Scuola Normale Superiore, Pisa
Classe di Scienze

Ph.D. Thesis

Existence of optimal transport maps and regularity of the transport density in mass transportation problems

Aldo Pratelli

Advisor: *prof. Luigi Ambrosio*

Year 2003

Contents

Introduction

0.1 Some hystory of the mass transport problem

1781–1999

"Étant donnés dans l'espace, deux volumes egaux entr'eux, & terminés chacune par une ou plusieurs surfaces courbes donnés; trouver dans le second volume le point où doit être transportée chaque molécule du premier, pour que la somme des produits des molecules multipliées chacune par l'espace parcouru soit un minimum" ([60]).

With these words, Gaspard Monge proposed in 1781 a problem that, in its various developments, is still arousing a deep interest among many people in various fields of mathematics and physics. His idea was to consider a given a pile of soil and an excavation of the same volume, and he wanted to find how to move the earth from the pile to the excavation minimizing the work done. We can formalize this problem in a modern terminology as follows: the data are two real functions defined on $\mathbb{R}^2$, g and h, which have to be thought as the height of the pile and the depth of the excavation. A way to move the mass is a function $t : \mathbb{R}^2 \to \mathbb{R}^2$, which will be referred to as *transport map*, and the fact that this is effectively a way to move the soil into the excavation can be expressed by the condition $t_\# g = h$, which means that $\int_{t^{-1}(A)} g(x)\, dx = \int_A h(x)\, dx$ for any Borel set $A \subseteq \mathbb{R}^2$. Since, according to Monge's formulation, the cost of moving a unit mass from the point x to the point y is the distance $|y - x|$, one can realize that the total cost for the transportation corresponding to the function $t : \mathbb{R}^2 \to \mathbb{R}^2$ is

$$\iint_{\mathbb{R}^2} |t(x) - x| g(x)\, dx. \tag{0.1}$$

The mass transportation problem consists then, given g and h, in finding the function t which minimizes the above cost among all the transport maps. Besides the formulation of the problem, it is also due to Monge the intuition that, if there is an optimal map t, then the direction of the vector $t(x) - x$ is given by the gradient $Du(x)$ of some function $u : \mathbb{R}^2 \to \mathbb{R}$.

After Monge's work, much timed passed before there were significant developments in the theory; one can remark that in the 1920's Appell ([7]) considered again the problem of the existence of such a function u, and he gave a formal proof of this existence writing down the Euler-Lagrange equation corresponding to the constraint $t_{\#}g = h$.

But the first true key ideas for studying the Monge's problem are due to Kantorovich ([53, 54]) in the 1940's: he generalized the setting of the problem considering, as ambient space, $\mathbb{R}^N$ instead of $\mathbb{R}^2$ (he tought also to a generic metric space X); moreover the initial and final positions of the mass were given by two general positive measures with the same total mass, f^+ and f^-, in place of the measures $g(x)\,dx$ and $h(x)\,dx$. The greatest generalization he proposed was to consider as admissible ways to move the mass all the measures γ defined in $\mathbb{R}^N \times \mathbb{R}^N$ admitting f^+ and f^- as marginals; each of these measures will be called *transport plan*. The meaning of this definition is to allow the splitting of masses; roughly speaking, consider the mass contained in a point x: according to Monge's formulation, it should be entirely moved in the point $t(x)$, while the Kantorovich's idea is to distribute it in $\mathbb{R}^N$ as desired, provided that the final distribution of the points results to be f^-; in practice, the new idea was to pass from a "deterministic" to a "stochastic" point of view. To any transport plan, generalizing (0.1), one can then associate the cost

$$\iint_{\mathbb{R}^N \times \mathbb{R}^N} |y - x|\, d\gamma(x, y); \tag{0.2}$$

note that a transport map t corresponds in an obvious sense to the transport plan $\gamma = (\mathrm{Id}, t)_{\#} f^+$, and moreover the cost is the same, so that the new problem is an extension of the first one. Since, at least in the Monge's setting, the infima of the costs of the transport maps and plans given by (0.1) and (0.2) are the same, to find a solution for the original problem one can try to solve the second one (which turns out to be much simpler) and then to select some particular optimal transport plan corresponding to a transport map (which is then an optimal transport map). Another fundamental contribution of Kantorovich was a correct proof of the existence of a map u determining the direction of the movement of the points, as foreseen by Monge and Appell: to be precise, he showed the existence of a $1-$Lipschitz function u such that the direction $y - x$ is determined by $u(x)$ for $\gamma-$almost all the pairs (x, y). To reach this result, he interpreted the new formulation of the transport problem as a linear programming problem with infinite dimension, and he showed that such a function $u : \mathbb{R}^N \to \mathbb{R}$ can be found as a maximizer of the functional

$$\int_{\mathbb{R}^N} u(x)\, d\big(f^+ - f^-\big)(x) \tag{0.3}$$

among the Lipschitz functions u with Lipschitz constant less than 1; in fact, the maximization of (0.3) and the minimization of (0.2) are some sort of "dual problems". The importance of Kantorovich's work is such that the function u is nowadays always referred to as *Kantorovich potential*; one can also recall that Kantorovich won the Nobel Prize in Economic Sciences in 1975 for his "contributions to the theory of optimum allocation of resources".

However, Kantorovich did not consider the problem to select, among the optimal transport plans, someone associated to a transport map, which then would be an optimal transport map solving Monge's original problem. The first attempt to go in this direction was due to Sudakov in 1979, who considered the mass transportation problem at the end of his book ([72]) about the decomposition of probability measures. His idea is quite easy to be explained: he first noticed that, given an optimal transport plan γ between f^+ and f^-, all the mass is moved along disjoint lines, which we will call *transport rays*; this happens because the mass must be moved along lines where the Kantorovich potential u (which is $1-$Lipschitz) is decreasing at rate 1, then two of these rays cannot cross.

Then, he noticed that the one-dimensional version of the problem is easy to solve, in fact if f^+ and f^- are two measures on $\mathbb{R}$ then there is always an optimal transport map, provided f^+ is non-atomic. Therefore, he suggested to decompose the measures f^+ and f^- on the different transport rays associated to the Kantorovich potential, to solve separately each of the consequent one-dimensional problems finding an optimal transport map on each ray, and finally to glue all the different maps, which are defined in disjoint sets, to obtain a single map, which he claimed to be an optimal transport map.

Sudakov also proposed to replace the Euclidean norm $|y-x|$ in (0.1) and (0.2) by any other norm $\|y-x\|$ –however, already Kantorovich had considered the general possibility to replace $|y-x|$ by a suitable function $c(x,y)$. Sudakov wrote that the same argument to show the existence of an optimal transport map should have worked for any norm: in this case, he guessed that the "transport rays" should have become cones where the norm is an affine function, and these cones are effectively lines only if the norm is strictly convex.

Unfortunately, Sudakov's proof was not formally correct, because he did not really consider the problem of the regularity of the rays in his decomposition, and in last years it has been shown with some counterexamples that his claims were not correct without some regularity assumption. However, his strategy was really powerful, and in fact all the existence proofs for the case of a cost $c(x,y)$ given by a norm (except the one by Evans and Gangbo we will mention later) follow in some way his construction.

In the years following Sudakov's work, there was a great development in studying the duality relationship we mentioned before between problems (0.2) and (0.3), even in the general case of a function $c(x,y)$ in place of $|y-x|$: a main ingredient was the extension of the notion of superdifferential for concave functions as proposed by Rockafellar ([65]), leading to the notions of $c-$concavity and $c-$superdifferential (see [55, 69, 68]).

A few years later, it came out an existence proof for the quadratic case, i.e. when the cost is given by $c(x,y) = |y-x|^2$, that was obtained independently by Knott and Smith ([56]) in 1984 and by Brenier ([16, 17]) in 1987; then, many generalizations of this result have been found for the case when $c(x,y)$ is a strictly convex function, see for example [47, 23] (where also the strictly concave case is considered). It is impressive the number of applications of the mass transportation problem in the strictly convex case, mainly in the quadratic case: for instance, one can give a look at the list given in [10]. A particular feature of the strictly convex problem is that, under suitable assumptions, there is a unique optimal transport plan, which turns out to correspond to an optimal transport map.

On the other hand, for a long time after Sudakov's work, there was not any progress in the study of the Euclidean case, i.e. when the cost is $c(x,y) = |y-x|$, therefore convex but not strictly convex, despite the fact that it was the situation closer to the intuition, and indeed the original idea of Monge; in fact, this case has probably been (and still it is) the most challenging one. One of the reasons of its difficulty is the fact that, due to the lack of strict convexity, the uniqueness of the optimal transport plan is never true, except in trivial situations, and moreover not all the optimal transport plans are actually transport maps (that is, they do not correspond to a transport map). Therefore, there is the additional trouble of selecting a particular optimal transport plan, which in fact comes from a map. Finally, in 1999 Evans and Gangbo ([39]) gave an existence proof for the Euclidean case: since the Sudakov proof is known to fail (see [4]), now we know (but it was still unknown in 1999) that this is actually the first correct proof of the existence of an optimal transport map in the original Monge's setting.

The construction of Evans and Gangbo used the approximation of the so-called $p-$laplacian: they considered the solutions u_p of the problem

$$-\text{div}\ \left(|Du_p|^{p-2}Du_p\right) = f^+ - f^-$$

and, passing to a limit for $p \to \infty$, they found a function u and a measure $\mu \ll \mathscr{L}$ with the property that

$$-\text{div}\ (\mu Du) = f^+ - f^- : \tag{0.4}$$

the function u was a Kantorovich potential, then an already well-known object, while the measure μ was the additional ingredient, that they used to recover enough information to move correctly the mass inside each transport ray. This measure (or, better, its density) was denoted *transport density* –see also [52].

2000-2003

The paper by Evans and Gangbo somehow restarted the interest about the transport problem in the non strictly convex case, and in last years there have been great developments and many stimulating generalizations came to the light.

Concerning the existence of optimal transport maps, the result has been strengthened a lot: in the work by Evans and Gangbo, it was considered the case when f^+ and f^- are absolutely continuous with respect to the Lebesgue measure $\mathscr{L}$ and their densities are Lipschitz and supported in disjoint compact sets. Afterwards, Caffarelli, Feldmann and McCann ([24]) and Trudinger and Wang ([73]) independently extended the result to the case when f^+ and f^- are only absolutely continuous and with compact supports. Then, Ambrosio ([1]) proved that it was sufficient the absolute continuity of the sole f^+, while f^- could be any measure. Finally, the compactness assumption on the supports of the measures was weakened ([2]). As anticipated by Sudakov, the interest has then moved in the direction of the generic norms in $\mathbb{R}^N$: the existence result is now known also in the case of a quasi-euclidean norm ([24, 2]), of a crystalline norm ([4]) and, in the particular case of $\mathbb{R}^2$, of any norm ([4]). At this moment, it is still open the case of a generic norm in $\mathbb{R}^N$, $N \geq 3$.

Another generalization is to consider a manifold as ambient space, as it was done in [43, 30, 62].

It has then also appeared the strong connection among different problems: we already mentioned the relationship between (0.2) and (0.3); moreover the equation (0.4) used by Evans and Gangbo had been independently considered and deeply generalized, in the context of the shape optimization problem, by Bouchitté, Buttazzo and Seppecher ([14, 15, 12]), therefore some relationship between these two problems became clear; more generally, the equation $-\mathrm{div}\ \nu = f^+ - f^-$ had been studied several years before by Iri ([51]) and Strang ([71]). Another problem which appeared to have correlations with the mass transportation was the evolution problem considered by Brenier ([18, 19, 20]). It turns out that, in all these connections, a main role is played by the transport density, which appears in each of the problems with different meanings (and it is called with different names).

The transport density has consequently been studied from many different points of view; it was first considered in [14] (see also [12, 13]), then its uniqueness under suitable hypotheses has been established in [1, 42]. Later, its absolute continuity was ensured in [61, 31, 42], and a sharp L^p estimate has been provided in [31, 33, 32]. Finally, its role in the connection of the problems has been established and generalized in [1, 62].

0.2 List of symbols

X, Y	Two Polish spaces;
$c : X \times Y \to \overline{R}^+$	The "cost function", a l.s.c. function;
$\lvert \cdot \rvert$, $\lVert \cdot \rVert$	the Euclidean norm and a generic norm in $\mathbb{R}^N$;
$[\![x, y]\!]$, $]\!]x, y[\![$	Closed and open segments in $\mathbb{R}^N$;
$\mathscr{H}^s$, $\mathscr{L}$	$s-$dimensional Hausdorff measure, Lebesgue measure on $\mathbb{R}^N$;
$\mathscr{B}\,(X)$	Borel $\sigma-$algebra on X;
$\mathscr{P}\,(X)$	Set of the subsets of X;
Δ	Diagonal of $X \times X$, i.e. $\{(x, x) : x \in X\}$;
$\mathcal{M}^+(X)$	Finite positive measures on the space X (see Section A.1);
$\mu \llcorner A$	Restriction of μ (a measure) to A (a set), i.e. $\mu \llcorner A(S) := \mu(A \cap S)$;
$\langle \cdot, \cdot \rangle$	Scalar product or, in general, duality;
$\mu \otimes \nu$, $\mu \boxtimes \nu$	Tensor (resp. rescaled tensor) product of measures (see Section A.3);
$a \wedge b$, $a \vee b$	$\min\{a, b\}$ and $\max\{a, b\}$ for a, $b \in \mathbb{R}$.
$t_\#$	Push-forward operator (see Section A.2)

0.3 Plan of the thesis

This thesis is divided in three large parts. The first one, which is the main one, is entirely devoted to the study of the mass transport problem; the intention is to give an idea, as much precise as possible, of the state of the art at the moment, with a particular interest in solving the original Monge's question, i.e. showing with suitable assumptions the existence of an optimal transport map. According to Sudakov's approach, we are mainly concerned in the situations where the cost is $c(x, y) = \|y - x\|$ where $\| \cdot \|$ is a norm in $\mathbb{R}^N$. We will try to give the definitions and the most known properties in a big generality, and we also try –as much as we could– to reach sharp results, providing a large number of counterexamples. This first part is meant to be self-contained, since we give a proof to all the assertion we are interested in; moreover, it contains several new results.

In Chapter 1 we set the Monge and the Kantorovich problems in a general Polish space.

In Chapter 2 we give two new results: the first one provides the existence of transport maps (it is easily seen that in general there might be no transport map at all); the second one shows that, in a great generality, it is true that the infimum of the costs of the transport maps equals the infimum of the costs of the transport plans, while a priori it is only known to be greater. Both these results extend the previous known ones and are shown to be sharp with examples.

In Chapter 3 we define the classical notions of $c-$cyclical monotonicity and we introduce the dual problem, leading to a Kantorovich potentials; the usual definitions are improved to allow considering a cost function whose value may be $+\infty$ somewhere.

In Chapter 4 we present a notion of tranport rays, transport set and other related tools: these notions have been widely used in the mass transportation problem, here we slightly refine the ones considered in [2, 4], which were more general than in the previous treatments; moreover we briefly discuss the one-dimensional case, and finally we present a couple of secondary variational problems, introduced in [2, 4], which will be later on our main "weapon" to attack the existence claim; also these secondary problems were already considered in [2, 4], and we improve the results found there.

Even though our main interest is in the case of the norms, in Chapter 5 we tell something about the quadratic case, we give a proof of the original result of Brenier and we briefly discuss some of the applications, among them the polar factorization, which was the principal goal of the construction of Brenier.

Chapter 6 is devoted to the existence Theorems for the case of the Euclidean norm: we first discuss the idea of Sudakov, underlining –also with a couple of counterexamples– where it fails; then we give a description of the Theorem of Evans and Gangbo, and a short discussion about the ones of Caffarelli–Felmann–McCann and of Trudinger–Wang; finally we present a complete proof of the Theorem of Ambrosio, which is the model on which our further existence proofs are built.

In Chapter 7 we show a general abstract existence Theorem, appeared in [2] and built on the pattern of the Theorem of Ambrosio.

With this result, in Chapter 8 we can show the existence result for the case of the quasi-euclidean norms (see Definition 8.1) and we provide a convergence result on the approximating transport maps.

Finally, again using the general abstract Theorem and an Euclidean approximation, we prove two new existence results, appeared in [4]: the case of the crystalline norms is considered in Chapter 9, while the case of a generic norm in $\mathbb{R}^2$ is in Chapter 10.

The second part of the thesis, taken from [62], is devoted to the generalization of the mass transport problem and of the equivalences between this problem, the shape optimization problem and the Brenier problem, to the case of a Riemannian manifold, non necessarily complete and possibly with a boundary. Moreover, we replace the Riemannian distance on the manifold by a more general finslerian semi-distance. Our aim is to show that the problems can be generalized to this setting, and they are still equivalent. We also relax these problems, that have no solutions in general, and we show that the relaxations admit solutions and that each of the problems has exactly the same associated transport densities as the others.

In Chapter 11 we introduce the classical problems and we describe our ambient space. In Chapter 12 we set the problems in our manifolds and we provide a relaxations. Finally, Chapter 13 contains all the results of the second part.

The third part of the thesis is devoted to the study of the regularity properties of the transport density in the classical situation of the Euclidean norm in a bounded subspace of $\mathbb{R}^N$ (however, we also discuss which of our results are still valid in the whole $\mathbb{R}^N$ and which are not).

In Chapter 14 we give an introductive discussion about the meaning of the transport density and we discuss in the detail some properties of it; we also provide a notion of dimension for measures and some geometrical results we use in the following. We also give some examples to show the regularity that one cannot expect, in order to infer that our following estimates will be sharp.

In Chapter 15 we give an estimate –which is shown to be sharp thanks to some counterexamples– about the dimension of the transport density μ depending on the dimensions of the data f^+ and f^-.

In Chapter 16 we give our first results about the L^p properties of the transport density: we first show that if at least one between f^+ and f^- belong to L^1, then the transport density is unique (a priori, each optimal transport plan may define a different transport density); this result is well known (see [1, 42]). Moreover, we prove that if at least one between f^+ and f^- is in L^1, then also the transport density (which is unique) is in L^1: this results has been established in [61], see also [31, 42]. We finally give some estimates about the L^p properties of the transport density; in particular we show that if both f^+ and f^- belong to L^p, then the transport density belong to each L^q with $q < p$: the case $q = p$ is reached only in the cases $p = 1$ and $p = +\infty$. All these results are taken from [31], and make use of the geometrical claims we proved in Chapter 14. Since easy examples show that the fact that f^+ and f^- are in L^p does not imply that μ is in L^q for any $q > p$, the only case left open is the implication $L^p \Rightarrow L^p$.

Then, in Chapter 17 we provide a PDE argument to show this implication for any $p \geq 2$, presented in [33], and in Chapter 18 we conclude giving an interpolation argument, taken from [32], that, making use of the L^1 and the L^∞ estimates, shows the sharp implication for any other $1 < p < +\infty$.

Acknowledgments

I am indebted to many people for the help I received during the preparation of this thesis: I should mention a lot of fruitful conversations, among them those with G. Buttazzo, L.C. Evans, B. Kirchheim. In particular, B. Kirchheim induced me to consider the problems that are developed in Chapter 2, and he also suggested me how to make the proof of Lemma 6.9 easier.

Moreover, there is no word to thank L. Ambrosio for his continuous help; very little of this work could have been done without his explanations, advice and support.

Part I

Monge transport problem and existence of transport maps

Chapter 1

Setting of the Monge–Kantorovich problem

1.1 Monge mass transport problem

Throughout this paper, X and Y will be two Polish spaces (see Section A.1 in the Appendix for the definition and the main properties) and f^+ and f^- will be two finite positive measures respectively on X and Y with the same total mass; this means $f^+ \in \mathcal{M}^+(X)$, $f^- \in \mathcal{M}^+(Y)$ and $\|f^+\| = f^+(X) = f^-(Y) = \|f^-\|$. The goal of the transport problem is to move f^+ into f^-: this means, roughly speaking, that one needs a function specifying where to move the mass. Taking then any measurable function $t : X \to Y$, we try to understand what should mean to consider it as a way to transport the distribution of mass given by f^+: the idea is that one should move all the mass which is in a point x into the point $t(x)$. Of course this simple "pointwise" point of view is not formally correct unless f^+ is the sum of countably many point masses, but it leads to the correct idea that the mass which will be in any Borel set $A \subseteq Y$ after the movement is $f^+\big(t^{-1}(A)\big)$. Since we want to move f^+ in f^-, it should happen that this mass equals $f^-(A)$. Following this intuitive argument, we give then the following

Definition 1.1. *A f^+-measurable function $t : X \to Y$ is said to be a* transport map *from f^+ to f^- if $t_\# f^+ = f^-$, where the push-forward is defined (see Section A.2) as*

$$t_\# f^+(A) := f^+\big(t^{-1}(A)\big) \qquad \textit{for any } A \in \mathscr{B}\,(Y).$$

Note that in the preceding definition it would be equivalent to ask t to be Borel instead of only f^+-measurable: in fact, thanks to Theorem A.18, if t is measurable we can change it in a f^+- negligible set (then t remains essentially the same map) in such a way that t becomes Borel.

Once we know what a transport map is, we are interested in finding a cheapest one; to explain what this means, we need to consider a l.s.c. function $c : X \times Y \to \mathbb{R}^+ \cup \{+\infty\}$ which will be called *cost function*: its meaning is that the cost of moving a unit mass from the point $x \in X$ to the point $y \in Y$ is $c(x, y)$. Recalling the meaning of moving the mass following t it is natural to consider, as the cost of the transport t, the number

$$\mathfrak{C}(t) := \int_X c(x, t(x))\, df^+(x); \tag{1.1}$$

to avoid misunderstandings, note that we call *cost* the function $\mathfrak{C}(t)$ and *cost function* the function $c(x, y)$.

The hypothesis of the cost function c to be l.s.c. is always taken, when dealing with transport problems, because otherwise the problem would be meaningless from many points of view.

The task will be to find an *optimal transport map*, according to the

Definition 1.2. *An* optimal transport map *is a map $t : X \to Y$ minimizing the cost (1.1) among all the transport maps.*

1.2 Some examples

Trying to solve Monge problem, one has to face many problems: a first problem is that it may happen that there are no optimal transport maps, or that there are no transport maps at all, as we will see with a couple of examples.

Example 1.3. It is possible that there are no transport maps: given three points x, y_1 and y_2 in $X = Y$, as in the left of Figure 1.1, take $f^+ := 2\delta_x$ and $f^- := \delta_{y_1} + \delta_{y_2}$. Given any function $t : X \to Y$, one has

$$t_\# f^+ = t_\# 2\delta_x = 2\delta_{t(x)} \neq f^-,$$

thus there exists no transport map for this configuration.

Example 1.4. It is possible that there are transport maps, but no one is optimal: take, as in the right of Figure 1.1, three parallel rectangles Π_1, Π_2 and Π_3 in $X = Y = \mathbb{R}^3$, and let $f^+ := 2\mathscr{H}^2 \llcorner \Pi_2$ and $f^- := \mathscr{H}^2 \llcorner \Pi_1 + \mathscr{H}^2 \llcorner \Pi_3$. One can show that there are transport maps: for example, as the arrows in the figure explain, one can move the upper half of Π_2 on Π_1 and the lower half on Π_3; more in general, Theorem 2.4 will show that there always exist transport maps if f^+ has no point masses. On the other hand, if $c(x, y) = |y - x|$

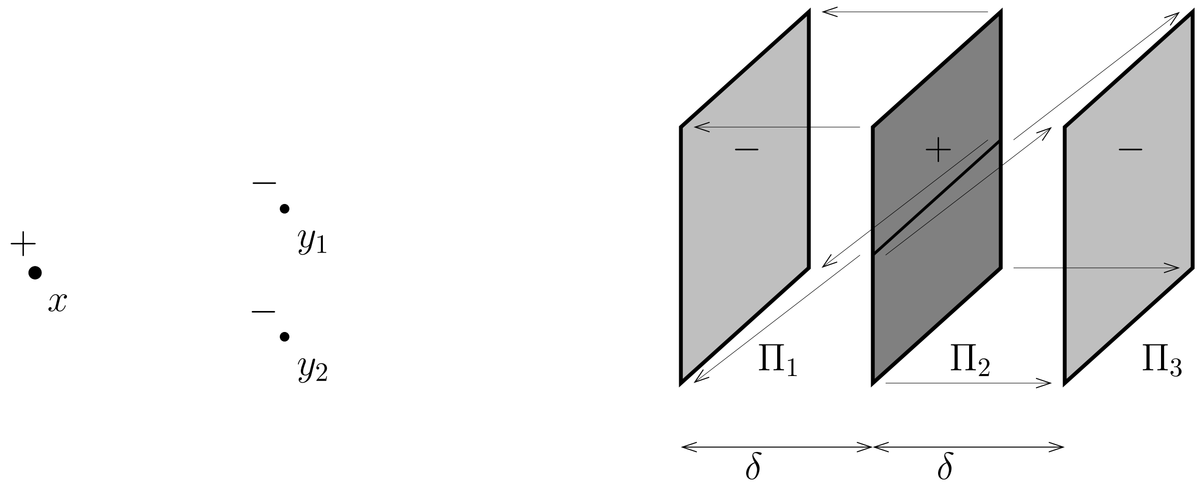


Figure 1.1: Two examples

then no transport map is optimal: if t is any transport, clearly $|t(x)-x| \geq \delta$ for f^+−a.e. x, thus

$$\mathfrak{C}(t) := \int_{\mathbb{R}^3} |t(x)-x| \, df^+(x) \geq \delta \int_{\mathbb{R}^3} df^+(x) = \delta.$$

We obtain $\mathfrak{C}(t) > \delta$ for any transport map t because it is not possible that the segment connecting x and $t(x)$ is horizontal for f^+−a.e. x; in fact, let A be the set of those points x such that $t(x)-x$ is an horizontal vector: we have $2f^-(t(A)) = f^+(A)$ by the geometry, but $t_\# f^+(t(A)) = f^+\big(t^{-1}(t(A))\big) \geq f^+(A)$, thus since t is a transport $f^+(A) \leq f^+(A)/2$ and then A is negligible. Finally, we can easily find transport maps moving f^+−a.e. point "almost horizontally" and then with cost arbitrarily close to δ, then the inf of the costs of the transport maps is δ and there are no optimal transport maps. Note carefully that the same happens if $c(x,y) = h(|y-x|)$ with any strictly increasing function h: this is really important, since one understands that it is not reasonable to expect a result of existence of optimal transport maps covering the case when f^+ has a positive mass in a 2−dimensional set and $X = \mathbb{R}^3$ or, with an obvious generalization of the last example, in a $(N-1)$−dimensional set and $X = \mathbb{R}^N$. In particular the existence Theorem 5.6, which says that there are optimal transport maps for the quadratic cost function $c(x,y) = |y-x|^2$ in $\mathbb{R}^N$ if $f^+(S) = 0$ whenever $\mathscr{H}^{N-1}(S) < +\infty$, is sharp (see also Remark 5.8).

1.3 The generalization of Kantorovich

Even if Monge problem is, as we saw in Section 1.1, quite easy to state, it is not easy at all to solve it and in general, as Example 1.4 showed, it may also have no solution.

However, let us consider carefully that example: one would have an optimal transport if he could somehow move each point x of Π_2 half on the left and half on the right. In Monge's setting this makes no sense, since the point x is entirely moved on $t(x)$, but this idea of splitting the masses of each point is useful because it leads us to the Kantorovich problem. Following then his work (see [53, 54]), we give the

Definition 1.5. *A* transport plan *between f^+ and f^- is a positive measure $\gamma \in \mathcal{M}^+(X \times Y)$ whose marginals are f^+ and f^-, i.e. for each $A \in \mathscr{B}(X)$ and $B \in \mathscr{B}(Y)$*

$$\pi_1\gamma(A) = \gamma(A \times Y) = f^+(A) \quad \text{and} \quad \pi_2\gamma(B) = \gamma(X \times B) = f^-(B).$$

The set of the transport plans between f^+ and f^- will be denoted by $\Pi(f^+, f^-)$.

In the preceding definition, as we will do throughout the paper, we used the convention to denote by $\pi_1 : X \times Y \to X$ and $\pi_2 : X \times Y \to Y$ the projections and, with a slight abuse, also by $\pi_1 : \mathcal{M}^+(X \times Y) \to \mathcal{M}^+(X)$ and $\pi_2 : \mathcal{M}^+(X \times Y) \to \mathcal{M}^+(Y)$ their push-forward, to avoid the uncomfortable writing $\pi_{1\#}$ and $\pi_{2\#}$.

Roughly speaking, while for a transport map t one can say that all the mass $f(x)$, which is initially in x, is moved in $t(x)$, for a transport plan γ one can say that only the mass $\gamma(x, y)$ from x has to be moved in y. To be more precise, making use of the Disintegration Theorem (Theorem A.15 in the Appendix) we can decompose γ with respect to the projection $\pi_1 : X \times Y \to X$ finding

$$\gamma = \gamma_x \otimes \pi_1\gamma = \gamma_x \otimes f^+,$$

where $\gamma_x \in \mathcal{P}(Y)$ for any $x \in X$ and the meaning of the preceding equation is that

$$\langle \gamma, \varphi \rangle = \int_X \langle \gamma_x, \varphi(x, \cdot) \rangle \, df^+(x) \qquad \forall \varphi \in C_c(X \times Y). \tag{1.2}$$

Recalling again Theorem A.15 and the notion of $\mu-$measurable (measure valued) maps given in Section A.4, we know that, given any transport plan γ, one has a f^+-measurable (measure valued) map $x \to \gamma_x$ such that (1.2) holds. On the other hand, given a f^+-measurable map $x \to \gamma_x$ with $\gamma_x \in \mathcal{P}(Y)$, one can recover a measure γ taking (1.2) as a definition; this measure is a transport plan if and only if its second marginal is f^-, since its first marginal is always f^+ by construction.

We can then explain better the meaning of a transport plan γ: while a transport map t "says" that the mass in x must be entirely moved on $t(x)$, γ says that the mass in x must be distributed in Y proportionally to the measure γ_x. Then we understand that the transport plans are a generalization of the transport maps, and γ "corresponds" to

t if $\gamma_x = \delta_{t(x)}$. In other words, to any transport map t we associate the transport plan $\gamma(t)$ defined by (1.2) setting $\gamma_x = \delta_{t(x)}$ for any x; as we saw, this makes sense if $x \to \delta_{t(x)}$ is a f^+-measurable function, that is easily seen to be true recalling Definition A.14. Moreover, recalling that $t_\# f^+ = f^-$ by definition, it is also straightforward to note that $\gamma(t)$ is a transport plan, which can be equivalently defined as

$$\gamma(t) := \big(\mathrm{Id}, t\big)_\# f^+. \tag{1.3}$$

The previous definition suggests to regard the transport maps as particular transport plans with the property that γ_x is a Dirac delta for any x; but also the converse is true, as the next well-known Lemma shows.

Lemma 1.6. *A transport plan γ is induced by a transport map –in the sense of (1.3)– if and only if γ_x is a Dirac delta for f^+-a.e. x, or equivalently if γ is concentrated in a $\sigma-$compact set S such that for f^+-a.e. x the set $\{y : (x,y) \in S\}$ consists of exactly a point.*

Proof. If γ comes from the transport t, by definition we know that $\gamma_x = \delta_{t(x)}$ for any x, then there is nothing to prove.

Conversely, suppose the existence of a set of $\gamma-$full measure S (which we can assume to be $\sigma-$compact thanks to Theorem A.17) such that, for each x belonging to a $\sigma-$compact set $\widetilde{X}$ of full f^+-measure, the set $\{y : (x,y) \in S\}$ contains a unique point, that we call $t(x)$; so, possibly removing from S a $\gamma-$negligible subset, we can assume that

$$S = \Big\{ \big(x, t(x)\big) : \ x \in \widetilde{X} \Big\} \tag{1.4}$$

and, as one can imagine, we want to prove that t is a transport map and that γ is induced by t. First of all we decompose as usual $\gamma = f^+ \otimes \gamma_x$: it is clear that each measure γ_x must be concentrated in S, but also in $\{x\} \times Y$: thus, for each $x \in \widetilde{X}$ –then for f^+-a.e. x– the probability measure γ_x is concentrated in $\{x, t(x)\}$, thus it must equal $\delta_{t(x)}$: we obtain that γ is induced by the map t, then we only need to prove that t is a transport map.

The fact that t is measurable is true, since for any closed set $B \subseteq Y$ one has

$$t^{-1}(B) = \big\{ x \in \widetilde{X} : t(x) \in B \big\} = \Big\{ x \in \widetilde{X} : \ \big(\{x\} \times B\big) \cap S \neq \emptyset \Big\} = \pi_1 \Big\{ S \cap \big(\widetilde{X} \times B\big) \Big\},$$

and the last set is $\sigma-$compact thus Borel. Finally, to prove that t is a transport map we only need to show $t_\# f^+ = f^-$: but for any Borel set $A \subseteq Y$ one has

$$t_\# f^+(A) = f^+\big(t^{-1}(A)\big) = \pi_1 \gamma\big(t^{-1}(A)\big) = \gamma\big(t^{-1}(A) \times Y\big) = \gamma\Big(\big(t^{-1}(A) \times Y\big) \cap S\Big);$$

moreover, (1.4) immediately implies that

$$\big(t^{-1}(A) \times Y\big) \cap S = \big(\widetilde{X} \times A\big) \cap S,$$

and then we infer

$$t_\# f^+(A) = \gamma\Big(\big(\widetilde{X} \times A\big) \cap S\Big) = \gamma\big(\widetilde{X} \times A\big) = f^-(A),$$

so $t_\# f^+ = f^-$ is established, therefore t is a transport map and the thesis follows. $\square$

The proof of the preceding Lemma shows that if γ is a transport plan concentrated in the graph Γ of t, then t is a transport map and γ is induced by t. Note that then, in this case (but not in general!), the transport plan γ is uniquely determined by Γ; in general, if a measure $\alpha \in \mathcal{M}^+(X \times Y)$ is concentrated in the graph of a function $h : X \times Y$, it is uniquely determined by its first projection: in fact, the argument of last lemma proves that $\alpha = (\mathrm{Id}, h)_\# \pi_1 \alpha$.

We need now to extend also the notion of cost of a transport from maps to plans: recalling that the cost function $c(x, y)$ represent the cost of moving a unit mass from x to y, it is immediate to give the following definition, extending (1.1):

$$\mathfrak{C}(\gamma) := \iint_{X\times Y} c(x,y)\, d\gamma(x,y). \tag{1.5}$$

The fact that (1.5) is an extension of (1.1) –considering the transport maps as special transport plans according to (1.3)– is ensured by the following

Lemma 1.7. *If t is a transport map, then $\mathfrak{C}\big(\gamma(t)\big) = \mathfrak{C}(t)$, where $\gamma(t)$ is the transport plan given by (1.3).*

Proof. It is an easy calculation using (1.5), (1.1) and the properties of the push-forward:

$$\begin{aligned} \mathfrak{C}\big(\gamma(t)\big) &= \int_{X\times Y} c(x,y)\, d\gamma(t)(x,y) = \int_{X\times Y} c(x,y)\, d\big(\mathrm{Id}, t\big)_\# f^+(x,y) \\ &= \int_X c(z, t(z))\, df^+(z) = \mathfrak{C}(t). \end{aligned}$$

$\square$

Following then Definition 1.2, we can generalize saying that

Definition 1.8. *An* optimal transport plan *is a measure $\gamma \in \mathcal{M}^+(X \times Y)$ minimizing the cost (1.5) among all the transport plans. The set of the optimal transport plans will be denoted by $\Pi'(f^+, f^-) \subseteq \Pi(f^+, f^-)$.*

The Kantorovich problem of finding an optimal transport plan is then a generalization of the Monge one of finding an optimal transport map, and this generalization is extremely useful for many reasons; let us briefly discuss some of them. First of all, the transport plans belong to the normed space $\mathcal{M}^+(X \times Y)$, and in particular we will see that they are a weak* closed and convex subset of it; moreover, the cost (1.5) is a linear function of a transport plan. On the other hand, the transport maps are only measurable functions, then there is not a useful structure on their set, and there is no immediate way to find a limit of a minimizing sequence of transport maps: consequently, it is much easier to compare different plans than different maps. Finally, another big difference is about the symmetry: for the Kantorovich problem, exchanging X and f^+ with Y and f^- does not have any effect, and it is completely equivalent to transport f^+ on f^- or f^- on f^+; for the Monge problem, this is absolutely not true, just notice that in Example 1.3 there is not any transport map from f^+ to f^- but there is a transport map from f^- to f^+; moreover, all the existence results we will give in the sequel about the Monge problem will have different hypotheses about f^+ and f^- (that is, much stronger hypotheses about f^+). This difference about the symmetry is the reason why we talk about transport maps *from f^+ to f^-* and transport plans *between f^+ and f^-*.

Other two differences that can convince that the Kantorovich problem is easier are given by the following two claims:

Remark 1.9. *There always exist transport plans (while Example 1.3 shows that there could be no transport map): in fact, the measure $f^+ \boxtimes f^-$ defined (see also Section A.2) by*

$$\mu \boxtimes \nu(A \times B) = \frac{\mu(A) \cdot \nu(B)}{\|\mu\|} \qquad \forall\, A \in \mathscr{B}(X),\ \forall\, B \in \mathscr{B}(Y) \tag{A.2}$$

is a transport plan between f^+ and f^-.

Remark 1.10. *The cost $\mathfrak{C}$ is weakly* l.s.c. on $\mathcal{M}^+(X \times Y)$, due to the lower semicontinuity of the cost function c: this is a well known property of the l.s.c. functions, since $\mathfrak{C}(\gamma) = \iint c\, d\gamma$.*

We can now realize that the Kantorovich problem is in fact "too easy", since it is immediately seen to admit a solution:

Theorem 1.11. *There exist optimal transport plans; in other words, $\Pi'(f^+, f^-)$ is never empty. In particular, both $\Pi(f^+, f^-)$ and $\Pi'(f^+, f^-)$ are non-empty, weak* closed, bounded and convex.*

Proof. Let us first prove the existence of optimal transport plans: in fact, since by Definition 1.5 $\|\gamma\| = \|f^+\| = \|f^-\|$ for any transport plan γ, $\Pi(f^+, f^-)$ is bounded in

$\mathcal{M}^+(X \times Y)$; moreover, by Ulam Theorem (see Section A.1), there are compact sets $X_0 \subseteq X$ and $Y_0 \subseteq Y$ such that $f^+(X \setminus X_0) < \varepsilon$ and $f^-(Y \setminus Y_0) < \varepsilon$. Then $X_0 \times Y_0$ is a compact subset of $X \times Y$ and, if $\gamma \in \Pi(f^+, f^-)$,

$$\gamma\big((X \times Y) \setminus (X_0 \times Y_0)\big) \le \gamma\big((X \setminus X_0) \times Y\big) + \gamma\big(X \times (Y \setminus Y_0)\big) = f^+\big(X \setminus X_0\big) + f^-\big(Y \setminus Y_0\big) \le 2\varepsilon :$$

then, by Prokhorov Theorem (see Section A.1), we infer that $\Pi(f^+, f^-)$ is weakly* sequentially relatively compact, therefore there is a sequence $\gamma_n \in \Pi(f^+, f^-)$ minimizing $\mathfrak{C}$ and weakly* converging to a measure $\gamma \in \Pi(f^+, f^-)$. Since the cost is l.s.c. as noticed in Remark 1.10 above, one derives that γ is optimal, and then there are optimal transport plans. In particular, both $\Pi(f^+, f^-)$ and $\Pi'(f^+, f^-)$ are non-empty and bounded. Moreover, $\Pi(f^+, f^-)$ is convex and weak* closed thanks to the linearity of the projections and by the fact that the projection of a weak* limit is the weak* limit of the projections. Finally, also $\Pi'(f^+, f^-)$ is immediately seen to be convex and weak* closed thanks to the linearity of the cost $\mathfrak{C}$ and its lower semicontinuity (recall that $\Pi'(f^+, f^-)$ is the lowest level of $\mathfrak{C}$). $\square$

We prove now that minimizing (1.5) is a local property: this means, we have the

Lemma 1.12. *Let us consider a measure $\gamma \in \Pi(f^+, f^-)$ and a Borel set $A \subseteq X \times Y$: if γ is an optimal transport plan with $\mathfrak{C}(\gamma) < +\infty$, then $\gamma \llcorner A$ is an optimal transport plan between its marginals.*

Proof. Since $\mathfrak{C}$ is a linear functional on $\mathcal{M}^+(X \times Y)$, $\mathfrak{C}(\gamma) = \mathfrak{C}(\gamma \llcorner A) + \mathfrak{C}(\gamma \llcorner {}^c A)$. If there would exist a measure $\tilde{\gamma}$ having the same marginals as $\gamma \llcorner A$ but with $\mathfrak{C}(\tilde{\gamma}) < \mathfrak{C}(\gamma \llcorner A)$, then $\tilde{\gamma} + \gamma \llcorner^c A$ would be a transport plan between f^+ and f^- with strictly less cost than γ (use the linearity of $\mathfrak{C}$), that is impossible if γ is optimal. $\square$

Let us consider now a much more general result, taken from [4], which gives the optimality of the disintegrated measures obtained by γ (actually, the preceding Lemma could be obtained from the next one with $g = \chi_A$ and $Z = \{0, 1\}$).

Lemma 1.13. *Let Z be a Polish space and let $g : X \times Y \longrightarrow Z$ be a Borel function; let moreover $\bar{\gamma}$ be an optimal transport plan with finite cost. If the disintegration of $\bar{\gamma}$ induced by g is $\bar{\gamma} = \gamma_z \otimes \eta$ with $\eta = g_\# \bar{\gamma}$, and f_z^+ and f_z^- are the marginals of γ_z, then for $\eta-$a.e. $z \in Z$ γ_z is an optimal transport plan between f_z^+ and f_z^-.*

Proof. In this proof we will make use of the Projection Theorem and of the Measurable Selection Theorem, which can be found in the book of Castaing and Valadier [25] and which are also stated in Section A.1; it is also useful to recall that a multi-function

$\alpha : X \to \mathscr{P}(Y)$ is said to be $\eta-$measurable if its graph is measurable as a subset of $X \times Y$ with respect to the $\sigma-$algebra $\mathscr{B}_\eta(X) \otimes \mathscr{B}(Y)$ on $X \times Y$, where $\mathscr{B}(Y)$ is the Borel $\sigma-$algebra on Y and $\mathscr{B}_\eta(X)$ is the completion of the Borel $\sigma-$algebra on X with respect to η.

First of all, we make for simplicity the assumption that $\|f^+\| = \|f^-\| = 1$, which is clearly non restrictive; moreover, we consider the unit ball E of $\mathcal{M}^+(X \times Y)$, and we recall that the weak* topology on E is induced by a metric since $\mathrm{C}_c(X \times Y)$ is separable (this is a well-known fact, see for example [22]); in particular, E is a Polish space. Then we recall that the function $\varphi : Z \to E$ given by $\varphi(z) = \gamma_z$ is $\eta-$measurable by the Disintegration Theorem A.15. Moreover the multi-function $\psi : E \to \mathscr{P}(E)$ given by

$$\psi(\gamma) = \{\beta \in E : \ \pi_1\beta = \pi_1\gamma, \pi_2\beta = \pi_2\gamma\} = \Pi\big(\pi_1\gamma, \pi_2\gamma\big)$$

can be easily seen to have a closed graph. Now, using Lusin Theorem and recalling Theorem A.17, we find that Z is, up to a $\eta-$negligible set, the union of countably many compact sets Z_i such that the restriction of φ to each Z_i is continuous; it follows that the graph of $\psi \circ \varphi \big| Z_i$ is a closed subset of $Z \times E$, and then η is concentrated on a $\sigma-$closed set Z_0 such that the set

$$\Theta := \Big\{(z, \gamma) \in Z \times E : \ z \in Z_0, \ \gamma \in \psi\big(\varphi(z)\big)\Big\}$$

is a countable union of closed sets, hence Borel; notice that $\psi\big(\varphi(z)\big) = \Pi(f_z^+, f_z^-)$ is the set of the transport plans between f_z^+ and f_z^-.

We consider now $s \in \mathbb{R}$: the set $\{\gamma : \ \mathfrak{C}(\gamma) \leq s\}$ is closed in E since, as noted in Remark 1.10, $\mathfrak{C}$ is a weakly* l.s.c. function on $\mathcal{M}^+(X \times Y)$; then the set $U_s \subseteq Z \times E$ given by

$$U_s := \Big\{(z, \gamma) \in Z \times E : \ z \in Z_0, \ \gamma \in \psi\big(\varphi(z)\big), \ \mathfrak{C}(\gamma) \leq s\Big\}$$

is Borel, since $U_s = Z \times \{\gamma : \ \mathfrak{C}(\gamma) \leq s\} \cap \Theta$. The Projection Theorem ensures that the projection of U_s on Z is $\eta-$measurable, and this projection is clearly $v^{-1}\big(]\!]-\infty, s]\!]\big)$ where $v : Z_0 \to \mathbb{R}^+$ is given by

$$v(z) := \inf\Big\{\mathfrak{C}(\gamma) : \ \gamma \in \Pi(f_z^+, f_z^-)\Big\}.$$

Then v is $\eta-$measurable, hence the set

$$\Psi := \Big\{(z, \gamma) \in Z \times E : \ z \in Z_0, \ \mathfrak{C}(\gamma) = v(z)\Big\}$$

is $\mathscr{B}_\eta(Z) \otimes \mathscr{B}(E)-$measurable. Thus the set $\Theta \cap \Psi$, which is the set of the pairs (z, γ) such that $z \in Z_0$ and γ is an optimal transport plan between f_z^+ and f_z^-, is measurable; moreover it intersects $\{z\} \times E$ for any $z \in Z_0$, thus $\eta-$a.e. z.

By the Measurable Selection Theorem, we can find an $\eta-$measurable function $z \to \tilde{\gamma}_z$ such that $\tilde{\gamma}_z$ is an optimal transport plan between f_z^+ andf_z^- for $\eta-$a.e. z. We can then define $\tilde{\gamma} = \tilde{\gamma}_z \otimes \eta$: making use of Lemma A.16, we see that this is a transport plan between f^+ and f^-, since the marginals of γ_z and of $\tilde{\gamma}_z$ are the same for $\eta-$a.e. z. By the properties of disintegration we know

$$\mathfrak{C}(\bar{\gamma}) = \int_Z \mathfrak{C}(\gamma_z)\, d\eta(z) \qquad \text{and} \qquad \mathfrak{C}(\tilde{\gamma}) = \int_Z \mathfrak{C}(\tilde{\gamma}_z)\, d\eta(z);$$

then, recalling that $\mathfrak{C}(\tilde{\gamma}_z) \leq \mathfrak{C}(\gamma_z)$ for $\eta-$a.e. z because $\tilde{\gamma}_z$ is an optimal transport plan, but also that $\mathfrak{C}(\bar{\gamma}) \leq \mathfrak{C}(\tilde{\gamma})$ because $\bar{\gamma}$ is optimal by hypothesis, we infer that γ_z is an optimal transport plan between f_z^+ and f_z^- for $\eta-$a.e. z, thus the thesis is reached. $\square$

1.4 Wasserstein distance

Here we give the notion of Wasserstein distance between two measures, which depends on the cost function c; it was introduced in a Probability setting, but in the last years has been successfully used also in other contexts, see for example [68, 48, 11]. Through this section, we will denote $\mathcal{W} = \big\{(\mu, \nu) \in \mathcal{M}^+(X) \times \mathcal{M}^+(Y) : \ \|\mu\| = \|\nu\| \geq 0\big\}$.

Definition 1.14. *Given $(f^+, f^-) \in \mathcal{W}$, we define the* Wasserstein distance *between f^+ and f^-, $d_{\mathcal{W}}(f^+, f^-)$, as the minimum of the Kantorovich problem associated to f^+ and f^-.*

We used the word distance, but it is clear that it makes no sense to ask whether or not $d_{\mathcal{W}}$ is a distance unless $X = Y$; however, in this case $d_{\mathcal{W}}$ is seen to be a semi-distance if the cost function c is symmetric, null on the diagonal $\Delta = \{(x, x) : x \in X\}$ and fulfills the triangular inequality $c(x, z) \leq c(x, y) + c(y, z)$; moreover it is a distance if $\{c(x, y) = 0\}$ is exactly the diagonal Δ. This is an easy application of the notion of composition between transport plans, which we will discuss in a more general setting in Section 12.6. An other interesting situation is if $X = Y$ and d is a distance on X: if the cost function is $c(x, y) = d(x, y)^p$ and $p > 1$, then one gives a different definition of $d_{\mathcal{W}}$, that is

$$d^p_{\mathcal{W}}(f^+, f^-) := \big(\min(1.5)\big)^{1/p} :$$

thanks to the $p-$th root, the function $d^p_{\mathcal{W}}$ turns out to be a distance.

Lemma 1.15. *The following inequality holds:*

$$d_{\mathcal{W}}(f + g, h + l) \leq d_{\mathcal{W}}(f, h) + d_{\mathcal{W}}(g, l).$$

Proof. The claim is straightforward to prove: given any transport plan γ_1 between f and h and any transport plan γ_2 between g and l, $\gamma_1 + \gamma_2$ is a transport plan between $f+g$ and $h+l$ with cost $\mathfrak{C}(\gamma_1) + \mathfrak{C}(\gamma_2)$. The thesis then follows. □

Let us note now that $\mathcal{W}$ is a weakly* closed subset of $\mathcal{M}^+(X) \times \mathcal{M}^+(Y)$; thus we can also prove the following fact:

Lemma 1.16. *The Wasserstein distance $d_{\mathcal{W}}$ is weakly* l.s.c. on $\mathcal{W}$.*

Proof. Let us assume $(\mu_n, \nu_n) \in \mathcal{W}$, $(\mu, \nu) \in \mathcal{W}$, $\mu_n \xrightharpoonup[\mathcal{M}^+(X)]{*} \mu$ and $\nu_n \xrightharpoonup[\mathcal{M}^+(Y)]{*} \nu$, and let γ_n be an optimal transport plan between μ_n and ν_n; since μ_n has a weak* limit, $\|\mu_n\|_X$ is bounded, hence $\{\gamma_n\}$ is a bounded sequence in $\mathcal{M}^+(X \times Y)$ and so, up to a subsequence, $\gamma_n \xrightharpoonup{*} \gamma$ in $\mathcal{M}^+(X \times Y)$: this implies that γ is a transport plan between μ and ν and the thesis follows by the lower semicontinuity of the cost $\mathfrak{C}$, already noticed in Remark 1.10. □

Chapter 2

Transport maps and equality of the extremals

In this chapter we will consider two questions that arise in a very natural way when the Monge-Kantorovich problem is considered. The first one concerns the existence of transport maps: in fact, it has easily been seen in last chapter that admissible transport plans always exist, unlike transport maps; however, we want to find some condition assuring the existence of transport maps. The second question concerns the equality between the infimum of the Monge problem, i.e. inf (1.1), and the minimum of the Kantorovich one, i.e. min (1.5): it is immediate to derive from Lemma 1.7 that $\min(1.5) \leq \inf(1.1)$, but of course one would like to know whether or not the equality holds, also because if not it is hard to imagine how the Kantorovich extension could give some help in studying the Monge problem.

