Scuola Normale Superiore di Pisa
Classe di Scienze

PhD Thesis

Optimization Problems for Transportation Networks

Alessio Brancolini
Advisor: *Prof. Giuseppe Buttazzo*

Year 2005

Contents

Introduction

Optimal transportation problems

The statement of the problem of the theory of Optimal Transportation was first posed by Monge in 1781 in [36] when he raised the following question: given two mass distributions f^+ and f^-, minimize the transport cost

$$\int_{\mathbb{R}^N} |x - t(x)| f^+(x) \, \mathrm{d}x$$

among all *transport maps* t, i.e. measurable maps such that the mass balance condition

$$\int_{t^{-1}(B)} f^+(x) \, \mathrm{d}x = \int_B f^-(y) \, \mathrm{d}y$$

holds for every Borel set B. The Monge's approach to this problem is quite simple to be stated: the unknown of the problem is the map t that tells that the infinitesimal amount of mass $\mathrm{d}x$ located at x will be placed in the point $t(x)$ at the end of the transportation and that the work done is given by $|x - t(x)| f^+(x)\mathrm{d}x$. Clearly, instead of two mass distributions f^+ and f^- one can consider two probability measures μ^+ and μ^- and minimize the functional given by

$$t \mapsto \int_{\mathbb{R}^N} |x - t(x)| \, \mathrm{d}\mu^+(x)$$

among the *transport maps* t, i.e. measurable maps such that $\mu^-(B) = \mu^+(t^{-1}(B))$ for any Borel set B. In spite of the easiness of its formulation and physical interpretation, the mathematical difficulties are great. In fact, because of the strong non-linearity in the unknown of the problem, the map t, in the general case neither the existence of a transport plan nor that of an optimal one is assured. So, Monge's formulation did not lead to significant advances up to 1940, when Kantorovich proposed his own formulation in his famous papers [30] and [31].

In modern notation, given two finite positive Borel measures μ^+ and μ^- on $\mathbb{R}^N$ such that $\mu^+(\mathbb{R}^N) = \mu^-(\mathbb{R}^N)$, Kantorovich was interested to minimize the functional

$$\mu \mapsto \int_{\mathbb{R}^N \times \mathbb{R}^N} c(x, y)\ \mathrm{d}\mu(x, y)$$

among all *transport plans* μ, i.e. positive Borel measures on $\mathbb{R}^N \times \mathbb{R}^N$ such that $\pi^+_\# \mu = \mu^+$ and $\pi^-_\# \mu = \mu^-$, where by $\#$ we denoted the *push-forward* operator (i.e. $h_\# \mu(E) = \mu(h^{-1}(E))$). The cost function c is a non-negative lower semicontinuous function defined on $\mathbb{R}^N \times \mathbb{R}^N$. It is easy to see that if t is a transport map between $\mu^+ = f^+ \mathcal{L}^N$ and $\mu^- = f^- \mathcal{L}^N$, then $\mu_t := (\mathrm{Id} \times t)_\# \mu^+$ is a transport plan and

$$\int_{\mathbb{R}^N} c(x, t(x))\ \mathrm{d}\mu^+(x) = \int_{\mathbb{R}^N \times \mathbb{R}^N} c(x, y)\ \mathrm{d}\mu_t(x, y).$$

So, Kantorovich's problem is a weak formulation of Monge's one. The advantages of this approach are evident: the set of transport plans is a non-empty, convex and weakly compact set, the functional is now linear in the unknown and, last but not least, the existence of minimizer is implied.

The study of optimal transport problems has received a great attention after the work by Sudakov on the existence of an optimal transport map (see [46]) and now the theory has applications in various subjects of research in Non-linear Partial Differential Equations, Calculus of Variations, Probability, Economics, Statistical Mechanics, Fluidodynamics and many other fields. These developments are accounted in various surveys and books such as [2], [25], [39], [40], and [47].

Plan of the work

This thesis consists in three chapters. The first one deals with the theory of Optimal Transportation, while in the second a model of urban planning based on a functional built upon Kantorovich functional is studied. In the third an alternative approach to transport probability measures is considered and studied.

Chapter 1, Optimal Transportation Problems

In this chapter we review some basic facts in the theory of Optimal Transportation. First we state Monge and Kantorovich's Problems in a sufficient general setting.

Problem (Monge Problem). Given two finite positive Borel measures μ^+ and μ^- on a metric space X such that $\mu^+(X) = \mu^-(X)$, the object of the minimization if the functional

$$M(t) = \int_X c(x, t(x)) \, \mathrm{d}\mu^+(x)$$

among all *transport maps* t, that is measurable maps $t : X \to X$ such that $\mu^-(B) = \mu^+(t^{-1}(B))$, that is $t_\#\mu^+ = \mu^-$. Here c is a generic *cost function*, that is to say a function $c : X \times X \to \mathbb{R}$ non-negative and lower semicontinuous.

Problem (Kantorovich Problem). Given two finite positive Borel measures μ^+ and μ^- on a metric space X such that $\mu^+(X) = \mu^-(X)$, the functional to minimize is

$$K(\mu) = \int_{X\times X} c(x, y) \, \mathrm{d}\mu(x, y)$$

among all *transport plans* μ, i.e. positive Borel measures on $X \times X$ such that $\mu^+(A) = \mu(A \times X)$ and $\mu^-(B) = \mu(X \times B)$, that is $\pi^+_\#\mu = \mu^+$ and $\pi^-_\#\mu = \mu^-$.

Then, we consider the question of the existence of an optimal transport map or of an optimal transport plan, showing that, while neither the existence of a transport map nor that of an optimal one are guaranteed, the existence of an optimal transport plan is proved (in Polish spaces).

In the sequel some other classic topics are considered, such as cyclical monotonicity with respect to a cost c.

Definition (c-cyclical monotonicity). A subset $S \subseteq X \times Y$ is said to be *c-cyclically monotone* if for any $n \in \mathbb{N}$ and for any couples $(x_i, y_i) \in S$, $i = 1, 2, \ldots, n$ and for any permutation of n elements $\sigma \in S_n$ we have that

$$\sum_{i=1}^{n} c(x_i, y_i) \leq \sum_{i=1}^{n} c(x_{\sigma(i)}, y_i).$$

The main result is the proof that the support of an optimal transport plan is c-cyclically monotone (Theorem 1.3.2), the generalized Rockafellar Theorem (Theorem 1.3.8) and the proof of the existence of an optimal transport plan in the quadratic case (Theorem 1.3.12).

In Section 1.4 we give the statement and the full proof of Kantorovich Duality Formula.

Then we go on with the study of the main properties of the p-Wasserstein distance, i.e. a distance on the set of probability measures.

Definition (Wasserstein distances). Let X be a metric space and d its distance. Given $\mu^+, \mu^- \in \mathcal{P}_p(X)$, the *Wasserstein distance of order* p is defined by

$$W_p(\mu^+, \mu^-) := \left[\inf_{\mu \in \mathcal{P}(\mu^+, \mu^-)} \int_{X \times X} [d(x, y)]^p \, \mathrm{d}\mu(x, y) \right]^{\min\left\{1, \frac{1}{p}\right\}}.$$

where μ ranges in the set of transport plans between μ^+ and μ^-.

The main basic properties are investigated, including the equivalence between the convergence with respect to the p-Wasserstein distance and the convergence in the duality with bounded continuous functions plus convergence of the momenta of order p.

Finally, we will introduce some tools to deal with *displacement convexity* (first introduced and studied in [35] and studied in the general case in [1]), that is convexity with respect to the *displacement interpolation* (given μ^+ and μ^- the displacement interpolation, in the case $p = 2$, is given by $\mu_t = [(1-t)\,\mathrm{Id} + tT]_\# \mu^+$, where T is the optimal transport between μ^+ and μ^-). In particular, we prove that certain kind of functionals (that arise in the modelling of an interacting gas, see [35]) are displacement convex (these results will be useful in Chapter 3).

Chapter 2, Optimal Networks for Mass Transportation Problems

In this chapter we study the generalization of a an urban planning problem already stated in [18] and [20]. The result of this chapter can also be found in [13].

We consider a bounded connected open subset Ω with Lipschitz boundary of $\mathbb{R}^N$ (the urban area) with $N > 1$ and two positive finite measures μ^+ and μ^- on $K := \overline{\Omega}$ (the distributions of working people and of working places). We assume that μ^+ and μ^- have the same mass that we normalize both equal 1, that is μ^+ and μ^- are probability measures on K.

The optimization problem for transportation networks considered is this: to every "urban network" Σ we associate a suitable "cost function" d_Σ which takes into account the geometry of Σ as well as the costs for customers to move with their own means and by means of the network. The cost functional will be then

$$T(\Sigma) = W_{d_\Sigma}(\mu^+, \mu^-),$$

where $W_{d_\Sigma}(\mu^+, \mu^-)$ is the "Wasserstein distance" $W_1(\mu^+, \mu^-)$ with respect to the pseudo-distance d_Σ, so that the optimization problem we deal with is

$$\min\{T(\Sigma) \ : \ \Sigma \text{ “admissible network”}\}.$$

The main result is to prove that, under suitable and very mild assumptions, and taking as admissible networks all connected, compact one-dimensional subsets Σ of K, the optimization problem we consider admits a solution. The tools we use to obtain the existence result are a suitable relaxation procedure to define the function d_Σ and a generalization of the classical Gołab Theorem (Theorem 2.2.2 and Theorem 2.2.3).

In order to introduce the distance d_Σ on the set $\overline{\Omega} \times \overline{\Omega}$ we consider a function $J : [0, +\infty]^3 \to [0, +\infty]$. For a given path γ in K the parameter a in $J(a, b, c)$ measures the length of γ outside Σ, b measures the length of γ inside Σ, while c represents the total length of Σ. The cost $J(a, b, c)$ is then the cost of a customer who travels for a length a by his own means and for a length b on the network, being c the length of the latter. For instance we could take $J(a, b, c) = A(a) + B(b) + C(c)$ and then the function $A(t)$ is the cost for travelling a length t by one's own means, $B(t)$ is the price of a ticket to cover the length t on Σ and $C(t)$ represents the cost of a network of length t.

For every closed connected subset Σ in K, we then define the cost function d_Σ as

$$d_\Sigma(x, y) := \inf\left\{J\left(\mathcal{H}^1(\gamma \setminus \Sigma), \mathcal{H}^1(\gamma \cap \Sigma), \mathcal{H}^1(\Sigma)\right) \ : \ \gamma \in \mathscr{C}_{x,y}\right\},$$

where $\mathscr{C}_{x,y}$ is the class of all closed connected subsets of K containing x and y.

The optimization problem we consider is then the minimization for the functional

$$\Sigma \mapsto T(\Sigma) = W_{d_\Sigma}(\mu^+, \mu^-)$$

where we take as *admissible networks* all closed connected subsets Σ of K with $\mathcal{H}^1(\Sigma) < +\infty$.

Chapter 3, Path Functionals over Wasserstein Spaces

The problem of transporting a source mass distribution onto a target mass distribution by keeping together as much mass as possible during the transport, from which tree-shaped configurations arise, has been very much studied (see, for example, [7], [34] or [48]). In the new approach to this problem presented in this chapter (and also in [14]) probability measures valued curves are considered, while the condition of keeping masses together is achieved considering only measures supported in discrete sets.

Given a source or initial probability measure μ_0 and a target or final probability measure μ_1 we look for a path γ in a Wasserstein space $\mathcal{W}_p(\Omega)$ that connects μ_0 to μ_1 and minimizes a suitable cost functional $\mathcal{J}(\gamma)$. We consider functionals of the form

$$\mathcal{J}(\gamma) = \int_0^1 J(\gamma(t))|\gamma'|(t)\mathrm{d}t$$

where $|\gamma'|$ is the metric derivative of γ in the Wasserstein space $\mathcal{W}_p(\Omega)$ and J is a lower semicontinuous functional defined on measures. Here J may be easily seen as the coefficient of a degenerate "Riemannian distance" on the space $\mathcal{W}_p(\Omega)$.

We restrict our analysis to the case of J being a local functional over measures, an important class of functionals extensively studied by Bouchitté and Buttazzo in [9], [10], and [11]. These functionals are the key tool in our approach, and among them we can find both functionals which are finite only on concentrated measures and functionals which are finite only on spread ones. In fact, a particular point of interest in our approach is the fact that also different kinds of "Riemannian distances" are allowed (for instance those which prefer spread measures) by a change of the functional J.

The analysis of existence results as well as the definition of the cost functionals is done in an abstract metric spaces framework, which can be used for future generalizations and developments.

In particular, we consider the two extreme cases, in which the functional J is chosen as one of the following:

$$G_r(\mu) = \begin{cases} \sum_{k\in\mathbb{N}}(a_k)^r & \text{if } \mu = \sum_{k\in\mathbb{N}} a_k\delta_{x_k} \\ +\infty & \text{otherwise} \end{cases} \qquad (0 \le r < 1)$$

whose domain is the space of purely atomic measures, or

$$F_q(\mu) = \begin{cases} \int_\Omega |u|^q \mathrm{d}x & \text{if } \mu = u \cdot \mathcal{L}^N \\ +\infty & \text{otherwise} \end{cases} \qquad (q > 1)$$

whose domain is the space $L^q(\Omega)$. We denote respectively by $\mathcal{G}_r$ the functional $\mathcal{J}$ with J replaced by G_r and by $\mathcal{F}_q$ the same functional with J replaced by F_q.

The first case is the one in which we get a "Riemannian distance" on probabilities which make paths passing through concentrated measures cheaper. The second case, on the contrary, allows only paths which lie on $L^q(\Omega)$.

In both cases we analyze the question of the existence of optimal paths γ_{opt} giving finite value to the functional. When the domain $\Omega \subset \mathbb{R}^N$ is compact we find for the first case:

- if μ_0 and μ_1 are atomic measures, then an optimal path γ_{opt} providing finite value to $\mathcal{G}_r$ always exists;
- if $r > 1 - 1/N$, then the same is true for any pair of measures;
- if $r \leq 1 - 1/N$, then there are measures μ_0 and μ_1 such that every path connecting them has an infinite cost.

Similarly, for the second case we find:

- if μ_0 and μ_1 are in $L^q(\Omega)$, then an optimal path γ_{opt} providing finite value to $\mathcal{F}_q$ always exists;
- if $q < 1 + 1/N$, then the same is true for any pair of measures;
- if $q \geq 1 + 1/N$, then there are measures μ_0 and μ_1 such that every path connecting them has an infinite cost.

It is not difficult to see that the model proposed is different and in general provides different solutions with respect to those proposed by Xia in [48] and by Maddalena, Morel and Solimini in [34]. However, among the different features our model supplies we may cite its mathematical simplicity and the possibility of performing standard numerical computations.

From the mathematical point of view, our model recalls the construction of Riemannian metrics as already pointed out, and the existence results for optimal paths is easy to prove.

The comparison with the results obtained by Xia and by Maddalena, Morel and Solimini will be important for future investigations. For instance, for the model proposed in [34] conditions to link two prescribed measures by a finite cost configuration have been studied in [23] (while in Chapter 3 or [14] and in [48] only conditions in order to link arbitrary measures are provided).

Chapter 1

Optimal Transportation Problems

In this chapter we discuss some aspects on the classical theory of mass transportation as it was proposed by Monge and subsequently developed by Kantorovich.

1.1 Original and relaxed formulation

The problem of Optimal Transportation can be simply set as follows: given two mass distributions f^+ and f^-, minimize the transport cost

$$\int_{\mathbb{R}^N} |x - t(x)| f^+(x) \, \mathrm{d}x$$

among all *transport maps* t, i.e. measurable maps such that the mass balance condition

$$\int_{t^{-1}(B)} f^+(x) \, \mathrm{d}x = \int_B f^-(y) \, \mathrm{d}y$$

holds for every Borel set B. In particular, taking $B = \mathbb{R}^N$, we must have

$$\int_{\mathbb{R}^N} f^+(x) \, \mathrm{d}x = \int_{\mathbb{R}^N} f^-(y) \, \mathrm{d}y,$$

that is f^+ and f^- carry the same mass.

Notation. Let X and Y be measure spaces and let $t : X \to Y$ be a measurable map. For any measure μ on X we define $t_\# \mu$ as the measure on Y given by

$$t_\# \mu(B) := \mu(t^{-1}(B)) \tag{1.1.1}$$

for any measurable set B.

In modern notation the problem of Monge can be set as follows.

Problem 1.1.1 (Monge's Problem). Given two finite positive Borel measures μ^+ and μ^- on a metric space X such that $\mu^+(X) = \mu^-(X)$, study the minimization of

$$M(\mu) := \int_X c(x, t(x)) \, \mathrm{d}\mu^+(x) \tag{1.1.2}$$

among all *transport maps* t, that is measurable maps $t : X \to X$ such that $\mu^-(B) = \mu^+(t^{-1}(B))$ for any measurable set B, that is $t_\# \mu^+ = \mu^-$. Here c is a generic *cost function*, that is a non-negative and lower semicontinuous function $c : X \times X \to \mathbb{R}$. We will denote by $\mathcal{M}(\mu^+, \mu^-)$ the set of transport maps between μ^+ and μ^-, and by $\mathcal{M}_{opt}(\mu^+, \mu^-)$ the subset of optimal ones.

The following proposition gives a quite raw result on the infimum value of Problem 1.1.1. We will provide better results in the next section.

Proposition 1.1.2. *The infimum of Problem 1.1.1 is larger than or equal to*

$$\sup \left\{ \int_X u \, \mathrm{d}(\mu^+ - \mu^-) \ : \ u \in \mathrm{Lip}_1(X, c) \right\}$$

where

$$\mathrm{Lip}_1(X, c) = \{ u : X \to \mathbb{R} \ : \ |u(x) - u(y)| \le c(x, y) \quad \forall x, y \in X \}.$$

Proof. Since u is 1-Lipschitz with respect to c we have

$$\begin{aligned} \int_X c(x, t(x)) \, \mathrm{d}\mu^+ &\ge \int_X |u(x) - u(t(x))| \, \mathrm{d}\mu^+ \\ &\ge \int_X (u(x) - u(t(x))) \, \mathrm{d}\mu^+ = \int_X u(x) \, \mathrm{d}(\mu^+ - \mu^-). \end{aligned}$$

Then

$$\inf_t \int_X c(x, t(x)) \, \mathrm{d}\mu^+ \ge \sup_u \int_X u(x) \, \mathrm{d}(\mu^+ - \mu^-)$$

where the infimum is taken among transport maps between μ^+ and μ^- and the supremum for $u \in \mathrm{Lip}_1(X, c)$. □

Because of the strong non-linearity in the unknown transport map t, Monge's formulation did not lead to significant advances up to 1940, when Kantorovich proposed his own formulation (see [30], [31]). Moreover, Monge's formulation shows some intrinsic difficulties as in the following examples.

Example 1.1.3 (Non-existence of transport maps). While the condition of equal total mass $\mu^+(X) = \mu^-(X)$ is a necessary condition to the existence of a transport map between μ^+ and μ^-, it is not sufficient. For example, consider $\mu^+ = \delta_{x_0}$ and $\mu^- = \frac{1}{2}\delta_{y_0} + \frac{1}{2}\delta_{y_1}$ (with $y_0 \neq y_1$). No transport maps can exist since $t_\#\delta_{x_0} = \delta_{t(x_0)}$.

Example 1.1.4 (Non-existence of the minimizer). Let S_0, S_1, S_2 be the subsets of $\mathbb{R}^2$ given by

$$\begin{aligned} S_0 &= \{(x,0)\ :\ 0 \leq x \leq 1\} \\ S_1 &= \{(x,d)\ :\ 0 \leq x \leq 1\} \\ S_2 &= \{(x,-d)\ :\ 0 \leq x \leq 1\} \end{aligned}$$

and let $\mu^+ = \mathcal{H}^1 \llcorner S_0, \mu^- = \frac{1}{2}\mathcal{H}^1 \llcorner S_1 + \frac{1}{2}\mathcal{H}^1 \llcorner S_2$. In this case the class of transport maps is not empty since the map $t : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$(x,0) \mapsto \begin{cases} (2x, d) & \text{if } 0 \leq x \leq \frac{1}{2} \\ (2x-1, -d) & \text{if } \frac{1}{2} < x \leq 1 \end{cases}$$

is a transport map. The optimal transport problem with cost $c(x,y) = |x-y|$ has no minima. This can be seen as follows. First of all, the cost of the optimal transport is at least d by Proposition 1.1.2 (use $u(x,y) = -|y|$). Actually, it is exactly d. Consider the sequence of transport maps $\{t_n\}_{n\in\mathbb{N}}$ such that t_n is linear non-decreasing between the intervals I_n^k and J_n^k if k is even and is linear non-decreasing between the intervals I_n^k and K_n^k if k is odd, where

$$\begin{aligned} I_n^k &= \left(\frac{k}{2n}, \frac{k+1}{2n}\right) \\ J_n^k &= \left(\frac{k/2}{n}, \frac{k/2+1}{n}\right) \\ K_n^k &= \left(\frac{(k-1)/2}{n}, \frac{(k+1)/2+1}{n}\right). \end{aligned}$$

It is easy to see that $\lim_n M(t_n) = d$.

On the other hand, the minimal value is achieved only if a point and its image lie on the same vertical line. Let S be the set of points lying on the same straight line. On one hand, both S_1 and S_2 carry a mass equal to $1/2$ each, so we have $\mu^-(t(S)) = \mu^+(S)/2$. On the other hand, $\mu^-(t(S)) = \mu^+(t^{-1}(t(S))) \geq \mu^+(S)$. So we must have $\mu^+(S) = 0$, that is the set of points lying on the same vertical line is negligible and the minimal value of the transport cannot be achieved.

The minimizers may not be unique (this may also happen in Kantorovich's statement of the problem) as it can be seen in the following example.

Example 1.1.5 (Non-uniqueness of the minimizer). Let us consider $\mu^+ = \frac{1}{2}\delta_{(0,0)} + \frac{1}{2}\delta_{(1,1)}$ and $\mu^- = \frac{1}{2}\delta_{(1,0)} + \frac{1}{2}\delta_{(0,1)}$ and $c(x,y) = |x-y|^p$. Then both the vertical transport and the horizontal one are optimal.

Here is another example of non-uniqueness of the optimal transport map.

Example 1.1.6 (Book shifting). Let us fix an integer number $n \geq 1$ and consider as the initial and final measures for the Monge's problem those given by $\mu^+ = \mathcal{L}^1\chi_{[0,n]}$ and $\mu^- = \mathcal{L}^1\chi_{[1,n+1]}$. From Proposition 1.1.2 (using as test function $u(t) = -t$) it follows that the value of the optimal transportation is at least n. The transportation cost for the map defined by $t(x) = x+1$ is n, so it is optimal. It can be also seen that the map

$$t(x) = \begin{cases} x+n & \text{if } t \in [0,1] \\ x & \text{if } t \in [1,n], \end{cases}$$

is optimal for every n, so we can conclude that in the case $n \geq 2$ there exist two distinct optimal transport maps. Actually, in the case $n = 1$ the map $t(x) = 2 - x$ is optimal, so even in the case $n = 1$ we do not have uniqueness of the optimal transport map.

In order to overcome the main difficulties that arise in the Monge's approach to the problem of mass transportation, Kantorovich proposed his own formulation. Kantorovich's approach is a generalization of Monge's Problem, but can also be viewed as a relaxation of it (as we will prove). In the following π^+ and π^- will be the projections of $X \times X$ on the first and on the second factor: $\pi^+(x,y) = x, \pi^-(x,y) = y$.

A great limitation in Monge's version of mass transportation problem is that the mass can be put together, but cannot be split. For example, as we have seen, no transport maps between a Dirac mass and a convex combination of Dirac masses can exist.

Problem 1.1.7 (Kantorovich's Problem). Given two finite positive Borel measures μ^+ and μ^- on a metric space X such that $\mu^+(X) = \mu^-(X)$, study the minimization of

$$K(\mu) := \int_{X\times X} c(x,y)\ \mathrm{d}\mu(x,y) \tag{1.1.3}$$

among all *transport plans* μ, i.e. positive Borel measures on $X \times X$ such that $\mu^+(A) = \mu(A \times X)$ and $\mu^-(B) = \mu(X \times B)$, that is $\pi^+_\# \mu = \mu^+$ and $\pi^-_\# \mu = \mu^-$. We will denote by $\mathcal{P}(\mu^+, \mu^-)$ the set of transport plans between μ^+ and μ^-, and by $\mathcal{P}_{opt}(\mu^+, \mu^-)$ the subset of optimal ones.

Thanks to the linearity of K with respect to μ, the subset of optimal plans $\mathcal{P}_{opt}(\mu^+, \mu^-)$ is a convex subset.

It is easy to see that if t is a transport map between μ^+ and μ^-, then $\mu_t := (\mathrm{Id} \times t)_\# \mu^+$ is a transport plan between the same measures and

$$\int_X c(x, t(x)) \, \mathrm{d}\mu^+(x) = \int_{X \times X} c(x, y) \, \mathrm{d}\mu_t(x, y).$$

So, Kantorovich's problem is a weak formulation of Monge's one. Of course, since not all transport plans are of the kind of μ_t for a suitable transport map t it may happen that the optimal value of Problem 1.1.7 is strictly less than the one of Problem 1.1.1 as in the following example (see also [38]).

Example 1.1.8. Consider again the situation of Example 1.1.4, but with a different cost function.

$$c(x, y) = \begin{cases} |x - y| & \text{if } x - y \text{ lies on a vertical line,} \\ 2|x - y| & \text{otherwise.} \end{cases}$$

In this case, the infimum of Kantorovich Problem is δ, while that of Monge Problem is 2δ since (as we have seen) a map cannot move the mass vertically.

1.2 Existence of an optimal transport plan

In this section and in the following ones we are going to prove some results in the theory of mass transportation, such as the existence of an optimal transport plan and the comparison between the infimum of Monge's Problem and Kantorovich's one.

Lemma 1.2.1. *Let f be a lower semicontinuous function defined on a metric space (X, d) with range in $[0, +\infty]$. Then the set of functions $\{g_t \; : \; t \geq 0\}$ defined by*

$$g_t(x) = \inf\{f(y) + td(x, y) \; : \; y \in X\}$$

satisfies the following properties:

- *$g_t \geq 0$;*
- *g_t is t-Lipschitz continuous;*
- *$g_t(x) \nearrow f(x)$ for every $x \in X$.*

Proof. Obviously, $g_t \geq 0$. For all y the function $x \mapsto f(y) + td(x, y)$ is t-Lipschitz continuous, so g_t is t-Lipschitz continuous too. Let now prove the third part of the Lemma. Since d is positive, the map $t \mapsto g_t$ is increasing and $g_t(x) \leq f(x)$ for every x. If $\sup_{t \geq 0} g_t(x) = +\infty$, then we have $f(x) = +\infty$ and the statement is proved. When $\sup_{t \geq 0} g_t(x) < +\infty$, let $x \in X$ and choose x_t such that

$$f(x_t) + td(x, x_t) < g_t(x) + 2^{-t}. \tag{1.2.1}$$

We have from (1.2.1)

$$\begin{aligned} td(x, x_t) &\leq g_t(x) - f(x_t) + 2^{-t} \\ &\leq g_t(x) + 2^{-t} \leq \sup_{t \geq 0} g_t(x) + 1 =: M(x) < +\infty. \end{aligned}$$

We then have $d(x, x_t) \leq M(x)/t$, so that $x_t \to x$. Passing to the limit as $t \to +\infty$ in (1.2.1), the semicontinuity of f yields

$$f(x) \leq \liminf_{t \to +\infty} f(x_t) \leq \lim_{t \to +\infty} g_t(x). \quad \square$$

Corollary 1.2.2. *In the same hypotheses of Lemma 1.2.1 there exists a sequence of continuous bounded non-negative functions h_t such that $h_t(x) \nearrow f(x)$ for every $x \in X$.*

Proof. Just define $h_t(x) = \inf\{g_t(x), t\}$. $\square$

This lemma will be useful in the next sections.

Lemma 1.2.3. *Let f be an upper semicontinuous function defined on a metric space (X, d). Assume also that f is bounded from above. Then the set of functions $\{h_t \ : \ t \geq 0\}$ defined by*

$$h_t(x) = \sup\{f(y) - td(x, y) \ : \ y \in X\}$$

satisfies the following properties:

- *$h_t \geq f$;*

- *h_t is t-Lipschitz continuous;*
- *$h_t(x) \searrow f(x)$ for every $x \in X$.*

Proof. The proof is very similar to that of Lemma 1.2.1. □

Kantorovich's Problem (Problem 1.1.7) has a solution: this is the result of the next Theorem 1.2.5. Before we prove Theorem 1.2.5 we need a lemma to establish the tightness of the set of transport plans.

Lemma 1.2.4 (Tightness of the set of transport plans). *Let μ^+ and μ^- be Borel probability measures on a Polish space X. Then the set of transport plans $\mathcal{P}(\mu^+, \mu^-)$ is tight.*

Proof. It is easy to check that $X \times X$ endowed with the product topology is still a Polish space. Thanks to Theorem A.1.2 (Ulam's Lemma), for every $\varepsilon > 0$ there exists a compact subset K_ε such that:

$$\mu^+(X \setminus K_\varepsilon) < \varepsilon, \quad \mu^-(X \setminus K_\varepsilon) < \varepsilon.$$

Let $\mu \in \mathcal{P}(\mu^+, \mu^-)$. Then

$$\begin{aligned} \mu((X \times X) \setminus (K_\varepsilon \times K_\varepsilon)) &\leq \mu((X \setminus K_\varepsilon) \times X) + \mu(X \times (X \setminus K_\varepsilon)) \\ &= \mu^+(X \setminus K_\varepsilon) + \mu^-(X \setminus K_\varepsilon) < 2\varepsilon. \end{aligned}$$

Since the compact K_ε depends only on μ^+ and μ^- (and ε of course), the proof is achieved. □

Theorem 1.2.5 (Existence of an optimal transportation plan). *Let μ^+ and μ^- be Borel probability measures on a Polish space X. Let $c : X \times X \to \overline{\mathbb{R}}_+$. Then there exists a measure μ_* such that:*

$$K(\mu_*) = \inf\{K(\mu) \ : \ \mu \in \mathcal{P}(\mu^+, \mu^-)\}.$$

Proof. Thanks to Lemma 1.2.4, the set of transport plans is tight and by Theorem A.1.4 (Prokhorov's Theorem) Kantorovich's functional is continuous when the cost function c is continuous and bounded.

In the general case of a lower semicontinuous cost function c the Kantorovich's functional 1.1.3 is lower semicontinuous. Let us consider the increasing sequence of non-negative bounded continuous functions $\{c_n\}_{n\in\mathbb{N}}$

given by Corollary 1.2.2. Thanks to the Monotone Convergence Theorem we have that

$$K_n(\mu) = \int_{X\times X} c_n(x,y)\ \mathrm{d}\mu \nearrow K(\mu) = \int_{X\times X} c(x,y)\ \mathrm{d}\mu$$

for every transport plan $\mu \in \mathcal{P}(\mu^+,\mu^-)$. K is then lower semicontinuous since it is the supremum of continuous functionals.

The existence of a minimizer is then a standard application of the direct method of the Calculus of Variations. The only thing we need now to prove is that the set of transport plans is *closed* under the weak topology. Let $\{\mu_n\}_{n\in\mathbb{N}}$ be a convergent sequence in $\mathcal{P}(\mu^+,\mu^-)$ and let us denote by μ the weak limit of $\{\mu_n\}_{n\in\mathbb{N}}$, that is $\mu_n \rightharpoonup \mu$. Then

$$\int_{X\times X} f\circ\pi^+\ \mathrm{d}\mu_n + \int_{X\times X} g\circ\pi^-\ \mathrm{d}\mu_n = \int_X f\ \mathrm{d}\mu^+ + \int_X g\ \mathrm{d}\mu^-$$

for every couple of functions $f,g\in\mathcal{C}_b(X)$. Passing to the limit as $\mu_n \rightharpoonup \mu$ we obtain:

$$\int_{X\times X} f\circ\pi^+\ \mathrm{d}\mu + \int_{X\times X} g\circ\pi^-\ \mathrm{d}\mu = \int_X f\ \mathrm{d}\mu^+ + \int_X g\ \mathrm{d}\mu^-$$

that is

$$\int_X f\ \mathrm{d}(\pi^+_\#\mu) + \int_X g\ \mathrm{d}(\pi^-_\#\mu) = \int_X f\ \mathrm{d}\mu^+ + \int_X g\ \mathrm{d}\mu^-,$$

that is $\pi^+_\#\mu = \mu^+$ and $\pi^-_\#\mu = \mu^-$.

Let $\{\mu_n\}_{n\in\mathbb{N}}$ be a minimizing sequence. By Theorem A.1.4 a convergent subsequence can be extracted, so we can suppose $\mu_n \rightharpoonup \mu_*$ for some Borel probability measure. Since the set of transport plans $\mathcal{P}(\mu^+,\mu^-)$ is closed with respect to the weak topology, we have $\mu_*\in\mathcal{P}(\mu^+,\mu^-)$ and thanks to the lower semicontinuity of the Kantorovich functional μ_* is a minimizer. □

The next theorem gives some conditions in order to assure the equality of the infima of Monge and Kantorovich's functionals.

Theorem 1.2.6. *Assume that X is a compact subset of $\mathbb{R}^N$. If the cost function c is continuous and real valued and μ^+ has no atoms, then Kantorovich functional is the lower semicontinuous envelope of Monge functional. In particular,*

$$\min_{\mu\in\mathcal{P}(\mu^+,\mu^-)} K(\mu) = \inf_{t\in\mathcal{M}(\mu^+,\mu^-)} M(t).$$

Proof. What we need to prove is that given $\mu \in \mathcal{P}(\mu^+, \mu^-)$ it is possible to provide a sequence of Borel maps $\psi_h \in \mathcal{M}(\mu^+, \mu^-)$ such that (with the notation used in the Appendix)

$$\int \delta_{\psi_n(x)} \otimes \mu^+ \rightharpoonup \mu$$

in $\mathcal{M}(X \times X)$. By Theorem A.3.1 we obtain a Borel map $\varphi : X \to X$ such that $\varphi_\# \mu^+$ is not atomic and approximates arbitrarily μ^- and $\int \delta_{\varphi(x)} \otimes \mu^+$ approximates arbitrarily μ (with respect to the weak convergence).

We can suppose that the cost function c is Lipschitz with constant L and that $|x - y| \leq c(x, y)$ (eventually multiplying c by a positive constant).

