde{u}^*(y) := \sup_{x \in X} \langle x, y \rangle - \tilde{u}(x).$$
  We simply have
  $$\tilde{v}(y) = \frac{1}{2}|y|^2 - u^c(y) = \sup_{x \in X} \frac{1}{2}|y|^2 - \frac{1}{2}|x - y|^2 + u(x)$$
  $$= \sup_{x \in X} \langle x, y \rangle - \big(\frac{1}{2}|x|^2 - u(x)\big) = \sup_{x \in X} \langle x, y \rangle - \tilde{u}(x).$$

- *A subset Γ of X^2 is c-monotone according to Definition* 6.1.3 *iff it is* cyclically monotone*, i.e. for every cyclical choice of points $(x_1^k, x_2^k) \in \Gamma$, $k = 0, \ldots, N$, with $(x_1^0, x_2^0) = (x_1^N, x_2^N)$, we have*
  $$\sum_{k=1}^{N} \langle x_1^k - x_1^{k-1}, x_2^k \rangle \geq 0. \tag{6.2.10}$$
  In particular, by Rockafellar theorem, c-monotone sets are always contained in the graph of the subdifferential
  $$\{(x, y) : \ y \in \partial\varphi(x)\}$$
  of a convex l.s.c function φ. Conversely, any subset of such a graph is c-monotone.

- *Suppose that μ, $\nu \in \mathscr{P}_2(X)$ and $\boldsymbol{\gamma} \in \Gamma(\mu, \nu)$. Then the following properties are equivalent:*

– $\boldsymbol{\gamma}$ *is optimal;*

– $\operatorname{supp}\boldsymbol{\gamma}$ *is cyclically monotone;*

– *there exists a convex, l.s.c. potential* $\tilde{\varphi} \in L^1(X,\mu)$ *such that*

$$\langle x, y\rangle = \tilde{\varphi}(x) + \tilde{\varphi}^*(y) \quad \boldsymbol{\gamma}\text{-a.e. in } X^2. \tag{6.2.11}$$

Equivalently, we can also state (6.2.11) by saying that $y \in \partial\varphi(x)$ for $\boldsymbol{\gamma}$-a.e. $(x,y) \in X^2$. In particular, if $\boldsymbol{\gamma} = (\boldsymbol{i} \times \boldsymbol{r})_{\#}\mu$ then there exists a l.s.c. convex functional φ such that $\boldsymbol{r}(x) \in \partial\varphi(x)$ for μ-a.e. $x \in X$.

- *Suppose that* $X = \mathbb{R}^d$ *and* $\mu \in \mathscr{P}_2^r(\mathbb{R}^d)$, $\nu \in \mathscr{P}_2(\mathbb{R}^d)$. *Then there exists a unique optimal transport plan and this plan is induced by a transport map* $\boldsymbol{r}$. *If* $\nu \in \mathscr{P}_2^r(\mathbb{R}^d)$ *as well, then* $\boldsymbol{r}$ *is* μ*-essentially injective and fulfills* (6.2.9).