These questions have already had some answer; in particular, in [46] the following result is proved:

Theorem 2.1 (Gangbo). *If f^+ and f^- are non-atomic (i.e. $f^+(\{x\}) = 0$ for any $x \in X$ and $f^-(\{y\}) = 0$ for any $y \in Y$) and c is continuous and bounded, then there are transport maps and moreover the equality inf(1.1)=min(1.5) holds.*

The proof of this Theorem is obtained recalling the abstract fact that any non-atomic finite positive measure on a Polish space is isomorphic to the Lebesgue measure on a closed real interval (see Theorem 9 in Chapter 15 of [66]), then one can reduce to the easier case of a real interval. Later, it was proved (see [1]) the following result, without the non-atomicity hypothesis on f^-:

Theorem 2.2. *If $X = Y$ is a convex and compact subset of $\mathbb{R}^N$ and f^+ is non-atomic, then there are some transport maps. In particular, if the cost function is continuous, then* *the equality inf(1.1)=min(1.5) holds.*

In this chapter we will assume the non-atomicity only for the measure f^+, and we will prove that the existence result is still true without any hypothesis on the l.s.c. cost function c and that the equality of the extremals is still true with the sole continuity of c. The result we will reach is then more general than the previously known ones; in particular it allows to consider cost functions which are not bounded (and that may also assume the value $+\infty$), which is one of the greatest difficulties in such kind of constructions, also because then the transport plans and maps can have an infinity cost. The proofs we will provide are constructive. The new results of this chapter are contained in the forthcoming paper [63].

2.1 Existence of transport maps

In this section we give an existence result for transport maps: the existence problem is one of the main differences between the original Monge problem and the Kantorovich extension, since Example 1.3 showed that for some f^+ and f^- there can be no transport map, while Remark 1.9 pointed out that there are always transport plans.

Let us consider again Example 1.3: in that case, there are no transport maps because of the point mass in x. In general, it is clear that a point mass can give problems because a transport map must move all the mass in the same place, while a transport plan may split the masses. We are now going to prove that point masses are the only obstacle to the existence, since if there are not point masses we build a transport map; moreover this map is better than $f^+ \boxtimes f^-$, which we already noticed to be always an admissible transport plan. Note also that the assumption we need is only about f^+, and this helps showing the asymmetry of the Monge problem we already discussed after Definition 1.8.

Since in the following we will use binary sequences, let us first give some notation: a *finite sequence* (resp. a *finite binary sequence*) is any $I \in \mathbb{N}^k$ (resp. $I \in \{0,1\}^k$) with $k \in \mathbb{N}$, while a *countable sequence* (resp. a *countable binary sequence*) is any $I \in \mathbb{N}^{\mathbb{N}}$ (resp. $I \in \{0,1\}^{\mathbb{N}}$); by $l(I)$ we denote the length (finite or $+\infty$) of any sequence, and given any $s \leq l(I)$ we write I_s to denote the subsequence made by the first s terms of I; if $I = (i_1, i_2, \ldots, i_k)$ is a finite sequence, we denote $(I, n) = (i_1, i_2, \ldots, i_k, n)$; moreover, we give a partial order saying that $I \leq J$ if $l(I) \leq l(J)$ and $J_{l(I)} = I$. Finally, we give also a (completely different) total order $\leq^k$ to $\{0,1\}^k$ saying that $I \leq^k J$ if, writing $I = (i_1, i_2, \ldots, i_k)$ and $J = (j_1, j_2, \ldots, j_k)$, the binary number $i_1 i_2 \ldots i_k$ is less than the binary number $j_1 j_2 \dots j_k$, i.e.

$$I = (i_1, \dots, i_k) \leq^k J = (j_1, \dots, j_k) \Longleftrightarrow \sum\nolimits_{n=1}^k 2^{k-n} i_n \leq \sum\nolimits_{n=1}^k 2^{k-n} j_n; \tag{2.1}$$

the set $\{0,1\}^k$, endowed with $\leq^k$, is trivially a finite totally ordered set.

To give consistence to our construction, we will need to consider also the finite binary sequences of length 0, where the unique element of $\{0,1\}^0$ is $\varnothing$.

We begin now with an important technical result:

Lemma 2.3. *Let Z be a Polish space and let $\mu \in \mathcal{P}(Z)$ be a probability measure. Then there are measures μ_I and sets $Z_I \subseteq Z$ for the finite binary sequences I with the following properties:*

(i) for any $k \in \mathbb{N}$, $\mu = \sum \big\{\mu_I : \; l(I) = k\big\}$;

(ii) for any finite sequence I of length $k = l(I)$, $\|\mu_I\| = 2^{-k}$;

(iii) for any finite sequence I, $\mu_I = \mu_{(I,0)} + \mu_{(I,1)}$;

(iv) for any finite sequence I, μ_I is concentrated in Z_I;

(v) for any $I \leq J$, $Z_J \subseteq Z_I$;

(vi) For any finite sequence $I \in \{0,1\}^k$, $I \neq (1, \dots, 1)$, the set $\cup\{Z_J : l(J) = k, \; J \leq^k I\}$ is closed;

(vii) for any countable binary sequence I except countably many, $\operatorname{diam}(Z_{I_k}) \xrightarrow[k \to +\infty]{} 0$;

(viii) in addition, if μ is non-atomic, then given $I \neq J \in \{0,1\}^k$, it is $Z_I \cap Z_J = \emptyset$.

Proof. We fix a metric on Z and we write (up to a $\mu-$negligible set) $Z = C_1 \cup C_2 \cup \cdots$ as a countable disjoint union of Borel sets with strictly positive $\mu-$measure, diameter less than $1/2$ and with the property that for any integer n the set $C_1 \cup \cdots \cup C_n$ is closed; this can be done for example taking a countable dense set $\{z_n\}$ in Z and recursively defining $C_i = \overline{B}(z_{n(i)}, 1/2) \setminus \cup_{j<i} C_j$ for the least integer $n(i)$ such that C_i is not $\mu-$negligible. Similarly, for each k we write, up to a $\mu-$negligible set, $C_k = C_{k,1} \cup C_{k,2} \cup \cdots$ as a countable disjoint union of Borel subsets with strictly positive $\mu-$measure, diameter less than $1/4$ and with the property that for any n the union

$$C_1 \cup C_2 \cup \cdots \cup C_{k-1} \cup C_{k,1} \cup C_{k,2} \cdots \cup C_{k,n}$$

is closed. Iterating, we define the Borel set C_I for any finite sequence I, with the property that any C_I has strictly positive $\mu-$measure, is the $\mu-$essential union of the sets $C_{(I,n)}$ for $n \in \mathbb{N}$ and has a diameter less than $1/2^{l(I)}$; moreover, the following property holds:

$$\begin{aligned}&\text{For any finite sequence } I=(i_1,i_2,\dots,i_k)\text{, the finite union}\\&\big(C_1\cup\cdots C_{i_1-1}\big)\cup\big(C_{i_1,1}\cup\cdots C_{i_1,i_2-1}\big)\cup\cdots\cup\big(C_{i_1,i_2,\dots,i_{k-1},1}\cup\cdots C_{i_1,i_2,\dots,i_{k-1},i_k}\big)\\&\text{is a closed set.}\end{aligned}\tag{2.2}$$

We must now define the sets Z_I and the measures μ_I for the binary sequences: to be consistent, we first set $Z_\varnothing := Z$ and $\mu_\varnothing := \mu$.

To write $\mu = \mu_0 + \mu_1$ following the requirements (i)-(viii), we let n_1 be the integer such that

$$\mu\big(C_1 \cup C_2 \cup \cdots \cup C_{n_1-1}\big) \le 1/2 \le \mu\big(C_1 \cup C_2 \cup \cdots \cup C_{n_1}\big),$$

which is well defined since $\mu(C_{n_1}) > 0$. Then we set

$$Z_0^1 := C_1 \cup C_2 \cup \cdots \cup C_{n_1-1} \qquad\qquad Z_1^1 := C_{n_1+1} \cup C_{n_1+2} \cup C_{n_1+3} \cdots$$

Then Z_0^1 and Z_1^1 are disjoint sets with $\mu(Z_0^1)$, $\mu(Z_1^1) \le 1/2$. Let now n_2 be the integer such that, writing $C_{n_1} = C_{n_1,1} \cup C_{n_1,2} \cup \cdots$ as before,

$$\mu\big(Z_0^1 \cup C_{n_1,1} \cup C_{n_1,2} \cup \cdots \cup C_{n_1,n_2-1}\big) \le 1/2 \le \mu\big(Z_0^1 \cup C_{n_1,1} \cup C_{n_1,2} \cup \cdots \cup C_{n_1,n_2}\big):$$

then we set

$$\begin{aligned}Z_0^2 &:= Z_0^1 \cup C_{n_1,1} \cup C_{n_1,2} \cup \cdots \cup C_{n_1,n_2-1},\\ Z_1^2 &:= Z_1^1 \cup C_{n_1,n_2+1} \cup C_{n_1,n_2+2} \cup C_{n_1,n_2+3} \cup \cdots,\end{aligned}$$

so that Z_0^2 and Z_1^2 are disjoint sets with $\mu(Z_0^2)$, $\mu(Z_1^2) \le 1/2$. We can then iterate, finding increasing sequences of disjoint sets Z_0^n and Z_1^n with $\mu(Z_0^n)$, $\mu(Z_1^n) \le 1/2$, and we define $\widehat{Z}_0 = \cup_n Z_0^n$ and $\widehat{Z}_1 = \cup_n Z_1^n$. Note now that $\mu \llcorner \big(Z \setminus (\widehat{Z}_0 \cup \widehat{Z}_1)\big)$ is concentrated in the intersection of the sets $C_{n_1,n_2,\dots,n_k}$, which has diameter 0 and then consists of at most a single point z. We define $Z_0 := \widehat{Z}_0 \cup \{z\}$; moreover if μ is non-atomic, or in general if $\mu(Z_0) = 1/2$, we define $Z_1 := \widehat{Z}_1$, so that $Z_0 \cap Z_1 = \emptyset$ and $\mu(Z_0) = \mu(Z_1) = 1/2$: therefore we also set $\mu_0 := \mu \llcorner Z_0$ and $\mu_1 := \mu \llcorner Z_1$. Otherwise we define $Z_1 := \widehat{Z}_1 \cup \{z\}$, $\mu_0 := \mu \llcorner \widehat{Z}_0 + \big(1/2 - \mu(\widehat{Z}_0)\big)\delta_z$ and $\mu_1 := \mu \llcorner \widehat{Z}_1 + \big(1/2 - \mu(\widehat{Z}_1)\big)\delta_z = \mu - \mu_0$. Note that by construction, recalling (2.2) and the fact that $\operatorname{diam}(C_{n_1,n_2,\dots,n_k}) \to 0$, the set Z_0 is closed.

We generalize now the preceding construction to define μ_I and Z_I for any finite binary sequence I. This is done recursively: if μ_J and Z_J are given, we repeat exactly the preceding construction replacing μ by μ_J, $1/2$ by $1/2^{l(J)+1}$ and using, instead of the sets

C_I, the intersections $\widetilde{C}_I = C_I \cap Z_J$ for all the sequences I such that $\mu_J(C_I \cap Z_J) > 0$; this leads us to find two measures, that we call respectively $\mu_{(I,0)}$ and $\mu_{(I,1)}$, and two sets, that we call $Z_{(I,0)}$ and $Z_{(I,1)}$. Again, the construction, (2.2) and the fact that the diameter of any C_J is less than $1/2^{l(J)}$ ensure that for any $k \in \mathbb{N}$ and for any $I \in \{0,1\}^k$ with $I \neq (1,1,\ldots,1)$ the finite union $\cup\{Z_J : \, l(J) = k, \, J \leq^k I\}$ is a closed set.

We have then defined the sets Z_I and the measures μ_I for any finite binary sequence I, and the properties (i), (ii), (iii), (iv), (v), (vi) and (viii) follow by construction. Finally, concerning the property (vii), note that at any step, passing from Z_I to $Z_{(I,0)}$ and $Z_{(I,1)}$, the number of sets C_k "used" becomes strictly smaller for at least one between $Z_{(I,0)}$ and $Z_{(I,1)}$ until it is only one; similarly, if Z_I is contained in only one C_k, the number of the sets $C_{k,i}$ "used" becomes strictly smaller for at least one between $Z_{(I,0)}$ and $Z_{(I,1)}$ until it is only one, and so on. Consequently, for all but countably many sequences $I \in \{0,1\}^{\mathbb{N}}$ and for any $l \in \mathbb{N}$, there is $J \in \mathbb{N}^l$ such that Z_{I_k} is definitively contained in C_J, whose diameter is less than 2^{-l}. Then also (vii) holds and the thesis follows. $\square$

After the preceding lemma, we are now ready to prove the main result, extending Theorems 2.1 and 2.2:

Theorem 2.4. *If f^+ is non-atomic there are transport maps. In addition, given any cost function c (i.e. any l.s.c. function from $X \times Y$ to $\overline{R}^+ = \mathbb{R}^+ \cup \{+\infty\}$), we can find a transport map t such that*

$$\mathfrak{C}(t) \leq \mathfrak{C}(f^+ \boxtimes f^-) \quad (\leq +\infty).$$

Proof. For simplicity (and without any loss of generality) we assume that $\|f^+\| = \|f^-\| = 1$; the idea is then to use the measures f^+_I and f^-_I defined from f^+ and f^- through the preceding lemma to define a sequence of transport plans weakly* converging to some transport map (i.e., converging to a transport plan induced by a transport map). To do that, we will define for each $k \in \mathbb{N}$ a one-to-one function $\varphi_k : \{0,1\}^k \to \{0,1\}^k$ and a measure γ_k by

$$\gamma_k := \sum\nolimits_{I \in \{0,1\}^k} f^+_I \boxtimes f^-_{\varphi_k(I)} \tag{2.3}$$

(for the definition of $\boxtimes$ see Section A.3 in the Appendix). The function φ_k will satisfy the following two requirements:

$$\mathfrak{C}(\gamma_k) \leq \mathfrak{C}(\gamma_{k-1}) \leq \cdots \leq \mathfrak{C}(\gamma_1) \leq \mathfrak{C}(\gamma_0) = \mathfrak{C}(f^+ \otimes f^-), \tag{2.4}$$

$$\forall\, 0 \leq r \leq s \leq k, \ \forall I \in \{0,1\}^s \text{ one has } \varphi_r(I_r) = \big(\varphi_s(I)\big)_r. \tag{2.5}$$

To do that we first define φ_0 as $\varphi(\varnothing) = \varnothing$ which, together with (2.3) and recalling $\|f^+\| = \|f^-\| = 1$, gives $\gamma_0 = f^+ \boxtimes f^- = f^+ \otimes f^-$: then (2.4) and (2.5) hold with $k = 0$.

To make the induction step of the construction, assume $\varphi_0,\ \varphi_1,\ \ldots,\ \varphi_k$ are given satisfying (2.4) and (2.5); to define φ_{k+1} in such a way that (2.5) continues to hold, we have 2^k choices to do: for any $I \in \{0,1\}^k$ we may set

$$\text{either} \qquad \begin{cases} \varphi_{k+1}(I,0) = \big(\varphi_k(I),0\big) \\ \varphi_{k+1}(I,1) = \big(\varphi_k(I),1\big) \end{cases} \qquad \text{or} \qquad \begin{cases} \varphi_{k+1}(I,0) = \big(\varphi_k(I),1\big) \\ \varphi_{k+1}(I,1) = \big(\varphi_k(I),0\big). \end{cases}$$

Since clearly these choices are independent and determine uniquely φ_{k+1}, there is a one-to-one correspondence between the functions $\varphi : \{0,1\}^{k+1} \to \{0,1\}^{k+1}$ satisfying (2.5) and the functions $\tau : \{0,1\}^k \to \{0,1\}$, where to any such τ we associate φ_τ given by

$$\varphi_\tau\big(I,i\big) = \begin{cases} \big(\varphi_k(I),i\big) & \text{if } \tau(I)=0 \\ \big(\varphi_k(I),1-i\big) & \text{if } \tau(I)=1 \end{cases}$$

for any $I \in \{0,1\}^k$ and $i = 0,1$. Thus, to any τ we can associate –recalling (2.3)– the measure γ_τ given by

$$\gamma_\tau := \sum\nolimits_{I\in\{0,1\}^{k+1}} f_I^+ \boxtimes f_{\varphi_\tau(I)}^-.$$

We are going now to complete the induction step selecting a particular function τ and defining $\varphi_{k+1} = \varphi_\tau$. Since (2.5) holds whichever τ we choose, we need only to take care of (2.4): but, recalling that –by property (iii) in Lemma 2.3– $f_I^+ = f_{(I,0)}^+ + f_{(I,1)}^+$ and $f_I^- = f_{(I,0)}^- + f_{(I,1)}^-$ and recalling also the linearity of the tensor product, one easily checks that

$$\gamma_k = \frac{1}{2^{2^k}} \sum\nolimits_\tau \gamma_\tau,$$

and since the possible τ are 2^{2^k} this means that γ is the arithmetic mean of the γ_τ: thus by the linearity of the cost for transport plans we can choose τ such that $\varphi_{k+1} := \varphi_\tau$ fulfills (2.4) and the inductive step is done.

Let us consider now the measures γ_k: each of them is a transport plan between f^+ and f^- thanks to properties (i) and (ii) in Lemma 2.3; define then $\varphi : \{0,1\}^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$ the unique function extending each φ_k, that is possible according to (2.5), and let γ be a weak* limit of some subsequence of $\{\gamma_k\}$, that makes sense since $\Pi(f^+,f^-)$ is weakly* closed –by the way, at the end it will be clear that the whole sequence $\{\gamma_k\}$ weakly* converges to γ. Since the cost $\mathfrak{C}$ is l.s.c. and recalling (2.4), the fact that $\mathfrak{C}(\gamma) \leq \mathfrak{C}(f^+ \otimes f^-)$ is true, thus we only need to prove that γ is induced by a transport map.

To do that, we call $X_k = \bigcup\big\{X_I : l(I) = k\big\}$, which is a set of full f^+-measure thanks to properties (i) and (iv) in Lemma 2.3, and let $\widetilde{X} = \cap_{k\in\mathbb{N}} X_k$, still a set of full f^+-measure. For any $x \in \widetilde{X}$ there is a unique countable binary sequence $\psi(x)$ such that $x \in X_{\psi(x)_k}$ for any $k \in \mathbb{N}$, recalling that f^+ is non-atomic and property (viii) in

Lemma 2.3; according with property (vii) in Lemma 2.3, we also let $A \subseteq \{0,1\}^{\mathbb{N}}$ (resp. $B \subseteq \{0,1\}^{\mathbb{N}}$) be the set, with countably many elements, of the countable binary sequences I such that $\operatorname{diam}(X_{I_k})$ (resp. $\operatorname{diam}(Y_{I_k})$) does not converge to 0 when k goes to $+\infty$. Then the set

$$\overline{X} := \left\{x \in \widetilde{X} :\ \psi(x) \notin A, \varphi(\psi(x)) \notin B\right\} \tag{2.6}$$

is still a set of full f^+-measure, since $\widetilde{X} \setminus \overline{X}$ contains countably many elements and f^+ is non-atomic. Moreover, by construction, γ_k is concentrated in

$$S_k := \bigcup \left\{X_I \times Y_{\varphi_k(I)} :\ I \in \{0,1\}^k\right\},$$

and in the same set is concentrated also each γ_h with $h \geq k$ thanks to properties (iii), (iv) and (v) in Lemma 2.3 and thanks to (2.5).

We write now, for simplicity, $\{0,1\}^k = \{0,1,2,\dots,2^k-1\}$ preserving the total order $\leq^k$ according to the definition (2.1). Since X_0 is closed by property (vi) in Lemma 2.3, then

$$\gamma\big(X_0 \times \overline{Y}_{\varphi_k(0)} \geq \limsup_{h\to+\infty} \gamma_h\big(X_0 \times Y_{\varphi_k(0)}\big).$$

But γ is a transport plan, then $\gamma\big(X_0 \times \overline{Y}_{\varphi_k(0)} \leq f^+(X_0)\big)$; on the other hand, according to our construction $\gamma_h\big(X_0 \times Y_{\varphi_k(0)}\big) = f^+(X_0)$ for any $h \geq k$. This implies that $\gamma \llcorner (X_0 \times Y)$ is concentrated in $X_0 \times \overline{Y}_{\varphi_k(0)}$: we want to infer the same property for any X_n with $n = 1, 2, \dots, 2^k-1$. The argument is similar: given any $0 < n \leq 2^k - 1$, we can write

$$\begin{aligned}\gamma\Big(X_0 \times \overline{Y}_{\varphi_k(0)}\ \cup\ \overline{X}_1 \times \overline{Y}_{\varphi_k(1)}\ \cup\ \cdots\ \cup\ \overline{X}_n \times \overline{Y}_{\varphi_k(n)}\Big) \geq& \\ \limsup_{h\to+\infty} \gamma_k\Big(X_0 \times Y_{\varphi_k(0)}\ \cup\ X_1 \times Y_{\varphi_k(1)}\ \cup\ \cdots\ \cup\ X_n \times Y_{\varphi_k(n)}\Big)&;\end{aligned}$$

but for any $h \geq k$ the number in the right is $f^+(X_0 \cup \cdots \cup X_n)$, while the number in the left is less than $f^+(X_0 \cup \overline{X}_1 \cup \cdots \cup \overline{X}_n)$; moreover, we claim that

$$f^+(X_0 \cup \overline{X}_1 \cup \cdots \cup \overline{X}_n) = f^+(X_0 \cup X_1 \cup \cdots \cup X_n) : \tag{2.7}$$

if $n < 2^k - 1$, this is true because $X_0 \cup \cdots \cup X_n$ is closed according to property (vi) in Lemma 2.3; on the other hand, if $n = 2^k - 1$ then $X_0 \cup \cdots \cup X_n$ is a set of full f^+-measure, then (2.7) holds anyhow. Summarizing, we obtain that $\gamma \llcorner \left(X_0 \cup \cdots \cup X_n\right)$ is concentrated in

$$X_0 \times Y_{\varphi_k(0)} \cup \overline{X}_1 \times \overline{Y}_{\varphi_k(1)} \cup \cdots \overline{X}_n \times \overline{Y}_{\varphi_k(n)}.$$

But the intersection of this set with $X_n \times Y$ is exactly $X_n \times \overline{Y}_{\varphi_k(n)}$, again using property (vi) in Lemma 2.3 to deduce that $X_0 \cup \cdots \cup X_{n-1}$ is closed. Then, since the different X_i are disjointed, we have finally proved that $\gamma \llcorner (X_I \times Y)$ is concentrated in $X_I \times \overline{Y}_{\varphi_k(I)}$ for any $I \in \{0,1\}^k$, then γ is concentrated in the set

$$\widetilde{S}_k := \bigcup \big\{ X_I \times \overline{Y}_{\varphi_k(I)} :\ I \in \{0,1\}^k \big\}$$

and consequently also in the set $\widetilde{S} = \cap \widetilde{S}_k$. We infer that for any $x \in \overline{X}$

$$\pi_2\Big(\widetilde{S} \cap \big(\{x\} \times Y\big)\Big) \subseteq \bigcap_{k\in\mathbb{N}} \Big\{ \overline{Y}_{\varphi_k(\psi(x)_k)} \Big\} = \bigcap_{k\in\mathbb{N}} \Big\{ \overline{Y}_{J_k} :\ J = \varphi\big(\psi(x)\big) \Big\},$$

and the latter set consists of at most a point, since the diameter of Y_{J_k} converges to 0 by property (vii) in Lemma 2.3 and recalling the definition (2.6). This implies that γ is induced by a transport map thanks to Lemma 1.6 –in particular the transport map is defined f^+–a.e. by associating to the point x the unique element of $\cap_k Y_{\psi(x)_k}$. Hence the thesis follows. □

2.2 Equality between Monge's infimum and Kantorovich's minimum

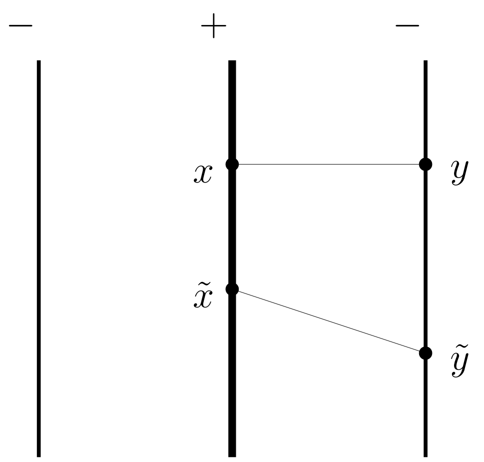
An interesting question is whether or not the equality

$$\inf(1.1) = \min(1.5) \tag{2.8}$$

holds, i.e. if the infimum of the costs of the transport maps equals the infimum –which we know thanks to Theorem 1.11 to be actually a minimum– of the costs of the transport plans. Lemma 1.7 gives the general inequality $\min(1.5) \leq \inf(1.1)$, but what about the other one? We know that in general the existence of optimal transport maps may fail, so the question is if one can find transport maps arbitrarily good, i.e. with costs converging to the minimum of the Kantorovich problem, or there is some Laurentieff phenomenon. First of all we give some examples when the equality (2.8) does not hold, after that we will prove the equality with suitable hypotheses.

Example 2.5. Let us consider once again Example 1.3: in that situation, there are no transport maps, so $\inf(1.1) = +\infty$ and then the minimum of the Kantorovich problem is strictly better, i.e. (2.8) does not hold, if $c(x, y_1) < +\infty$ and $c(x, y_2) < +\infty$.

Example 2.6. Let us consider the situation in Figure 2.1, a two dimensional version of Example 1.4: here f^+ and f^- are (up to the constant 2 for f^+) the one-dimensional Hausdorff measures on three parallel segments; it is clear that there are transport maps,



$$c_1(x,y) = \begin{cases} |y-x| & \text{if } x_2 = y_2 \\ +\infty & \text{otherwise} \end{cases}$$

$$c_2(x,y) = \begin{cases} |y-x| & \text{if } x_2 = y_2 \\ 2|y-x| & \text{otherwise} \end{cases}$$

Figure 2.1: Two examples

for example one can apply the general Theorem 2.4. Given any transport map t, however, it does not move the points horizontally: more precisely,

$$f^+\left(\left\{x : (t(x))_2 = x_2\right\}\right) = 0 \tag{2.9}$$

(argue exactly as in Example 1.4). This does not happen for the transport plans, for instance we call $\bar{\gamma}$ the transport plan moving the central segment half on the left and half on the right, thus everything horizontally: formally, if we call S the central segment and set the distance between the segments to be unitary,

$$\bar{\gamma} = \int_S \frac{\delta_{(x,x-(1,0))} + \delta_{(x,x+(1,0))}}{2}\, dx.$$

Set then

$$c(x,y) = \begin{cases} |y-x| & \text{if } x_2 = y_2 \\ +\infty & \text{otherwise.} \end{cases}$$

It is easy to note that $\mathfrak{C}(\bar{\gamma}) = 1$, while $\mathfrak{C}(t) = +\infty$ for any transport map t using (2.9). Then again (2.8) does not hold.

Example 2.7. In the preceding example, the true obstacle to the equality (2.8) was not the $+\infty$ in c and in $\mathfrak{C}$, but the non continuity of c. In fact, let us consider the same geometry but set

$$c(x,y) = \begin{cases} |y-x| & \text{if } x_2 = y_2 \\ 2|y-x| & \text{otherwise.} \end{cases}$$

Then it is still true that $\mathfrak{C}(\bar{\gamma}) = 1$, but all the transport maps have finite cost; however, recalling again (2.9), it is trivial to note that $\inf(1.1) = 2 > 1 = \min(1.5)$. Then it becomes clear that the lower semicontinuity of c is not sufficient if one wants to prove the equality (2.8), but the "natural" condition on c is the continuity.

After the preceding examples, one could think that the only importance of the non-atomicity hypothesis is to obtain existence of transport maps, and that maybe the equality (2.8) is true whenever c is continuous and there exist optimal transport maps. We give then this last

Example 2.8. Let us consider $X = Y = \mathbb{R}^2$ and take $f^+ = \delta_{x^+} + \mathscr{L} \llcorner B^+$ and $f^- = \delta_{x^-} + \mathscr{L} \llcorner B^-$ where, as in figure 2.2, x^+ and x^- are two points in and B^+ and B^- are two balls of equal volume. If we take $c(x,y) = |y-x|$ and if, as in the figure, x^+ and B^- are close to each other but far from x^- and B^+, it is clear that the unique optimal transport plan $\bar{\gamma}$ is the one moving the mass of x^+ on B^- and that of B^+ on x^-. On

Figure 2.2: Example 2.8

the other hand, there are also transport maps, but they need to move x^+ on x^- and B^+ on B^-, so their cost is finite but much greater than that of $\bar{\gamma}$. In particular, the Monge problem admits a solution, but (2.8) does not hold.

The examples we presented suggest that, in order to prove the validity of (2.8), it is essential to assume the continuity of c (Examples 2.6 and 2.7) and the non atomicity of f^+ (Example 2.8). We claim that these two assumptions are sufficient, and the remaining part of this chapter will be devoted to show this claim:

Theorem 2.9. *Assume that f^+ is non atomic and c is continuous: then (2.8) holds.*

Proof. If $\min(1.5) = +\infty$, the equality (2.8) is clearly true; otherwise, we fix a metric on X and on Y, take an optimal transport plan γ (so $\mathfrak{C}(\gamma) < +\infty$) and fix $\varepsilon > 0$: we want to find a transport map t such that $\mathfrak{C}(t) \leq \mathfrak{C}(\gamma) + \varepsilon \|\gamma\|$. This will be done in three steps:

Step I. *Preliminary definition of A_I, M_I, m_I.*
Let $K := \{(x,y) \in X \times Y : c(x,y) = +\infty\}$, which is a closed set thanks to the continuity of the cost function c; then, for any set $S \subseteq X \times Y$ not completely contained in K, we call *range of S* the number

$$rg(S) := \sup\{c(x,y) : (x,y) \in S\} - \inf\{c(x,y) : (x,y) \in S\} \in \mathbb{R}^+ \cup \{+\infty\}.$$

Take now two sequences $\{x_i\}$ and $\{y_i\}$ dense in X and Y respectively; for any $(i,j) \in \mathbb{N}^2$ such that $c(x_i,y_j) < +\infty$, let r be the maximum real number such that $B_X(x_i,r) \times B_Y(y_j,r)$ (the product of a ball in X and one in Y) have range less than ε: notice that $r > 0$ thanks to the continuity of c, then define also

$$C_{i,j} := B(x_i,r) \qquad\qquad D_{i,j} := B(y_j,r).$$

Notice that, thanks to the definition and the density of the sequences $\{x_i\}$ and $\{y_i\}$,

$$\bigcup_{i,j\in\mathbb{N}} C_{i,j} \times D_{i,j} = X \times Y \setminus K;$$

therefore, since γ has finite cost and thus is concentrated in $(X \times Y) \setminus K$, it follows

$$\gamma\Big((X \times Y) \setminus \bigcup_{i,j\in\mathbb{N}} C_{i,j} \times D_{i,j}\Big) = 0. \tag{2.10}$$

We want to define now the sets A_I and B_I for any finite sequence I. Let us begin with the sequences of length 1: first of all we set

$$A_1 := C_{i,j} \qquad\qquad B_1 := D_{i,j}$$

where the pair (i,j) is chosen such that $\gamma(A_1 \times B_1)$ is almost maximal, in the sense that $\gamma(A_1 \times B_1) \geq \gamma(C_{m,n} \times D_{m,n})/2$ for any (m,n). Then we proceed by induction: if A_h and B_h have been defined for any $h < k$, we set

$$A_k := C_{i,j} \setminus \bigcup_{h=1}^{k-1} A_h \qquad\qquad B_k := D_{i,j},$$

where again the pair (i,j) is such that $\gamma(A_k \times B_k)$ is almost maximal. Then we have defined A_k and B_k for any integer k, and we know that $A_i \cap A_j = \emptyset$ if $i \neq j$. We claim also that

$$f^+(X \setminus \cup_k A_k) = 0; \tag{2.11}$$

if not, since the sets $C_{i,j} \times D_{i,j}$ cover $R = \big((X \setminus \cup_k A_k) \times Y\big) \setminus K$ and $\gamma(R) = f^+(X \setminus \cup_k A_k) > 0$ by (2.10), there would be a pair (m,n) such that $\gamma\big((C_{m,n} \times D_{m,n}) \cap R\big) = \delta > 0$. This implies that $\gamma(A_k \times B_k) \geq \delta/2$ for any k, otherwise one would have not chosen A_k and B_k in that way; but then, since the different A_i are disjoint, it follows $\gamma(\cup_k A_k \times B_k) = +\infty$ and this is impossible: then (2.11) is proved. Finally, we define $\gamma_0 := \gamma$ and

$$\gamma_1 := \gamma \llcorner \Big(\big(X \times Y\big) \setminus \big(\cup_k A_k \times B_k\big)\Big).$$

Let us consider now the sequences of length 2: given $m \in \mathbb{N}$ we define

$$A_{m,1} := C_{i,j} \cap A_m \qquad\qquad B_{m,1} := D_{i,j} \setminus B_m$$

for a pair (i,j) such that $\gamma_1(A_{m,1}\times B_{m,1})$ is almost maximal, as before; then, by induction, we set also

$$A_{m,k} := \big(C_{i,j}\cap A_k\big)\setminus \bigcup_{h=1}^{k-1} A_{m,h} \qquad\qquad B_{k,1} := D_{i,j}\setminus B_k$$

for an almost maximizing pair (i,j); we have then all the sets $A_{m,k}$ and $B_{m,k}$, and moreover $A_{m,k}\subseteq A_m$ and $B_{m,k}\cap B_m=\emptyset$. With the same argument as before, (2.11) can be improved with

$$\pi_1\gamma_1(X\setminus \cup_{i,j}A_{i,j}) = 0. \tag{2.12}$$

Finally we define, as before,

$$\gamma_2 := \gamma \llcorner \Big(\big(X\times Y\big)\setminus \big(\bigcup_{I:\, l(I)\le 2} A_I\times B_I\big)\Big).$$

With an obvious induction argument, we define all the sets A_I and B_I for finite sequences I, we set

$$\gamma_p := \gamma \llcorner \big(X\times Y\big)\setminus \big(\bigcup_{I:\, l(I)\le p} A_I\times B_I\big), \tag{2.13}$$

and (2.12) can be generalized as

$$\pi_1\gamma_{k-1}(X\setminus \bigcup_{l(I)=k} A_I) = 0. \tag{2.14}$$

We also claim that

$$\forall\, I,\ rg(A_I\times B_I) < \varepsilon; \tag{2.15}$$
$$I\le J \Longrightarrow A_I\supseteq A_J; \tag{2.16}$$
$$A_I\cap A_J=\emptyset \text{ unless } I\le J \text{ or } J\le I; \tag{2.17}$$
$$B_I\cap B_J=\emptyset \text{ if } I<J \text{ or } J<I; \tag{2.18}$$
$$\gamma\Big(X\times Y\setminus \bigcup_I \big(A_I\times B_I\big)\Big) = 0. \tag{2.19}$$

Properties (2.15)-(2.18) are clear from the construction, so let us show (2.19): recalling (2.10), if (2.19) is not true, we can find a pair (i,j) with $\gamma(\Lambda)=\delta>0$ where

$$\Lambda := \big(C_{i,j}\times D_{i,j}\big)\setminus \bigcup_I \big(A_I\times B_I\big);$$

but then, since the pair (i,j) was not chosen at the first step, it follows

$$\gamma(A_1\times B_1)\ge \delta/2,$$

then a fortiori

$$\gamma\Big(\bigcup_{I:\, l(I)=1} A_I\times B_I\Big) \ge \delta/2. \tag{2.20}$$

Let us define now $\delta_k = \gamma\big(\Lambda \cap (A_k \times Y)\big)$: thanks to (2.11), one has $\sum \delta_k = \delta$; then as before, since the pair (i,j) was not chosen to define $A_{k,1}$ and $B_{k,1}$ and since clearly $\gamma(\Lambda) = \gamma_1(\Lambda)$, it means that

$$\gamma_1(A_{k,1} \times B_{k,1}) = \gamma(A_{k,1} \times B_{k,1}) \geq \delta_k/2;$$

adding the preceding inequality for $k \in \mathbb{N}$, we derive that a fortiori

$$\gamma\Big(\bigcup\nolimits_{I:\, l(I)=2} A_I \times B_I\Big) \geq \delta/2. \tag{2.21}$$

Recalling that (2.11) holds for any length, according to (2.14), it is trivial to repeat the preceding argument generalizing (2.20) and (2.21) to any length. Since it would follow that $\|\gamma\| = +\infty$, the absurd shows the validity of (2.19).

Now we set

$$M_I := f^+(A_I), \qquad\qquad m_I := \gamma(A_I \times B_I); \tag{2.22}$$

it is clear that $m_I \leq M_I \;\forall\, I$; in particular,

$$M_I \geq \sum \{m_J : J \geq I\}, \tag{2.23}$$

$$M_I = \sum \{m_J : J \geq I\} \text{ if } l(I) = 1, \tag{2.24}$$

$$\sum\nolimits_I m_I = \|\gamma\| : \tag{2.25}$$

the first two properties follow recalling (2.14) and (2.17), while the third one is an immediate consequence of (2.19).

Step II. *Inductive definition of the sets X_I^p.*
We want now to define, by means of an inductive construction, the sets X_I^p for each finite sequence I and each integer $p \geq l(I)$; in particular, we will build $X_I^{l(I)}$ with

$$f^+\big(X_I^{l(I)}\big) = m_I, \tag{2.26}$$

then we will build $X_I^{p,+}$, $X_I^{p,-}$ for $p \geq l(I)$, and finally we will take as a definition for X_I^p with $p > l(I)$ the following:

$$X_I^p := X_I^{l(I)} \bigcup \Big(\bigcup\nolimits_{l(I) \leq s < p} X_I^{s,+}\Big) \setminus \Big(\bigcup\nolimits_{l(I) \leq s < p} X_I^{s,-}\Big). \tag{2.27}$$

The sets $X_I^{p,\pm}$ will satisfy the following properties:

For any p, the sets X_I^{p+1} with $l(I) \leq p+1$ are disjoint w.r.t. f^+; (2.28) for any $l(I) \le p+1$, $X_I^{p+1} \subseteq A_I$; (2.29)

for any $l(I) \le p+1$, $f^+(X_I^{p+1}) = m_I$; (2.30)

for any $l(I) \le p$, $X_I^{p,+} \cap \left(X_I^p \cup \bigcup \left\{X_J^{p+1} : \ l(I) < l(J) \le p+1\right\}\right) = \emptyset$; (2.31)

for any $l(I) \le p$, $X_I^{p,-} = X_I^p \cap \bigcup \left\{X_J^{p+1} : \ l(I) < l(J) \le p+1\right\}$; (2.32)

for any $l(I) \le p$, $f^+(X_I^{p,+}) = f^+(X_I^{p,-})$; (2.33)

for any $l(I) \le p$, $f^+(X_I^{p,-}) \le \sum_{l(J)=p+1} m_J$. (2.34)

Let us begin the construction: first of all, for each index $i \in \mathbb{N}$ we choose arbitrarily a set $X_i^1 \subseteq A_i$ with mass $f^+(X_i^1) = m_i$ –so that (2.26) holds: this is possible since $m_i \le M_i$ and f^+ is non atomic. Then we choose arbitrarily a set $X_{i,j}^2 \subseteq A_{i,j}$ for each $(i,j) \in \mathbb{N}^2$ with mass $f^+(X_{i,j}^2) = m_{i,j}$, again according to (2.26). Now we want to define X_i^2 still being contained in A_i and with mass m_i, but non intersecting any of the $X_{i,j}^2$: to do that, set

$$X_i^{1,-} := X_i^1 \cap \left(\bigcup\nolimits_{j\in\mathbb{N}} X_{i,j}^2\right);$$

since $m_i + \sum_j m_{i,j} \le M_i$ thanks to (2.23) and

$$f^+\left(X_i^1 \cup \left(\cup_j X_{i,j}^2\right)\right) = m_i + \sum\nolimits_{j\in\mathbb{N}} m_{i,j} - f^+(X_i^{1,-}),$$

we can choose $X_i^{1,+} \subseteq A_i$ with $f^+(X_i^{1,+}) = f^+(X_i^{1,-})$ non intersecting X_i^1 nor any of the $X_{i,j}^2$. Finally set $X_i^2 = X_i^1 \cup X_i^{1,+} \setminus X_i^{1,-}$ according to (2.27). One can easily note that the equations (2.28)–(2.34) hold for $p=1$; then we start the induction.

Let us assume, after the $n-$th step of the construction, the existence of sets $X_I^{l(I)}$ for the finite sequences I of length $l(I) \le n+1$ and of sets $X_I^{p,\pm}$ for $l(I) \le p \le n$ such that using the definition (2.27) the equations (2.28)–(2.34) hold for each $p \le n$. We have to define the sets X_I^{n+2} fulfilling (2.26) for $l(I) = n+2$, and the sets $X_I^{n+1,\pm}$ for $l(I) \le n+1$ in such a way that the equations (2.28)–(2.34) hold with $p = n+1$.

First of all we choose arbitrarily the sets $X_I^{n+2} \subseteq A_I$ for the sequences I of length $n+2$ with mass $f^+(X_I^{n+2}) = m_I$. Then, to define $X_I^{n+1,\pm}$, we will proceed recursively, starting from the sequences I of length $n+1$ and arriving until the sequences of length 1.

Let us do that taking care only that (2.29), (2.30), (2.31), (2.32) and (2.33) hold: if $1 \le s \le n+1$ and we have already done it for all the sequences I of length $s < l(I) \le n+1$, we consider now the sequences of length s. We take (2.32) with $p = n+1$ as definition for $X_I^{n+1,-}$, and we note that to define $X_I^{n+1,+}$, we have just to choose a set of f^+-measure equal to $f^+(X_I^{n+1,-})$ contained in A_I and non intersecting X_I^{n+1} nor any of the X_J^{n+2} for $l(I) < l(J) \le n+2$: in fact, if we do that then (2.29), (2.30), (2.31) and (2.33) hold. We can then choose that set $X_I^{n+1,+}$ if "there is enough space", i.e. if

$$f^+\left(A_I \setminus \left(X_I^{n+1} \cup \bigcup\{X_J^{n+2}:\ l(I) < l(J) \leq n+2\}\right)\right) \geq f^+(X_I^{n+1,-}). \tag{2.35}$$

But recall that, given J such that $l(I) < l(J) \leq n+2$, $A_I \cap X_J^{n+2} = \emptyset$ if $J \not\geq I$, by (2.17) and (2.29), which we already know to hold since $l(J) > s$. Moreover recalling (2.30), which holds for $p = n$ and for $p = n+1$ and $l(J) > s$, it is

$$\begin{aligned} f^+ &\left(A_I \setminus \left(X_I^{n+1} \cup \bigcup\{X_J^{n+2}:\ l(I) < l(J) \leq n+2\}\right)\right) = \\ &\qquad = M_I - \left(\sum\left\{m_J:\ J \geq I,\ l(J) \leq n+2\right\} - f^+(X_I^{n+1,-})\right). \end{aligned}$$

Recalling (2.23), then, (2.35) holds. Thus the step of the sequences of length s is done and the we have recursively defined all the $X_I^{n+1,\pm}$ satisfying (2.29), (2.30), (2.31), (2.32) and (2.33). We have then only to prove that also (2.28) and (2.34) hold.

Regarding (2.28), take $I \neq J$: we must prove that $X_I^{n+2} \cap X_J^{n+2} = \emptyset$. If $l(I) = l(J)$ then $A_I \cap A_J = \emptyset$ by (2.17), thus the non intersection of X_I^{n+2} and X_J^{n+2} follows from (2.29); otherwise we can assume $l(J) > l(I)$, and then the validity of (2.28) with $p = n$ together with (2.32), (2.31) and the definition (2.27) give the searched non intersection.

Concerning (2.34), consider that (2.32) can be read as

$$X_I^{n+1,-} = X_I^{n+1} \cap \left(\bigcup\nolimits_{l(J)=n+2} X_J^{n+2} \cup \bigcup\nolimits_{l(I)<l(J)<n+2} X_J^{n+2}\right). \tag{2.36}$$

But if $l(J) < n+2$ then $X_J^{n+2} = X_J^{n+1} \cup X_J^{n+1,+} \setminus X_J^{n+1,-}$; on the other hand, each $X_J^{n+1,-}$ is contained in

$$\bigcup\nolimits_{l(J)=n+2} X_J^{n+2} \cup \bigcup\nolimits_{l(J)<l(K)<n+2} X_K^{n+1,+},$$

and then (2.36) can be rewritten as

$$X_I^{n+1,-} = X_I^{n+1} \cap \left(\bigcup\nolimits_{l(J)=n+2} X_J^{n+2} \cup \bigcup\nolimits_{l(I)<l(J)<n+2} X_J^{n+1} \cup \bigcup\nolimits_{l(I)<l(J)<n+2} X_J^{n+1,+}\right).$$

But we know the validity of (2.28) for $p = n$, so $X_I^{n+1} \cap X_J^{n+1} = \emptyset$ for $I \neq J$, then we can also write

$$X_I^{n+1,-} = X_I^{n+1} \cap \left(\bigcup\nolimits_{l(J)=n+2} X_J^{n+2} \cup \bigcup\nolimits_{l(I)<l(J)<n+2} X_J^{n+1,+}\right).$$

Now, consider that the intersection between X_I^{n+1} and each $X_J^{n+1,-}$ with $l(I) < l(J) < n+2$ is empty since X_I^{n+1} does not intersect X_J^{n+1} which contains $X_J^{n+1,-}$. Then we can further change (2.36) writing

$$\begin{aligned} X_I^{n+1,-} &= X_I^{n+1} \cap \left(\bigcup\nolimits_{l(J)=n+2} X_J^{n+2} \cup \bigcup\nolimits_{l(I)<l(J)<n+2} X_J^{n+1,+} \Big\backslash \bigcup\nolimits_{l(I)<l(J)<n+2} X_J^{n+1,-}\right) = \\ &= X_I^{n+1} \cap (A \cup B \setminus C) = X_I^{n+1} \cap D \end{aligned}$$

Finally, note that A and B are disjoint sets by (2.31), while C is contained in $A \cup B$ by (2.28) and (2.32), thus $f^+(D) = f^+(A) + f^+(B) - f^+(C)$. But $f^+(B) = f^+(C)$ since both B and C are union of disjoint sets and thanks to (2.33); we infer, by (2.30),

$$f^+\left(X_I^{n+1,-}\right) \leq f^+(D) = f^+(A) = \sum\nolimits_{l(J)=n+2} m_J,$$

thus the validity of (2.34).

Step III. *Conclusion.*
Fix $p \in \mathbb{N}$ and note that, recalling the definition (2.22) and (2.13), it is

$$\rho_p := \sum\nolimits_{l(I)=p} m_I = \|\gamma_p - \gamma_{p-1}\|.$$

but since γ_p is a decreasing sequence of measures, strongly converging to 0 thanks to (2.19), it follows

$$\sum\nolimits_{p\in\mathbb{N}} \rho_p = \|\gamma\| < +\infty. \tag{2.37}$$

Then (2.34) gives $f^+(X_I^{p,-}) \leq \rho_p$; so, since by (2.37) one has $\sum_{p=m}^{+\infty} \rho_p \to 0$ when $m \to +\infty$, the property (2.33) and the definition (2.27) ensure that, for each finite sequence I, the sequence X_I^m strongly converges in the f^+−norm to some set X_I, which is contained in A_I thanks to (2.29), and such that $f^+(X_I) = m_I$ thanks to (2.30). In addition, (2.28) implies that the X_I's are essentially disjoint sets (with respect to f^+).

Let us now define

$$\gamma_I := \gamma \llcorner (A_I \times B_I)\,,$$

so that $\gamma = \sum_I \gamma_I$ by (2.18) and (2.19) and then by linearity

$$\mathfrak{C}(\gamma) = \sum\nolimits_I \mathfrak{C}(\gamma_I). \tag{2.38}$$

Consider now the measure

$$\tilde{\gamma}_I := f^+ \llcorner X_I \boxtimes \pi_2\gamma_I\,:$$

this is a transport plan between $f^+ \llcorner X_I$ and $\pi_2\gamma_I$, thanks to the definition (2.22) and the fact that $f^+(X_I) = m_I$. Notice also that $\|\tilde{\gamma}_I\| = \|\gamma_I\| = m_I$ and then, recalling (2.15),

$$\left|\mathfrak{C}\left(\tilde{\gamma}_I\right) - \mathfrak{C}\left(\gamma_I\right)\right| \leq m_I\,\varepsilon\,: \tag{2.39}$$

in fact, both $\mathfrak{C}(\gamma_I)$ and $\mathfrak{C}(\tilde{\gamma}_I)$ are greater than $m_I \inf\{c(x,y) :\ (x,y) \in A_I \times B_I\}$ and less than $m_I \sup\{c(x,y) :\ (x,y) \in A_I \times B_I\}$.

We apply now Theorem 2.4 to find a transport plan t_I between $f^+ \llcorner X_I$ and $\pi_2\gamma_I$ such that, recalling (2.39)

$$\mathfrak{C}(t_I) \leq \mathfrak{C}(\tilde{\gamma}_I) \leq \mathfrak{C}(\gamma_I) + m_I\,\varepsilon. \tag{2.40}$$

Since the sets X_I are essentially disjoint (w.r.t. f^+), it is well defined the transport map t which equals t_I on each set X_I, and this map t is a transport between $\sum_I f^+ \llcorner X_I$ and $\sum_I \pi_2\gamma_I$. But the first measure is f^+ since the sets X_I are disjoint and cover X up to a f^+-negligible set (recall (2.25)); on the other hand, the second measure is $\pi_2(\sum_I \gamma_I) = \pi_2\gamma = f^-$. Thus t is a transport map between f^+ and f^- and its cost, according to (2.40) and (2.38), is

$$\mathfrak{C}(t) \leq \mathfrak{C}(\gamma) + \varepsilon \|\gamma\| :$$

this finishes the proof. $\square$

Chapter 3

Cyclical monotonicity, duality and Kantorovich potential

In this chapter we present and study the notion of $c-$cyclical monotonicity, which is a necessary and –under suitable assumptions– sufficient condition for optimal transport plans. We also introduce and discuss a dual problem, which leads to the important concept of Kantorovich potential.