Now, let us fix an integer n. We will rename μ^+ with μ_0^0 and choose $\varphi_0 : X \to X$ such that $\varphi_{0\#}\mu_0^0$ is not atomic, $W_1(\mu^-, \varphi_{0\#}\mu_0^0) < 2^{-n}$ and

$$\int_X c(\varphi_0(x), x) \, \mathrm{d}\mu_0^0 < W_1(\mu^-, \mu_0^0) + 2^{-n},$$

where by $W_1(\mu^+, \mu^-)$ we mean infimum value of Kantorovich functional with respect to the cost c. Then, setting $\mu_0^1 = \varphi_{0\#}\mu_0^0$ we find $\varphi_1 : X \to X$ such that $\varphi_{1\#}\mu_0^1$ has no atom, $W_1(\mu^-, \varphi_{1\#}\mu_0^1) < 2^{-(n+1)}$

$$\int_X c(\varphi_1(x), x) \, \mathrm{d}\mu_0^1 < W_1(\mu^-, \mu_0^1) + 2^{-(n+1)}.$$

Setting $\mu_0^k = \varphi_{k-1\#}\mu_0^{k-1}$ we can then build by induction a Borel function $\varphi_k : X \to X$ such that $\varphi_{k\#}\mu_0^k$ has no atom, $W_1(\mu^-, \varphi_{k\#}\mu_0^k) < 2^{-(n+k)}$

$$\int_X c(\varphi_k(x), x) \, \mathrm{d}\mu_0^k < W_1(\mu^-, \mu_0^k) + 2^{-(n+k)}.$$

Now we set $\phi_0(x) = x$ and $\phi_k = \varphi_{k-1} \circ \cdots \circ \varphi_0$ for $k \geq 1$, so that $\mu_0^k = \phi_{k\#}\mu^+$. The sequence $\{\phi_k\}_{k \geq 0}$ is a Cauchy sequence in $L^1(X, \mu^+; X)$. In fact,

$$\begin{aligned} \sum_{k=0}^{+\infty} \int_X |\phi_{k+1}(x) - \phi_k(x)| \, \mathrm{d}\mu^+(x) &= \sum_{k=0}^{+\infty} \int_X |\varphi_k(y) - y| \, \mathrm{d}\mu_0^k(y) \\ &\leq 2^{1-n} + \sum_{k=0}^{+\infty} W_1(\mu^-, \mu_0^k) < +\infty. \end{aligned}$$

Now, set $\lim_k \phi_k = \psi_n$. We have $\psi_{n\#}\mu^+ = \mu^-$ and

$$\begin{aligned}\int_X c(\phi_k(x), x)\, \mathrm{d}\mu^+(x) \\ &\leq \int_X c(\varphi_0(x), x)\, \mathrm{d}\mu^+(x) + L \sum_{i=1}^{k} \int_X |\phi_i(x) - \phi_{i-1}(x)|\, \mathrm{d}\mu^+(x) \\ &\leq W_1(\mu^+, \mu^-) + 2^{-n} + L \sum_{i=1}^{k} \int_X |\varphi_i(y) - y|\, \mathrm{d}\mu_0^i(y) \\ &\leq W_1(\mu^+, \mu^-) + 2^{-n}(1 + 2L).\end{aligned}$$

Passing to the limit as $k \to +\infty$ we obtain

$$\int_X c(\psi_n(x), x)\, \mathrm{d}\mu^+(x) \leq W_1(\mu^+, \mu^-) + 2^{-n}(1 + 2L),$$

which is the result. □

1.3 Cyclical monotonicity and regularity of optimal transport plans

In this section we are going to study some concepts deeply studied in various papers, such as [27], [4], [32], [43].

In this section we will consider rather general assumptions: X and Y will be locally compact, σ-compact metric spaces. The cost function $c : X \times Y \to \overline{\mathbb{R}}$ will be a positive function. We will show some results when c is continuous and real valued, but also in a more general setting allowing c to be only lower semicontinuous and assuming extended real values.

The notions of *cyclical monotonicity* and *concavity* with respect to a cost c we are going to introduce are generalization of those given by Rockafellar in [41] in the quadratic case and in Euclidean spaces, that is when $c(x, y) = \langle x, y\rangle$.

Notation. In this section we will denote by S_n the *symmetric group*, that is the set of permutations of a set with cardinality n, $n \in \mathbb{N}$.

Definition 1.3.1 (c-cyclical monotonicity). A subset $S \subseteq X \times Y$ is said to be *c-cyclically monotone* if for any $n \in \mathbb{N}$ and for any couples $(x_i, y_i) \in S$,

$i = 1, 2, \ldots, n$ and for any permutation of n elements $\sigma \in S_n$ it is true that

$$\sum_{i=1}^{n} c(x_i, y_i) \leq \sum_{i=1}^{n} c(x_{\sigma(i)}, y_i). \tag{1.3.1}$$

Since any $\sigma \in S_n$ is the product of cycles, inequality (1.3.1) implies Definition 1.3.1 whenever it is true for any cycle σ. Moreover, by suitable rearrangement of indexes it can be seen that Definition 1.3.1 is equivalent to

$$\sum_{i=1}^{n} c(x_i, y_i) \leq \sum_{i=1}^{n} c(x_{i+1}, y_i) \tag{1.3.2}$$

for any n and any couples $(x_i, y_i) \in S$, $i = 1, 2, \ldots, n$, (with $x_{n+1} = x_1$).

Through c-cyclical monotonicity we can give a characterization of optimal measures. Recall that

$$\int_Y f \,\mathrm{d}(g_{\#}\mu) = \int_X f \circ g \,\mathrm{d}\mu \tag{1.3.3}$$

whenever $g : X \to Y$ is a measurable map between the measure space $(X, \mathcal{T}_X, \mu)$ and the measurable space $(Y, \mathcal{T}_Y)$ and $f : Y \to \overline{\mathbb{R}}$ is a non-negative measurable function.

Theorem 1.3.2. *Let $c : X \times Y \to \mathbb{R}$ be a continuous and non-negative cost. Suppose that the transport plan μ_* is optimal for Kantorovich's Problem (Problem 1.1.7). Moreover, suppose that the infimum of Kantorovich's functional* (1.1.3) *is finite, $K(\mu_*) < +\infty$. Then, the support of μ_* is c-cyclically monotone.*

Proof. Suppose on the contrary that $\operatorname{spt}\mu$ is not c-cyclically monotone. Then, there exist an integer n, a permutation σ, and couples of points $(\tilde{x}_i, \tilde{y}_i) \in \operatorname{spt}\mu_* \subseteq X \times Y$, $i \in \{1, 2, \ldots, n\}$, such that the function

$$f(x_1, y_1, \ldots, x_n, y_n) = \sum_{i=1}^{n} c(x_{\sigma(i)}, y_i) - c(x_i, y_i)$$

takes negative value on $(\tilde{x}_1, \tilde{y}_1, \ldots, \tilde{x}_n, \tilde{y}_n)$. Since f is continuous, there exist neighbourhoods $U_i \subseteq X$ and $V_i \subseteq Y$ such that f is negative on the set $\prod_{i=1}^{n} U_i \times V_i$. Moreover, we can choose the sets U_i and V_i to be relatively compact and disjoint. Let us set

$$\lambda := \inf\{\mu_*(U_i \times V_i) \ : \ i \in \{1, 2, \ldots, n\}\} > 0$$

(it is a positive number since $(\tilde{x}_i, \tilde{y}_i) \in \operatorname{spt} \mu_*$ for $i \in \{1, 2, \ldots, n\}$) and

$$\mu_i(B) = \frac{\mu_*(B \cap (U_i \times V_i))}{\mu_*(U_i \times V_i)}$$

that is the normalized restriction to $U_i \times V_i$ of μ_*. Let us now consider the product probability space

$$\left(\Omega = \prod_{i=1}^{n} U_i \times V_j,\ \mathcal{B}(\Omega),\ \eta = \otimes_{i=1}^{n} \mu_i\right)$$

and projections $\pi_i^X : \Omega \to X$ and $\pi_i^Y : \Omega \to Y$ defined as

$$\pi_i^X(x_1, y_1, \ldots, x_n, y_n) = x_i,$$
$$\pi_i^Y(x_1, y_1, \ldots, x_n, y_n) = y_i.$$

Finally, let μ' be the non-negative measure defined by:

$$\mu' = \mu_* + \lambda n^{-1} \sum_{i=1}^{n} (\pi_{\sigma(i)}^X \times \pi_i^Y)_{\#}\eta - (\pi_i^X \times \pi_i^Y)_{\#}\eta.$$

It is easy to check via formula (1.3.3) that

$$K(\mu') - K(\mu_*) = \lambda n^{-1} \int_{\Omega} c(\pi_{\sigma(i)}^X, \pi_i^Y) - c(\pi_i^X, \pi_i^Y)\ \mathrm{d}\eta < 0.$$

That implies $K(\mu') < K(\mu^*)$, in spite of the minimality of μ_*. $\square$

Not only the support of an optimal transport plan is c-cyclically monotone, but also the union of the supports of all the optimal plans is so. This is a consequence of the convexity of the set of optimal measures with fixed marginals.

Corollary 1.3.3. *The union of the supports of all the optimal transport plans is c-cyclically monotone.*

Proof. Let S be the union of the supports of all optimal transport plans. Let $(x_i, y_i) \in S$ for $i \in \{1, \ldots, n\}$, and let μ_i be an optimal measure such that $(x_i, y_i) \in \operatorname{spt} \mu_i$. It is easy to check that the measure given by

$$\mu = \frac{1}{n} \sum_{i=1}^{n} \mu_i$$

is still an optimal transport plan between μ^+ and μ^-. Since μ is an optimal measure its support is c-cyclically monotone, and by construction $(x_i, y_i) \in \operatorname{spt} \mu$. This implies that condition (1.3.1) is satisfied. $\square$

Definition 1.3.4. A function $u : X \to \mathbb{R} \cup \{-\infty\}$ is said to be *c-concave* if it is of the kind

$$u(x) = \inf_{(y,\lambda)\in\mathcal{A}} c(x,y) + \lambda$$

for some subset $\mathcal{A} \subseteq X \times \mathbb{R}$.

The usual definition of *concavity* for a function $u : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\}$ can be recovered by the one of *c-concavity*. In fact, it is easy to see that for the cost functional $c(x,y) = \frac{1}{2}|x-y|^2$ the c-concavity of u reduces to the concavity of $u(x) - \frac{1}{2}|x|^2$. Moreover, if c is a continuous function, then u is upper semicontinuous.

Before we prove Theorem 1.3.8, we just recall some definitions about the *c-transform* which will become useful in the Section 1.4. A deeper account on these facts and their proofs can be found in [41], [24] or [47].

Definition 1.3.5 (c-transform). Let X and Y be non-empty sets and $c : X \times Y \to \mathbb{R}$. Given $u : X \to \mathbb{R} \cup \{-\infty\}$, the *c-transform of* u is defined by

$$u^c(y) := \inf_{x\in X}[c(x,y) - u(x)]. \tag{1.3.4}$$

Of course, a similar definition can be given for functions $v : Y \to \mathbb{R} \cup \{-\infty\}$.

The following facts are an easy consequence of Definition 1.3.5.

Theorem 1.3.6 (Generalized Legendre duality). *Let X and Y be non-empty sets and $c : X \times Y \to \mathbb{R}$. Given a function $u : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\}$, we have*

- $u(x) + u^c(y) \leq c(x,y)$ *for every* $(x,y) \in X \times Y$;
- $u^{cc} \geq u, u^{ccc} = u^c$;
- $u^{cc} = u$ *if and only if* u *is c-concave.*

Definition 1.3.7 (c-superdifferential). Given a function $u : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\}$, the *c-superdifferential* $\partial^c u$ is the set

$$\partial^c u := \{(x,y) \in X \times Y \ : \ u(v) \leq u(x) + c(v,y) - c(x,y) \text{ for all } v \in X\}. \tag{1.3.5}$$

Moreover, we define $\partial^c u(x)$ as the set of those y such that $(x,y) \in \partial^c u$:

$$\partial^c u(x) := \{y \in Y \ : \ u(v) \leq u(x) + c(v,y) - c(x,y) \text{ for all } v \in X\}. \tag{1.3.6}$$

Theorem 1.3.8 (Rockafellar Theorem for general costs). *The following statements are equivalent for a subset $S \subseteq X \times Y$:*

1. *There exists a c-concave function $u : X \to \mathbb{R} \cup \{-\infty\}$ such that $S \subseteq \partial^c u$;*

2. *S is c-cyclically monotone.*

Proof.

- $1 \Rightarrow 2$. Let us consider $(x_i, y_i) \in \partial^c u$ for $i \in 1, \dots n$. Since u is c-concave we can assert the existence of a point $\tilde{x}$ such that u is finite at $\tilde{x}$. Thanks to the c-concavity of u

$$u(\tilde{x}) \leq u(x_i) + c(\tilde{x}, y) - c(x_i, y)$$

that is $\tilde{x}_i$ cannot be $-\infty$. Thanks to c-concavity of u again:

$$u(x_{\sigma(i)}) - u(x_i) \leq c(x_{\sigma(i)}, y_i) - c(x_i, y_i) \tag{1.3.7}$$

Summing from $i = 1, \dots, n$ inequalities (1.3.7) we get

$$0 \leq \sum_{i=1}^{n} c(x_{\sigma(i)}, y_i) - \sum_{i=1}^{n} c(x_i, y_i)$$

which is Definition 1.3.1.

- $2 \Rightarrow 1$. Fix $(x_0, y_0) \in S$ and let u be given by (recall the equivalence of Definition 1.3.1 with inequality (1.3.2)):

$$u(x) := \inf_{n,(x_i,y_i)\in S} \left(c(x, y_n) - c(x_n, y_n) + \sum_{i=1}^{n-1} [c(x_{i+1}, y_i) - c(x_i, y_i)] \right).$$

where $n \in \mathbb{N}$, and $(x_i, y_i) \in S$ for $i \in \{1, \dots, n\}$. We now prove that if $(x', y') \in S$, then $(x', y') \in \partial^c u$. By definition, u is c-concave and $u(x_0) = 0$ (the infimum is attained when $n = 1$ and $(x_1, y_1) = (x_0, y_0)$). By definition of u we also have that for every $\varepsilon > 0$

$$\left(c(x', y_n) - c(x_n, y_n) + \sum_{i=1}^{n-1} [c(x_{i+1}, y_i) - c(x_i, y_i)] \right) < u(x') + \varepsilon.$$

Setting $x_{m+1} = x'$ and $y_{m+1} = y'$ again by definition of u we have

$$u(x) \leq \left(c(x, y_{n+1}) - c(x_{n+1}, y_{n+1}) + \sum_{i=1}^{n} [c(x_{i+1}, y_i) - c(x_i, y_i)] \right)$$

so that

$$u(x) \leq c(x, y_{n+1}) - c(x_{n+1}, y_{n+1}) + u(x') + \varepsilon.$$

By the arbitrary choice of ε we get

$$u(x) - u(x') \leq c(x, y') - c(x', y').$$

Since $u(x_0) = 0$, $u(x') > -\infty$ and $y' \in \partial^c u(x')$, that is $S \subseteq \partial^c u$. □

Remark 1.3.9. Recall that the support of an optimal transport plan and the union of the supports of all optimal transport plans are c-monotone sets, so they are contained in the superdifferential of a suitable c-monotone function.

As a consequence of Theorem 1.3.8 we prove the following Corollary, which is true for continuous costs c which are a distance on $X \times X$.

Corollary 1.3.10. *Suppose that the continuous cost c is a distance on X, and let $\mu \in \mathcal{P}(X \times X)$ be a transport plan between fixed marginals μ^+ and μ^-. Then μ is optimal if and only if there exists $u : X \to \mathbb{R}$ such that*

$$|u(x) - u(y)| \leq c(x, y) \text{ for all } (x, y) \in X \times X, \tag{1.3.8}$$

$$u(x) - u(y) = c(x, y) \text{ for all } (x, y) \in \operatorname{spt} \mu. \tag{1.3.9}$$

Moreover, there exists a function u such that equations (1.3.8) *and* (1.3.9) *holds for any optimal plan μ.*

Proof.

- *Sufficiency.* Let μ^* be an admissible transport plan. Thanks to (1.3.8) and (1.3.9) we have

$$\begin{aligned} K(\mu^*) &\geq \int_{X \times X} u(x) - u(y) \, \mathrm{d}\mu^* = \int_X u \, \mathrm{d}\mu^+ - \int_X u \, \mathrm{d}\mu^- \\ &= \int_X u(x) - u(y) \, \mathrm{d}\mu = K(\mu). \end{aligned}$$

- *Necessity.* Let S be the union of the supports of all the optimal transport plans. Thanks to Theorem 1.3.8 there exists a c-concave function u such that $S \subseteq \partial^c u$. Equation (1.3.8) simply follows from the fact that the infimum of c-Lipschitz functions is itself c-Lipschitz, while equation (1.3.9) follows from $S \subseteq \partial^c u$. We have

  $$u(y) - u(x) \leq c(y, y) - c(x, y) = -c(x, y)$$

  for any $(x, y) \in \operatorname{spt} \mu$ because $S \subseteq \partial^c(u)$ and, by equation (1.3.8),

  $$u(x) - u(y) \leq c(x, y)$$

  for any $(x, y) \in X \times X$, that is equation (1.3.9). $\square$

A simple consequence of Corollary 1.3.10 is the following result.

Corollary 1.3.11 (Linear case). *Suppose that the continuous cost c is a distance on X. Then, the function in Corollary 1.3.10 is a maximizer of*

$$u \mapsto \int_X u \; \mathrm{d}(\mu^+ - \mu^-)$$

among 1-Lipschitz maps with respect to the distance c. Moreover, we have

$$\inf\{K(\mu) \; : \; \mu \in \mathcal{P}(X \times X)\} = \max\left\{\int_X u \; \mathrm{d}(\mu^+ - \mu^-) \; : \; u \in \operatorname{Lip}_1(X, c)\right\}. \tag{1.3.10}$$

Proof. For every $u \in \operatorname{Lip}_1(X, c)$, we have:

$$\begin{aligned}\int_X u(x) \; \mathrm{d}(\mu^+ - \mu^-) &= \int_X u(x) \; \mathrm{d}\mu^+ - \int_X u(y) \; \mathrm{d}\mu^- \\ &= \int_{X \times X} u(x) \; \mathrm{d}\mu - \int_{X \times X} u(y) \; \mathrm{d}\mu \\ &= \int_{X \times X} u(x) - u(y) \; \mathrm{d}\mu \leq \int_{X \times X} c(x, y) \; \mathrm{d}\mu.\end{aligned}$$

We now prove that, when μ is an optimal transport plan and u is a Kantorovich potential, equality in the previous inequality holds. In fact, we have:

$$\begin{aligned}\int_{X \times X} c(x, y) \; \mathrm{d}\mu &= \int_{\operatorname{spt} \mu} c(x, y) \; \mathrm{d}\mu = \int_{\operatorname{spt} \mu} u(x) - u(y) \; \mathrm{d}\mu \\ &= \int_{X \times X} u(x) - u(y) \; \mathrm{d}\mu = \int_{X \times X} u(x) \; \mathrm{d}\mu - \int_{X \times X} u(y) \; \mathrm{d}\mu \\ &= \int_X u(x) \; \mathrm{d}\mu^+ - \int_X u(y) \; \mathrm{d}\mu^- = \int_X u(x) \; \mathrm{d}(\mu^+ - \mu^-),\end{aligned}$$

and equality in equation (1.3.10) must hold. $\square$

Another useful consequence of Theorem 1.3.8 is the following corollary. It was proved first by Brenier in [15] and generalized in [27]. We will prove it in a slight weaker form we will use in the following when the notion of displacement convexity will be introduced.

Corollary 1.3.12 (Quadratic case). *Let us consider the Kantorovich functional associated to the cost function $c(x, y) = \frac{1}{2}|x - y|^2$ and assume that μ^+ is absolutely continuous. Then there exists a unique transport plan μ. Moreover, μ is induced by an optimal transport map which is the gradient of a convex function.*

Proof. Let $\psi : X \to \mathbb{R}$ be a c-concave function such that the graph Γ of its superdifferential contains the support of any optimal transport plan μ. Since we are considering the quadratic cost $c(x, y) = |x - y|^2/2$ it can be seen that

$$(x_0, y_0) \in \Gamma \Longleftrightarrow y_0 \in \partial^+ v(x_0)$$

where v is the concave function given by $v(x) = u(x) - |x|^2/2$. Since a concave function is almost everywhere differentiable with respect to Lebesgue measure and hence with respect to μ^+ we get that for μ^+-a.e. $x_0 \in X$ there exists a unique point y_0 such that (x_0, y_0) (i.e. $y_0 = \nabla v(x_0)$). Since $\operatorname{spt} \mu \subset \Gamma$, we have $\mu = (\mathrm{Id} \times v)_{\#}\mu^+$. □

Corollary 1.3.12 can be generalized following Gangbo and McCann in [27] or Ambrosio, Gigli and Savaré [1] for the Hilbert space setting.

Theorem 1.3.13 (Stricly convex cost case). *Let us consider the Kantorovich functional associated to a cost $c(x, y) = h(x - y)$ with h strictly convex and superlinear. Assume that μ^+ is absolutely continuous and that the minimum of Kantorovich functional is finite. Then there exists a unique transport plan μ. Moreover, μ is induced by an optimal transport map T uniquely determined μ^+-a.e. requiring that*

- $T_{\#}\mu^+ = \mu^-$;
- $T(x) = x - \nabla c^*(\nabla \varphi(x))$ *for some c-concave function φ; c^* is the Legendre Transform of c (see Definition 1.4.3).*

1.4 Kantorovich Duality Formula

In this section we will prove both the general Kantorovich duality and the particular case when the cost function is itself a distance (which will lead to the Kantovich-Rubinstein Theorem 1.4.8)

1.4.1 General duality

Kantorovich Problem admits a dual formulation, that is the minimum value of Kantorovich functional can be related to the supremum of

$$J(\varphi, \psi) = \int_X \varphi(x)\ \mathrm{d}\mu^+(x) + \int_Y \psi(y)\ \mathrm{d}\mu^-(y). \tag{1.4.1}$$

on a suitable subset of $L^1(X) \times L^1(Y)$ (or $\mathcal{C}_b(X) \times \mathcal{C}_b(Y)$ as we will see that it leads to the same value). Let us denote by $\Phi(\mu^+, \mu^-, c)$ the subset of $L^1(X) \times L^1(Y)$ given by

$$\Phi(\mu^+, \mu^-, c) := \left\{(\varphi, \psi) \in L^1(X) \times L^1(Y)\ :\ \mu^+(B^+_{\varphi,\psi}) = 0,\ \mu^-(B^-_{\varphi,\psi}) = 0\right\},$$

where we set

$$B_{\varphi,\psi} := \{(x, y) \in X \times Y\ :\ \varphi(x) + \psi(y) > c(x, y)\},$$

and

$$B^+_{\varphi,\psi} = \pi^+(B_{\varphi,\psi}), \quad B^-_{\varphi,\psi} = \pi^-(B_{\varphi,\psi}).$$

Finally, we set $\Phi_{\mathcal{C}_b}(\mu^+, \mu^-, c) = \Phi(\mu^+, \mu^-, c) \cap (\mathcal{C}_b(X) \times \mathcal{C}_b(Y))$.

Before the proof, we state the following simple lemma.

Lemma 1.4.1. *Let $(\varphi, \psi) \in \Phi(\mu^+, \mu^-, c)$. Then there exist functions $\tilde{\varphi}$ and $\tilde{\psi}$ such that $\tilde{\varphi} = \varphi$ and $\tilde{\psi} = \psi$ almost everywhere with respect to μ^+ and μ^- respectively such that $\tilde{\varphi} + \tilde{\psi} \le c$ is point-wise true.*

Proof. Let us indicate by χ_S the indicator function of the set S (defined by $\chi_S(x) = 1$, if $x \in S$, and $\chi_S(x) = 0$, if $x \notin S$). Given a couple $(\varphi, \psi) \in \Phi(\mu^+, \mu^-, c)$, set

$$\tilde{\varphi} = (1 - \chi_{B^+_{\varphi,\psi}})\varphi, \quad \tilde{\psi} = (1 - \chi_{B^-_{\varphi,\psi}})\psi.$$

By definition $\tilde{\varphi} = \varphi$ and $\tilde{\psi} = \psi$ almost everywhere with respect to μ^+ and μ^- respectively, and the inequality

$$\tilde{\varphi}(x) + \tilde{\psi}(y) \le c(x, y)$$

is point-wise true. □

We now prove Kantorovich Duality Formula.

Theorem 1.4.2 (Kantorovich Duality Formula). *Let X and Y be Polish spaces, let $\mu^+ \in \mathcal{P}(X)$ and $\mu^- \in \mathcal{P}(Y)$ be probability measures on X and Y respectively. Finally, let $c : X \times Y \to \mathbb{R}_+ \cup \{+\infty\}$ be a lower semicontinuous function. Then*

$$\begin{aligned}\inf\{K(\mu) \ : \ \mu \in \mathcal{P}(\mu^+, \mu^-)\} &= \sup\{J(\varphi, \psi) \ : \ (\varphi, \psi) \in \Phi(\mu^+, \mu^-, c)\} \\ &= \sup\{J(\varphi, \psi) \ : \ (\varphi, \psi) \in \Phi_{\mathcal{C}_b}(\mu^+, \mu^-, c)\}.\end{aligned} \tag{1.4.2}$$

Proof of Theorem 1.4.2 (Part I). Let $\tilde{\varphi}$ and $\tilde{\psi}$ be the functions of Lemma 1.4.1. If $\mu \in \mathcal{P}(\mu^+, \mu^-)$ we have:

$$\begin{aligned}\int_X \varphi \ \mathrm{d}\mu^+ + \int_Y \psi \ \mathrm{d}\mu^- &= \int_X \tilde{\varphi} \ \mathrm{d}\mu^+ + \int_Y \tilde{\psi} \ \mathrm{d}\mu^- \\ &= \int_{X\times Y} \tilde{\varphi} + \tilde{\psi} \ \mathrm{d}\mu \leq \int_{X\times Y} c(x, y) \ \mathrm{d}\mu.\end{aligned}$$

Since the previous inequality holds for every couple $(\varphi, \psi) \in \Phi(\mu^+, \mu^-, c)$ and every transport plan $\mu \in \mathcal{P}(\mu^+, \mu^-)$, taking the supremum on the left-hand side and the infimum on the right-hand side we get

$$\sup\{J(\varphi, \psi) \ : \ (\varphi, \psi) \in \Phi(\mu^+, \mu^-, c)\} \leq \inf\{K(\mu) \ : \ \mu \in \mathcal{P}(\mu^+, \mu^-)\} \tag{1.4.3}$$

which gives the first desired inequality. The other inequality to be proved, that is

$$\begin{aligned}\sup\{J(\varphi, \psi) \ : \ (\varphi, \psi) \in \Phi_{\mathcal{C}_b}(\mu^+, \mu^-, c)\} \leq \\ \sup\{J(\varphi, \psi) \ : \ (\varphi, \psi) \in \Phi(\mu^+, \mu^-, c)\},\end{aligned} \tag{1.4.4}$$

is trivial. □

To go on in the proof of Theorem 1.4.2 we need some basic results on convex analysis. We begin with the definition of *Legendre-Fenchel transform.*

Definition 1.4.3 (Legendre-Fenchel transform). Let X be a normed vector space and $f : X \to \mathbb{R} \cup \{+\infty\}$ be a convex function. The *Legendre-Fenchel transform* of f is the function $f^* : X^* \to \mathbb{R} \cup \{+\infty\}$ defined on the topological dual of X by:

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

Theorem 1.4.4 (Fenchel-Rockafellar Duality). *Let X be a normed vector space and let $F, G : X \to \mathbb{R} \cup \{+\infty\}$ be convex functionals. Let us suppose that there exists $x_0 \in X$ such that*

- *$F(x_0) < +\infty$ and $G(x_0) < +\infty$;*
- *F is continuous at x_0.*

Then,

$$\inf_{x \in X}[F(x) + G(x)] = \max_{x^* \in X^*}[-F^*(-x^*) - G^*(x^*)]. \tag{1.4.5}$$

Proof. The proof is an application of Hahn-Banach Theorem and can be found, for example, in [16] or [42]. □

Proof of Theorem 1.4.2 (Part II). We split what remains to be proved in three parts. In the first one we assume that X and Y are compact metric spaces and the cost function c is continuous, in the second we will drop compactness, but c will be uniformly continuous and bounded and in the third one we will use some approximation arguments to reach the general case.

Part II-I. Let $E = \mathcal{C}_b(X \times Y)$ be the set of continuous functions on $X \times Y$ (these functions are all bounded since the product $X \times Y$ is compact) equipped with the $\|\cdot\|_\infty$ norm. Let us define the functionals F and G on $\mathcal{C}_b(X \times Y)$ by

$$F(u) = \begin{cases} 0 & \text{if } u(x,y) \le -c(x,y) \\ +\infty & \text{otherwise} \end{cases}$$

and

$$G(u) = \begin{cases} \displaystyle\int_X \varphi \,\mathrm{d}\mu^+ + \int_Y \psi \,\mathrm{d}\mu^- & \text{if } u \in S \\ +\infty & \text{if } u \notin S, \end{cases}$$

where S is the subset of $\mathcal{C}_b(X \times Y)$ given by

$$S = \{u \in \mathcal{C}_b(X \times Y) \ : \ \exists \varphi \in \mathcal{C}_b(X), \psi \in \mathcal{C}_b(Y), \ u(x,y) = \varphi(x) + \psi(y)\}.$$

Just before going on in the proof, note that the functional G is well-defined, that is its value does not depend on the expression of u as the sum of a function of the variable x and of a function of the variable y. In other words if $u(x,y) = \varphi(x) + \psi(y) = \varphi'(x) + \psi'(y)$, then $\varphi(x) - \varphi'(x) = \psi'(y) - \psi(y)$ must be constant so that

$$\int_X \varphi(x)\ \mathrm{d}\mu^+ + \int_Y \psi(y)\ \mathrm{d}\mu^- = \int_X \varphi'(x)\ \mathrm{d}\mu^+ + \int_Y \psi'(y)\ \mathrm{d}\mu^-.$$

F and G are obviously convex and it is easy to see that the function F is continuous at $u(x,y) = 1$, so we can use Theorem 1.4.4. Let us now compute the left-hand side and right-hand side of equation (1.4.5). The left-hand side is given by

$$\inf\left\{\int_X \varphi\ \mathrm{d}\mu^+ + \int_Y \psi\ \mathrm{d}\mu^- \ :\ \varphi(x) + \psi(y) \geq -c(x,y)\right\}.$$

Let us now compute the Legendre-Fenchel transform of F and G. We have:

$$F^*(-\mu) = \sup_{u\in\mathcal{C}_b(X\times Y)}\left\{-\int_{X\times Y} u(x,y)\ \mathrm{d}\mu(x,y)\ :\ u(x,y) \geq -c(x,y)\right\} \tag{1.4.6}$$

$$= \sup_{u\in\mathcal{C}_b(X\times Y)}\left\{\int_{X\times Y} u(x,y)\ \mathrm{d}\mu(x,y)\ :\ u(x,y) \leq c(x,y)\right\}. \tag{1.4.7}$$

If μ is not a non-negative measure, then we can find a continuous function $u \in \mathcal{C}_b(X \times Y)$ such that $v \leq 0$ and $\int_{X\times Y} v\ \mathrm{d}\mu > 0$. The supremum taken over the functions given by $u = \lambda v$ with $\lambda > 0$ is $+\infty$, so $F^*(-\mu) = +\infty$ when μ is a non-negative measure. When μ is a non-negative measure, it is easy to see that the supremum is given by $\int_{X\times Y} c(x,y)\ \mathrm{d}\mu$. Then the Legendre-Fenchel transform of F is given by

$$F^*(-\mu) = \begin{cases} \displaystyle\int_{X\times Y} c(x,y)\ \mathrm{d}\mu & \text{if } \mu \in \mathcal{M}_+(X\times Y), \\ +\infty & \text{else.} \end{cases}$$

Let us now compute the Legendre-Fenchel transform of the functional G. We have

$$G^*(\mu) = \sup\left[\int_{X\times Y} u(x,y)\ \mathrm{d}\mu - \int_X \varphi(x)\ \mathrm{d}\mu^+ - \int_Y \psi(y)\ \mathrm{d}\mu^-\right] \tag{1.4.8}$$

where the supremum is taken on the subset of $\mathcal{C}_b(X \times Y)$ of functions $u \in S$. Recall then that, when μ is a transport plan between μ^+ and μ^-,

$$\int_{X\times Y} \varphi(x) + \psi(y)\ \mathrm{d}\mu = \int_X \varphi(x)\ \mathrm{d}\mu^+ + \int_Y \psi(y)\ \mathrm{d}\mu^-, \tag{1.4.9}$$

and, viceversa, when (1.4.9) holds for $(\varphi, \psi) \in \mathcal{C}_b(X) \times \mathcal{C}_b(Y)$, then μ is a transport plan between μ^+ and μ^-. When μ is a transport plan between μ^+ and μ^-, then supremum in (1.4.8) is zero. In the other cases, we can find $(\varphi', \psi') \in \mathcal{C}_b(X) \times \mathcal{C}_b(Y)$ such that

$$\int_{X \times Y} \varphi'(x) + \psi'(y) \, \mathrm{d}\mu - \int_X \varphi'(x) \, \mathrm{d}\mu^+ - \int_Y \psi'(y) \, \mathrm{d}\mu^- \neq 0.$$

The supremum of (1.4.8) on the class of functions u given by $u(x, y) = \lambda\varphi(x) + \lambda\psi(y), \lambda \in \mathbb{R}$ is then $+\infty$. Thanks to Theorem 1.4.4 we get

$$\inf\{K(\mu) \ : \ \mu \in \mathcal{P}(\mu^+, \mu^-)\} = \max\{J(\varphi, \psi) \ : \ (\varphi, \psi) \in \Phi_{\mathcal{C}_b}(\mu^+, \mu^-, c)\}. \tag{1.4.10}$$

The equality (1.4.10) combined with inequalities (1.4.3) and (1.4.4) proves Kantorovich Duality in compact metric spaces for continuous costs.

Part II-II. Let us suppose that the cost function c is bounded and uniformly continuous. Let μ_* be an optimal transport plan given by Theorem 1.2.5. Since μ_* is tight, for every $\delta > 0$ there exists compact sets $X_0 \subset X$ and $Y_0 \subset Y$ such that $\mu^+(X \setminus X_0) \leq \delta$ and $\mu^-(Y \setminus Y_0) \leq 0$. Then, it is easy to see that $\mu_*((X \times Y) \setminus (X_0 \times Y_0)) \leq 2\delta$. Let us define μ_{*0} as

$$\mu_{*0}(B) = \frac{\mu_*(B \cap (X_0 \times Y_0))}{\mu_*(X_0 \times Y_0)}.$$

μ_{*0} is a probability measure on $X_0 \times Y_0$ whose marginals on X_0 and Y_0 will be respectively indicated by μ_0^+ and μ_0^-. Up to the end of the proof we will consider Kantorovich functional on the space $X \times Y$ and on the space $X_0 \times Y_0$ and we will denote them by K and K_0 respectively, that is

$$K(\mu) = \int_{X \times Y} c(x, y) \, \mathrm{d}\mu, \quad K_0(\mu_0) = \int_{X_0 \times Y_0} c(x, y) \, \mathrm{d}\mu_0,$$

where μ and μ_0 are measures respectively on $\mathcal{M}(X \times Y)$ and $\mathcal{M}_0(X_0 \times Y_0)$. Let us consider a measure $\tilde{\mu}_0$ optimal for Kantorovich functional on $X_0 \times Y_0$, that is

$$K_0(\tilde{\mu}_0) = \inf K_0(\mu_0),$$

where μ_0 ranges among the set of transport plans between μ_0^+ and μ_0^-. We now consider the following transport plan between μ^+ and μ^-:

$$\tilde{\mu} = \mu_*(X_0 \times Y_0)\tilde{\mu}_0 + \chi_{(X_0 \times Y_0)^c}\mu_*.$$

Then, from

$$K(\tilde{\mu}) = \mu_*(X_0 \times Y_0)K_0(\tilde{\mu}_0) + \int_{(X_0\times Y_0)^c} c(x,y)\ \mathrm{d}\mu_* \leq K_0(\tilde{\mu}_0) + 2\delta\|c\|_\infty$$
$$= \inf K_0 + 2\delta\|c\|_\infty,$$

it easily follows that

$$\inf_{\mu\in\mathcal{P}(\mu^+,\mu^-)} K(\mu) \leq K_0(\tilde{\mu}_0) + 2\|c\|_\infty\delta.$$

Let now consider the analogous of functional J, but defined on $X_0 \times Y_0$, that is

$$J_0(\varphi_0,\psi_0) = \int_{X_0} \varphi_0\ \mathrm{d}\mu_0^+ + \int_{Y_0} \psi_0\ \mathrm{d}\nu_0^-,$$

which is defined on $L^1(X,\mu_0^+) \times L^1(Y,\mu_0^-)$. Thanks to Part II-I of the proof, $\inf K_0 = \sup J_0$. In particular, there exists admissible functions $\tilde{\varphi}_0, \tilde{\psi}_0$ such that

$$J_0(\tilde{\varphi}_0,\tilde{\psi}_0) \geq \sup J_0 - \delta.$$

Thanks to Lemma 1.4.1 we can suppose that $\tilde{\varphi}_0(x)+\tilde{\psi}_0(y) \leq c(x,y)$ is pointwise true. Since $J_0(0,0) = 0$, we have $\sup J_0 \geq 0$. In particular, if μ_0 is any admissible measure we can write

$$J_0(\tilde{\varphi}_0,\tilde{\psi}_0) = \int_{X\times Y} \tilde{\varphi}_0(x) + \tilde{\psi}_0(y)\ \mathrm{d}\mu_0,$$

and then the existence of (x_0,y_0) such that $\tilde{\varphi}(x_0) + \tilde{\psi}(y_0) \geq -1$ is assured. Moreover, with a careful choice of the couple $(\tilde{\varphi},\tilde{\psi})$, we get

$$\tilde{\varphi}(x_0) \geq -\frac{1}{2}, \quad \tilde{\psi}(y_0) \geq -\frac{1}{2}.$$

As a consequence, for every $(x,y) \in X_0 \times Y_0$,

$$\tilde{\varphi}_0(x) \leq c(x,y_0) - \tilde{\psi}_0(y_0) \leq c(x,y_0) + \frac{1}{2},$$
$$\tilde{\psi}_0(y) \leq c(x_0,y) - \tilde{\varphi}_0(x_0) \leq c(x_0,y) + \frac{1}{2}.$$

Let us now define for every $x \in X$

$$\overline{\varphi}_0(x) = \inf_{y\in Y_0}[c(x,y) - \tilde{\psi}_0(y)].$$

It can be easily seen that $\tilde{\varphi}_0 \leq \overline{\varphi}_0$ on X_0 and $J_0(\overline{\varphi}_0, \tilde{\psi}_0) \geq J_0(\tilde{\varphi}_0, \tilde{\psi}_0)$. Moreover, the following estimates are true

$$\overline{\varphi}_0(x) \geq \inf_{y \in Y_0} [c(x,y) - c(x_0,y)] - \frac{1}{2},$$

$$\overline{\varphi}_0(x) \leq c(x,y_0) - \tilde{\psi}_0(y_0) \leq c(x,y_0) + \frac{1}{2}.$$

We now define, for $y \in Y$,

$$\overline{\psi}_0(y) = \inf_{x \in X} [c(x,y) - \overline{\phi}_0(x)].$$

Then, $(\overline{\varphi}_0, \overline{\psi}_0) \in \Phi(\mu^+, \mu^-, c)$, $J_0(\overline{\varphi}_0, \overline{\psi}_0) \geq J_0(\overline{\varphi}_0, \tilde{\psi}_0) \geq J_0(\tilde{\varphi}_0, \tilde{\psi}_0)$ and

$$\overline{\psi}_0(y) \geq \inf_{x \in X} [c(x,y) - c(x,y_0)] - \frac{1}{2},$$

$$\overline{\psi}_0(y) \leq c(x_0,y) - \overline{\varphi}_0(x_0) \leq c(x_0,y) - \tilde{\varphi}_0(x_0) \leq c(x_0,y) + \frac{1}{2}.$$

From the inequalities above we get

$$\overline{\varphi}_0(x) \geq -\|c\|_\infty - \frac{1}{2},$$

$$\overline{\psi}_0(y) \geq -\|c\|_\infty - \frac{1}{2}.$$

Then we have:

$$\begin{aligned}
J(\overline{\varphi}_0, \overline{\psi}_0) &= \int_X \overline{\varphi}_0 \, \mathrm{d}\mu^+ + \int_Y \overline{\psi}_0 \, \mathrm{d}\mu^- = \int_{X \times Y} [\overline{\varphi}_0(x) + \overline{\psi}_0(y)] \, \mathrm{d}\mu_* \\
&= \mu_*(X_0 \times Y_0) \int_{X_0 \times Y_0} [\overline{\varphi}_0(x) + \overline{\psi}_0(y)] \, \mathrm{d}\mu_{*0} \\
&\qquad\qquad + \int_{X \times Y} [\overline{\varphi}_0(x) + \overline{\psi}_0(y)] \, \mathrm{d}\mu_{*0} \\
&\geq (1 - 2\delta) \left(\int_{X_0} \overline{\varphi}_0 \, \mathrm{d}\mu_0^+ + \int_{Y_0} \overline{\psi}_0 \, \mathrm{d}\mu_0^- \right) \\
&\qquad\qquad\qquad - (2\|c\|_\infty + 1)\mu_*((X_0 \times Y_0)^c) \\
&\geq (1 - 2\delta) J_0(\overline{\varphi}_0, \overline{\psi}_0) - 2(2\|c\|_\infty + 1)\delta \\
&\geq (1 - 2\delta) J_0(\tilde{\varphi}_0, \tilde{\psi}_0) - 2(2\|c\|_\infty + 1)\delta \\
&\geq (1 - 2\delta)(\inf K_0 - \delta) - 2(2\|c\|_\infty + 1)\delta \\
&\geq (1 - 2\delta)(\inf I - (2\|c\|_\infty + 1)\delta) - 2(2\|c\|_\infty + 1)\delta.
\end{aligned}$$

Thanks to the arbitrary choice of δ, we finally get $\sup J = \inf K$. Note that the uniform continuity of c implies the uniform continuity of $\overline{\varphi}_0$ and $\overline{\psi}_0$, then the supremum of J can be either taken over $\Phi(\mu^+,\mu^-,c)$ or $\Phi_{\mathcal{C}_b}(\mu^+,\mu^-,c)$.

Part II-III, general case. Let us now drop the uniform continuity of the cost function c. X and Y will be Polish spaces and the cost c will only be lower semicontinuous and non-negative. Thanks to Corollary 1.2.2, there exists a non-decreasing sequence of bounded non-negative and uniformly continuous costs c_n such that $c = \sup_n c_n$. Let us define K_n by

$$K_n(\mu) = \int_{X\times Y} c_n \, \mathrm{d}\mu.$$

Thanks to Part II-II, Kantorovich Duality holds for the cost c_n. Since inequality

$$\inf_{\mu\in\mathcal{P}(\mu^+,\mu^-)} K(\mu) \geq \sup_{(\varphi,\psi)\in\Phi(\mu^+,\mu^-,c)} J(\varphi,\psi)$$

is already proved, the only inequality left to show is the opposite one. This will be done if we prove first that

$$\inf_{\mu\in\mathcal{P}(\mu^+,\mu^-)} K(\mu) = \sup_n \inf_{\mu\in\mathcal{P}(\mu^+,\mu^-)} K_n(\mu) \tag{1.4.11}$$

$$\sup_{(\varphi,\psi)\in\Phi(\mu^+,\mu^-,c_n)} J(\varphi,\psi) \leq \sup_{(\varphi,\psi)\in\Phi(\mu^+,\mu^-,c)} J(\varphi,\psi) \tag{1.4.12}$$

With these two inequalities, combined with the known part of Kantorovich Duality for the costs c_n, we will get the one we are looking for.

Inequality (1.4.12) is very easy since $\Phi(\mu^+,\mu^-,c_n) \subseteq \Phi(\mu^+,\mu^-,c)$ and $\Phi_{\mathcal{C}_b}(\mu^+,\mu^-,c_n) \subseteq \Phi_{\mathcal{C}_b}(\mu^+,\mu^-,c)$.

$c_n \leq c$ implies $K_n \leq K$ and inequality

$$\inf_{\mu\in\mathcal{P}(\mu^+,\mu^-)} K(\mu) \geq \sup_n \inf_{\mu\in\mathcal{P}(\mu^+,\mu^-)} K_n(\mu)$$

follows immediately. Let us now prove

$$\inf_{\mu\in\mathcal{P}(\mu^+,\mu^-)} K(\mu) \leq \sup_n \inf_{\mu\in\mathcal{P}(\mu^+,\mu^-)} K_n(\mu).$$

Let μ_n be an optimal measure for functional K_n. Thanks to Lemma 1.2.4 and Theorem A.1.4 (Prokhorov Theorem), μ_n has a convergent subsequence. Suppose that $\mu_{n_k} \rightharpoonup \mu_*$. If $n \geq m$, we have $\inf K_n = K_n(\mu_n) \geq K_m(\mu_n)$. Then, passing to the limit with respect to n

$$\sup_n \inf_{\mathcal{P}(\mu^+,\mu^-)} K_n \geq \limsup_{n\to+\infty} K_m(\mu_n) \geq K_m(\mu_*). \tag{1.4.13}$$

By the Monotone Convergence Theorem

$$K_m(\mu_*) \to K(\mu_*) \geq \inf_{\mu \in \mathcal{P}(\mu^+,\mu^-)} K(\mu). \tag{1.4.14}$$

Gluing together inequalities (1.4.13) and (1.4.14) we finally achieve (1.4.11) and the proof of Kantorovich Duality is finally complete. □

Remark 1.4.5. As a consequence of the proof of Theorem 1.4.2, we have that, when the cost function c is bounded, the supremum in equation (1.4.2) can be taken on pairs $(\varphi^{cc}, \varphi^c)$ with φ bounded of *conjugate c-concave functions.*

1.4.2 Some other results

The following results are due to Ambrosio and Pratelli. They can be found in [4]. Theorem 1.4.6 adds some necessary and sufficiency optimality conditions.

Theorem 1.4.6. *The following facts are true.*

1. *If $\mu \in \mathcal{P}(\mu^+, \mu^-)$ is optimal and $K(\mu) < +\infty$, then μ is concentrated on a c-monotone Borel subset of $X \times Y$.*

2. *Assume that c is real-valued, $\mu \in \mathcal{P}(\mu^+, \mu^-)$ is concentrated on a c-cyclically monotone Borel subset Γ of $X \times Y$ and*

$$\mu\left(\left\{x \in X \ : \ \int_Y c(x,y)\ \mathrm{d}\mu^-(y) < +\infty\right\}\right) > 0, \tag{1.4.15}$$

$$\nu\left(\left\{y \in Y \ : \ \int_X c(x,y)\ \mathrm{d}\mu^+(x) < +\infty\right\}\right) > 0. \tag{1.4.16}$$

Then μ is optimal, $K(\mu) < +\infty$, and there exists a maximizing pair (φ, ψ) where φ is a c-concave function and $\psi = \varphi^c$.

Proof. Let (φ_n, ψ_n) be a maximizing sequence for the functional J and set $c_n = c - \varphi_n - \psi_n$. By definition, $c_n \geq 0$ and

$$\int_{X\times Y} c_n\ \mathrm{d}\mu \to 0.$$

So, we can find a sequence $c_{n(k)}$ and a Borel set Γ on which μ is concentrated, c is finite on Γ and $c_{n(k)} \to 0$ on Γ. Let us consider a finite subset of points

$\{(x_i, y_i)\}_{1\le i\le p} \subseteq \Gamma$ and a permutation σ. We have:

$$\sum_{i=1}^{p} c(x_i, y_{\sigma(i)}) \ge \sum_{i=1}^{p} \left[\varphi_{n(k)}(x_i) + \psi_{n(k)}(y_{\sigma(i)})\right] = \sum_{i=1}^{p} \left[\varphi_{n(k)}(x_i) + \psi_{n(k)}(y_i)\right]$$
$$= \sum_{i=1}^{p} \left[c(x_i, y_i) - c_{n(k)}(x_i, y_i)\right].$$

The c-cyclical monotonicity of Γ follows now as $k \to +\infty$.

In order to show the second part of the theorem, first make a partition of Γ in subsets Γ_k such that $\Gamma = \cup_k \Gamma_k$ and $c_{|\Gamma_k}$ is continuous. We now build a c-concave Borel function $\varphi : X \to [-\infty, +\infty)$ such that for μ^+-a.e. $x \in X$ we have

$$\varphi(x') \le \varphi(x) + c(x', y) - c(x, y) \text{ for every } x' \in X \text{ and } (x, y) \in \Gamma. \quad (1.4.17)$$

This fact is achieved with a generalized Rockafellar construction as in [43]. Set

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

where $(x_0, y_0) \in \Gamma_1$ is fixed and the infimum runs among all integers p and collections $\{(x_i, y_i)\}_{1\le i\le p} \subseteq \Gamma$. φ is a Borel function since it can be written as

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

where $\varphi_{p,m,l}$ is defined as the infimum among all finite subsets of p points $\{(x_i, y_i)\}_{1\le i\le p}$ of Γ_m as

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

and $\varphi_{p,m,l}$ is lower semicontinuous. With the same argument of the proof of Theorem 1.3.8 it can be checked that $\varphi(x_0) = 0$ and inequality (1.4.17) holds. Moreover, choosing $x' = x_0$ we obtain that $\varphi > -\infty$ on $\pi_X(\Gamma)$. Since μ is concentrated on Γ, $\pi_X(\Gamma)$ has full measure with respect to μ^+, that is $\varphi(x) \in \mathbb{R}$ for μ^+-a.e. x.

Now, we have to show that $\psi := \varphi^c$ is μ^--measurable, real-valued μ^--a.e. and that $\varphi + \psi = c$ on Γ. It is sufficient to study ψ on $\pi_Y(\Gamma)$ since $\mu^- = \pi_{Y\#}\mu$

is concentrated on the Borel set $\pi_Y(\Gamma)$. Thanks to inequality (1.4.17), for $y \in \pi_Y(\Gamma)$ we have

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

Consider now the disintegration of $\mu = \int_X \mu_y \otimes \mu^-$ of μ with respect to y and observe that μ_y is concentrated on Γ_y for μ^--a.e. y, therefore

$$\psi(y) = \int_X [c(x,y) - \psi(x)] \,\mathrm{d}\mu_y(x), \quad \mu^-\text{-a.e. } y \in Y.$$

Since $y \mapsto \mu_y$ is a Borel measure-valued map we obtain that ψ is μ^--measurable.

Let us now prove that φ^+ and ψ^+ are integrable with respect to μ^+ and μ^-. By condition (1.4.15) choose x such that

$$\int_Y c(x,y) \,\mathrm{d}\mu^-(y) < +\infty$$

and $\varphi(x) \in \mathbb{R}$. Since $\psi^+(y) \le c(x,y) + \varphi^-(x)$, by integration on Y, we obtain $\psi^+ \in L^1(Y,\mu^-)$. In the same way, we can prove that $\varphi \in L^1(X,\mu^+)$.

From $\varphi^+ \in L^1(X,\mu^+)$ and $\psi^+ \in L^1(Y,\mu^-)$ we deduce

$$\int_{X\times Y} (\varphi + \psi) \,\mathrm{d}\tilde{\mu} = \int_X \varphi \,\mathrm{d}\mu^+ \int_Y \psi \,\mathrm{d}\mu^- \in \mathbb{R} \cup \{-\infty\}, \quad \forall \tilde{\mu} \in \mathcal{P}(\mu^+,\mu^-).$$

Choosing $\tilde{\mu} = \mu$ and recalling that $\varphi + \psi = c$ we obtain $\int_{X\times Y} c(x,y) \,\mathrm{d}\mu < +\infty$, $\varphi \in L^1(X,\mu^+)$ and $\psi \in L^1(Y,\mu^-)$. Moreover,

$$\begin{aligned}\int_{X\times Y} c \,\mathrm{d}\tilde{\mu} \ge \int_{X\times Y} (\varphi + \psi) \,\mathrm{d}\tilde{\mu} &= \int_X \varphi \,\mathrm{d}\mu^+ + \int_Y \psi \,\mathrm{d}\mu^- \\ &= \int_{X\times Y} (\varphi + \psi) \,\mathrm{d}\mu = \int_\Gamma (\varphi + \psi) \,\mathrm{d}\mu = \int_{X\times Y} c \,\mathrm{d}\mu,\end{aligned}$$

that is μ is optimal for K and the pair (φ,ψ) is optimal for J. □

A Borel function $\varphi \in L^1(X,\mu^+)$ is a *maximal Kantorovich potential* if the pair (φ,φ^c) is a maximizer for the functional J introduced in (1.4.1). Thanks to Theorem 1.4.6 we can state the optimality of a transport plan via the optimality of a maximal Kantorovich potential. This is the content of the next theorem.

Theorem 1.4.7. *Let $\mu^+ \in \mathcal{P}(X)$ and $\mu^- \in \mathcal{P}(Y)$. Assume that c is real valued,* $\sup J = \inf K < +\infty$ *and that conditions* (1.4.15) *and* (1.4.16) *are satisfied. Then there exists a maximizing pair (φ, φ^c) for the functional J. Moreover, $\mu \in \mathcal{P}(\mu^+, \mu^-)$ is optimal if and only if*

$$\varphi(x) + \varphi^c(y) = c(x, y), \quad \mu\text{-a.e. in } X \times Y. \tag{1.4.18}$$

Proof. The existence of a maximizing pair is a consequence of the last part of the proof of Theorem 1.4.6. Suppose now that $\mu \in \mathcal{P}(\mu^+, \mu^-)$ is an optimal transport plan. From Kantorovich Duality it follows that

$$\int_{X\times Y} [c(x, y) - \varphi(x) - \varphi^c(y)] \mathrm{d}\mu = 0. \tag{1.4.19}$$

Since the integrand is positive, we must have

$$\varphi(x) + \varphi^c(y) = c(x, y), \quad \mu\text{-a.e. in } X \times Y.$$

Viceversa, if equation (1.4.18) holds then also equation (1.4.19) holds. This means $J(\varphi, \varphi^c) = K(\mu)$, that is μ has to be optimal (since (φ, φ^c) is optimal). □

1.4.3 Duality when the cost function is a distance

Recall that if $\varphi \in \mathrm{Lip}(X)$, then

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

Theorem 1.4.8 (Kantorovich-Rubinstein Theorem). *Let X be a Polish space and let d be a lower semicontinuous metric on X. Fix $\mu^+, \mu^- \in \mathcal{P}(X)$. Let us consider the Kantorovich functional associated to the cost function d, that is*

$$\mu \mapsto K_d(\mu) = \int_{X\times X} d(x, y)\ \mathrm{d}\mu.$$

Then,

$$\begin{aligned} &\inf\{K_d(\mu) \ : \ \mu \in \mathcal{P}(\mu^+, \mu^-)\} \\ &\quad = \sup\left\{\int_X \mathrm{d}(\mu - \nu) \ : \ \varphi \in \mathrm{Lip}(X) \cap L^1(\|\mu^+ - \mu^-\|), \|\varphi\|_{\mathrm{Lip}} \leq 1\right\}. \end{aligned} \tag{1.4.20}$$

Moreover, the supremum in formula (1.4.20) *does not change if we add the condition of φ bounded.*

Proof. Set

$$d_n := \frac{d}{1 + d/n}.$$

d_n is again a distance and satisfies $d_n(x,y) \nearrow d(x,y)$. Arguing as in Part II-III of Theorem 1.4.2 we reduce to prove Theorem 1.4.8 for d bounded. Then in the following we replace d by d_n and suppose d bounded. We have to show that

$$\sup_{(\varphi,\psi)\in\Phi(\mu^+,\mu^-,d)} J(\varphi,\psi) = \sup\left\{\int_X \varphi \; \mathrm{d}(\mu^+ - \mu^-) \;:\; \|\varphi\|_{\mathrm{Lip}} \le 1)\right\}.$$

Thanks to Remark 1.4.5

$$\sup_{(\varphi,\psi)\in\Phi(\mu^+,\mu^-,d)} J(\varphi,\psi) = \sup_{\varphi\in L^1(X,\mu^+)} J(\varphi^{dd},\varphi^d).$$

By definition, φ^d is 1-Lipschitz. Since φ^d is 1-Lipschitz, we have

$$-\varphi^d(x) \le d(x,y) - \varphi^d(y) \tag{1.4.21}$$

and then

$$-\varphi^d(x) \le \inf_y [d(x,y) - \varphi^d(y)].$$

On the other hand, setting $x = y$ in the left-hand side of inequality (1.4.21)

$$\inf_y [d(x,y) - \varphi^d(y)] \le d(x,x) - \varphi^d(x) = -\varphi^d(x).$$

Then, summing up,

$$-\varphi^d(x) \le \inf_{y\in X} [d(x,y) - \varphi^d(y)] \le -\varphi^d(x),$$

that is, $\varphi^{dd} = -\varphi^d$. Finally,

$$\begin{aligned}\sup_{(\varphi,\psi)\in\Phi(\mu^+,\mu^-,d)} J(\varphi,\psi) &\le \sup_{\varphi\in L^1(X,\mu^+)} J(\varphi^{dd},\varphi^d)\\ &= \sup_{\varphi\in L^1(X,\mu^+)} J(-\varphi^d,\varphi^d) \le \sup_{\|\varphi\|_{\mathrm{Lip}}\le 1} J(\varphi,-\varphi)\\ &\le \sup_{(\varphi,\psi)\in\Phi(\mu^+,\mu^-,d)} J(\varphi,\psi),\end{aligned}$$

which concludes the proof. $\square$

1.5 Wasserstein distances

1.5.1 Definition and basic properties

In this section we review some basic concepts and properties about Wasserstein distances. In particular we will relate the convergence with respect to Wasserstein distance to the weak convergence of measures.

Let X be a Polish space and let d be a complete distance which metrizes X. Given $p \geq 0$, we will denote by $\mathcal{P}_p(X)$ the set of Borel probability measures with finite momentum of order p, that is all measures $\mu \in \mathcal{P}(X)$ such that

$$\int_X [d(x_0, x)]^p \,\mathrm{d}\mu(x) < +\infty \tag{1.5.1}$$

(for $p = 0$ we define $[d(x, y)]^0$ as $\lim_{p\to 0+}[d(x, y)]^p = \chi_{\{x\neq y\}}$) for a given point x_0 (actually the finiteness of integral in formula (1.5.1) is independent on the choice of the point x_0). It is clear that when d is bounded (for example when X is compact) $\mathcal{P}(X) = \mathcal{P}_p(X)$.

Definition 1.5.1 (Wasserstein distances). Given $\mu^+, \mu^- \in \mathcal{P}_p(X)$, the *Wasserstein distance of order* p is defined by

$$W_p(\mu^+, \mu^-) := \left[\inf_{\mu\in\mathcal{P}(\mu^+,\mu^-)} \int_{X\times X} [d(x, y)]^p \,\mathrm{d}\mu(x, y)\right]^{\min\left\{1,\frac{1}{p}\right\}}. \tag{1.5.2}$$

Example 1.5.2. Via Theorem 1.4.8 the Wasserstein distance of order $p = 1$ is given also by

$$W_1(\mu^+, \mu^-) = \sup_{\|\varphi\|_{\mathrm{Lip}}\leq 1} \int_X \varphi \,\mathrm{d}(\mu^+ - \mu^-).$$

Via Theorem 1.4.8 again it can be proved that the Wasserstein distance of order $p = 0$ is just half of the total variation of $\mu^+ - \mu^-$, that is

$$W_0(\mu^+, \mu^-) = \frac{1}{2}\|\mu^+ - \mu^-\|.$$

In the rest of this section we want to prove some useful properties of this family of distances. But before we start, let us prove that Wasserstein distances are actually distances.

Theorem 1.5.3. *Given $\mu_0, \mu_1 \in \mathcal{P}_p(X)$, we have always $W_p(\mu_0, \mu_1) < +\infty$. Moreover, given $\mu_0, \mu_1, \mu_2 \in \mathcal{P}_p(X)$*

- $W_p(\mu_0, \mu_1) = 0$ *if and only if* $\mu_0 = \mu_1$*;*
- $W_p(\mu_0, \mu_1) = W_p(\mu_1, \mu_0)$*;*
- $W_p(\mu_0, \mu_2) \leq W_p(\mu_0, \mu_1) + W_p(\mu_1, \mu_2)$.

The next step before the proof of Theorem 1.5.3 is the following "gluing lemma" which we are going to prove in a quite general setting.

Lemma 1.5.4 (Gluing Lemma). *Let μ_1, μ_2, μ_3 be probability measures on the spaces on X_1, X_2, X_3 respectively and let $\mu_{12} \in \mathcal{P}(\mu_1, \mu_2)$ and $\mu_{23} \in \mathcal{P}(\mu_2, \mu_3)$. Then there exists a probability measure $\mu \in \mathcal{P}(X_1 \times X_2 \times X_3)$ such that its marginals on $X_1 \times X_2$ and $X_2 \times X_3$ are μ_{12} and μ_{23} respectively.*

Proof. Thanks to Theorem A.2.2, if μ is a probability measure on $X \times Y$ with marginal μ_0 on X there exists a map $x \in X \mapsto \mu_x \in \mathcal{P}(Y)$ such that for every $u \in \mathcal{C}_b(X \times Y)$,

$$\int_{X\times Y} u(x,y)\ \mathrm{d}\mu(x,y) = \int_X \left[\int_Y u(x,y)\ \mathrm{d}\mu_x(y)\right] \mathrm{d}\mu_0(x),$$

or, shortly,

$$\mu = \int_X (\delta_x \otimes \mu_x)\ \mathrm{d}\mu_0(x).$$

Let us now consider the disintegration of μ_{12} and μ_{23}, that is maps $X_2 \to \mathcal{P}(X_1)$ and $X_2 \to \mathcal{P}(X_3)$ denoted by $x_2 \mapsto \mu_{12,x_2}$ and $x_2 \mapsto \mu_{23,x_2}$ respectively such that

$$\mu_{12} = \int_{X_2} \mu_{12,x_2} \otimes \delta_{x_2}\ \mathrm{d}\mu_2(x_2),$$

$$\mu_{23} = \int_{X_2} \delta_{x_2} \otimes \mu_{23,x_2}\ \mathrm{d}\mu_2(x_2).$$

The measure μ we are seeking is then given by

$$\mu = \int_{X_2} \mu_{12,x_2} \otimes \delta_{x_2} \otimes \mu_{23,x_2}\ \mathrm{d}\mu_2(x_2),$$

as it can be easily checked. $\square$

Proof of Theorem 1.5.3. Clearly, $W_p(\mu_0, \mu_0) = 0$ since in this case the identity map is admissible and the transport plan induced by it realizes the infimum. Let us suppose on the contrary that $W_p(\mu_0, \mu_1) = 0$ and let μ be an optimal transport plan between μ_0 and μ_1. Then, the diagonal set in $X \times X$ has to be charged with full measure by μ and given $\varphi \in \mathcal{C}_b(X)$

$$\begin{aligned}\int_X \varphi(x)\ \mathrm{d}\mu_0(x) &= \int_{X\times X} \varphi(x)\ \mathrm{d}\mu(x,y)\\ &= \int_{X\times X} \varphi(y)\ \mathrm{d}\mu(x,y) = \int_X \varphi(y)\ \mathrm{d}\mu_1(y),\end{aligned}$$

which implies $\mu_0 = \mu_1$.

Note that if μ is a plan between μ_0 and μ_1, and $S : X \times X \to X \times X$ is the map given by $S(x,y) = (y,x)$, then $S_\#\mu$ is a plan between μ_1 and μ_0. Moreover, thanks to the symmetric properties of d,

$$\int_{X\times X} d(x,y)\ \mathrm{d}\mu(x,y) = \int_{X\times X} d(x,y)\ \mathrm{d}(S_\#\mu)(x,y).$$

The symmetric property $W_p(\mu_0,\mu_1) = W_p(\mu_1,\mu_0)$ is then straightforward.

Let us now prove the triangular inequality and the finiteness of W_p on $\mathcal{P}_p(X) \times \mathcal{P}_p(X)$. When $0 \le p \le 1$ both easily follows from the fact that $(x+y)^p \le x^p + y^p$ for every $x, y \ge 0$, while when $p > 1$ both are consequences of the Gluing Lemma 1.5.4. So we skip the proof for the case $0 \le p \le 1$ and go straight into the one of the case $p > 1$. Let μ_{01} and μ_{12} be optimal transport plans between the pairs μ_0, μ_1 and μ_1, μ_2. Let $\mu \in \mathcal{P}(X \times X \times X)$ be the measure given by the Gluing Lemma 1.5.4 and μ_{02} the image of μ onto the first and third factor of $X \times X \times X$. Then we have:

$$\begin{aligned}W_p(\mu_0,\mu_1) &\le \left(\int_{X\times X} [d(x_0,x_2)]^p\ \mathrm{d}\mu_{01}(x_0,x_2)\right)^{1/p}\\ &= \left(\int_{X\times X\times X} [d(x_0,x_2)]^p\ \mathrm{d}\mu(x_0,x_1,x_2)\right)^{1/p}\\ &\le \left(\int_{X\times X\times X} [d(x_0,x_1)+d(x_1,x_2)]^p\ \mathrm{d}\mu(x_0,x_1,x_2)\right)^{1/p}.\end{aligned}$$

Thanks to Minkowski inequality

$$\begin{aligned}&\left(\int_{X\times X\times X} [d(x_0,x_1)+d(x_1,x_2)]^p\ \mathrm{d}\mu(x_0,x_1,x_2)\right)^{1/p}\\ &\qquad\le \left(\int_{X\times X\times X} [d(x_0,x_1)]^p\ \mathrm{d}\mu(x_0,x_1,x_2)\right)^{1/p}\\ &\qquad\qquad + \left(\int_{X\times X\times X} [d(x_1,x_2)]^p\ \mathrm{d}\mu(x_0,x_1,x_2)\right)^{1/p}\end{aligned}$$

$$= \left(\int_{X\times X} [d(x_0,x_1)]^p \, \mathrm{d}\mu_{01}(x_0,x_1)\right)^{1/p}$$
$$+ \left(\int_{X\times X} [d(x_1,x_2)]^p \, \mathrm{d}\mu_{12}(x_1,x_2)\right)^{1/p}$$
$$= W_p(\mu_0,\mu_1) + W_p(\mu_1,\mu_2).$$

The triangular inequality

$$W_p(\mu_0,\mu_1) \le W_p(\mu_0,\mu_1) + W_p(\mu_1,\mu_2)$$

is then proved. Note that this inequality also implies the finiteness of W_p on $\mathcal{P}_p(X) \times \mathcal{P}_p(X)$ since for every $\mu \in \mathcal{P}(X)$

$$W_p(\mu,\delta_Q) = \left(\int_X [d(x,Q)]^p \, \mathrm{d}\mu\right)^{1/p},$$

so that when $(\mu^+,\mu^-) \in \mathcal{P}_p(X) \times \mathcal{P}_p(X)$

$$\begin{aligned} W_p(\mu_0,\mu_1) &\le W_p(\mu_0,\delta_Q) + W_p(\delta_Q,\mu_2) \\ &= \int_X [d(x,Q)]^p \, \mathrm{d}\mu^+ + \int_X [d(x,Q)]^p \, \mathrm{d}\mu^- < +\infty, \end{aligned}$$

which concludes the proof. □

Remark 1.5.5 *(Equivalence of Wasserstein distances).* Note that Jensen's inequality implies that $W_p \le W_{p'}$ whenever $p \le p'$. On the other hand it is easy to check that when $d \le R$ we have $(W_{p'}/R)^{p'} \le (W_p/R)^p$.

1.5.2 Topological properties

In this subsection the main result about Wasserstein spaces is proved. Theorem 1.5.6 will relate the convergence of probability measures with respect to the Wasserstein distance to the weak convergence. The convergence with respect to the Wasserstein distance will be proved to be equivalent to the weak one in compact spaces.

Theorem 1.5.6 (Wasserstein distances and weak convergence). *Let $1 \le p < +\infty$, let $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence of probability measures in $\mathcal{P}_p(X)$ and let $\mu \in \mathcal{P}(X)$. Then, the following statements are equivalent:*

1. $W_p(\mu_n,\mu) \to 0$ for $n \to +\infty$;

2. *$\mu_n \to \mu$ in the weak sense and the following* tightness condition *is satisfied*

$$\lim_{R\to+\infty} \limsup_{n\to+\infty} \int_{X\setminus B(x_0,R)} [d(x_0,x)]^p \, \mathrm{d}\mu_n(x) = 0 \tag{1.5.3}$$

for some point (and then for any) $x_0 \in X$;

3. *$\mu_n \to \mu$ in weak sense and the p-momenta converge, that is for some point (and then for any) $x_0 \in X$*

$$\int_X [d(x,x_0)]^p \, \mathrm{d}\mu_n(x) \to \int_X [d(x,x_0)]^p \, \mathrm{d}\mu(x) \tag{1.5.4}$$

for $n \to +\infty$;

4. *if $\varphi \in \mathcal{C}(X)$ satisfies the growth condition*

$$|\varphi(x)| \leq C[1 + d(x_0,x)^p] \tag{1.5.5}$$

for some $x_0 \in X$ and $C \in \mathbb{R}$, then

$$\int_X \varphi \, \mathrm{d}\mu_n \to \int_X \varphi \, \mathrm{d}\mu \tag{1.5.6}$$

for $n \to +\infty$.

We split the proof of Theorem 1.5.6 for convenience.

Proof of Theorem 1.5.6, Part I. First, let us reduce the proof of the theorem to the equivalence of 1 and 3.

$3 \Rightarrow 2$. We have

$$\lim_{k\to+\infty} \int_X [\inf\{d(x_0,x),R\}]^p \, \mathrm{d}\mu_k(x) = \int_X [\inf\{d(x_0,x),R\}]^p \, \mathrm{d}\mu(x);$$

on the other hand, by the Monotone Convergence Theorem,

$$\lim_{R\to+\infty} \int_X [\inf\{d(x_0,x),R\}]^p \, \mathrm{d}\mu(x) = \int_X [d(x_0,x)]^p \, \mathrm{d}\mu(x).$$

Combining the previous inequalities with (1.5.4) we get

$$\lim_{R\to+\infty} \lim_{k\to+\infty} \int_X d(x_0,x)^p - \inf\{[d(x_0,x)]^p, R^p\} \, \mathrm{d}\mu_k(x) = 0.$$

When $d(x_0,x) \geq 2R$, $d(x_0,x)^p - R^p \geq (1-2^{-p})d(x_0,x)^p$. It then follows that (1.5.3) is satisfied.