3.1 Cyclical monotonicity and necessity for c continuous

In this section we will discuss the property of the cyclical monotonicity and we will see that it is a necessary condition for the optimality of the transport plans in the case when c is continuous. First of all, we give the

Definition 3.1. *A set $S \subseteq X \times Y$ is said* $c-$cyclically monotone *if, given any $k \in \mathbb{N}$, any finite set of pairs $(x_1, y_1), \ldots, (x_k, y_k) \in S$ and any permutation $\sigma \in \mathfrak{S}_k$, the condition*

$$\sum_{i=1}^{k} c(x_i, y_i) \leq \sum_{i=1}^{k} c(x_i, y_{\sigma(i)}) \tag{3.1}$$

holds.

This property is a generalization of the cyclical monotonicity introduced by Rockafellar in [65], and it was first considered by Knott and Smith in the paper [57]; a detailed discussion can be found in [47]. The aim of this section is to prove that the support of any optimal transport plan needs to be $c-$cyclically monotone at least when c is continuous; the assert for a general cost function c, which is still true, will be deduced later, in Section 3.3. Note that this fact, that the supports of the optimal transport plans are $c-$cyclically monotone, can be true only with the hypothesis that the min in (1.5) is finite, since otherwise any admissible transport plan is optimal, regardless of any property of its support. This result with the continuity hypothesis on c is well-known (see for example [47], but also [65] and [57]).

Proposition 3.2. *If c is continuous and* $\min(1.5) < +\infty$*, then the support of any optimal transport plan is $c-$cyclically monotone.*

Proof. Let us suppose that the assert is false: then there would exist an optimal transport plan γ, a finite set of pairs $(x_1, y_1), \dots, (x_k, y_k) \in \operatorname{spt}\gamma$ and a permutation $\sigma \in \mathfrak{S}_k$ such that

$$\rho_\sigma\big((x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)\big) := \sum_{i=1}^{k} c(x_i, y_i) - c(x_i, y_{\sigma(i)}) > 0$$

(recall that $c : X \times Y \to [0, +\infty]$); we assume, without loss of generality, that $\sigma(i) \neq i$ for any $i = 1, \dots, k$. The function $\rho_\sigma : (X \times Y)^k \to \mathbb{R}$ is continuous by construction, since c is continuous: then, recalling also that each (x_i, y_i) belongs to $\operatorname{spt}\gamma$, there are neighborhoods U_i and V_i of the points x_i and y_i respectively such that

$$\rho_\sigma \llcorner \Big(\big(U_1 \times V_1\big) \times \dots \times \big(U_k \times V_k\big)\Big) > 0 \tag{3.2}$$

and for any $i = 1, \dots, k$

$$\gamma(U_i \times V_i) = \varepsilon > 0.$$

Set now $\gamma_i = \gamma \llcorner \big(U_i \times V_i\big)$: then $\gamma - \gamma_i$ is a positive measure, and the same is true for $\gamma - \big(\sum_{i=1}^k \gamma_i\big)/k$, where we divide by k because the sets $U_i \times V_i$ might intersect each other. Defining also f_i^+ and f_i^- the marginals of the measure γ_i, the mass of any measure f_i^+ and any f_i^- is the same, ε. We define then (recall Section A.3)

$$\tilde{\gamma}_i := f_i^+ \boxtimes f_{\sigma(i)}^-,$$

a positive measure with marginals f_i^+ and $f_{\sigma(i)}^-$. Then,

$$\tilde{\gamma} = \gamma - \frac{\sum_{i=1}^k \gamma_i - \sum_{i=1}^k \tilde{\gamma}_i}{k}$$

is a positive measure, and its marginal are f^+ and f^-, since the sum of the marginals of the γ_i's and of those of the $\tilde{\gamma}_i$'s are the same. By construction, and recalling that $\sigma(i) \neq i$ for any i, one can check that

$$\begin{aligned}
\mathfrak{C}(\tilde{\gamma}) - \mathfrak{C}(\gamma) &= -\frac{1}{k} \iint_{X \times Y} c(x, y)\, d\big(\gamma_1 + \dots + \gamma_k - \tilde{\gamma}_1 - \dots - \tilde{\gamma}_k\big)(x, y) \\
&= -\frac{1}{k} \int_{(X \times Y)^k} \rho_\sigma\big((x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)\big)\, d\big(\gamma_1 \boxtimes \gamma_2 \boxtimes \dots \boxtimes \gamma_k\big);
\end{aligned}$$

but $\gamma_1 \boxtimes \cdots \boxtimes \gamma_k$ is a positive measure concentrated in $\big(U_1 \times V_1\big) \times \cdots \times \big(U_k \times V_k\big)$, then by (3.2) we find $\mathfrak{C}(\tilde{\gamma}) < \mathfrak{C}(\gamma)$, which is not possible because γ is optimal. The thesis follows. $\square$

Corollary 3.3. *If c is continuous and* $\min(1.5) < +\infty$*, then also the union of the supports of the optimal transport plans is $c-$cyclically monotone.*

Proof. Let us denote by Γ the union of the supports of the optimal transport plans, let $(x_1, y_1), \dots, (x_k, y_k)$ be a finite number of pairs in Γ, and let $\sigma \in \mathfrak{S}_k$ be a permutation. By definition, for any $i = 1, \dots, k$ there is an optimal plan γ_i such that $(x_i, y_i) \in \operatorname{spt} \gamma_i$; let moreover $\gamma = \big(\gamma_1 + \cdots + \gamma_k\big)/k$: it is an optimal transport plan, since $\Pi'(f^+, f^-)$ is, as shown in Theorem 1.11, a convex subset of $\mathcal{M}^+(X \times Y)$. Then the support of γ, which contains each (x_i, y_i), is $c-$cyclically monotone, thus (3.1) holds thanks to Proposition 3.2 and then the thesis follows. $\square$

3.2 Duality formulation, $c-$concavity, $c-$superdifferential and $c-$transform

We present here a maximization problem whose solution turns out to be related to that of the mass transport problem; as the generalization of the original Monge problem, it was first proposed by Kantorovich in the case $c(x, y) = |y - x|$ (see [53, 54]). One can find many discussions about this problem, among them we refer for example to [37], [2], [74], [57] and [64]. The problem is the following:

$$\sup\left\{\int_X u(x)\, df^+(x) + \int_Y v(y)\, df^-(y) :\ (u, v) \in \mathcal{A}\right\}, \tag{3.3}$$

where $\mathcal{A} = \{(u, v) \in L^1_{f^+}(X) \times L^1_{f^-}(Y) :\ u(x) + v(y) \leq c(x, y)\}$. The relationship with the Kantorovich problem is the fact that $\sup(3.3) = \min(1.5)$, that is what we want to prove in this section.

First of all, we need some technical tools, namely the $c-$concavity and the $c-$superdifferential of functions $u : X \to \overline{R}$. To begin, we recall the classical definition of superdifferential:

Definition 3.4. *If $X \subseteq \mathbb{R}^N$ is an open convex set, given $u : X \to \overline{R}$ one defines the* superdifferential $\partial^\bullet u \subseteq X \times \mathbb{R}^N$ *as*

$$\partial^\bullet u := \left\{(x, y) \in X \times \mathbb{R}^N : u(x) \in \mathbb{R},\ u(x + \xi) \leq u(x) + \langle \xi, y \rangle + o(\xi)\right\}.$$

Moreover, y is said a supergradient *of u at x if $(x, y) \in \partial^{\bullet} u$, and u is said* superdifferentiable *at x if there is some supergradient of u at x.*

A property showing the connection between the superdifferential and the concavity is given by the following

Lemma 3.5. *If $X \subseteq \mathbb{R}^N$ is an open convex and $u : X \to \overline{R}$ is concave, then $u(\xi)-\langle \xi, y\rangle \leq u(x) - \langle x, y\rangle$ for any $\xi \in X$ if and only if $(x, y) \in \partial^{\bullet} u$.*

Proof. Take $(x, y) \in \partial^{\bullet} u$ and $\xi \in X$, and define $\alpha = u(\xi) - u(x) - \langle \xi - x, y\rangle$ so that we need to show $\alpha \leq 0$. The concavity of u gives

$$\begin{aligned} u\big(x + \varepsilon\,(\xi - x)\big) &= u\big((1-\varepsilon)\,x + \varepsilon\xi\big) \geq (1-\varepsilon)\,u(x) + \varepsilon\,u(\xi) \\ &= (1-\varepsilon)\,u(x) + \varepsilon\,u(x) + \varepsilon\,\langle \xi - x, y\rangle + \varepsilon\,\alpha = u(x) + \varepsilon\,\langle \xi - x, y\rangle + \varepsilon\alpha; \end{aligned}$$

from Definition (3.4) we infer $\varepsilon\,\alpha \leq o(\varepsilon)$, then $\alpha \leq 0$ as wanted and an implication is shown. The other one is trivial. □

We generalize now the notion of concavity to the case of a general space X, and by means of the cost function c: recalling that a concave function is the inferior envelope of a family of affine functions (for example, all those greater than the function itself), we give the following

Definition 3.6. *A function $u : X \to \overline{R} \setminus \{+\infty\}$ whose value is not constantly $-\infty$ is said to be $c-$*concave *if there is a set $\mathfrak{F} \subseteq Y \times \mathbb{R}$ such that*

$$u(x) = \inf_{(y,\lambda)\in\mathfrak{F}} c(x, y) + \lambda;$$

Remark 3.7. *If $X = Y \subseteq \mathbb{R}^N$ and $c(x, y) = |y - x|^2$, then an easy calculation shows that u is $c-$concave if and only if $u(x) - |x|^2$ is concave, i.e. if and only if u is semiconcave with constant less than 1.*

One can extend also the notion of superdifferential in order to take into account the cost function c; it is standard (see for example [57]) to take a $c-$concave function $u : X \to \overline{R} \setminus \{+\infty\}$ and to define the *$c-$superdifferential* $\partial^c u$ as the set

$$\partial^c u = \{(x, y) \in X \times Y : u(x) \in \mathbb{R},\ \forall\,\xi \in X,\ u(\xi) - c(\xi, y) \leq u(x) - c(x, y)\}\,, \qquad (3.4)$$

but we will give a slightly more general definition in a moment. Before doing that, we give the following

Definition 3.8. *Given $u : X \to \overline{R}$, the $c-$*transform *$u^c : Y \to \overline{R}$ of u is*

$$u^c(y) := \inf\left\{c(\xi, y) - u(\xi) : \xi \in X,\ u(\xi) < +\infty\right\}. \tag{3.5}$$

Similarly, given $v : Y \to \overline{R}$, one defines $v^c : X \to \overline{R}$. A pair of functions (u, v), where $u : X \to \overline{R}$ and $v : Y \to \overline{R}$, is said then a dual pair *if $u^c = v$ and $v^c = u$.*

Putting together (3.4) and (3.5), we can rewrite the first one as

$$\partial^c u = \{(x, y) \in X \times Y : u(x) \in \mathbb{R},\ u(x) + u^c(y) = c(x, y)\} :$$

in fact, $(x, y) \in \partial^c u$ if and only if $u(x) \in \mathbb{R}$ and $u(x) + u^c(y) \geq c(x, y)$; on the other hand, for any pair (x, y) such that $u(x) \in \mathbb{R}$ it is $u(x) + u^c(y) \leq c(x, y)$ by (3.5). We give then the following notion of $c-$superdifferential:

Definition 3.9. *Given a function $u : X \to \overline{R}$, we define the $c-$superdifferential of u as the set*

$$\partial^c u := \{(x, y) \in X \times Y : u(x) \in \mathbb{R},\ u(x) + u^c(y) = c(x, y)\}. \tag{3.6}$$

The reason of this modified definition is to allow also functions u which are not $c-$concave: this will be fundamental, as we will discuss in Remark 3.26.

It is to be noticed that if $X \subseteq \mathbb{R}^N$ is an open convex set and $c(x, y) = \langle x, y\rangle$, the notions of $c-$concavity, $c-$superdifferential and $c-$transform reduce to the classical concavity, superdifferential and Fenchel-Moreau transform. Note also that the $c-$superdifferential of a function u is a subset of $X \times Y$, and we already gave a notion for subsets of $X \times Y$, namely the $c-$cyclical monotonicity: we prove now a fundamental result, which states a connection between $c-$cyclical monotonicity and $c-$superdifferential:

Theorem 3.10. *Take a set $\Gamma \subseteq X \times Y$ such that $\Gamma \subseteq \{(x, y) : c(x, y) < +\infty\}$ and there exists $y_0 \in \pi_2(\Gamma)$ with $c(x, y_0) < +\infty$ for any $x \in X$. Then Γ is $c-$cyclically monotone if and only if there is a $c-$concave function u such that $\Gamma \subseteq \partial^c u$.*

Note that the preceding statement is much simpler in the case when the cost function is real-valued, since in that case one directly says that $\Gamma \subseteq X \times Y$ is $c-$cyclically monotone is and only if it is contained in some superdifferential. However, we handle this more general case because we are deeply interested in cost functions possibly assuming the value $+\infty$.

Proof. (of Theorem 3.10): **Sufficiency:** Suppose $\Gamma \subseteq \partial^c u$ for some function $u : X \to \overline{R}$. Then take a finite number of pairs $(x_1, y_1), \ldots, (x_k, y_k) \in \Gamma$ and a permutation $\sigma \in \mathfrak{S}_k$:

recalling that $c(x_i, y_i) = u(x_i) + u^c(y_i)$ and applying (3.5) to y_i with $\xi = x_{\sigma(i)}$ (note that $u(x_{\sigma(i)}) < +\infty$ because $(x_{\sigma(i)}, y_{\sigma(i)}) \in \partial^c u$), we find

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

note that in the preceding inequality $u(x_i)$, $u(x_{\sigma(i)})$ and $c(x_i, y_i)$ are real numbers thanks to the assumptions, while $c(x_{\sigma(i)}, y_i)$ may be $+\infty$. Adding now the preceding inequality with $i = 1, \dots, k$ and recalling that σ is a permutation thus $\sum u(x_i) = \sum u(x_{\sigma(i)})$, we find

$$\sum c(x_i, y_i) \leq \sum c(c_i, y_{\sigma(i)}),$$

then Γ is $c-$cyclically monotone; in particular so is each $\partial^c u$.

Necessity: Let us take a $c-$cyclically monotone set Γ as stated: we need to find a suitable function u; to do that, following the construction of Rüschendorf in [69] (see also [64]), we fix $(x_0, y_0) \in \Gamma$ with y_0 as in the hypothesis and set

$$\begin{aligned} u(x) := \inf \Big\{ & c(x, y_k) - c(x_k, y_k) + c(x_k, y_{k-1}) - c(x_{k-1}, y_{k-1}) + \cdots \\ & \cdots + c(x_1, y_0) - c(x_0, y_0) \;:\; (x_1, y_1),\, (x_2, y_2),\, \dots,\, (x_k, y_k) \in \Gamma,\ k \in \mathbb{N} \Big\}; \end{aligned} \tag{3.7}$$

note that the sum makes always sense since all the subtracting terms are finite by hypothesis; moreover $u(x) < +\infty$ for any x because (3.7) gives $u(x) \leq c(x, y_0) - c(x_0, y_0) < +\infty$; finally, we can exclude the possibility $u \equiv -\infty$ because the $c-$cyclical monotonicity of Γ immediately gives $u(x_0) = 0$. It follows that u is a $c-$concave function. Take now $(x, y) \in \Gamma$: we need to check that $(x, y) \in \partial^c u$; thanks to (3.7), we can find $(x_1, y_1), \dots, (x_k, y_k) \in \Gamma$ such that

$$u(x) \geq c(x, y_k) - c(x_k, y_k) + c(x_k, y_{k-1}) - c(x_{k-1}, y_{k-1}) + \cdots + c(x_1, y_0) - c(x_0, y_0) - \varepsilon;$$

but, again using (3.7), for any $\xi \in X$ it is

$$\begin{aligned} u(\xi) &\leq c(\xi, y) - c(x, y) + c(x, y_k) - c(x_k, y_k) + \cdots + c(x_1, y_0) - c(x_0, y_0) \\ &\leq c(\xi, y) - c(x, y) + u(x) + \varepsilon, \end{aligned}$$

that gives

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

recalling (3.5), this implies $(x, y) \in \partial^c u$. □

Remark 3.11. *This Theorem was first proved by Rockafellar in [65] in the case of cyclically monotone sets and concave functions, while the generalization to $c-$cyclically monotone sets and $c-$concave functions is due to Knott and Smith [57].*

Remark 3.12. *As the sufficiency part in last Theorem shows, given any function $u : X \to \overline{R}$ and any l.s.c. function c one has that $\partial^c u$ is $c-$cyclically monotone, as well as any subset of it.*

We can finally prove the equality we stated before, i.e. that $\sup(3.3) = \min(1.5)$: we will begin with the well-known case when c is continuous and bounded (see for example [64] or [74]), then we will prove the general assert.

Lemma 3.13. *If c is continuous and bounded, then* $\sup(3.3) = \min(1.5)$.

Proof. An inequality is trivial: take any transport plan γ and any couple $(u, v) \in \mathcal{A}$; recalling that f^+ and f^- are the marginals of γ, it is

$$\iint_{X\times Y} c(x,y)\, d\gamma \geq \iint_{X\times Y} u(x) + v(y)\, d\gamma = \int_X u(x)\, df^+ + \int_Y v(y)\, df^-,$$

then $\min(1.5) \geq \sup(3.3)$. On the other hand, take an optimal transport plan γ: since the cost function c is bounded, $\min(1.5) \leq \|c\|_{L^\infty}\|f^+\| < +\infty$, then we can apply Proposition 3.2 to find that $\operatorname{spt}\gamma$ is $c-$cyclically monotone. Thus Theorem 3.10 implies the existence of a $c-$concave function u such that $\operatorname{spt}\gamma \subseteq \partial^c u$. We immediately point out that u is u.s.c. because it is $c-$concave and then the infimum of continuous functions; we define now v the $c-$transform of u, according to (3.5): then, v is also the infimum of some continuous function and so is u.s.c.. Moreover, since c is bounded, u must be bounded from above; in fact, take $y_0 \in Y$ with $u^c(y_0) > -\infty$, which is possible because otherwise $\partial^c u$ would be empty: then, given any x, $u(x) < +\infty$ since u is $c-$concave and so $u(x) \leq c(x, y_0) - u^c(y_0) \leq \sup c - u^c(y_0)$ by (3.5). Then, since u is bounded from above, it is clear from (3.5) that v is bounded, both from above and below. From (3.6) it finally follows that u is bounded from below on $\operatorname{spt} f^+$, since for any $x \in \operatorname{spt} f^+$ there is some y such that $(x, y) \in \operatorname{spt}\gamma \subseteq \partial^c u$. Therefore u and v are bounded u.s.c. functions and $u(x) + v(y) \leq c(x, y)$ for any $(x, y) \in X \times Y$ by definition of v: it follows $(u, v) \in \mathcal{A}$ and also $u(x) + v(y) = c(x, y)$ on $\operatorname{spt}\gamma \subseteq \partial^c u$, thus

$$\int_X u(x)\, df^+ + \int_Y v(y)\, df^- = \iint_{X\times Y} u(x) + v(y)\, d\gamma = \iint_{X\times Y} c(x,y)\, d\gamma = \mathfrak{C}(\gamma),$$

so we have $\sup(3.3) \geq \min(1.5)$ and the thesis follows. □

By an approximation method already used in [2] we can now get rid of the hypothesis of continuity of c, showing the

Theorem 3.14. $\sup(3.3) = \min(1.5)$.

Proof. First of all, let us note that there is an increasing sequence of continuous and bounded functions c_j converging to c from below: this is a standard fact, it suffices to take distances d_X and d_Y on X and Y according to the definition of Polish spaces, and then to define

$$c_j(x,y) := \inf \left\{ c(\tilde{x},\tilde{y}) + j\big(d_X(x,\tilde{x}) + d_Y(y,\tilde{y})\big) \right\} \wedge j.$$

Let now γ_j be an optimal transport plan with respect to c_j, and let γ be a weak* limit of some subsequence of γ_j. Since c_j is continuous, by Lemma 3.13 there is a pair (u,v) with $\int u\, df^+ + \int v\, df^- \geq \int c_j d\gamma_j - \delta$ and $u + v \leq c_j \leq c$, then $\sup(3.3) \geq \int c_j\, d\gamma_j - \delta$; we fix then j_0 and, since c_j is increasing, we find for any $j \geq j_0$ that $\sup(3.3) \geq \int c_{j_0}\, d\gamma_j - \delta$ and, since $\gamma_j \stackrel{*}{\rightharpoonup} \gamma$ one infers

$$\sup(3.3) \geq \int_{X\times Y} c_{j_0}\, d\gamma - \delta.$$

Passing to the limit for $j \to +\infty$ recalling Beppo-Levi Theorem, and then sending $\delta \to 0$, we obtain that $\sup(3.3) \geq \min(1.5)$. The opposite inequality is trivial, exactly as in Lemma 3.13. $\square$

The fact that X and Y are metrizable (since they are Polish spaces) was fundamental in the preceding proof: in fact, the property that any l.s.c. function is the pointwise limit of an increasing sequence of continuous function is not true for general topological spaces, even for locally compact and separable spaces, as the next example shows.

Example 3.15. Let $Z = \overline{R}$ and define $\tau \subseteq \mathscr{P}(Z)$ by the property that $A \subseteq Z$ belongs to τ if and only if $\{+\infty\} \in A$ and $\{-\infty\} \notin A$, or if $A = Z$ or $A = \emptyset$. It is immediately seen that τ is a $\sigma-$algebra and that the topological space (Z,τ) is compact (thanks to $-\infty$) and separable (thanks to $+\infty$). Moreover, a function $\varphi : Z \to \mathbb{R}$ is continuous if and only if it is constant, while it is l.s.c. if and only if for any $z \in Z$ it is $\varphi(-\infty) \leq \varphi(z) \leq \varphi(+\infty)$: consequently, a l.s.c. function which is not continuous is never a limit of continuous functions.

3.3 Necessity for a general cost function c

We prove now Proposition 3.2 with the sole lower semicontinuity of c, a result first considered in [2].

Theorem 3.16. *If* $\min(1.5) < +\infty$*, then any optimal transport plan is concentrated in a $c-$cyclically monotone and $\sigma-$compact set on which the cost function c is finite.*

Proof. Let us consider the optimal transport plan γ; take then a maximizing sequence (u_n, v_n) for (3.3) and define

$$c_n(x,y) := c(x,y) - u_n(x) - v_n(y).$$

Theorem 3.14 gives

$$\int_X u_n(x)\, df^+(x) + \int_Y v_n(y)\, df^-(y) \to \int_{X\times Y} c(x,y)\, d\gamma(x,y),$$

and then, recalling that $\min(1.5) < +\infty$, $\int c_n\, d\gamma \to 0$. Since c_n is a positive function because by definition $u_n(x)+v_n(y) \le c(x,y)$, up to a subsequence we can assume $c_n(x,y) \to 0$ pointwise on a $\sigma-$compact set of full $\gamma-$measure Γ. Take now $(x_1,y_1), \dots (x_k,y_k) \in \Gamma$ and a permutation $\sigma \in \mathfrak{S}_k$. Since $c \ge u_n + v_n$,

$$\sum_{i=1}^k c(x_i, y_{\sigma(i)}) \ge \sum_{i=1}^k u_n(x_i) + v_n(y_{\sigma(i)}) = \sum_{i=1}^k u_n(x_i) + v_n(y_i);$$

but for any i we know by construction that $c_n(x_i,y_i) \to 0$, which means $u_n(x_i)+v_n(y_i) \to c(x_i,y_i)$. The preceding inequality then becomes

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

which gives the $c-$cyclical monotonicity. The fact that c is finite on Γ can be achieved with an intersection with the set $\{(x,y) : c(x,y) < +\infty\}$, which is of $\gamma-$full measure because $\mathfrak{C}(\gamma) < +\infty$, and $\sigma-$closed because c is l.s.c.; since passing to a smaller set does not effect the $c-$cyclical monotonicity, the thesis follows. □

In Proposition 3.2 we proved, under the hypothesis that c is continuous, that the support of any optimal transport plan is $c-$cyclically monotone, while in last Theorem, with the sole lower semicontinuity of c, we proved that any optimal transport plan is concentrated on a $c-$cyclically monotone set: in fact, it is easily seen that the whole support of some optimal transport plan may be not $c-$cyclically monotone; moreover, one can easily build examples where any optimal transport plan has a non $c-$cyclically monotone support (see example 3.17 below).

In particular, in the case when c is continuous we saw the existence of a $c-$cyclically monotone set containing the support of all the optimal transport plans with the easy Corollary 3.3 of Proposition 3.2; but in the l.s.c. case, one cannot immediately infer from Theorem 3.16 the existence of a $c-$cyclically monotone set in which each optimal transport plan is concentrated. However, this is what we will do, under suitable hypotheses, in Corollary 3.24.

Example 3.17. Let $X = Y = \mathbb{R}$ and let $f^+ = \mathscr{L} \llcorner [0,1]$ and $f^- = \mathscr{L} \llcorner [1,2]$; let moreover the cost function be given by

$$c(x,y) := \begin{cases} 0 & \text{if } (x,y) = (0,2), \\ |y-x|^2 & \text{otherwise:} \end{cases}$$

the function c is l.s.c. but not continuous and the only optimal transport plan is $\gamma = (\mathrm{Id}, \mathrm{Id}+1)_\# \mathscr{L} \llcorner [0,1]$. According to Theorem 3.16, the measure γ is concentrated in the $c-$cyclically monotone set $]\!]0,1[\![\times]\!]1,2[\![$; however, the support of γ is $[\![0,1]\!] \times [\![1,2]\!]$, which is not $c-$cyclically monotone (take the pairs $(0,1)$ and $(1,2)$).

We point out now that the inequality (3.1), in the particular case $k = 2$, reads as $c(x_1,y_1) + c(x_2,y_2) \le c(x_1,y_2) + c(x_2,y_1)$: this has the intuitive meaning that if an optimal transport plan moves x_1 in y_1 and x_2 in y_2, then this must be more convenient than moving x_1 in y_2 and x_2 in y_1. An even more particular case of this observation will be very interesting in the sequel:

Lemma 3.18. *If $X = Y = \mathbb{R}^N$ and $c(x,y) = |y-x|$, if γ is an optimal transport plan and $(x_1,y_1),(x_2,y_2) \in \operatorname{spt}\gamma$ then the closed segments $[\![x_1,y_1]\!]$ and $[\![x_2,y_2]\!]$ cannot intersect, unless they are parallel or the intersection point is the left extrem of both them of the right extreme of both them. In particular, then, the closed segment $[\![x_1,y_1]\!]$ and the open segment $]\!]x_2,y_2[\![$ cannot intersect unless they are parallel. The same statement holds whenever $c(x,y) = \|y-x\|$ and $\|\cdot\|$ is a strictly convex norm in $\mathbb{R}^N$.*

Proof. The inequality (3.1), in the above considered case and for $k = 2$, says

$$|y_1 - x_1| + |y_2 - x_2| \le |y_2 - x_1| + |y_1 - x_2|; \tag{14.6}$$

by the strict triangle inequality, this is not possible if the segments $[\![x_1,y_1]\!]$ and $[\![x_2,y_2]\!]$ intersect in a single point z which is in the interior of at least one of them: in fact, in that case z is not contained in $[\![x_1,y_2]\!]$ nor in $[\![x_2,y_1]\!]$ and $|y_1 - x_1| = |y_1 - z| + |z - x_1|$ and $|y_2 - x_2| = |y_2 - z| + |z - x_2|$; thus

$$|y_2 - x_1| + |y_1 - x_2| < |y_2 - z| + |z - x_1| + |y_1 - z| + |z - x_1| = |y_1 - x_1| + |y_2 - x_2|,$$

which is absurd. On the other hand, if $[\![x_1,y_1]\!]$ and $[\![x_2,y_2]\!]$ intersect in a single point which is the left extreme of a segment and the right extreme of the other one, again the strict triangle inequality leads to an absurd if the segment are not parallel.

This argument clearly works if the Euclidean norm $|\cdot|$ is replaced by any strictly convex norm $\|\cdot\|$. $\square$

We also point out that the preceding "no-crossing" property is particularly interesting in $\mathbb{R}^2$, because in that case the directions of the segments with endpoints in $\operatorname{spt}\gamma$ are strongly constrained by this property. This is why we will able to show an existence Theorem for general norms in $\mathbb{R}^2$, in Chapter 10, while the case of general norms in $\mathbb{R}^N$, $N \geq 3$ is still open.

3.4 Sufficiency of the cyclical monotonicity and Kantorovich potential

In this section we will provide a converse to Theorem 3.16, saying that a transport plan concentrated in a $c-$cyclical monotone set is optimal; in particular, we will prove the existence of a maximizing pair for (3.3) and, using it, we will define a $c-$cyclically monotone set on which each optimal transport plan is concentrated. This will lead us to the definition of the *Kantorovich potential*, a function with remarkable properties which has been deeply used throughout the last years and which has always had a great importance in the mass transport problem: Monge himself ([60]) guessed the existence of such a function, and Kantorovich ([53, 54]) first considered it formally. We will provide the results of this section under some hypotheses on c, f^+ and f^-; then in next section we will give two counterexamples showing that actually some hypotheses have to be made.

First of all we give the following result:

Lemma 3.19. *Let γ be a transport plan concentrated in a $c-$cyclically monotone set Γ, and assume that $\{(x, y) \in X \times Y : c(x, y) < +\infty\}$ contains Γ as well as $X_0 \times \{y_0\}$ for some X_0 of full f^+-measure and $y_0 \in \pi_2(\Gamma)$. Then there exists a Borel dual pair (u, v) such that γ is concentrated in $\partial^c u$.*

Proof. Thanks to Theorem 3.10 applied replacing X by X_0, since Γ is $c-$cyclically monotone there exists a $c-$concave function $u_0 : X_0 \to \overline{R} \setminus \{+\infty\}$ such that $\Gamma_0 := \big(\Gamma \cap (X_0 \times Y)\big) \subseteq \partial^c u_0$; we extend then u_0 to X setting $u_0 \equiv -\infty$ on $X \setminus X_0$ and we define $v = u_0^c$ and $u = v^c$. It is trivial to notice that (as in the standard polar transform) (u, v) is a dual pair; moreover, since $u \geq u_0$ and $u^c = u_0^c = v$, recalling (3.6) it is clear that $\partial^c u \supseteq \partial^c u_0 \supseteq \Gamma_0$. Since clearly γ is concentrated in Γ_0 because X_0 has full f^+-measure hence $X_0 \times Y$ has full $\gamma-$measure, we have only to prove that u and v are Borel function: to do that, since c is l.s.c. we take an increasing sequence of continuous functions c_j pointwise converging to c and define

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

as infimum of continuous function ($c_j(x,y)-u(x)$ is a continuous function in the unknown y), v_j is u.s.c.; moreover, it is clear that v_j is an increasing sequence of functions and that it is pointwise converging to v, so $v = u^c$ is the supremum of countably many u.s.c. functions, thus it is Borel. The same holds for $u = v^c$, hence the assert follows. $\square$

It is reasonable to look at the pair (u, v) found in last Lemma as an optimal pair for (3.3): in fact, since $u(x)+v(y) = c(x,y)$ for any pair $(x,y) \in \partial^c u$ and then for $\gamma-$a.e. pair, recalling that the marginals of γ are f^+ and f^- one would say that

$$\iint_{X\times Y} c(x,y)\, d\gamma(x,y) = \iint_{X\times Y} u(x)+v(y)\, d\gamma(x,y) = \int_X u(x)\, df^+(x) + \int_Y v(y)\, df^-(y),$$

that thanks to Theorem 3.14 would give at the same moment the optimality of (u, v) and of γ. In fact, the preceding argument works if $u \in L^1_{f^+}$ and $v \in L^1_{f^-}$ (or at least if u and v are semi-integrable), while otherwise the last equation does not make sense because one can not split the integral: then we will assume some hypotheses to ensure the needed summability of u and v. Following [2], we set then the following

Definition 3.20. *We say that c is slow $-(f^+, f^-)-$ increasing if*

$$\begin{aligned} f^+\Big(\Big\{x \in X : \int_Y c(x,y)\, df^-(y) < +\infty\Big\}\Big) > 0 \\ f^-\Big(\Big\{y \in Y : \int_X c(x,y)\, df^+(x) < +\infty\Big\}\Big) > 0. \end{aligned} \tag{3.8}$$

and $\min(1.5) < +\infty$.

The importance of property (3.8) is, in some sense, to ensure that c does not increase too much with respect to f^+ and f^-; note that the hypothesis

$$\iint_{X\times Y} c(x,y)\, df^+ \boxtimes f^-(x,y) < +\infty,$$

corresponding to the request that the transport plan $f^+ \boxtimes f^-$ has a finite cost, is stronger than the property of being *slow $-$ increasing*; even stronger is the assumption $c(x,y) \le a(x)+b(y)$ with $a \in L^1_{f^+}$ and $b \in L^1_{f^-}$. We point out now an important fact which will be used often in the sequel.

Remark 3.21. *If $X = Y$ is a normed space and $c(x,y) = \|y-x\|$, then it is easily noticed that c is slow $-(f^+, f^-)-$ increasing if and only if f^+ and f^- have finite first order moments $\int_X \|x\|\, f^+(x)$ and $\int_X \|y\|\, f^-(y)$; for this reason in [2] the assumption that f^+ and f^- have finite first order moments was made. However, if c is a generic l.s.c. function then the finiteness of the first order moments and the fact of c being slow $-(f^+, f^-)-$ increasing are clearly independent, also in the particular case when $X = Y = \mathbb{R}$ with the Euclidean norm.*

We prove now the converse of Theorem 3.16:

Theorem 3.22. *If c is slow $-(f^+, f^-)-$ increasing, then there exists an optimal dual pair (u, v) for (3.3). Moreover a transport plan is optimal if and only if it is concentrated in a $c-$cyclically monotone set. In particular, a transport plan γ is optimal if and only if it is concentrated in the $c-$cyclically monotone set $\partial^c u$.*

Proof. Let us take an optimal transport plan γ; since by hypothesis $\mathfrak{C}(\gamma) < +\infty$, by Theorem 3.16 γ is concentrated in a $c-$cyclically monotone set Γ on which c is finite. Moreover by (3.8) we can find some $y_0 \in \pi_2(\Gamma)$ such that $c(x, y_0) < +\infty$ for each x contained in a set X_0 of full f^+-measure; thanks to Lemma 3.19 we have a Borel dual pair (u, v) such that γ is concentrated on $\partial^c u$; we claim that (u, v) is optimal for (3.3), so we have first to prove that $u \in L^1_{f^+}$ and $v \in L^1_{f^-}$.

Since γ is concentrated in $\partial^c u$, again thanks to (3.8) we find $y_1 \in \pi_2(\partial^c u)$ such that $\int_X c(x, y_1)\, df^+(x) < +\infty$. Moreover, since $y_1 \in \partial^c u$, it is clearly $u^c(y_1) > -\infty$. Then, $u = v^c$ ensures that for any x it is

$$u(x) \leq c(x, y_1) - v(y_1);$$

then, setting $A = \{x : u(x) \geq 0\}$, we have

$$\int_X u^+(x)\, df^+(x) = \int_A u(x)\, df^+(x) \leq \int_A c(x, y_1) - v(y_1)\, df^+(x)$$

and the last integral has a real value because it is positive and $v(y_1) > -\infty$ (by the way, one also infers from last inequality that $v(y_1) < +\infty$). It follows $u^+ \in L^1_{f^+}$, and the same is similarly true for v^+ and f^-.

Since $c(x, y) = u(x) + v(y)$ in $\partial^c u$ and then $\gamma-$a.e. and since u and v are semi-integrable with respect to f^+ and f^-, we can write

$$\mathfrak{C}(\gamma) = \int_{X\times Y} c(x, y)\, d\gamma(x, y) = \int_X u(x)\, f^+(x) + \int_Y v(y)\, df^-(y),$$

and we know the last number to belong to $\mathbb{R} \cup \{-\infty\}$. Since clearly $\mathfrak{C}(\gamma) \geq 0$ we infer that $u \in L^1_{f^+}$ and $v \in L^1_{f^-}$; at the same time, then, we deduce that (u, v) is optimal for (3.3) thanks to Theorem 3.14.

Finally, let us take any transport plan η: we have

$$\mathfrak{C}(\eta) = \int_{X\times Y} c(x, y)\, d\eta(x, y) \geq \int_{X\times Y} u(x) + v(y)\, d\eta(x, y) = \min(1.5),$$

so η is optimal if and only if $c(x, y) = u(x) + v(y)$ for $\gamma-$a.e. (x, y), then if and only if it is concentrated in $\partial^c u$. $\square$

The function u in the Theorem above is usually referred to as a Kantorovich potential; formally, we give the following definition:

Definition 3.23. *A* Kantorovich potential *is any function* $u : X \to \overline{R}$ *such that* (u, u^c) *is a dual pair which is optimal for (3.3).*

Let us now note some non-trivial consequences of the last theorem: first of all, the same argument we used works for any transport plan γ concentrated in a $c-$cyclically monotone set where c is finite provided that (3.8) holds: the fact that γ is optimal and that $\mathfrak{C}(\gamma) < +\infty$ is then a consequence. In particular, if c is real-valued then the sole hypothesis (3.8) and the existence of some transport plan concentrated in a $c-$cyclically monotone set imply min(1.5) to be finite, while the same is not true without the hypothesis (3.8). Moreover, in the hypotheses of the last Theorem, a transport plan is concentrated in a $c-$cyclically monotone set if and only if it is concentrated in $\partial^c u$! This immediately gives the following extension of Corollary 3.3:

Corollary 3.24. *If* c *is* $slow - (f^+, f^-) - increasing$*, then there is a* $c-$*cyclically monotone set* Γ *on which each optimal transport plan is concentrated.*

In last corollary, the hypothesis that c is $slow - (f^+, f^-) - increasing$ is fundamental, as the next elementary Example shows:

Example 3.25. Let $f^+ = f^- = \delta_0 + \delta_1 + \delta_2$ and $c(i, j)$ is 0 if $i = 0$ or $j = 0$ and $+\infty$ otherwise: it is easy to notice that (3.8) holds, c is not $slow-(f^+, f^-)-increasing$ because $\min(1.5) = +\infty$, (3.3) does not have solutions and there is not a $c-$cyclically monotone set S such that $\gamma\big((X \times Y) \setminus S\big) = 0$ for any optimal transport plan γ. In particular, there are optimal transport plans which are not concentrated in a $c-$cyclically monotone set.

Remark 3.26. *We can now discuss the reason why we gave in Definition 3.9 a notion of* $c-$*superdifferential slightly more general than the usual one: the only difference is that we allow to compute the* $c-$*superdifferential of any function* $u : X \to \overline{R}$ *instead of only the* $c-$*concave ones. But we deal only with dual pairs* (u, v)*, and recalling the definitions it is clear that any such* u *is in fact* $c-$*concave unless it takes the value* $+\infty$ *somewhere: this is essential, because there are situations where the solutions of (3.3) take the value* $+\infty$*. However, there is no difference between the classical notions and our ones in the case when the cost function is real-valued, which is usually the only case considered in transport problems.*

Another consequence of Theorem 3.22 is the following particular case:

Corollary 3.27. *Assuming that $X = Y$ and c is a semi-distance on X, if c is slow $-$ (f^+, f^-) $-$ increasing then there is a function $u : X \to \mathbb{R}$ which is $1-$Lipschitz with respect to c and such that $(u, -u)$ is an optimal dual pair for (3.3).*

Proof. More in general, we prove that for any dual pair (u, v) it is $v \equiv -u$ if c is a semi-distance (the fact that u is $1-$Lipschitz w.r.t. c is then an immediate consequence of (3.5)): take a dual pair (u, v) and notice that, since $u(x) + v(y) \leq c(x, y)$ and c is a semi-distance, in particular it follows

$$u(x) + v(x) \leq 0$$

(this is in general true whenever $X = Y$ and $c = 0$ on the diagonal Δ of $X \times X$). Let us then assume the existence of some $x \in X$ such that $u(x) + v(x) = -\varepsilon < 0$, and take y and z (thanks to the definition of $c-$transform) with the property

$$u(x) > c(x, y) - v(y) - \varepsilon/3 \qquad\qquad v(x) > c(x, z) - u(z) - \varepsilon/3;$$

Adding these inequalities and recalling that $u(z) + v(y) \leq c(z, y)$, we find

$$-\varepsilon = u(x)+v(x) > c(x, y)+c(x, z)-\big(u(z)+v(y)\big)-2\varepsilon/3 \geq c(x, y)+c(x, z)-c(z, y)-2\varepsilon/3,$$

thus it follows $c(z, y) > c(z, x) + c(x, y)$, which is absurd by the triangle inequality of the semi-distance c. Then $u(x) + v(x) = 0$ for any $x \in X$ and the thesis follows. ☐

In order to have a quick reference for later, we collect some of the preceding results in the

Theorem 3.28. *If c is slow $-$ (f^+, f^-) $-$ increasing, then there exists a Kantorovich potential $u : X \to \overline{R}$. Defining $v = u^c$, then, it is $u(x) + v(y) \leq c(x, y)$ for any pair (x, y) and a transport plan γ is optimal if and only if $u(x) + v(y) = c(x, y)$ for $\gamma-$a.e. (x, y), hence if and only if it is concentrated on $\partial^c u = \{(x, y) : u(x) + v(y) = c(x, y)\}$.*

3.5 Two counterexamples

We proved the results of last section under the hypothesis of c being *slow* $-$ (f^+, f^-) $-$ *increasing*; we then conclude this Chapter with a couple of counterexamples showing that some hypotheses have to be made: in particular, in the first one (already appeared in [2]) we show the existence of a transport plan which has finite cost and whose support is $c-$cyclically monotone but which is not optimal; in the second one, we give an example where $\min(1.5) < +\infty$ but there is no optimal pair for (3.3).

Example 3.29. Let $X = Y = [0,1] \subseteq \mathbb{R}$ and $f^+ = f^- = \mathscr{L} \llcorner [0,1]$; let moreover $\alpha \in [0,1] \setminus \mathbb{Q}$ and define the cost function as

$$c(x,y) := \begin{cases} 0 & \text{if } y = x \in [0,1] \\ 1 & \text{if } y = x \oplus \alpha \in [0,1] \\ +\infty & \text{otherwise,} \end{cases} \tag{3.9}$$

where $\oplus : [0,1] \times [0,1] \to [0,1]$ is the sum modulo 1; the cost function is shown in Figure 3.1; it is clear that the unique optimal transport plan is $\bar{\gamma} = (\mathrm{Id}, \mathrm{Id})_\# f^+$, which

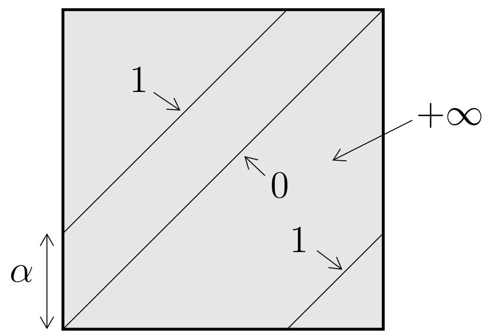


Figure 3.1: The cost function c in Example 3.29

has a null cost, and it is also clear that (3.8) does not hold. Consider now the transport plan $\gamma = \big(\mathrm{Id}, \mathrm{Id} \oplus \alpha\big)_\# f^+$: this is clearly a transport plan, but it is not optimal since its cost is 1; nevertheless, we check now that its support is $c-$cyclically monotone. Otherwise, there would be a minimal set of couples (x_1, y_1), $(x_2, y_2), \dots, (x_n, y_n)$ in $\operatorname{spt} \gamma$ with the property that

$$n = c(x_1, y_1) + \cdots + c(x_n, y_n) > c(x_2, y_1) + c(x_3, y_2) + \cdots + c(x_1, y_n);$$

but $(x_i, y_i) \in \operatorname{spt} \gamma$ implies $y_i = x_i \oplus \alpha$: moreover, since the preceding inequality ensures $c(x_{i+1}, y_i) < +\infty$, we can infer from (3.9) that $y_i = x_{i+1}$ or $y_i = x_{i+1} \oplus \alpha$, and then $x_{i+1} = x_i \oplus \alpha$ or $x_{i+1} = x_i$. Since the second possibility is incompatible with the minimality of the set, it must be $x_{i+1} = x_i \oplus \alpha$; applying this equality n times, we find $x_1 = x_1 \oplus n\alpha$, which is impossible since α is not a rational number. Then γ is a non-optimal transport plan with a $c-$cyclically monotone support.

It is interesting to notice that also in very specific but important cases, like the quadratic cost $c(x,y) = |y - x|^2$ in $\mathbb{R}^N$ (see also Chapter 5), it is not clear at the moment whether or not the sole cyclical monotonicity of a transport plan γ with $\mathfrak{C}(\gamma) < +\infty$ implies its optimality, see the Open problem 16 in Chapter 3 of [74].

In last Example, there still was an optimal pair for (3.3), namely $u = v \equiv 0$; now we see an example where there are no solutions of (3.3), even though $\min(1.5) < +\infty$

Example 3.30. Let $X = Y = \mathbb{R}$ and $c(x, y) = |y - x|^2$; let moreover f^+ be a probability measure diffused in the whole $\mathbb{R}$, i.e. $f^+(A) > 0$ for any open set $A \subseteq \mathbb{R}$, and let $f^- = \tau_{1\#} f^+$, where τ_1 is the translation $\tau_1(x) = x + 1$. The unique optimal transport plan is $\gamma = (\mathrm{Id}, \tau_1)_\# f^+$, i.e. the plan shifting all the mass on the right for a length 1: this is an immediate consequence of the general properties of the $1-$dimensional case we will study in Section 4.2. Moreover, $\mathfrak{C}(\gamma) = 1$ since $c(x, y) = 1$ for $\gamma-$a.e. (x, y), hence $\min(1.5) = 1 < +\infty$.

Let us suppose now that there is an optimal pair $(u, v) \in \mathcal{A}$ for (3.3); first of all we notice that it must be $v = u^c$: this happens because $v \leq u^c$ by definition, but if $v < u^c$ on a set of positive f^+-measure, then the pair $(u, \tilde{v})$ is strictly better than (u, v) for (3.3) whenever $v \leq \tilde{v} \leq u^c$, $\tilde{v} > v$ in some set of positive measure and $\tilde{v} \in L^1_{f^-}$, against the optimality of (u, v). Take then $x \in \mathbb{R}$: since $\operatorname{spt}\gamma \subseteq \partial^c u$, we infer that $v(x+1) = u^c(x+1) = |x+1-x|^2 - u(x)$ (recall that f^+ is diffuse, then $(x, x+1)$ belongs to $\operatorname{spt}\gamma$). Therefore the function $z \to |x + 1 - z|^2 - u(z)$ attains its minimum in $z = x$. This gives

$$1 - u(x) \leq 1 - 2\,\varepsilon + \varepsilon^2 - u(x + \varepsilon),$$

and analogously

$$1 - u(x + \varepsilon) \leq 1 + 2\,\varepsilon + \varepsilon^2 - u(x)\,:$$

we infer, for an arbitrary $x \in \mathbb{R}$ and $\varepsilon > 0$

$$-2 - \varepsilon \leq \frac{u(x + \varepsilon) - u(x)}{\varepsilon} \leq -2 + \varepsilon.$$

This gives that $u(x) = -2x + \lambda$ for some constant λ, and similarly $v(y) = 2y - \lambda - 1$. But one can easily find a probability measure f^+ diffuse on $\mathbb{R}$ and such that for any λ the function $-2x + \lambda$ is not in $L^1_{f^+}$, so that the pair (u, v) does not belong to $\mathcal{A}$. So, there are no optimal pairs for (3.3).

Chapter 4

Some useful tools in transport problems

In this chapter we will discuss some definitions and properties, that will be needed to prove our main results in next chapters. First we present a refined version of the notions of transport rays, transport set and other similar definitions; these definitions can be found very often in the literature, but in many more or less different ways: we also discuss them and explain the meaning of the changes we made. Then we study the case when $X = Y = \mathbb{R}$, which is an important particular case and will be also used in the following to derive results in $\mathbb{R}^N$. Finally, we introduce a couple of secondary problems that allow to select optimal transport plans with particular properties and we will show some results about them, providing also two counterexamples to show that the hypotheses are needed.

4.1 Transport rays and transport set

In this section we define some extremely useful tools in the study of the mass transport problem. The notions of transport rays and transport set have been first introduced by Evans and Gangbo in [39] (even if Monge himself thought to something similar), and they have been used a lot since (for example, in [24, 1, 31]). Their idea was the following: given a $1-$Lipschitz function u in $\mathbb{R}^N$, a transport ray is each closed segment $[\![x, y]\!]$ such that $u(x) - u(y) = |y - x|$, and the transport set is the union of all the transport rays. The reason of these definitions was that, if u is a Kantorovich potential for (f^+, f^-) with the Euclidean cost function $c(x, y) = |y - x|$, then an optimal transport map has to move the mass along the transport rays. Here we present a different notion, more general, starting from a subset Γ of $X \times Y$, which refines the one already introduced and used in [2] (see also [4]).

In this chapter, we consider the case $X = Y = \mathbb{R}^N$. Suppose we are given a set $\Gamma \subseteq \mathbb{R}^N \times \mathbb{R}^N$ (in our applications, we will always consider the case when Γ is $\sigma-$compact); we give the following definitions:

(Transport ray) A *transport ray* is any open segment $]\!]x, y[\![$ such that $(x, y) \in \Gamma \setminus \Delta$, where $\Delta = \{(x, x) : x \in \mathbb{R}^N\}$ is the diagonal of $\mathbb{R}^N \times \mathbb{R}^N$. The reason for the assumption $(x, y) \notin \Delta$ is only to avoid considering the empty open segments $]\!]x, x[\![$.

(Transport set) The *transport set* T_Γ is simply the union of all the transport rays, i.e.

$$\mathrm{T}_\Gamma := \bigcup \left\{ \,]\!]x, y[\![: \, (x, y) \in \Gamma \setminus \Delta \right\}.$$

(Ray direction) The *direction* of any transport ray $]\!]x, y[\![$ is

$$\tilde{\tau}_\Gamma(x, y) := \frac{y - x}{|y - x|};$$

as in the rest of the paper, $| \cdot |$ denotes the Euclidean norm in $\mathbb{R}^N$.

(Left-extreme points) A *left-extreme point* is any point contained in the set

$$\mathrm{E}^l_\Gamma := \bigcup \left\{ \, [\![x, y[\![: \, (x, y) \in \Gamma \right\} \setminus \bigcup \left\{ \,]\!]x, y[\![: \, (x, y) \in \Gamma \right\},$$

where we set $[\![x, x[\![= \emptyset$; in words, x is a left-extreme point if there is some $y \neq x$ such that $(x, y) \in \Gamma$ and x is not contained in any transport ray –recall that, since the transport rays are open, x is not contained in $]\!]x, y[\![$.

(Fixed points) A *fixed point* is any point $x \notin \mathrm{T}_\Gamma \cup \mathrm{E}^l_\Gamma$ such that $(x, x) \in \Gamma$. The set of the fixed points will be denoted by F_Γ.

(Maximal transport ray) A (possibly unbounded) open interval $S \subseteq \mathbb{R}^N$ is said a *maximal transport ray* if any point $z \in S$ is contained in the closure of some transport ray parallel to S and there is not an interval strictly containing S with the same property.

As the preceding definitions clearly show, T_Γ, E^l_Γ and F_Γ are three disjoint subsets of $\mathbb{R}^N$ covering $\pi_1(\Gamma)$. Notice also that the maximal transport rays need not to be transport rays: for instance, in the case when $X = Y = \mathbb{R}$ and $\Gamma = \{(1/(n+2), 1/n) : n \in \mathbb{N}\}$, the open interval $]\!]0, 1[\![$ is a maximal transport ray but not a transport ray.

Let us fix now some notation: we denote by $\mathcal{S}_c$ the space of the (possibly unbounded) oriented closed segments in $\mathbb{R}^N$, and let us define a metric on it: if for any $n \in \mathbb{N}$

we denote by $[\![x_n, y_n]\!]$ the intersection of the closed segment $[\![x, y]\!]$ with the closed ball $\{z \in \mathbb{R}^N : |z| \leq n\}$, the metric is given by

$$d\big([\![x, y]\!], [\![x', y']\!]\big) := \sum_{n\in\mathbb{N}} 2^{-n} \frac{|x_n - x'_n| + |y_n - y'_n|}{|x_n - x'_n| + |y_n - y'_n| + 1};$$

note that this metric induces the "obvious" topology on $\mathcal{S}_c$, and makes it complete.

We consider now a very important property:

Definition 4.1. *The set* $\Gamma \subseteq \mathbb{R}^N \times \mathbb{R}^N$ *is said to fulfill the* no-crossing condition *if the following hold:*

$$\text{\textit{If S and T are two different maximal transport rays and $\bar S$ intersects $\bar T$, then the intersection point is the left-extreme of both $\bar S$ and $\bar T$ or the right-extreme point of both $\bar S$ and $\bar T$.}} \tag{4.1}$$

This property is very useful, and it will be of primary importance in the sequel. Under the assumption (4.1), we can give some other useful definitions.

(Doubling points) We call $x \in \mathbb{R}^N$ a *doubling point* if there are y_1 and y_2 such that (x, y_1) and (x, y_2) belong to Γ and $\tilde\tau_\Gamma(x, y_1) \neq \tilde\tau_\Gamma(x, y_2)$. The set of the doubling points (which is easily seen to be a subset of E^l_Γ) will be denoted by D_Γ.

(Extended transport set) We define *extended transport set* the union of T_Γ with $\mathrm{E}^l_\Gamma \setminus \mathrm{D}_\Gamma$, and we will denote this set by ET_Γ.

Thanks to the preceding definitions, we can reconsider what noticed after the first group of definition with the

Remark 4.2. $\pi_1(\Gamma) = \mathrm{T}_\Gamma \cup \mathrm{E}^l_\Gamma \cup \mathrm{F}_\Gamma$ *and, under hypothesis (4.1),* $\pi_1(\Gamma) = \mathrm{ET}_\Gamma \cup \mathrm{D}_\Gamma \cup \mathrm{F}_\Gamma$*; in both cases, this is a disjoint union.*

It is worth to be noticed that the main difference between the definitions we are giving now and those already considered in [2] is the introduction of the set of doubling points; in the sequel, the set D_Γ will be that of "bad points", while in [2] the "bad points" were those in E^l_Γ. But clearly $\mathrm{D}_\Gamma \subseteq \mathrm{E}^l_\Gamma$, and it may happen that the first set is much smaller than the second: this is why this new definition can be a bit more powerful. The meaning of this assertion will be much more clear in the sequel, in particular we will focus on it in Remark 7.6. Another consequence of the introduction of the set of the doubling points is that it makes much easier some proofs: as we will point out in Remark 8.6 and in Remark 10.3, there are a couple of situations when showing $\mathscr{L}(\mathrm{D}_\Gamma) = 0$ is trivial while $\mathscr{L}(\mathrm{E}^l_\Gamma) = 0$ is still true but requires a bit of additional work.

We notice now that, under assumption (4.1), the different maximal transport rays are disjoint and any point in ET_Γ is associated to a unique maximal transport ray: if $x \in \mathrm{T}_\Gamma$, then x is contained in the closure C of a unique maximal transport ray (in particular, x is contained in the interior of a unique maximal transport ray and is not contained in the closure of any other maximal transport ray); on the other hand, if $x \in \mathrm{E}^l_\Gamma \setminus \mathrm{D}_\Gamma$ then x may happen to belong to the closure of more than one maximal transport ray, but there is exactly one of them, call it C, containing a transport ray which has x as the left endpoint. We can then define the function $r_{\Gamma} : \mathrm{ET}_\Gamma \to \mathcal{S}_c$ which associates to any point $x \in \mathrm{ET}_\Gamma$ the closed segment C as before. Hence, if $x \in \mathrm{ET}_\Gamma$ and $(x,y) \in \Gamma$ one has that $\tilde{\tau}_{\Gamma}(x,y)$ is the direction of $r_{\Gamma}(x)$ (or equivalently of $r_{\Gamma}(y) = r_{\Gamma}(x)$ if $y \in \mathrm{ET}_\Gamma$).

(Pointwise direction) The *pointwise direction* $\tau_{\Gamma}(x)$ of Γ at any point $x \in \mathrm{ET}_\Gamma$ is the direction of the maximal transport ray $r_{\Gamma}(x)$. Then $\tau_{\Gamma} : \mathrm{ET}_\Gamma \to \mathbb{S}^{N-1}$ and in particular $\tau_{\Gamma}(z) = \tilde{\tau}_{\Gamma}(x,y)$ for any $z \in \mathrm{ET}_\Gamma$ and any transport ray (x,y) such that $z \in [\![x,y[\![$.

We also remark that, in the definition of maximal transport rays, we asked z to be contained in the closure of some transport ray parallel to S, while in [2] and [4] it was asked to be contained simply in a transport ray parallel to S. We made this change for many reasons: first of all, this is essential in order to deal with concave cost functions and non-increasing transport plans (however, in this paper we will deal only with the non-decreasing transport plans, associated to convex cost functions). Moreover this definition lowers the number of the maximal transport rays, and this helps when one has to consider separately the maximal transport rays: for example, if $X = Y = \mathbb{R}$ and $\Gamma = \{(1/(n+1),(1/n))\}$, our definition gives the unique maximal transport ray $]\!]0,1[\![$, while the other one gives countably many maximal transport rays. Finally, it makes the proof of Lemma 4.4 easier.

As we will better discuss in Section 6.1, to build an optimal transport map we will decompose an optimal transport plan in its associated maximal transport rays, then we will solve a one-dimensional transport problem finding an optimal transport map for each maximal transport ray, finally we will glue the different optimal transport maps. It is then easy to guess that it will be of primary importance the Borel property for all the tools we defined: we then first notice the

Lemma 4.3. *If Γ is $\sigma-$compact, the sets T_Γ, E^l_Γ and F_Γ are Borel; if the no-crossing condition (4.1) holds, the same is true for D_Γ and ET_Γ.*

Proof. If Γ is compact, then T_Γ is $\sigma-$compact, since it is the union of the closed sets $\{z \in \mathbb{R}^N : \exists\,(x,y) \in \Gamma : \ z \in]\!]x,y[\![, \ |z-x| \geq 1/n, \ |y-z| \geq 1/n\}$; then in general, since Γ is $\sigma-$compact, T_Γ is still $\sigma-$compact, thus Borel. Similarly, E^l_Γ is the difference of two

$\sigma-$compact sets, thus it is Borel; finally F_Γ is the difference between $\pi_1(\Gamma\cap\Delta)$, which is the projection of a $\sigma-$compact set, thus a $\sigma-$compact set, and $\mathrm{T}_\Gamma\cup\mathrm{E}^l_\Gamma$, which is Borel.

Suppose now that (4.1) holds: if Γ is compact, then D_Γ is the union of the closed sets

$$\Big\{x\in\mathbb{R}^N:\exists\, y_1,\, y_2:\ (x,y_1),\,(x,y_2)\in\Gamma,\ |\tilde{\tau}_\Gamma(x,y_2)-\tilde{\tau}_\Gamma(x,y_1)|\geq 1/n,\\ |y_1-x|\geq 1/n,\ |y_2-x|\geq 1/n,\Big\}$$

then it is $\sigma-$compact; in general, if Γ is the union of the compact sets Γ_i then D_Γ is the union of the corresponding D_{Γ_i}, thus it is still $\sigma-$compact; finally, ET_Γ is the difference between two Borel sets, thus it is Borel. □

Now we are ready to face the following measurability result, already proved in [2] where it was considered the restriction $r_\Gamma:\mathrm{T}_\Gamma\to\mathcal{S}_c$ in place of $r_\Gamma:\mathrm{ET}_\Gamma\to\mathcal{S}_c$:

Lemma 4.4. *If Γ is $\sigma-$compact and (4.1) holds, then r_Γ is a Borel map.*

Proof. Since Γ is $\sigma-$compact, we write $\Gamma=\bigcup_{n\in\mathbb{N}}\Gamma_n$ where $\{\Gamma_i\}$ is an increasing sequence of compact sets. Moreover we choose a compact set $K\subset\mathcal{S}_c$: the thesis will follow if we show that ${r_\Gamma}^{-1}(K)$ is a Borel subset of R^N.

Step I. *Definition of the function φ.*
First of all, we denote by $\varphi:\mathscr{P}(\mathcal{S}_c)\to\mathscr{P}(\mathbb{R}^N)$, where $\mathscr{P}(X)$ are the subsets of any set X, the union of the closed segments, i.e. $\varphi(J)$ is the union of the closed segments $C\in J$ for any $J\in\mathscr{P}(\mathcal{S}_c)$. It is immediate to notice that if J is closed, then so is $\varphi(J)$.

Step II. *Definition of $\widetilde{\Gamma}_n$, G_q, Sub and R_ε.*
Here we define the subsets $\widetilde{\Gamma}_n$ and G_q of $\mathcal{S}_c$, the map Sub from $\mathscr{P}(\mathcal{S}_c)$ to itself, and the map R_ε from the set of the bounded subsets of $\mathcal{S}_c$ to $\mathscr{P}(\mathcal{S}_c)$ –we say that $J\in\mathscr{P}(\mathcal{S}_c)$ is "bounded" if any element of J is a bounded closed segment.

Concerning $\widetilde{\Gamma}_n$, the closed segment $[\![x,y]\!]$ belongs to it if there are $1\leq s\leq n$ pairs $(x_i,y_i)\in\Gamma_n$ such that $[\![x,y]\!]$ is the union of the s segments $[\![x_i,y_i]\!]$ and the length of the intersection $[\![x_i,y_i]\!]\cap[\![x_{i+1},y_{i+1}]\!]$ is greater than $1/n$ for any $1\leq i<s$.

Concerning G_q, it is simply the subset of $\mathcal{S}_c$ made by the closed segments of length q.

Given $J\subseteq\mathcal{S}_c$, we define $Sub(J)\subseteq\mathcal{S}_c$ as the set of all the subsegments of the elements of J: this means that the closed segment C belongs to $Sub(J)$ if and only if there is $C'\in J$ containing C.

Finally, concerning R_ε, given a bounded subset J of $\mathcal{S}_c$ we set

$$R_\varepsilon(J):=\bigcup\Big\{\Big[\!\!\Big[x+\delta_1\frac{y-x}{|y-x|},y-\delta_2\frac{y-x}{|y-x|}\Big]\!\!\Big]:\ |\delta_1|,|\delta_2|\leq\varepsilon,\ |y-x|\geq\delta_1+\delta_2,\ [\![x,y]\!]\in J\Big\}.$$

It is straightforward to note that G_q is closed for any q, as well as $\widetilde{\Gamma}_n$ –recall that Γ_n is compact. Moreover it is also clear that, whenever J is a bounded compact subset of $\mathcal{S}_c$, so are $Sub(J)$ and $R_\varepsilon(J)$.

Step III. *Conclusion.*

We define

$$G := \bigcap_{1/\varepsilon\in\mathbb{N}} \bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} \bigcup_{q\in\mathbb{Q}^+} \varphi\Big(Sub\big(\widetilde{\Gamma}_n\big)\cap G_q \cap R_\varepsilon(K)\Big)\Big\backslash\varphi\Big(Sub\big(\widetilde{\Gamma}_n\big)\cap G_{q+2\varepsilon}\Big)$$

and we claim that ${r_\Gamma}^{-1}(K) = G\cap \mathrm{ET}_\Gamma$, which clearly leads to the conclusion thanks to the claims of the first two steps and Lemma 4.3. To show the preceding equality, first recall that $\{Sub(\widetilde{\Gamma}_n)\}$ is an increasing sequence of closed and bounded subsets of $\mathcal{S}_c$ converging to $Sub(M)$, where M is the set of the closures of the maximal transport rays related to Γ.

Let us then consider a maximal transport ray $A\in M$, whose closure is C; for n sufficiently large, C contains a non-empty maximal transport ray of $\widetilde{\Gamma}_n$, whose closure we denote by B and whose length by l. Then the intersection of C with

$$\varphi\Big(Sub\big(\widetilde{\Gamma}_n\big)\cap G_q \cap R_\varepsilon(K)\Big)\Big\backslash\varphi\Big(Sub\big(\widetilde{\Gamma}_n\big)\cap G_{q+\varepsilon}\Big)$$

is empty unless $q\leq l < q+\varepsilon$, and so the intersection of C with

$$\bigcup_{q\in\mathbb{Q}^+} \varphi\Big(Sub\big(\widetilde{\Gamma}_n\big)\cap G_q \cap R_\varepsilon(K)\Big)\Big\backslash\varphi\Big(Sub\big(\widetilde{\Gamma}_n\big)\cap G_{q+\varepsilon}\Big)$$

is a segment contained in B and whose endpoints are $\varepsilon-$close to those of B if $R_\varepsilon(K)$ contains some segment $\varepsilon-$close to B and contained in B; otherwise, it is empty. It is now clear that $C\cap G=\emptyset$ if $C\notin K$, while $A\subseteq C\cap G\subseteq C$ if $C\in K$; in particular, if $C\in K$ then C is bounded, and its left endpoint x belongs to $C\cap G$ if and only if $x\in\pi_1(\Gamma_n)$ for some n, therefore if and only if $x\in\pi_1(\Gamma)$. It follows that ${r_\Gamma}^{-1}(K) = G\cap\mathrm{ET}_\Gamma$, as claimed, then the thesis follows. ☐

We prove now also the following measurability result:

Lemma 4.5. *If Γ is $\sigma-$compact and (4.1) holds, then the set $MTR\subseteq\mathcal{S}_c$ of the closures of the maximal transport rays is Borel.*

Proof. If Γ is $\sigma-$compact, so is the set $T\subseteq\mathcal{S}_c$ of the closures of the transport rays, then so is $S=Sub(T)$ according to Step II in last lemma. Now, given $C\in\mathcal{S}_c$ and $1/2>\varepsilon>0$, we define

$$C_\varepsilon := \left[\!\left[x-\varepsilon\,\frac{y-x}{|y-x|},\, y+\varepsilon\,\frac{y-x}{|y-x|}\right]\!\right] \qquad C_{-\varepsilon} := \left[\!\left[x+\varepsilon\,\frac{y-x}{|y-x|},\, y-\varepsilon\,\frac{y-x}{|y-x|}\right]\!\right]$$

if $C = [\![x, y]\!]$ is bounded, while if C is unbounded we set $C_\varepsilon = C$ and $C_{-\varepsilon}$ the intersection of C with the closed ball $B(0, 1/\varepsilon)$. Setting then

$$MTR_\varepsilon := \Big\{ C \in \mathcal{S}_c : \; C_{-\varepsilon} \in S, \; C_\varepsilon \notin S \Big\},$$

it is again easy to note that MTR_ε is the difference between two $\sigma-$compact sets, so it is Borel. Finally, the thesis follows since $MTR = \cap\{MTR_\varepsilon : \varepsilon = 1/n, \; n \in \mathbb{N}\}$. $\square$

We give now two other definitions, which will be also of a great importance in the following.

Definition 4.6. *Under assumption (4.1), we say that* $\tau_\Gamma : \mathrm{ET}_\Gamma \to \mathbb{S}^{N-1}$ *is* countably Lipschitz *(resp. countably Lipschitz with respect to* $\mu \in \mathcal{M}^+(\mathbb{R}^N)$*) if* ET_Γ *is the union (resp. the union up to a* $\mu-$*negligible set) of countably many sets* L_i *such that the restriction of* τ_Γ *to each* L_i *is Lipschitz continuous.*

Definition 4.7. *A set* $\Gamma \subseteq \mathbb{R}^N \times \mathbb{R}^N$ *is said to fulfill the* non-inclusion condition *if*

$$(x, y) \in \Gamma, \; (x', y') \in \Gamma \quad \Longrightarrow \quad [\![x, y]\!] \not\subseteq]\!]x', y'[\![\tag{4.2}$$

The importance of this condition relies essentially on the fact that, as we will see in next section, it is a sufficient and necessary condition for the optimality for the strict convex case in $\mathbb{R}$.

4.2 The $1-$dimensional case

In this section we consider the one-dimensional case, i.e. when $X = Y = \mathbb{R}$: in the sequel, to show existence in $\mathbb{R}^N$ we will often reduce to this case. First of all, we give a couple of definitions:

Definition 4.8. *A transport plan* γ *between* f^+ *and* f^- *is said* non-decreasing *if its support* $\Gamma \subseteq \mathbb{R} \times \mathbb{R}$ *fulfills the non-inclusion condition.*

We also define

$$\mathfrak{M} := \big\{ \gamma \in \Pi(f^+, f^-) : \; \forall\, (x, y) \in \operatorname{spt} \gamma, \; y \geq x \big\}, \tag{4.3}$$

which is the set of the transport plans moving all the mass from the left to the right.

All we need to know about the $1-$dimensional case is collected in the following

Theorem 4.9. *There exists only one non-decreasing transport plan $\bar\gamma$, which is induced by a transport map if f^+ is non-atomic. Moreover*

$$\mathfrak{M} \neq \emptyset \Longleftrightarrow \bar\gamma \in \mathfrak{M}. \tag{4.4}$$

Finally, if $\mathfrak{M} \neq \emptyset$, $\min(1.5) < +\infty$ and $c(x, y) = \psi(y - x)$ for a convex (resp. strictly convex) function ψ, then $\bar\gamma$ is a transport plan (resp. the unique transport plan) minimizing the cost $\mathfrak{C}$ in $\mathfrak{M}$.

Proof. **Step I.** *Existence and uniqueness.*
Let γ be a non-decreasing transport plan (if any), and take $(x, y) \in \operatorname{spt}\gamma$. We claim that

$$f^+\big(]\!] - \infty, x[\![\big) \leq f^-\big(]\!] - \infty, y]\!]\big) \qquad f^+\big(]\!] - \infty, x]\!]\big) \geq f^-\big(]\!] - \infty, y[\![\big). \tag{4.5}$$

To show it, assume that the first inequality is false: so

$$\begin{aligned} f^-\big(]\!] - \infty, y]\!]\big) &< f^+\big(]\!] - \infty, x[\![\big) = \gamma\big(]\!] - \infty, x[\![\times\mathbb{R}\big) \\ &\leq f^-\big(\{\tilde y : \exists\, \tilde x \in]\!] - \infty, x[\![:\ (\tilde x, \tilde y) \in \operatorname{spt}\gamma\}\big) = f^-(B), \end{aligned}$$

then it is not possible that $B \subseteq]\!] - \infty, y]\!]$; therefore $\exists\, (\tilde x, \tilde y) \in \operatorname{spt}\gamma$ with $\tilde x < x$, $\tilde y > y$, against the fact that γ is non-decreasing. Then the first inequality in (4.5) is proved, and the proof of the other one is similar. As a consequence, it follows that for any $a, b \in \mathbb{R}$ it is

$$\gamma\big(]\!] - \infty, a[\![\times]\!] - \infty, b[\![\big) = f^+\big(]\!] - \infty, a[\![\big) \wedge f^-\big(]\!] - \infty, b[\![\big);$$

analogously,

$$\gamma\big(]\!] - a, \infty[\![\times]\!] - b, \infty[\![\big) = f^+\big(]\!] - a, \infty[\![\big) \wedge f^-\big(]\!] - b, \infty[\![\big).$$

We can then recover a quite expressive formula for γ: given $a < b$, $\tilde a < \tilde b$ in $\mathbb{R}$, and defining

$$m := f^+\big(]\!] - \infty, a]\!]\big),\quad M := f^+\big(]\!] - \infty, b[\![\big),\quad \widetilde m := f^+\big(]\!] - \infty, \tilde a]\!]\big),\quad \widetilde M := f^+\big(]\!] - \infty, \tilde b[\![\big),$$

the two preceding equations give

$$\gamma\Big(]\!]a, b[\![\times]\!]\tilde a, \tilde b[\![\Big) = \Big(\big(M \wedge \widetilde M\big) - \big(m \vee \widetilde m\big)\Big) \vee 0. \tag{4.6}$$

This formula gives the uniqueness of a non-decreasing transport plan. Concerning the existence, we can argue as in [4]: if f^+ and f^- have compact supports, then the infimum of the mass transport problem with the continuous cost $c(x, y) = \sqrt{1 + |y - x|^2}$ is finite, then any optimal transport plan γ must have a $c-$cyclically monotone support by Proposition 3.2; but since $c(x, y)$ is a strictly convex function of $|y - x|$, the $c-$cyclical monotonicity of $\operatorname{spt}\gamma$ immediately implies the validity of the non-inclusion condition, then the existence of a non-decreasing transport plan is established under the assumption that f^+ and f^- have compact supports. For general measures f^+ and f^- it suffices to take measures f_k^+ and f_k^- with the same total mass, compact supports and weakly* converging to f^+ and f^- respectively: then there exist non-decreasing transport plans γ_k between f_k^+ and f_k^-, and it is easilly checked that any limit point of $\{\gamma_k\}$ is a non-decreasing transport plan between f^+ and f^-. Then both the existence and the uniqueness of a non-decreasing transport plan, which we will denote by $\bar\gamma$ in the rest of this proof, are proved.

Step II. *Non-decreasing transport map.*

We define the following two functions:

$$\begin{aligned} t_>(x) &:= \inf\big\{y \in \mathbb{R} : \ f^+\big(\rrbracket -\infty, x\rrbracket\big) < f^-\big(\rrbracket -\infty, y\llbracket\big)\big\}, \\ t_<(x) &:= \sup\big\{y \in \mathbb{R} : \ f^+\big(\rrbracket -\infty, x\llbracket\big) > f^-\big(\rrbracket -\infty, y\rrbracket\big)\big\}; \end{aligned} \tag{4.7}$$

$t_>$ and $t_<$ are non-decreasing maps from $\mathbb{R}$ to itself and that $t_< \le t_>$. Moreover, comparing with (4.5), it is immediately seen that whenever $(x, y) \in \operatorname{spt}\bar\gamma$ it is $t_<(x) \le y \le t_>(x)$. We finally note that a consequence of (4.7) is that $t_<(x+\varepsilon) \ge t_>(x)$ for any $x \in \mathbb{R}$ and $\varepsilon > 0$, thus $t_<(x) = t_>(x)$ for all real x but countably many. If f^+ is non-atomic, then, these countably many points are f^+−negligible and then, applying Lemma 1.6, we infer that $t_< = t_>$ (w.r.t. f^+) is a transport map inducing γ.

Step III. *Proof of (4.4).*

Assume now that the non-decreasing transport plan $\bar\gamma$ is not contained in $\mathfrak{M}$: this implies by (4.5) the existence of some $\tilde x > \tilde y$ in $\mathbb{R}$ such that $f^+\big(\rrbracket -\infty, x\llbracket\big) < f^-\big(\rrbracket -\infty, y\llbracket\big)$. Given any transport plan γ and arguing as when showing the validity of (4.5), we see that there are some $x \ge \tilde x$ and $y \le \tilde y$ such that $(x, y) \in \operatorname{spt}\gamma$: thus also γ is not contained in $\mathfrak{M}$ and then the non-trivial implication of (4.4) is proved.

Step IV. *Optimality of $\bar\gamma$.*

We finally assume that $\mathfrak{M} \neq \emptyset$, $\min(1.5) < +\infty$ and $c(x, y) = \psi(y-x)$ for a strictly convex function ψ; take then any $\gamma \in \mathfrak{M}$: it is clear that γ fails to be non-decreasing (thus fails to equal $\bar\gamma$) if and only if there are (x, y), $(\tilde x, \tilde y)$ in $\operatorname{spt}\gamma$ such that $\llbracket\tilde x, \tilde y\rrbracket \subseteq \llbracket x, y\rrbracket$. But in this case

$$c(x, y) + c(\tilde x, \tilde y) > c(\tilde x, y) + c(x, \tilde y)$$

is an immediate consequence of the strict convexity of ψ, and the same inequality must hold also in a small neighborhood of the points x, y, $\tilde x$, $\tilde y$; since (x, y) and $(\tilde x, \tilde y)$ are in $\operatorname{spt}\gamma$, one derives that any set of full γ−measure is not c−cyclically monotone, thus γ is not optimal by Theorem 3.16. Since $\mathfrak{M}$ is weak* closed, there must be minimizers of $\mathfrak{C}$ in $\mathfrak{M}$, then $\bar\gamma$ is the unique transport plan minimizing $\mathfrak{C}$ in $\mathfrak{M}$ under the hypothesis that

ψ is strictly convex. By approximation, we infer that $\bar{\gamma}$ is a minimizer (possibly not the unique one) even in the case when the function ψ is only assumed to be convex. $\square$

Another easy but useful result concerning the $1-$dimensional case, which also underlines that the fact that the uniqueness part in the above Theorem is not true without the strict convexity hypothesis, is the following

Lemma 4.10. *If the cost function is $c(x,y)=|y-x|$ and c is slow$-(f^+,f^-)-$increasing , then all the transport plans belonging to $\mathfrak{M}$ (if there are) have the same cost.*

Proof. If γ belongs to $\mathfrak{M}$, then for $\gamma-$a.e. (x,y) one has $y \geq x$ and then $c(x,y)=|y-x|=y-x$. It follows that

$$\mathfrak{C}(\gamma)=\iint_{\mathbb{R}\times\mathbb{R}} y-x\,d\gamma(x,y).$$

Moreover, by Remark 3.21 f^+ and f^- have finite first order moments, that means $\int |x|\,df^+(x)<+\infty$ and $\int |y|\,df^-(y)<+\infty$: then the above integral splits giving

$$\mathfrak{C}(\gamma)=\int_{\mathbb{R}} y\,df^- - \int_{\mathbb{R}} x\,df^+,$$

thus the cost is not dependent on γ, provided it is a plan belonging to $\mathfrak{M}$. $\square$

4.3 Secondary variational problems

In this section we restrict ourselves to the case when the cost function $c(x,y)=\|y-x\|$ is induced by some norm $\|\cdot\|$ in $\mathbb{R}^N$. Then we will define two secondary minimization problems, which allow us to select some particular optimal transport plans; in particular, the minimizers of these problems will be optimal transport maps in some situations, namely the Euclidean norm, the quasi-euclidean norm and the crystalline norm (see Chapters 6, 8 and 9).

We begin briefly explaining why one should try to find transport plans with some additional property. In last section we discussed the no-crossing condition (4.1) and we saw some consequences of its validity, and we will make strongly use of this property in the next chapters: when we will build an optimal transport map, we will take an optimal transport plan γ and our idea will be to move the mass along the maximal transport rays associated to the set $\Gamma=\operatorname{spt}\gamma$; the validity of (4.1) will then be essential since we will need the different maximal transport rays to be disjoint. Lemma 3.18 already ensures that, if the norm is strictly convex, then $\operatorname{spt}\gamma$ fulfills the no-crossing condition whenever γ is optimal; however, there are some interesting situations where the norm is not strictly convex, for example the crystalline norms that we will see in Chapter 9. To understand what could happen in the non strictly convex case, we discuss a simple example: consider four points x_1, x_2, y_1 and y_2 as in Figure 4.1 and let $f^+ = \delta_{x_1} + \delta_{x_2}$ and $f^- = \delta_{y_1} + \delta_{y_2}$.

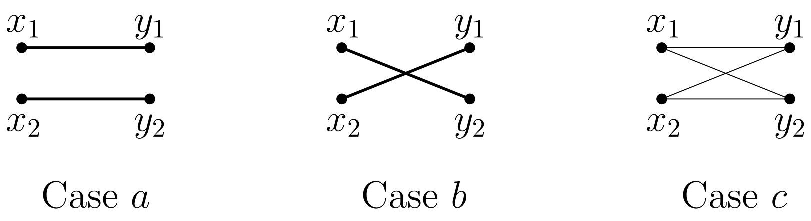


Figure 4.1: The transport rays for γ_a, γ_b and γ_c

Suppose that $c(x_i, y_j) = 1$ for $i = 1, 2$ and $j = 1, 2$, that is clearly possible with a norm, but only if it is not strictly convex. Consider then the three possible transport plans γ_a, γ_b and γ_c defined as follows:

$$\gamma_a = \delta_{x_1} \otimes \delta_{y_1} + \delta_{x_2} \otimes \delta_{y_2}, \qquad \gamma_b = \delta_{x_1} \otimes \delta_{y_2} + \delta_{x_2} \otimes \delta_{y_1}, \qquad \gamma_c = \frac{\gamma_1 + \gamma_2}{2};$$

the transport rays associated to these transport plans are shown in Figure 4.1: it is clear that all these plans are optimal, but only the first one has a support fulfilling hypothesis (4.1). In general, we may not expect that (4.1) holds for the support Γ of *any* optimal transport plan when c is induced by a non strictly convex norm; on the other hand, it is reasonable to hope that, as in the preceding example, there is *some* optimal transport plan whose support fulfills the no-crossing condition. The first secondary problem we present has then the purpose to select an optimal transport plan fulfilling (4.1).

We begin then setting the following minimization problem

$$\inf \left\{ \mathfrak{C}_{\mathrm{eucl}}(\gamma) := \iint_{\mathbb{R}^N \times \mathbb{R}^N} |y - x| \, d\gamma(x, y) : \ \gamma \in \Pi'(f^+, f^-) \right\}; \tag{4.8}$$

recall that by $|\cdot|$ we always denote the Euclidean norm in $\mathbb{R}^N$; we also recall that, as defined in Chapter 1, $\Pi(f^+, f^-)$ is the set of the transport plans between f^+ and f^- and $\Pi'(f^+, f^-)$ is its non-empty subset made by all the optimal transport plans. We will then denote by $\Pi''(f^+, f^-)$ the set of the minimizers in (4.8). As one can imagine, the main difference between the minimization of (1.5) and (4.8) is the validity of (4.1); indeed, if the norm is strictly convex then $\Pi' = \Pi''$, as we will prove in Lemma 4.16.

Solving problem (4.8) leads to a measure as we were searching for, thanks to the following

Theorem 4.11. *The following facts hold:*

1: The set $\Pi''(f^+, f^-)$ is contained in $\Pi'(f^+, f^-)$ and it is non-empty, weak closed, bounded and convex; in particular, there are minimizers of (4.8).*

2: If c is slow $-(f^+, f^-)-$ increasing then any $\gamma \in \Pi''(f^+, f^-)$ is concentrated in a $c-$cyclically monotone $\sigma-$compact set Γ'_γ fulfilling the no-crossing condition (4.1).

3: There is a $c-$cyclically monotone Borel set Γ' fulfilling the no-crossing condition with the property that any minimizer of (4.8) is concentrated in Γ'.

4: If there is a set Γ' as before which moreover is $\sigma-$compact and such that $f^+(\mathrm{D}'_\Gamma) = 0$, then in fact any transport plan between f^+ and f^- is a minimizer of (4.8) if and only if it is concentrated in Γ'.

Proof. **Step I.** *Properties of* $\Pi''(f^+, f^-)$.
We already know (recall Theorem 1.11) that $\Pi'(f^+, f^-)$ is a non-empty, weak* closed, bounded and convex subset of $\mathcal{M}^+(\mathbb{R}^N \times \mathbb{R}^N)$; moreover the functional $\mathfrak{C}_{\mathrm{eucl}}$ is clearly weakly* l.s.c. and linear on $\mathcal{M}^+(\mathbb{R}^N \times \mathbb{R}^N)$. The fact that $\Pi''(f^+, f^-)$ is non-empty, weakly* closed and convex is then straightforward; moreover $\Pi''(f^+, f^-)$ is included in $\Pi'(f^+, f^-)$ by definition, then it is bounded because so is $\Pi'(f^+, f^-)$.
Step II. *Equivalent transport problem.*
Assume now that c is *slow* $-(f^+, f^-)-$ *increasing*: according to Corollary 3.27, we can take a Kantorovich potential u and we know that γ is an optimal transport plan (with respect to the cost function $c(x, y) = \|y - x\|$) if and only if $c(x, y) = u(x) - u(y)$ for $\gamma-$a.e. (x, y). We claim that a transport plan is a minimizer for (4.8) if and only if it is optimal with respect to the auxiliary cost function

$$c'(x, y) := \begin{cases} |y - x| & \text{if } (x, y) \in \partial^c u \\ +\infty & \text{otherwise:} \end{cases} \tag{4.9}$$

note that the function c' is l.s.c. because $\partial^c u$ is closed: in fact, c is continuous by definition, and u is $1-$Lipschitz with respect to c, then also continuous. To show our claim, note that the cost $\mathfrak{C}'(\gamma)$ of a transport plan γ with respect to c' is $+\infty$ if $c(x, y) > u(x) - u(y)$ for a set of pairs (x, y) of positive $\gamma-$measure, thus if γ is non optimal with respect to c; otherwise, $\mathfrak{C}'(\gamma) = \mathfrak{C}_{\mathrm{eucl}}(\gamma)$ if $\gamma \in \Pi'(f^+, f^-)$. To deduce that minimizing (4.8) is equivalent to minimize $\mathfrak{C}'$, we have then only to exclude that $\min(4.8) = +\infty$: otherwise the minimizers of $\mathfrak{C}'$ would be all the elements of $\Pi(f^+, f^-)$, while $\Pi''(f^+, f^-) = \Pi'(f^+, f^-)$ could be in principle strictly smaller. But since $\|\cdot\|$ and $|\cdot|$ are two norms in $\mathbb{R}^N$, there exists a constant ε such that $\varepsilon \|\cdot\| \le |\cdot| \le \varepsilon^{-1} \|\cdot\|$: then $\inf(4.8) \le \varepsilon^{-1} \min(1.5) < +\infty$ and our claim is proved.

Step III. *The $c'-$cyclical monotonicity implies the no-crossing property.*
Let us assume that a set $\Lambda \subseteq \mathbb{R}^N \times \mathbb{R}^N$ is contained in $\partial^c u$ and it is $c'-$cyclically monotone; assume moreover that there are two pairs (x, y) and $(\tilde{x}, \tilde{y})$ in Λ against the no-crossing property of Λ: this means that the two segments $[\![x, y]\!]$ and $[\![\tilde{x}, \tilde{y}]\!]$ cross in a point z which is not the left extreme of both the segments, nor the right one. The $c'-$cyclical monotonicity and the fact that (x, y) and $(\tilde{x}, \tilde{y})$ belong to Λ imply

$$|y - x| + |\tilde{y} - \tilde{x}| \leq c'(x, \tilde{y}) + c'(\tilde{x}, y). \tag{4.10}$$

Moreover, recalling that $\Lambda \subseteq \partial^c u$ and that c is induced by the norm $\| \cdot \|$, one has that $\|y - x\| = u(x) - u(y)$; but in general, since u is a Kantorovich potential,

$$\|y - z\| \geq u(z) - u(y) \qquad\qquad \|z - x\| \geq u(x) - u(z),$$

so one derives that both the equalities hold. Since the same argument works with $\tilde{x}$ and $\tilde{y}$, we know

$$\begin{aligned} \|y - z\| &= u(z) - u(y), & \|z - x\| &= u(x) - u(z), \\ \|\tilde{y} - z\| &= u(z) - u(\tilde{y}), & \|z - \tilde{x}\| &= u(\tilde{x}) - u(z). \end{aligned} \tag{4.11}$$

On the other hand, again recalling that u is a Kantorovich potential, we have

$$\|\tilde{y} - x\| \geq u(x) - u(\tilde{y}) \qquad\qquad \|y - \tilde{x}\| \geq u(\tilde{x}) - u(y); \tag{4.12}$$

finally, the triangle inequality ensures

$$\|\tilde{y} - x\| \leq \|\tilde{y} - z\| + \|z - x\| \qquad\qquad \|y - \tilde{x}\| \leq \|y - z\| + \|z - \tilde{x}\|. \tag{4.13}$$

Putting together (4.11), (4.12) and (4.13), one obtains that all the inequalities are equalities, therefore in particular both $(x, \tilde{y})$ and $(x, \tilde{y})$ belong to $\partial^c u$. Then (4.10) reads as

$$|y - x| + |\tilde{y} - \tilde{x}| \leq |\tilde{y} - x| + |y - \tilde{x}|,$$

which is false if the segments $[\![x, y]\!]$ and $[\![\tilde{x}, \tilde{y}]\!]$ contradict the no-crossing property. We derived that any $c'-$cyclically monotone subset of $\partial^c u$ fulfills the no-crossing property.

In particular, we immediately derive the second claim of the Theorem: in fact, if $\gamma \in \Pi''(f^+, f^-)$ then it is an optimal transport plan with respect to the cost function c', as proved in Step II: then, since we noticed that $\mathfrak{C}'(\gamma) < +\infty$, according to Theorem 3.16 we derive that γ is concentrated in a $c'-$cyclically monotone and $\sigma-$compact set Γ'_γ on which $c' < +\infty$, that means $\Gamma'_\gamma \subseteq \partial^c u$. Finally, Γ'_γ is $c-$cyclically monotone because so is $\partial^c u$, and it fulfills the no-crossing property because it is $c'-$cyclically monotone.
Step IV. *Definition of* Γ'.
We want now build a set Γ' as in the claim of the Theorem: to do that, let $\{M_n\}$ be

an increasing sequence of integers (to be determined later) and define the approximating cost functions

$$c_n(x,y) := \Big(|y-x| \cdot \big(1 + M_n \cdot \big(\|y-x\| - u(x) + u(y)\big)\big)\Big) \wedge M_n|y-x| : \tag{4.14}$$

note that c_n is an increasing sequence of continuous functions converging from below to c': moreover, since c is $slow-(f^+,f^-)-increasing$, $|\cdot| \le \varepsilon^{-1}\|\cdot\|$ and $c_n(x,y) \le M_n|y-x|$, we infer that also c_n is $slow-(f^+,f^-)-increasing$: then there exists an optimal dual pair (u_n, v_n) for the problem (3.3) with the cost function c_n by Theorem 3.22. We want to show that the sequence $\{(u_n, v_n)\}$ is optimal for the problem (3.3) relative to the cost function c'. In fact, $u_n(x) - u_n(y) \le c_n(x,y) \le c'(x,y)$ for any $n \in \mathbb{N}$ and any pair (x,y), therefore we need to show that

$$\min \iint c_n(x,y)\, d\gamma(x,y) \longrightarrow \min \iint c'(x,y)\, d\gamma(x,y) = \min(4.8). \tag{4.15}$$

But if $\gamma \in \Pi''(f^+,f^-)$, then it is concentrated in $\partial^c u$, and in this set the functions c' and c_n are equal: this means that

$$\limsup_{n\to\infty} \min \iint c_n(x,y)\, d\gamma(x,y) \le \min(4.8).$$

On the other hand, assume the existence of $\delta > 0$ and of transport plans γ_n such that $\iint c_n\, d\gamma_n \le \min(4.8) - \delta$: take then a weak* limit γ of a subsequence (not relabeled) of $\{\gamma_n\}$, fix $m \in \mathbb{N}$, and then evaluate for $n \ge m$ (recalling that $c_n \ge c_m$ and that c_m is continuous)

$$\min(4.8) - \delta \ge \iint c_n(x,y)\, d\gamma_n(x,y) \ge \iint c_m(x,y)\, d\gamma_n(x,y) \xrightarrow[n\to+\infty]{} \iint c_m(x,y)\, d\gamma(x,y).$$

But $\iint c_m\, d\gamma \to \iint c'\, d\gamma$ by the Monotone Convergence Theorem, and this leads to an absurd because $\iint c'\, d\gamma \ge \min(4.8)$: thus (4.15) is established; since (u_n, v_n) is an optimal dual pair for the problem (3.3) with the cost function c_n, we know that

$$\min \iint c_n(x,y)\, d\gamma(x,y) = \int u_n(x)\, df^+(x) + \int v_n(y)\, df^-(y) :$$

then, comparing with (4.15) and recalling (4.14), we can finally choose the sequence of integers $\{M_n\}$ in such a way that

$$\int u_n(x)\, df^+(x) + \int v_n(y)\, df^-(y) \ge \min(4.8) - \frac{1}{2^n}. \tag{4.16}$$

Take now a point $x \in \mathbb{R}^N$, and choose $y \in \mathbb{R}^N$ such that $u_n(x) \ge c_n(x,y) - v_n(y)$: for any $\tilde{x} \in \mathbb{R}^N$ one has

$$u_n(x) - u_n(\tilde{x}) \ge c_n(x,y) - c_n(\tilde{x},y) - \varepsilon.$$

Since $c_n(\cdot,y)$ is continuous at x, we derive that $u_n(x)-u_n(\tilde{x})\geq -2\varepsilon$ if $|\tilde{x}-x|\ll 1$: then u_n is u.s.c., and analogously so is v_n. Let us then define the "error"

$$r_n(x,y):=c'(x,y)-u_n(x)-v_n(y)\geq 0,$$

and finally the set

$$\Gamma':=\Big\{(x,y)\in\mathbb{R}^N\times\mathbb{R}^N:\ r_n(x,y)\xrightarrow[n\to+\infty]{}0\Big\}.$$

Step V. *Properties of* Γ'.
We want to show that Γ' has all the properties stated: concerning the Borel property, recall that by construction r_n is l.s.c. –because sum of three l.s.c. functions–, so each set $\{r_n\leq\varepsilon\}$ is closed, therefore $\Gamma'=\cap_\varepsilon\cup_n\cap_{m\geq n}\{r_m\leq\varepsilon\}$ is a countable intersection of $\sigma-$*compact* sets, thus a Borel set. Concerning the $c-$cyclical monotonicity, it suffices to notice that $r_n\equiv+\infty$ out of $\partial^c u$, then Γ' is a subset of $\partial^c u$, thus a $c-$cyclical monotone set.

We show now that Γ' is $c'-$cyclically monotone: in fact, given (x,y) and $(\tilde{x},\tilde{y})$ in Γ', first of all we notice that $c'(x,y)$ and $c'(\tilde{x},\tilde{y})$ are finite because $\Gamma'\subseteq\partial^c u$; then also $r_n(x,y)$ and $r_n(\tilde{x},\tilde{y})$ are finite numbers. We need to show

$$c'(x,y)+c'(\tilde{x},\tilde{y})\leq c'(x,\tilde{y})+c'(\tilde{x},y), \tag{4.17}$$

and the inequality is trivial unless both $c'(x,\tilde{y})$ and $c'(\tilde{x},y)$ are finite. In this case, also $r_n(x,\tilde{y})$ and $r_n(\tilde{x},y)$ are finite, and then we can write

$$c'(x,y)+c'(\tilde{x},\tilde{y})-r_n(x,y)-r_n(\tilde{x},\tilde{y})=c'(x,\tilde{y})+c'(\tilde{x},y)-r_n(x,\tilde{y})-r_n(\tilde{x},y),$$

since both the terms equal $u_n(x)+u_n(\tilde{x})+v_n(y)+v_n(\tilde{y})$. When n goes to $+\infty$, $r_n(x,y)$ and $r_n(\tilde{x},\tilde{y})$ converge to 0 by hypothesis, while $r_n(x,\tilde{y})$ and $r_n(\tilde{x},y)$ are always positive. Then (4.17) holds, and so the $c'-$cyclical monotonicity follows. By Step III, also the no-crossing property of Γ' is then established.

To finish the proof of the third claim of the Theorem, we only need to show that any $\gamma\in\Pi''(f^+,f^-)$ is concentrated in Γ': but for any $\gamma\in\Pi(f^+,f^-)$ one has

$$\iint c'(x,y)\,d\gamma(x,y)=\int u_n(x)\,df^+(x)+\int v_n(y)\,df^-(y)+\iint r_n(x,y)\,d\gamma(x,y).$$

Moreover we already pointed out that $\int u_n\,df^+ + \int v_n\,df^-$ converges to the minimum of $\mathfrak{C}'$, then $\gamma\in\Pi''(f^+,f^-)$ if and only if $\iint r_n\,d\gamma\to 0$; since $r_n\geq 0$, if the preceding property is true then γ is concentrated in the set of the points where r_n converges to 0: this is usually true passing to a subsequence, but in this case there is no need of it thanks to the estimate (4.16). Summarizing, if γ is a minimizer of (4.8) then it is concentrated in Γ', so that also the third claim is proved. Note that, in principle, the converse is not true in general, i.e. there could be a measure γ concentrated in Γ' such that $\iint r_n \, d\gamma$ does not converge to 0.

Step VI. *Conclusion.*

We finally prove the last claim of the Theorem: assume there is a set Γ' with the properties of the third claim and with the additional properties to be $\sigma-$compact and to have negligible doubling points, i.e. $f^+(\mathrm{D}'_\Gamma) = 0$; let moreover γ be a transport plan between f^+ and f^- concentrated in Γ'. The plan γ must leave fixed any point of $\mathrm{F}_{\Gamma'}$, and the same is done by any transport plan minimizing (4.8), then we can directly consider the case when $f^+(\mathrm{F}_{\Gamma'}) = 0$: in this case, $\mathrm{ET}_{\Gamma'}$ is a set of full f^+-measure, then the map ${r_{\Gamma}}' : \mathrm{ET}_{\Gamma'} \to \mathcal{S}_c$, which is Borel thanks to Lemma 4.4, is defined f^+-a.e.. We can then consider the disintegration of f^+ with respect to ${r_{\Gamma}}'$, leading to the decomposition $f^+ = f^+_C \otimes \nu$, where $\nu = {r_{\Gamma}}'_\# f^+$ is a measure on $\mathcal{S}_c$. Since γ is concentrated in Γ', the map ${r_{\Gamma}}' \circ \pi_1$ associates to $\gamma-$a.e. pair (x, y) the closure of the maximal transport ray containing $[\![x, y]\!]$, therefore recalling the properties of the disintegration we can write $\gamma = \gamma_C \otimes \nu$ and each γ_C is a transport plan on the segment C. The first projection of γ_C is the measure f^+_C mentioned above, which depends on f^+ and Γ', but not on γ, provided it is concentrated in Γ'. One can prove that also the second projection of γ_C, call it f^-_C, does not depend on the plan γ concentrated in Γ', but only on f^+, f^- and Γ': we do not prove it now, since it will be done carefully in Lemma 4.15.

Summarizing, any transport plan γ concentrated in Γ' is nothing else than

$$\gamma = \gamma_C \otimes \nu,$$

where the measure ν is given and any plan γ_C is a transport plan between the given measures f^+_C and f^-_C on C; in particular, since the maximal transport rays are oriented, the plan γ_C moves the mass from left to right in the sense of the definition (4.3). Note now that the cost function in each ray C is the Euclidean one; moreover, the fact that c is *slow* $-\,(f^+, f^-)\,-$ *increasing* means, by Remark 3.21, that the first order moments of f^+ and f^- with respect to $\|\cdot\|$ are finite: then the same is true for the first order moments with respect to $|\cdot|$, since $|\cdot| \leq 1/\varepsilon \|\cdot\|$, and this implies that for $\nu-$a.e. C the first order moments of f^+_C and f^-_C with respect to the Euclidean norm are finite: then, according to Lemma 4.10, one obtains that also the euclidean cost $\mathfrak{C}(C)$ of γ_C does not depend on γ for $\nu-$a.e. C. Since the Euclidean cost of γ is

$$\mathfrak{C}_{\mathrm{eucl}}(\gamma) = \mathfrak{C}'(\gamma) = \int_{\mathcal{S}_c} \mathfrak{C}(C) \, d\nu(C),$$

it follows that all the transport plans concentrated in Γ' have the same Euclidean cost, then in particular they are all minimizers of (4.8) since all the minimizers of (4.8) are concentrated in Γ'. The thesis follows. $\square$

We are now ready to set the other secondary variational problem (which is then a "tertiary" problem). To understand the reason why we do that, consider that our main intention is to solve the Monge transport problem; with the first variational problem, we already reduced ourselves to the case of optimal transport plans that move mass only on the disjoint maximal transport rays. So it is reasonable that we need first of all to find optimal transport maps on the rays, which are segments. In Section 4.2 we already saw that, working on a segment, an optimal transport plan is induced by a map if f^+ is non-atomic and the cost is strictly convex; on the other hand, if we take the Euclidean distance there are optimal transport plans which are not induced by any transport map, even if f^+ is non-atomic. In Section 4.2 we also observed that, in dimension 1, the strict convexity of the cost implies the non-inclusion condition (4.2) for the support of any optimal transport map and conversely a transport plan fulfilling (4.2) is induced by a transport map if f^+ is non-atomic. Then we may try to select, from the measures contained in $\Pi''(f^+, f^-)$, some measure concentrated in a set realizing (4.2).

We set then the following minimization problem

$$\inf\left\{\mathfrak{C}_{\text{conv}}(\gamma) := \iint_{\mathbb{R}^N\times\mathbb{R}^N} \alpha(y-x)\,d\gamma(x,y) :\ \gamma \in \Pi''(\mu,\nu)\right\} \tag{4.18}$$

with $\alpha(v) = \sqrt{1+|v|^2}$. Note that α is a strictly convex real function of $\mathbb{R}^N$, and that $\inf(4.18) < +\infty$ if c is *slow* $-$ (f^+, f^-) $-$ *increasing*, because $\alpha(v) \le |v| + 1$ and then any minimizer of (4.8) has finite cost $\mathfrak{C}_{\text{conv}}$: provided that it is strictly convex and that $\inf(4.18) < +\infty$, there is *usually* not any difference if we take another α in the definition of problem (4.18); the meaning of this claim will be clear after Lemma 4.15 and Example 4.18. We define then, in analogy with Π, Π' and Π'', the set $\Pi'''(f^+, f^-)$ of the minimizers of (4.18). We see now that solving problem (4.18) leads to the measure we were searching for, thanks to the following

Lemma 4.12. *The set $\Pi'''(f^+, f^-)$ is contained in $\Pi''(f^+, f^-)$ and it is non-empty, weakly* closed, bounded and convex; in particular, there are minimizers of (4.18). Moreover, if c is slow $-$ (f^+, f^-) $-$ increasing then for any $\gamma \in \Pi'''(f^+, f^-)$ there is a $c-$cyclically monotone and Borel set Γ'' fulfilling the no-crossing condition (4.1) and the non-inclusion condition (4.2) such that γ is concentrated in Γ''.*

Proof. The existence of minimizers for (4.18) is straightforward: thanks to the fact that $\alpha(y-x)$ is a continuous function of (x,y), since $\Pi''(f^+,f^-)$ is non-empty, weakly* closed, bounded and convex so is also $\Pi'''(f^+,f^-) \subseteq \Pi''(f^+,f^-)$.