2 $\Rightarrow$ 4. Consider a function φ satisfying the growth condition (1.5.5). Then we can write

$$\varphi = \varphi_R + \psi_R$$

where $\varphi_R(x) = \inf\{\varphi(x), C(1+R^p)\}$ and $\psi_R(x) = \varphi(x) - \varphi_R(x)$. Note that

$$\psi_R(x) \leq C d(x_0,x)^p \chi_{\{x\in X \,:\, d(x_0,x)\geq R\}}.$$

Then,

$$\begin{aligned} \left| \int_X \varphi \, \mathrm{d}\mu_k - \int_X \varphi \, \mathrm{d}\mu \right| \leq \left| \int_X \varphi_R \, \mathrm{d}(\mu_k - \mu) \right| \\ + C \int_{\{x\in X \,:\, d(x_0,x)\geq R\}} d(x_0,x)^p \, \mathrm{d}\mu_k(x) \\ + C \int_{\{x\in X \,:\, d(x_0,x)\geq R\}} d(x_0,x)^p \, \mathrm{d}\mu(x). \end{aligned}$$

Then,

$$\begin{aligned} \limsup_{k\to+\infty} \left| \int_X \varphi(x) \, \mathrm{d}\mu_k - \int_X \varphi(x) \, \mathrm{d}\mu \right| \\ \leq C \limsup_{k\to+\infty} \int_{\{x\in X \,:\, d(x_0,x)\geq R\}} d(x_0,x)^p \mathrm{d}(\mu_k + \mu)(x). \end{aligned}$$

When $R \to +\infty$ condition (1.5.6) is achieved.

4 $\Rightarrow$ 3. This claim is trivial. □

This shows that what is left to prove is only the equivalence of 1 and 3. We now deal with the case the metric on X is bounded. In this case the condition of convergence of the momenta is a consequence of the weak convergence.

Proof of Theorem 1.5.6, Part II. Let us then suppose that the distance is bounded, $d \leq 1$. Recall that by Remark 1.5.5 all Wasserstein distances are equivalent to W_1, so we just need to prove the theorem for $p = 1$.

(1) $\Rightarrow$ (3). Let us assume that $W_1(\mu_n,\mu) \to 0$. Note that by Theorem 1.4.8, this is equivalent to

$$\lim_{n\to+\infty} \left(\sup_{\|\varphi\|_{\mathrm{Lip}}\leq 1} \int_X \varphi \, \mathrm{d}(\mu_n - \mu) \right) = 0. \tag{1.5.7}$$

From condition (1.5.7) we get

$$\lim_{n\to+\infty}\int_X \varphi \, \mathrm{d}\mu_n = \int_X \varphi \, \mathrm{d}\mu \tag{1.5.8}$$

for any Lipschitz function φ. But thanks to Lemma 1.2.1 and Lemma 1.2.3 any $\varphi \in \mathcal{C}_b(X)$ can be approximated by sequences of Lipschitz and uniformly bounded functions $\underline{\varphi}_n$ and $\overline{\varphi}_n$ point-wise increasing and decreasing such that

$$\lim_{n\to+\infty}\underline{\varphi}_n = \varphi = \lim_{n\to+\infty}\overline{\varphi}_n,$$

it easily follows that (1.5.8) is true also for any $\varphi \in \mathcal{C}_b(X)$.

(3) $\Rightarrow$ (1). Let us assume that $\mu_n \rightharpoonup \mu$ and fix $x_0 \in X$. If we prove that

$$\lim_{n\to+\infty}\left(\sup_{\varphi\in \mathrm{Lip}_{1;x_0}(X)}\int_X \varphi \, \mathrm{d}(\mu_n-\mu)\right) = 0,$$

where Lip_{1,x_0} is the set of Lipschitz function vanishing at x_0 and with a Lipschitz constant less or equal than 1, we are done since condition (1.5.7) is implied. From Theorem A.1.4 we obtain a sequence of compact sets K_n such that $\sup_k \mu_k(K_n) \le 1/n$ and $\mu(K_n^c) \le 1/n$. We assume that $x_0 \in K_1$. Then, for every $n \ge 1$ the set

$$\{\varphi\chi_{K_n} \ : \ \varphi \in \mathrm{Lip}_{1,x_0}\}$$

is a subset of Lip_{1,x_0} and by Ascoli-Arzelà Theorem it is a compact subset of $\mathcal{C}_b(K_n)$. Via a diagonal argument it can be then proved that for any sequence of $\varphi_n \in \mathrm{Lip}_{1,x_0}(X)$ we can extract a subsequence uniformly convergent on each K_n to some bounded Lipschitz (because uniform limit of a sequence of uniformly bounded and uniformly Lipschitz functions) function φ_∞ defined on $\cup_n K_n$. We will apply this statement to a family φ_n such that

$$\sup_{\varphi\in \mathrm{Lip}_{1,x_0}(X)}\int_X \varphi \, \mathrm{d}(\mu_n-\mu) \le \int_X \varphi_n \, \mathrm{d}(\mu_n-\mu) + \frac{1}{n},$$

finding a convergent subsequence φ_n (not relabelled) converging uniformly on each K_n to a function $\varphi_\infty \in \mathrm{Lip}(\cup_n K_n)$. The function φ_∞ can be extended to a function in $\mathrm{Lip}_1(X)$. We then have that

$$\begin{aligned}\int_X \varphi_k \, \mathrm{d}(\mu_k-\mu) \le \left|\int_{K_n}(\varphi_k-\varphi_\infty)\, \mathrm{d}(\mu_k-\mu)\right| + \left|\int_{K_n^c}(\varphi_k-\varphi_\infty)\, \mathrm{d}(\mu_k-\mu)\right| \\ + \left|\int_X \varphi_\infty \, \mathrm{d}(\mu_k-\mu)\right|.\end{aligned}$$

For any n, thanks to the uniform convergence of φ_k to φ_∞,

$$\lim_{k\to+\infty}\left|\int_{K_n}(\varphi_k-\varphi_\infty)\,\mathrm{d}(\mu_k-\mu)\right|=0.$$

Since the functions φ_k and φ_∞ are uniformly bounded in k, then

$$\left|\int_{K_n^c}(\varphi_k-\varphi_\infty)\,\mathrm{d}(\mu_k-\mu)\right|\leq C(\mu_k(K_n^c)+\mu(K_n^c))\leq\frac{2C}{n},$$

then as n approaches $+\infty$ the second addendum goes to 0 uniformly with respect to k. The third addendum goes to zero because of the weak convergence of μ_k to μ. The proof is then achieved passing to the limit $n\to+\infty$ and then $k\to+\infty$. □

Lemma 1.5.7. *Let X be a Polish space and d its distance. Given $R>0$, let W_p and $W_{p,R}$ be Wasserstein distances with respect to the distance d and $d_R:=\inf\{d,R\}$. Let μ_n and μ_0 be in $\mathcal{P}_p(X)$. Then,*

$$W_p(\mu_n,\mu_0)\to 0\iff W_{p,R}(\mu_n,\mu_0)\to 0.$$

Proof. Since $d\geq d_R$ from $W_p(\mu_n,\mu_0)\to 0$ follows that $W_{p,R}(\mu_n,\mu_0)\to 0$. To prove the other claim, let us consider a transport plan between μ_n and μ_0

$$\begin{aligned}
W_p^p(\mu_n,\mu_0)&\leq\int_{X\times X}[d(x,y)]^p\,\mathrm{d}\mu\\
&=\int_{\{d\leq R\}}[d(x,y)]^p\,\mathrm{d}\mu+\int_{\{d\leq R\}^c}[d(x,y)]^p\,\mathrm{d}\mu\\
&\leq\int_{\{d\leq R\}}[d(x,y)]^p\,\mathrm{d}\mu\\
&\qquad+2^{p-1}\left[\int_{\{d>R\}}[d(x,z_0)]^p\,\mathrm{d}\mu_n+\int_{\{d>R\}}[d(z_0,y)]^p\,\mathrm{d}\mu_0\right]\\
&\leq\int_{X\times X}[d_R(x,y)]^p\,\mathrm{d}\mu\\
&\qquad+2^{p-1}\left[\int_{\{d>R\}}[d(x,z_0)]^p\,\mathrm{d}\mu_n+\int_{\{d>R\}}[d(z_0,y)]^p\,\mathrm{d}\mu_0\right],
\end{aligned}$$

and passing to the infimum on the transport plan μ we get

$$W_p^p(\mu_n,\mu_0)\leq W_{p,R}^p(\mu_n,\mu_0)+\left[\int_{\{d>R\}}[d(x,z_0)^p]\,\mathrm{d}\mu_n+\int_{\{d>R\}}[d(z_0,y)^p]\,\mathrm{d}\mu_0\right].$$

Recall that, by Part II of the proof of Theorem 1.5.6, the convergence with respect to $W_{p,R}$ is equivalent to the weak convergence. Passing to the limit as $n\to+\infty$ and then to the limit as $R\to+\infty$ we prove the other claim. □

We finally deal with the case of an unbounded distance.

Proof of Theorem 1.5.6, Part III. First of all, note that $\mu_k \rightharpoonup \mu$ implies

$$\int_X [d(x_0,x)]^p \,\mathrm{d}\mu(x) = \lim_{R\to+\infty}\lim_{k\to+\infty}\int_X [\inf\{d(x_0,x),R\}]^p \,\mathrm{d}\mu_k(x)$$
$$\leq \liminf_{k\to+\infty}\int_X [d(x_0,x)]^p \,\mathrm{d}\mu_k(x).$$

So, the convergence of the momenta is equivalent to

$$\limsup_{k\to+\infty}\int_X [d(x_0,x)]^p \,\mathrm{d}\mu_k(x) \leq \int_X [d(x_0,x)]^p \,\mathrm{d}\mu(x).$$

The proof will be then complete if we prove that the convergence with respect W_p implies the above inequality. Recall that given $\varepsilon > 0$ there exists a constant C_ε such that for all nonnegative real numbers a, b we have

$$(a+b)^p \leq (1+\varepsilon)a^p + C_\varepsilon b^p.$$

We then have

$$d(x_0,x)^p \leq (1+\varepsilon)[d(x_0,y)]^p + C_\varepsilon[d(x,y)]^p.$$

Now let us consider a sequence μ_k such that $W_p(\mu_k,\mu)\to 0$ and as sequence of optimal transport plans π_k between μ_k and μ. We then have

$$\int_X [d(x_0,x)]^p \,\mathrm{d}\mu_k(x) \leq (1+\varepsilon)\int_X [d(x_0,y)]^p \,\mathrm{d}\mu(y)$$
$$+ C_\varepsilon \int_{X\times Y} [d(x,y)]^p \,\mathrm{d}\pi_k(x,y).$$

The second addendum goes to zero being equal to $W_p^p(\mu_k,\mu)$ up to a constant. The inequality reduces to

$$\limsup_{k\to+\infty}\int_X [d(x_0,x)]^p \,\mathrm{d}\mu_k(x) \leq (1+\varepsilon)\int_X [d(x_0,x)]^p \,\mathrm{d}\mu(x),$$

and passing to the limit $\varepsilon\to 0$ the proof is achieved. $\square$

1.6 Displacement convexity

In this section we will present some useful results due to McCann (see [35]) and generalized by Ambrosio, Gigli and Savaré (see [1]). The results we are going to show refers to the case $p = 2$ even though in Chapter 3 we will use it in the general case $p > 1$. The missing proofs can be found [1].

1.6.1 The case $p = 2$

Given two absolutely continuous (with respect to the Lebesgue measure) μ_0, μ_1 by Theorem 1.3.12 there exists a unique gradient of a convex function ϕ such that $(\nabla\phi)_\# \mu_0 = \mu_1$ which is optimal for Monge problem. Thanks to ϕ we can build an interpolation curve between μ_0 and μ_1 in a particular way which will be very useful.

Definition 1.6.1 (Displacement interpolation). Given two absolutely continuous (with respect to the Lebesgue measure) μ_0, μ_1, let $\nabla\phi$ the function given by Theorem 1.3.12. The *displacement interpolation* between μ_0, μ_1 is defined by

$$[\mu_0, \mu_1]_t := [(1-t)\,\mathrm{Id} + t\nabla\phi]_\# \mu_0. \tag{1.6.1}$$

Remark 1.6.2. Since $(1-t)\,\mathrm{Id} + t\nabla\phi = \nabla\phi_t$ with $\phi_t = [(1-t)|\cdot|^2/2 + t\phi]$ is the gradient of a convex function, it is the optimal map between μ_0 and $\mu_t := [\mu_0, \mu_1]_t$. Then, we have:

$$\begin{aligned} W_2^2(\mu_0, \mu_t) &= \int_{\mathbb{R}^n} |x - [(1-t)x + t\nabla\phi(x)]|^2 \, \mathrm{d}\mu^+(x) \\ &= t^2 \int_{\mathbb{R}^n} |x - \nabla\phi(x)|^2 \, \mathrm{d}\mu^+(x) = t^2 W_2^2(\mu_0, \mu_1). \end{aligned}$$

Then, $W_2(\mu_0, \mu_t) = tW_2(\mu_0, \mu_1)$, that is the curve $t \mapsto \mu_t$ is a geodesic between μ_0 and μ_1.

Proposition 1.6.3 (Basic properties). *The following statements are true:*

1. $[\mu, \nu]_t = [\nu, \mu]_{1-t}$*;*
2. $[[\mu, \nu]_t, [\mu, \nu]_{t'}]_s = [\mu, \nu]_{(1-s)t+st'}$*;*
3. *if μ or ν are absolutely continuous with respect to the Lebesgue measure, then also $[\mu, \nu]_t$ is so for every $t \in (0, 1)$.*

Proof. We will recall some properties of convex functions.

1. First, note that $\nabla\phi^* = (\nabla\phi)^{-1}$, so $\mu_0 = (\nabla\phi^*)_\# \mu_1$. Then we have:

$$\begin{aligned} [\mu_0, \mu_1]_t &= [(1-t)\,\mathrm{Id} + t\nabla\phi]_\# \mu_0 = [(1-t)\,\mathrm{Id} + t\nabla\phi]_\# (\nabla\phi^*)_\# \mu_1 \\ &= [(1-t)\nabla\phi^* + t\,\mathrm{Id}]_\# \mu_1 = [\mu_1, \mu_0]_{1-t}. \end{aligned}$$

2. It follows from a straightforward computation.

3. Our claim is that if B is a Borel set of zero Lebesgue measure, then $(\nabla\phi_t)^{-1}(B) = 0$. Let us consider the function ϕ_t defined by $\phi_t(x) = (1-t)|x|^2/2+t\phi(x)$. The convexity of ϕ implies the strict convexity of ϕ_t so that $\nabla\phi$ must be a single-valued function on its domain. Moreover, since

$$\begin{aligned}|\nabla\phi_t(x) - \nabla\phi_t(y)||x-y| &\geq \langle\nabla\phi_t(x) - \nabla\phi_t(y), x-y\rangle \\ &= (1-t)|x-y|^2 + t\langle\nabla\phi(x) - \nabla\phi(y), x-y\rangle.\end{aligned}$$

This shows that $(\nabla\phi_t)^{-1}$ is $(1-t)^{-1}$-Lipschitz, so by standard geometric measure results $(\nabla\phi_t)^{-1}(B) = 0$ for every Borel set B of null Lebesgue measure.

The proof is then concluded. □

Remark 1.6.4. An alternative way to define the *displacement interpolation* would be to consider the optimal transport plan μ of Kantorovich Problem (the uniqueness of the optimal transport plan is guaranteed by Theorem 1.3.12 in the case of an absolutely continuous pair of measures) and then define the displacement interpolation as

$$[\mu_0, \mu_1]_t := (\Phi_t)_\#\mu, \tag{1.6.2}$$

where $(\Phi_t)_\#$ is the map defined by $\Phi_t(x,y) = (1-t)x + ty$. Of course, in this case the displacement interpolation between them may not be unique. This alternative notion will be useful in the generalization to the case $p > 1$ in Section 1.6.3. We will go deeper in the next section.

We now go on with the notion of convexity we are interested in, that is *displacement convexity.*

Definition 1.6.5 (Displacement convex sets). Let us denote by $\mathcal{P}_{ac}(\mathbb{R}^N)$ the subset of $\mathcal{P}(\mathbb{R}^N)$ which are absolutely continuous with respect to the Lebesgue measure. A subset of $\mathcal{P} \subseteq \mathcal{P}_{ac}(\mathbb{R}^N)$ is said to be *displacement convex* if for all $\mu_0, \mu_1 \in \mathcal{P}$ the displacement interpolation satisfies $[\mu_0, \mu_1]_t \in \mathcal{P}$.

Note that $\mathcal{P}_{ac}(\mathbb{R}^N)$ is a displacement convex set.

Definition 1.6.6 (Displacement convex functionals). Let $\mathcal{P} \subseteq \mathcal{P}_{ac}$ a displacement convex set and $F : \mathcal{P} \to \mathbb{R} \cup \{+\infty\}$. F is said to be *displacement convex* on $\mathcal{P}$ if for every $\mu_0, \mu_1 \in \mathcal{P}$ the function

$$t \mapsto F([\mu_0, \mu_1]_t)$$

is convex on the interval $[0, 1]$.

Remark 1.6.7. In view of the alternative definition of displacement interpolation given by formula (1.6.2), Definition 1.6.5 and Definition 1.6.6 can be generalized requiring that the characterizing condition holds for at least one displacement interpolation (recall that the uniqueness is no more assured).

Definition 1.6.8 (Displacement convexity, again). A functional F defined on a displacement convex subset $\mathcal{P} \subseteq \mathcal{P}_{ac}$ is said to be *strictly displacement convex* if for every couple $\mu_0, \mu_1 \in \mathcal{P}_{ac}(\mathbb{R}^N)$ such that $\mu_0 \neq \mu_1$ the function

$$t \mapsto F([\mu_0, \mu_1]_t)$$

is strictly convex on $[0, 1]$. It is said to be λ*-uniformly displacement convex* for some $\lambda \in \mathbb{R}_+$ if given μ_0, μ_1 we have

$$\frac{d^2}{dt^2} F([\mu_0, \mu_1]_t) \geq \lambda W_2^2(\mu_0, \mu_1)$$

It is said to be *displacement semiconvex* if for some constant $C > 0$ we have

$$\frac{d^2}{dt^2} F([\mu_0, \mu_1]_t) \geq -C W_2^2(\mu_0, \mu_1).$$

In the next part we will study some functionals which will come out to be displacement convex. These results have been developed by McCann in [35].

1.6.2 Displacement convex functionals, case $p = 2$

The functionals we are going to consider are of three types.

The first type is the *internal energy functional* $\mathcal{U}$ and is defined as $\mathcal{U}(\mu) := +\infty$ on $\mu \in \mathcal{P}_2(\mathbb{R}^N) \setminus \mathcal{P}_{ac}(\mathbb{R}^N)$ and as

$$\mathcal{U}(\mu) := \int_{\mathbb{R}^N} U(\rho(x))\mathrm{d}x \tag{1.6.3}$$

if $\mu = \rho\mathcal{L}^N$ is an absolutely continuous measure given by the density ρ. The function U is called *internal energy density* and will be a measurable function $U : \mathbb{R}_+ \to \mathbb{R}_+ \cup \{+\infty\}$ such that $U(0) = 0$ and U not identically $+\infty$ on $\mathbb{R}_+ \setminus \{0\}$.

The second kind is the *potential energy functional* $\mathcal{V}$ which is defined as $\mathcal{V}(\mu) := +\infty$ on $\mu \in \mathcal{P}_p(\mathbb{R}^N) \setminus \mathcal{P}_{ac}(\mathbb{R}^N)$ and as

$$\mathcal{V}(\mu) := \int_{\mathbb{R}^N} V(x)\rho(x)\mathrm{d}x \tag{1.6.4}$$

if $\mu = \rho\mathcal{L}^N$ is an absolutely continuous measure and its density is given by ρ. The function V will be a measurable function $\mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ bounded from below.

The last kind is then the *interaction energy functional* $\mathcal{W}$ defined by $\mathcal{W}(\mu) := +\infty$ on $\mathcal{P}_p(\mathbb{R}^N) \setminus \mathcal{P}_{ac}(\mathbb{R}^N)$ and by

$$\mathcal{W}(\mu) := \frac{1}{2}\int_{\mathbb{R}^N} W(x-y)\rho(x)\rho(y)\mathrm{d}x\mathrm{d}y \tag{1.6.5}$$

if $\mu = \rho\mathcal{L}^N$ is an absolutely continuous measure with respect to the Lebesgue measure with density ρ. The function W will be a measurable function $\mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ bounded from below.

The next three theorems are main results due to McCann in [35].

Theorem 1.6.9 (Displacement convexity of internal energy). *Let $\mathcal{P}$ be a displacement convex subset of $\mathcal{P}_{ac}(\mathbb{R}^N)$. Then, if U satisfies $U(0) = 0$ and the map*

$$r \mapsto \psi(r) := r^N U(r^{-N})$$

is convex non-increasing on $]0, +\infty[$, then $\mathcal{U}$ is displacement convex on $\mathcal{P}$.

Theorem 1.6.10 (Displacement convexity of potential energy). *Let $\mathcal{P}$ be a displacement convex subset of $\mathcal{P}_2(\mathbb{R}^N)$. Convexity, strict convexity, λ-uniform convexity, semiconvexity of constant C of V imply respectively the displacement convexity, strict displacement convexity, λ-uniform displacement convexity, displacement semiconvexity of constant C of $\mathcal{V}$.*

Theorem 1.6.11 (Displacement convexity of interaction energy). *Let $\mathcal{P}$ be a displacement convex subset of $\mathcal{P}_2(\mathbb{R}^N)$. Convexity, semiconvexity of constant C of W imply respectively the displacement convexity, displacement semiconvexity of constant C of $\mathcal{W}$. Strict convexity, λ-uniform convexity of W respectively imply strict displacement convexity and λ-uniform convexity of $\mathcal{W}$ on the subspace $\mathcal{P}_m$ of $\mathcal{P}$ of measures having a given $m \in \mathbb{R}^N$ as mean.*

Notation. Let μ_0 and μ_1 be probability measures, ρ_0 and ρ_1 their densities. In the following we will write the displacement interpolation as

$$\mu_t = (\mathrm{Id} - t\theta)_{\#}\mu_0,$$

with $\theta = \mathrm{Id} - \nabla\varphi$ and φ convex.

Proof of Theorem 1.6.9, internal energy. As a consequence of Theorem 4.4 of [35], we have

$$\mathcal{U}(\mu_t) = \int_{\mathbb{R}^N} U\left(\frac{\rho_0}{\det(\mathrm{Id} - t\nabla\theta(x))}\right) \det(\mathrm{Id} - t\nabla\theta(x))\,\mathrm{d}x \qquad (1.6.6)$$

Note that the integrand in equation (1.6.6) is the composition of $t \mapsto \lambda = \det(\mathrm{Id} - tS)^{1/N}$ and $\lambda \mapsto U(r/\lambda^N)\lambda^N$, where $r = \rho(x)$ and $S = \nabla\theta(x)$ is a symmetric matrix and $S \leq \mathrm{Id}$. The convexity of the integrand of equation (1.6.6) and of the whole integral will be established if we prove the concavity of the map $t \mapsto \det(\mathrm{Id} - tS)^{1/N}$. This will follow from Lemma 1.6.12. $\square$

Lemma 1.6.12 (Arithmetic-geometric inequality). *The following statements are true.*

1. *Let $\{x_i\}_{1\leq i\leq n}$ and $\{\lambda_i\}_{1\leq i\leq n}$ be real numbers satisfying $x_i \geq 0, \lambda_i \geq 0, \sum_{i=1}^n \lambda_i = 1$. Then,*

$$\sum_{i=1}^n \lambda_i x_i \geq \prod_{i=1}^n x_i^{\lambda_i}. \qquad (1.6.7)$$

2. *Let A and B be two nonnegative symmetric matrices and $\lambda \in [0,1]$. Then*

$$\det(\lambda A + (1-\lambda)B)^{1/N} \geq \lambda(\det A)^{1/N} + (1-\lambda)(\det B)^{1/N}. \quad (1.6.8)$$

3. *Let A and B be two nonnegative symmetric matrices and $\lambda \in [0,1]$. Then*

$$\det(\lambda A + (1-\lambda)B) \geq (\det A)^{\lambda}(\det B)^{1-\lambda}. \qquad (1.6.9)$$

Proof. For sake of completeness we give a full proof of these inequalities.

1. It is an immediate consequence of the concavity of the logarithmic function.

2. First, it is sufficient to prove

$$\det(A+B)^{1/N} \geq (\det A)^{1/N} + (\det B)^{1/N}$$

Thanks to a density argument, we may also suppose that A is invertible. So, we must prove that

$$\det(\mathrm{Id} + C)^{1/N} \geq \det(\mathrm{Id})^{1/N} + (\det C)^{1/N} = 1 + (\det C)^{1/N},$$

with C symmetric and nonnegative. C can be diagonalized and if $c_i, 1 \leq i \leq N$ are its eigenvalues, then our inequality reduces to

$$\prod_{i=1}^{N} (1+c_i)^{1/N} \geq 1 + \left(\prod_{i=1}^{N} c_i \right)^{1/N},$$

which is a consequence of the arithmetic-geometric inequality (1.6.7).

3. Consider inequality (1.6.8) and apply the well known inequality

$$\lambda x + (1-\lambda) y \geq x^{\lambda} y^{1-\lambda}$$

(inequality (1.6.7) with $n = 2$). Setting $x = (\det A)^{1/N}$ and $y = (\det B)^{1/N}$. We then have

$$\begin{aligned} \det(\lambda A + (1-\lambda)B)^{1/N} &\geq \lambda (\det A)^{1/N} + (1-\lambda)(\det B)^{1/N} \\ &\geq (\det A)^{\lambda/N} (\det B)^{(1-\lambda)/N}. \end{aligned}$$

We achieve the third claim raising to the N-th power.

This concludes the proof. □

For sake of completeness we will prove Theorems 1.6.10 and 1.6.11, even though we will not use these results later.

Proof of Theorem 1.6.10, potential energy. This is the easiest case. We have:

$$\mathcal{V}(\rho_t) = \int_{\mathbb{R}^N} V(x) \, \mathrm{d}\mu_t(x) = \int_{\mathbb{R}^N} V(x - t\theta(x)) \, \mathrm{d}\mu_0(x), \qquad (1.6.10)$$

so, the convexity of V implies that of $\mathcal{V}$. All the other convexity properties (strict displacement convexity, λ-uniform displacement convexity, displacement semiconvexity) follow easily from equation (1.6.10).

If V is strictly convex, the function $t \mapsto \mathcal{V}(\rho_t)$ is not strictly convex only if $\theta(x) = 0$ for μ-a.e. $x \in \mathbb{R}^N$, that is $\mu_0 = \mu_1$.

If V is λ-uniformly convex, then

$$\begin{aligned}
&\sigma\mathcal{V}(\rho_{t_1}) + (1-\sigma)\mathcal{V}(\rho_{t_2}) - \mathcal{V}(\rho_{\sigma t_1+(1-\sigma)t_2}) = \\
&\qquad = \int_{\mathbb{R}^N} [\sigma V(x - t_1\theta(x)) + (1-\sigma)V(x - t_2\theta(x)) \\
&\qquad\quad - V(\sigma(x - t_1\theta(x)) + (1-\sigma)(x - t_2\theta(x))] \, \mathrm{d}\mu_0(x) \\
&\qquad \geq \lambda \frac{\sigma(1-\sigma)}{2} \int_{\mathbb{R}^N} |(x - t_1\theta(x)) - (x - t_2\theta(x))|^2 \, \mathrm{d}\mu_0(x) \\
&\qquad = \lambda \frac{\sigma(1-\sigma)}{2} \left[\int_{\mathbb{R}^N} |\theta(x)|^2 \, \mathrm{d}\mu_0(x) \right] (t_1 - t_2)^2.
\end{aligned}$$

Since $\int |\theta(x)|^2 \, \mathrm{d}\mu(x) = W_2^2(\mu_0, \mu_1)$, the map $t \mapsto \mathcal{V}(\mu_t)$ is displacement convex of constant λ. $\square$

Proof of Theorem 1.6.11, interaction energy. First of all, note that we may replace W with its symmetric part, $W^S(z) = (W(z) + W(-z))/2$. Let us write

$$\mathcal{W}(\rho_t) = \frac{1}{2} \int_{\mathbb{R}^N} W([x-y] - t[\theta(x) - \theta(y)])\rho_0(x)\rho_0(y)\mathrm{d}x\mathrm{d}y.$$

So $\mathcal{W}$ is convex as soon as W is convex.

If W is strictly convex, then the function $t \mapsto \mathcal{W}(\rho_t)$ may be not strictly convex in the only case for some θ_0 we have $\theta(x) = \theta_0$ for μ-a.e. $x \in \mathbb{R}^N$. This means that μ_0 and μ_1 are obtained one from the other through a translation (which is excluded by the fact that the center of mass is fixed).

If W is λ-uniformly convex, we can write:

$$\begin{aligned}
&\sigma\mathcal{W}(\rho_{t_1}) + (1-\sigma)\mathcal{W}(\rho_{t_2}) - \mathcal{W}(\rho_{\sigma t_1+(1-\sigma)t_2}) \\
&\qquad \geq \frac{1}{2}\lambda \frac{\sigma(1-\sigma)}{2} \left[\int_{\mathbb{R}^N \times \mathbb{R}^N} |\theta(x) - \theta(y)|^2 \rho_0(x)\rho_0(y)\mathrm{d}x\mathrm{d}y \right] (t_1 - t_2)^2.
\end{aligned}$$

Since the center of mass is fixed, we have

$$\int_{\mathbb{R}^N \times \mathbb{R}^N} |\theta(x) - \theta(y)|^2 \, \mathrm{d}\mu_0(x)\mathrm{d}\mu_0(y) = 2\int_{\mathbb{R}^N} |\theta(x)|^2 \, \mathrm{d}\mu_0(x) = 2W_2^2(\mu_0, \mu_1).$$

Indeed, since $(\mathrm{Id} - \theta)_\# \mu_0 = \mu_1$, we have

$$\int_{\mathbb{R}^N} (x - \theta(x)) \, \mathrm{d}\mu_0(x) = \int_{\mathbb{R}^N} y \, \mathrm{d}\mu_1(y).$$

Since μ_0 and μ_1 have the same center of mass, we obtain

$$\int_{\mathbb{R}^N} \theta(x)\ \mathrm{d}\mu_0(x) = 0.$$

Finally,

$$\int_{\mathbb{R}^N\times\mathbb{R}^N} |\theta(x)-\theta(y)|^2\ \mathrm{d}\mu_0(x)\mu_0(y) =$$
$$2\int_{\mathbb{R}^N} |\theta(x)|^2\ \mathrm{d}\mu_0(x) - 2\left|\int_{\mathbb{R}^N} \theta(x)\ \mathrm{d}\mu_0\right|^2 = 2\int_{\mathbb{R}^N} |\theta(x)|^2\ \mathrm{d}\mu_0(x),$$

and the proof is concluded. □

1.6.3 The case $p > 1$

The ideas developed by McCann in [35] have been generalized by Ambrosio, Gigli and Savaré in [1]. In this section we will make just a sketch of the results without going into details, trying to highlight the differences introduced in this more general approach.

In the following we will consider a separable Hilbert space X. Recall that $\mathcal{P}_p(X)$ is the space of probability measures on X such that the p-momentum is finite.

Definition 1.6.13. A curve $t \mapsto \mu_t$ from $[0,1]$ to $\mathcal{P}(X)$ is a *constant speed geodesic* if

$$W_{\mu_s,\mu_t} = (t-s)W_{\mu_0,\mu_1}.$$

Let $N \geq 2$, $1 \leq i,j,k \leq N$, $t \in [0,1]$, and $\mu \in \mathcal{P}(X^N)$ be given. We define the maps $\pi_t^{i\to j} : X^N \to X$, $\pi_t^{i\to j,k} : X^N \to X^2$ and the curves $t \to \mu_t^{i\to j} \in \mathcal{P}(X)$, $t \to \mu_t^{i\to j,k} \in \mathcal{P}(X^2)$ as

$$\pi_t^{i\to j} := (1-t)\pi^i + t\pi^j, \tag{1.6.11}$$
$$\pi_t^{i\to j,k} := (1-t)\pi^{i,k} + t\pi^{j,k}, \tag{1.6.12}$$

and

$$\mu_t^{i\to j} := (\pi_t^{i\to j})_{\#}\mu, \tag{1.6.13}$$
$$\mu_t^{i\to j,k} := (\pi_t^{i\to j,k})_{\#}\mu, \tag{1.6.14}$$

where $\pi_i : X^N \to X$ is the projection onto the i-th factor and $\pi_{j,k} : X^N \to X^2$ is the projection onto the j-th and k-th factors.

It will be useful to recover the alternative definition of displacement convexity we already introduced in Definition (1.6.2).

Definition 1.6.14 (λ-convexity along geodesics). Let X be a separable Hilbert space and $F : \mathcal{P}_p(X) \to (-\infty, +\infty]$. Given $\lambda \in \mathbb{R}$, F is said to be λ-geodesically convex in $\mathcal{P}_p(X)$ if for every couple $\mu^1, \mu^2 \in \mathcal{P}_p(X)$ there exists an optimal plan μ between μ^1 and μ^2 such that

$$F(\mu_t^{1\to 2}) \le (1-t)F(\mu^1)+tF(\mu^2)-\frac{\lambda}{2}t(1-t)W_p^2(\mu^1,\mu^2) \quad \forall t \in [0,1], \tag{1.6.15}$$

where $\mu_t^{1\to 2} = (\pi_t^{1\to 2})_{\#}\mu$ (as in (1.6.13)).

We now go on introducing the notion of *generalized geodesic*. We will denote by $\mathcal{P}(\mu^1, \mu^2, \mu^3)$ the set of measures $\mu \in \mathcal{P}(X^3)$ such that $\pi^i_{\#}\mu = \mu^i, i = 1, 2, 3$ (the proof of the existence of such a measure is the so called Gluing Lemma, Lemma 1.5.4).

Definition 1.6.15 (Generalized geodesics). Let X be a separable Hilbert space. A *generalized geodesic* joining $\mu^2, \mu^3 \in \mathcal{P}_p(X)$ with base $\mu_1 \in \mathcal{P}_p(X)$ is a curve whose expression can be given by

$$\mu_t^{2\to 3} := (\pi_t^{2\to 3})_{\#}\mu \qquad t \in [0,1], \tag{1.6.16}$$

where $\mu \in \mathcal{P}(\mu^1, \mu^2, \mu^3)$ and $\pi^{1,2}_{\#}\mu, \pi^{1,3}_{\#}\mu$ are optimal plans between μ^1, μ^3 and μ^1, μ^3 respectively.

Remark 1.6.16. Recall that when μ^1 is absolutely continuous it can be proved that there exists a unique generalized geodesic joining μ^2 and μ^3 with base μ^1 (via Theorem 1.3.13 and the fact that the plan μ satisfying the conditions of $\pi^{1,2}_{\#}\mu, \pi^{1,3}_{\#}\mu$ being optimal is unique). If t_2 and t_3 are the optimal maps between μ^1 and $\mu^i, i = 2, 3$ respectively, then μ is given by the formula $\mu = (\mathrm{Id} \times t_2 \times t_3)_{\#}\mu^1$.