Assume now that c is *slow* $-$ (f^+,f^-) $-$ *increasing*, and select a transport plan $\gamma \in \Pi'''(f^+,f^-)$: by Theorem 4.11, we already know the existence of a set Γ', depending on γ, which is $\sigma-$compact, fulfills the no-crossing condition and in which γ is concentrated. We want to find a Borel subset Γ'' of Γ' which fulfills also the non-inclusion condition and on which γ is still concentrated. To do that, first of all we notice that we can assume without loss of generality that the set $\mathrm{F}_{\Gamma'}$ is empty, since the plan γ must keep fixed all the mass contained in that set; then, we want to decompose γ on the different maximal transport rays. By Lemma 4.5, we know that the set $MTR \subseteq \mathcal{S}_c$ of the closures of the maximal transport rays associated to Γ' is Borel: we consider then the map $r : \Gamma' \to MTR$ associating to each pair $(x,y) \in \Gamma'$ the closure of the maximal transport ray containing $[\![x,y]\!]$: this is a Borel map since it is the composition of the Borel map ${r_{\Gamma}}' : \mathrm{ET}_{\Gamma'} \to MTR$ with the Borel map which associates to any pair $(x,y) \in \Gamma'$ the point $(x+y)/2 \in \mathrm{ET}_{\Gamma'}$. We can then disintegrate $\gamma = \gamma_C \otimes r_\#\gamma$; let us consider now any segment C: recalling the $1-$dimensional case we discussed in Section 4.2, we know that all the plans contained in $\mathfrak{M}$ between $\pi_1\gamma_C$ and $\pi_2\gamma_C$ have the same Euclidean cost, while $\bar{\gamma}_C$, the unique non-decreasing plan, is the unique minimizer of the cost $c(x,y) = \alpha(y-x)$ since α is strictly convex. By Lemma 1.13, we infer that $\gamma_C = \bar{\gamma}_C$ for $r_\#\gamma-$a.e. C; moreover, again from Section 4.2 we know that the support of any $\bar{\gamma}_C$ fulfills the non-inclusion condition. We can then select a Borel subset Λ of MTR of full $r_\#\gamma-$measure such that $\gamma_C = \bar{\gamma}_C$ for any $C \in \Lambda$, and finally $\Gamma'' = \bigcup \big\{ \operatorname{spt} \gamma_C,\ C \in \Lambda \big\}$ is clearly a subset of Γ' of full $\gamma-$measure and fulfilling the non-inclusion condition. The fact that Γ'' is Borel is proved with an approximation argument in Lemma 4.6 of [4]. $\square$

We briefly point out that our secondary problems can be introduced in different ways: we used the variational formulation in (4.8) and (4.18), but we also saw that (4.8) can be seen as a standard transport problem with the auxiliary cost function c'; moreover, there is another interesting approach, already used in [2]: the idea is to make use of the theory of $\Gamma-$asymptotic expansions in $\Gamma-$convergence (see [5, 6, 28]) to define functionals ${\mathfrak{C}_\varepsilon}'$ (resp. ${\mathfrak{C}_\varepsilon}''$) on $\Pi(f^+,f^-)$ whose $\Gamma-$limit admits as minimizers exactly the solutions of (4.8) (resp. (4.18)). We will discuss and use this approach in Section 8.2.

An easy consequence of the preceding result, which will be useful in Chapters 8 and 9 to derive uniqueness of optimal transport maps, is the following:

Lemma 4.13. *If each measure solving (4.18) is a transport map, then there is a unique solution of (4.18).*

Proof. This is a really easy consequence of the convexity of the set $\Pi'''(f^+, f^-)$: suppose that there are two solutions γ_1 and γ_2, which by hypothesis must be induced by the transport maps t_1 and t_2; then, after a disintegration, we have $\gamma_1 = \gamma_{1x} \otimes f^+$ and $\gamma_2 = \gamma_{2x} \otimes f^+$, where $\gamma_{1x} = \delta_{t_1(x)}$ and $\gamma_{2x} = \delta_{t_2(x)}$ for f^+-a.e. x; let now $\gamma = (\gamma_1 + \gamma_2)/2$: it is also a solution of (4.18) since $\Pi'''(f^+, f^-)$ is convex, then it is induced by a transport map t. This means that $\gamma_x = \delta_{t(x)}$ for f^+-a.e. x; but since $\gamma_x = (\delta_{t_1(x)} + \delta_{t_2(x)})/2$ for f^+-a.e. x, this is not possible unless $\gamma_1 = \gamma_2$. $\square$

As we said right after setting problem (4.18), under suitable integrability hypotheses any strictly convex function α such that $\inf(4.18) < +\infty$ leads to the same solutions as $\alpha(v) = \sqrt{1 + |v|^2}$ (for example, in [2] it was used the function $\alpha(v) = |v| \ln |v|$, the reason for this choice will become clear in Section 8.3). Let us make now this statement more clear: first of all, we assume that c is *slow* $- (f^+, f^-) -$ *increasing* and we take the set Γ' as in Theorem 4.11: we know that all the plans in $\Pi''(f^+, f^-)$ move the mass only along the closures of the maximal transport rays of Γ'; on the other hand, we know from Section 4.2 that on a segment there is a unique solution of the mass transport problem, which is the unique non-decreasing transport plan. Then we may guess that there is a unique solution of (4.18), which is the unique optimal transport plan on each maximal transport ray, and this does not depend on the choice of α, provided that it is strictly convex (to effectively select the non-decreasing transport plan) and that $\inf(4.18) < +\infty$ (to avoid the ill-posed situation where each $\gamma \in \Pi''(f^+, f^-)$ is optimal for (4.18)). As we will see now, there is another hypothesis to be made, namely the existence of a set Γ' as in the fourth claim of Theorem 4.11: this is reasonable, since for the doubling points there is not a unique maximal transport ray along which to be moved; moreover, we will show with Example 4.18 that this hypothesis is sharp.

Remark 4.14. *It is also to be noticed that if α is convex and $\gamma \in \Pi(f^+, f^-)$, then α is semi-integrable with respect to γ, since otherwise (4.18) makes no sense. But in fact, if c is slow $- (f^+, f^-) -$ increasing then f^+ and f^- have finite first order moments thanks to Remark 3.21 and then, since α is greater than some affine function, it follows that α is $\gamma-$integrable and $\iint \alpha \, d\gamma > -\infty$. More generally, the same argument shows also $\iint \alpha \, d\gamma \geq K > -\infty$ for some K depending on f^+ and f^- but not on $\gamma \in \Pi(f^+, f^-)$; therefore,* $\inf(4.18) > -\infty$.

Lemma 4.15. *Let us assume that c is slow$-(f^+, f^-)-$increasing and that there is a set Γ' as in the fourth claim of Theorem 4.11; then the set $\Pi'''(f^+, f^-)$ does not depend on the* *choice of α in (4.18), provided it is a strictly convex function such that* $\inf(4.18) < +\infty$. *In particular,* $\Pi'''(f^+, f^-)$ *contains a unique element, which does not depend on* α.

Proof. First of all let us notice that any solution of (4.8) must leave fixed any point $x \in \mathrm{F}_{\Gamma'}$ because it is concentrated in Γ': this ensures that $f^+ \llcorner \mathrm{F}_{\Gamma'} = f^- \llcorner \mathrm{F}_{\Gamma'}$, and that if γ is a solution of (4.8) then $\gamma \llcorner \left(\mathrm{F}_{\Gamma'} \times \mathbb{R}^N\right) = \gamma \llcorner \left(\mathbb{R}^N \times \mathrm{F}_{\Gamma'}\right) = (1,1)_\# f^+ \llcorner \mathrm{F}_{\Gamma'}$; thus, it is not restrictive to assume that $\mathrm{F}_{\Gamma'} = \emptyset$.

Thanks to the hypothesis $f^+(\mathrm{D}_{\Gamma'}) = 0$, we are then reduced to consider the case when f^+ is concentrated in $\mathrm{ET}_{\Gamma'}$; choose then any strictly convex function α such that the inf in (4.18) is finite and take a solution γ of (4.18). Since in particular $\gamma \in \Pi''(f^+, f^-)$, we can decompose γ with respect to $r_{\Gamma'} \circ \pi_1 : \mathbb{R}^N \times \mathbb{R}^N \to \mathcal{S}_c$ –where $\pi_1 : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^N$ is the projection on the first argument– that is Borel thanks to Lemma 4.4; the decomposition gives

$$\gamma = \gamma_C \otimes \nu;$$

recalling Lemma A.16, calling f_C^+ and f_C^- the projections of γ_C we find

$$f^+ = f_C^+ \otimes \nu \qquad\qquad f^- = f_C^- \otimes \nu. \tag{4.19}$$

We note now that thanks to the properties of disintegration the measure $\nu = r_{\Gamma'} \circ \pi_{1\#}\gamma = r_{\Gamma'\#} f^+$ does not depend on γ: roughly speaking, given a Borel set $A \subseteq \mathcal{S}_c$ the number $\nu(A)$ is the total mass which is moved along the maximal transport rays contained in A. In particular, $f^+ = f_C^+ \otimes \nu$ in (4.19) is also the disintegration of f^+ with respect to $r_{\Gamma'}$, then also f_C^+ do not depend on γ. We claim that also f_C^- do not depend on C, that cannot be trivially inferred with another disintegration in case f^- is not concentrated in $\mathrm{ET}_{\Gamma'}$; to show this fact, we write $f_C^- = f_C^1 + f_C^2$ with $f_C^1 = f_C^- \llcorner \mathrm{ET}_\Gamma$, so clearly $f^1 = f^- \llcorner \mathrm{ET}_\Gamma = f_C^1 \otimes \nu$; on the other hand, $r_{\Gamma'}$ is defined almost everywhere with respect to $f^- \llcorner \mathrm{ET}_\Gamma$, then we can operate a disintegration finding $f^- \llcorner \mathrm{ET}_\Gamma = \tilde{f}_C \otimes \tilde{\nu}$, where $\tilde{f}_C$ and $\tilde{\nu}$ clearly do not depend on γ. But $\tilde{\nu} = r_{\Gamma'\#} f^- \llcorner \mathrm{ET}_\Gamma = r_{\Gamma'} \circ \pi_{1\#}\left(\gamma \llcorner \mathbb{R}^N \times \mathrm{ET}_\Gamma\right) \leq \nu$ and then one can write

$$\tilde{\nu} = \beta\nu$$

where $\beta : \mathcal{S}_c \to [0,1]$ is a Borel function. Therefore, the uniqueness part of Theorem A.15 easily ensures that

$$f_C^1 = \beta(C)\tilde{f}_C$$

for $\tilde{\nu}$–a.e. C. But $f_C^- = f_C^1 + f_C^2$ is a probability measure on C, f_C^1 is its restriction to ET_Γ and f_C^2 is its restriction to $C \setminus \mathrm{ET}_\Gamma$, which is empty if $C \subseteq \mathrm{ET}_\Gamma$ and consists of the sole right endpoint y_C of C otherwise. Then, for any C such that $\beta(C) < 1$, it must be $f_C^2 = \big(1 - \beta(C)\big)\delta_{y_C}$: then our claim that also f_C^- do not depend on γ is proved.

Recalling Lemma 1.13 and the fact that $\inf(\mathfrak{C}_{\mathrm{conv}}) < +\infty$ by hypothesis and $\inf(\mathfrak{C}_{\mathrm{conv}}) > -\infty$ by Remark 4.14, γ is a solution of (4.18) if and only if for $\nu-$a.e. C the measure γ_C is a minimizer of

$$\int_{\mathbb{R}^N\times\mathbb{R}^N} \alpha(y-x)\, d\gamma_C(x,y)$$

in $\Pi(f_C^+, f_C^-)$; but recalling Theorem 4.9 and since f_C^+ and f_C^- are concentrated in the closed segment C, the strict convexity of α implies that $\gamma\in\Pi'''(f^+,f^-)$ if and only if for $\nu-$a.e. C γ_C is the unique non-decreasing transport plan between f_C^+ and f_C^-. Since ν, f_C^+ and f_C^- do not depend on γ nor on α, the thesis follows. $\square$

The importance of last lemma is that in some interesting situations the unique element of $\Pi'''(f^+,f^-)$ is a transport map: this is what will happen for the quasi-euclidean norms (see Chapter 8) and for the crystalline norms (see Chapter 9).

We can also note another useful consequence of the fact that the doubling points are f^+-negligible: assume that the norm $\|\cdot\|$ is strictly convex, so that arguing as in Theorem 4.11 one finds that $\Gamma=\partial^c u$ fulfills the no-crossing condition and then the function r_Γ and the set D_Γ are defined. Then, as mentioned after stating problem (4.8), the first variational problem we studied does not make any selection between the optimal transport plans; it is easily noticed that, also for this claim, the hypothesis $f^+(\mathrm{D}_\Gamma)=0$ is needed.

Lemma 4.16. *If c is slow $-(f^+,f^-)-$ increasing, $\|\cdot\|$ is strictly convex and $f^+(\mathrm{D}_\Gamma)=0$ then $\Pi'(f^+,f^-)=\Pi''(f^+,f^-)$; moreover one can choose $\Gamma'=\Gamma$ in the last step of Theorem 4.11.*

Proof. We take two elements γ_1 and γ_2 of $\Pi'(f^+,f^-)$, and we want to prove that $\mathfrak{C}_{\mathrm{eucl}}(\gamma_1)=\mathfrak{C}_{\mathrm{eucl}}(\gamma_2)$; the same argument of last Lemma allows also to assume that $\mathrm{F}_\Gamma=\emptyset$, since both γ_1 and γ_2 are concentrated in Γ and then leave fixed any element of F_Γ. Since by hypothesis $f^+(\mathrm{D}_\Gamma)=0$, the set ET_Γ is of full f^+-measure and then as in last Lemma we can operate the disintegration of γ_1 and γ_2 with respect to $r_\Gamma\circ\pi_1$, finding

$$\gamma_1=\gamma_{1,C}\otimes\nu_1 \qquad\qquad \gamma_2=\gamma_{2,C}\otimes\nu_2;$$

exactly as already noticed in the proof of last Lemma, it is proved that $\nu_1=\nu_2=\nu$ and for $\nu-$a.e. C both the measures $\gamma_{1,C}$ and $\gamma_{2,C}$ are transport plans between f_C^+ and f_C^-, two measures not depending on γ_1 and γ_2. Let us then consider a single segment C: since it is one-dimensional, there is a constant $K>0$ such that $|\cdot|=K\|\cdot\|$ on C –of course K depends on C. Thus $\mathfrak{C}_{\mathrm{eucl}}(\gamma_{1,C})=K\mathfrak{C}(\gamma_{1,C})$. But $\mathfrak{C}(\gamma_{1,C})=\mathfrak{C}(\gamma_{2,C})$ for

$\nu-$a.e. C according to Lemma 1.13, thus it follows $\mathfrak{C}_{\rm eucl}(\gamma_1) = \mathfrak{C}_{\rm eucl}(\gamma_2)$ as stated, and the first assert follows. The fact that $\Gamma' = \Gamma$ fulfills all the properties of the last claim in Theorem 4.11 is then straightforward. □

4.4 Two counterexamples

Lemma 4.13 and Lemma 4.15 ensure that $\Pi'''(f^+, f^-)$ contains only a point in two different situations; in particular, the second one is a quite general case. In this section we show two examples which will ensure that the hypotheses of the two lemmas are essential; in both the examples, we will consider the particular case when $\alpha(v)$ depends only on $|v|$, so we will directly consider, with a slight abuse of notation, α as a strictly convex function from $\mathbb{R}^+$ to $\mathbb{R}$.

Example 4.17. The first example we give is to show that it is not always true that the solution of (4.18) is unique; let us consider the celebrated example of the four Dirac masses, where f^+ and f^- are two Dirac masses situated in the vertexes of a square as in the left of Figure 4.2. Then it is immediate to realize that the transport plans are given by

$$\gamma_\lambda := \lambda\big(\delta_{x_+} \otimes \delta_{y_-} + \delta_{z_+} \otimes \delta_{w_-}\big) + (1-\lambda)\big(\delta_{x_+} \otimes \delta_{w_-} + \delta_{z_+} \otimes \delta_{y_-}\big)$$

with $0 \leq \gamma \leq 1$. If we take $\|\cdot\| = |\cdot|$ as the Euclidean distance, so that the first variational problem coincides with the usual mass transport problem, we see that each γ_λ has Euclidean cost

$$\mathfrak{C}(\gamma_\lambda) = \mathfrak{C}_{\rm eucl}(\gamma_\lambda) = \lambda \cdot 2 + (1-\lambda)\cdot 2 = 2,$$

then all the transport plans are optimal; on the other hand, it is also clear that $\mathfrak{C}_{\rm conv}(\gamma_\lambda) = 2\,\alpha(1)$ for any λ, so that all the transport plans γ_λ are also optimal for (4.18). Thus the solution of (4.18) is not unique.

The last example showed that it can happen that there are different solutions of (4.18); however, in many situations, as in last example, it is still true that, even if it contains more than a single measure, the set $\Pi'''(f^+, f^-)$ does not depend on α, provided it is strictly convex and $\inf(4.18) < +\infty$. The next example, a bit surprising, shows that it can also happen that different functions α give rise to different sets $\Pi'''(f^+, f^-)$.

Example 4.18. This example is a modified version of the preceding one which makes use of 8 Dirac masses (it is not difficult to check that $\Pi'''(f^+, f^-)$ is uniquely determined, i.e. not depending on α, if there are up to 6 Dirac masses). Let us then consider f^+ and f^- given by the sum of 4 Dirac masses situated in the points x_i and y_i respectively, and let

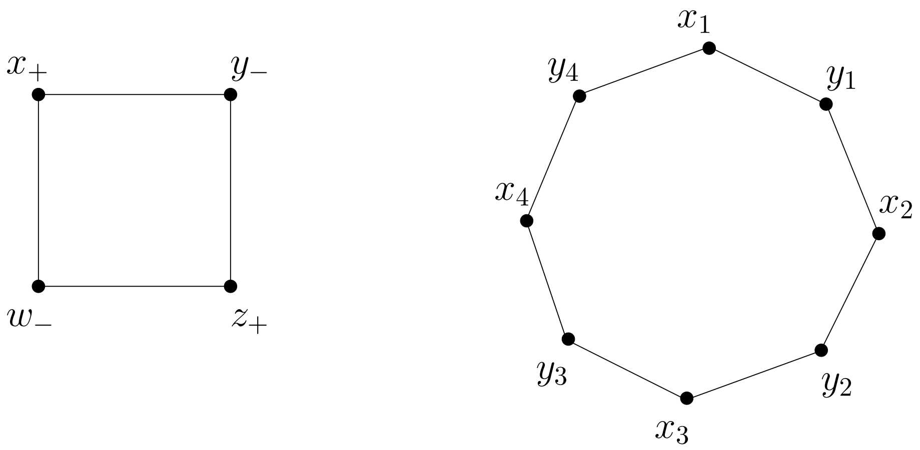


Figure 4.2: The counterexamples

the 8 points be, as in the right Figure 4.2, in the vertexes of an "almost regular" octagon: defining $d_{ij} = |y_j - x_i|$, we ask

$$\begin{array}{llll} d_{11} = 101 & d_{22} = 104 & d_{33} = 104 & d_{44} = 107 \\ d_{14} = 102 & d_{21} = 102 & d_{32} = 106 & d_{43} = 106, \end{array} \tag{4.20}$$

which is clearly possible. If the octagon is not too far from being regular, it is clear that the optimal transport plans with respect to $c(x,y) = \|y - x\| = |y - x|$ can move the mass only along the sides of the octagon. Then the "interesting" transport plans can be indexed by a single number $0 \leq \lambda \leq 1$: in fact, if λ is the mass that is transported from x_1 to y_1 by a transport plan working only along the sides, then the plan must be

$$\gamma_\lambda := \lambda\Big(\delta_{x_1} \otimes \delta_{y_1} + \delta_{x_2} \otimes \delta_{y_2} + \delta_{x_3} \otimes \delta_{y_3} + \delta_{x_4} \otimes \delta_{y_4}\Big) + \\ (1-\lambda)\Big(\delta_{x_1} \otimes \delta_{y_4} + \delta_{x_2} \otimes \delta_{y_1} + \delta_{x_3} \otimes \delta_{y_2} + \delta_{x_4} \otimes \delta_{y_3}\Big).$$

This transport plan has then a cost

$$\mathfrak{C}(\gamma_\lambda) = \mathfrak{C}_{\mathrm{eucl}}(\gamma_\lambda) = \lambda\Big(d_{11} + d_{22} + d_{33} + d_{44}\Big) + (1-\lambda)\Big(d_{14} + d_{21} + d_{32} + d_{43}\Big) \tag{4.21}$$

Recalling (4.20), this gives

$$\mathfrak{C}(\gamma_\lambda) = \lambda \cdot 416 + (1-\lambda) \cdot 416 = 416,$$

so that the optimal transport plans are all and only the γ_λ with $0 \leq \lambda \leq 1$. Concerning the convex cost $\mathfrak{C}_{\mathrm{conv}}$, similarly to (4.21) it is

$$\mathfrak{C}_{\mathrm{conv}}(\gamma_\lambda) = \lambda\Big(\alpha(d_{11}) + \alpha(d_{22}) + \alpha(d_{33}) + \alpha(d_{44})\Big) + (1-\lambda)\Big(\alpha(d_{14}) + \alpha(d_{21}) + \alpha(d_{32}) + \alpha(d_{43})\Big);$$

we fix $\alpha(s) = \beta(s-104)$ so that, up to a constant, the cost of each γ_λ is

$$\mathfrak{C}_{\text{conv}}(\gamma_\lambda) = \lambda\Big(\beta(-3) - 2\beta(-2) + 2\beta(0) - 2\beta(2) + \beta(3)\Big).$$

It is easy to notice that the term between the parentheses in last equation can either be positive, negative, or null for a strictly convex function β; it follows that $\Pi'''(f^+, f^-)$ can consist of the unique plan γ_0, or of the unique plan γ_1, or of all the plans γ_λ, and then it is not always made by a single element, and it depends on α.

Chapter 5

Optimal transport maps for the quadratic case and some applications

As discussed in the Introduction, in the original idea of Monge the ambient space $X = Y$ was a compact subspace of $\mathbb{R}^2$, and the cost function was the Euclidean distance $c(x, y) = |y - x|$; however, as we will discuss in Chapter 6, it has been really hard to give some answer to this problem. The main difficult of the Euclidean case is the fact that the cost $c(x, y) = |y - x|$ is convex but not strictly convex: on the other hand, a strongly different situation is when one considers a strictly convex cost. In fact, the first general existence result for the Monge problem has been proved for the most popular strictly convex cost, i.e. the quadratic cost $c(x, y) = |y - x|^2$: it was obtained independently in 1984 by Knott and Smith, [56], and in 1987 by Brenier, [16, 17]; in the context of Calculus of Variations and PDE one usually refers to Brenier, while in the context of Statistics and Economics to Knott and Smith. After their first results, many generalizations came out, see for example [47], [64], [10]). As far as we know, at this moment the most general result for the strictly convex case is the following one, proved by Gangbo and McCann in [47]:

Theorem 5.1 (Gangbo-McCann). *Assume $c(x, y) = h(y - x)$ and that*

- *h is strictly convex and with superlinear growth (i.e. $h(v)/|v| \to +\infty$ as $|v| \to +\infty$);*
- *given any bounded cone $C_{\theta,r} = \{z \in \mathbb{R}^N : |z| \cos\theta \leq \langle z, e_N\rangle \leq r\}$, for any $x \in \mathbb{R}^N$ with $|x|$ sufficiently large there is a cone obtained from $C_{\theta,r}$ by a translation and a rotation such that its vertex is in x and on which the function h attains it maximum at x*

and one of the following additional conditions holds:

a) $f^+ \ll \mathscr{L}$;

b) $f^+(S) = 0$ *whenever* $\mathscr{H}^{N-1}(S) < +\infty$ *and* $h \in \mathrm{C}^{1,1}_{\mathrm{loc}}$.

Then there is an unique optimal transport plan between f^+ *and* f^-*, and it is also induced by an optimal transport map.*

Note that this result can be applied to any $f^+ \ll \mathscr{L}$ in the particular case when $c(x, y) = h(|y - x|)$ with $h : [0, +\infty) \to \mathbb{R}^+$ strictly convex and such that $h(0) = 0$ and $h' \to +\infty$.

In this chapter, we will consider only the specific case of the quadratic cost $c(x, y) = |y - x|^2$: we will prove the Brenier's existence result, with an easier proof given several years later (see [45, 39]); then we will discuss some of the many applications that are related to this case.

5.1 The quadratic case

In this Section, we assume $X = Y \subseteq \mathbb{R}^N$ to be a compact subspace of $\mathbb{R}^N$ and the cost function to be the square of the norm, $c(x, y) = |y - x|^2$; we will also denote by $x \cdot y$ the scalar product between x and y. We try now to develop the dual formulation introduced in Section 3.2, beginning with an useful variation of the functions in (3.3): if we define

$$\varphi(x) = |x|^2 - u(x) \qquad\qquad \psi(y) = |y|^2 - v(y),$$

it is easily seen that $u(x) + v(y) \leq c(x, y)$ if and only if $\varphi(x) + \psi(y) \geq 2x \cdot y$; since u and v in $\mathcal{A}$ are uniquely determined by φ and ψ in

$$\mathcal{B} = \{(\varphi, \psi) \in L^1_{f^+}(X) \times L^1_{f^-}(Y) : \ \varphi(x) + \psi(y) \geq x \cdot y\},$$

and since $\int |x|^2 \, df^+$ and $\int |y|^2 \, df^-$ are finite and do not depend on u and v, solving the maximization problem (3.3) is equivalent to solving the minimization problem

$$\inf \left\{ \int_X \varphi(x) \, df^+(x) + \int_Y \psi(y) \, df^-(y) : \ (u, v) \in \mathcal{B} \right\}. \tag{5.1}$$

Let us see now that the last problem admits a solution: we recall (see [35]) that two functions φ and ψ are said dual convex if $\varphi^* = \psi$ and $\psi^* = \varphi$, where the standard polar transform of a function $\omega : \mathbb{R}^N \to \overline{R}$ is given by

$$\omega^*(y) := \sup_x x \cdot y - \omega(x).$$

Lemma 5.2. *There exists an optimal pair* $(\bar\varphi, \bar\psi)$ *for (5.1); moreover* $\bar\varphi$ *and* $\bar\psi$ *are dual convex.*

Proof. Let us take any pair $(\varphi, \psi) \in \mathcal{B}$. Since by definition $\varphi(x) \geq x \cdot y - \psi(y)$ for any y, we infer $\varphi \geq \psi^*$; moreover, inequalities $\psi \geq \psi^{**}$ and $\psi^*(x) + \psi^{**}(y) \geq x \cdot y$ are always true. Then $(\psi^*, \psi^{**}) \in \mathcal{B}$ and $\psi^* \leq \varphi$, $\psi^{**} \leq \psi$, thus

$$\int_X \psi^*(x)\, df^+(x) + \int_Y \psi^{**}(y)\, df^-(y) \leq \int_X \varphi(x)\, df^+(x) + \int_Y \psi(y)\, df^-(y).$$

So to minimize (5.1) we can restrict to dual convex pairs: given such a pair (φ, φ^*) and x, $x' \in X$, due to the compactness of X it is $\varphi(x) = \varphi^{**}(x) = x \cdot y - \varphi^*(y)$ for some y. Then

$$\varphi(x') - \varphi(x) \geq x' \cdot y - \varphi^*(y) - \varphi(x) = (x' - x) \cdot y \leq \operatorname{diam}(X)\, |x' - x|,$$

thus if (φ_n, φ_n^*) is an optimal sequence for (5.1) the functions φ_n are uniformly Lipschitz continuous. But since $\|f^+\| = \|f^-\|$ the pair (φ, φ^*) is equivalent to $\big(\varphi + K, (\varphi + K)^*\big) = (\varphi + K, \varphi^* - K)$, then it is not restrictive to assume that the functions φ_n have 0 mean: thus, by Ascoli–Arzelà Theorem, up to a subsequence we have $\varphi_n \to \bar\varphi$ uniformly. One immediately infers that $(\bar\varphi, \bar\varphi^*)$ is also a dual convex pair belonging to $\mathcal{B}$, and that it realizes the minimum in (5.1). $\square$

In view of last Lemma, in the sequel of this Section we will denote by $(\bar\varphi, \bar\psi)$ a dual convex pair optimal for (5.1).

Lemma 5.3. *If $\bar\varphi$ is differentiable at x, then $D\bar\varphi(x) \in X$ and $\bar\varphi(x) = x \cdot D\bar\varphi(x) - \bar\psi\big(D\bar\varphi(x)\big)$.*

Proof. Due to the compactness of X and the fact that $\bar\varphi = \bar\psi^*$, there is some $z \in X$ such that $\bar\varphi(x) = x \cdot z - \bar\psi(z)$; on the other hand, since $\bar\psi = \bar\varphi^*$ it is $\bar\psi(z) \geq y \cdot z - \bar\varphi(y)$ for any y, then

$$\bar\varphi(y) \geq \bar\varphi(x) + (y - x) \cdot z :$$

this gives, if $\bar\varphi$ is differentiable at x, that $D\bar\varphi(x) = z$, and then the thesis. $\square$

Remark 5.4. *In last Lemma, we proved that if $\bar\varphi(x) = x \cdot z - \bar\psi(z)$ and $\bar\varphi$ is differentiable at x, then $D\bar\varphi(x) = z$; conversely, it is also true that $\bar\psi(z) = z \cdot x - \bar\varphi(x)$, then if $\bar\psi$ is differentiable at z it is $D\bar\psi(z) = x$.*

We can then define the function which will be the optimal transport map between f^+ and f^-,

$$t(x) := D\bar\varphi(x). \tag{5.2}$$

It is clear that in order to be a transport map, t needs to be defined f^+–a.e., otherwise it makes no sense asking whether or not $t_\# f^+ = f^-$ because the first object is not defined:

keeping in mind the fact that a convex function (such as $\bar\varphi$) is differentiable up to a set S of Hausdorff dimension $N-1$, we know that t is defined f^+-a.e. if $f^+(S)=0$. Let us prove now that t is a transport map, then we will see that it is optimal.

Proposition 5.5. *If $D\bar\varphi$ is defined f^+-a.e. (thus, in particular, if $f^+(S)=0$ whenever $\mathscr{H}^{N-1}(S)<+\infty$), then t is a transport map between f^+ and f^-.*

Proof. By definition, $t_\# f^+=f^-$ means that for any $h\in \mathrm{C}_0^\infty(X)$ it is

$$\int_X h\big(t(x)\big)\,df^+(x)=\int_Y h(y)\,df^-(y), \tag{5.3}$$

so we need to prove the preceding equality.

Let us fix then a generic $h\in \mathrm{C}_0^\infty(X)$ and define

$$\psi_\varepsilon(y):=\bar\psi(y)+\varepsilon\,h(y), \qquad\qquad \varphi_\varepsilon:=\psi_\varepsilon^*; \tag{5.4}$$

then

$$\|\psi_\varepsilon-\bar\psi\|_{L^\infty}\le\varepsilon\|h\|_{L^\infty}, \qquad\qquad \|\varphi_\varepsilon-\bar\varphi\|_{L^\infty}\le\varepsilon\|h\|_{L^\infty}: \tag{5.5}$$

the first inequality is trivial; concerning the second one, given $x\in X$ there is $y\in Y$ with the property $\bar\varphi(x)=x\cdot y-\bar\psi(y)$, so

$$\varphi_\varepsilon(x)\ge x\cdot y-\psi_\varepsilon(y)=x\cdot y-\bar\psi(y)-\varepsilon\,h(y)\ge\bar\varphi(x)-\varepsilon\|h\|_{L^\infty}$$

and then $\varphi_\varepsilon-\bar\varphi\ge\varepsilon\|h\|_{L^\infty}$. To show $\bar\varphi-\varphi_\varepsilon\ge\varepsilon\|h\|_{L^\infty}$ it suffices to repeat the same argument interchanging the roles of $\bar\varphi$ and φ_ε.

Let us take now $x\in X$ such that $D\bar\varphi(x)=t(x)$ exists, and then by Lemma 5.3 it is $\bar\varphi(x)=x\cdot D\bar\varphi(x)-\bar\psi\big(D\bar\varphi(x)\big)$; let moreover $y_\varepsilon\in Y$ be a point such that

$$\varphi_\varepsilon(x)=x\cdot y_\varepsilon-\psi_\varepsilon(y_\varepsilon), \tag{5.6}$$

which is possible since $X=Y$ is compact and then $\varphi_\varepsilon=\psi_\varepsilon^*$ is continuous. Putting together (5.6), (5.5) and again the compactness of $X=Y$, one has that y_ε converges to some $\bar y$ with the property $\bar\varphi(x)=x\cdot\bar y-\bar\psi(\bar y)$: recalling Remark 5.4, we infer $y_\varepsilon\to D\bar\varphi(x)=t(x)$. Note now that (5.6), (5.4) and the fact that, since $\bar\varphi=\bar\psi^*$, $\bar\varphi(x)+\bar\psi(y)\ge x\cdot y$ give

$$\bar\varphi(x)-\varphi_\varepsilon(x)=\bar\varphi(x)+\psi_\varepsilon(y_\varepsilon)-x\cdot y_\varepsilon\ge\varepsilon\,h(y_\varepsilon).$$

On the other hand

$$\bar\varphi(x)-\varphi_\varepsilon(x)\le\big(x\cdot\bar y-\bar\psi(\bar y)\big)-\big(x\cdot\bar y-\psi_\varepsilon(\bar y)\big)=\varepsilon\,h(\bar y),$$

thus we infer

$$\eta_\varepsilon(x) := \frac{\bar\varphi(x) - \varphi_\varepsilon(x)}{\varepsilon} \leq \|h\|_{L^\infty} \qquad \eta_\varepsilon(x) \to h\big(D\bar\varphi(x)\big) = h\big(t(x)\big); \tag{5.7}$$

remember that this holds for any $x \in X$ such that $D\bar\varphi(x)$ exists, then by hypothesis for f^+−a.e. x.

We use now the fact that, since $\varphi_\varepsilon = \psi_\varepsilon^*$, $\varphi_\varepsilon(x) + \psi_\varepsilon(y) \geq x \cdot y$ and then $(\varphi_\varepsilon, \psi_\varepsilon) \in \mathcal{B}$; but $(\bar\varphi, \bar\psi)$ is optimal for (5.1), thus

$$\int_X \bar\varphi(x)\, df^+(x) + \int_Y \bar\psi(y)\, df^-(y) \leq \int_X \varphi_\varepsilon(x)\, df^+(x) + \int_Y \psi_\varepsilon(y)\, df^-(y) \Longrightarrow$$
$$\int_X \bar\varphi(x) - \varphi_\varepsilon(x)\, df^+(x) \leq \varepsilon \int_Y h(y)\, df^-(y) \Longrightarrow$$
$$\int_X \eta_\varepsilon(x)\, df^+(x) \leq \int_Y h(y)\, df^-(y).$$

Thanks to (5.7), which holds f^+−a.e., we can apply the Dominated Convergence Theorem and definition (5.2) to infer

$$\int_X h\big(t(x)\big)\, df^+(x) \leq \int_Y h(y)\, df^-(y).$$

Replacing h with $-h$ we understand that the preceding inequality is in fact an equality, so (5.3) holds and the assert is proved. □

We prove now Brenier's result, i.e. that, with the same hypothesis of the preceding Proposition, the map s is also an optimal transport map: note that this is a particular case of the general Theorem 5.1.

Theorem 5.6 (Brenier). *If $f^+(S) = 0$ whenever $\mathscr{H}^{N-1}(S) < +\infty$ (more generally, if $D\bar\varphi$ is defined f^+−a.e.), then there is an optimal transport map, which corresponds to the unique optimal transport plan. Moreover, if $f^-(S) = 0$ whenever $\mathscr{H}^{N-1}(S) < +\infty$ (more generally, if $D\bar\psi$ is defined f^-−a.e.), then the transport map is invertible and its inverse is a transport map and the unique optimal transport plan from f^- to f^+.*

Proof. Let us take the map t defined in (5.2), which we know to be a transport map thanks to Proposition 5.5; we write γ_t to denote the transport plan induced by the transport map t in the sense of (1.3), i.e. $\gamma_t = \big(\mathrm{Id}, t\big)_\# f^+$; recall also that $\mathfrak{C}(\gamma_t) = \mathfrak{C}(t)$ according to Lemma 1.7. The cost of any other transport plan γ is

$$\begin{aligned}\mathfrak{C}(\gamma) &= \iint_{X\times Y} |y-x|^2\, d\gamma(x,y) = \iint_{X\times Y} |x|^2 + |y|^2 - 2x\cdot y\, d\gamma(x,y) \\ &= \int_X |x|^2\, df^+(x) + \int_Y |y|^2\, df^-(y) - 2\iint_{X\times Y} x\cdot y\, d\gamma(x,y) \\ &= K - 2\iint_{X\times Y} x\cdot y\, d\gamma(x,y),\end{aligned}$$

where K is a real constant depending on f^+ and f^-, but not on γ. It follows

$$\mathfrak{C}(\gamma) \le \mathfrak{C}(t) \Longleftrightarrow \iint_{X\times Y} x\cdot y\, d\gamma(x,y) \ge \iint_{X\times Y} x\cdot y\, d\gamma_t(x,y) \tag{5.8}$$

Recall now that $x\cdot y \le \bar{\varphi}(x)+\bar{\psi}(y)$ and if $\bar{\varphi}$ is differentiable at x then the equality holds if and only if $y=t(x)$. But, thanks to the hypothesis, $\bar{\varphi}(x)$ is defined for f^+-a.e. x, thus also for $\gamma-$a.e. (x,y): then, recalling that $\gamma_t = \big(\mathrm{Id}, t\big)_{\#}f^+$ and that $t_\# f^+ = f^-$,

$$\begin{aligned}\iint_{X\times Y} x\cdot y\, d\gamma_t(x,y) &= \int_X x\cdot t(x)\, df^+(x) = \int_X \bar{\varphi}(x)+\bar{\psi}\big(t(x)\big)\, df^+(x)\\ &= \int_X \bar{\varphi}(x)\, df^+(x) + \int_Y \bar{\psi}(y)\, df^-(x)\\ &= \iint_{X\times Y} \bar{\varphi}(x)+\bar{\psi}(y)\, d\gamma(x,y) \ge \iint_{X\times Y} x\cdot y\, d\gamma(x,y).\end{aligned}$$

Comparing with (5.8), one infer that any transport plan γ has a cost greater than t unless $y=t(x)$ for a.e. $(x,y)\in \mathrm{spt}\,(\gamma)$, which is true if and only if γ is the transport plan induced by t thanks to Lemma 1.6. This implies that γ_t is the unique optimal transport plan and the first part of the thesis follows.

If $D\bar{\psi}$ is defined f^-–a.e., the same argument implies that there is a transport map s, defined –according to (5.2)– by

$$s(y) := D\bar{\psi}(y),$$

which is optimal and corresponds to the unique transport plan between f^- and f^+. Recall now the fact that, for f^+-a.e. x and f^--a.e. y, $x\cdot y = \bar{\varphi}(x)+\bar{\psi}(y)$ is true if and only if $y = D\bar{\varphi}(x) = t(x)$, and similarly if and only if $x = D\bar{\psi}(y) = s(y)$. It follows that t and s are invertible and each one is the inverse of the other. $\square$

Remark 5.7. *To show the second part of last Theorem, one could also just notice that t is f^+-essentially injective, since if $t^{-1}(y)$ contains more than one point then $\bar{\psi}$ is not differentiable at y. Then the support of the unique optimal transport plan between f^+ and f^-, which is the graph of $t : X \to Y$, is also the graph of some function $s : Y \to X$, which is the inverse of t. Then, according to the discussion after Definition 1.8, it follows that s is a transport map from f^- to f^+, and it also corresponds to the unique optimal transport plan between f^+ and f^-.*

Remark 5.8. *Thanks to the discussion in Example 1.4, we know that there may be no optimal transport maps if f^+ "sees" a set of finite $\mathscr{H}^{N-1}-$measure: thus the dimensional hypothesis in the preceding Theorem is sharp.*

5.2 Some applications

In this section we will briefly describe some applications of the existence and uniqueness Theorem 5.6; one can see also [16, 17, 37, 10, 18, 19, 20, 74] for related descriptions. In particular, see [10] for an impressive list of applications of the mass transport problem in the quadratic case to a large number of fields covering continuum mechanics, astrophysics, quantum chemistry, meteorology and many others. The main sources for this section are [16], [17], [37] and [74], but these arguments can be found also elsewhere; in particular, the polar factorization was the main goal of the Brenier existence Theorem 5.6 ([16, 17]).

Polar factorization and Helmoltz decomposition of vector fields

Suppose that $X \subseteq \mathbb{R}^N$ is bounded and let $r : X \to \mathbb{R}^N$ be a bounded measurable vector field such that $r_\# \mathscr{L} \ll \mathscr{L}$, i.e. $\mathscr{L}\left(r^{-1}(A)\right) = 0$ for any $A \subseteq \mathbb{R}^N$ such that $\mathscr{L}(A) = 0$: this is what Brenier called "nondegeneracy condition". The polar factorization ([16, 17]) consists in writing r as the composition of the gradient of a convex function with a measure-preserving map. To find such a factorization of r, let $\overline{X}$ be a bounded ball in $\mathbb{R}^N$ containing both X and $r(X)$, that is possible since r is bounded. Then, consider the mass transport problem between $f^+ = \mathscr{L} \llcorner X$ and $f^- = r_\# f^+$: since by hypothesis f^+, $f^- \ll \mathscr{L}$, we can apply Theorem 5.6 and find a dual convex pair $(\bar\varphi, \bar\psi)$ such that $D\bar\varphi$ (resp. $D\bar\psi$) is an optimal transport map from f^+ to f^- (resp. from f^- to f^+). In particular, $D\bar\psi_\# f^- = f^+$, which means

$$\left(D\bar\psi \circ r\right)_\# \mathscr{L} \llcorner X = \mathscr{L} \llcorner X.$$

We define then $\alpha = D\bar\psi \circ r$, which is by construction a measure-preserving map between X and itself, and note that $r = D\bar\varphi \circ \alpha$, thus we decomposed r as wanted. Moreover, this decomposition is unique up to an additive constant for $\bar\varphi$, since any such decomposition leads to an optimal transport map, which is unique according to Theorem 5.6. The first proof of this decomposition was by Brenier in [17], and it was later simplified by Gangbo [45]

This polar factorization is also correlated to the Helmoltz decomposition, which consists in writing a vector field as the sum of a gradient and a divergence-free field. Let X be a bounded and smooth subset of $\mathbb{R}^N$ and $a : X \to \mathbb{R}^N$ a smooth vector field: one wants to write $a = Db + c$ with div $c = 0$: it is known that such a decomposition exists, and it is unique under the further request that c is parallel to ∂X (the uniqueness immediately follows from the unicity of the solution of a Neumann problem). We show now that, at least formally, this unique decomposition can be obtained via a mass transport problem: for any $\varepsilon \geq 0$, define

$$r_\varepsilon(x) := x + \varepsilon\, a(x)$$

so that, using the polar factorization as before, we can find a sufficiently big open ball $\overline{X}$ and smooth functions $\bar{\varphi}_\varepsilon : \overline{X} \to \mathbb{R}$ and $\alpha_\varepsilon : X \to X$, smoothly depending on ε, such that

$$r_\varepsilon = D\bar{\varphi}_\varepsilon \circ \alpha_\varepsilon. \tag{5.9}$$

Since $r_0 = \mathrm{Id}$, it follows $\bar{\varphi}_0(x) = |x|^2/2$ (up to an additive constant) and $\alpha_0 = \mathrm{Id}$; moreover, the regularity implies the existence of a scalar function $b : \overline{X} \to \mathbb{R}$ and a vectorial function $c : X \to X$ such that

$$\bar{\varphi}_\varepsilon(x) = \bar{\varphi}_0(x) + \varepsilon\, b(x) + o(\varepsilon) \qquad\qquad \alpha_\varepsilon(x) = x + \varepsilon\, c(x) + o(\varepsilon) : \tag{5.10}$$

note also that c is parallel to the boundary of X since α_ε maps X onto itself. Now, we have that (5.9) reads as

$$x + \varepsilon\, a(x) = D\big(\bar{\varphi}_0 + \varepsilon\, b + o(\varepsilon)\big) \circ \big(\mathrm{Id} + \varepsilon\, c + o(\varepsilon)\big) = \mathrm{Id} + \varepsilon\,(Db + c) + o(\varepsilon),$$

and it follows $a = Db + c$. To conclude, we only need to check $\operatorname{div} c = 0$. In fact, since α_ε is measure-preserving, it follows $\mathrm{Det}(D\alpha_\varepsilon) = 1$; recalling (5.10), it is

$$D\alpha_\varepsilon = \begin{pmatrix} 1 + \varepsilon \dfrac{\partial c_1}{\partial x_1} & \cdots & \varepsilon \dfrac{\partial c_1}{\partial x_N} \\ \vdots & \ddots & \vdots \\ \varepsilon \dfrac{\partial c_N}{\partial x_1} & \cdots & 1 + \varepsilon \dfrac{\partial c_N}{\partial x_N} \end{pmatrix} + o(\varepsilon) :$$

it follows $1 = \mathrm{Det}(D\alpha_\varepsilon) = 1 + \varepsilon\big(\partial c_1/\partial x_1 + \cdots + \partial c_N/\partial x_N\big) + o(\varepsilon)$, which implies $\operatorname{div} c = 0$. Then we found the Helmoltz decomposition which was, in some sense, the linearization of the polar factorization.

We just mention that the polar factorization $r = D\bar{\varphi}\circ\alpha$ is also related to the monotone rearrangement of vector fields: we refer to [17] for the study of this subject and the proof that $D\bar{\varphi}$ is the unique such rearrangement of r.

Nonlinear interpolation

Another application of the quadratic case in the mass transportation problem is the following, studied in [37]: let us consider a gas in $\mathbb{R}^N$, whose density is given by a probability measure $\rho \ll \mathscr{L}$ (then ρ is in fact a L^1 function: in this section it will be often useful to consider ρ both as a measure and a function); its energy will be considered as

$$E(\rho) = E_1(\rho) + E_2(\rho) = \int_{\mathbb{R}^N} \rho^\delta(x)\,dx + \iint_{\mathbb{R}^N\times\mathbb{R}^N} V(y-x)\,d\rho(x)\,d\rho(y) :$$

the first part is the internal energy, where $\delta > 1$ is a given constant, while the second part is the interaction energy, and V is a given positive and strictly convex function with $V(x) \to +\infty$ when $|x| \to +\infty$. We want to prove that there is a unique minimizer of the energy, up to a traslation. First we prove the following

Lemma 5.9. *There are minimizers for E in $\mathcal{P}(\mathbb{R}^N) \cap L^1(\mathbb{R}^N)$.*

Proof. Since $E_2 \geq 0$, E is coercive in L^δ, then if ρ_n is a minimizing sequence the norm $\|\rho_n\|_{L^\delta}$ is bounded. The hypotheses on V ensure that, up to a translation of the densities ρ_n –which does not change their energy– for any ε we can find R_ε such that

$$\int_{B(0,R_\varepsilon)} \rho_n \geq 1-\varepsilon \qquad \text{for any integer } n. \tag{5.11}$$

Up to a subsequence, then, we can assume $\rho_n \xrightarrow{L^\delta} \rho$: this usually implies $\rho_n \xrightarrow{L^1_{\mathrm{loc}}} \rho$ but in this case (5.11) ensures that it is also $\rho_n \xrightarrow{L^1} \rho$ and then $\int \rho = 1$.

The thesis will then be reached if we show that $E(\rho) \leq \liminf E(\rho_n)$; since the lower semicontinuity of E_1 is obvious, let us consider only that of E_2. To this aim, define

$$\rho_{n,\varepsilon} = \rho_n \llcorner B(0,R_\varepsilon), \qquad \varphi_{n,\varepsilon}(y) = \int_{\mathbb{R}^N} V(y-x)\,d\rho_{n,\varepsilon}(x),$$

$$\rho_\varepsilon = \rho \llcorner B(0,R_\varepsilon), \qquad \varphi_\varepsilon(y) = \int_{\mathbb{R}^N} V(y-x)\,d\rho_\varepsilon(x) :$$

since V is bounded on the compact sets, it is clear that $\varphi_{n,\varepsilon} \to \varphi_\varepsilon$ pointwise and that $\|\varphi_{n,\varepsilon}\|_{L^\infty}$ is bounded by a constant depending on ε but not on n: thus $\varphi_{n,\varepsilon} \xrightarrow{L^p} \varphi_\varepsilon$ for any $p < \infty$ and in particular $\varphi_{n,\varepsilon} \xrightarrow{L^{\delta'}} \varphi_\varepsilon$. Then

$$\iint V(y-x)\,d\rho_{n,\varepsilon}(x)\,d\rho_{n,\varepsilon}(y) = \int \varphi_{n,\varepsilon}(y)\,d\rho_{n,\varepsilon}(y) \xrightarrow[n\to+\infty]{} \int \varphi_\varepsilon(y)\,d\rho_\varepsilon(y)$$

for strong-weak convergence. For any ε, it follows

$$E_2(\rho_\varepsilon) = \iint V(y-x)\,d\rho_\varepsilon(x)\,d\rho_\varepsilon(y) = \int \varphi_\varepsilon(y)\,d\rho_\varepsilon(y) = \lim_{n\to+\infty} E_2(\rho_{n,\varepsilon}) \leq \liminf_{n\to+\infty} E_2(\rho_n);$$

when ε goes to 0, last inequality gives

$$E_2(\rho) \leq \liminf_{n\to+\infty} E_2(\rho_n),$$

thus the searched lower semicontinuity and then the thesis follow. $\square$

We show now that the minimizer of the energy is unique up to translation; a standard approach could be to check if $\sigma \to E\big(\sigma\rho_0 + (1-\sigma)\rho_1\big)$ is a strictly convex function of σ when ρ_0 and ρ_1 are two measures, but unfortunately this is in general not true.