Definition 1.6.17 (λ-convexity along generalized geodesics). Given $\lambda \in \mathbb{R}$, F is said to be λ-convex along generalized geodesics if for any $\mu^1, \mu^2, \mu^3 \in \mathrm{Dom}(F)$ there exists a generalized geodesic $\mu_t^{2\to 3}$ induced by a plan $\mu \in \mathcal{P}(\mu^1, \mu^2, \mu^3)$ such that

$$F(\mu_t^{2\to 3}) \le (1-t)F(\mu^2)+tF(\mu^3)-\frac{\lambda}{2}t(1-t)W_\mu^2(\mu^2,\mu^3) \quad \forall t \in [0,1], \tag{1.6.17}$$

where

$$W_\mu^2(\mu^2,\mu^3) := \int_{X^3} |x_2 - x_3|^2 \, \mathrm{d}\mu(x_1,x_2,x_3) \ge W_2^2(\mu^2,\mu^3).$$

1.6.4 Displacement convex functionals, case $p > 1$

Internal energy

Let $U : [0, +\infty) \to (-\infty, \infty]$ be a proper, lower semicontinuous convex function such that

$$U(0) = 0, \quad \liminf_{s \to 0^+} \frac{U(s)}{s^\alpha} > -\infty \quad \text{for some } \alpha > \frac{N}{N+p}. \tag{1.6.18}$$

The functional we are interested in is, as usual, $\mathcal{U} : \mathcal{P}_p(\mathbb{R}^N) \to (-\infty, +\infty]$ defined by

$$\mathcal{U}(\mu) := \int_{\mathbb{R}^N} U(\rho(x)) \, \mathrm{d}x,$$

if $\mu = \rho \mathcal{L}^N$ is an absolutely continuous with respect to the Lebesgue measure given by the density ρ and by $\mathcal{U}(\mu) := +\infty$ otherwise. Note that equation (1.6.18) implies that U^- is integrable. It can be shown (see [3], [17], [9], [29]) that $\mathcal{U}$ coincides with its lower semicontinuous envelope $\mathcal{U}^*$ on the subset of $\mathcal{P}_p(\mathbb{R}^N)$ of absolutely continuous measures and on the whole $\mathcal{P}_p(\mathbb{R}^N)$ if U has superlinear growth.

Theorem 1.6.18 (Displacement convexity of internal energy). *If the map*

$$r \mapsto \psi(r) := r^N U(r^{-N})$$

is convex non-increasing on $]0, +\infty[$*, then* $\mathcal{U}$ *is convex along geodesics in* $\mathcal{P}_p(\mathbb{R}^N)$ *(Definition 1.6.14) and along generalized geodesics when* $p = 2$ *(Definition 1.6.17).*

Sketch of the proof. We will skip the proof of generalized geodesic convexity and we will prove only the convexity along geodesics; all details can be found in [1]. First, it is sufficient to check the geodesic convexity of $\mathcal{U}$. So, let us suppose that $\mu^1 = \rho^1 \mathcal{L}^N, \mu^2 = \rho^2 \mathcal{L}^N \in \mathrm{Dom}(\mathcal{U})$ and that r is the optimal transport map between μ^1, μ^2 for the p-Kantorovich functional (such a map exists thanks to Theorem 1.3.13). It can be shown that r is approximately differentiable μ^1-a.e., $\tilde{\nabla} r$ is diagonalizable and its eigenvalues are non-negative. Since μ^2 is absolutely continuous, it follows that $\det \tilde{\nabla} r(x) > 0$ for μ^1-a.e. $x \in \mathbb{R}^N$; as a consequence $r_t = (1-t)\,\mathrm{Id} + tr$ is diagonalizable with positive eigenvalues. The measure $\mu_t^{1 \to 2} = (r_t)_\# \mu^1$ is absolutely continuous and its density ρ_t is given by

$$\rho_t(r_t(x)) = \frac{\rho^1(x)}{\det \tilde{\nabla} r_t(x)}$$

for μ^1-a.e. $x \in \mathbb{R}^N$. Then, it follows that

$$\mathcal{U}(\mu_t^{1\to 2}) = \int_{\mathbb{R}^N} U(\rho_t(y))\ \mathrm{d}y = \int_{\mathbb{R}^N} U\left(\frac{\rho^1(x)}{\det \tilde{\nabla} r_t(x)}\right) \det \tilde{\nabla} r_t(x)\ \mathrm{d}x. \quad (1.6.19)$$

Since it can be viewed as the composition of the convex non-increasing map $s \mapsto s^N F(\rho_1(x)/s^N)$ and of the concave map $t \mapsto \det((1-t)\,\mathrm{Id} + t\tilde{\nabla} r(x))$, the integrand of equation (1.6.19) is convex in t and then the integral is convex. $\square$

Potential energy

Let $V : X \to (-\infty, +\infty]$ be a proper, lower semicontinuous function such that for some constants $A, B \in \mathbb{R}$

$$V(x) \geq -A - B|x|^p \quad \forall x \in X. \quad (1.6.20)$$

The functional we are interested in is then

$$\mathcal{V}(\mu) := \int_X V(x)\ \mathrm{d}\mu(x).$$

The functional $\mathcal{V}$ is finite on Dirac masses, so it is proper. Moreover, thanks to inequality (1.6.20) we gain the lower semicontinuity.

Theorem 1.6.19 (Displacement convexity of potential energy). *If V is λ-convex, i.e. for every $x_1, x_2 \in X$*

$$V((1-t)x_1 + tx_2) \leq (1-t)V(x_1) + tV(x_2) - \frac{\lambda}{2} t(1-t)|x_1 - x_2|^2, \quad (1.6.21)$$

then for every $\mu^1, \mu^2 \in \mathrm{Dom}(\mathcal{V})$ and $\mu \in \mathcal{P}(\mu^1, \mu^2)$ we have

$$V(\mu_t^{1\to 2}) \leq (1-t)V(\mu^1) + tV(\mu^2) - \frac{\lambda}{2} t(1-t) \int_{X^2} |x_1 - x_2|^2\ \mathrm{d}\mu(x_1, x_2). \quad (1.6.22)$$

As a consequence, if $p = 2$ the functional $\mathcal{V}$ is λ-convex on generalized geodesics (Definition 1.6.17). Finally, if $p \leq 2, \lambda \geq 0$ or $p \geq 2, \lambda \leq 0$, then $\mathcal{V}$ is λ-geodesically convex in $\mathcal{P}_p(X)$ (Definition 1.6.14).

We now go on with the proof which is quite similar to that of Theorem 1.6.10.

Proof. Thanks to inequality (1.6.20) the function $\mathcal{V}$ is well-defined on $\mathcal{P}_p(X)$. Fix a transport plan $\mu \in \mathcal{P}(\mu^1, \mu^2)$ with $\mu^1, \mu^2 \in \mathrm{Dom}(\mathcal{V})$. We then have:

$$\begin{aligned}\mathcal{V}(\mu_t^{1\to 2}) &= \int_{X^2} V((1-t)x_1 + tx_2)\,\mathrm{d}\mu(x_1, x_2)\\ &\le \int_{X^2} (1-t)V(x_1) + tV(x_2) - \frac{\lambda}{2}t(1-t)|x_1 - x_2|^2\,\mathrm{d}\mu(x_1, x_2)\\ &= (1-t)\mathcal{V}(\mu^1) + t\mathcal{V}(\mu^2) - \frac{\lambda}{2}t(1-t)\int_{X^2} |x_1 - x_2|^2\,\mathrm{d}\mu(x_1, x_2),\end{aligned}$$

which is the convexity along generalized geodesics, when $p = 2$. If $p \neq 2$, choose an optimal $\mu \in \mathcal{P}(\mu^1, \mu^2)$: if $p > 2$ we use the fact that the inequality

$$\int_{X^2} |x_1 - x_2|^2\,\mathrm{d}\mu(x_1, x_2) \le W_p^2(\mu^1, \mu^2)$$

is true, while if $p < 2$ we use the opposite one. □

Interaction energy

Fix an integer $k > 1$ and a lower semicontinuous function $W : X^k \to (-\infty, +\infty]$ such that W^- satisfies

$$W(x) \ge -A - B|x|^p \quad \forall x \in X, \tag{1.6.23}$$

for some $A, B \in \mathbb{R}$ and all $x \in \mathbb{R}^N$. Consider the functional

$$\mathcal{W}_k(\mu) := \int_{X^k} W(x_1, x_2, \dots, x_k)\,\mathrm{d}\mu^{\otimes k}(x_1, x_2, \dots, x_k).$$

Theorem 1.6.20 (Displacement convexity of interaction energy). *If W is convex, then the functional $\mathcal{W}_k$ is convex along any interpolating curve $\mu_t^{1\to 2} = (\pi_t^{1\to 2})_\#\mu$ with μ transport plan (not necessarily optimal) between μ^1 and μ^2.*

Proof. Recall that if $f_i : X_i \to Y_i$ are measurable maps and μ_i are measures on X_i for $1 \le i \le k$, then

$$\otimes_{i=1}^k (f_{i\#}\mu_i) = (\otimes_{i=1}^k \mu_i)_\#(\otimes_{i=1}^k f_i),$$

where $(\otimes_i f_i)(x_1, \dots, x_k) = (f_1(x_1), \dots, f_k(x_k))$. Let us now consider a transport plan μ between μ^1 and μ^2 and the curve $\mu_t^{1\to 2} = (\pi_t^{1\to 2})_\#\mu$. Then,

$$\begin{aligned}&\mathcal{W}(\pi_t^{1\to 2})_\#\mu) = \int_{X^k} W\,\mathrm{d}((\pi_t^{1\to 2})_\#\mu)^{\otimes k}\\ &= \int_{(X\times X)^k} W((1-t)x_1 + ty_1, \dots, (1-t)x_k + ty_k)\,\mathrm{d}\mu^{\otimes k}(x_1, y_1, \dots, x_k, y_k).\end{aligned}$$

Now it is clear that the convexity properties of W will reflect on $\mathcal{W}$. $\square$

Chapter 2

Optimal Networks for Mass Transportation Problems

In this chapter we study the network problem proposed in [18] and investigated from a qualitative point of view in [20]. Here a more general cost functional is considered. The results of this chapter can also be found in [13].

2.1 The Optimal Network Problem

We consider a bounded connected open subset Ω with Lipschitz boundary of $\mathbb{R}^N$ (the urban area) with $N > 1$ and two positive finite measures μ^+ and μ^- on $K := \overline{\Omega}$ (the distributions of working people and of working places). We assume that μ^+ and μ^- have the same mass that we normalize both equal 1, that is μ^+ and μ^- are probability measures on K.

The optimization problem for transportation networks is the following: to every "urban network" Σ we associate a suitable "cost function" d_Σ which takes into account the geometry of Σ as well as the costs for customers to move with their own means and by means of the network. The cost functional will be then

$$T(\Sigma) = W_{d_\Sigma}(\mu^+, \mu^-),$$

where $W_{d_\Sigma}(\mu^+, \mu^-)$ is the Wasserstein distance $W_1(\mu^+, \mu^-)$ with respect to the pseudo-distance d_Σ, so that the optimization problem we deal with is

$$\min\{T(\Sigma) \ : \ \Sigma \text{ "admissible network"}\}. \tag{2.1.1}$$

The main result is the proof that, under suitable and very mild assumptions, and taking as admissible networks all connected, compact one-dimensional subsets Σ of K, the optimization problem (2.1.1) admits a solution. The tools we use to obtain the existence result are a suitable relaxation procedure to define the function d_Σ (Theorem 2.3.2) and a generalization of the Gołab Theorem (Theorem 2.2.3), also obtained by Dal Maso and Toader in [21].

In order to introduce the distance d_Σ on the set $\overline{\Omega} \times \overline{\Omega}$ we consider a function $J : [0, +\infty]^3 \to [0, +\infty]$. For a given path γ in K the parameter a in $J(a, b, c)$ measures the length of γ outside Σ, b measures the length of γ inside Σ, while c represents the total length of Σ. The cost $J(a, b, c)$ is then the cost of a customer who travels for a length a by his own means and for a length b on the network, being c the length of the latter. For instance we could take $J(a, b, c) = A(a) + B(b) + C(c)$ and then the function $A(t)$ is the cost for travelling a length t by one's own means, $B(t)$ is the price of a ticket to cover the length t on Σ and $C(t)$ represents the cost of a network of length t.

For every closed connected subset Σ in K, we define the cost function d_Σ as

$$d_\Sigma(x, y) := \inf \left\{ J\left(\mathcal{H}^1(\gamma \setminus \Sigma), \mathcal{H}^1(\gamma \cap \Sigma), \mathcal{H}^1(\Sigma)\right) \ : \ \gamma \in \mathscr{C}_{x,y} \right\},$$

where $\mathscr{C}_{x,y}$ is the class of all closed connected subsets of K containing x and y. The optimization problem we consider is then (2.1.1) where we take as *admissible networks* all closed connected subsets Σ of K with $\mathcal{H}^1(\Sigma) < +\infty$. We also define, for every closed connected subset γ of K

$$L_\Sigma(\gamma) := J\left(\mathcal{H}^1(\gamma \setminus \Sigma), \mathcal{H}^1(\gamma \cap \Sigma), \mathcal{H}^1(\Sigma)\right).$$

We assume that J satisfies the following conditions:

- J is lower semicontinuous,
- J is non-decreasing, i.e.

$$a_1 < a_2, b_1 < b_2, c_1 < c_2 \Longrightarrow J(a_1, b_1, c_1) \leq J(a_2, b_2, c_2),$$

- $J(a, b, c) \geq G(c)$ with $G(c) \to +\infty$ when $c \to +\infty$,
- J is continuous in its first variable.

A *curve* joining two points $x, y \in K$ is an element of the set

$$\mathscr{C}_{x,y} := \{\gamma \text{ closed connected}, \{x,y\} \subseteq \gamma \subseteq K\}$$

while an element of $\mathscr{C}$ will be, by definition, a closed connected set in K:

$$\mathscr{C} := \{\gamma \text{ closed connected}, \gamma \subseteq K\}.$$

We associate to every admissible network $\Sigma \in \mathscr{C}$ the cost function

$$d_\Sigma(x,y) = \inf\{L_\Sigma(\gamma) \ : \ \gamma \in \mathscr{C}_{x,y}\}.$$

We are interested in the functional T given by

$$\Sigma \mapsto T(\Sigma) := W_{d_\Sigma}(\mu^+, \mu^-)$$

which is defined on the class $\mathscr{C}$.

Finally by $\overline{L}_\Sigma^{x,y}$ we denote the lower semicontinuous envelope of L_Σ with respect to the Hausdorff convergence on $\mathscr{C}_{x,y}$ (see Section 2.2 for the main definitions). In other words, for every $\gamma \in \mathscr{C}_{x,y}$ we set

$$\overline{L}_\Sigma^{x,y}(\gamma) = \begin{cases} \min\{\liminf_n L_\Sigma(\gamma_n) \ : \gamma_n \to \gamma, \gamma_n \in \mathscr{C}_{x,y}\} & \text{if } \gamma \in \mathscr{C}_{x,y} \\ +\infty & \text{if } \gamma \notin \mathscr{C}_{x,y}, \end{cases}$$

where we fix the condition $x, y \in \gamma$. Moreover, we define $\overline{L}_\Sigma$ as

$$\overline{L}_\Sigma(\gamma) = \min\left\{\liminf_{n\to+\infty} L_\Sigma(\gamma_n) \ : \gamma_n \to \gamma, \gamma_n \in \mathscr{C}\right\},$$

that is to say, the lower semicontinuous envelope of L_Σ with respect to the Hausdorff convergence on the class of closed connected sets of K.

2.2 The Gołab Theorem and its extensions

In this section X will be a set endowed with a distance function d, i.e. (X,d) is a metric space. We assume for simplicity X to be *compact*. By $\mathscr{C}(X)$ we indicate the class of all closed subsets of X.

Given two closed subsets C and D, the *Hausdorff distance* between them is defined by

$$d_{\mathcal{H}}(C,D) := 1 \wedge \inf\{r \in [0,+\infty[\ : \ C \subseteq D_r, D \subseteq C_r\}$$

where

$$C_r := \{x \in X \ : \ d(x, C) < r\}.$$

It is easy to see that $d_{\mathcal{H}}$ is a distance on $\mathscr{C}(X)$, so $(\mathscr{C}(X), d_{\mathcal{H}})$ is a metric space. We remark the following well-known facts (see for example [5]):

- (X, d) compact $\Longrightarrow (\mathscr{C}(X), d_{\mathcal{H}})$ compact,
- (X, d) complete $\Longrightarrow (\mathscr{C}(X), d_{\mathcal{H}})$ complete.

In the rest of the chapter we will use the notation $C_n \to C$ to indicate the convergence of a sequence $\{C_n\}_{n\in\mathbb{N}}$ to C with respect to the distance $d_{\mathcal{H}}$.

Proposition 2.2.1. *Let $\{C_n\}_{n\in\mathbb{N}}$ be a sequence of compact connected subsets in X such that $C_n \to C$ for some compact subset C. Then C is connected.*

Proof. Suppose, on the contrary, that there exist two closed non-void separated subsets F_1 and F_2 such that $C = F_1 \cup F_2$. Since F_1 and F_2 are compact, $d(F_1, F_2) = d > 0$. Let us choose $\varepsilon = d/4$. By the definition of Hausdorff convergence, there exists a positive integer N such that

$$n \geq N \Longrightarrow C_n \subseteq (C)_\varepsilon, \ C \subseteq (C_n)_\varepsilon.$$

Since C_N is connected, we must have either $C_N \subseteq (F_1)_\varepsilon$ or $C_N \subseteq (F_2)_\varepsilon$. Let us suppose, for example, that $C_N \subseteq (F_1)_\varepsilon$. On one side by the Hausdorff convergence it is $F_2 \subseteq (C_N)_\varepsilon$, on the other by the choice of ε we have $(C_N)_\varepsilon \cap F_2 = \varnothing$, a contradiction. $\square$

The *Hausdorff 1-dimensional measure* in (X, d) of a Borel set B is defined by

$$\mathcal{H}^1(B) := \lim_{\delta \to 0^+} \mathcal{H}^{1,\delta}(B),$$

where

$$\mathcal{H}^{1,\delta}(B) := \inf \left\{ \sum_{n\in\mathbb{N}} \operatorname{diam} B_n : \operatorname{diam} B_n < \delta, B \subseteq \bigcup_{n\in\mathbb{N}} B_i \right\}.$$

The measure $\mathcal{H}^1$ is Borel regular and if (X, d) is the 1-dimensional Euclidean space, then $\mathcal{H}^1$ is just the Lebesgue measure $\mathcal{L}^1$.

The Gołab classical Theorem states that in a metric space, the measure $\mathcal{H}^1$ is sequentially lower semicontinuous with respect to the Hausdorff convergence over the class of all compact connected subsets of X.

Theorem 2.2.2 (Gołab). *Let X be a metric space. If $\{C_n\}_{n\in\mathbb{N}}$ is a sequence of compact connected subsets of X and $C_n \to C$ for some compact connected subset C, then*

$$\mathcal{H}^1(C) \le \liminf_{n\to+\infty} \mathcal{H}^1(C_n). \tag{2.2.1}$$

Actually, this result can be strengthened.

Theorem 2.2.3. *Let X be a metric space, $\{\Gamma_n\}_{n\in\mathbb{N}}$ and $\{\Sigma_n\}_{n\in\mathbb{N}}$ be two sequences of compact subsets such that $\Gamma_n \to \Gamma$ and $\Sigma_n \to \Sigma$ for some compact subsets Γ and Σ. Let us also suppose that Γ_n is connected for all $n \in \mathbb{N}$. Then*

$$\mathcal{H}^1(\Gamma \setminus \Sigma) \le \liminf_{n\to+\infty} \mathcal{H}^1(\Gamma_n \setminus \Sigma_n). \tag{2.2.2}$$

A proof of this result has been given by Dal Maso and Toader in [21]; for sake of completeness, we include the proof here below. It is in fact based on the following two rectifiability theorems whose proof can be found in [5].

Theorem 2.2.4. *Let X be a metric space and C a closed connected subset of finite length, i.e. $\mathcal{H}^1(C) < +\infty$. Then C is compact and connected by injective rectifiable curves.*

Theorem 2.2.5. *Let C be a closed connected subset in a metric space X such that $\mathcal{H}^1(C) < +\infty$. Then there exists a sequence of Lipschitz curves $\{\gamma_n\}_{n\in\mathbb{N}}$, $\gamma_n : [0,1] \to C$, such that*

$$\mathcal{H}^1(C \setminus \bigcup_{n\in\mathbb{N}} \gamma_n([0,1])) = 0.$$

The first step in the proof of Theorem 2.2.3 is a localized form of the classical Gołab Theorem. To this aim we need the following lemma.

Lemma 2.2.6. *Let C be a closed connected subset of X and let $x \in C$. If $r \in [0, \frac{1}{2}\operatorname{diam} C]$, then*

$$\mathcal{H}^1(C \cap B_r(x)) \ge r.$$

Proof. See for instance Lemma 4.4.2 of [5] or Lemma 3.4 of [26]. □

Remark 2.2.7. Lemma 2.2.6 yields the following estimate from below for the upper density:

$$\overline{\theta}(C,x) := \limsup_{r\to 0^+} \frac{\mathcal{H}^1(C \cap B_r(x))}{2r} \ge \frac{1}{2}.$$

We recall that for every measure μ the *upper density* is defined by

$$\overline{\theta}(\mu, x) := \limsup_{r \to 0^+} \frac{\mu(B_r(x))}{2r}.$$

We also recall that $\overline{\theta}(\mu, x) \geq t$ for all $x \in X$ implies $\mu(B) \geq t\mathcal{H}^1(B)$ for every Borel set B (see Theorem 2.4.1 in [5]).

We are now in a position to obtain the localized version of the Gołab Theorem.

Theorem 2.2.8. *Let X be a metric space. If $\{C_n\}_{n \in \mathbb{N}}$ is a sequence of compact connected subsets of X such that $C_n \to C$ for some compact connected subset C, then for every open subset U of X*

$$\mathcal{H}^1(C \cap U) \leq \liminf_{n \to +\infty} \mathcal{H}^1(C_n \cap U).$$

Proof. We can suppose that $L := \lim_n \mathcal{H}^1(C_n \cap U)$ exists, is finite and $\mathcal{H}^1(C_n \cap U) \leq L + 1$. Let $d_n = \operatorname{diam}(C_n \cap U)$. We can suppose up to a subsequence that $d_n \to d > 0$. Let us consider the sequence of Borel measures defined by

$$\mu_n(B) := \mathcal{H}^1(B \cap C_n \cap U)$$

for every Borel set B. Up to a subsequence we can assume that $\mu_n \rightharpoonup^* \mu$ for a suitable μ. We choose $x \in C \cap U$ and $r' < r < \operatorname{diam}(C \cap U)/2$. Then, by Lemma 2.2.6,

$$\begin{aligned} \mu(B_r(x)) \geq \mu(\overline{B}_{r'}(x)) &\geq \limsup_{n \to +\infty} \mu_n(\overline{B}_{r'}(x)) \\ &= \limsup_{n \to +\infty} \mathcal{H}^1(C_n \cap \overline{B}_{r'}(x) \cap U) \geq r'. \end{aligned}$$

Since r' was chosen arbitrarily we get

$$\mu(B_r(x)) \geq r$$

for every $x \in C \cap U$ and $r < \operatorname{diam}(C \cap U)/2$. This implies $\overline{\theta}(C, x) \geq 1/2$. By Remark 2.2.7

$$\mathcal{H}^1(C \cap U) \leq 2\mu(X) \leq 2 \liminf_{n \to +\infty} \mu_n(X) = 2 \liminf_{n \to +\infty} \mathcal{H}^1(C_n \cap U) = 2L.$$

By Theorem 2.2.5 for $\mathcal{H}^1$-almost all $x_0 \in C \cap U$ there exists a Lipschitz curve γ whose range is in $C \cap U$ such that $x_0 = \gamma(t_0)$ and $t_0 \in]0, 1[$. We can also suppose that

$$\lim_{h \to 0^+} \frac{d(\gamma(t_0 + h), \gamma(t_0 - h))}{2|h|} = 1.$$

We choose arbitrarily $\sigma \in]0,1[$. If h is small, then

$$d(\gamma(t_0+h),\gamma(t_0-h)) \geq (2-\sigma)|h|$$

and

$$(1-\sigma)|h| \leq d(\gamma(t_0 \pm h),\gamma(t_0)) \leq (1+\sigma)|h|.$$

Let us also suppose that $|h| < \sigma/(1+\sigma)$ and put

$$y := \gamma(t_0-h), \quad z := \gamma(t_0+h), \quad r := \max\{d(y,x_0), d(z,x_0)\}.$$

We get

$$r < (1+\sigma)|h| < \sigma, \quad d(y,z) \geq (2-\sigma)|h| \geq \frac{2-\sigma}{2+\sigma} r.$$

Let $r' := (1+\sigma)r$. Since $C_n \to C$, then (see Proposition 4.4.3 in [5]) there exist subsequences $\{y_n\}_{n\in\mathbb{N}}$ and $\{z_n\}_{n\in\mathbb{N}}$ such that $y_n, z_n \in C_n \cap U$, $y_n \to y$ and $z_n \to z$. One must have $y_n, z_n \in B_{r'}(x_0)$ for n large enough and

$$\mu_n(\overline{B_{r'}(x)}) = \mathcal{H}^1(C_n \cap \overline{B_{r'}(x)} \cap U) \geq d(z,y_n).$$

Taking the limsup

$$\begin{aligned} \mu(\overline{B_{r'}(x)}) &\geq \limsup_{n\to+\infty} \mathcal{H}^1(C_n \cap \overline{B_{r'}(x)} \cap U) \geq \limsup_{n\to+\infty} d(z,y_n) \\ &= d(z,y) \geq \frac{2-\sigma}{2+\sigma} r = \frac{2-\sigma}{(2+\sigma)(1+\sigma)} r'. \end{aligned}$$

Since σ was arbitrary, we get $\overline{\theta}(\mu,x_0) \geq 1$ for $\mathcal{H}^1$-almost all $x_0 \in C \cap U$. Then, by Remark 2.2.7

$$\mathcal{H}^1(C\cap U) \leq \mu(X) \leq \liminf_{n\to+\infty} \mu_n(X) = \liminf_{n\to+\infty} \mathcal{H}^1(C_n \cap U). \quad \square$$

Proof of Theorem 2.2.3. Let $A = \Gamma \cap \Sigma$. Thanks to the equality

$$\bigcup_{\varepsilon>0} (\Gamma \setminus \overline{A}_\varepsilon) = \Gamma \setminus \Sigma$$

we have

$$\lim_{\varepsilon\to 0^+} \mathcal{H}^1(\Gamma \setminus \overline{A}_\varepsilon) = \mathcal{H}^1(\Gamma \setminus \Sigma).$$

Recalling that the following inclusion of sets holds for large values of n

$$\Gamma_n \setminus \overline{A}_\varepsilon \subseteq \Gamma_n \setminus A_n \subseteq \Gamma_n \setminus \Sigma_n$$

by the localized form of Gołab Theorem (Theorem 2.2.8) we deduce

$$\mathcal{H}^1(\Gamma \setminus \overline{A}_\varepsilon) \leq \liminf_{n\to+\infty} \mathcal{H}^1(\Gamma_n \setminus \overline{A}_\varepsilon) \leq \liminf_{n\to+\infty} \mathcal{H}^1(\Gamma_n \setminus \Sigma_n).$$

Taking the limit as $\varepsilon \to 0^+$, we obtain

$$\mathcal{H}^1(\Gamma \setminus \Sigma) \leq \liminf_{n\to+\infty} \mathcal{H}^1(\Gamma_n \setminus \Sigma_n). \quad \square$$

Remark 2.2.9. It is easy to see that if the number of connected components of C_n is bounded from above by a positive integer independent on n, then the localized form of Gołab Theorem is still valid. All details can be found in [21].

2.3 Relaxation of the cost function

We can give an explicit expression for the lower semicontinuous envelopes $\overline{L}_\Sigma$ and $\overline{L}_\Sigma^{x,y}$ in terms of J. In order to achieve this result it is useful to introduce the function:

$$\overline{J}(a,b,c) = \inf\{J(a+t, b-t, c) \ : \ 0 \leq t \leq b\}.$$

The following lemma is an important step to establish Theorem 2.3.2.

Lemma 2.3.1. *Let γ and Σ be closed connected subsets of K. Let also suppose that Σ has a finite length. Then for every $t \in [0, \mathcal{H}^1(\gamma \cap \Sigma)]$ we can find a sequence $\{\gamma_n\}_{n\in\mathbb{N}}$ in $\mathscr{C}$ such that*

- $\gamma_n \to \gamma$,
- $\lim_n \mathcal{H}^1(\gamma_n) = \mathcal{H}^1(\gamma)$,
- $\mathcal{H}^1(\gamma_n \cap \Sigma) \nearrow \mathcal{H}^1(\gamma \cap \Sigma) - t$.

Moreover, if $x, y \in \gamma$ then the sequence $\{\gamma_n\}_{n\in\mathbb{N}}$ can be chosen in $\mathscr{C}_{x,y}$.

Proof. The set $\gamma \cap \Sigma$ is closed and with a finite length. By the second rectifiability result (Theorem 2.2.5) it follows the existence of a sequence of curves $\sigma_n \in \mathrm{Lip}([0,1], K)$ such that

$$\mathcal{H}^1\left((\gamma \cap \Sigma) \setminus \bigcup_{n\in\mathbb{N}} \sigma_n([0,1])\right) = 0.$$

We can also suppose that the subsets $\sigma_n([0,1])$ are disjoint up to subsets of negligible length. Fix a sufficiently small $\delta > 0$ and choose a sequence of intervals $I_n = [a_n, b_n]$ such that

$$\sum_{n\in\mathbb{N}} \mathcal{H}^1(\sigma_n(I_n)) = t + \delta.$$

For every sequence $\underline{v} = \{v_n\}_{n\in\mathbb{N}}$ of unit vectors of $\mathbb{R}^N$ such that v_n is not tangent to $\gamma \cap \Sigma$ in $\sigma_n(a_n)$ and $\sigma_n(b_n)$, and every sequence $\underline{\varepsilon} = \{\varepsilon_n\}_{n\in\mathbb{N}}$ of positive real numbers, let us consider

$$\begin{aligned}
A_{\underline{v},\underline{\varepsilon}} &= \bigcup_{n\in\mathbb{N}} \sigma_n([0,a_n] \cup [b_n,1]),\\
B_{\underline{v},\underline{\varepsilon}} &= \bigcup_{n\in\mathbb{N}} (\sigma_n(a_n) + \varepsilon_n V_n),\\
C_{\underline{v},\underline{\varepsilon}} &= \bigcup_{n\in\mathbb{N}} (v_n + \sigma_n(I_n)),\\
D_{\underline{v},\underline{\varepsilon}} &= \bigcup_{n\in\mathbb{N}} (\sigma_n(b_n) + \varepsilon_n V_n)\\
\gamma_{\underline{v},\underline{\varepsilon}} &= (\gamma \setminus \Sigma) \cup A_{\underline{v},\underline{\varepsilon}} \cup B_{\underline{v},\underline{\varepsilon}} \cup C_{\underline{v},\underline{\varepsilon}} \cup D_{\underline{v},\underline{\varepsilon}}
\end{aligned}$$

where $V_n = \{tv_n \ : \ t \in [0,1]\}$ (see Figure 2.1).

Since Σ is closed and with a finite length, the class of $\gamma_{\underline{v},\underline{\varepsilon}}$ that have not $\mathcal{H}^1$-negligible intersection with Σ is at most countable. Out of that set we can choose sequences $\delta_m \searrow 0$, and $\{\gamma_{\underline{v}_m,\underline{\varepsilon}_m}\}_{m\in\mathbb{N}}$ such that $\|\underline{\varepsilon}_m\| \searrow 0$, where by $\|\underline{\varepsilon}\|$ we denote the quantity $\sum_n \varepsilon_n$. The sequence $\{\gamma_{\underline{v}_m,\underline{\varepsilon}_m}\}_{m\in\mathbb{N}}$ is the one we were looking for. $\square$

Theorem 2.3.2. *For every closed connected subset $\gamma \in \mathscr{C}_{x,y}$ we have*

$$\overline{L}^{x,y}_{\Sigma}(\gamma) = \overline{J}(\mathcal{H}^1(\gamma \setminus \Sigma), \mathcal{H}^1(\gamma \cap \Sigma), \mathcal{H}^1(\Sigma)).$$

Moreover, if $\gamma \in \mathscr{C}_{x,y}$ then

$$\overline{L}^{x,y}_{\Sigma}(\gamma) = \overline{L}_{\Sigma}(\gamma).$$

Proof. Let γ be a fixed curve in $\mathscr{C}_{x,y}$. First we establish that

$$\overline{L}^{x,y}_{\Sigma}(\gamma) \geq \overline{J}(\mathcal{H}^1(\gamma \setminus \Sigma), \mathcal{H}^1(\gamma \cap \Sigma), \mathcal{H}^1(\Sigma)).$$

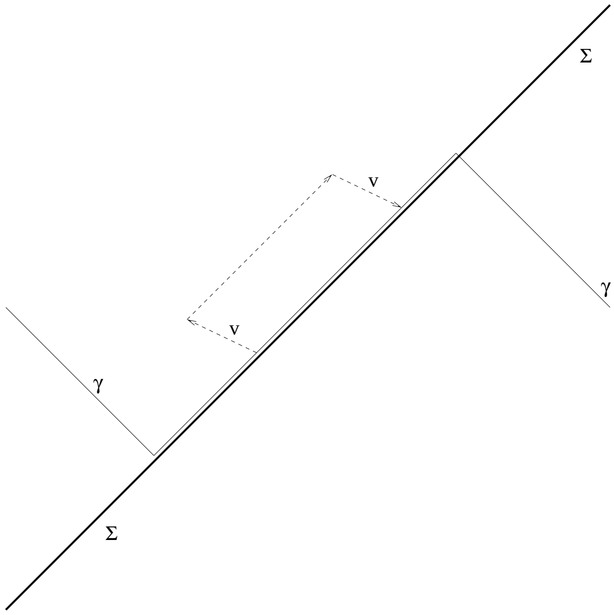

Figure 2.1: The approximating curves γ_n.