Another neat approach giving the uniqueness using transportation tools is the following: take two measures ρ_0 and ρ_1 in L^1, take the optimal transport map t from ρ_0 to ρ_1, which exists and is unique thanks to Theorem 5.6, and define $\rho_\sigma = \tau_{\sigma\#}\rho_0$, where $\tau_\sigma = \sigma t + (1-\sigma)\mathrm{Id}$; note that for $\sigma = 0,\,1$ this is consistent with the given data ρ_0 and ρ_1, and that apriori ρ_t is a only probability measure. The uniqueness of the minimizer up to a translation comes from the following

Lemma 5.10. *The function $\sigma \to E(\rho_\sigma)$ is convex, and strictly convex unless ρ_0 and ρ_1 are traslate of each other (the fact that $\rho_\sigma \in L^1 \cap \mathcal{P}$ is part of the proof).*

Proof. Since both ρ_0 and ρ_1 are absolutely continuous measures, thanks to Theorem 5.6 we infer that the map t is invertible and that $\bar\varphi$ and $\bar\psi$ are strictly convex. Then for any $x \neq y$ it is

$$(y-x)\cdot\big(D\bar\varphi(y) - D\bar\varphi(x)\big) > 0 :$$

recalling that according to (5.2) $t = D\bar\varphi$, the preceding inequality implies in particular that for any σ

$$(1-\sigma)(y-x) + \sigma\big(t(y) - t(x)\big) \neq 0$$

which means $\tau_\sigma(x) \neq \tau_\sigma(y)$, and then τ_σ is one-to-one; moreover $D\tau_\sigma = (1-\sigma)\mathrm{Id} + \sigma D^2\bar\varphi$ is strictly positive defined because $\bar\varphi$ is strictly convex. Taken then $A \subseteq \mathbb{R}^N$, we can evaluate with a change of coordinates

$$\rho_\sigma(A) = \tau_{\sigma\#}\rho_0(A) = \rho_0\big(\tau_\sigma^{-1}(A)\big) = \int_{\tau_\sigma^{-1}(A)} \rho_0(x)\,dx = \int_A \rho_0\big(\tau_\sigma^{-1}(x)\big)\,\big|\mathrm{Det}\big(D\tau_\sigma^{-1}(x)\big)\big|\,dx :$$

this is possible since, by the definition of τ_σ and again recalling that $\bar\varphi$ is strictly convex, τ_σ^{-1} is seen to be Lipschitz with constant less than $1/(1-\sigma)$. It follows that $\rho_\sigma \ll \mathscr{L}$ so we proved that $\rho_\sigma \in L^1$, while $\rho_\sigma \in \mathcal{P}(\mathbb{R}^N)$ is clear since ρ_σ is obtained from ρ_0 via a push-forward and in general the push-forward preserves the total mass. We will consider now separately the convexity of $\sigma \to E_1(\rho_\sigma)$ and of $\sigma \to E_2(\rho_\sigma)$.

Concerning E_1, we first evaluate, for any $\varphi \in \mathrm{C}_0^\infty(\mathbb{R}^N)$,

$$\int \varphi\big(\tau_\sigma(x)\big)\rho_0(x)\,dx = \int \varphi(x)\rho_\sigma(x)\,dx = \int \varphi\big(\tau_\sigma(x)\big)\rho_\sigma\big(\tau_\sigma(x)\big)\,\big|\mathrm{Det}\big(D\tau_\sigma(x)\big)\big|\,dx$$

from which we infer

$$\rho_\sigma\big(\tau_\sigma(x)\big) = \frac{\rho_0(x)}{\big|\mathrm{Det}\big(D\tau_\sigma(x)\big)\big|}.$$

Then

$$E_1(\rho_\sigma) = \int \rho_\sigma^\delta(x)\,dx = \int \rho_\sigma^\delta\big(\tau_\sigma(x)\big)\,\big|\mathrm{Det}\big(D\tau_\sigma(x)\big)\big|\,dx = \int \rho_0^\delta(x)\,\big|\mathrm{Det}\big(D\tau_\sigma(x)\big)\big|^{1-\delta}\,dx$$

and we are reduced to prove the convexity of $\beta(\sigma) = \alpha(\sigma)^{N(1-\delta)}$, setting

$$\alpha(\sigma) = \big|\mathrm{Det}\big(D\tau_\sigma(x)\big)\big|^{1/N}.$$

The reason why we gave this definition of α is the fact that $A \to \big(\mathrm{Det}(A)\big)^{1/N}$ is a concave function in the set of the positively defined symmetric $N \times N$ real matrices: this is a well known fact, that can be obtained via elementary linear algebra tools. Note now that $\sigma \to D\tau_\sigma = \sigma D^2\bar\varphi + (1-\sigma)\mathrm{Id}$ is a linear matrix-valued function of σ, and it is a positively defined symmetric $N \times N$ matrix for $\sigma \in [0,1]$. To show the convexity of β, then, it suffices to calculate

$$\beta'(\sigma) = N(1-\delta)\alpha(\sigma)^{N(1-\delta)-1}\alpha'(\sigma)$$

and consequently

$$\beta''(\sigma) = \underbrace{N(1-\delta)}_{\leq 0}\,\underbrace{\big(N(1-\delta)-1\big)}_{\leq 0}\,\underbrace{\alpha(\sigma)^{N(1-\delta)-2}}_{\geq 0}\,\underbrace{\alpha'(\sigma)^2}_{\geq 0} + \underbrace{N(1-\delta)}_{\leq 0}\,\underbrace{\alpha(\sigma)^{N(1-\delta)-1}}_{\geq 0}\,\underbrace{\alpha''(\sigma)}_{\leq 0} \geq 0,$$

recalling the concavity of $\sigma \to \alpha(\sigma)$: thus the convexity of β, so that of E_1, is proved.

Concerning E_2, finally, one has

$$\begin{aligned}
E_2(\rho_\sigma) &= \iint V(y-x)\,d\rho_\sigma(x)\,d\rho_\sigma(y)\\
&= \iint V\Big(\big(\sigma t(y) + (1-\sigma)y\big) - \big(\sigma t(x) + (1-\sigma)x\big)\Big)\,d\rho_0(x)\,d\rho_0(y)\\
&= \iint V\Big(\sigma\big(t(y)-t(x)\big) + \big(1-\sigma\big)(y-x)\Big)\,d\rho_0(x)\,d\rho_0(y)\\
&\leq \sigma\left(\iint V(\big(t(y)-t(x)\big)\,d\rho_0(x)\,d\rho_0(y)\right) + \big(1-\sigma\big)\left(\iint (y-x)\,d\rho_0(x)\,d\rho_0(y)\right)\\
&= \sigma E_2(\rho_1) + (1-\sigma)E_2(\rho_0),
\end{aligned}$$

and this gives the convexity. Moreover, since V is strictly convex, the function $\sigma \to E_2(\sigma)$ is strictly convex unless $t(y)-t(x) = y-x$ for any x, y, and then unless $t(x)-x$ is constant, that happens if and only if ρ_1 is a translation of ρ_0. This proves that if there are two minimizers, they must be one the translation of the other; since, on the other hand, the energy is not affected by a translation, the thesis follows. $\square$

Chapter 6

Existence results for the Euclidean case

In this chapter we consider the Euclidean case: this means that $X = Y = \mathbb{R}^N$ and the cost function is $c(x, y) = |y - x|$. This is probably the most challenging case in mass transportation, and it also follows the original idea of Monge; despite this importance, and despite the very natural formulation of the problem, there was not a formally correct existence Theorem until the proof by Evans and Gangbo in 1999 ([39]).

In this chapter, we will first discuss the idea of solution elaborated by Sudakov in late '70s, which is really interesting and which is somehow the basis of most existence proofs given in last years; then we will see why his argument does not effectively work and we will give two counterexamples showing that in fact his claim was too strong. Then we will present the main existence results proved for the Euclidean distance: we give a brief description of the work of Evans and Gangbo [39], which is the first correct one, an of the works of Caffarelli, Feldman, and McCann and of Trudinger and Wang [24, 73]; finally, we will discuss and proof a result of Ambrosio [1], since it is the basis of the deep generalization we will consider in the next chapter.

6.1 Sudakov's idea

In his major work [72], Sudakov dedicated the last chapter to the Monge–Kantorovich problem, developing a strategy to build an optimal transport map. The idea was, roughly speaking, to consider separately many one-dimensional problems, solve each of them, and then glueing the different transport maps.

The first point was then to divide the space $\mathbb{R}^N$ in many segments, as with the decomposition in maximal transport rays introduced in Section 4.1: this was done by Sudakov considering the segments on which a Kantorovich potential u (which is a $1-$Lipschitz function as noted in Corollary 3.27) increases with slope 1; these segments are, as we already discussed, disjoint except possibly in their endpoints. The second step was to solve a one-dimensional problem on each segment; to do that, expressing with today's notation his idea, one has to disintegrate an optimal transport γ with respect to the inclusion in maximal transport rays, writing then $\gamma = \gamma_C \otimes \nu$ where ν is a measure in $\mathcal{S}_c$; finally the problem to solve is to find an optimal transport map from f^+_C to f^-_C in each segment C. The third and last step was to glue the different transport maps found on each C to obtain a unique transport map from f^+ to f^-, and this should be possible since the different transport rays are disjoint; Sudakov then claimed that the map built in this way was an optimal transport map.

We can now see that, in order to make this strategy formal, one needs to solve some problems: first of all, it is not difficult to find an optimal Kantorovich potential u and then to individuate some open segments in $\mathbb{R}^N$ (the maximal transport rays) along which trying to move the mass; the problem is, some of the mass to be moved may not be contained in the union of these segments, since there could be mass in the endpoints: for instance, in the well-known example of the 4 Dirac masses (see Example 4.17) all the mass f^+ is contained in the vertices of the square, while the maximal transport rays are the (open) sides of the square; however, it seems reasonable that, if f^+ is sufficiently regular, then almost all the mass is in the interior of the maximal transport rays. Another problem is that, thanks to the $1-$dimensional theory we developed in Section 4.2, and which was also well-known at Sudakov's time, a one-dimensional Monge problem can always be solved, provided f^+ is non-atomic: hence, to solve the one-dimensional problems given by disintegration, one needs to show that f^+_C is non-atomic, at least for a.e. segment C (with respect to $r_{\mathrm{r}\#}f^+$). And this also seems to be reasonable if f^+ is sufficiently regular.

The main ingredient that Sudakov used to overcome these difficulties is the following result, which was Proposition 78 in [72]:

Claim 6.1. *Take a Borel set $T \subseteq \mathbb{R}^N$ and a Borel map $r : T \to \mathcal{S}_c$ such that $x \in r(x)$ and $r(x) \cap r(y) = \emptyset$ unless $r(x) = r(y)$; then, given a measure $f^+ \ll \mathscr{L}$ and operating a disintegration $f^+ = f^+_C \otimes r_\# \mu$ with respect to r, f^+_C is non-atomic for $r_\# f^+-$a.e. C.*

The assertion of this Claim follows the intuition: in fact, one can imagine that almost each measure f^+_C is negligible with respect to $\mathscr{H}^1 \llcorner C$, that is in fact what happens with suitable regularity hypothesis (see Lemma 6.14). As a consequence of the preceding Claim, Sudakov stated the following existence result:

Claim 6.2. *Let $X = Y = \mathbb{R}^N$ and $c(x, y) = |y - x|$; assume, moreover, that $f^+ \ll \mathscr{L}$. Then there exists an optimal transport map $t : \mathbb{R}^N \to \mathbb{R}^N$ from f^+ to f^-.*

In his work, Sudakov suggested that the preceding result should be true not only for the Euclidean norm, but also for any norm in $\mathbb{R}^N$, and in the case of a non strictly convex norm one should have decomposed not only in segments but also in affine sets of dimension greater than 1.

Today it is known that the two preceding claims are not correct without further regularity assumptions, and in the two following sections we will present two counterexamples showing this fact. However, the proof which we will give in Section 6.6 follows the approach thought by Sudakov, filling the gap in his proof thanks to suitable stronger assumptions.

6.2 A Nikodym set in $\mathbb{R}^3$

In this Section we present an example, first announced by Alberti, Kirchheim and Preiss and then appeared in an improved form in [4], which shows that the assert of Claim 6.1 is false without further assumptions on the map r and on the measure f^+; this example also improves an older one by Larman [58]. What we have in mind is the following

Theorem 6.3. *There exist a Borel subset M of the cube $[-1, 1]^3$ of full Lebesgue measure and a Borel map $g : M \to [-2, 2]^2 \times [-2, 2]^2$ such that, denoting by l_x the closed segment connecting $\big(g_1(x), -2\big)$ to $\big(g_2(x), 2\big)$, for any x it is $l_x \cap M = \{x\}$ and for any $x \neq y$ it is $l_x \cap l_y = \emptyset$.*

Let us notice that this claim may sound amazing: we have a family of pairwise disjoint segments, $\{l_x\}$, and we take an interior point for each of them; so one can imagine that the set of these points has Hausdorff dimension at most $N - 1$, and this is indeed what happens if the direction of the segments is sufficiently regular. In this example, on the countrary, the set of these point has full measure in the cube, so the set of all the remaining parts, which is the union of all the whole segments without a point, is Lebesgue negligible in the cube. Of course this is a counterexample to Claim 6.1 if $T = M$, $f^+ = \mathscr{L} \llcorner M$ and r is the function $x \to l_x$.

One of the main ingredients in the construction of this example is the following 2−dimensional remarkable result of Davies [29]:

Theorem 6.4 (Davies). *Take a parallelogram $T = \overline{PQRS}$ in $\mathbb{R}^2$ and a closed set $K \subseteq T$. Then for any $\varepsilon > 0$ there is a finite family $\{T_i = \overline{P_i Q_i R_i S_i},\ i = 1, \dots, k\}$ of parallelograms* *whose bases $\overline{P_iQ_i}$ and $\overline{R_iS_i}$ are contained in the bases $\overline{PQ}$ and $\overline{RS}$ of T such that $K \subseteq \cup_i T_i$ and $\mathscr{H}^2\big(\cup_i T_i \setminus K\big) < \varepsilon$.*

Now we will prove Theorem 6.3; since the proof is quite technical, we will only describe the different steps of the construction: for the formally correct proof we refer to [4]. First of all we give the

Definition 6.5. *A* beam *B is the image of the unit cube $[0,1]^3$ of $\mathbb{R}^3$ under an affine map, so we will identify it with the pair (P_B, a_B), where $a_B \in \mathbb{R}^3$, P_B is an invertible 3×3 matrix and $B = \{P_B(x) + a_B : x \in [0,1]^3\}$. We will call the segment $P_B(e_3)$ the* axis *of the beam, and $P_B\big([0,1]^2 \times \{i\}\big)$, with $i = 0, 1$, will be called respectively the* bottom *and the* top *of the beam.*

We will give also the following

Definition 6.6. *Given two beams B_1 and B_2, we will write $B_1 \prec B_2$ if bottom$(B_1) \subseteq$ bottom(B_2) and top$(B_1) \subseteq$ top(B_2). Moreover, given two families $\mathcal{F}_1$ and $\mathcal{F}_2$ of beams, we will write $\mathcal{F}_1 \prec \mathcal{F}_2$, and we will say that $\mathcal{F}_1$* refines *$\mathcal{F}_2$, if for each $B_1 \in \mathcal{F}_1$ there is $B_2 \in F_2$ such that $B_1 \prec B_2$.*

In the following, whenever $\mathcal{F}$ is a family of beams, we will also write $\cup\mathcal{F}$ to denote $\cup\{B : B \in \mathcal{F}\}$. In order to proceed with the construction, we will mainly make use of the following geometrical result which allows, taking a beam and a finite family of beams, to separate them taking away only an arbitrary small volume:

Lemma 6.7. *Given a beam B_0 and a finite family $\mathcal{F}$ of beams, and given $\varepsilon > 0$, there are finite families $\mathcal{F}_0' \prec \{B_0\}$ and $\mathcal{F}' \prec \mathcal{F}$ such that*

(i) $\bigcup \mathcal{F}_0' \cap \bigcup \mathcal{F}' = \emptyset$,

(ii) $\big|(B_0 \cup \bigcup \mathcal{F}) \setminus (\bigcup \mathcal{F}_0' \cup \bigcup \mathcal{F}')\big| < \varepsilon$.

The proof of this Lemma is a careful construction which makes use of the planar result of Davies; once one has this result, which involves a beam and a finite family, it is easy to generalize to the case of a finite number of finite families, with the

Corollary 6.8. *Given the finite families of beams $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_k$, and given $\varepsilon > 0$, there are finite families $\mathcal{F}_i' \prec \mathcal{F}_i$ such that*

(i) $\bigcup \mathcal{F}_i' \cap \bigcup \mathcal{F}_j' = \emptyset$ *for any $i \neq j$;*

(ii) $\Big|\Big(\bigcup_{i=1}^k \bigcup \mathcal{F}_i\Big) \setminus \Big(\bigcup_{i=1}^k \bigcup \mathcal{F}_i'\Big)\Big| < \varepsilon$.

Proof. We argue by induction: suppose the thesis is not true, then take a counterexample with k minimal and, between all the counterexamples with k minimal, with $\#(\mathcal{F}_1)$ minimal. We then choose $B_0 \in \mathcal{F}_1$ and we define $\mathcal{F} = \mathcal{F}_2 \cup \mathcal{F}_3 \cup \cdots \cup \mathcal{F}_k$; applying Lemma 6.7 to B_0 and $\mathcal{F}$, we find $\mathcal{F}'_0 \prec \{B_0\}$ and $\mathcal{F}' \prec \mathcal{F}$ with $\bigcup \mathcal{F}'_0 \cap \bigcup \mathcal{F}' = \emptyset$ and $\left|\left(B_0 \cup \bigcup \mathcal{F}\right) \setminus \left(\bigcup \mathcal{F}'_0 \cup \bigcup \mathcal{F}'\right)\right| < \varepsilon/2$. We can clearly also write $\mathcal{F}' = \cup \widehat{\mathcal{F}}_i$ with $\widehat{\mathcal{F}}_i \prec \mathcal{F}_i$; now we apply the induction hypothesis to $\mathcal{F}_1 \setminus \{B_0\}$, $\widehat{\mathcal{F}}_2, \ldots, \widehat{\mathcal{F}}_k$, getting families $\widetilde{\mathcal{F}}_1, \widetilde{\mathcal{F}}_2, \ldots, \widetilde{\mathcal{F}}_k$ with the properties

(i) $\bigcup \widetilde{\mathcal{F}}_i \cap \bigcup \widetilde{\mathcal{F}}_j = \emptyset$ for any $i \neq j$;

(ii) $\left|\left(\bigcup \mathcal{F}_1 \setminus \{B_0\} \cup \bigcup_{i=2}^k \bigcup \widehat{\mathcal{F}}_i\right) \setminus \left(\bigcup_{i=1}^k \bigcup \widetilde{\mathcal{F}}_i\right)\right| < \varepsilon/2.$

Finally, the proof is achieved setting $\mathcal{F}'_1 = \widetilde{\mathcal{F}}_1 \cup \mathcal{F}'_0$ and $\mathcal{F}'_i = \widetilde{\mathcal{F}}_i$ for any $i = 2, \ldots, k$, recalling the general set inclusion

$$(A \cup B) \setminus (C \cup D) \subseteq \big((A \cup E) \setminus C\big) \cup \big(B \setminus (D \cup E)\big).$$

□

Now we have all the tools we need to proof Theorem 6.3; what one makes then is the following iterative construction: defining Q and B the cubes $[-1,1]^3$ and $[-2,2]^3$ respectively, it is possible to define families $\mathcal{F}_k \prec \{B\}$, $\mathcal{N}_k \prec \{B\}$, $\mathcal{D}_k \prec \{B\}$ and $\mathcal{C}_k$ fulfilling the properties below (we call "width" of a beam the diameter of its top and bottom); the starting point is $\mathcal{N}_0 = \emptyset$, $\mathcal{D}_0 = \{B\}$ and $\mathfrak{C}_0 = Q$ ($\mathcal{F}_0$ is already part of the construction).

$\mathcal{C}_k$ is a disjoint family of beams; (6.1)

the widths of the elements of $\mathcal{D}_k$ become arbitrarily small when $k \to +\infty$; (6.2)

heights and widths of elements of $\mathcal{C}_k$ become arbitrarily small when $k \to +\infty$; (6.3)

$\mathcal{D}_{k+1} \prec \mathcal{F}_k \cup \mathcal{N}_k$; (6.4)

$C,\, D \in \mathcal{D}_k,\ C \cap D \neq \emptyset$ imply that the angle between the axes is at most $1/k$; (6.5)

$\forall\, C \in \mathcal{C}_k$ there exists $E(C) \in \mathcal{D}_k$ s.t. $C \in \operatorname{int}(Q \cap E(C))$; (6.6)

$\mathcal{F}_k = \displaystyle\bigcup_{C \in \mathcal{C}_k} \mathcal{F}_k(C)$ and $\mathcal{F}_k(C) \prec E(C)$; (6.7)

$\forall\, C \in \mathcal{C}_{k+1}$ s.t. $C \cap \mathcal{C}_k \neq \emptyset,$ there exists $D(C) \in \mathcal{C}_k:\ C \subseteq D(C)$; (6.8)

$\mathscr{L}\left(Q \setminus \bigcup \mathcal{C}_k\right) < 1/2^{k-1}; \qquad \mathscr{L}\left(\bigcup \mathcal{N}_k\right) < 3/2^{k+1}; \qquad \mathscr{L}(T_k) < 1/2^{k+1},$ (6.9)

where $T_k := \bigcup_{C \in \mathcal{C}_k}\left(\bigcup \mathcal{F}_k(C) \setminus C\right)$. The construction of the families with all the above properties is made by a careful induction argument, which makes use also of Theorem 6.4

and Corollary 6.8; once we have these families, we finally define

$$M := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \Big(\bigcup \mathcal{C}_k \setminus \Big(T_k \cup \bigcup \mathcal{N}_k \Big) \Big).$$

The fact that M is a Borel subset of Q of full Lebesgue measure is clear by construction and thanks to (6.9); let us now take $x \in M$: then, recalling (6.1), there is a unique sequence C_k for $k \geq k(x)$ such that $x \in C_k \in \mathcal{C}_k$; thanks to (6.8), $D(C_{k+1}) = C_k$, hence C_k is a decreasing sequence of beams; since $x \in C_k$ for any $k \geq k(x)$ and recalling (6.3), it follows $\cap_{k \geq k(x)} C_k = \{x\}$. Moreover we have, according to (6.6), the sequence of beams $E_k = E(C_k) \in \mathcal{D}_k$ such that $C_k \subseteq E_k$; we claim that also

$$E_{k+1} \subseteq E_k \quad \text{for } k \geq k(x) : \tag{6.10}$$

in fact, thanks to (6.4) we have $E_{k+1} \prec F$ for some $F \in \mathcal{F}_k \cup \mathcal{N}_k$; but since $x \in C_{k+1} \subseteq E_{k+1}$ and $x \notin \bigcup \mathcal{N}_k$, $F \in \mathcal{F}_k$; thus, by (6.7), $F \in \mathcal{F}_k(C')$ for some $C' \in \mathcal{C}_k$. It must be $C' = C_k$, because otherwise $C' \cap C_k = \emptyset$ so $x \in \mathcal{F}_k(C') \setminus C' \subseteq T_k$, against the definition. So $E_{k+1} \subseteq F \subseteq \bigcup \mathcal{F}_k(C_k) \subseteq E_k$ again thanks to (6.7), so (6.10) is established. Recalling that $E_k \prec B$ for any $k \geq k(x)$, we can define $l_x = \cap E_k$, which is a segment thanks to (6.2); moreover $l_x \subseteq E_{k+1} \subseteq \bigcup \mathcal{F}_k(C_k) \subseteq C_k \cup T_k$, thus $l_x \cap M \subseteq C_k$: since this holds for any $k \geq k(x)$, we derive $l_x \cap M = \{x\}$. This gives that $l_x \cap l_y = \emptyset$ whenever $x \neq y$ are two points in M, using again (6.6) and (6.5): one can then set $g(x)$ in such a way that it defines the correct l_x according to Theorem 6.3. The Borel assertion about g is easy to establish writing g as a pointwise limit of locally constant functions g_k, thus Theorem 6.3 is finally proved.

Then we built an example showing that Claim 6.1 is not correct.

6.3 Non-existence for dimension smaller than N

Here we present an example, first given in [2], showing that Claim 6.2 is not correct without suitable regularity assumptions.

Let us consider $X = Y = \mathbb{R}^2$, take a continuous function $\varphi : [0,1] \to [0,1]$ and define

$$f^+ := (\varphi, \mathrm{Id})_\# \left(\mathscr{H}^1 \llcorner [0,1] \right) \qquad\qquad f^- = \mathscr{L} \llcorner \left([2,3] \times [0,1] \right); \tag{6.11}$$

then f^+ and f^- are two probability measure: as shown in Figure 6.1, the first one is concentrated on the graph of φ (horizontally instead of vertically for the simplicity of the figure), while the second one on the square $[2,3] \times [0,1]$. Assuming, as in the rest of this chapter, that the cost function is the Euclidean distance $c(x,y) = |y - x|$, we prove now an intuitive result:

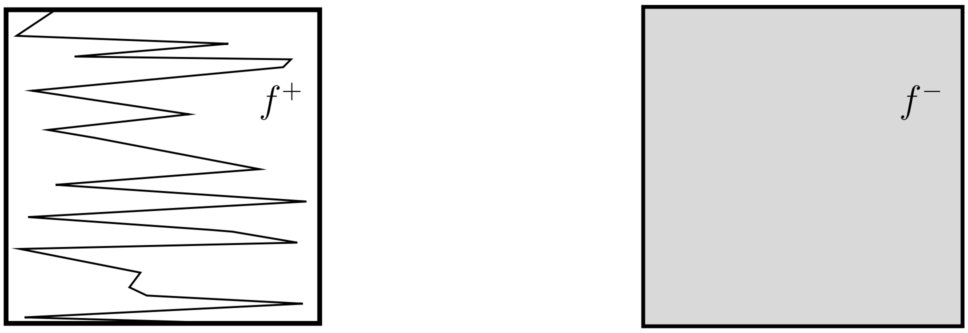


Figure 6.1: The measures f^+ and f^- in Section 6.3

Lemma 6.9. *The unique optimal transport plan is*

$$\bar{\gamma} := \int_{s=0}^{1} \delta_{(\varphi(s),s)} \otimes \mathscr{H}^1 \llcorner \big([2,3] \times \{s\}\big)\, ds,$$

which is the plan moving all the mass horizontally (i.e. all the transport rays are horizontal).

Proof. Let us take any transport plan γ: if we write each point $x \in \mathbb{R}^2$ as $x = (x_1, x_2)$, the cost of γ results

$$\mathfrak{C}(\gamma) = \iint_{\mathbb{R}^2\times\mathbb{R}^2} |y-x|\, d\gamma(x,y) \geq \iint_{\mathbb{R}^2\times\mathbb{R}^2} y_1 - x_1\, d\gamma(x,y) = \int_{\mathbb{R}^2} y_1\, df^-(y) - \int_{\mathbb{R}^2} x_1\, df^+(x)$$

and the last number does not depend on γ. Since by definition $|y-x| = y_1 - x_1$ for $\bar{\gamma}-$a.e. (x,y), it follows that $\bar{\gamma}$ is optimal. Moreover $\bar{\gamma}$ is the unique optimal transport plan, because if a transport plan γ between f^+ and f^- has the property that $|y-x| = y_1 - x_1$ $\gamma-$a.e., then it easily follows $\gamma = \bar{\gamma}$. □

In particular, from last lemma it follows that there are no optimal transport maps, since the unique optimal transport plan is clearly not induced by any map. This is apparently not a contradiction with Claim 6.2, since whenever φ is Lipschitz then f^+ is concentrated on a $N-1$ dimensional set (the graph of φ); however, this is no more true for highly non-regular continuous functions φ, thanks to the following known result (see Theorem 8.2 in [40]):

Lemma 6.10. *For any $1 < s < 2$ there is a continuous function $\varphi : [0,1] \to [0,1]$ whose graph is an $s-$dimensional set; as a consequence, if f^+ is defined as in (6.11), $f^+ \ll \mathscr{H}^\sigma$ for any $\sigma < s$.*

It follows that the existence result of Claim 6.2 is not correct without further assumptions; in particular, if $s < N$ one cannot infer the existence of optimal transport maps from the fact that $f^+ \ll \mathscr{H}^s$, while this can be done if $s = N$ as we will see in Theorem 6.17. This is a great difference with the quadratic case, i.e. when $c(x, y) = |y-x|^2$; in fact, recall that in that case it is sufficient that $s > N-1$, as we saw in Theorem 5.6: then, in particular, if f^+ and f^- are given as in (6.11) for φ as in Lemma 6.10, Theorem 5.6 and Lemma 6.9 ensure that there is a unique optimal transport plan γ_1 (resp. γ_2) for the Euclidean (resp. quadratic) case: but γ_1 is "horizontal" according to Lemma 6.9, while γ_2 is induced by a transport map.

Then this example shows a big difference between the Euclidean case and the strict convex case, and ensures that the hypothesis $f^+ \ll \mathscr{L} = \mathscr{H}^N$ in Theorem 6.17 is sharp.

6.4 Evans-Gangbo approach

In this section we will briefly see how Evans and Gangbo proved their existence result in 1999 ([39]); in view of the counterexamples to Sudakov's proof we discussed in Sections 6.2 and 6.3, now we know –even though this was not known when the work came out– that this is the first correct existence proof for the Euclidean cost.

They first consider the $p-$laplacian, i.e. the solution of

$$-\text{div } \left(|Du|^{p-2}Du\right) = f = f^+ - f^- : \tag{6.12}$$

the name comes from the fact that for $p = 2$ this problem reduces to the equation $-\Delta u = f$; the $p-$laplacian has been deeply studied, for its connection with the mass optimization see also [13]. Since the Euler–Lagrange associated functional is

$$\frac{1}{p}\int |Du|^p - uf, \tag{6.13}$$

which is strictly convex, the unique minimizer of (6.13) is the unique solution of problem (6.12): according to [39], let us call it u_p.

They prove then that there is some $u \in \mathrm{W}^{1,1}$ such that, up to a subsequence, u_p uniformly converge to u and Du_p weakly* converge in L^∞ to Du; moreover $|Du_p|^{p-2}Du_p$ weakly* converge in L^∞ to $a\,Du$ where a is a nonnegative and bounded function: then in particular

$$-\text{div } \left(aDu\right) = f.$$

This function a is referred to as the *transport density.* This is a really important function in mass transportation, which had been already considered also by Janfalk (see [52]), and in the work of Evans and Gangbo it is one of the main tools to build the transport map. The transport density represents also the connection between the mass transportation and other problems, and its properties will be carefully studied in the second and in the third Part of this thesis. The importance of this function is such that the notation "transport density" has been kept also for more general situations studied later, where it is only a measure and then the name "density" would actually not be appropriate.

Once having the functions u and a one tries to solve, at least a.e., the time-dependent system

$$\begin{cases} x'(x_0,s) = \dfrac{-a\big(x(x_0,s)\big)Du\big(x(x_0,s)\big)}{(1-s)f^+\big(x(x_0,s)\big)+sf^-\big(x(x_0,s)\big)} \\ x(x_0,0)=x_0 \end{cases} \tag{6.14}$$

for $0\leq s\leq 1$, see also [26]. The solution of (6.14) gives the missing information one needs to build a transport map; let us in fact consider for any time $0\leq s\leq 1$ the function

$$f_s = x(\cdot,s)_\# f^+ :$$

it is clear that $f_0=f^+$; under the appropriate hypotheses, it happens also that $f_1=f^-$, thus $t:=x(\cdot,1)$ is a transport map: moreover, t is an optimal transport map. The final result they prove is then

Theorem 6.11 (Evans–Gangbo). *Assume that φ and ψ are positive Lipschitz functions in $\mathbb{R}^N$ with supports X and Y, and assume that*

$$\int_X \varphi(x)\,dx = \int_Y \psi(y)\,dy,$$

that X and Y are disjoint compact sets with smooth boundary, and that φ and ψ are strictly positive in the interior of X and Y respectively. Then there exists an optimal transport map t from $f^+=\varphi(x)\,dx$ to $f^-=\psi(y)\,dy$.

What we gave is just a very rough idea of how the construction of Evans and Gangbo works; however, they needed to solve a lot of technical problems, and their proofs contain many involved calculations. Just to give an idea of what one has to face, we remark that a formal solution of (6.14) does not exist since if the supports of f^+ and f^- are disjoint then for each x_0 there is a division by 0 in a non-empty interval of times: in fact, they solve a slightly modified version of (6.14) and then pass to a limit.

It is worth to be mentioned that in [39] it is also first introduced the concept of transport set and transport rays, which has been deeply used and generalized in last years and have been introduced in Section 4.1.

Let us also finally mention the fact that the time-dependent approach of the construction is related to the configurations method introduced by Brenier [10, 18], which will be studied in the Second part of this thesis.

6.5 The results of Caffarelli-Feldman-McCann and Trudinger-Wang

In this section we will briefly describe two existence results for optimal transport maps with the Euclidean cost, which were indepentenly (and almost contemporarily) found by Caffarelli, Feldman and McCann and by Trudinger and Wang.

Caffarelli-Feldman-McCann

In [24], the authors build an optimal transport map t considering the Euclidean cost $|y-x|$ as the limit of $|y-x|^{1+\varepsilon}$ when $\varepsilon \to 0$. As we already discussed in last chapter, the strictly convex cost functions are easier to deal with than the convex ones, such as the norms in $\mathbb{R}^N$: in fact, it is clear that any norm is a convex but not strictly convex function; analogously, among the norms, the strict convexity of the norm is a useful fact, since for example it implies the no-crossing condition as shown in Lemma 3.18. Then the authors consider the easy-to-solve problems

$$\inf \iint_{\mathbb{R}^N \times \mathbb{R}^N} \|y-x\|^{1+\varepsilon} \, d\gamma(x,y),$$

and they are able to infer from them the informations which are needed in order to find an optimal transport map t: more precisely, they use the duality we discussed in Chapter 3, finding a particular Kantorovich potential for the Euclidean problem as a limit of some sequence of Kantorovich potentials for the approximating problems. It is worth to be noticed that they pass also through a decomposition in transport rays and through a solution in each ray, using then the Sudakov's idea but making it rigorous. The result they prove is then the following

Theorem 6.12 (Caffarelli-Feldman-McCann). *Assume $X = Y = \mathbb{R}^N$ and $c(x,y) = |y-x|$ is the Euclidean norm (or also a quasi-euclidean norm); assume moreover that f^+ and f^- are absolutely continuous with respect to $\mathscr{L}$ and have compact support. Then there exists an optimal transport map t from f^+ to f^-.*

The notion of quasi-euclidean norm will be presented and discussed in Chapter 8: as we will see, it is an hypothesis that has been used in mass transportation problems more than once.

In [43] the preceding result is sharpened to prove some existence result also on a Riemannian manifold, and the authors use the absolute continuity of the sole f^+.

Trudinger-Wang

In [73], the authors provide a proof of the existence of optimal transport maps in the Euclidean case, and they reach the same result as Caffarelli, Feldman and McCann (but without any extension to other cost functions, as quasi-euclidean norms):

Theorem 6.13 (Trudinger-Wang). *If f^+ and f^- are absolutely continuous and have compact support in $\mathbb{R}^N$, then there exists an optimal transport map from f^+ to f^- with the Euclidean cost function $c(x, y) = |y - x|$.*

The authors underlined the fact that their result was a bit stronger than the one of Evans and Gangbo but quite weaker than the one of Sudakov, and then they regarded it as a simplification of the proof, not as a true step forward in the theory. However, since Sudakov's claim does not hold, today we know that the Theorem proved by Trudinger and Wang was also stronger that all the other results known at that time.

The construction of Trudinger and Wang is somehow similar to the one of Caffarelli, Feldman and McCann: they also consider the problems with cost functions $c_\varepsilon(x, y) = |y - x|^{1+\varepsilon}$, and then they find a Kantorovich potential for the Euclidean problem as the uniform limit of a subsequence of the potentials for the approximated problems (recall that the region they have to consider is compact). The transport map is then separately built on the different transport rays (thus again following Sudakov idea), and on each transport ray it is the unique non-decreasing transport map we already presented in Section 4.2. As in the works of Evans and Gangbo and of Caffarelli, Feldman and McCann, to bypass the problems we encountered and discussed about Sudakov work they used some Lipschitz estimates; until now, all the works about the existence for the Euclidean setting have used, in different ways, some Lipschitz-type property (such as the countable Lipschitz property we defined in Chapter 4).

6.6 The result of Ambrosio

In this section we will prove an existence result for the Euclidean cost function $c(x, y) = |y - x|$ in $\mathbb{R}^N$, which states the existence of an optimal transport plan whenever $f^+ \ll \mathscr{L}$. This claim was proved by Ambrosio in 2000, [1], a few months after the proofs of

Caffarelli-Feldman-McCann and of Trudinger-Wang we discussed in last section; here we will essentially follow that proof.

First of all, we prove an important technical result holding under the assumption of the countable Lipschitz property; it does not depend on a particular choice of the cost function c, and it will be used in the sequel for cost functions different from the Euclidean one. To state and prove it, one needs to remind the definitions given in Section 4.1. Note that this lemma is the Claim 6.1 of Sudakov, which is now true thanks to the additional countable Lipschitz hypothesis.

Lemma 6.14. *Take a $\sigma-$compact set $\Gamma \subseteq \mathbb{R}^N \times \mathbb{R}^N$ fulfilling the no-crossing condition (4.1); take then a measure $\mu \ll \mathscr{L} \llcorner \mathrm{ET}_\Gamma$ such that τ_Γ is countably Lipschitz with respect to μ: writing $\mu = \mu_C \otimes {r_\Gamma}_\# \mu$ with a disintegration with respect to r_Γ, $\mu_C \ll \mathscr{H}^1 \llcorner C$ for ${r_\Gamma}_\#\mu-$a.e. closed segment $C \in \mathcal{S}_c$.*

Before to prove this lemma, we need to recall something about the coarea formula, an extension of Fubini Theorem first proved by Federer (see for example Section 2.12 in [3], or Theorem 3.2.22 in [41], or [70]): given a Lipschitz function $\varphi : \mathbb{R}^N \to \mathbb{R}^{N-1}$, for $\mathscr{L}-$a.e. point $x \in \mathbb{R}^N$ one defines the *coarea factor*

$$\mathbf{C}\varphi(x) := \sqrt{\sum_A \mathrm{Det}^2 A}, \tag{6.15}$$

where the sum runs on all the $(N-1) \times (N-1)$ minors A of $\nabla\varphi(x)$. Then $\mathbf{C}\varphi$ is a $\mathscr{L}-$measurable function, $\varphi^{-1}(y)$ is countably $\mathscr{H}^1-$rectifiable for $\mathscr{H}^{N-1}-$a.e. y and the *coarea formula*

$$\mathbf{C}\varphi \mathscr{H}^N = \mathscr{H}^1 \llcorner \varphi^{-1}(y) \otimes \mathscr{H}^{N-1}(y) \tag{6.16}$$

holds. One can easily note that the coarea formula reduces to standard Fubini Theorem when φ is the linear projection on a $(N-1)-$dimensional hyperplane. We can now prove the above stated lemma:

Proof. *(of Lemma 6.14):* First of all we note that the disintegration of μ with respect to r_Γ is possible since r_Γ is a Borel map according to Lemma 4.4; moreover, up to a disjoint countable union and a rigid motion, we can assume that

(i) each transport ray intersects $\{x : x_1 = 0\}$;

(ii) τ_Γ is Lipschitz on a set $K \subseteq \mathrm{ET}_\Gamma$ and μ is concentrated in K;

(iii) $\tau_\Gamma(x) \cdot \mathrm{e}_1 \geq 1/2$ for any $x \in \mathrm{ET}_\Gamma$;

(iv) ET_Γ is contained in the strip $\{x \in \mathbb{R}^N : x \cdot \mathrm{e}_1 = x_1 \in [-\varepsilon, \varepsilon]\}$ with some $\varepsilon \ll \mathrm{Lip}(\tau_\Gamma)$ to be determined later.

Thanks to (i) and (iii), we can define the function $\varphi : \mathrm{ET}_\Gamma \to \mathbb{R}^{N-1}$ as $\varphi(x) = y$, where $y \in \mathbb{R}^{N-1}$ is the unique point such that $(0, y)$ belongs to the closed segment $r_\Gamma(x)$.

We want now to show that φ is Lipschitz and then to evaluate the coarea factor of φ: to do that, let us write $\tau_\Gamma = (\tau_1, \tau_y)$ so that

$$\varphi(s, y) = y - s\,\frac{\tau_y(s, y)}{\tau_1(s, y)};$$

since τ_Γ is Lipschitz according to (ii) and $\tau_1(s, y) \geq 1/2$ thanks to (iii), it follows that φ is Lipschitz. Moreover for each $(s, y) \in \mathrm{ET}_\Gamma$ one easily calculates

$$\nabla_y \varphi(s, y) = \mathrm{Id}_{N-1} - \frac{s}{\tau_1}\,\nabla_y \tau_y - \frac{s}{\tau_1^2}\,\tau_y \times \nabla_y \tau_1 : \tag{6.17}$$

note that the preceding equation makes sense since $\nabla_y \varphi(s, y)$ is a $(N-1) \times (N-1)$ matrix, as Id_{N-1} and $\nabla_y \tau_y$; s and τ_1 are real numbers; τ_y and $\nabla_y \tau_1$ are $(N-1)-$vectors, then their product $\tau_y \times \nabla_y \tau_1$ is also a $(N-1) \times (N-1)$ matrix. Recalling again that $\tau_1(s, y) \geq 1/2$ and that $|s| \leq \varepsilon$ thanks to (iv), (6.17) gives $\mathrm{Det}\big(\nabla_y \varphi(s, y)\big) \geq \delta > 0$ provided ε is sufficiently small with respect to $\mathrm{Lip}(\tau_\Gamma)$. This implies, recalling (6.15), that the coarea factor $\mathbf{C}\varphi$ is greater than δ for $\mathscr{L}-$a.e. point in ET_Γ. Then, using coarea formula (6.16), we have (recall that $\mu \ll \mathscr{L}$, thus the measure μ is in particular a real function)

$$\mu = \frac{\mu}{\mathbf{C}\varphi}\,\mathbf{C}\varphi \mathscr{H}^N = \frac{\mu}{\mathbf{C}\varphi}\,\mathscr{H}^1 \llcorner \varphi^{-1}(y) \otimes \mathscr{H}^{N-1}(y) = \mu_y \otimes \nu;$$

we have then $\mu = \mu_C \otimes r_{\Gamma\#}\mu = \mu_y \otimes \nu$, and we want to derive that the two decomposition are "similar". To give a meaning to this assertion we note that, also thanks to the no-crossing condition (4.1), $y \to \varphi^{-1}(y)$ is a one-to-one function between the image of φ (a Borel subset of $\mathbb{R}^{N-1}$) and the set of M the maximal transport rays (a Borel subset of $\mathcal{S}_c$); then in the sequel of the proof we will identify for simplicity y with $C = \varphi^{-1}(y)$. We define now $g(y) = \|\mu_y\|$ and

$$\nu' := g\nu, \qquad\qquad \mu'_y := \frac{\mu_y}{g(y)} :$$

then μ'_y is a probability measure as μ_C and $\mu = \mu'_y \otimes \nu' = \mu_C \otimes r_{\Gamma\#}\mu$; we also claim that

$$\nu' = r_{\Gamma\#}\mu, \qquad\qquad \mu'_y = \mu_C \tag{6.18}$$

for $r_{\Gamma\#}\mu-$a.e. C: note that, as we identify y with $C = \varphi^{-1}(y)$, we also identify ν', a measure on $\mathbb{R}^{N-1}$, with $\varphi^{-1}_{\#}\nu'$, a measure on $\mathcal{S}_c$: thus the first equality in (6.18) makes sense. This will lead to the conclusion, since the measure μ'_y equals, up to the real number $g(y)$, the measure μ_y, which is absolutely continuous with respect to $\mathscr{H}^1 \llcorner C$ by definition.

To show (6.18), take a Borel set $D \subseteq M \subseteq \mathcal{S}_c$ and define $\psi : \mathbb{R}^N \to \mathbb{R}$ as

$$\psi(x) := \begin{cases} 1 & \text{if } x \in C \text{ for some } C \in D, \\ 0 & \text{otherwise.} \end{cases}$$

Since μ'_y and μ_C are probability measures for any y (hence any C), $\langle \mu'_y, \psi \rangle = \langle \mu_C, \psi \rangle$ is 1 if $C \in D$ and 0 otherwise. Then $\nu'(D) = \langle \mu, \psi \rangle = r_{\Gamma\#}\mu(D)$ and, by the generality of D, we infer the first equality in (6.18). By the uniqueness part in Theorem A.15, it immediately follows the second equality, so (6.18) is proved and the thesis follows. □

From geometric considerations and making use of the preceding result, there are two important consequences:

Corollary 6.15. *Given a* $1-$*Lipschitz function* u *and defining* $\Gamma := \partial^c u$*, the set* D_Γ *is* $\mathscr{L}-$*negligible.*

Corollary 6.16. *Given a* $1-$*Lipschitz function* u *and defining* $\Gamma := \partial^c u$*, the countable Lipschitz property with respect to* $\mathscr{L}$ *holds.*

We do not prove here Corollary 6.16, since it will be done in a greater generality with Lemmas 8.4 and 8.3 in Chapter 8. We just point out that, as we said in last section, all the existence results proved for the Euclidean case make use of some Lipschitz-type property to bypass the problems in Sudakov approach; then there are several claims similar to Corollary 6.16, achieved with different proofs (one can for example take a look of the third step in Proposition 4.1 of [39], or of Lemma 3.1 of [73]). An alternative proof of our claim can be obtained via geometrical arguments: in fact, Corollary 6.16 is an immediate consequence of Lemma 14.16, and also the more general Lemma 8.3 can be obtained from the same result by means of a simple argument.

We proof now the main result of this section, taken from [1]:

Theorem 6.17 (Ambrosio). *If* f^+ *and* f^- *have finite first order moments (equivalently, if* c *is slow* $-(f^+, f^-)-$ *increasing, thanks to Remark 3.21) and* $f^+ \ll \mathscr{L}$*, then there exists an optimal transport map from* f^+ *to* f^- *with the Euclidean cost* $c(x, y) = |y - x|$*.*

Proof. This proof follows essentially the original idea of Sudakov, making it rigorous.

Let us take a Kantorovich potential u, which exists thanks to Theorem 3.28, and take an optimal transport plan γ, which we know to be concentrated in $\Gamma := \partial^c u$; in the first four steps we will consider the case when

$$f^+\big(\mathbb{R}^N \setminus \mathrm{ET}_\Gamma\big) = 0 : \tag{6.19}$$

we will remove this assumption in the last step.

Step I. *Disintegration of γ.*

Let us consider the function $r_{\Gamma} : \mathrm{ET}_\Gamma \to \mathcal{S}_c$ which associates to any point x the closure of the maximal transport ray containing x: this map has been already considered in Section 4.1, and it is Borel as shown by Lemma 4.4. Thanks to assumption (6.19), $r_{\Gamma} \circ \pi_1(x, y)$ is defined for $\gamma-$a.e. (x, y), and it associates to any pair (x, y) a maximal transport ray containing both the points x and y. We can then operate a disintegration of γ with respect to the map $r_{\Gamma} \circ \pi_1$, which allows to write, according to Theorem A.15,

$$\gamma = \gamma_C \otimes \nu \tag{6.20}$$

where $\nu := (r_{\Gamma} \circ \pi_1)_{\#}\gamma$: the probability measure γ_C is a transport plan between its marginals f_C^+ and f_C^- (and in particular for $\nu-$a.e. C it is optimal according to Lemma 1.13). Moreover, recalling Lemma A.16, we know

$$f^+ = f_C^+ \otimes \nu \qquad\qquad f^- = f_C^- \otimes \nu. \tag{6.21}$$

Step II. *Non-atomicity of f_C^+.*

We want to apply Lemma 6.14: this is possible since τ_{Γ} is countably Lipschitz with respect to $\mathscr{L}$ (thus w.r.t. f^+) thanks to Corollary 6.16. It follows that f_C^+ is absolutely continuous with respect to $\mathscr{H}^1 \llcorner C$, and then in particular non-atomic, for $\nu-$a.e. C. Then $\mathrm{T}_\Gamma = \mathrm{ET}_\Gamma$ up to a f^+-negligible set, since $\mathrm{ET}_\Gamma \setminus \mathrm{T}_\Gamma$ contains at most a point for each C and recalling (6.21); then (6.19) reads as

$$f^+\big(\mathbb{R}^N \setminus \mathrm{T}_\Gamma\big) = 0. \tag{6.22}$$

Step III. *Construction of the map t.*

For any C such that f_C^+ is non-atomic (and then for $\nu-$a.e. C), we can consider the mass transport problem with data f_C^+ and f_C^-; recalling Theorem 4.9, there exists a unique optimal transport map t_C, which is the non-decreasing one. Recall now that the closures of different maximal transport rays can intersect only at a common endpoint, then – thanks to (6.22), which follows by hypothesis (6.19) as noted at the end of last step– they are essentially disjoint with respect to f^+: we can then glue the different transport maps, i.e. there exists a map $t : \mathbb{R}^N \to \mathbb{R}^N$ such that t equals t_C on each segment C up to a f^+-negligible set. It is reasonable to guess that this map t must be an optimal transport map, but this is not completely trivial: a priori, we do not even know if $t_{\#}f^+ = f^-$.

We will show now that t is f^+-measurable; this will be done applying Lemma A.19 with $X = \mathbb{R}^N$, $Y = \mathcal{S}_c$, $Z = \mathbb{R}^N$, $\pi = r_{\Gamma}$, $\mu = f^+$ and $g = t$: if it is possible, we will deduce that t equals a Borel map out of a f^+-negligible set, then in particular t is f^+-measurable.

The fact that t is f^+_C measurable for $\nu-$a.e. C is obvious, since $t = t_C$ up to a f^+_C negligible set and t_C is an optimal transport map between f^+_C and f^-_C.