It is enough to show that for every sequence $\{\gamma_n\}_{n\in\mathbb{N}}$ in $\mathscr{C}_{x,y}$ converging to γ with respect to the Hausdorff metric, there exists $t \in [0, \mathcal{H}^1(\gamma \cap \Sigma)]$ such that

$$J(\mathcal{H}^1(\gamma \setminus \Sigma) + t, \mathcal{H}^1(\gamma \cap \Sigma) - t, \mathcal{H}^1(\Sigma)) \le \liminf_{n\to+\infty} L_\Sigma(\gamma_n).$$

Up to a subsequence we can suppose the following equalities hold true:

$$\begin{aligned}\liminf_{n\to+\infty} L_\Sigma(\gamma_n) &= \lim_{n\to+\infty} L_\Sigma(\gamma_n),\\ \liminf_{n\to+\infty} \mathcal{H}^1(\gamma_n) &= \lim_{n\to+\infty} \mathcal{H}^1(\gamma_n),\\ \liminf_{n\to+\infty} \mathcal{H}^1(\gamma_n \setminus \Sigma) &= \lim_{n\to+\infty} \mathcal{H}^1(\gamma_n \setminus \Sigma).\end{aligned}$$

Moreover, by Gołab Theorems (Theorem 2.2.2 and Theorem 2.2.3)

$$\begin{aligned}\mathcal{H}^1(\gamma) &\le \lim_{n\to+\infty} \mathcal{H}^1(\gamma_n),\\ \mathcal{H}^1(\gamma \setminus \Sigma) &\le \lim_{n\to+\infty} \mathcal{H}^1(\gamma_n \setminus \Sigma).\end{aligned}$$

Choose $t = \lim_n \mathcal{H}^1(\gamma_n \setminus \Sigma) - \mathcal{H}^1(\gamma \setminus \Sigma)$. Then $\mathcal{H}^1(\gamma \setminus \Sigma) + t = \lim_n \mathcal{H}^1(\gamma_n \setminus \Sigma)$. We have

$$\begin{aligned}\mathcal{H}^1(\gamma_n) &= \mathcal{H}^1(\gamma_n \setminus \Sigma) + \mathcal{H}^1(\gamma_n \cap \Sigma)\\ &= [\mathcal{H}^1(\gamma_n \setminus \Sigma) - t] + [\mathcal{H}^1(\gamma_n \cap \Sigma) + t].\end{aligned}$$

Taking the limit as $n \to +\infty$ gives

$$\mathcal{H}^1(\gamma) \leq \lim_{n\to+\infty} \mathcal{H}^1(\gamma_n) = [\mathcal{H}^1(\gamma \setminus \Sigma) + t] + \lim_{n\to+\infty} \mathcal{H}^1(\gamma_n \cap \Sigma)$$

so that

$$\mathcal{H}^1(\gamma \cap \Sigma) - t \leq \lim_{n\to+\infty} \mathcal{H}^1(\gamma_n \cap \Sigma).$$

It follows by the semicontinuity and monotonicity of J in the first two variables

$$J\left(\mathcal{H}^1(\gamma \setminus \Sigma) + t, \mathcal{H}^1(\gamma \cap \Sigma) - t, \mathcal{H}^1(\Sigma)\right) \leq \liminf_{n\to+\infty} J\left(\mathcal{H}^1(\gamma_n \setminus \Sigma), \mathcal{H}^1(\gamma_n \cap \Sigma), \mathcal{H}^1(\Sigma)\right).$$

Now, we have to establish the opposite inequality:

$$\overline{L}^{x,y}_\Sigma(\gamma) \leq \overline{J}\left(\mathcal{H}^1(\gamma \setminus \Sigma), \mathcal{H}^1(\gamma \cap \Sigma), \mathcal{H}^1(\Sigma)\right).$$

In the same way as before, it is enough to show that for every $t \in [0, \mathcal{H}^1(\gamma\cap\Sigma)]$ we can find a sequence $\{\gamma_n\}_{n\in\mathbb{N}}$ in $\mathscr{C}_{x,y}$ which converges to γ such that

$$\liminf_{n\to+\infty} L_\Sigma(\gamma_n) \leq J\left(\mathcal{H}^1(\gamma \setminus \Sigma) + t, \mathcal{H}^1(\gamma \cap \Sigma) - t, \mathcal{H}^1(\Sigma)\right).$$

Given t, let $\{\gamma_n\}_{n\in\mathbb{N}}$ be the sequence given by Lemma 2.3.1. Then we get

$$\lim_{n\to+\infty} \mathcal{H}^1(\gamma_n \setminus \Sigma) = \mathcal{H}^1(\gamma) - \mathcal{H}^1(\gamma \cap \Sigma) + t = \mathcal{H}^1(\gamma \setminus \Sigma) + t.$$

Thanks to $\mathcal{H}^1(\gamma_n \cap \Sigma) \leq \mathcal{H}^1(\gamma \cap \Sigma) - t$, we have

$$J\left(\mathcal{H}^1(\gamma_n \setminus \Sigma), \mathcal{H}^1(\gamma_n \cap \Sigma), \mathcal{H}^1(\Sigma)\right) \leq J\left(\mathcal{H}^1(\gamma_n \setminus \Sigma), \mathcal{H}^1(\gamma \cap \Sigma), \mathcal{H}^1(\Sigma)\right)$$

and by the continuity of J in the first variable

$$\liminf_{n\to+\infty} J\left(\mathcal{H}^1(\gamma_n \setminus \Sigma), \mathcal{H}^1(\gamma_n \cap \Sigma), \mathcal{H}^1(\Sigma)\right) \leq J\left(\mathcal{H}^1(\gamma \setminus \Sigma) + t, \mathcal{H}^1(\gamma \cap \Sigma) - t, \mathcal{H}^1(\Sigma)\right)$$

which implies the inequality we looked for. The proof of the second statement of the Theorem is analogous and hence omitted. $\square$

The next proposition is a consequence of Theorem 2.3.2.

Proposition 2.3.3. *For every $x, y \in K$ we have*

$$d_\Sigma(x,y) = \inf\{\overline{L}_\Sigma(\gamma) \ : \ \gamma \in \mathscr{C}_{x,y}\}.$$

Proof. By a general result of relaxation theory (see for instance [17]), the infimum of a function is the same as the infimum of its lower semicontinuous envelope, so

$$d_\Sigma(x,y) = \inf\{\overline{L}_\Sigma^{x,y}(\gamma) \ : \ \gamma \in \mathscr{C}_{x,y}\}.$$

It is then enough to prove that

$$\inf\{\overline{L}_\Sigma^{x,y}(\gamma) \ : \ \gamma \in \mathscr{C}_{x,y}\} = \inf\{\overline{L}_\Sigma(\gamma) \ : \ \gamma \in \mathscr{C}_{x,y}\},$$

which is a consequence of Theorem 2.3.2. □

It is more convenient to introduce the function whose variables a, b, c now represent the length $\mathcal{H}^1(\gamma \setminus \Sigma)$ covered by one's own means, the path length $\mathcal{H}^1(\gamma)$, and the length of the network $\mathcal{H}^1(\Sigma)$:

$$\Theta(a,b,c) = \overline{J}(a, b-a, c).$$

Obviously, Θ satisfies

$$\Theta(\mathcal{H}^1(\gamma \setminus \Sigma), \mathcal{H}^1(\gamma), \mathcal{H}^1(\Sigma)) = \overline{J}(\mathcal{H}^1(\gamma \setminus \Sigma), \mathcal{H}^1(\gamma \cap \Sigma), \mathcal{H}^1(\Sigma)).$$

We now study some properties of Θ.

Proposition 2.3.4. Θ *is monotone, non-decreasing with respect to each of its variables.*

Proof. The monotonicity in the third variable is straightforward. The one in the first variable can be obtained observing that

$$\Theta(a,b,c) = \inf_{a \le s \le b} J(s, b-s, c) \tag{2.3.1}$$

and that the right-hand side of (2.3.1) is a non-decreasing function of a. The monotonicity in the second variable is obtained in a similar way, still relying on (2.3.1) and paying attention to the sets where the infimum is taken. □

Proposition 2.3.5. Θ *is lower semicontinuous.*

Proof. We have to show that

$$\Theta(a,b,c) \leq \liminf_{n\to+\infty} \Theta(a_n,b_n,c_n)$$

when $a_n \to a$, $b_n \to b$ and $c_n \to c$. Let us consider for every real positive number ε and for every positive integer n a real number s_n such that $a_n \leq s_n \leq b_n$ and

$$J(s_n, b - s_n, c_n) \leq \Theta(a_n, b_n, c_n) + \varepsilon.$$

Up to a subsequence, we can suppose that

$$\liminf_{n\to+\infty} \Theta(a_n,b_n,c_n) = \lim_{n\to+\infty} \Theta(a_n,b_n,c_n).$$

We can also suppose that $s_n \to s$, where $a \leq s \leq b$. Thanks to the semicontinuity of J

$$\Theta(a,b,c) \leq J(s, b-s, c) \leq \liminf_{n\to+\infty} J(s_n, b_n - s_n, c_n) \leq \liminf_{n\to+\infty} \Theta(a_n,b_n,c_n) + \varepsilon.$$

Letting $\varepsilon \to 0^+$ yields the desired inequality. □

2.4 Existence theorem

In this section we continue to develop the tools we will use to prove Theorem 2.4.5.

Proposition 2.4.1. *Let $\{x_n\}_{n\in\mathbb{N}}$ and $\{y_n\}_{n\in\mathbb{N}}$ be sequences in K such that $x_n \to x$ and $y_n \to y$. If $\{\Sigma_n\}_{n\in\mathbb{N}}$ is a sequence of closed connected sets such that $\Sigma_n \to \Sigma$, then*

$$d_\Sigma(x,y) \leq \liminf_{n\to+\infty} d_{\Sigma_n}(x_n,y_n). \tag{2.4.1}$$

Proof. First, up to a subsequence, we can suppose that

$$\liminf_{n\to+\infty} d_{\Sigma_n}(x_n,y_n) = \lim_{n\to+\infty} d_{\Sigma_n}(x_n,y_n).$$

Given $\varepsilon > 0$, we choose a sequence $\{\gamma_n\}_{n\in\mathbb{N}}$ such that $\gamma_n \in \mathscr{C}_{x_n,y_n}$ and

$$\Theta(\mathcal{H}^1(\gamma_n \setminus \Sigma_n), \mathcal{H}^1(\gamma_n), \mathcal{H}^1(\Sigma_n)) \leq d_{\Sigma_n}(x_n,y_n) + \varepsilon.$$

Up to a subsequence we can suppose that $\gamma_n \to \gamma$ (it is easy to check that $x_n \to x$ and $y_n \to y$ imply $\gamma \in \mathscr{C}_{x,y}$) and

$$\begin{aligned}\mathcal{H}^1(\gamma \setminus \Sigma) &\leq \lim_n \mathcal{H}^1(\gamma_n \setminus \Sigma_n),\\ \mathcal{H}^1(\gamma) &\leq \lim_n \mathcal{H}^1(\gamma_n),\\ \mathcal{H}^1(\Sigma) &\leq \lim_n \mathcal{H}^1(\Sigma_n).\end{aligned}$$

Using the semicontinuity and monotonicity of Θ (Propositions 2.3.4 and 2.3.5), we obtain

$$\begin{aligned}d_\Sigma(x,y) &\leq \Theta\left(\mathcal{H}^1(\gamma \setminus \Sigma), \mathcal{H}^1(\gamma), \mathcal{H}^1(\Sigma)\right)\\ &\leq \Theta\left(\lim_{n\to+\infty} \mathcal{H}^1(\gamma_n \setminus \Sigma_n), \lim_{n\to+\infty} \mathcal{H}^1(\gamma_n), \lim_{n\to+\infty} \mathcal{H}^1(\Sigma_n)\right)\\ &\leq \liminf_{n\to+\infty} \Theta\left(\mathcal{H}^1(\gamma_n \setminus \Sigma_n), \mathcal{H}^1(\gamma_n), \mathcal{H}^1(\Sigma_n)\right)\\ &\leq \liminf_{n\to+\infty} d_{\Sigma_n}(x_n, y_n) + \varepsilon.\end{aligned}$$

The arbitrary choice of ε gives then inequality (2.4.1). □

As a consequence of Proposition 2.4.1 we have the following Corollary.

Corollary 2.4.2. *Let $\{x_n\}_{n\in\mathbb{N}}$ and $\{y_n\}_{n\in\mathbb{N}}$ be sequences in K such that $x_n \to x$ and $y_n \to y$. If Σ is a closed connected set, then*

$$d_\Sigma(x,y) \leq \liminf_{n\to+\infty} d_\Sigma(x_n, y_n).$$

In other words, d_Σ is a lower semicontinuous function on $K \times K$.

Proposition 2.4.4 will play a crucial role in the proof of our main existence result. We split its proof in the next two lemmas for convenience.

Lemma 2.4.3. *Let X be a compact metric space, $\{f_n\}_{n\in\mathbb{N}}$ a sequence of positive real valued functions defined on X. Let also g be a continuous positive real valued function defined on X. Then, the following statements are equivalent:*

1. $\forall \varepsilon > 0\ \exists N\ :\ \forall n \geq N\ \forall x \in X \quad g(x) \leq f_n(x) + \varepsilon$,

2. $\forall x \in X\ \forall x_n \to x \quad g(x) \leq \liminf_n f_n(x_n)$.

Proof.

- Let $x_n \to x$. Then
$$g(x_n) = f_n(x_n) + (g(x_n) - f_n(x_n)) \leq f_n(x_n) + \varepsilon$$
By the continuity of g, taking the lower limit we achieve
$$g(x) \leq \liminf_{n\to+\infty} f_n(x_n) + \varepsilon. \tag{2.4.2}$$
Then (1) $\Rightarrow$ (2) is established when $\varepsilon \to 0^+$.

- Let us now prove that (2) $\Rightarrow$ (1). Suppose on the contrary that there exists a positive ε and an increasing sequence of positive integers $\{n_k\}_k$ such that
$$g(x_{n_k}) \geq f_{n_k}(x_{n_k}) + \varepsilon \tag{2.4.3}$$
for a suitable x_{n_k}. Thanks to the compactness of X we can suppose up to a subsequence that $x_{n_k} \to x$. Define
$$x_n = \begin{cases} x_{n_k} & \text{if } n = n_k \text{ for some } k \\ x & \text{otherwise} \end{cases}$$
Then $x_n \to x$, and $g(x) \leq \liminf_n f_n(x_n)$. From (2.4.3) it follows,
$$g(x) \geq \liminf_{k\to+\infty} f_{n_k}(x_{n_k}) + \varepsilon \geq \liminf_{n\to+\infty} f_n(x_n) + \varepsilon \geq g(x) + \varepsilon$$
which is false. □

Proposition 2.4.4. *Let $\{f_n\}_{n\in\mathbb{N}}$ and f be non-negative lower semicontinuous functions, all defined on a compact metric space (X, d). Let $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence of nonnegative measures on X such that $\mu_n \rightharpoonup^* \mu$. Suppose that*
$$\forall x \in X\ \forall x_n \to x \quad f(x) \leq \liminf_{n\to+\infty} f_n(x_n).$$
Then
$$\int_X f \ \mathrm{d}\mu \leq \liminf_{n\to+\infty} \int_X f_n \ \mathrm{d}\mu_n.$$

Proof. Let ψ be a continuous function with compact support such that $0 \leq \psi \leq 1$. Let g_t be the function of Lemma 1.2.1; since g_t satisfies the hypothesis of Lemma 2.4.3 with $g = g_t$, we have $g_t \leq f_n + \varepsilon$ for n large enough and then
$$\int_X g_t\psi \ \mathrm{d}\mu = \lim_{n\to+\infty} \int_X g_t\psi \ \mathrm{d}\mu_n \leq \liminf_{n\to+\infty} \int_X f_n \ \mathrm{d}\mu_n.$$
Taking the supremum in t and ψ, we obtain
$$\int_X f \ \mathrm{d}\mu \leq \liminf_{n\to+\infty} \int_X f_n \ \mathrm{d}\mu_n. \quad \square$$

We may now state and prove our existence result.

Theorem 2.4.5. *The problem*

$$\min\{T(\Sigma)\ :\ \Sigma \in \mathscr{C}\}$$

admits a solution.

Proof. First, let us prove that for every $l > 0$ the class

$$\mathscr{D}_l := \{\Sigma\ :\ \Sigma \in \mathscr{C},\ \mathcal{H}^1(\Sigma) \le l\}$$

is a compact subset of the metric space $(\mathscr{C}(K), d_{\mathcal{H}})$. Since $(\mathscr{C}(K), d_{\mathcal{H}})$ is a compact space, it is enough to show that $\mathscr{D}_l$ is closed. We already know that the Hausdorff limit of a sequence of closed connected set is a closed connected set. If $\{\Sigma_n\}_{n\in\mathbb{N}}$ is a sequence of closed connected sets such that $\mathcal{H}^1(\Sigma_n) \le l$

$$\Sigma_n \to \Sigma \Longrightarrow \mathcal{H}^1(\Sigma) \le \liminf_{n\to+\infty} \mathcal{H}^1(\Sigma_n) \le l$$

by Gołab Theorem (Theorem 2.2.2).

Second, by our assumption on the function J

$$d_\Sigma(x,y) \ge G(\mathcal{H}^1(\Sigma))$$

so that

$$T(\Sigma) \ge G(\mathcal{H}^1(\Sigma)).$$

Then, if $\{\Sigma_n\}_{n\in\mathbb{N}}$ is a minimizing sequence, the sequence of 1-dimensional Hausdorff measures $\{\mathcal{H}^1(\Sigma_n)\}_{n\in\mathbb{N}}$ must be bounded, i.e. $\mathcal{H}^1(\Sigma_n) \le l$, for some $l > 0$.

If we prove that the functional $\Sigma \mapsto T(\Sigma)$ is sequentially lower semicontinuous on the class $\mathscr{D}_l$, then then existence of an optimal Σ will be a consequence of the fact that a sequentially lower semicontinuous function takes a minimum on a compact metric space. Let $\{\Sigma_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathscr{D}_l$ such that $\Sigma_n \to \Sigma$. Let $\{\mu_n\}_{n\in\mathbb{N}}$ be an optimal transport plan for the transport problem

$$\min\left\{\int_{K\times K} d_{\Sigma_n}(x,y)\mathrm{d}\mu\ :\ \pi^+_\#\mu = \mu^+, \pi^-_\#\mu = \mu^-\right\}.$$

Up to a subsequence we can suppose $\mu_n \rightharpoonup^* \mu$ for a suitable μ. It is easy to see that μ is a transport plan between μ^+ and μ^-.

Since by Proposition 2.4.1 $d_\Sigma(x,y) \le \liminf_n d_{\Sigma_n}(x_n, y_n)$ for all $x_n \to x$ and $y_n \to y$, by Lemma 2.4.4 we have

$$\int_{K\times K} d_\Sigma(x,y)\ \mathrm{d}\mu \le \liminf_{n\to+\infty} \int_{K\times K} d_{\Sigma_n}(x,y)\ \mathrm{d}\mu_n. \tag{2.4.4}$$

Then by (2.4.4) we have

$$T(\Sigma) \le \int_{K\times K} d_\Sigma(x,y)\ \mathrm{d}\mu \le \liminf_{n\to+\infty} \int_{K\times K} d_{\Sigma_n}(x,y)\ \mathrm{d}\mu_n = \liminf_{n\to+\infty} T(\Sigma_n). \quad \square$$

We end with the following remark.

Remark 2.4.6. Note that if Σ_n is a minimizing sequence, then the measure μ obtained in the proof of Theorem 2.4.5 is an optimal transport plan for the transport problem

$$\min\left\{\int_{K\times K} d_\Sigma(x,y)\ \mathrm{d}\mu\ :\ \pi^+_\#\mu = \mu^+, \pi^-_\#\mu = \mu^-\right\}.$$

Chapter 3

Path Functionals over Wasserstein Spaces

3.1 Introduction

The problem of transporting a source mass distribution onto a target mass distribution by keeping together as much mass as possible during the transport, from which tree-shaped configurations arise, has been very much studied, for instance in [6], [7], [34], and [48]. In the approach to this problem proposed in this chapter and in [14] probability measures valued curves are considered, while the condition of keeping masses together is achieved considering only measures supported in discrete sets.

Given a source or initial probability measure μ_0 and a target or final probability measure μ_1 we look for a path γ in a Wasserstein space $\mathcal{W}_p(\Omega)$ that connects μ_0 to μ_1 and minimizes a suitable cost functional $\mathcal{J}(\gamma)$. We consider functionals of the form

$$\mathcal{J}(\gamma) = \int_0^1 J(\gamma(t))|\gamma'|(t)\mathrm{d}t \tag{3.1.1}$$

where $|\gamma'|$ is the metric derivative of γ in the Wasserstein space $\mathcal{W}_p(\Omega)$ and J is a lower semicontinuous functional defined on measures. Here J may be easily seen as the coefficient of a degenerate "Riemannian distance" on the space $\mathcal{W}_p(\Omega)$.

We restrict our analysis to the case of J being a local functional over measures, an important class of functionals extensively studied by Bouchitté and Buttazzo in [9], [10], and [11]. These functionals are the key tool in our approach, and among them we can find both functionals which are finite only on concentrated measures (we can see an application of them in [19] and [44]) and functionals which are finite only on spread ones. In fact, a particular point of interest in our approach is the fact that also different kinds of "Riemannian distances" are allowed (for instance those which prefer spread measures) by a change of the functional J.

In particular, we consider the two extreme cases, in which the functional J is chosen as one of the following:

$$G_r(\mu) = \begin{cases} \sum_{k\in\mathbb{N}}(a_k)^r & \text{if } \mu = \sum_{k\in\mathbb{N}} a_k\delta_{x_k} \\ +\infty & \text{otherwise} \end{cases} \qquad (0 \le r < 1)$$

whose domain is the space of purely atomic measures, or

$$F_q(\mu) = \begin{cases} \int_\Omega |u|^q \mathrm{d}x & \text{if } \mu = u\cdot\mathcal{L}^N \\ +\infty & \text{otherwise} \end{cases} \qquad (q > 1)$$

whose domain is the space $L^q(\Omega)$. We denote respectively by $\mathcal{G}_r$ the functional in (3.1.1) with J replaced by G_r and by $\mathcal{F}_q$ the same functional with J replaced by F_q.

The first case is the one in which we get a "Riemannian distance" on probabilities which make paths passing through concentrated measures cheaper. The second case, on the contrary, allows only paths which lie on $L^q(\Omega)$.

In both cases we analyze the question of the existence of optimal paths γ_{opt} giving finite value to the functional. When the domain $\Omega \subset \mathbb{R}^N$ is compact we find for the first case:

- if μ_0 and μ_1 are atomic measures, then an optimal path γ_{opt} providing finite value to $\mathcal{G}_r$ always exists;
- if $r > 1 - 1/N$, then the same is true for any pair of measures;
- if $r \le 1-1/N$, then there are measures μ_0 and μ_1 such that every path connecting them has an infinite cost.

Similarly, for the second case we have:

- if μ_0 and μ_1 are in $L^q(\Omega)$, then an optimal path γ_{opt} providing finite value to $\mathcal{F}_q$ always exists;
- if $q < 1 + 1/N$, then the same is true for any pair of measures;

- if $q \geq 1+1/N$, then there are measures μ_0 and μ_1 such that every path connecting them has an infinite cost.

In subsection 3.3.2 we also discuss the case of unbounded domains such as $\Omega = \mathbb{R}^N$.

The analysis of existence results as well as the definition of the cost functionals is made in Section 3.2 in an abstract metric spaces framework, which can be used for future generalizations and developments.

In relation to the papers already mentioned it is not difficult to see that the model proposed is different and in general provides different solutions. However, among the different features our model supplies we may cite its mathematical simplicity and the possibility of performing standard numerical computations.

From the mathematical point of view, our model recalls the construction of Riemannian metrics as already pointed out, and the existence results for optimal paths are quite easy to prove.

As far as numerics is concerned, when discretizing the metric derivative the cost functional becomes a weighted sum of Wasserstein distances among couples of atomic probability measures which can be evaluated by well-known algorithms such as the simplex method.

Taking into account the comparison with the results presented by Xia in [48] and by Maddalena, Morel and Solimini in [34] will be important for future investigations. For instance, for the model proposed in [34] conditions to link two prescribed measures by a finite cost configuration have been studied in [23] (while here and in [48] only conditions in order to link arbitrary measures are provided): we do not know if similar conditions can be achieved in our case.

3.2 The Metric Framework

In this section a generic metric space X with distance d is considered. Under the assumption that closed bounded subsets of X are compact, we will prove an existence result (Theorem 3.2.1) for variational problems with functionals of the type

$$\mathcal{J}(\gamma) = \int_0^1 J(\gamma(t))|\gamma'|(t)\ \mathrm{d}t$$

where $\gamma : [0,1] \to X$ ranges among all Lipschitz curves such that $\gamma(0) = x_0$ and $\gamma(1) = x_1$. We will refer to the value of $\mathcal{J}$ in γ as the *energy* of γ. By $|\gamma'|(t)$ we denote the *metric derivative* of γ at the point $t \in (0,1)$, i.e.

$$|\gamma'|(t) = \lim_{s \to t} \frac{d(\gamma(s), \gamma(t))}{|s-t|}.$$

As a consequence of Rademacher Theorem it can be seen (see [5]) that for any Lipschitz curve the metric derivative exists in almost every point (with respect to Lebesgue measure). Another useful result is that the variation of γ can be written in terms of the metric derivative in integral form:

$$\mathrm{Var}(\gamma) = \int_0^1 |\gamma'|(t)\ \mathrm{d}t.$$

By this formula it follows easily that $|\gamma'| \leq M$ if and only if γ is M-Lipschitz, since when $s < t$

$$d(\gamma(t), \gamma(s)) \leq \mathrm{Var}(\gamma, [s,t]) = \int_s^t |\gamma'|(\tau)\ \mathrm{d}\tau \leq M|t-s|,$$

the converse implication being immediate.

Theorem 3.2.1. *Let X be a metric space such that any closed bounded subset of X is compact and $J : X \to [0, +\infty]$ be a lower semicontinuous function and x_0, x_1 arbitrary points in X. Then the functional*

$$\mathcal{J}(\gamma) = \int_0^1 J(\gamma(t))|\gamma'|(t)\ \mathrm{d}t$$

achieves a minimum value among all Lipschitz curves $\gamma : [0,1] \to X$ such that $\gamma(0) = x_0$ and $\gamma(1) = x_1$, provided the following two assumptions are satisfied:

(H1): there exists a curve γ_0 such that $\mathcal{J}(\gamma_0) < +\infty$;

(H2): it holds

$$\int_0^\infty \inf_{B_r(x_0)} J\ \mathrm{d}r = +\infty.$$

The proof of Theorem 3.2.1 relies on the following reparametrization lemma whose proof can be found for example in [5].

Lemma 3.2.2. *Let $\gamma \in \mathrm{Lip}([0,1], X)$ and $L = \mathrm{Var}(\gamma)$ be its total variation. Then there exists a Lipschitz curve $\tilde{\gamma} \in \mathrm{Lip}([0,L], X)$ such that $|\tilde{\gamma}'| = 1$ almost everywhere in $[0,L]$ and $\tilde{\gamma}$ is a parametrization of γ.*

Proof of Theorem 3.2.1. Let $\{\gamma_n\}_{n\in\mathbb{N}}$ be a minimizing sequence and set $L_n = \mathrm{Var}(\gamma_n)$. Then the sequence $\{\mathcal{J}(\gamma_n)\}_{n\in\mathbb{N}}$ is bounded by a finite number M. By Lemma 3.2.2 there exists a sequence of curves $\tilde{\gamma}_n : [0, L_n] \to X$ parametrized with unit velocity, reparametrizing the given curves. We have:

$$M \geq \mathcal{J}(\gamma_n) = \int_0^{L_n} J(\tilde{\gamma}_n(t))\ \mathrm{d}t \geq \int_0^{L_n} \left(\inf_{B_t(x_0)} J \right)\ \mathrm{d}t.$$

Then $\{L_n\}_{n\in\mathbb{N}}$ is bounded otherwise, by assumption H2, the right hand side would be unbounded. We can reparametrize each curve γ_n at constant speed L_n, thus obtaining a new sequence $\{\hat{\gamma_n}\}_{n\in\mathbb{N}}$ in $\mathrm{Lip}([0,1], X)$, which is still a minimizing sequence, thanks to the equality $\mathcal{J}(\gamma_n) = \mathcal{J}(\hat{\gamma_n})$. Being $\{L_n\}_{n\in\mathbb{N}}$ bounded, we get that this new minimizing sequence is uniformly bounded and uniformly Lipschitz. By Ascoli-Arzelà Theorem we can suppose up to a subsequence that $\hat{\gamma}_n \to \hat{\gamma}$ uniformly for some L-Lipschitz curve $\hat{\gamma}$ where we have taken $L = \liminf_n L_n$. By recalling the link between Lipschitz conditions and metric derivative we have

$$|\hat{\gamma}'|(t) \leq L \quad \text{for a.e. } t \in [0,1].$$

Now by using the lower semicontinuity of the functional J, we obtain

$$\begin{aligned}\mathcal{J}(\hat{\gamma}) = \int_0^1 J(\hat{\gamma}(t))|\hat{\gamma}'|(t)\ \mathrm{d}t &\leq L \int_0^1 \liminf_{n\to+\infty} J(\hat{\gamma}_n(t))\ \mathrm{d}t \\ &\leq \liminf_{n\to+\infty} L_n \int_0^1 J(\hat{\gamma}_n(t))\ \mathrm{d}t = \liminf_{n\to+\infty} \mathcal{J}(\hat{\gamma}_n),\end{aligned}$$

that is the lower semicontinuity of $\mathcal{J}$ on the considered sequence, which achieves the proof. □

Remark 3.2.3. Notice that the integral assumption H2 is always true if $J \geq c$ for a suitable strictly positive constant. Moreover Theorem 3.2.1 still holds if condition H2 is replaced by the weaker assumption that there exists a curve γ_0 such that

$$\mathcal{J}(\gamma_0) < \int_0^{+\infty} \inf_{B(x_0,r)} J\,\mathrm{d}r.$$

We give here a slightly refined version of Theorem 3.2.1, which will be useful in the last section. The goal here is to weaken the compactness assumption on bounded subsets of X.

Theorem 3.2.4. *Let (X, d, d') be a metric space endowed with two different distances, such that:*

(K1): $d' \leq d$;

(K2): *all $d-$bounded sets in X are relatively compact with respect to d';*

(K3): *the mapping $d : X \times X \to \mathbb{R}^+$ is a lower semicontinuous function with respect to the distance $d' \times d'$.*

Let $J : X \to [0, +\infty]$ be lower semicontinuous with respect to d'. Consider the functional, defined on the set of $d-$Lipschitz curves $\gamma : [0,1] \to X$, given by

$$\mathcal{J}(\gamma) = \int_0^1 J(\gamma(t))|\gamma'|_d(t)\mathrm{d}t,$$

where $|\gamma'|_d(t)$ stands for the metric derivative of γ with respect to d. Then, with the same hypotheses H1 and H2 (where $B_r(x_0)$ are in the d-sense) of Theorem 3.2.1, there exists a minimum for $\mathcal{J}$.

Proof. We can take a minimizing sequence $\{\gamma_n\}_n$ and, as in Theorem 3.2.1, reparametrize it to obtain a sequence $\{\hat{\gamma}_n\}_n$ in which every curve has constant speed L_n. Hypothesis H2 gives us the boundedness of L_n. Hence the sequence $\{\hat{\gamma}_n\}_n$ is composed by $d-$equicontinuous functions from $[0,1]$ to a $d-$bounded subset of X. If we endow X with the distance d' we have an equicontinuous (thanks to assumption K1) sequence of functions whose images are contained in a compact set. We can consequently use Ascoli-Arzelà Theorem to choose a subsequence (not relabelled), such that $\hat{\gamma}_n \to \gamma$, for a suitable curve γ (uniformly in the $d'-$sense).

The lower semicontinuity of J with respect to d' allows us to use Fatou Lemma and shows that γ minimizes $\mathcal{J}$, as far as we can show that γ is $d-$Lipschitz with a Lipschitz constant not exceeding $\liminf_n L_n$. To do this we use assumption K3. Taken two points s, t we have in fact:

$$d(\gamma(s), \gamma(t)) \leq \liminf_n d(\hat{\gamma}_n(s), \hat{\gamma}_n(t)) \leq \liminf_n L_n|s-t|,$$

which shows the required Lipschitz property. □

3.3 The Case of Wasserstein Spaces

In this section we consider a compact metric space Ω equipped with a distance function c and a positive finite non-atomic Borel measure m. We consider the *p-Wasserstein metric space* $\mathcal{W}_p(\Omega)$. As we have seen in Section 1.5, this is the space of Borel probability measures μ on Ω with finite momentum of order p with respect to a point x_0

$$\int_\Omega c(x, x_0)^p \, \mathrm{d}\mu < +\infty,$$

equipped with the *p-Wasserstein distance*

$$W_p(\mu_1, \mu_2) = \inf \left(\int_{\Omega\times\Omega} c(x, y)^p \, \lambda(\mathrm{d}x, \mathrm{d}y) \right)^{1/p}$$

where the infimum is taken on all transport plans λ between μ_1 and μ_2, that is on all probability measures λ on $\Omega \times \Omega$ whose marginals $\pi^+_\#\lambda$ and $\pi^-_\#\lambda$ coincide with μ_1 and μ_2 respectively.

Notice that, since the distance c is bounded, the space $\mathcal{W}_p(\Omega)$ consists of all probability measures. We consider functions J on $\mathcal{W}_p(\Omega)$ that can be represented in the following form:

$$J(\mu) = \int_\Omega f\left(\frac{\mathrm{d}\mu}{\mathrm{d}m}\right) \, \mathrm{d}m + \int_{\Omega\setminus A_\mu} f^\infty \left(\frac{\mathrm{d}\mu^s}{\mathrm{d}|\mu^s|}\right) \, \mathrm{d}|\mu^s| + \int_{A_\mu} g(\mu(x)) \, \mathrm{d}\#(x)$$

where

- $\mathrm{d}\mu/\mathrm{d}m$ is Radon-Nikodym derivative of μ with respect to m,
- $f : \mathbb{R} \to [0, +\infty]$ is convex, lower semicontinuous and proper (i.e. not identically $+\infty$),
- μ^s is the singular part of μ with respect to m according to the Radon-Nikodym decomposition theorem;
- f^∞ is the *recession function*

  $$f^\infty(s) := \lim_{t\to+\infty} \frac{f(s_0 + ts)}{t}$$

  (the limit is independent of the choice of s_0 in the domain of f, i.e. the set of points where f is finite),

- A_μ is the set of atoms of μ, i.e. the points such that $\mu(x) := \mu(\{x\}) > 0$,
- $g : \mathbb{R} \to [0, +\infty]$ is a lower semicontinuous subadditive function such that $g(0) = 0$
- $\#$ is the counting measure.