We need then to show that $C \to (\mathrm{Id}, t_C)_\# f^+_C$ is a $r_{\Gamma\#}f^+-$measurable measure valued map in the sense of Definition A.14. Let us take a rectangle $B = B_1 \times B_2$ in $\mathbb{R}^N \times \mathbb{R}^N$, where B_1 and B_2 are closed cubes in $\mathbb{R}^N$: it will be sufficient to show that

$$C \longrightarrow (\mathrm{Id}, t_C)_\# f^+_C(B)$$

is a real $r_{\Gamma\#}f^+-$measurable function in $\mathcal{S}_c$. If we call $D \subseteq \mathcal{S}_c$ the set of the closed

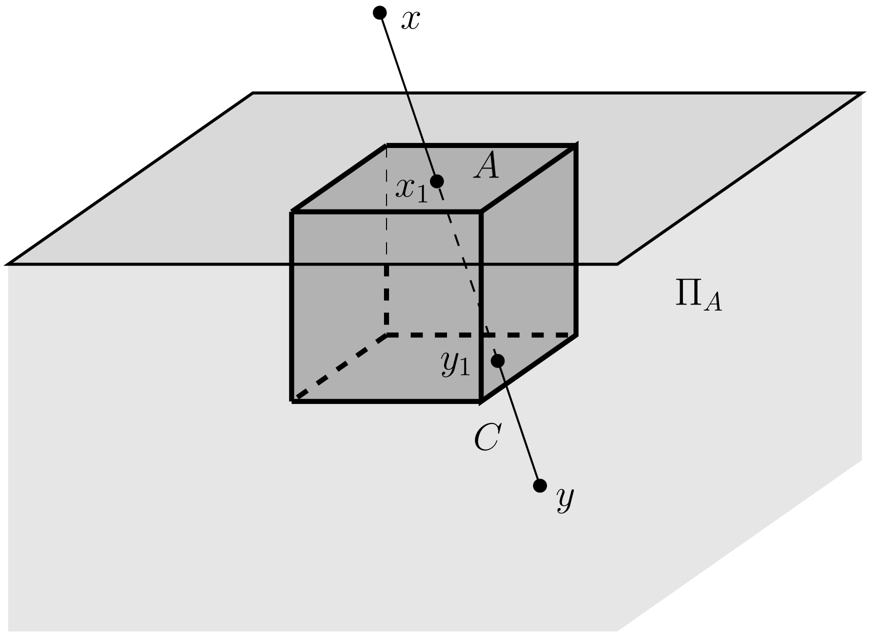


Figure 6.2: Borel property for the map t

segments intersecting both B_1 and B_2, it is clear that D is a closed subset of $\mathcal{S}_c$ and that $(\mathrm{Id}, t_C)_\# f^+_C(B) = 0$ if $C \notin D$. On the other hand, for any $C \in D$ we can use the formula (4.6) for the non-decreasing transport plan that we established in Theorem 4.9: it follows that

$$(\mathrm{Id}, t_C)_\# f^+_C(B) = \Big(\big(M_C \wedge \widetilde{M}_C \big) - \big(m_C \vee \widetilde{m}_C \big) \Big) \vee 0$$

calling $[\![x_1, y_1]\!]$ and $[\![x_2, y_2]\!]$ the intersections of C with B_1 and B_2 and setting

$$m_C := f^+_C\big([\![x, x_1[\![\big), \qquad M_C := f^+_C\big(]\!]x, y_1]\!]\big);$$
$$\widetilde{m}_C := f^-_C\big([\![x, x_2[\![\big), \qquad \widetilde{M}_C := f^-_C\big(]\!]x, y_2]\!]\big).$$

We will have our claim, then, showing that m_C, $\widetilde{m}_C$, M_C and $\widetilde{M}_C$ are Borel functions in D. We consider here m_C, the other three are similar; first of all, we reduce to the case of the closed segments in C whose first point of intersection with B_1 lies on one of its faces, say A (recall that the elements of $\mathcal{S}_c$ are oriented segments); this is possible dividing the Borel subset D of $\mathcal{S}_c$ in 2^N Borel subsets. If we call Π_A the half-space delimited from above by the hyperplane extending A, as in Figure 6.2, it is then $m_C = 1 - f_C^+(\Pi_A)$. Finally, the claim follows since $C \to f_C^+$ is a $r_{\Gamma\#}f^+-$measurable measure-valued map according to Theorem A.15.

Step IV. *Optimality of the transport map t.*

We show now that t is a transport map and it is optimal: first of all we define the measure $\bar{\gamma} := (\mathrm{Id}, t)_\# f^+$, which is concentrated in $\mathrm{ET}_\Gamma \times \mathbb{R}^N$ by construction: then $r_\Gamma \circ \pi_1$ is defined $\bar{\gamma}-$a.e., thus we can operate the disintegration of $\bar{\gamma}$ with respect to $r_\Gamma \circ \pi_1$, which gives, similarly to (6.20),

$$\bar{\gamma} = \bar{\gamma}_C \otimes \bar{\nu}. \tag{6.23}$$

We would like to prove that $\bar{\nu} = \nu$ and that $\bar{\gamma}_C$ is the transport plan between f_C^+ and f_C^- induced by t_C: in fact, $\bar{\gamma}_C$ is concentrated in the graph of t and also in the segment C, thus in their intersection which is the graph of t_C. Then, recalling the discussion after Lemma 1.6, we have $\bar{\gamma}_C = (\mathrm{Id}, t_C)_\# \pi_1 \bar{\gamma}_C$: on the other hand, the transport plan between f_C^+ and f_C^- induced by t_C is $(\mathrm{Id}, t_C)_\# f_C^+$, thus we need to check that $\pi_1 \bar{\gamma}_C = f_C^+$.

To show this fact, we first use that $\pi_1 \bar{\gamma} = \pi_1 \gamma = f^+$ by construction thus, applying Lemma A.16 to (6.20) and (6.23),

$$f^+ = (\pi_1 \gamma_C) \otimes \nu = (\pi_1 \bar{\gamma}_C) \otimes \bar{\nu} : \tag{6.24}$$

since ν and $\bar{\nu}$ are measures on $\mathcal{S}_c$ concentrated on a set of essentially (with respect to f^+) disjoint segments, i.e. the set MTR of the closures of the maximal transport rays –which is Borel by Lemma 4.5– we want to infer from (6.24) that $\nu = \bar{\nu}$ and $\pi_1 \bar{\gamma}_C = \pi_1 \gamma_C$ $\nu-$a.e.; in fact, take any Borel subset K of $\mathcal{S}_c$, and define $K_0 = K \cap MTR$ and $A = \cup\{C : C \in K_0\}$. Recalling (6.24), we have

$$f^+(A) = \Big((\pi_1 \gamma_C) \otimes \nu\Big)(A) = \int_{C \in \mathcal{S}_c} \pi_1 \gamma_C(A)\, d\nu(C).$$

But the definition of A and the fact that each $\pi_1 \gamma_C$ is a probability measure concentrated on C give $\pi_1 \gamma_C(A) = 1$ for any $C \in K_0$; on the other hand, if $C \notin K_0$ (but $C \in MTR$, otherwise it is not seen by ν) then $C \cap A$ is contained in the extreme points of C, thus its intersection with T_Γ is empty: recalling (6.22), we infer $\pi_1 \gamma_C(A) = 0$ for $\nu-$a.e. $C \notin K_0$. Summarizing, we have $f^+(A) = \nu(K_0) = \nu(K)$: similarly, again thanks to (6.24), we have also $f^+(A) = \bar\nu(K)$, and since K is a generic Borel subset of $\mathcal{S}_c$ we infer $\nu = \bar\nu$. As a consequence, from (6.21) and Lemma A.16 it follows that for $\nu-$a.e. C it is $\pi_1\bar\gamma_C = \pi_1\gamma_C = f_C^+$, then $\bar\gamma_C$ is the transport plan between f_C^+ and f_C^- induced by t_C as claimed.

Therefore, $\pi_2\bar\gamma_C = f_C^-$ and then, recalling (6.21), we infer $\pi_2\bar\gamma = f^-$: then $\bar\gamma$ is a transport plan between f^+ and f^- and so t is a transport map from f^+ to f^-.

We use now that $\bar\gamma_C$ is an optimal transport plan between f_C^+ and f_C^-, thus

$$\mathfrak{C}(\bar\gamma_C) \leq \mathfrak{C}(\gamma_C). \tag{6.25}$$

Moreover, thanks to (6.20) and the properties of disintegration it is

$$\mathfrak{C}(\gamma) = \int_{\mathcal{S}_c}\int_{\mathbb{R}^N\times\mathbb{R}^N} c(x,y)\,d\gamma_C(x,y)\,d\nu = \int_{\mathcal{S}_c}\mathfrak{C}(\gamma_C)\,d\nu$$

and similarly, thanks to (6.23)

$$\mathfrak{C}(t) = \mathfrak{C}(\bar\gamma) = \int_{\mathcal{S}_c}\mathfrak{C}(\bar\gamma_C)\,d\nu.$$

Recalling (6.25), we finally infer the optimality of $\bar\gamma$ and of t.

Step V. *General case.*

Here we remove hypothesis (6.19). We first recall that, according to Remark 4.2, each point $x \in \operatorname{spt} f^+ = \pi_1(\Gamma)$ must be either in ET_Γ, or in D_Γ, or in F_Γ: in words, it is contained in the transport set, or it is a doubling point, or it is a fixed point. Thanks to Corollary 6.15, we know that $f^+(\mathrm{D}_\Gamma) = 0$ and we can write

$$\gamma = \gamma \llcorner \left(\pi_1(\mathrm{ET}_\Gamma)\times\mathbb{R}^N\right) + \gamma\llcorner\left(\pi_1(\mathrm{F}_\Gamma)\times\mathbb{R}^N\right) = \gamma_1 + \gamma_2.$$

Using Lemma 1.12, we know that both γ_1 and γ_2 are optimal transport plans for the problem given by their marginals. We consider then separately γ_1 and γ_2.

Concerning γ_1, it is concentrated by construction in $\Gamma_1 = \Gamma \setminus \Delta$, where Δ is the diagonal of $\mathbb{R}^N\times\mathbb{R}^N$; moreover, the property (6.19) holds for $f_1^+ = \pi_1\gamma_1$ and Γ_1, then applying the first four steps we find an optimal transport map $t_1 : \mathrm{ET}_\Gamma \to \mathbb{R}^N$ from f_1^+ to $f_1^- = \pi_2\gamma_1$. On the other hand, γ_2 is concentrated in the diagonal, so the identity map $t_2 : \Delta\setminus\mathrm{ET}_\Gamma \to \mathbb{R}^N$ is an optimal transport map because it has a null cost. Thanks to the fact that ET_Γ and $\Delta\setminus\mathrm{ET}_\Gamma$ are disjoint sets, we can glue t_1 and t_2 in a transport map $t : \mathbb{R}^N \to \mathbb{R}^N$ from f^+ to f^-: since the cost of t is less than that of γ, which is optimal, the optimality of t and then the thesis follow. □

Remark 6.18. *As in the quadratic case (see Theorem 5.6), in last Theorem if it is also $f^- \ll \mathscr{L}$ then the optimal transport map t built in this way is invertible and the inverse* *is an optimal transport map from f^- to f^+; in fact, consider any maximal transport ray C: recalling the one-dimensional situation studied in Section 4.2, the only non-decreasing transport plan between f^+_C and f^-_C is induced by a transport map from f^+_C to f^-_C since $f^+ \ll \mathscr{L}$, but it is also induced by a transport map from f^-_C to f^-_C since $f^- \ll \mathscr{L}$. More in general, see Corollary 8.7.*

Chapter 7

A general existence result

In this chapter we give a general existence result, slightly refining Theorem 6.1 of [2]. As better discussed in Remark 7.5, this assert covers all the previously known existence theorems for the case when $c(x,y) = \|y-x\|$ is some norm in $\mathbb{R}^N$; moreover, it allows to deduce new results. In fact, making use of it, in the next chapters we will consider the cases of the crystalline norms and of the general norms in $\mathbb{R}^2$.

Our aim is to present a general result to show the existence of a particular transport plan which is induced by a transport map and which, in some situations we are interested in, is optimal: the idea is to generalize Theorem 6.17, so first of all we need to regard its proof to identify the main ingredients. The first step consisted in operating a disintegration of γ with respect to $r_\Gamma \circ \pi_1$: to do that it is necessary that r_Γ is Borel, and it was ensured by Lemma 4.4 using the no-crossing property (4.1). The second step consisted in verifying that each f_C^+ is non-atomic, and it was true thanks to the fact that $f^+ \ll \mathscr{L}$ and to Lemma 6.14, for which the countable Lipschitz property of τ_Γ was necessary: recall that also to define τ_Γ one needs that (4.1) holds. The third and the fourth step consisted in defining a map t, showing that it is Borel and it is a transport map, and proving that it is optimal; they did not use anything coming from the particular choice of the Euclidean cost. The last step consisted in handling the points not contained in ET_Γ, and it was necessary the fact that the set D_Γ of the doubling points was f^+-negligible.

We will see now that these four hypotheses, namely the validity of (4.1), the absolute continuity of f^+, the countable Lipschitz property of τ_Γ and the f^+-negligibility of D_Γ, are in fact sufficient to provide the same construction of Theorem 6.17, leading to a measure $\bar{\gamma}$ induced by a transport map and "better" than γ.

Theorem 7.1. *Let $X = Y = \mathbb{R}^N$ and let $\Gamma \subseteq \mathbb{R}^N \times \mathbb{R}^N$ be a $\sigma-$compact set such that the following is true:*

(i) the no-crossing condition (4.1) holds;

(ii) f^+ is absolutely continuous with respect to $\mathscr{L}$;

(iii) D_Γ is f^+-negligible;

(iv) τ_Γ has the countable Lipschitz property with respect to f^+.

Then, if γ is a transport plan between f^+ and f^- concentrated in Γ, there exists another transport plan $\bar\gamma$ such that

(a) $\bar\gamma$ is induced by a transport map t.

(b) $\bar\gamma$ is concentrated on the set of pairs (x, y) such that either $x = y$ or $[\![x, y]\!]$ is contained in the closure of a maximal transport ray of Γ.

(c) For any convex function $\phi : \mathbb{R}^N \to \mathbb{R}$ we have

$$\int_{\mathbb{R}^N\times\mathbb{R}^N} \phi(y-x)\,d\bar\gamma \le \int_{\mathbb{R}^N\times\mathbb{R}^N} \phi(y-x)\,d\gamma$$

and, if ϕ is strictly convex, the above inequality is strict unless $\gamma = \bar\gamma$.

Proof. As in Theorem 6.17, we begin assuming that

$$f^+\big(X \setminus \mathrm{ET}_\Gamma\big) = 0 \tag{7.1}$$

and this hypothesis will be removed in the last step.

Step I. *Disintegration of γ.*

Let us consider again the Borel function $r_\Gamma : \mathrm{ET}_\Gamma \to \mathcal{S}_c$ defined in Section 4.1; let us also consider $r_\Gamma \circ \pi_1(x, y) = r_\Gamma(x)$: it is defined for γ−a.e. (x, y) thanks to assumption (7.1) and then, exactly as in Theorem 6.17, we can operate the disintegration of γ with respect to it, leading to $\gamma = \gamma_C \otimes \nu$ with $\nu := (r_\Gamma \circ \pi_1)_\#\gamma$: moreover γ_C is a transport plan between its marginals f^+_C and f^-_C and, recalling Lemma A.16, it is

$$f^+ = f^+_C \otimes \nu, \qquad\qquad f^- = f^-_C \otimes \nu.$$

Step II. *Non-atomicity of f^+_C.*

We apply again Lemma 6.14, which is possible thanks to the hypotheses, thus we infer that f^+_C is absolutely continuous with respect to $\mathscr{H}^1 \llcorner C$, and then in particular it is non-atomic, for ν−a.e. C. We derive that $\mathrm{ET}_\Gamma = \mathrm{T}_\Gamma$ up to an f^+−negligible set, or equivalently that $f^+(\mathrm{E}^l_\Gamma \setminus \mathrm{D}_\Gamma) = 0$.

Step III. *Construction of the map t.*

As in Theorem 6.17, we take a non-decreasing transport map t_C from f_C^+ to f_C^- whenever f_C^+ is non-atomic, making use of Theorem 4.9. Since the closures of different transport rays cannot intersect except in the extreme points, which are f^+-negligible as noticed in last step, there is a map $t : \mathbb{R}^N \to \mathbb{R}^N$ which, up to a f^+-negligible set, equals t_C on each C. This map is f^+-measurable: to show this fact, exactly as we made in Step III of Theorem 6.17, it suffices to apply Lemma A.19, since t is f_C^+ measurable for $\nu-$a.e. C and $C \to (\mathrm{Id}, t_C)_\# f_C^+$ is a $r_{\Gamma\#} f^+-$measurable measure-valued map according to Definition A.14.

Step IV. *Optimality of the transport plan $\bar{\gamma}$.*

Here we show that the measure $\bar{\gamma} = (\mathrm{Id}, t)_\# f^+$ belongs to $\Pi(f^+, f^-)$ and fulfills properties (a), (b) and (c) claimed.

The first fact can be done as in Theorem 6.17: since $r \circ \pi_1(x, y)$ is defined for $\bar{\gamma}-$a.e. (x, y), one can operate a disintegration of $\bar{\gamma}$ with respect to $r \circ \pi_1$, which leads to the decomposition $\bar{\gamma} = \bar{\gamma}_C \otimes \nu$, where $\bar{\gamma}_C = (\mathrm{Id}, t_C)_\# f_C^+$ is the non-decreasing transport plan between f_C^+ and f_C^- induced by t_C. It follows, using Lemma A.16, that $\pi_2 \bar{\gamma} = f_C^- \otimes \nu = f^-$, thus $\bar{\gamma} \in \Pi(f^+, f^-)$.

Property (a) is clear by definition; concerning property (b), it is clear recalling that $\bar{\gamma}$ is concentrated on the graph of t: in fact, the construction of t ensures that each open segment $\rrbracket x, y \llbracket$ with $(x, y) \in \operatorname{spt} \bar{\gamma}$ is contained in some maximal transport ray of Γ. Concerning property (c), arguing by disintegration we can write

$$\int_{\mathbb{R}^N \times \mathbb{R}^N} \varphi(y - x)\, d\gamma(x, y) = \int_{\mathcal{S}_c} \left(\int_{\mathbb{R}^N \times \mathbb{R}^N} \varphi(y - x)\, d\gamma_C(x, y) \right) d\nu \tag{7.2}$$

$$\int_{\mathbb{R}^N \times \mathbb{R}^N} \varphi(y - x)\, d\bar{\gamma}(x, y) = \int_{\mathcal{S}_c} \left(\int_{\mathbb{R}^N \times \mathbb{R}^N} \varphi(y - x)\, d\bar{\gamma}_C(x, y) \right) d\nu. \tag{7.3}$$

Let us consider a single closed segment C: since $\bar{\gamma}_C$ is the non-decreasing transport plan between f_C^+ and f_C^-, it is clear by Theorem 4.9 that

$$\int_{\mathbb{R}^N \times \mathbb{R}^N} \varphi(y - x)\, d\bar{\gamma}_C(x, y) \leq \int_{\mathbb{R}^N \times \mathbb{R}^N} \varphi(y - x)\, d\gamma_C(x, y),$$

thus property (c) follows from (7.2) and (7.3).

Step V. *General case.*

In this last step we remove hypothesis (7.1). Recalling that $\mathbb{R}^N \setminus \big(\mathrm{ET}_\Gamma \cup \mathrm{F}_\Gamma \cup \mathrm{D}_\Gamma\big)$ is f^+-negligible thanks to Remark 4.2 and since D_Γ is f^+-negligible by hypothesis, we can write $\gamma = \gamma_1 + \gamma_2$ where

$$\gamma_1 := \gamma \llcorner \big(\pi_1(\mathrm{ET}_\Gamma) \times \mathbb{R}^N\big), \qquad \gamma_2 := \gamma \llcorner \Big(\pi_1(\mathrm{F}_\Gamma) \times \mathbb{R}^N\Big).$$

We can then apply the first four steps to γ_1, for which hypothesis (7.1) easily holds with $\Gamma_1 = \Gamma \setminus \Delta$, since hypotheses (i)-(iv) are easily true for Γ_1 as well as for Γ. So we find a transport plan $\bar{\gamma}_1$ fulfilling properties (a), (b) and (c) with γ_1 in place of γ. Concerning γ_2, it is concentrated in the diagonal –which is the graph of the identity function– then the measure $\bar{\gamma}_2 = \gamma_2$ fulfills properties (a), (b) and (c) with γ_2 in place of γ. Finally, the thesis immediately follows taking $\bar{\gamma} = \bar{\gamma}_1 + \bar{\gamma}_2$, since $\pi_1\gamma_1$ and $\pi_1\gamma_2$ are essentially disjoint with respect to f^+. $\square$

It is important to point out some facts, now:

Remark 7.2. *Notice that the in the preceding abstract result, one does not ask finiteness of the first order moments; however, usually we will make this assumptions when we will use this Theorem, in order to show the validity of the hypotheses (i)-(iv).*

Remark 7.3. *The construction of the preceding Theorem does not depend on the cost function c; in fact, the function c was not even mentioned in the statement of the Theorem, nor in the proof.*

Remark 7.4. *In the preceding Theorem, if in particular $c(x, y) = h(y - x)$ for some convex function $h : \mathbb{R}^N \to \mathbb{R}$, then $\mathfrak{C}(t) = \mathfrak{C}(\bar{\gamma}) \leq \mathfrak{C}(\gamma)$: this is an immediate application of property (c) with $\varphi = h$.*

However, inequality $\mathfrak{C}(\bar{\gamma}) \leq \mathfrak{C}(\gamma)$ is not true in general, and in particular it may happen that γ is an optimal transport plan with respect to c while $\bar{\gamma}$ is not optimal. For example, this is the case when $c(x, y) = h(y - x)$ and h is concave: this possibility is not meaningless, since the costs related to real "transport" situations are usually concave. Anyway, note that Theorem 7.1 and Remark 7.4 are still true replacing everywhere the word "convex" with "concave": in fact, one can make exactly the same construction replacing the unique non-decreasing transport map with the unique non-increasing one, that we already mentioned in Section 4.2.

Remark 7.5. *What noticed in Remark 7.4 gives an idea of the main application of Theorem 7.1 we have in mind: we would like to prove existence of an optimal transport map in the case when $c(x, y) = \|y - x\|$ is a norm in $\mathbb{R}^N$, or at least a strictly convex norm; in fact, this method applies, as we will see in next chapters, to all the previously known existence results with costs given by a norm, and it works also for new situations.*

The idea is to take an optimal measure γ and to apply Theorem 7.1: it is clear that the measure $\bar{\gamma}$ given by the Theorem is an optimal transport map; in particular, if we take γ being optimal for (4.18), it comes out also that $\bar{\gamma} = \gamma$. Therefore, we are able to find *an optimal transport map in the case when c comes from a norm if we are able to apply Theorem 7.1: thus, what is needed is to prove that the hypotheses (i)-(iv) hold.*

However, the case of the norms seems to be the only one when the preceding Theorem can be applied with some power: in fact, if c is not given by a norm there is no reason for the no-crossing condition to hold. In particular, for the quadratic case the hypothesis (ii) would be too strong, since we know that it suffices $f^+(S) = 0$ *whenever* $\mathscr{H}^{N-1}(S) < +\infty$*; on the other hand, hypothesis (i) usually does not hold –for example, it does not hold for the situation considered in Section 6.3.*

Remark 7.6. *As we just pointed out, an existence Theorem for the case of a generic norm would be proved by Theorem 7.1 if we could show that the hypotheses (i)-(iv) hold for a set* Γ *with the same properties as in Lemma 4.12. In particular, the condition (i) holds by Lemma 4.12, the condition (ii) has clearly to be taken as an hypothesis, and we know that it is almost sharp also for the Euclidean norm. The property (iii) is always true if* $\|\cdot\|$ *is strictly convex, as we will easily notice in Remark 8.5: then the main problem is to obtain a particular set* Γ *having also the countable Lipschitz property.*

Chapter 8

Quasi-euclidean norms and convergence of transport maps

In this chapter we introduce the quasi-euclidean norms and we prove an existence result of optimal transport maps for the case when $c(x,y) = \|y-x\|$ is given by a quasi-euclidean norm. Then we will give a short presentation of some tools concerning the $\Gamma-$convergence and the asymptotic developments; with them, in the last section of this chapter we will prove the convergence of some approximating optimal transport plans to a unique particular optimal transport map, which turns out to be the unique element of $\Pi'''(f^+, f^-)$. This result is finally seen to solve a convergence problem left open in [24].

8.1 Existence for the quasi-euclidean norms

In this section we will use Theorem 7.1 to generalize the well known existence result for the Euclidean norm to the case of the "quasi-euclidean" norms, as it has been already done in [2] –as we already mentioned in Section 6.5, an existence result for the quasi-euclidean norms were found also in [24]. First of all we give the

Definition 8.1. *The norm* $\|\cdot\|$ *in* $\mathbb{R}^N$ *is said to be a* quasi-euclidean norm *if* $\|\cdot\|^2 \in \mathrm{C}^{1,1}$ *and there are two constants* $c_2 \geq c_1 > 0$ *such that*

$$c_1 \leq \frac{\partial^2}{\partial \xi^2} \|\cdot\|^2 \leq c_2 \qquad \forall\, \xi \in \mathbb{S}^{N-1}. \tag{8.1}$$

Throughout this section we assume then that $X = Y = \mathbb{R}^N$ and that $c(x,y) = \|y-x\|$ for some quasi-euclidean norm $\|\cdot\|$ on $\mathbb{R}^N$; the existence and uniqueness result we want to prove is the

Theorem 8.2. *If f^+ and f^- have finite first order moments (or equivalently if c is slow $-(f^+, f^-)-$ increasing, see Remark 3.21) and $f^+ \ll \mathscr{L}$, then there is an optimal transport map t from f^+ to f^-.*

The proof of this result essentially relies on the two following technical claims; we point out that similar statements were proved in a different way in [24].

Lemma 8.3. *Given a function $u : \mathbb{R}^N \to \mathbb{R}$ which is $1-$Lipschitz with respect to $\|\cdot\|$, the countable Lipschitz property with respect to $\mathscr{L}$ holds for $\Gamma = \partial^c u$.*

Proof. Let us define $S_{\|\cdot\|}$ and $S^*_{\|\cdot\|}$ the unit spheres in $(\mathbb{R}^N, \|\cdot\|)$ and $(\mathbb{R}^N, \|\cdot\|)^*$ respectively; let us moreover consider $x \in \mathrm{ET}_\Gamma$ such that u is differentiable at x; since u is $1-$Lipschitz, $\nabla u(x) \in S^*_{\|\cdot\|}$ and $(\nabla u(x))^*$ is the unique element $v \in S_{\|\cdot\|}$ such that $\nabla u(x) \cdot v = 1$, denoting by $v \cdot w$ the standard scalar product between v and w: the "graphical" relation between $\nabla u(x)$ and $(\nabla u(x))^*$ is illustrated in the left of Figure 8.1. By definition, $(\nabla u(x))^*$ is the direction of maximal slope of u, thus $-\tau_\Gamma(x)$ is the vector of unitary *Euclidean* norm in the same direction of $(\nabla u(x))^*$ –it is $-\tau_\Gamma(x)$ because u is decreasing along the transport rays.

We assume for a moment that $x \to \nabla u(x)$ is defined and Lipschitz; then, to prove that $x \to \tau_\Gamma(x)$ is Lipschitz it suffices to show that the function $S^*_{\|\cdot\|} \ni v \to v^*/|v^*| \in \mathbb{S}^{N-1}$ is Lipschitz; on the other hand, the functions $S_{\|\cdot\|} \ni v \to v/|v| \in \mathbb{S}^{N-1}$ and $S^*_{\|\cdot\|} \ni v \to v/|v| \in \mathbb{S}^{N-1}$ are easily bilipschitz because $\mathbb{R}^N$ is finite dimensional: then it suffices to show that the inverse of $\psi : S_{\|\cdot\|} \to \mathbb{S}^{N-1}$ is Lipschitz, where $\psi(v)$ is the outer normal of $S_{\|\cdot\|}$ at v. To do that, consider the right of Figure 8.1, where v_1 and v_2 belong to $S_{\|\cdot\|}$, $\eta = \psi(v_1)$ and ξ is the versor in the direction $v_2 - v_1$: by the convexity of the norm $\|\cdot\|$ it follows

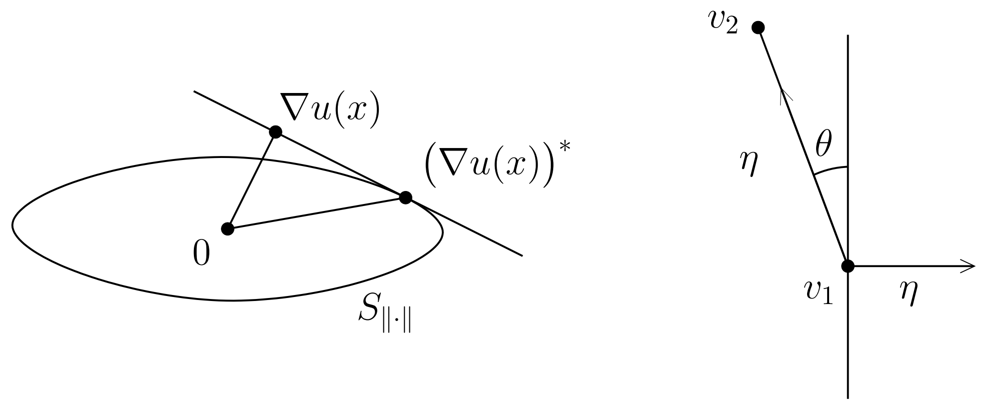


Figure 8.1: Situation in Lemma 8.3

$$|\psi(v_2) - \psi(v_1)| \geq \theta; \tag{8.2}$$

(by example, notice that $|\psi(v_2) - \psi(v_1)| = 2\theta$ if $\|\cdot\|$ is the Euclidean norm). On the other hand, on the segment $[\![v_1, v_2]\!]$ the function $\|v\|^2$ is strictly convex with second derivative greater than c_1 by (8.1); then, since $\|v_1\|^2 = \|v_2\|^2 = 1$,

$$\frac{\partial(\|v\|^2)}{\partial\xi}(v_1)\,|v_2 - v_1| + \frac{c_1}{2}\,|v_2 - v_1|^2 \leq 0 \Longrightarrow \frac{\partial(\|v\|^2)}{\partial\xi}(v_1) \leq -\frac{c_1}{2}\,|v_2 - v_1|. \tag{8.3}$$

Moreover, since η is the outer normal of $S_{\|\cdot\|}$ at v_1 one has

$$\frac{\partial(\|v\|^2)}{\partial\xi}(v_1) = -\frac{\partial(\|v\|^2)}{\partial\eta}(v_1)\,\operatorname{sen}\theta; \tag{8.4}$$

now take any $\lambda \in \mathbb{R}$: again by (8.1) on the segment $[\![v_1, v_1 + \lambda\eta]\!]$, similarly to (8.3) one finds

$$0 \leq \|v_1 + \lambda\eta\|^2 \leq 1 + \lambda\,\frac{\partial(\|v\|^2)}{\partial\eta}(v_1) + \lambda^2 c_2/2:$$

since this is true for any $\lambda \in \mathbb{R}$, one has

$$\frac{\partial(\|v\|^2)}{\partial\eta}(v_1) \leq \sqrt{2c_2}. \tag{8.5}$$

Finally, putting together (8.3), (8.4), (8.5), (8.2) and the fact that $\theta \geq \operatorname{sen}\theta$ one finds

$$|v_2 - v_1| \leq \frac{\sqrt{8c_2}}{c_1}\,|\psi(v_2) - \psi(v_1)|$$

and then the claim that the inverse of ψ is Lipschitz is proved.

We have then proved that τ_Γ is Lipschitz if so is $x \to \nabla u(x)$, and recall that we need to show that τ_Γ is countably Lipschitz with respect to $\mathscr{L}$. Then, defining

$$C = C_{a,b,\xi,n} := \big\{z \in \mathbb{R}^N :\ z \in]\!]x, y[\![,\ (x, y) \in \Gamma,\ z\cdot\xi \leq b < a \leq y\cdot\xi,\ |z| \leq n\big\}$$

where a, $b \in \mathbb{R}$, $n \in \mathbb{N}$ and $\xi \in \mathbb{S}^{N-1}$ are given, the thesis will follow if we show that for any $\varepsilon > 0$ there is a set C_ε such that $C_\varepsilon \ni x \to \nabla u(x) \in \mathbb{R}^N$ is defined and Lipschitz and $\mathscr{L}(C \setminus C_\varepsilon) < \varepsilon$. Since this property is known to be true for BV functions (see for instance Theorem 5.34 in [3]), it will be sufficient to prove that $\nabla u = w$ $\mathscr{L}$−a.e. on C for some BV_{loc} function w. To do that, set

$$\tilde u(x) := \inf\{u(y) + \|y - x\| :\ y\cdot\xi \geq a\}:$$

thanks to (8.1) and the fact that $b < a$, for K sufficiently large $x \to u(y) + \|y - x\| - K|x|^2$ is concave in $\{x :\ x\cdot\xi \leq b\} \supseteq C$ for any y such that $y\cdot\xi \geq a$, then $\tilde u - K|x|^2$ is concave on $\{x :\ x\cdot\xi \leq b\}$. Moreover since $\Gamma = \partial^c u$ it is clear that $u = \tilde u$ on C (while $u \leq \tilde u$

in general); then $\mathscr{L}$ −a.e. on C it is $\nabla u = \nabla \tilde{u}$ (recall that Lipschitz functions admit gradient $\mathscr{L}$ −a.e. and since $\mathbb{R}^N$ is finite dimensional 1−Lipschitz functions with respect to $\|\cdot\|$ are Lipschitz); finally, it is well known that gradients of concave functions are BV_{loc}, thus $\nabla\big(\tilde{u} - K|x|^2\big)$ is BV_{loc} and so is $\nabla\tilde{u}$. □

Lemma 8.4. *Given a function $u : \mathbb{R}^N \to \mathbb{R}$ which is 1−Lipschitz with respect to $\|\cdot\|$, setting $\Gamma = \partial^c u$ the no-crossing condition (4.1) hold and $\mathscr{L}(\mathrm{D}_\Gamma) = 0$.*

Proof. The validity of the no-crossing condition is always true for $\Gamma = \partial^c u$ if u is 1−Lipschitz with respect to $c(x, y) = \|y - x\|$ and $\|\cdot\|$ is a strictly convex norm: this can be obtained with the same argument of Lemma 3.18. Moreover we claim that a quasi-euclidean norm is strictly convex: in fact, taking any two points x, y in the unit sphere, the function $\|v\|^2$ is strictly convex in the segment $[\![x, y]\!]$ by the lower bound in (8.1), then since $\|x\|^2 = \|y\|^2 = 1$ it follows that $\|v\|^2 < 1$, thus $\|v\| < 1$, for any $v \in]\!]x, y[\![$. Then the open segment $]\!]x, y[\![$ is contained in the open unit ball, and this means that the unit ball is strictly convex: the claim follows since the strict convexity of a norm is equivalent to the strict convexity of its unit ball. Then (i) is proved.

Concerning the fact that D_Γ is $\mathscr{L}$ −negligible, it is really easy: in fact, it is immediate to notice that $\mathrm{D}_\Gamma \subseteq \Sigma$, where Σ is the set of the points where u is not differentiable; moreover, this set is $\mathscr{L}$ −negligible since u is Lipschitz with respect to $\|\cdot\|$, and so with respect to the Euclidean norm $|\cdot|$, since $\mathbb{R}^N$ is finite-dimensional. □

Remark 8.5. *It is important to notice that, with the same proof as in last Lemma, $\mathscr{L}(\mathrm{D}_\Gamma) = 0$ is always true with the usual Γ whenever $\|\cdot\|$ is any strictly convex norm, since the Kantorovich potential is Lipschitz also with respect to $|\cdot|$ and thus it is differentiable a.e..*

Remark 8.6. *In the hypotheses of preceding Lemma, it is true also that E^l_Γ is $\mathscr{L}$ −negligible; however, the proof of this fact (see [2]) is not completely trivial, and we do not need this stronger hypothesis (recall the discussion after Remark 4.2, and see also Remark 10.3).*

The proof of Theorem 8.2 is now an easy consequence:

Proof. *(of Theorem 8.2):* Since c is *slow* $-(f^+, f^-)-$ *increasing*, according to Corollary 3.27 there is a Kantorovich potential u; thus, given an optimal transport plan γ, it is concentrated in $\Gamma = \partial^c u$ by Theorem 3.28. We can then apply Theorem 7.1, whose hypotheses hold thanks to Lemmas 8.3 and 8.4 (concerning property (iv), recall that $f^+ \ll \mathscr{L}$), finding a transport plan $\bar{\gamma}$ induced by a transport map t; moreover using property (c) with $\phi(v) = \|v\|$ we find that $\mathfrak{C}(\bar{\gamma}) \leq \mathfrak{C}(\gamma)$, and since γ is optimal so is $\bar{\gamma}$. It follows that t is an optimal transport map. $\square$

We can now derive a very useful corollary, making use also of Lemma 8.4 and Lemma 4.15, namely the

Corollary 8.7. *If f^+ and f^- have finite first order moments and $f^+ \ll \mathscr{L}$, then there exists an optimal transport map $t_{\|\cdot\|}$ such that for any strictly convex function α with the property that* $\inf(4.18) < +\infty$ *the set $\Pi'''(f^+, f^-)$ consist of the unique element $\gamma_{\|\cdot\|}$, the transport plan induced by $t_{\|\cdot\|}$. If moreover $f^- \ll \mathscr{L}$, there exists also an optimal transport map $\tilde{t}_{\|\cdot\|}$ from f^- to f^+ and $\gamma_{\|\cdot\|}$ is also induced by $\tilde{t}_{\|\cdot\|}$; finally, $t_{\|\cdot\|}$ and $\tilde{t}_{\|\cdot\|}$ are inverse.*

Proof. We first note that $\Pi''(f^+, f^-) = \Pi'(f^+, f^-)$ and $\Gamma = \Gamma'$ by Lemma 4.16 (recall that a quasi-euclidean norm is strictly convex as pointed out in the proof of Lemma 8.4); then the elements of $\Pi'''(f^+, f^-)$ are all the measure minimizing, among the optimal transport plan with respect to c, the cost $\mathfrak{C}_{conv}$ related to the cost function $c_{conv}(x, y) = \alpha(y - x)$, where $\alpha(v) = \sqrt{1 + |v|^2}$ Moreover $\mathrm{D}_\Gamma = \mathrm{D}'_\Gamma$ implies, by Lemma 4.15, that $\Pi'''(f^+, f^-)$ consists of a unique element, which we denote by $\gamma_{\|\cdot\|}$.

We apply then Theorem 7.1 to $\gamma_{\|\cdot\|}$ as we did in the proof of Theorem 8.2, finding an optimal transport plan $\bar{\gamma}$ which is induced by the optimal transport map $t_{\|\cdot\|}$. According to property (c) in Theorem 7.1, one derives that $\gamma_{\|\cdot\|} = \bar{\gamma}$, then the thesis follows. If $f^- \ll \mathscr{L}$, the same argument gives the existence of the optimal transport map $\tilde{t}_{\|\cdot\|}$ from f^- to f^+, and $\gamma_{\|\cdot\|}$ is induced by $\tilde{t}_{\|\cdot\|}$. This ensures also that $t_{\|\cdot\|}$ and $\tilde{t}_{\|\cdot\|}$ are inverse, i.e. $\tilde{t}_{\|\cdot\|} \circ t_{\|\cdot\|} = \mathrm{Id}$ (f^+-almost everywhere) and $t_{\|\cdot\|} \circ \tilde{t}_{\|\cdot\|} = \mathrm{Id}$ (f^-−almost everywhere). $\square$

8.2 $\Gamma-$convergence and asymptotic developments

In this section we briefly recall some things about the $\Gamma-$convergence (for a precise reference on this subject, one can refer to [28]) and about the asymptotic developments (see [5, 6]). As we said in Section 4.3, this leads to an equivalent way to define Π'' and Π'''.

Given a compact space H and a sequence of l.s.c. functions $g_n : H \to \overline{R}$, one sais that g_n $\Gamma-$converges to $g : H \to \overline{R}$ if

(*i*) $g(x) \leq \liminf g_n(x_n)$ for any sequence $x_n \to x$; (8.6i)

(*ii*) for any $x \in H$ there is a sequence $x_n \to x$ such that $g_n(x_n) \to g(x)$; (8.6ii)

it is clear that, if such a function g exists, it is also l.s.c.; the first condition in (8.6) is usually referred to as *liminf inequality*, while the second one is said *limsup inequality* since it can also be rewritten as the existence of a sequence such that $\limsup g_n(x_n) \le g(x)$. One of the most important facts about the $\Gamma-$convergence is that it is induced by a metric that makes the space of the l.s.c. functions compact: then, given any sequence of l.s.c. functions, it is always true that a subsequence $\Gamma-$convergences somewhere. A useful application of the $\Gamma-$convergence is when H is a closed and bounded subset of the positive measures $\mathcal{M}^+(W)$ and W is a separable space, since it is well known that in this case the weak* convergence is induced by a compact metric (see for example [22]). It is easily noticed that if g_n $\Gamma-$converges to g then $m_n := \min g_n$ converges to $m := \min g$; moreover, any limit point of minimizers of g_n is a minimizer of g, thus

$$\limsup \operatorname{Argmin}(g_n) \subseteq \operatorname{Argmin} g; \tag{8.7}$$

one of the interesting facts about the $\Gamma-$convergence is the possibility to select particular minimizers, since the last inclusion can be strict: for example, if $H = [0,1]$ and $g_n(x) = x/n$, then $g_n \xrightarrow{\Gamma} g = 0$ but the unique element of the set in the left in (8.7) is $x = 0$, while the set in the right is the whole segment H. An idea to give a better estimate of $\limsup \operatorname{Argmin}(g_n)$ than in (8.7) is the following result, due to Anzellotti and Baldo:

Lemma 8.8 (Anzellotti-Baldo). *Given a sequence $\varepsilon_n \searrow 0$ and defining $g'_n = (g_n - m)/\varepsilon_n$, if $g'_n \xrightarrow{\Gamma} g'$ then $g' \equiv +\infty$ out of $\operatorname{Argmin} g$ and, if $g' \not\equiv +\infty$,*

$$\limsup \operatorname{Argmin}(g_n) \subseteq \operatorname{Argmin} g' \subseteq \operatorname{Argmin} g. \tag{8.8}$$

Proof. If $g'(x) < +\infty$, by condition (ii) there is a sequence x_n such that $g'_n(x_n)$ converges to $g'(x)$, thus $g_n(x_n) \to m$ and by (i) $x \in \operatorname{Argmin} g$, so the first assertion is proved. If $g' \not\equiv +\infty$, this implies the second inclusion in (8.8), while the first one is an immediate consequence of (8.7) since clearly $\operatorname{Argmin} g'_n = \operatorname{Argmin} g_n$. $\square$

This method can be used to give another interpretation of the variational problems we presented in Section 4.3: in fact, let us define

$$\mathfrak{C}_n(\gamma) := \mathfrak{C}(\gamma) + \mathfrak{C}_{\mathrm{eucl}}(\gamma)/n.$$

It is clear that $\mathfrak{C}_n \xrightarrow{\Gamma} \mathfrak{C}$; moreover, if we take $\varepsilon_n = 1/n$, it is easily verified that $\mathfrak{C}'_n \xrightarrow{\Gamma} \mathfrak{C}'$, where

$$\mathfrak{C}'(\gamma) = \begin{cases} \mathfrak{C}_{\mathrm{eucl}}(\gamma) & \text{if } \gamma \in \Pi'(f^+, f^-) \\ +\infty & \text{otherwise,} \end{cases}$$

thus this asymptotic development provides another interpretation of (4.8), since by Lemma 8.8 any weak* limit of minimizers of $\mathfrak{C}_n$ is a solution of (4.8). The same argument can be used to provide another interpretation also to (4.18) since it is similarly verified, applying twice the argument of Anzellotti and Baldo, that solutions of (4.18) can be found as weak* limits of minimizers of

$$\widetilde{\mathfrak{C}}_n := \mathfrak{C}(\gamma) + \mathfrak{C}_{\mathrm{eucl}}(\gamma)/n + \mathfrak{C}_{\mathrm{conv}}(\gamma)/n^2.$$

8.3 Convergence of the approximate transport maps

In this section we are given a quasi-euclidean norm $\|\cdot\|$ and we approximate the cost function $c(x,y) = \|y-x\|$ with some costs $c_\varepsilon \to c$; we then take optimal transport maps t_ε with respect to c_ε and we show the convergence in measure of t_ε to $t_{\|\cdot\|}$.

We consider nondecreasing and strictly convex functions $\alpha_\varepsilon : \mathbb{R}^N \to \mathbb{R}^+$ for $\varepsilon > 0$ with the two following properties:

(i) $\varepsilon \to \alpha_\varepsilon$ is convex and $\alpha_\varepsilon(v) \to \|v\|$ as $\varepsilon \to 0$;

(ii) the right derivative
$$\alpha(v) := \lim_{\varepsilon \to 0} \frac{\alpha_\varepsilon(v) - \|v\|}{\varepsilon}$$
is a well defined strictly convex function.

We define then the approximated cost functions $c_\varepsilon(x,y) = \alpha_\varepsilon(y-x)$ and we want to study the associated mass transport problems. According to Theorem 5.1 of Gangbo and McCann we know that there is a unique optimal transport plan γ_ε related to c_ε, which is induced by a transport map t_ε if the condition of Theorem 5.1 hold (in particular, this is true if f^+ does not charge the $(N-1)-$dimensional sets and $\alpha_\varepsilon(v) = \alpha_\varepsilon(|v|)$ with $\alpha_\varepsilon(0) = 0$ and $\alpha'_\varepsilon \to +\infty$).

We recall, from Section 6.5, that this approximation is for example what is done in [24] with $\alpha_\varepsilon(v) = \|v\|^{1+\varepsilon}$, where an optimal transport map is found passing to the limit for some sequence $\varepsilon_i \to 0$; in this case, as used in [2], it is $\alpha(v) = |v| \ln(|v|)$. What we have in mind is to prove the following

Theorem 8.9. *Assume that f^+ and f^- have finite first order moments and that $\inf_\varepsilon \mathfrak{C}_\varepsilon(\gamma) < +\infty$ for any transport plan γ optimal for $\mathfrak{C}$. Then, taking any optimal transport plan γ_ε related to the cost function c_ε, $\gamma_\varepsilon \xrightarrow{*} \gamma_{\|\cdot\|}$ as $\varepsilon \to 0$. If moreover there are optimal transport maps t_ε (for example, if Theorem 5.1 can be applied), then t_ε converge in measure to $t_{\|\cdot\|}$.*

Proof. We define $\mathfrak{C}_\varepsilon$ the cost associated to the cost function $c_\varepsilon(x,y) = \alpha_\varepsilon(y-x)$ and, according to the method of Anzellotti and Baldo, we also define

$$\mathfrak{C}'_\varepsilon := \frac{\mathfrak{C}_\varepsilon - \inf \mathfrak{C}}{\varepsilon}.$$

Step I. $\Gamma-$*convergence of* $\mathfrak{C}_\varepsilon$.

We prove here that $\mathfrak{C}_\varepsilon \xrightarrow{\Gamma} \mathfrak{C}$ in $\Pi(f^+, f^-)$, which we know to be a convex, weak* closed, bounded and non-empty subset of $\mathcal{M}^+(\mathbb{R}^N \times \mathbb{R}^N)$, thus metrizable and compact with the weak* topology. Concerning the lim inf inequality in (8.6), it easily comes from the convexity hypothesis (i) about α_ε; in fact one can evaluate, being given $\gamma_\varepsilon \xrightarrow{*} \gamma$,

$$\begin{aligned}\liminf_{\varepsilon\to 0} \mathfrak{C}_\varepsilon(\gamma_\varepsilon) &\geq \liminf_{\varepsilon\to 0} \iint \alpha_\varepsilon(y-x)\, d\gamma_\varepsilon(x,y)\\ &\geq \liminf \iint \|y-x\| + \varepsilon\alpha(y-x)\, d\gamma_\varepsilon(x,y)\\ &\geq \liminf \iint \|y-x\|\, d\gamma_\varepsilon(x,y) \geq \mathfrak{C}(\gamma),\end{aligned}$$

recalling that $\iint \alpha\, d\gamma_\varepsilon$ is bounded from below according to Remark 4.14; therefore, the lim inf inequality follows. Concerning the lim sup one, if $\mathfrak{C}(\gamma) < +\infty$ (otherwise the inequality is clearly true) it suffices to consider $\gamma_\varepsilon = \gamma$ for each ε: by hypothesis, there is $\bar\varepsilon$ such that $\mathfrak{C}_{\bar\varepsilon}(\gamma) < +\infty$; moreover, the convexity hypothesis (i) about α_ε gives

$$\|y-x\| \geq \alpha_\varepsilon(y-x) - \varepsilon\, \frac{\alpha_{\bar\varepsilon}(y-x) - \|y-x\|}{\bar\varepsilon},$$

and thus

$$\mathfrak{C}(\gamma) \geq \mathfrak{C}_\varepsilon(\gamma) - \frac{\varepsilon}{\bar\varepsilon}\big(\mathfrak{C}_{\bar\varepsilon}(\gamma) - \mathfrak{C}(\gamma)\big),$$

from which the lim sup inequality immediately follows.

Step II. $\Gamma-$*convergence of* $\mathfrak{C}'_\varepsilon$.

We prove here that $\mathfrak{C}'_\varepsilon \xrightarrow{\Gamma} \mathfrak{C}'$, where

$$\mathfrak{C}'(\gamma) = \begin{cases} \iint \alpha(y-x)\, d\gamma(x,y) & \text{if } \gamma \text{ minimizes } \mathfrak{C},\\ +\infty & \text{otherwise.}\end{cases}$$

Recall that $\mathfrak{C}'(\gamma)$ is well defined and in particular $\iint \alpha\, d\gamma > -\infty$ for any $\gamma \in \Pi(f^+, f^-)$, again according to Remark 4.14.

Concerning the lim inf inequality, if $\gamma_\varepsilon \xrightarrow{*} \gamma$ and $\liminf \mathfrak{C}'_\varepsilon(\gamma_\varepsilon) < +\infty$, then in particular $\liminf \mathfrak{C}_\varepsilon(\gamma_\varepsilon) = \inf \mathfrak{C}$, thus γ minimizes $\mathfrak{C}$; moreover, again the convexity of $\varepsilon \to \alpha_\varepsilon$ gives $\alpha_\varepsilon(v) \geq \|v\| + \varepsilon\alpha(v)$, and it follows

$$\mathfrak{C}_\varepsilon(\gamma_\varepsilon) \geq \mathfrak{C}(\gamma_\varepsilon) + \varepsilon \iint \alpha(y-x)\, d\gamma_\varepsilon(x,y);$$

then

$$\begin{aligned}\liminf_{\varepsilon\to 0} \mathfrak{C}'_\varepsilon(\gamma_\varepsilon) &= \liminf \frac{\mathfrak{C}_\varepsilon(\gamma_\varepsilon) - \inf \mathfrak{C}}{\varepsilon} \geq \liminf \frac{\mathfrak{C}_\varepsilon(\gamma_\varepsilon) - \mathfrak{C}(\gamma_\varepsilon)}{\varepsilon}\\ &\geq \liminf \iint \alpha(y-x)\, d\gamma_\varepsilon(x,y) \geq \mathfrak{C}'(\gamma),\end{aligned}$$

since we noted that γ minimizes $\mathfrak{C}$: then the lim inf inequality holds.