Note that our functional can be written in a simpler form since in our case $\mathrm{d}\mu^s/\mathrm{d}|\mu^s| = 1$ for $|\mu^s|$-a.e. of x, being μ a positive measure:

$$J(\mu) = \int_\Omega f\left(\frac{\mathrm{d}\mu}{\mathrm{d}m}\right)\,\mathrm{d}m + f^\infty(1)|\mu^s|(\Omega \setminus A_\mu) + \int_{A_\mu} g(\mu(x))\,\mathrm{d}\#(x).$$

By the results that can be found in [9] and [10], these functionals are lower semicontinuous for the weak-$*$ convergence of measures (and represent all local functionals with this semicontinuity property) whenever

$$g_0(s) := \sup_{t>0} \frac{g(st)}{t} = f^\infty(s).$$

Theorem 3.3.1. *Suppose that $f(s) > 0$ for $s > 0$ and $g(1) > 0$. Then we have $J \geq c > 0$. In particular, the functional $\mathcal{J}$ defined on the set of Lipschitz curves $\gamma : [0,1] \to \mathcal{W}_p(\Omega)$ with given starting and ending point achieves a minimum, provided that there exists a curve with finite cost.*

Proof. Let us fix some notation. By μ^a we mean the absolutely continuous part of μ with respect to the measure m, and by $\mu^s, \mu^\#, \mu^c$ respectively the singular part, the atomic part and the singular diffused part of μ. Then we have $\mu = \mu^a + \mu^s = \mu^a + \mu^c + \mu^\#$. Since f is convex, by Jensen inequality we have

$$\int_\Omega f\left(\frac{\mathrm{d}\mu}{\mathrm{d}m}\right)\,\mathrm{d}m \geq m(\Omega) f\left(\frac{1}{m(\Omega)}\int_\Omega \frac{\mathrm{d}\mu}{\mathrm{d}m}\,\mathrm{d}m\right) = m(\Omega) f\left(\frac{\mu^a(\Omega)}{m(\Omega)}\right). \tag{3.3.1}$$

Since μ is a positive measure and f^∞ is $1-$homogeneous

$$\int_{\Omega\setminus A_\mu} f^\infty\left(\frac{\mathrm{d}\mu^s}{\mathrm{d}|\mu^s|}\right)\,\mathrm{d}|\mu^s| = |\mu^s|(\Omega \setminus A_\mu) f^\infty(1) = m(\Omega) f^\infty\left(\frac{\mu^c(\Omega)}{m(\Omega)}\right). \tag{3.3.2}$$

Since g is a subadditive function

$$\int_{A_\mu} g(\mu(x))\,\mathrm{d}\#(x) =$$

$$\sum_{x\in A_\mu} g(\mu(x)) \geq g\left(\sum_{x\in A_\mu} \mu(x)\right) = g(\mu^{\#}(\Omega)). \quad (3.3.3)$$

For the recession function f^∞ it holds

$$f^\infty(x) \geq f(x+y) - f(y) \text{ for all } x, y \in \mathbb{R},$$

and so the sum of the first two terms, i.e. those given by (3.3.1) and (3.3.2), can be estimated from below by

$$m(\Omega) f\left(\frac{\mu^a(\Omega) + \mu^c(\Omega)}{m(\Omega)}\right).$$

Therefore summing up (3.3.1), (3.3.2) and (3.3.3) we obtain

$$J(\mu) \geq m(\Omega) f\left(\frac{\mu^a(\Omega) + \mu^c(\Omega)}{m(\Omega)}\right) + g(\mu^{\#}(\Omega)).$$

We set $a = \mu^{\#}(\Omega)$ and $1 - a = \mu^a(\Omega) + \mu^c(\Omega)$. Since the function $a \mapsto m(\Omega) f((1-a)/m(\Omega)) + g(a)$ is lower semicontinuous, it attains a minimum in the interval $[0,1]$. Thanks to our hypothesis this sum is always positive, and so we have

$$\min_{0\leq a\leq 1} m(\Omega) f\left(\frac{1-a}{m(\Omega)}\right) + g(a) = c > 0,$$

that is, we have $J(\mu) \geq c > 0$. $\square$

3.3.1 Bounded domains

We now study some special cases of the functional we defined above. In the rest of this section Ω will be a *compact convex* subset of $\mathbb{R}^N$ and the measure m will be the Lebesgue measure $\mathcal{L}^N$ on it.

First case: $f = +\infty,\ g(z) = |z|^r\ (0 \le r < 1)$.

In this case we will denote the functional J by G_r and the corresponding functional $\mathcal{J}$ on Lipschitz paths will be called $\mathcal{G}_r$. This is the case when G_r is finite only on purely atomic measures.

We are now going to consider the question whether there exists a curve connecting two given measures keeping finite our functional. First we prove that if both the initial and the final measure are atomic the answer is positive. Then we prove that for r in a suitable subinterval of $[0,1]$ every measure can be connected to a Dirac mass, hence every measure can be connected to every other measure by a path of finite energy. Finally we show that this is not possible in general for every $r \in [0,1]$.

Theorem 3.3.2. *Let μ_0 and μ_1 be convex combinations of Dirac masses, i.e.,*

$$\mu_0 = \sum_{k=1}^{m} a_k \delta_{x_k}, \quad \mu_1 = \sum_{l=1}^{n} b_l \delta_{y_l}$$

with $a_k, b_l > 0$, $\sum_k a_k = \sum_l b_l = 1$. Then, there exists a Lipschitz curve $\gamma : [0,1] \to \mathcal{W}_p(\Omega)$ such that $\gamma(0) = \mu_0, \gamma(1) = \mu_1$ and

$$\mathcal{G}_r(\gamma) = \int_0^1 G_r(\gamma(t))|\gamma'|(t)\ \mathrm{d}t < +\infty.$$

Proof. It is sufficient to prove the theorem when $a_1 = 1$, i.e. $\mu_0 = \delta_{x_1}$ since in the general case one connects the first measure μ_0 to a Dirac mass supported in an arbitrary point, then one connects that Dirac mass to the final measure μ_1. If one can keep finite the functional in both steps, then the result is proved in the general case.

We now prove that the curve $\gamma : [0,1] \to \mathcal{W}_p(\Omega)$ given by:

$$\gamma(t) = \sum_{l=1}^{n} b_l \delta_{x_1 + t(y_l - x_1)}.$$

is W_p-Lipschitz and $\mathcal{G}_r(\gamma) < +\infty$. Let t_1 and t_2 be time instants such that $t_1 < t_2$. Then, the transport plan between the probability measures given by $\sum_l b_l \delta_{x_1 + t_1(y_l - x_1)}$ and $\sum_l b_l \delta_{x_1 + t_2(y_l - x_1)}$ induced by the map $T(x_1 + t_1(y_l -$

$x_1)) := x_1 + t_2(y_l - x_1)$ gives:

$$
\begin{aligned}
W_p(\gamma(t_2),\gamma(t_1)) &\le \left(\int_\Omega |x - T(x)|^p \,\mathrm{d}\gamma(t_1)\right)^{1/p} \\
&= \left(\sum_{l=1}^n b_l |t_2 - t_1|^p |y_l - x_1|^p\right)^{1/p} = |t_2 - t_1| \left(\sum_{l=1}^n b_l |y_l - x_1|^p\right)^{1/p}.
\end{aligned}
$$

Hence the metric derivative with the respect to the Wasserstein p-distance is given by:

$$|\gamma'|(t) \le \left(\sum_{l=1}^n b_l |y_l - x_1|^p\right)^{1/p} = W_p(\mu_0, \mu_1).$$

On the other hand we have:

$$G_r(\mu) = \begin{cases} \sum_{x \in A_\mu} |\mu(x)|^r & \text{if } \mu^a = \mu^c = 0 \\ +\infty & \text{otherwise.} \end{cases}$$

Then

$$G_r(\gamma(t)) = \begin{cases} 1 & \text{if } t = 0 \\ \sum_{l=1}^n b_l^r & \text{if } t > 0. \end{cases}$$

Hence

$$
\begin{aligned}
\mathcal{G}_r(\gamma) = \int_0^1 G_r(\gamma(t))|\gamma'|(t)\,\mathrm{d}t \le \\
\sum_{l=1}^n |b_l|^r \left(\sum_{l=1}^n b_l |y_l - x_1|^p\right)^{1/p} < +\infty. \quad \square
\end{aligned}
$$

Remark 3.3.3. By repeating the proof of Theorem 3.3.2 one obtains that the statement still holds for infinite sums of Dirac masses (i.e. $m = n = +\infty$) provided $G_r(\mu_0)$ and $G_r(\mu_1)$ are finite, that is $\sum_k a_k^r < +\infty$ and $\sum_l b_l^r < +\infty$.

The proof of the next theorem is related to the one of Proposition 3.1 of [48].

Theorem 3.3.4. *Let $1 - 1/N < r \le 1$. Then given two arbitrary μ_0 and μ_1 in $\mathcal{W}_p(\Omega)$, there exists a curve joining them such that the functional $\mathcal{G}_r$ is finite.*

Proof. It is sufficient to prove that every measure can be joined to a Dirac mass in an arbitrary point. We prove first the statement for $\Omega = [0,1]^N$.

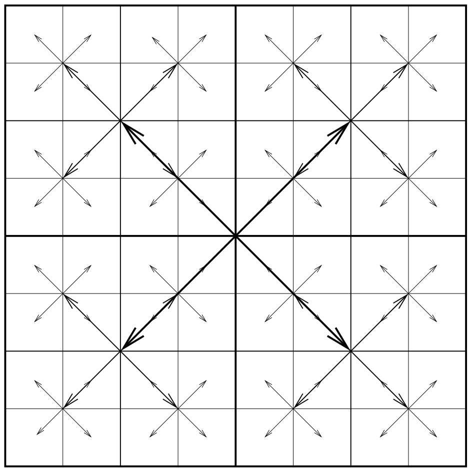

Figure 3.1: Approximation at step $k = 3$.

The *dyadic subdivision of order* k of $Q = [0,1]^N$ is given by the family of closed N-dimensional cubes $\{Q_h^k\}_{h\in I_k}$ where $I_k = \{1, 2, 3, \ldots, 2^k\}^N$ obtained by Q dividing each edge into 2^k pieces of equal length. We will refer to the elements of $\{Q_h^k\}_{h\in I_k}$ as k-cubes. To every Borel regular finite measure μ we associate the following sequence of measures:

$$\mu_k = \sum_{h\in I_k} b_h^k \delta_{y_h}$$

where $b_h = \mu(Q_h^k)$ and y_h is the center of Q_h^k. It is straightforward to see that $\mu_k \rightharpoonup^* \mu$ as $k \to +\infty$.

The idea is now simple (see Figure 3.1): first join μ_k to μ_{k+1} with an arc length parametrization γ_k, second put together all these curves to obtain a path from a Dirac mass to the measure μ. At every step a k-cube is divided in 2^N parts which are $(k+1)$-cubes. To bring the Dirac mass in the centre of the k-cube to the 2^N centres of the $(k+1)$-cubes with the right weights at each centre one splits the centre of the k-cube into 2^N parts moving towards the centres of the adjacent $(k+1)$-cubes in such a way that each point moves with unitary speed. At each step (see Figure 3.1 where the first three steps are represented) we obtain a curve γ_k defined on an interval of length $(1/2)^k d/2$ (d is the diagonal of Q) such that $|\gamma_k'|(t) = 1$ for all t.

Let us now compute the value of the functional on the curve γ made by joining all curves γ_k above. Since the function $f(x_1, \ldots, x_n) = \sum_i^n x_i^r$ with the constraint $\sum_i^n x_i = 1$ reaches its maximum at point $(1/n, \ldots, 1/n)$ we have:

$$\mathcal{G}_r(\gamma) = \sum_{k=1}^{+\infty} \left(\frac{1}{2} \frac{1}{2^k} d \sum_{h \in I_k} (b_h^k)^r \right) \leq \sum_{k=1}^{+\infty} \left(\frac{1}{2} \frac{1}{2^k} d 2^{Nk} \left(\frac{1}{2^{Nk}} \right)^r \right).$$

Since $1 - 1/N < r \leq 1$ the sum considered above is convergent.

In the case of a general Ω it is sufficient to consider a large cube containing the support of the measure μ such that the centre is contained in Ω. $\square$

The bound given by $r > 1 - 1/N$ is sharp. We have in fact the following result.

Theorem 3.3.5. *Suppose $r \leq 1 - 1/N$. Then there exists a probability measure μ on Ω such that every non-constant W_p-Lipschitz path γ such that $\gamma(0) = \mu$ is such that $\mathcal{G}_r(\gamma) = +\infty$.*

Proof. Let Ω be the cube $[0,1]^N$ and μ the Lebesgue measure on it. We want to estimate from below

$$\inf \{ G_r(\nu) \,|\, W_p(\mu, \nu) \leq t \}$$

and we will show it to be larger than $ct^{-N(1-r)}$. Therefore, if γ is a W_p-Lipschitz path with constant speed which originates from μ, the integral defining $\mathcal{G}_r$ diverges. We can simply consider $t = 2^{-k}$. To estimate $G_r(\nu)$, when ν is such that $W_p(\mu, \nu) \leq t$, consider a partition of Ω by small cubes of side ε. Let k be the number of those cubes Q_i such that $\nu(Q_i) \leq \mu(Q_i)/2 = \varepsilon^N/2$. In all these cubes we have a zone in which the optimal transport map s between μ and ν must take values outside the cube; this zone, given by $Q_i \setminus s^{-1}(Q_i)$, has a measure of at least $\varepsilon^N/2$. We want to estimate from below the contribute of this zone to the total transport cost between μ and ν. For this contribute we may write

$$\begin{aligned}
\int_{Q_i \setminus s^{-1}(Q_i)} d(x, \partial Q_i)^p \mathrm{d}x &= \int_0^{(\varepsilon/2)^p} \left| \left(Q_i \setminus s^{-1}(Q_i) \right) \cap \{ d(x, \partial Q_i)^p > \tau \} \right| \mathrm{d}\tau \\
&\geq \int_0^{(\varepsilon/2)^p} \left(\frac{\varepsilon^N}{2} - \left| \{ d(x, \partial Q_i)^p \leq \tau \} \right| \right) \mathrm{d}\tau \\
&\geq \int_0^{B^p \varepsilon^p} \left(\frac{\varepsilon^N}{2} - \left| \{ d(x, \partial Q_i) \leq B\varepsilon \} \right| \right) \mathrm{d}\tau \\
&\geq c_1 \varepsilon^p \varepsilon^N,
\end{aligned}$$

where B is sufficiently small and c_1 is a positive constant. By recalling that the total transport cost (i.e. the $p-$th power of the distance W_p) is less than t^p, we have

$$kc_1\varepsilon^{N+p} \leq t^p. \tag{3.3.4}$$

On the other hand, the value of G_r can be estimated from below by means of the other cubes and we have

$$G_r(\nu) \geq (\varepsilon^{-N} - k)c_2\varepsilon^{Nr}.$$

Let us now choose $\varepsilon = mt$ with m an integer such that $c_1 m^p > 1$ and, by using (3.3.4), we have

$$G_r(\nu) \geq t^{-N}(m^{-N} - m^{-N-p}/c_1)c_2 m^{Nr} t^{Nr} = c_3 t^{-N(1-r)},$$

where the constant c_3 is positive.

For general Ω we can simply use a cube contained in Ω and show that the Lebesgue measure on it, rescaled to a probability measure, cannot be reached keeping finite the value of the integral. □

Example 3.3.6 (Y-shaped paths versus V-shaped paths). Consider the example in Figure 3.2, where we suppose that l and h are fixed. We define for $0 \leq t \leq l_0$

$$x(t) = (t, 0)$$

and for $l_0 \leq t \leq l_0 + \sqrt{l_1^2 + h^2}$

$$x_1(t) = \left(l_0 + l_1 \frac{t - l_0}{\sqrt{l_1^2 + h^2}}, h\frac{t - l_0}{\sqrt{l_1^2 + h^2}} \right)$$
$$x_2(t) = \left(l_0 + l_1 \frac{t - l_0}{\sqrt{l_1^2 + h^2}}, -h\frac{t - l_0}{\sqrt{l_1^2 + h^2}} \right).$$

Let us consider the curve $\gamma : [0, l_0 + \sqrt{l_1^2 + h^2}] \to \mathcal{W}_p(\Omega)$ defined by

$$\gamma(t) = \begin{cases} \delta_{x(t)} & \text{if } 0 \leq t < l_0 \\ \frac{1}{2}\delta_{x_1(t)} + \frac{1}{2}\delta_{x_2(t)} & \text{if } l_0 \leq t \leq l_0 + \sqrt{l_1^2 + h^2}. \end{cases}$$

It easy to see that $|\gamma'|(t) = 1$ and that

$$\mathcal{G}_r(\gamma) = l_0 + 2^{1-r}\sqrt{(1 - l_0)^2 + h^2}.$$

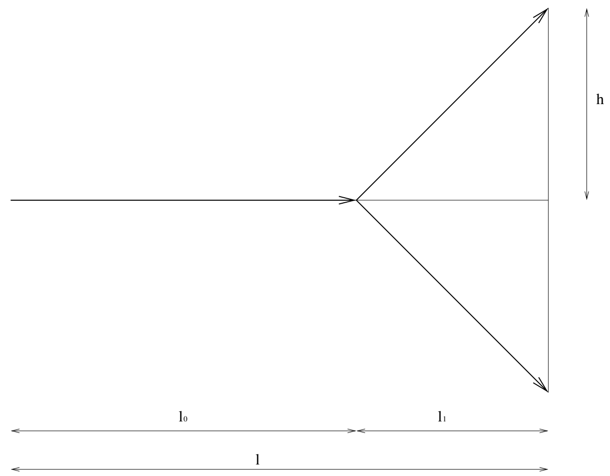

Figure 3.2: A Y-shaped path for $r = 1/2$.

Then the minimum is achieved for

$$l_0 = l - \frac{h}{\sqrt{4^{1-r} - 1}}.$$

In particular, when $r = 1/2$ we have a Y-shaped path (similar to the one of Figure 3.2) when $l > h$, while the path is V-shaped when $l \leq h$.

Remark 3.3.7. The result given by Theorem 3.3.4 can clearly be improved for particular choices of μ_0 and μ_1. For instance, we can connect a Dirac mass to the k-dimensional Hausdorff measure on a smooth k-surface for all $r \in [1 - 1/k, 1]$ (see also [34]).

Second case: $f(z) = |z|^q$ $(q > 1), g = +\infty$.

We follow the same structure of the previous section. In this case we will denote the functional J by F_q and $\mathcal{J}$ by $\mathcal{F}_q$.

We start by proving that when $F_q(\mu_0)$ and $F_q(\mu_1)$ are finite, that is μ_0 and μ_1 are measures with $L^q(\Omega)$ densities, the optimal path problem admits a solution with finite energy.

Theorem 3.3.8. *Assume that* $\mu_0 = u_0 \cdot \mathcal{L}^N, \mu_1 = u_1 \cdot \mathcal{L}^N$ *with* u_0, $u_1 \in L^q(\Omega)$. *Then* μ_0 *and* μ_1 *can be joined by a finite energy path.*

The proof of this result relies on the notion of *displacement convexity* which has been introduced in Chapter 1, Section 1.6. Recall that given μ_0

and μ_1 absolutely continuous probability measures on Ω and $T : \Omega \to \Omega$ optimal transport map (unique if $p > 1$) between μ_0 and μ_1 with respect to the cost function $|x-y|^p$, the map $\gamma^T : [0,1] \to \mathcal{W}_p(\Omega)$ given by

$$t \mapsto \gamma^T(t) := [(1-t)\mathrm{Id} + tT]_{\#}\mu_0 \tag{3.3.5}$$

is called a *displacement interpolation.*

Remark 3.3.9. It is well-known (see [1] or Remark 1.6.2 for the case $p = 2$) that the curve defined in (3.3.5) is a geodesic in $\mathcal{W}_p(\Omega)$, parametrized in such a way that

$$|(\gamma^T)'|(t) = W_p(\mu_0, \mu_1) \text{ for a.e. } t.$$

A functional F defined on all absolutely continuous measures (with respect to the Lebesgue measure) of $\mathcal{W}_p(\Omega)$ is said to be *displacement convex* if for every choice of μ_0, μ_1 absolutely continuous measures there exists an optimal transport map T such that

$$t \mapsto F(\gamma^T(t))$$

is convex on $[0,1]$.

Proof of Theorem 3.3.8. By Theorem 1.6.9 and Theorem 1.6.18 the functional F_q is displacement convex, so that

$$F_q(\gamma^T(t)) \leq (1-t)F_q(\mu_0) + tF_q(\mu_1).$$

Then

$$\int_0^1 F_q(\gamma^T(t))|(\gamma^T)'|(t)\ \mathrm{d}t \leq$$
$$W_p(\mu_0,\mu_1)\int_0^1 [(1-t)F_q(\mu_0) + tF_q(\mu_1)]\ \mathrm{d}t =$$
$$\frac{1}{2}(F_q(\mu_0) + F_q(\mu_1))W_p(\mu_0,\mu_1).$$

Since $F_q(\mu_0)$ and $F_q(u_1)$ are finite, we have that the path $t \mapsto \gamma^T(t)$ provides a finite value for the energy functional $\mathcal{F}_q$. $\square$

Next step will be the existence of an admissible path for arbitrary extremal measures, if q satisfies some additional constraints.

Recall that if μ_0 and μ_1 are probability measures given by L^1 densities (u_0 and u_1 respectively) and T is a transport map between them with sufficient regularity we have:

$$u_1(y) = u_0(T^{-1}(y))|\det \mathrm{D}T^{-1}(y)|.$$

Lemma 3.3.10. *Let $q < 1 + 1/N$. Let also $\mu = u \cdot \mathcal{L}^N$ with $u \in L^q(\Omega)$ and $\nu = \sum_{j=1}^k b_j \delta_{y_j}$ with $\sum_{j=1}^k b_j = 1$. Then there exists a path between μ and ν with finite energy.*

Proof. Let T be an optimal transport map between μ and ν. Let $B_j := T^{-1}(y_j)$. We now show that the path γ^T has a finite energy. Let us set $T_t = (1-t)\mathrm{Id}+tT$. If $x \in B_j$, then $T_t(x) = (1-t)x+ty_j$ and $\det \mathrm{D}T_t(x) = (1-t)^N$. Let u_t be the density of the measure $(T_t)_{\#}\mu$, that is to say:

$$u_t(y) = u(T_t^{-1}(y))|\det \mathrm{D}T_t^{-1}(y)|.$$

We then have:

$$\begin{aligned}
\int |u_t(y)|^q \, \mathrm{d}y &= \sum_{j=1}^k \int \left| u\left(\frac{y - ty_j}{1-t}\right)\right|^q \frac{1}{(1-t)^{Nq}} \, \mathrm{d}y \\
&= \sum_{j=1}^k \int |u(z)|^q (1-t)^{N(1-q)} \, \mathrm{d}z \\
&= (1-t)^{N(1-q)} \int |u(z)|^q \, \mathrm{d}z.
\end{aligned}$$

Moreover, thanks to Remark 3.3.9, the metric derivative $|\gamma'|(t)$ is constantly equal to the Wasserstein distance $W_p(\mu, \nu)$. Then,

$$\mathcal{F}_q(\gamma) = W_p(\mu, \nu) \int_0^1 \int |u_t(y)|^q \, \mathrm{d}y \, \mathrm{d}t = \frac{W_p(\mu,\nu)}{N+1-Nq} \int_\Omega |u|^q \, \mathrm{d}x$$

which is finite since $q < 1 + 1/N$. □

Theorem 3.3.11. *Let $q < 1 + 1/N$. Then every couple of measures can be joined by a path with finite energy.*

Proof. It is enough to link any measure ν to a fixed L^q measure μ (for instance, the normalized Lebesgue measure) with a finite energy path. Let

$\{\nu_k\}_{k\in\mathbb{N}}$ be a sequence of atomic measures approximating ν in the Wasserstein distance W_p. By Lemma 3.3.10, for every k there is a path γ_k with energy

$$\mathcal{F}_q(\gamma_k) = CW_p(\mu, \nu_k)$$

where C is a constant which only depends on N, q, Ω (and of course μ). Extracting a convergent subsequence of $\{\gamma_k\}_{k\in\mathbb{N}}$ provides a path γ such that, by repeating the lower semicontinuity argument of Theorem 3.2.1, gives

$$\mathcal{F}_q(\gamma) \le \liminf_{k\to+\infty} \mathcal{F}_q(\gamma_k) = \lim_{k\to+\infty} CW_p(\mu, \nu_k) = CW_p(\mu, \nu).$$

Since γ_k connects μ to ν_k, then γ connects μ to ν and the result is established. □

As in the previous section, we show that the previous result is sharp, as it can be seen from the following statement which is valid in a more general setting. In fact, we prove an estimate which holds for every W_p-Lipschitz curve not only valued in $\mathcal{P}(\Omega)$, but also in $\mathcal{P}(\mathbb{R}^N)$.

Theorem 3.3.12. *Suppose $q \ge 1 + 1/N$. Then there exists $\mu \in \mathcal{W}_p(\Omega)$ such that every non-constant W_p-Lipschitz path γ with $\gamma(0) = \mu$ gives $\mathcal{F}_q(\gamma) = +\infty$.*

Proof. Let us choose $\mu = \delta_0$ (supposing, up to a translation, that $0 \in \Omega$). It is sufficient to prove that

$$\inf\{F_q(\nu) \,|\, \nu \in \mathcal{P}(\Omega),\, W_p(\mu, \nu) \le t\} \ge Ct^{-N(q-1)}, \tag{3.3.6}$$

with $C > 0$. In fact, by reparametrization, it is sufficient to prove that the functional is infinite on constant speed paths. Taken such a path γ, with constant speed $L > 0$, we then have

$$\mathcal{F}_q(\gamma) = L\int_0^1 F_q(\gamma(t))\mathrm{d}t \ge L\int_0^1 C(Lt)^{-N(q-1)}\mathrm{d}t = +\infty,$$

where the integral diverges thanks to the assumption on q. To prove (3.3.6) we can suppose that $\Omega = \mathbb{R}^N$, which is the worst case. This shows that the result does depend neither on the compactness nor on the convexity of Ω.

By considering the map that associates to every probability measure ρ the measure $\nu = (m_t)_\#\rho$, where $m_t(x) = tx$, one has a one-to-one correspondence between the probabilities whose Wasserstein distance from δ_0 is less than 1 and those whose distance is less than t. It is easy to see that ν is L^q if and only if the same happens for ρ and that the density of ν is the function $x \mapsto t^{-N}u(x/t)$, where u is the density of ρ. Therefore

$$F_q(\nu) = \int \frac{u^q(x/t)}{t^{Nq}}\mathrm{d}x = \int u^q(y)t^{-Nq}t^N\mathrm{d}y = F_q(\rho)t^{-N(q-1)}.$$

Consequently, it is now sufficient to evaluate the infimum in (3.3.6) when $t = 1$, and this number will be the constant C we are looking for. We will show that this infimum is in fact a minimum, thus obtaining that it is strictly positive. This problem is quite similar to those studied in [44]. To get the existence of a minimum we recall that the functional F_q is sequentially lower semicontinuous with respect to weak-$*$ topology on probability measures, while the set $\{\nu \in \mathcal{P}(\mathbb{R}^n) \,|\, W_p(\delta_0, \nu) \leq 1\}$ is sequentially compact with respect to the same topology (in fact every sequence in it turns out to be tight). $\square$

Remark 3.3.13. As in the previous case, it is possible that two measures could be connected by a finite energy path even when q is greater than $1 + 1/N$. For instance, with $N = 2$, the path given by

$$\gamma(t) = \frac{1}{4t}\mathbb{1}_{[-1,1]\times[-t,t]} \cdot \mathcal{L}^2$$

is a Lipschitz path in $\mathcal{W}_p([-1,1] \times [-1,1])$ joining $\gamma_0 = 1/2\mathcal{H}^1 \llcorner [-1,1]$ to $\gamma_1 = 1/4\mathcal{L}^2$ (it is in fact a Wasserstein geodesic between them). The energy is finite as far as

$$\int_0^1 \frac{4t}{(4t)^q}\mathrm{d}t < +\infty.$$

This condition is fulfilled when $1 - q > -1$, i.e. when $q < 2$, instead of the condition $q < 1 + 1/2$ found in Theorem 3.3.11.

3.3.2 Unbounded domains

The existence results of the previous section were based on two important facts: the compactness of Wasserstein spaces $\mathcal{W}_p(\Omega)$ when Ω itself is compact and $1 \leq p < +\infty$, and the estimate like $F_q \geq c > 0$, proven in Theorem 3.3.1, that can be obtained when $|\Omega| < +\infty$. Both the facts do not hold when $\Omega = \mathbb{R}^N$, for instance. This is the reason why we developed in Section 3.1 some tools giving the existence of optimal paths under weaker assumption, even in the abstract metric setting. To replace the compactness of Ω we need to use Theorem 3.2.4, while to deal with the fact that we do not have $F_q(\nu) \geq c > 0$ in the case where ν runs over all $\mathcal{W}_p(\mathbb{R}^N)$ we can use the weaker assumption given by hypothesis H2.

In this section we only deal with the case of $\mathcal{F}_q$-like functionals studied in the compact case in Section 3.3.1; the case of atomic measures and $\mathcal{G}_r$-like functionals of Section 3.3.1 still present some extra difficulties when Ω is unbounded. We stress the fact that most of the techniques we use can be adapted to deal with several different cases, i.e. Ω unbounded but not necessarily the whole space, or the space $\mathcal{W}_\infty(\Omega)$ (where the distance is given by transport costs computed in a supremal way instead of an integral one). Notice that the use of Theorem 3.2.4 is necessary because in general, if Ω is not compact, the corresponding Wasserstein spaces are not even locally compact (and the same happens when we take Ω compact but we choose to consider the space $\mathcal{W}_\infty(\Omega)$), thus we cannot have the compactness of closed balls.

First, we show some lemmas in order to use Theorem 3.2.4.

Lemma 3.3.14. *The weak topology (i.e. the one induced by the duality with the space $C_b(\Omega)$ of bounded continuous functions on Ω) on the space $\mathcal{W}_p(\Omega)$ can be metrized by a distance d' such that $d' \leq W_1 \leq W_p$.*

Proof. The usual distance metrizing the weak topology is given by

$$d(\mu,\nu) = \sum_{k=1}^{\infty} 2^{-k} \left| \int \phi_k \mathrm{d}(\mu-\nu) \right|,$$

where $(\phi_k)_k$ is a dense sequence in the unit ball of $C_b(\Omega)$. We can choose these functions to be Lipschitz continuous and let, for every index k, c_k be the Lipschitz constant of ϕ_k. Then

$$d'(\mu,\nu) = \sum_{k=1}^{\infty} \frac{2^{-k}}{1+c_k} \left| \int \phi_k \mathrm{d}(\mu-\nu) \right|$$

is a distance which metrizes the same topology. Being $\phi_k/(1+c_k)$ a $1-$Lipschitz function, thanks to the dual formulation of Monge's problem, we have

$$\left| \int \frac{\phi_k}{1+c_k} \mathrm{d}(\mu-\nu) \right| \leq W_1(\mu,\nu),$$

and so, by summing up on k, we get $d' \leq W_1$ as required. □

The following two lemmas are well known.

Lemma 3.3.15. *The distance W_p is lower semicontinuous on $\mathcal{W}_p(\Omega) \times \mathcal{W}_p(\Omega)$ endowed with the weak $\times$ weak convergence.*

Proof. Take $\mu_n \rightharpoonup \mu$ and $\nu_n \rightharpoonup \nu$. Let γ_n be an optimal transport plan for the cost $|x-y|^p$ between μ_n and ν_n: the sequence of this plans turns out to be tight thanks to tightness of the sequence of the marginal measures, and so we may suppose $\gamma_n \rightharpoonup \gamma$. We can now see that γ is a transport plan between μ and ν and so it holds

$$\begin{aligned} W_p(\mu,\nu) &\leq \left(\int |x-y|^p \mathrm{d}\gamma\right)^{1/p} \\ &\leq \liminf_{n\to+\infty} \left(\int |x-y|^p \mathrm{d}\gamma_n\right)^{1/p} = \liminf_{n\to+\infty} W_p(\mu_n,\nu_n). \quad \square \end{aligned}$$

Lemma 3.3.16. *All bounded sets in $\mathcal{W}_p(\mathbb{R}^N)$ are relatively compact with respect to weak topology.*

Proof. Just notice that, in a bounded set, every sequence of probability measures turns out to be tight. The limits up to subsequences (that exist in the weak sense) still belong to the space $\mathcal{W}_p(\mathbb{R}^N)$ as a consequence of the lower semicontinuity of the functional $\mu \mapsto W_p(\mu,\delta_0)$ (which is nothing but the $p-$th momentum of the measure). $\square$

We can give now our result.

Theorem 3.3.17. *Let F_q and $\mathcal{F}_q$ be defined as in Section 3.3.1 respectively on $\mathcal{W}_p(\mathbb{R}^N)$ and on the set of Lipschitz path in $\mathcal{W}_p(\mathbb{R}^N)$ joining two measures μ_0 and μ_1. Then*

- *if $q < 1+1/N$ for every μ_0 and μ_1 there exists a path giving finite and minimal value to $\mathcal{F}_q$;*
- *if $q \geq 1+1/N$ there exist measures μ_0 such that $\mathcal{F}_q = +\infty$ on every non-constant path starting from μ_0.*

Proof. Let us start by the case $q < 1+1/N$: thanks to Lemma 3.3.15 and 3.3.16 we can use Theorem 3.2.4 and so we just need to verify the two assumptions H1 and H2. The existence of a finite-energy path can be achieved in the same way as in Theorem 3.3.11, by passing through a fixed L^q probability measure. Notice that, in order to have the convergence of a subsequence and the lower semicontinuity in the approximation by atomic measures, we will argue as in the proof of Theorem 3.2.4 instead of Theorem 3.2.1.

In order to estimate the integral in H2 we will use the same estimate given in Theorem 3.3.12, to achieve

$$\inf \{F_q(\nu) \mid \nu \in \mathcal{P}(\Omega),\, W_p(\mu, \nu) \le t\} \ge Ct^{-N(q-1)},$$

so that the integral diverges as far as $q < 1 + 1/N$.

By repeating the arguments of Theorem 3.3.12, we can then prove also the second part of our result, because $\mu = \delta_0$ cannot be joined to any other probability measure by a finite energy path. □

Remark 3.3.18. In the previous theorem we did not mention the possibility to link, for arbitrary $q > 1$, two measures $\mu_0, \mu_1 \in L^q(\mathbb{R}^N)$. It is easy to check that the same construction used in Theorem 3.3.8 can be used in this setting too. We get in such a way the existence of a path providing a finite value to $\mathcal{F}_q$, but some problems arise when we look for a minimal one. In fact, for arbitrary q, condition H2 is no longer fulfilled and this prevents us from applying the general existence results.

To conclude this section, we highlight the difference between the case we dealt with (about the functional $\mathcal{F}_q$) and the other important case, represented by the functional $\mathcal{G}_r$. In this latter case it is not necessary to pass through the divergence of the integral in assumption H2, because we actually have $G_r \ge 1$, as already shown.

On the other hand, some difficulties arise in verifying assumption H1. In fact the construction we made to build a finite energy path linking δ_0 to a probability measure μ strongly uses the compactness of the support of μ. In order to get a similar construction for the case $\Omega = \mathbb{R}^N$ we would need an estimate like

$$\inf\{W_p(\mu, \nu) \mid \#\operatorname{spt}(\nu) \le k\} \le C(\mu)k^{-1/N},$$

where $C(\mu)$ is a finite constant depending on the measure μ. It is easy to get a similar estimate when μ has compact support, but the constant may depend on the diameter of its support. The existence of a similar estimate for arbitrary measures μ is linked to the asymptotics of the rescaled location problem in $\mathbb{R}^N$.

A theory on this asymptotic problem has been explicitly developed (for instance in [12]) only in the case of compact support. However, it leads to a condition like $\mu^{N/(N+p)} \in L^1$, which is always fulfilled for μ compactly supported, while it may fail for general probability measures in $\mathcal{W}_p(\mathbb{R}^N)$. A positive answer to this estimate would easily imply a theorem similar to Theorem 3.3.17 for the case of $\mathcal{G}_r$, but we cannot say this to be necessary.

3.4 An alternative model for tree structures

Path functionals are not the only alternative model of optimal transportation. In the past few years two models (formulated in a mathematical different way which turns out to be the same) have been proposed in various papers. The first one deals with *transport paths* and was proposed by Xia in [48]. The second one deals with *irrigation trees* and *traffic plans* and appeared in [34] and [7].

Interesting papers concerning the possibility of computing the optima of these two models can be found in [28], [33] and [50]. In particular in [28] a construction with ruler and compass for the nodes of an optimal graph is provided, while in [50] an algorithm to optimize a graph with a given topology is presented.

These model came out from the study of drainage networks, plants, trees and their root systems, bronchial and cardiovascular systems. These models seem also to be suitable to study the formation of a tree leaf (see [49]) or the shaping of a land due to the flow of its rivers (see [45]).

In the next few pages we will describe the main properties of these two models.

3.4.1 Xia's *transport paths*

The first thing we want to define is how *atomic* two probability measures are transported. So, we fix a compact subset $\Omega \subset \mathbb{R}^N$ and consider as initial and final measures μ^+ and μ^- convex combinations of Dirac measures in Ω:

$$\mu^+ = \sum_{i=1}^{m} a_i \delta_{x_i}, \quad \mu^- = \sum_{j=1}^{n} b_j \delta_{y_j},$$

with $a_i, b_j \geq 0$, $\sum_{i=1}^m a_i = \sum_{j=1}^n b_j = 1$.

Let us consider a graph G (we respectively indicate the set of vertexes and the set of the edges of G by $V(G)$ and $E(G)$) such that each edge is oriented and is carried by a positive real number (i.e. a weight function $w : E(G) \to]0, +\infty[$ is given). For such a graph to be a *transport path* between μ^+ and μ^- we require that:

- $x_i \in V(G)$ for all $1 \leq i \leq m$, $y_j \in V(G)$ for all $1 \leq j \leq n$;
- for each initial vertex x_i the sum of the weights of the edges coming out from x_i is equal to a_i, the mass carried by x_i. Denoting by e^+ the starting point of the oriented edge $e \in E(G)$, we must have
$$a_i = \sum_{\substack{e \in E(G), \\ e^+ = x_i}} w(e);$$
- for each final vertex y_j the sum of the weights of the edges coming in y_j is equal to b_j, the mass carried by y_j. Denoting by e^- the ending point of the oriented edge $e \in E(G)$, we must have
$$b_j = \sum_{\substack{e \in E(G), \\ e^- = y_j}} w(e);$$
- for any interior vertex $v \in V(G) \setminus \{x_1, \ldots, x_m, y_1 \ldots, y_n\}$ the Kirchoff's Law must hold:
$$\sum_{\substack{e \in E(G), \\ e^+ = v}} w(e) = \sum_{\substack{e \in E(G), \\ e^- = v}} w(e).$$

The idea under the transport paths is very simple. At the beginning we have the initial measure μ^+. Then the mass starts to flow inside the edges of the graph G until it comes to the final measure μ^-. The conditions above simply guarantee that no mass is created or disappears when it is split in two or more edges.

The point is now to provide to each transport path G a suitable cost that makes keeping the mass together cheaper. The right cost function is then
$$M^\alpha(G) := \sum_{e \in E(G)} [w(e)]^\alpha l(e), \tag{3.4.1}$$
where $l(e)$ is the length of the edge e and $0 \leq \alpha \leq 1$ is fixed. This cost takes advantage of the subadditivity of the function $t \mapsto t^\alpha$ in order to make more economic the tree-shaped graphs.

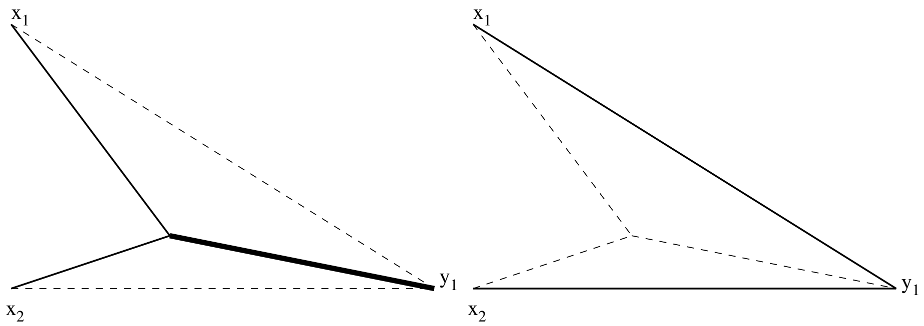

Figure 3.3: Y shape versus V shape.

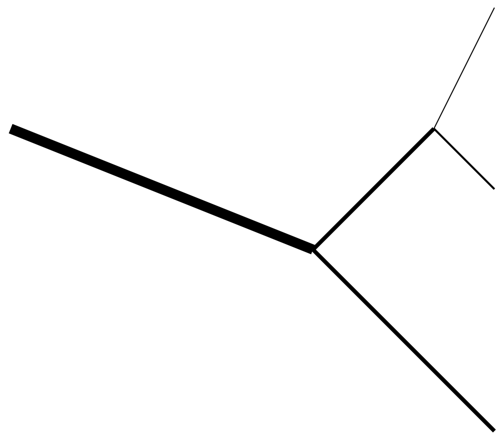

Figure 3.4: A more general tree.

Example 3.4.1. Let $\mu^+ = m\delta_{x_1} + (1-m)\delta_{x_2}$ and $\mu^- = \delta_{y_1}$. Thanks to the subadditivity of $t \mapsto t^\alpha$ it may happen (depending on the mass m carried by x_1, the value of the parameter α and the positions of the points x_1, x_2 and y_1) that a Y-shaped graph will be more efficient that a V-shaped one in the case of Figure 3.3. In the general case, an optimal graph will look like that of Figure 3.4.

To deal with the case of general initial and final measures (i.e. Borel probability measures) we must write the four conditions defining a transport path in simplified and more compact form. The only fact we have to note is that to each oriented edge e of a graph we can associate the vector measure given by $\mu_e = (\mathcal{H}^1 \llcorner e)\hat{e}$ ($\hat{e}$ is the unit vector with the same orientation as e) and that

$$\operatorname{div} \mu_e = \delta_{e^+} - \delta_{e^-}$$

in distributional sense. In this way, instead a weighted directed graph we can consider the vectorial measure given by

$$G = \sum_{e \in E(G)} \mu_e$$

and the four conditions are summarized simply requiring that $\operatorname{div} G = \mu^+ - \mu^-$. In this case it is easy to see that $M^1(G) = \|G\|(X)$.

In the general case a *transport path* between measures μ^+ and μ^- is a vectorial measure T such that:

- there exists a sequence of atomic probability measures $\{\mu_i^+\}_{i\geq 1}$ such that $\mu_i^+ \rightharpoonup \mu^+$;
- there exists a sequence of atomic probability measures $\{\mu_i^-\}_{i\geq 1}$ such that $\mu_i^- \rightharpoonup \mu^-$;
- there exists a sequence of transport paths $\{T_i\}_{i\geq 1}$ between μ_i^+ and μ_i^- such that $T_i \rightharpoonup T$.

The M^α cost for a generalized transport path is then defined as the lower semicontinuous envelope of M^α as defined on graphs in Definition 3.4.1:

$$M^\alpha(T) = \inf_{\{T_i\}_{i\geq 1}} \liminf_{i\to+\infty} M^\alpha(T_i), \tag{3.4.2}$$

where $\{T_i\}_{i\geq 1}$ ranges in the set of transport paths satisfying the three conditions above.

The main result regarding the existence of an optimal transport path is quite analogous to Theorem 3.3.4. In fact it can be shown that if $1 - 1/N < \alpha \leq 1$, then the existence of an optimal transport path is assured whatever the initial and final measures are. Conditions to link arbitrary measures have been found in [23]. Moreover, many questions about the regularity of an optimal transport path are open.

Remark 3.4.2. Let $\mu^+ = m\delta_{x_1} + (1-m)\delta_{x_2}$ and $\mu^- = \delta_{y_1}$. With a simple minimization of a function of two real variables it is easy to see that the position x of the node of an optimal Y-shaped graph between μ^+ and μ^- is given by:

$$m^\alpha \frac{x - x_1}{|x - x_1|} + (1-m)^\alpha \frac{x - x_2}{|x - x_2|} + \frac{x - y_1}{|x - y_1|} = 0. \tag{3.4.3}$$

Consider now the same problem for the path functional $\mathcal{G}_r$ (note that the parameter r corresponds exactly to α). With another simple minimization it is easy to see that if we consider an optimal curve between μ^+ and μ^- the masses will join in a point x satisfying (in the case $p = 1$, but a formula can be easily obtained for a general $p > 1$)

$$[m^r + (1-m)^r] \left(m \frac{x - x_1}{|x - x_1|} + (1-m) \frac{x - x_2}{|x - x_2|} \right) + \frac{x - y_1}{|x - y_1|} = 0. \quad (3.4.4)$$

Equalities (3.4.3) and (3.4.4) do not provide in general the same point, so in the general case the minimizers will "look" different.

3.4.2 Maddalena, Morel, Solimini's *irrigation trees*

The following approach is based on a kind of fluidodynamics approach. The idea is to consider a set Ω and a set "fibres" coming out from a given point S. The set of fibres will be obtained associating to each $\omega \in \Omega$ a curve $\chi_\omega : [0, +\infty] \to \Omega$ such that $\chi_\omega(0) = S$. Since all the fibres start from the same point S in this model the initial measure cannot be other than δ_S. Then, as time passes, the fibres separate one from each other until they reach the final measure.

To be more clear, let $(\Omega, \mathcal{M}, \mu)$ be a probability space representing the reference configuration of a fluid incompressible material body. Let $S \in \mathbb{R}^N$ be a given point of the Euclidean space of dimension N.

Definition 3.4.3 (Set of fibres). A *set of fibres of* Ω *with source* S is a mapping

$$\chi : \Omega \times \mathbb{R}_+ \to \mathbb{R}^N$$

such that:

- for μ-a.e. $\omega \in \Omega$, the curve given by χ_ω

$$t \mapsto \chi_\omega(t) := \chi(\omega, t)$$

  is Lipschitz continuous and $\mathrm{Lip}(\chi_\omega) \le 1$;

- for μ-almost-every $\omega \in \Omega$, $\chi_\omega(0) = S$.

We will denote by $\mathbf{C}_S(\Omega)$ the set of such functions.

Definition 3.4.4 (χ-vessels at time t). Given $t \in \mathbb{R}_+$, the χ-vessels at time t will be the equivalence classes of the equivalence relation defined by:

$$\omega_1 \simeq_t \omega_2 \iff \chi_{\omega_1} = \chi_{\omega_2} \text{ on } [0,t].$$

In simpler words, a χ-vessel at a certain time instant t identifies all the flux lines which build a certain tube at time t. As time goes on, the fibres separate and the number of χ-vessels increases.

Definition 3.4.5 (Absorption time). Given $\chi \in \mathbf{C}_S(\Omega)$, the function $\sigma_\chi : \Omega \to \mathbb{R}_+$ given by

$$\sigma_\chi(\omega) := \inf\{t \in \mathbb{R}_+ \ : \ \chi_\omega \text{ constant on } [t,+\infty]\}$$

is the *absorption time*. A point $\omega \in \Omega$ is *absorbed* if $\sigma(\omega) < +\infty$, while it is *absorbed at time* t if $\sigma(\omega) \le t$. We will denote by $A_t(\chi)$ the set of absorbed points at time t:

$$A_t(\chi) := \{\omega \in \Omega \ : \ \sigma_\chi(\omega) \le t\},$$

and by $M_t(\chi)$ its complementary:

$$M_t(\chi) := \Omega \setminus A_t(\chi) = \{\omega \in \Omega \ : \ \sigma_\chi(\omega) > t\}.$$

The *irrigation measure* induced by the set of fibres χ is the measure $\mu_\chi = i_{\chi\#}\mu$, where i_χ is defined by $i_\chi(\omega) := \chi(\omega, \sigma_\chi(\omega))$. This is the measures that is reached from δ_S through the set of fibres χ.

Definition 3.4.6 (MMS functional). Let $\alpha \in [0,1]$. The *MMS functional* is defined by

$$MMS(\chi) := \int_0^{+\infty} \left[\int_{M_t(\chi)} [\mu([\omega]_t)]^{\alpha-1} \, \mathrm{d}\mu(\omega) \right] \mathrm{d}t.$$

In [34] it is proved that in a suitable subset of $\mathbf{C}_S(\Omega)$ functionals of the type

$$\chi \mapsto MMS(\chi) + F(\mu_\chi),$$

where F is a lower semicontinuous functional defined on positive measures, admit a minimizer. A case of particular interest is then that of a functional F finite on a certain measure μ^- and infinity elsewhere. In this case a comparison between the model of Xia can be made, and the two models come out to be the same (see [34]).

3.4.3 Bernot, Caselles and Morel's *traffic plans*

Let X be a compact subset of $\mathbb{R}^N$. Consider the space $K := \mathrm{Lip}_1(\mathbb{R}_+, X)$ of 1-Lipschitz maps $\gamma : \mathbb{R}_+ \to X$ endowed by the distance

$$d(\gamma_1, \gamma_2) := \sup_{k \geq 1} \left[\frac{1}{k} \|\gamma_1 - \gamma_2\|_{L^\infty[0,k]} \right].$$

It is easy to check that, via Ascoli-Arzelà Theorem, $K = \mathrm{Lip}_1(\mathbb{R}_+, X)$ is a compact metric space. The metric d endows K with a topology, so we can consider the Borel σ-algebra $\mathcal{B}$ on K.

One point is to define the *stopping time* of a curve of the set K. It will be the last time instant before the curve becomes constant.

Definition 3.4.7 (Stopping time of a curve γ). Given a curve $\gamma \in K = \mathrm{Lip}_1(\mathbb{R}_+, X)$, its *stopping time* T is defined by

$$T(\gamma) := \inf\{t \in \mathbb{R}_+ \ : \ \gamma_{|[t,+\infty[} \text{ is constant}\}. \tag{3.4.5}$$

The function T given in Definition 3.4.7 is lower semicontinuous (and, in particular, measurable).

Definition 3.4.8 (Traffic plans). A *traffic plan* is a probability measure on the space $(K = \mathrm{Lip}_1(\mathbb{R}_+, X), \mathcal{B})$ such that

$$\int_K T(\gamma) \,\mathrm{d}\mu(\gamma) < +\infty. \tag{3.4.6}$$

The set of traffic plans on X will be denoted by $TP(X)$.

In other words, the integral appearing in Definition 3.4.8 is the transportation time which is requested not to be infinite.

Definition 3.4.9 (Transport plan associated to a traffic plan). Given a traffic plan μ, the transport plan associated to it is the measure acting on the functions $\phi \in \mathcal{C}(X \times X)$ as

$$\langle \pi_\mu, \phi \rangle := \int_K \phi(\gamma(0), \gamma(T(\gamma))) \,\mathrm{d}\mu(\gamma).$$

The subset of $TP(X)$ of traffic plans μ such that $\pi_\mu = \pi$, π given, is denoted by $TP(X, \pi)$.

Since, formally,

$$\begin{aligned}\langle \pi_\mu, \chi_{A\times B}\rangle &:= \int_K \chi_{A\times B}(\gamma(0), \gamma(T(\gamma)))\ \mathrm{d}\mu(\gamma)\\ &= \mu(\{\gamma \in K\ :\ \gamma(0) \in A, \gamma(T(\gamma)) \in B\}),\end{aligned}$$

$\pi_\mu(A \times B)$ is the amount of mass which is moved from A to B.

Definition 3.4.10 (Irrigating and irrigated measure). If μ is a traffic plan, we define the *irrigating measure* μ^+ and the *irrigated measure* μ^- as

$$\langle \mu^+, \phi_1\rangle := \langle \pi_\mu, \phi_1 \circ \pi_+\rangle, \quad \langle \mu^-, \phi_2\rangle := \langle \pi_\mu, \phi_2 \circ \pi_-\rangle.$$

where $\phi_1, \phi_2 \in \mathcal{C}(X)$. The set of traffic plans with prescribed irrigating and irrigated measure (say ν^+ and ν^-) will be denoted by $TP(\nu^+, \nu^-)$.

In the next definition we consider the measure space $(\Omega = [0,1], \mathcal{B}(\Omega), \lambda = \mathcal{L}^1 \llcorner [0,1])$.

Definition 3.4.11 (Parametrization of a traffic plan). Let μ be a probability measure on K. A *parametrization* of μ is a measurable application $\chi : \Omega \to K$ such that $\chi_{\#}\lambda = \mu$.

Definition 3.4.12 (Multiplicity and path class). Let μ be a traffic plan. We call *multiplicity* of μ at a point $x \in \mathbb{R}^N$ the number

$$|x|_\mu := \mu(\{\gamma\ :\ \exists t \in \mathbb{R}_+, \gamma(t) = x\}).$$

Given a parametrization of μ, we define the *path class* of $x \in \mathbb{R}^N$ as the set

$$[x]_\chi := \{\omega\ :\ \exists t \in \mathbb{R}_+, \chi(\omega)(t) = x\}.$$

Since $\chi_{\#}\lambda = \mu$, we have that $|[x]_\chi| = |x|_\mu$.

Given a parametrization χ of μ, the energy $E(\mu)$ of a traffic plan μ will be given by Definition 3.4.13.

Definition 3.4.13 (BCM functional). *BCM functional* is defined by

$$BCM(\chi) := \int_\Omega \left[\int_{\mathbb{R}_+} [\mu([\omega]_t)]^{\alpha-1}]|\dot{\chi}_\omega(t)|\ \mathrm{d}t\right] \mathrm{d}\mu(\omega). \tag{3.4.7}$$

The idea of traffic plan comes out from the necessity to track the movement of the mass. In fact, a traffic plan is a measure on a set of curves along which the mass moves. Roughly speaking, the way through the mass moves is then established by the traffic plan: if no mass is carried on a certain set of curves, then the mass will not go along that set of curves. The next theorem proves the existence of a minimum for the BCM functional 3.4.7 whenever we fix the irrigating and irrigated measure (i.e. we need to transport two given distribution of masses ν^+ and ν^-), or a transport plan π is given (i.e. we need to transport two given distribution of masses, but we already know that an amount of mass given by $\pi(A \times B)$ of the initial mass placed in a certain set A will have to be placed in a certain set B).

Proposition 3.4.14 (Existence of minimizers). *The following facts are true:*

- *given an irrigating measure ν^+ and an irrigated measures ν^-, the functional BCM admits a minimizer in $TP(\nu^+, \nu^-)$;*
- *given a transport plan π, the functional BCM admits a minimizer in $TP(\pi)$.*

Definition 3.4.15 (Loop free traffic plan). A traffic plan μ is *loop free* if there is a parametrization χ of μ such that for almost all $\omega \in \Omega$, the curve $\chi(\omega)$ is injective.

Definition 3.4.16 (Geometric embedding of a traffic plan). Let μ be a traffic plan. We define the *geometric embedding* as

$$G_\mu := \{x \in \Omega \ : \ |x|_\mu \neq 0\}.$$

Proposition 3.4.17. *Let μ be a traffic plan such that $E(\mu) < +\infty$. Then, there exists a loop free traffic plan $\tilde{\mu}$ such that $G_{\tilde{\mu}} \subseteq G_\mu$ and $\pi_{\tilde{\mu}} = \pi_\mu$.*

The following theorem relates the two functional 3.4.6 and 3.4.13.

Theorem 3.4.18 (MMS-BCM comparison). *Let χ be a parametrization of a non trivial traffic plan μ with finite energy. Then, $BCM(\chi) \geq MMS(\chi)$; moreover, $BCM(\chi) = MMS(\chi)$, if χ is loop free.*

Appendix A

Polish spaces and measure theory

A.1 Polish spaces

The aim of this appendix is just to recall some results on measure theory and Polish spaces and to give some references on these subjects for the interested reader.

Let X be a topological space. A *Borel probability measure* on X is a positive measure of unitary total mass defined on the Borel σ-algebra, that is the smallest σ-algebra which contains all open sets of X. A *Polish space* is a separable topological space such that its topology is induced by a complete metric. It is easy to see that if X and Y are Polish spaces, then $X \times Y$ is a Polish space.

A useful result is that Borel probability measures on a Polish space are always regular.

Proposition A.1.1. *Let μ be a Borel probability measures on a Polish space X. Then μ is regular, that is for any Borel set B the following equalities are true:*

$$\mu(B) = \sup\{\mu(K) \ : \ K \ compact, \ K \subseteq A\},$$
$$\mu(B) = \inf\{\mu(U) \ : \ U \ open, \ A \subseteq U\}.$$

The next result is known as *Ulam's Lemma.*

Theorem A.1.2 (Ulam's Lemma). *A probability measure μ on a Polish space is concentrated on a σ-compact subset, that is there exist countably many compact subsets K_n, $n \in \mathbb{N}$, such that $\mu(\cup_n K_n) = 1$.*

In other words, the content of Ulam's Lemma is that a Borel probability measure on a Polish space is *tight*, that is for every $\varepsilon > 0$ there exists a compact subset K_ε such that $\mu(X \setminus K_\varepsilon) < \varepsilon$.

Another useful property of Polish spaces is the following theorem (known as Prokhorov's Theorem).

Definition A.1.3 (Tight set of Borel probability measures). A set S of Borel probability measures on a topological space X is *tight* if for every $\varepsilon > 0$ there exists a compact subset K_ε such that

$$\sup_{\mu \in S} \mu(X \setminus K_\varepsilon) < \varepsilon.$$

Theorem A.1.4 (Prokhorov's Theorem). *Let S be a tight set of Borel probability measures on a Polish space X. Then S is relatively sequentially compact with respect to the weak convergence, that is given a sequence of $\{\mu_n\}_{n\in\mathbb{N}}$ in S there exists a Borel probability measure μ such that for a suitable subsequence μ_{n_k} we have*

$$\lim_{k\to+\infty} \int_X \varphi \, \mathrm{d}\mu_{n_k} = \int_X \varphi \, \mathrm{d}\mu$$

for every $\varphi \in \mathcal{C}_b(X)$.

A.2 Disintegration of measures

Let X and Y be locally compact and separable metric spaces. Let us consider a map $Y \to [\mathcal{M}(X)]^m$ which we will denote by $y \mapsto \lambda_y$. By definition, $y \mapsto \lambda_y$ is a *Borel map* if for any open subset A of X the map $Y \to \mathbb{R}^m$ given by $y \mapsto \lambda_y(A)$ is a Borel map in the usual sense.

Recall that, given a set X, a *Dynkin class* $\mathcal{D}$ is a class of subsets of X such that $X \in \mathcal{D}$, $\mathcal{D}$ is closed under the union of an increasing sequence, $A \setminus B \in \mathcal{D}$ whenever $A, B \in \mathcal{D}$ and $B \subseteq A$. According to Dynkin Lemma (whose proof can be found in [8] or [37]), if $\mathcal{D}$ contains a class closed under finite intersection, then it contains the σ-algebra generated by it. Recall also that a vector space $\mathcal{V}$ of real functions defined on a measurable space is said to be a *monotone vector space* if the point-wise limit of a sequence of functions in $\mathcal{V}$ bounded by above by a function in $\mathcal{V}$ is still in $\mathcal{V}$. Again, according to [8] and [37], a monotone vector space containing all constant functions and characteristic functions of the sets in a basis (a *basis* is a subset which generates the σ-algebra and is closed under finite intersection) contains all bounded measurable functions.

Then, the following result is true.

Proposition A.2.1. *Let X and Y be locally compact and separable metric spaces. Let us consider a map $Y \to [\mathcal{M}(X)]^m$ which we will denote by $y \mapsto \lambda_y$. Then, $y \mapsto \lambda_y$ is Borel if and only if $y \mapsto \lambda_y(A)$ is Borel (in the usual sense) for every Borel set A. Moreover, the map*

$$y \mapsto \int_X \varphi(x, y) \, \mathrm{d}\lambda_y(x)$$

is Borel for any bounded Borel function $\varphi : X \times Y \to \mathbb{R}$.

Proof. For the first statement it is sufficient to apply Dynkin Lemma to the class of Borel sets $A \subseteq X$ such that $y \mapsto \lambda_y(A)$ is a Borel function and to note that the class of open subsets is closed under finite intersection. For the second it is sufficient to consider the monotone vector space of the functions φ such that

$$y \mapsto \int_X \varphi(x, y) \, \mathrm{d}\lambda_y(x)$$

is Borel and to note that it contains constant functions and characteristics of Borel rectangles. $\square$

The proof of the following result can be found on [22]

Theorem A.2.2 (Disintegration of measures). *Let X and Y be locally compact and separable metric spaces and let $\pi : X \to Y$ be a Borel map. Let $\lambda \in [\mathcal{M}(X)]^m$ and $\mu = \pi_\#|\lambda| \in \mathcal{M}_+(Y)$. Then, there exists a family of measures $\lambda_y \in [\mathcal{M}(X)]^m$ such that:*

- *$y \mapsto \lambda_y$ is a Borel map and $|\lambda_y|$ is a probability measure in X for μ-a.e. $y \in Y$;*
- *$\lambda = \int_Y \lambda_y \otimes \mu$, that is for every $A \in \mathcal{B}(X)$*

$$\lambda(A) = \int_Y \lambda_y(A) \, \mathrm{d}\mu(y);$$

- *$|\lambda_y|(X \setminus \pi^{-1}(y)) = 0$ for μ-a.e. $y \in Y$.*

Theorem A.2.2 will be useful in the following form.

Corollary A.2.3 (Disintegration of measures). *Let X and Y be locally compact and separable metric spaces and let $\pi : X \times Y \to Y$ be a Borel map. Let $\lambda \in [\mathcal{M}(X \times Y)]^m$ and $\mu = \pi_\#|\lambda| \in \mathcal{M}_+(Y)$. Then, there exists a family of measures $\lambda_y \in [\mathcal{M}(X \times Y)]^m$ such that:*

- *$y \mapsto \lambda_y$ is a Borel map and $|\lambda_y|$ is a probability measure in $X \times Y$ for μ-a.e. $y \in Y$;*
- *$\lambda = \int_Y \lambda_y \otimes \mu$, that is for every $A \in \mathcal{B}(X \times Y)$*
$$\lambda(A) = \int_Y \lambda_y(A) \; \mathrm{d}\mu(y);$$
- *$|\lambda_y|$ is concentrated on the set $X \times \{y\}$ for μ-a.e. $y \in Y$.*

Theorem A.2.4 (Uniqueness of the disintegration). *Let X, Y, π be as in Theorem A.2.2. Suppose that $\lambda \in \mathcal{M}_+(X), \mu \in \mathcal{M}_+(Y)$. Suppose that $y \mapsto \eta_y$ be a Borel function $Y \to \mathcal{M}_+(X)$ such that*

- *$\lambda = \int_Y \eta_y \otimes \mu$, that is*
$$\eta(A) = \int_Y \eta_y(A) \; \mathrm{d}\mu(y);$$
- *η_y is concentrated on $\pi^{-1}(y)$ for μ-a.e. $y \in Y$.*

Then the map $y \mapsto \eta_y$ is uniquely determined up to a set negligible with respect to μ.

A.3 Young measures

Assume that $\psi_n : X \to Y$ is a sequence of Borel maps between the compact metric spaces X, Y. We now consider the measures given by

$$\mu_{\psi_n} = (\mathrm{Id} \times \psi_n)_\# \mu_0 = \int \delta_{\psi_n(x)} \; \mathrm{d}\mu_0(x).$$

Assume also that $\mu_{\psi_n} \rightharpoonup \mu$. Since $\pi_{0\#}\mu_{\psi_n} = \mu_0$, we have also that $\pi_{0\#}\mu = \mu_0$. According to Theorem A.2.2, the measure μ can then be written as

$$\mu = \int \mu_n \otimes \mu_0$$

for a suitable Borel map of probability measures $x \mapsto \mu_x$ (which is referred to as the *Young limit* of the sequence ψ_h).

We will use the following result.

Theorem A.3.1. *Let $\mu \in \mathcal{M}_+(X \times Y)$ and set $\mu_0 = \pi_{0\#}\mu$. Let $\mu = \mu_x \otimes \mu_0$ be its disintegration. Then, if μ is not atomic we can find a sequence of Borel maps $\psi_n : X \to Y$ such that*

$$\mu = \mu_x \otimes \mu_0 = \lim_{n \to +\infty} \delta_{\psi_n(x)} \otimes \mu.$$

Moreover, the functions ψ_n can be chosen is such a way $\psi_{n\#}\mu$ is not atomic.

Bibliography

[1] L. Ambrosio, N. Gigli, and G. Savarè. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, volume 7 of *Lectures in Mathematics, ETH Zürich*. Birkhäuser, Basel, 2005.

[2] Luigi Ambrosio. *Lecture Notes on Optimal Transport Problems*, volume 1812 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2003. Lectures given at the C.I.M.-C.I.M.E. Joint Euro-Summer School on "Mathematical aspects of evolving interfaces" held in Funchal, July 3–9, 2000.

[3] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. The Clarendon Press Oxford University Press, New York, 2000.

[4] Luigi Ambrosio and Aldo Pratelli. Existence and stability results in the L^1 theory of optimal transportation. In *Optimal transportation and applications (Martina Franca, 2001)*, volume 1813 of *Lecture Notes in Math.*, pages 123–160. Springer, Berlin, 2003.

[5] Luigi Ambrosio and Paolo Tilli. *Topics on Analysis in Metric Spaces*, volume 25 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2004.

[6] M. Bernot, V. Caselles, and J. M. Morel. Are there infinite irrigation trees? Preprint available at `http://www.cmla.ens-cachan.fr/Cmla/`, 2003.

[7] Marc Bernot, Vicent Caselles, and Jean-Michel Morel. Traffic plans. *Publ. Mat.*, 49(2):417–451, 2005.

[8] Patrick Billingsley. *Convergence of probability measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[9] G. Bouchitté and G. Buttazzo. New Lower Semicontinuous Results for Nonconvex Functionals defined on Measures. *Nonlinear Analysis, Theory, Methods & Applications*, 15(7):679–692, 1990.

[10] G. Bouchitté and G. Buttazzo. Integral Representation of Nonconvex Functionals defined on Measures. *Ann. Inst. H. Poincaré Anal. Non Linéaire.*, 9(1):101–117, 1992.

[11] G. Bouchitté and G. Buttazzo. Relaxation for a Class of Nonconvex Functionals defined on Measures. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 10(3):345–361, 1993.

[12] G. Bouchitté, C. Jimenez, and M. Rajesh. Asymptotique d'un Problème de Positionnement Optimal. *C. R. Acad. Sci. Paris*, 335:853–858, 2002.

[13] A. Brancolini and G. Buttazzo. Optimal networks for mass transportation problems. *ESAIM Control Optim. Calc. Var.*, 11(1):88–101, 2005.

[14] A. Brancolini, G. Buttazzo, and F. Santambrogio. Path Functionals over Wasserstein Spaces. To appear on the Journal of the European Mathematical Society, available at `http://cvgmt.sns.it/`.

[15] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.

[16] Haïm Brezis. *Analyse fonctionnelle.* Collection Mathématiques Appliquées pour la Maîtrise. [Collection of Applied Mathematics for the Master's Degree]. Masson, Paris, 1983.

[17] G. Buttazzo. *Semicontinuity, Relaxation and Integral Representation in the Calculus of Variations*, volume 207 of *Pitman Research Notes in Mathematics Series.* Longman Scientific & Technical, Harlow, 1989.

[18] G. Buttazzo, E. Oudet, and E. Stepanov. Optimal Transportation Problems with Free Dirichlet Regions. *Progress in Nonlinear Differential Equations and their Applications*, 51:41–65, 2002. Proceedings in *"Variational Methods for Discontinuous Structures"*, Cernobbio 2001.

[19] G. Buttazzo and F. Santambrogio. A model for the Optimal Planning of an Urban Area. *SIAM J. Math. Anal.*, 37(2):514–530 (electronic), 2005.

[20] G. Buttazzo and E. Stepanov. Optimal Transportation Networks as Free Dirichlet Regions for the Monge-Kantorovich Problem. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.*, 2:631–678, 2003.

[21] G. Dal Maso and R. Toader. A Model for the Quasi-Static Growth of Brittle Fractures: Existence and Approximation Results. *Arch. Rational Mech. Anal.*, 162:101–135, 2002.

[22] C. Dellacherie and P. Meyer. *Probabilities and Potential*, volume 29 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1978.

[23] G. Devillanova and S. Solimini. On the Dimension of an Irrigable Measure. Preprint in preparation.

[24] Ivar Ekeland and Roger Temam. *Convex analysis and variational problems*. North-Holland Publishing Co., Amsterdam, 1976. Translated from the French, Studies in Mathematics and its Applications, Vol. 1.

[25] Lawrence C. Evans. Partial differential equations and Monge-Kantorovich mass transfer. In *Current developments in mathematics, 1997 (Cambridge, MA)*, pages 65–126. Int. Press, Boston, MA, 1999.

[26] K. J. Falconer. *The geometry of fractal sets*, volume 85 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1986.

[27] Wilfrid Gangbo and Robert J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.

[28] E. N. Gilbert. Minimum Cost Communication Networks. *Bell System Tech. J.*, 46:2209–2227, 1967.

[29] Casper Goffman and James Serrin. Sublinear functions of measures and variational integrals. *Duke Math. J.*, 31:159–178, 1964.

[30] L.V. Kantorovich. On the Transfer of Masses. *Dokl. Akad. Nauk. SSSR*, 37:227–229, 1942.

[31] L.V. Kantorovich. On a Problem of Monge. *Uspekhi Mat. Nauk.*, 3:225–226, 1948.

[32] M. Knott and C. Smith. On the optimal mapping of distributions. *J. Optim. Theory Appl.*, 43(1):39–49, 1984.

[33] D. H. Lee. Low Cost Drainage Networks. *Networks*, 6:351–371, 1976.

[34] F. Maddalena, J.M. Morel, and S. Solimini. A Variational Model of Irrigation Patterns. *Interfaces and Free Boundaries*, 5:391–415, 2003.

[35] R. J. McCann. A Convexity Principle for Interacting Gases. *Adv. Math.*, 128(1):153–159, 1997.

[36] G. Monge. Mémoire sur la Théorie des Déblais et des Remblais. *Histoire de l'Acad. des Sciences de Paris*, pages 666–704, 1781.

[37] N. Pintacuda. *Probabilità*. Zanichelli, 1994.

[38] A. Pratelli. On the Equality between Monge's infimum and Kantorovich's minimum in Optimal Mass Transportation. *Ann. Inst. H. Poincaré Probab. Statist.*, 2005.

[39] Svetlozar T. Rachev and Ludger Rüschendorf. *Mass Transportation Problems. Vol. I*. Springer-Verlag, New York, 1998. Theory.

[40] Svetlozar T. Rachev and Ludger Rüschendorf. *Mass Transportation Problems. Vol. II*. Springer-Verlag, New York, 1998. Applications.

[41] R. Tyrrell Rockafellar. *Convex analysis*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997. Reprint of the 1970 original, Princeton Paperbacks.

[42] Walter Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, 1987.

[43] Ludger Rüschendorf. On c-optimal random variables. *Statist. Probab. Lett.*, 27(3):267–270, 1996.

[44] F. Santambrogio. Misure Ottime per Costi di Trasporto e Funzionali Locali, 2003. Tesi di Laurea. Available at `http://cvgmt.sns.it/`.

[45] F. Santambrogio. Optimal Channel Networks, Landscape Function and Branched Transport. Available at `http://cvgmt.sns.it/`, 2006.

[46] V. N. Sudakov. Geometric problems in the theory of infinite-dimensional probability distributions. *Proc. Steklov Inst. Math.*, 2(2):i–v, 1–178, 1979. Cover to cover translation of Trudy Mat. Inst. Steklov **141** (1976).

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

[48] Q. Xia. Optimal Paths Related to Transport Problems. *Comm. Cont. Math.*, 5(2):251–279, 2003.

[49] Q. Xia. The Formation of a Tree Leaf. Preprint, 2005.

[50] G. Xue, T. P. Lillys, and D. E. Dougherty. Computing the Minimum Cost Pipe Network Interconnecting one Sink and many Sources. *SIAM J. Optim.*, 10(1):22–42 (electronic), 1999.