Concerning the lim sup one, it is again sufficient to consider $\gamma_\varepsilon = \gamma$ for any ε; then, if $\mathfrak{C}(\gamma) = \inf \mathfrak{C}$ (otherwise $\mathfrak{C}'(\gamma) = +\infty$ and there is nothing to prove)

$$\mathfrak{C}'_\varepsilon(\gamma) = \frac{\mathfrak{C}_\varepsilon(\gamma) - \inf \mathfrak{C}}{\varepsilon} = \iint \frac{\alpha_\varepsilon(y-x) - \|y-x\|}{\varepsilon}\, d\gamma(x,y);$$

but the integrand in last equation converges pointwise to $\alpha(y-x)$ and by the convexity of $\varepsilon \to \alpha_\varepsilon$ it is non-increasing when $\varepsilon \searrow 0$; thus it follows that $\mathfrak{C}'_\varepsilon(\gamma)$ converges to $\mathfrak{C}'(\gamma)$, then also the lim sup inequality is proved and the claimed $\Gamma-$convergence follows. Note that another consequence of the monotonicity argument just used is the fact that $\mathfrak{C}'(\gamma) < +\infty$ for any optimal transport plan γ (recall $\mathfrak{C}_\varepsilon(\gamma) < +\infty$ for some $\varepsilon > 0$).

Step III. *First conclusion.*
Recalling Lemma 8.8 and the first two steps, any weak* limit γ of measures γ_ε optimal for $\mathfrak{C}_\varepsilon$ is seen to be a minimizer of the cost $\mathfrak{C}'$ between the optimal transport plans; moreover, recalling (4.18), Lemma 4.16 and Lemma 8.4, it is clear that $\gamma \in \Pi'''(f^+, f^-)$. According to the remark at the end of last step, it is clear that $\inf(4.18) < +\infty$ with the function α defined in this section: thus, by Corollary 8.7 it follows $\gamma = \gamma_{\|\cdot\|}$. This implies that in fact there is full convergence of the measures γ_ε to $\gamma_{\|\cdot\|}$.

Step IV. *Convergence in measure.*
Let us consider now the case when there are optimal transport maps t_ε, and set each γ_ε to be induced by the corresponding t_ε. This means that $(\mathrm{Id}, t_\varepsilon)_\# f^+ \overset{*}{\rightharpoonup} (\mathrm{Id}, t_{\|\cdot\|})_\# f^+$; take then, by Lusin Theorem, a compact set K such that $t_{\|\cdot\|} \llcorner K$ is continuous and $f^+({}^cK) \leq \delta$; take also a continuous positive function $\varphi : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ such that whenever $x \in K$ it is $\varphi(x,y) = 1$ if $|y - t(x)| \geq \delta$ and $\varphi(x,y) = 0$ if $y = t(x)$. This allows to evaluate

$$\begin{aligned}\limsup_{\varepsilon\to 0} f^+\big(\{|t_\varepsilon(x) - t(x)| \geq \delta\}\big) &\leq \delta + \limsup f^+\big(K \cap \{|t_\varepsilon(x) - t(x)| \geq \delta\}\big)\\ &\leq \delta + \limsup \iint \varphi(x,y)\, d\gamma_\varepsilon(x,y) = \delta + \iint \varphi(x,y)\, d\gamma(x,y) = \delta :\end{aligned}$$

this implies the claimed convergence in measure of t_ε to $t_{\|\cdot\|}$. $\square$

Remark 8.10. *In the case considered by Caffarelli, Feldman and McCann ([24], see Section 6.5), it is used the approximation with $\alpha_\varepsilon(v) = \|v\|^{1+\varepsilon}$ for some quasi-euclidean norm $\|\cdot\|$; it is immediately noticed that Theorem 5.1 applies, and that also the hypotheses of this section hold with $\alpha(v) = \|v\|\ln(\|v\|)$. It was left open the problem whether or not there is convergence of the maps t_ε to t, possibly for some subsequence $\varepsilon_n \searrow 0$: the last Theorem ensures that in fact the maps t_ε have full convergence in measure to $t_{\|\cdot\|}$; in particular, the map found by Caffarelli, Feldman and McCann is $t_{\|\cdot\|}$.*

Chapter 9

The crystalline case

In this Chapter we will prove an existence result, taken from [4], which considers the case when the cost function $c(x, y)$ is a crystalline norm, that is a norm whose unit ball is delimited by finite many hyperplanes. The easiest example of a crystalline norm is the L^∞ norm in $\mathbb{R}^2$, which is $\|(x, y)\| = \max(|x|, |y|)$.

The crystalline norms are interesting in many different fields in mathematics, also for their connection with some physical models, among them the models for studying crystal structures (from which the name comes).

It may seem a bit surprizing that we are able to show the existence result of optimal transport maps with this class of norms: in fact, together with the results of last chapter, we are now able to infer existence for those norms whose unit ball is the most smooth possible (i.e. the quasi-euclidean norms, whose unit balls are C^2-regular and uniformly round) or the least smooth possible (i.e. the crystalline norms, whose unit balls are flat and have sharp corners). Still open is the case of all the other norms, with an "intermediate behavior".

We assume then $X = Y = \mathbb{R}^N$ and that we are given $M \geq N$ vectors $v_1, \ldots, v_M$ spanning the whole space: the norm is defined by

$$\|w\| := \max \{|v_i \cdot w| \ : \ i = 1, \ldots, N\} ; \tag{9.1}$$

recall that, as already said, we keep the notation $|\cdot|$ for the Euclidean norm. Using (9.1), we set then the cost to be

$$c(x, y) := \|y - x\|.$$

Let us give now a few notation. We will write $v_{-i} = -v_i$ for $i = 1, \ldots, M$ and we will call "multiindex" any unordered subset of $\{\pm 1, \ldots, \pm M\}$; moreover, given a multiindex

$I = \{i_1, \dots, i_m\}$, we will say that the couple $(x,y) \in \mathbb{R}^N \times \mathbb{R}^N$ *is of the type* I if

$$c(x,y) = v_{i_1} \cdot (y-x) = \dots = v_{i_m} \cdot (y-x) > v_j \cdot (y-x) \quad \forall j \notin I.$$

By definition, each couple (x,y) with $x \neq y$ is of a unique type, so that the types induce a partition of $\left(\mathbb{R}^N \times \mathbb{R}^N\right) \setminus \Delta$ in finite disjoint subsets. We will call *simple types* the ones corresponding to a multiindex with only an index.

Remark 9.1. *It is clear that if (x,y) is of the simple type $I = \{i\}$, the same is true for a neighborhood of (x,y). Indeed, for each $j \neq i$ it holds $v_i \cdot (y-x) > v_j \cdot (y-x)$, and then the same inequalities continue to hold in a neighborhood of (x,y).*

We state and prove now the result of this chapter:

Theorem 9.2. *If f^+ and f^- have finite first order moments and $f^+ \ll \mathscr{L}$, then there is an optimal transport map, which corresponds to the unique solution of (4.18).*

Proof. Let us fix a solution γ of (4.18) and a Borel set Γ on which γ is concentrated which has the properties stated in Lemma 4.12 –in particular, it has the no-crossing condition (4.1) and the non-inclusion condition (4.2). Setting then

$$\Gamma_I := \Gamma \cap \left\{(x,y) \in \left(\mathbb{R}^N \times \mathbb{R}^N\right) \setminus \Delta : \ (x,y) \text{ is of the type } I\right\},$$

all the sets Γ_I are also Borel. We define for each multiindex I the measure $\gamma_I := \gamma \llcorner \Gamma_I$, which is a transport plan between $f_I^+ = \pi_1 \gamma_I$ and $f_I^- = \pi_2 \gamma_I$ and which, in view of Lemma 1.12, is optimal.

Step I. *γ_I is locally a transport map for $I = \{i\}$.*
Here we prove that, for each simple type $I = \{i\}$, γ_I is locally a transport map, i.e. given $(x,y) \in \Gamma_I$ there exists an open neighborhood U such that $\gamma_I \llcorner U$ is induced by a transport map. Indeed, let $U = Z \times W$ be a product of two open balls such that only points of type I are contained in U (this is possible thanks to Remark 9.1): since by Lemma 1.12 one can immediately derive that minimizing (4.18) is a local property, it follows that $\gamma \llcorner U$ solves (4.18) between the transport plans connecting $\pi_1 \gamma_I \llcorner U$ and $\pi_2 \gamma_I \llcorner U$. Let now η be any such transport plan: then it is concentrated on U and

$$\begin{aligned} C(\eta) &= \iint_U c(z,w)\, d\eta(z,w) = \iint_U v_i \cdot (w-z)\, d\eta(z,w) \\ &= \int_W v_i \cdot w\, d\pi_2 \eta - \int_Z v_i \cdot z\, d\pi_1 \eta = \int_W v_i \cdot w\, d\pi_2 \gamma_I \llcorner U - \int_Z v_i \cdot z\, d\pi_1 \gamma_I \llcorner U : \end{aligned}$$

this shows that the cost is independent of η and then it follows that $\gamma \llcorner U$ is a transport plan minimizing the Euclidean cost among all the plans connecting $\pi_1 \gamma_I \llcorner U$ and $\pi_2 \gamma_I \llcorner U$ and, among all the plans η minimizing the Euclidean cost, minimizes the quantity

$$\iint_{\mathbb{R}^N \times \mathbb{R}^N} \alpha(y-x)\, d\eta(x,y)$$

with $\alpha(v) = \sqrt{1+|v|^2}$. Thus $\gamma_I \llcorner U$ is the unique element of the set $\Pi'''(\pi_1 \gamma_I \llcorner U, \pi_2 \gamma_I \llcorner U)$ made with the Euclidean cost function $\tilde{c}(x,y) = |y-x|$ instead of $c(x,y) = \|y-x\|$: then we can apply Corollary 8.7 and we obtain that $\gamma_I \llcorner U$ is induced by a transport map –note that we could not apply Corollary 8.7 using the cost function c because it is clearly not quasi-euclidean. The remaining of this step is devoted to show that the set $\mathrm{E}^l_{\Gamma_I \cap U}$ of the left-extreme points of $\Gamma_I \cap U$ is f^+-negligible. In fact, there is a Kantorovich potential u for the Monge problem associated to the Euclidean cost function $\tilde{c}(x,y) = |y-x|$ and the measures $\pi_1 \gamma_I \llcorner U$ and $\pi_2 \gamma_I \llcorner U$; so $\gamma_I \llcorner U$ is concentrated both in $\widetilde{\Gamma} = \partial^{\tilde{c}} u$ and in $\Gamma_I \cap U$. The Borel set $\Gamma_I \cap U \setminus \widetilde{\Gamma}$ is then $\gamma_I \llcorner U-$negligible: as a consequence, since $\gamma \llcorner U = \gamma_I \llcorner U$,

$$\gamma\Big(\big(\Gamma_I \cap U\big) \setminus \widetilde{\Gamma}\Big) = 0. \tag{9.2}$$

In addition, according to Corollary 6.16, the countable Lipschitz property with respect to $\mathscr{L}$ holds for $\widetilde{\Gamma}$, then by Lemma 6.14 we can disintegrate $f^+ \llcorner \mathrm{ET}_{\widetilde{\Gamma}}$ with respect to $r_{\widetilde{\Gamma}}$ obtaining

$$f^+ \llcorner \mathrm{ET}_{\widetilde{\Gamma}} = f^+_C \otimes \nu \tag{9.3}$$

and for $\nu-$a.e. C the measure f^+_C is absolutely continuous with respect to $\mathscr{H}^1 \llcorner C$. Consider now the subset Ψ of $\mathrm{E}^l_{\Gamma_I \cap U}$ made by those points x for which there is not $y \neq x$ with $(x,y) \in \widetilde{\Gamma}$: Ψ is f^+-negligible by (9.2). On the other hand, if we call MTR the set of the closures of the maximal transport rays associated to $\widetilde{\Gamma}$, in any $C \in MTR$ there can be at most a point $x \in \mathrm{E}^l_{\Gamma_I \cap U} \setminus \Psi$: therefore, $f^+_C(\mathrm{E}^l_{\Gamma_I \cap U} \setminus \Psi) = 0$ for $\nu-$a.e. C, since $\nu-$a.e. C is in MTR, and as a consequence of (9.3) we derive $f^+ \llcorner \mathrm{ET}_{\widetilde{\Gamma}}(\mathrm{E}^l_{\Gamma_I \cap U} \setminus \Psi) = 0$. Finally, a point $x \in \mathrm{E}^l_{\Gamma_I \cap U} \setminus \Psi$ which is not in $\mathrm{ET}_{\widetilde{\Gamma}}$ must be a doubling point of $\widetilde{\Gamma}$: but $\mathrm{D}_{\widetilde{\Gamma}}$ is $\mathscr{L}-$negligible thanks to Corollary 6.15, and then it is also f^+-negligible and our claim is proved.

Step II. *γ_I is a transport map for $I = \{i\}$.*

Now we prove that, for each single type $I = \{i\}$, γ_I is induced by a transport map and the family of the left-extreme points of Γ_I is f^+-negligible. Using Step I we can cover $\mathbb{R}^N$, up to a γ_I-negligible set, with countably many rectangles $U_j = Z_j \times W_j$ such that $\gamma_I \llcorner U_j$ is induced by a transport t_j; in particular, we take a family of rectangles so large that for any $\varepsilon > 0$ the subfamily of the rectangles U_j such that $\mathrm{diam}(Z_j) \vee \mathrm{diam}(W_j) < \varepsilon$ also covers $\mathbb{R}^N$ up to a γ_I-negligible set.

The fact that γ_I is induced by a transport map will be obtained simply gluing the different t_j if we show that they match f_I^+-a.e.; recall from Step I that, for any j, the set $\mathrm{E}^l_{\Gamma_I \cap U}$ of the left-extreme points of $\Gamma_I \cap U_j$ is f^+-negligible: for simplicity, we denote this set by Θ_j. It is then $f^+(\Theta) = 0$ with $\Theta = \cup\Theta_j$, then γ_I will be shown to be induced by a transport map if we prove that there is only one y such that $(x, y) \in \Gamma_I$ for any $x \in \pi_1(\Gamma_I) \setminus \Theta$.

Assume then that (x, y_1) and (x, y_2) both belong to Γ_I with $y_1 \neq y_2$ and $|y_1 - x| \leq |y_2 - x|$; by construction, we can choose j such that $(x, y_2) \in U_j$ and $(x, y_1) \notin U_j$. We can also find $(\tilde{x}, \tilde{y}) \in \Gamma_I \cap U_j$ such that $x \in]\!]\tilde{x}, \tilde{y}[\![$, unless $x \in \Theta_j \subseteq \Theta$. Then, since the no-crossing condition holds for Γ, it follows that both $[\![x, y_1]\!]$ and $[\![x, y_2]\!]$ are parallel to $[\![\tilde{x}, \tilde{y}]\!]$ and therefore they are parallel. So we have only to exclude the case when x, y_1 and y_2

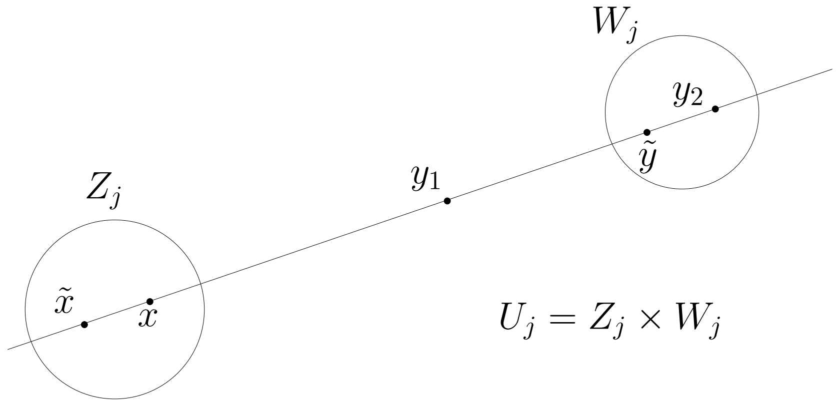


Figure 9.1: Geometric picture for Step II

are aligned. But then (see Figure 9.1) it is also $[\![x, y_1]\!] \subseteq]\!]\tilde{x}, \tilde{y}[\![$, and this is a contradiction with the non-inclusion property.

We have then proved that γ_I is a transport plan; concerning its left-extreme points, we claim that they are contained in the f^+-negligible set Θ. In fact, if x is an extreme point of Γ_I, there is some j such that $x \in \pi_1(U_j)$ and then x is a fortiori an extreme point for $\Gamma_I \cap U_j$, thus $x \in \Theta_j \subseteq \Theta$.

Step III. *γ_I is a transport map for each I.*

Here we generalize Step II to non-simple types: we take any multiindex I and we want to show that the plan γ_I is induced by a transport map t_I and the left-extreme points of Γ_I are f^+-negligible. First of all, we define the plane

$$P_I := \{x : x \cdot (v_i - v_j) = 0 \quad \forall i,\, j \in I\} \tag{9.4}$$

and we set $m := \dim(P_I) \geq 1$, $K := P_I^\perp$ and $\mathrm{P}_I^\omega = P_I + \omega$ for each $\omega \in K$. We project now γ_I on the planes P_I^ω: to do that, we denote by Proj_K the orthogonal projection on K, we define $\varphi : \mathbb{R}^N \times \mathbb{R}^N \to K$ as $\varphi(x, y) := \mathrm{Proj}_K(x)$ and finally we consider the disintegration of γ_I with respect to φ, $\gamma_I = \gamma_\omega \otimes \varphi_\# \gamma_I$. By definition, each measure γ_ω is concentrated on the set $\{(x, y) \in \Gamma_I : \varphi(x, y) = \omega\}$; but if the pair (x, y) is of the type I, by (9.4) it is

$$\varphi(x, y) = \omega \iff \mathrm{Proj}_K(x) = \omega \iff x \in \mathrm{P}_I^\omega \iff y \in \mathrm{P}_I^\omega;$$

since by definition γ_I–a.e. pair (x, y) is of the type I, for $\varphi_\#\gamma_I$–a.e. ω the measure γ_ω is concentrated on $\mathrm{P}_I^\omega \times \mathrm{P}_I^\omega$, and then we can consider it as a probability measure in $\mathrm{P}_I^\omega \times \mathrm{P}_I^\omega$.

We apply now, as in the first step, Lemma 1.13 to infer that, for $\varphi_\#\gamma_I$–a.e. ω, the transport plan γ_ω is a solution of the problem (4.18) on the plane P_I^ω with marginals $\pi_1\gamma_\omega$ and $\pi_2\gamma_\omega$. We want then apply Step II to γ_ω on P_I^ω: to do this, we first note that $\pi_1\gamma_\omega = f_\omega^+$, the disintegration of f_I^+ with respect to Proj_K, thanks to Lemma A.16. Then, since $f^+ \ll \mathscr{L}$, for $\varphi_\#\gamma_I$–a.e. ω it is $\pi_1\gamma_\omega \ll \mathscr{L}_\omega$, where the last measure is the disintegration of $\mathscr{L}$ with respect to Proj_K. But it is clear that $\mathscr{L}_\omega = \mathscr{H}^m \llcorner \mathrm{P}_I^\omega$, then finally we have that for a.e. ω $\pi_1\gamma_\omega$ is absolutely continuous with respect to the Lebesgue measure on P_I^ω.

Observe now that the distance on P_I^ω is clearly induced by the crystalline norm associated to the projections $\tilde{v}_i$ of the vectors v_i on P_I: but the vectors $\tilde{v}_i$ with $i \in I$ are all equal, and then we call $\tilde{v}_I$ that unique vector; in other words, the crystalline norm on P_I^ω is induced by the vectors $\{\tilde{v}_j : j \notin I\} \cup \{\tilde{v}_I\}$, so the multiindex I has been contracted to a single index. Moreover each $(x, y) \in \Gamma_I \cap (\mathrm{P}_I^\omega \times \mathrm{P}_I^\omega)$ is of type $\{I\}$, which is a simple type on P_I^ω: it is then possible to apply Step II to the planes P_I^ω to find that γ_ω is induced by a transport map for $\varphi_\#\gamma_I$–a.e. ω. Since the different planes are disjoint, γ_I is concentrated on some graph, thus according to Lemma 1.6 it is in fact induced by a transport, which will be call t_I.

We finally consider the set of the extreme points of Γ_I: for $\varphi_\#\gamma_I$–a.e. ω, Step II already gives that its intersection with P_I^ω is f_ω^+–negligible, writing $f^+ = f_\omega^+ \otimes \varphi_\# f^+$, and then it follows that this set is f^+–negligible.

Step IV. *γ is a transport map.*
Here we will show that the whole measure γ is induced by a transport map. Since this is true, by Step III, for each γ_I, we begin gluing all the different transport maps t_I: this is possible if the points x with the property that there exist $J \neq K$ and two points y_J and y_K with $(x, y_J) \in \Gamma_J$ and $(x, y_K) \in \Gamma_K$ are f^+–negligible. In fact, if such a point x exists, it must be an extreme point both for Γ_J and for Γ_K, since otherwise by the no-crossing condition the closed segments $[\![x, y_J]\!]$ and $[\![x, y_K]\!]$ should be parallel and this is impossible if $J \neq K$: since Step III gives that the extreme points for each Γ_I are f^+–negligible, it

follows that the different transport maps can be glued and then $\sum_I \gamma_I$ is induced by a transport map $\tilde{t}$.

Consider now the fact that $\gamma = \left(\sum_I \gamma_I\right) + \gamma_F$, where $\gamma_F = \gamma \llcorner \Delta$ is the part relative to the diagonal. By the non-inclusion property, $\pi_1(\Delta \cap \Gamma) \cap \mathrm{T}_\Gamma = \emptyset$; but γ_F is clearly induced by the identity map on $\pi_1(\Delta \cap \Gamma)$, while the map $\tilde{t}$ is defined on T_Γ, since the left-extreme points are f^+-negligible. Thus γ is in fact induced by a transport map, which is then optimal.

Step V. *Uniqueness.*
We have only to prove that there is a unique solution of (4.18): this is immediate thanks to Lemma 4.13. □

The following remark was kindly pointed out by M. Feldman; the reader should already know the main results of the third part of this thesis, at least Definition 14.1 and Theorems 16.2 and 16.4

Remark 9.3. *From the above construction it follows that the transport density corresponding to the transport plan γ is in L^1. This is an interesting property: it is easily seen that the transport density is in general not unique, and this is a strong difference between the crystalline norm and the Euclidean norm; more precisely, the transport density associated to the measure γ as in the above Theorem is in L^1, but in general it may happen that another optimal transport plan for the crystalline norm is associated to a singular measure: to give an example of this fact, one can simply use the construction of the Nikodym set we discussed in Section 6.2.*

Chapter 10

The norms in $\mathbb{R}^2$

In this chapter, we will see our last existence result, already appeared in [4] with a slightly different proof, which states the existence of an optimal transport map when $c(x,y) = \|y-x\|$ is any norm in $\mathbb{R}^2$. As already discussed, the no-crossing condition is particularly strong in the case when $X = Y = \mathbb{R}^2$, because it forces the direction of the transport rays more than in $\mathbb{R}^N$ with $N \geq 3$ (or, at least, in an easier-to-use way). Then, what is left open is the case of the general –i.e. non quasi-euclidean nor crystalline– norms in $\mathbb{R}^N$ with $N \geq 3$.

We will use Theorem 7.1 as discussed in Remark 7.5, then we need to show that hypotheses (i)-(iv) hold for some Γ: we will take a set Γ as in Lemma 4.12, so that property (i) is true; property (ii) will be of course taken as hypothesis, and we know by the counterexample in Section 6.3 that this is close to be sharp even if c is the Euclidean norm. Then we need to consider (iii) and (iv): we prove now that both these properties are always true in $\mathbb{R}^2$ if the no-crossing condition holds.

Lemma 10.1. *Let $\Gamma \subseteq \mathbb{R}^2 \times \mathbb{R}^2$ be a $\sigma-$compact set fulfilling the no-crossing condition (4.1). Then the map τ_Γ defined in Section 4.1 is countably Lipschitz with respect to $\mathscr{L}$.*

Proof. Up to a countable union and a rigid motion, we can easily restrict to the case when Γ is compact and prove the countable Lipschitz property on the set

$$T := \Big\{ z \in \mathrm{ET}_\Gamma : \tau_\Gamma(z) \cdot e_1 \geq \sqrt{2}/2,\ r_\Gamma(z) = C = [\![x,y]\!],\ -l_0 \leq z_1 \leq 0 < l \leq y_1 \Big\},$$

where $-l_0 < 0 < l$ are given and we denote $w_1 = w \cdot e_1$ for any point $w \in \mathbb{R}^2$. We set now $-l \leq -s \leq 0$ and $\varepsilon > 0$, and we first claim that $\mathrm{ET}_\Gamma^s = \mathrm{ET}_\Gamma \cap \{z : z_1 = s\}$ is, up to a set of $\mathscr{H}^1-$measure less than ε, a Borel set on which τ_Γ is Lipschitz with constant $k = k(\varepsilon)$.

To show this fact, consider two points z_1 and z_2 in ET^s_Γ such that $|\tau_\Gamma(z_1) - \tau_\Gamma(z_2)| \geq k|z_1 - z_2|$: as a consequence, thanks to the no-crossing condition and according to Figure 10.1 we have $|p_1 - p_2| > lk|z_1 - z_2|$; we can also assume that $|z_1 - z_2|$ is almost maximal (up to a factor $1/2$) between such pairs (z_1, z_2), and then we remove the segment $[\![z_1, z_2]\!]$ from ET^s_Γ and the corresponding segment $[\![p_1, p_2]\!]$ from $\{w_1 = l\}$. If we repeat this action countably many times, at the end we will have removed a Borel set Σ of length σ from ET^s_Γ and a set of length greater than $lk\sigma$ from $\{w_1 = l\}$: since it is clear that τ_Γ is Lipschitz with constant k in $\mathrm{ET}^s_\Gamma \setminus \Sigma$ and since, by our compactness assumption on Γ, it is

$$\mathscr{H}^1(\Sigma) \leq \frac{M}{lk}$$

for a constant $M \in \mathbb{R}^+$ depending only on the diameter of Γ, our claim is proved.

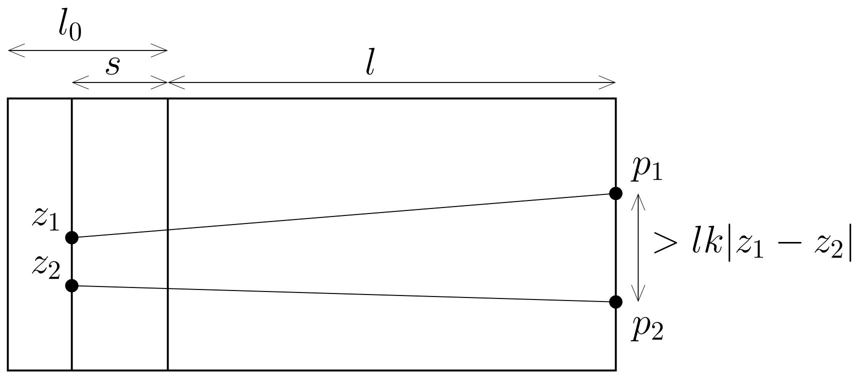


Figure 10.1: Construction in Lemma 10.1

To any s and any ε, then, we associate a set $T_{s,\varepsilon} \in \mathscr{B}(\mathbb{R}^2)$ following

$$T_{s,\varepsilon} := \Big\{ z \in T : \ r_\Gamma(z) = C = [\![x, y]\!], \ \exists\, w \in \mathrm{ET}^s_\Gamma \setminus \Sigma, \ z \in [\![x, z]\!] \Big\},$$

where Σ is as before: roughly speaking, we transport the "good" set $\mathrm{ET}^s_\Gamma \setminus \Sigma$ on the right following the transport rays. What is important, is that τ_Γ is Lipschitz with constant $k(l + l_0)/l$ on the whole set $T_{s,\varepsilon}$, as it follows by an immediate geometric argument again recalling the no-crossing condition.

It is finally easy to finish the proof: we take a set $T_1 = T_{s_1,\varepsilon_1}$ such that $\mathscr{L}(T_1)$ is almost maximal, then a set $T_2 = T_{s_2,\varepsilon_2}$ such that $\mathscr{L}(T_2 \setminus T_1)$ is almost maximal and so on. After a countable union, we have reached a set $T_\infty = \cup T_i$ which is the countable union of sets on which the restrictions of τ_Γ are Lipschitz, and which is the whole T up to an $\mathscr{L}$−negligible set. □

We prove now property (iii), i.e. the fact that the family of the doubling points is $\mathscr{L}$−negligible.

Lemma 10.2. *Let $\Gamma \subseteq \mathbb{R}^2 \times \mathbb{R}^2$ be a $\sigma-$compact set fulfilling the no-crossing condition (4.1). Then D_Γ is $\mathscr{L}-$negligible.*

Proof. Up to a countable union, we can assume that all the closures of the maximal transport rays intersecate the vertical line $\{w_1 = \lambda\}$. We restrict then our attention to the points of $\mathrm{D}_\Gamma \cap \{w_1 = \beta\}$ for some $\beta \neq \lambda$; it is clear that each such point can be associated to a non-empty open interval of $\{w_1 = \lambda\}$ by definition of doubling points and different open intervals are disjointed by the no-crossing property: then $\mathrm{D}_\Gamma \cap \{w_1 = \beta\}$ consists of at most countably many points. By Fubini Theorem, it follows that D_Γ is $\mathscr{L}-$negligible. $\square$

Remark 10.3. *As already underlined with Remark 8.6, also in this case the introduction of the notion of doubling points has been really useful: in fact, the proof that $\mathscr{L}(\mathrm{D}_\Gamma) = 0$ has been trivial, while the fact that $\mathscr{L}(\mathrm{E}^l_\Gamma) = 0$, which is still true, requires a quite more careful proof (see [2]).*

Thanks to the two preceding Lemmas, we can now easily prove the existence result:

Theorem 10.4. *Assume that $X = Y = \mathbb{R}^2$ and that $c(x, y) = \|y - x\|$ is any norm; if f^+ and f^- have finite first order moments and $f^+ \ll \mathscr{L}$, then there exists an optimal transport map t. If moreover $f^- \ll \mathscr{L}$, then t is invertible and $s = t^{-1}$ is an optimal transport map from f^- to f^+.*

Proof. Let us take a solution γ of (4.18), and let Γ be a $c-$cyclically monotone and $\sigma-$compact set on which γ is concentrated as in Theorem 4.11: our idea is to apply Theorem 7.1 with γ and Γ.

In fact, the validity of (i) is directly given by Lemma 4.12, (ii) is true by hypothesis and, since (i) holds, (iii) is given by Lemma 10.2 and (iv) by Lemma 10.1.

We apply then Lemma 7.1 finding a transport plan $\bar{\gamma}$ induced by a transport map t, which is optimal since Γ is $c-$cyclically monotone.

Concerning the second part, it can be obtained with the same argument as in Corollary 8.7. $\square$

Part II

Equivalence between three different problems

Chapter 11

Introduction and ambient space

The secont Part of this thesis is devoted to the study of three problems, related to the classical Monge's mass transportation and known to be equivalent in the case when the ambient space is an open, convex and bounded subset of $\mathbb{R}^N$; to each of these problems it corresponds a particular measures (the transport density in the Monge–Kantorovich case) and it is known that these measures are the same for each of the problems. The aim of this second Part is to generalize the setting of the three problems and the resulting definitions of the particular measures to the case of a Riemannian manifold endowed with a finslerian semidistance; we will also show that the equivalences still hold.

11.1 Standard setting of the problems

Here we give a brief presentation of the basic setting of the problems that we will generalize in this second Part of the thesis; through this section, the ambient space $X = Y$ will be a convex and bounded subset of $\mathbb{R}^N$.

The mass transport problem

We already presented and studied the mass transportation problem in the first Part of the thesis, so there is no need to restate it. We just recall that, since $X = Y$ (otherwise there is not a notion of transport density) the cost (1.5) of a transport plan becomes

$$\mathfrak{C}(\gamma) := \iint_{X\times X} c(x,y)\, d\gamma(x,y), \tag{11.1}$$

and a transport plan is any measure $\gamma \in \mathcal{M}^+(X \times X)$ such that its marginals are f^+ and f^-, two assigned positive measure of equal total mass. Moreover, it is important to assume that the l.s.c. cost function c is symmetric, i.e.

$$c(x,y) = c(y,x) \qquad \forall\, x,\, y\, \in X.$$

We recall here some of the results we proved in the first Part of the thesis, which will be needed in the sequel:

Theorem 11.1. *There exists an optimal transport plan γ, i.e. a transport plan minimizing (11.1). If moreover c is a semidistance and it is bounded, there exists a Kantorovich potential u, i.e. a function such that $u(x) - u(y) \le c(x,y)$ for any couple $(x,y) \in X \times X$ and $u(x) - u(y) = c(x,y)$ for $\gamma-$a.e. $(x,y) \in X \times X$ whenever γ is an optimal transport plan.*

The preceding statement is in fact quite weaker than Theorem 1.11 and Corollary 3.27, just note that if c is bounded then it is *slow* $-\,(f^+, f^-)\,-$ *increasing* for any couple of measures f^+, f^- on X with the same total mass.

We conclude with the definition of the transport density, a measure that we already mentioned in the first Part of the thesis, and which plays a crucial role in the connection with the other problems that we are going to study: given a transport plan γ, the *transport density* associated to γ is the measure $\mu \in \mathcal{M}^+(X)$ defined by

$$\langle \mu, \varphi \rangle := \iint_{X\times X} \left(\int_X \varphi(z)\, d\mathscr{H}^1_{xy}(z) \right) d\gamma(x,y) \qquad \forall\, \varphi \in \mathrm{C}_b(X); \tag{14.1}$$

with $\mathscr{H}^1_{xy}$ we denote the restriction of the one-dimensional Hausdorff measure $\mathscr{H}^1$ to the segment xy. Recalling that the transport plan γ takes a mass $\gamma(A\times B)$ from A and transports it in B, the transport density is the distribution in X of the work done by the transport plan γ. We will call *optimal transport density* each transport density corresponding to an optimal transport plan γ.

The Brenier problem

This second problem was proposed by Brenier (see [18] and [10]), and consists in finding the best configuration which transports f^+ onto f^-. A *configuration* is a set of scalar measures f_t and vectorial measures $\overline{E}_t$ on X, with $t \in [0,1]$, such that the compatibility condition $\dot{f}_t + \operatorname{div} \overline{E}_t = 0$ holds in the distribution sense; by saying that $\{f_t, \overline{E}_t\}$ "transports f^+ onto f^-" we intend that $f_0 = f^+$ and $f_1 = f^-$. The "goodness" of a configuration is given by its cost

$$\mathfrak{C}\left(\{f_t, \overline{E}_t\}\right) := \int_0^1 \|\overline{E}_t\|\; dt, \tag{11.2}$$

where $\|\overline{E}_t\| = |\overline{E}_t|(X)$ is the norm of $\overline{E}_t$ in the space of the vectorial measures on X; of course a configuration is better if its cost is lower. The connection between the configurations and the mass transport problem can be roughly described as follows: we imagine that a transport from f_0 to f_1 is made in an unitary time, and then f_t and $\overline{E}_t$ respectively say, at time t, which is the distribution of mass and the speed with which it has to be moved. For each configuration $\{f_t, \overline{E}_t\}$ we consider the *work density*, i.e. the measure μ given by

$$\mu := \int_0^1 |\overline{E}_t| \, dt. \tag{11.3}$$

We will call *optimal work density* each work density corresponding to a configuration $\{f_t, \overline{E}_t\}$ of minimal cost.

About the interest and the different physical interpretations of this problem see the introduction of [10].

The shape optimization problem

The last problem we consider is the *shape optimization*, which can be reformulated (see [14, 15]) as the search of the extremals of the following dual problems:

$$\sup\left\{\int_X u \, d(f^+ - f^-) : \ u \text{ is } 1\text{–Lipschitz}\right\} \tag{11.4}$$

$$\inf\left\{\int_X d|\nu| : \ -\operatorname{div}\,\nu = f^+ - f^-\right\} \tag{11.5}$$

–note that the first problem is strictly related to (3.3) in view of Corollary 3.27. The *optimal shapes* are the measures $\mu = |\nu|$ where ν is a solution of (11.5); in particular, the problem (11.5) was already considered in [51, 71]. In [14, 15] it was also shown the connection between the shape optimization problem and the following PDE equation, whose formulation in the context of the Monge–Kantorovich problem is due to Evans and Gangbo ([39], see also Section 6.4):

$$\begin{cases} -\operatorname{div}\,(\mu Du) = f^+ - f^- & \text{on } X \\ |Du| = 1 & \mu - \text{a.e.}, \end{cases} \tag{11.6}$$

where $\mu \in \mathcal{M}^+(X)$ and u is a 1–Lipschitz function on X; we will discuss the connections which hold in our generalized setting in Section 12.5.

Before the paper [62] (that is the source of this second Part of the thesis), the equivalence between these problems in the case of the Euclidean cost function $c(x, y) = |x - y|$ was known as follows (see [14, 1]):

Theorem 11.2. $\inf(11.1) = \inf(11.2) = \inf(11.5)$. *Moreover the set of the optimal transport densities, the set of the optimal work densities and the set of the optimal shapes are the same.*

Let us discuss for a moment the second part of the last assert: in general, there are several optimal transport densities, corresponding to different optimal transport plans, as there are several optimal work densities and optimal shapes; however, a measure μ on X is an optimal transport density if and only if it is an optimal work density and if and only if it is an optimal shape. Moreover, the uniqueness of the transport density is true under suitable additional hypotheses, see Theorem 16.2.

Some discussions about the equivalences between the problems and the proofs of the asserts of the preceding Theorem can be found in [1], [12], [31] and [61], under the hypothesis that X is a bounded and convex subset of $\mathbb{R}^N$; an important generalization that has been already considered (see for example [12]) is to imagine a closed Dirichlet region $\Sigma \in X$ on which the transport is free, while the case of a C^3 regular Riemannian manifold was considered in [43] (see also [30, 50]).

11.2 The ambient space

Here we present the ambient space we will consider: we are interested in the case of a Riemannian manifold M, on which the Riemannian distance is replaced by a more general finslerian semidistance.

Let M be an $N-$dimensional connected Riemannian manifold, non necessarily complete and possibly with boundary, whose Riemannian metric is denoted by d_M. We want define a finslerian semidistance on M (to find a general discussion on finslerian metrics, refer to [9]): to do this, we are given a measurable function $j : TM \longrightarrow \mathbb{R}^+$ $1-$homogeneous, i.e. $j(x, \lambda v) = |\lambda| j(x, v)$. We remark that, as just done, if $v \in T_x M$ we will always write $j(x, v)$ instead of simply $j(v)$: the second writing would be the formally correct one, since $v \in TM$, but the first one is more useful because it specifies the point of M associated to the vector v.

We denote by S the set of the paths in M which are Lipschitz with respect to d_M, and we regard S as a metric space by means of the distance

$$d_S(\sigma, \tau) := \max_{t\in[0,1]} d_M(\sigma(t), \tau(t)) : \tag{11.7}$$

here we think, as we will always do, $\sigma : [0, 1] \longrightarrow M$ as the parameterization at constant speed of any element $\sigma \in S$. Now we define the length of each path $\sigma \in S$ as

$$l(\sigma) := \int_0^1 j\big(\sigma(t), \sigma'(t)\big)\, dt,$$

that is well defined (possibly $+\infty$) since j is nonnegative; finally, the finslerian semidistance d between two points is simply the infimum of the lengths of all the paths that connect them, i.e.

$$d(x, y) := \inf \{ l(\sigma) : \ \sigma \in S, \ \sigma(0) = x, \ \sigma(1) = y \} .$$

Note that d reduces to the standard Riemannian distance if $j(x, v) = |v|$, while the case of the Dirichlet region Σ cited above is the situation in which $j(x, v) = 0$ if $x \in \Sigma$ and $j(x, v) = |v|$ otherwise.

We will call M a *manifold of type 0* if the following holds:
0-a) The metric space (M, d) is bounded.
0-b) j is locally bounded, i.e. for each $x \in M$ there exist a neighborhood U of x and a constant K such that $j(z, v) \leq K|v|$ for each $z \in U$ and $v \in T_z M$.
0-c) For each $x \in M$ the restriction $j(x, \cdot)$ is a convex function.

If M is of type 0, we will call it a *manifold of type 1* if the following holds:
1-a) M is a complete manifold, possibly with boundary;
1-b) j is l.s.c.;
1-c) there exists an idempotent map $ret : S \longrightarrow S$ (i.e. $ret^2 = ret$) such that $ret(\sigma)$ has the same extremal points of σ, $\mathscr{H}^1_j \llcorner ret(\sigma) \leq \mathscr{H}^1_j \llcorner \sigma$ –see (12.1)– and there are two constants C and α such that, denoted by l_M the Riemannian length of the paths, for any $\sigma \in S$ it holds

$$l_M(ret(\sigma)) \leq C + \alpha\, l(\sigma) : \tag{11.8}$$

we will call "retraction" the function ret and its push-forward $ret_\# : \mathcal{M}^+(S) \to \mathcal{M}^+(S)$ and "retracted" a path $\sigma \in S$ or a measure $\eta \in \mathcal{M}^+(S)$ if $ret(\sigma) = \sigma$ or $ret_\#(\eta) = \eta$.

Finally, if M is of type 0, we will call it a *manifold of type 2* if, denoted by $(\hat{M}, d_M)$ and $\big(\hat{S}, d_S\big)$ the completions of (M, d_M) and (S, d_S) respectively, the following holds:
2-a) $\hat{M}$ is a manifold of type 1, where we extend j from TM to the whole $T\hat{M}$ by lower semicontinuity; that is, the new function $\hat{j} : T\hat{M} \to \mathbb{R}^+$ is the supremum of all the l.s.c. functions whose restriction to TM is lower than j: it is equivalent to set

$$\hat{j}\,(x, v) = \inf \Big\{ \liminf j(x_n, v_n) : \ x_n \in M, \ (x_n, v_n) \xrightarrow{T\hat{M}} (x, v) \Big\},$$

and it is a standard fact that $\hat{j}$ is l.s.c. and coincides with j on $TM \subseteq T\hat{M}$: therefore, we will for simplicity use the letter j also to denote the extension to the whole $T\hat{M}$.
2-b) For each $\sigma \in \hat{S}$, there exists a sequence $\sigma_i \to \sigma$ in S such that $l(\sigma_i) \to l(\sigma)$.

It is easy to show that $\hat{S}$ can be equivalently defined as the metric space of the paths on $\hat{M}$ which are Lipschitz with respect to d_M, where the distance is given as in (11.7).

11.3 About the hypotheses

Here we briefly discuss the hypotheses we made: first of all, let us note that our hypotheses are very weak; in fact, they cover much more situations than the ones which have been already considered for the Monge–Kantorovich problem, and in particular all the compact manifolds with boundary, equipped with any Riemannian metric and possibly with a Dirichlet region. In the manifolds of type 0, the more general we can work with, the problems presented in Section 11.1 will be suitably restated, and we will prove that the extremals are still equal. Note that asking M to be a bounded metric space allows us to avoid situations in which all the extremals are $+\infty$, which is not particularly interesting for our purposes and forces to make a lot of boring distinctions; the fact that j is locally bounded ensures that the distance d is continuous with respect to the Riemannian topology; finally, the convexity assumption is standard when dealing with a finslerian semidistance.

It is easy to imagine that we can not hope the extremals to be reached without some completeness hypotheses: in fact, we will prove the existence of solutions and the fact that the optimal measures of the problems are the same for the manifold of type 1, which are complete, and for the manifold of type 2, in which we will consider a relaxed setting of the problems passing to the completion. To convince ourselves that these hypotheses are nearly essential to get the existence of solutions, we point out that:
1) if j were not be l.s.c., we could not hope to have paths of minimal length, which is clearly essential to reach the extremals of the problems; moreover standard convexity arguments show that the lower semicontinuity of j implies the one of $l : S \to \mathbb{R}^+$.
2) The hypothesis *1-c)* seems at first glance a bit strange, but it is quite natural: the idea is that if for example σ contains some closed subpath, then $ret(\sigma)$ is without them; or if in σ there is a long –in the Riemannian sense– part in a Dirichlet region Σ (it does not effect $l(\sigma)$ since the part in Σ has null finslerian length), then in $ret(\sigma)$ it will be replaced by a shorter one with the same extremal points. This hypothesis is useful because if there are short (w.r.t. l) paths with great Riemannian length l_M, we need some corresponding paths which are short also in the Riemannian sense; if not, there could lack paths of minimal

length. However, if for example on M there is a Dirichlet region Σ and $j(x,v) \geq \varepsilon|v|$ for $x \notin \Sigma$, then (11.8) holds when $\alpha = \varepsilon^{-1}$ and C is the sum of the diameters of the arcwise connected components of Σ (if finite). Vice versa, if there is no ε such that $j(x,v) \geq \varepsilon v$ there are counterexamples to the existence of minimizers.

3) The hypothesis *2-b)* is fundamental to made a good relaxation of the problems in the non-complete case since, if it does not hold, the extremals could change passing to the relaxed formulations: in fact, there could be paths in $\hat{M}$ connecting two points with length strictly smaller than the distance between them. However, note that this is not a strong hypothesis: it holds, for example, if j is continuous on $\hat{M} \setminus M$.

Chapter 12

Setting of the problems and relaxation

In this section we will set the problems in the general context of the manifolds of type 0; then we will relax them in the non-complete case of the manifolds of type 2; then, we discuss the possibility to generalize problem (11.6); finally, we present some operations which can be done with the transports. In the sequel, f^+ and f^- will be positive measures on M of equal total mass, and we will denote $f = f^+ - f^-$.

12.1 The Monge–Kantorovich problem

Let us recall that, in the classical case of a convex subset X of $\mathbb{R}^N$, a transport is simply a measure γ on $X \times X$, whose meaning is that a mass $\gamma(x, y)$ has to be moved from x to y; γ does not say *how* to move this mass, but this does not give difficulties since clearly the best way is to follow the segment from x to y. In our new setting the situation is no more so easy, because given x and y in M there could be more than one shortest path from x to y, or there could be no shortest path at all: to say that a certain mass must travel from x to y is then not sufficient, we must specify which path has to be used. The best definition for a transport is then the following:

Definition 12.1. *Given f^+, $f^- \in \mathcal{M}^+(M)$ of equal total mass, a* transport *between f^+ and f^- is a measure $\eta \in \mathcal{M}^+(S)$ such that $\pi_0\eta = f^+$ and $\pi_1\eta = f^-$.*

By $\pi_t : S \longrightarrow M$ we denote the function defined by $\pi_t(\sigma) = \sigma(t)$, recalling that we think always σ as parameterized at constant speed. It is easy to understand how do we think a measure η in Definition 12.1 as a "transport": $\eta(\sigma)$ indicates the quantity of mass which must be moved following the path σ, and then $\pi_0\eta$ and $\pi_1\eta$ are the distributions of mass that start and arrive respectively. We can note that in this new setting, where giving a transport means saying precisely how to move the mass, the transports are much more similar to the configurations (in the sense of Brenier). In fact, the difference between transports and configurations will be better discussed in Remark 13.4.

Having defined the transports, we have now to specify their cost; since $\eta(\sigma)$ is the quantity of mass which is moved through the path σ, it is not difficult to generalize (11.1) defining the cost of η, which can also be $+\infty$, as

$$\mathfrak{C}(\eta) := \int_S l(\sigma)\ d\eta(\sigma). \tag{MK}$$

Note that what we are presenting is a generalization of the classical transport problem in the space M with the semidistance d as cost function, and we have inf(11.1)=inf(MK): in fact, we can associate to the "new" transport η the "old" transport $\gamma = (\pi_0, \pi_1)_\# \eta$, and it is clearly $\mathfrak{C}(\gamma) \leq \mathfrak{C}(\eta)$ since whenever $\sigma \in S$ connects x and y it is $d(x, y) \leq l(\sigma)$; on the other hand, if $\beta : M \times M \longrightarrow S$ is a measurable selection of paths which are optimal up to ε and we associate to the "old" γ the "new" $\eta = \beta_\# \gamma$, it is $\mathfrak{C}(\eta) \leq \mathfrak{C}(\gamma) + \varepsilon \|\gamma\|$. This argument also shows that a minimizing transport (if there exists) is concentrated on geodesics for the semidistance d.

Remark 12.2. *Note, in particular, that if the ambient space is $\mathbb{R}^N$ and the distance is the Euclidean one, there is no difference between the optimal transport plans studied in the first Part and the optimal transports according to our new notations: in fact, any transport plan corresponds to the transport which follows the straight lines between the points, and in the same way an optimal transport is associated to a unique optimal transport plan since it must be concentrated in straight lines by optimality.*

In order to introduce the transport densities, as it can be understood recalling (14.1), we have first to define $\mathscr{H}^1_j \llcorner \sigma$ for $\sigma \in S$, which is the 1−dimensional Hausdorff measure on the path σ weighed with j: formally, if $\varphi \in \mathrm{C}_b(M)$, it can be computed by

$$\langle \mathscr{H}^1_j \llcorner \sigma, \varphi \rangle := \int_0^1 \varphi\left(\sigma(t)\right) j\big(\sigma(t), \sigma'(t)\big)\ dt. \tag{12.1}$$

We can now define the *transport density* associated to a transport η: the direct generalization of (14.1) turns out to be

$$\mu := \int_S \mathscr{H}^1_j \llcorner \sigma\ d\eta(\sigma); \tag{12.2}$$

note also that $l(\sigma) = \langle \mathscr{H}^1_j \llcorner \sigma, 1 \rangle$ and then $\mathfrak{C}(\eta) = \langle \mu, 1 \rangle = \mu(M)$.

12.2 The Brenier problem

The new formulation of the Brenier problem will be the same as the old one, and we will need only to extend the definitions of the objects we work with to our new setting. Recalling the definitions given in Appendix B, we start defining *configuration* any set $\left\{f_t \in \mathcal{M}^+(M),\, \overline{E}_t \in \mathcal{M}^N(M), t \in [0,1]\right\}$ such that the compatibility condition

$$\dot{f}_t + \operatorname{div} \overline{E}_t = 0 \tag{12.3}$$

holds. The meaning of (12.3) is that for each $\varphi \in \mathrm{C}_b^\infty(M)$ the equality

$$\frac{d}{dt} \langle f_t, \varphi \rangle = \langle \overline{E}_t, D\varphi \rangle$$

holds in the weak sense: therefore, if $\left\{f_t, \overline{E}_t\right\}$ is a configuration, for any $\varphi \in \mathrm{C}_b^\infty(M)$ the function $t \to \langle \overline{E}_t, D\varphi \rangle$ is in $L^1_{\mathrm{loc}}([0,1])$ and $t \to \langle f_t, \varphi \rangle$ is in $W^{1,1}_{\mathrm{loc}}([0,1])$, so in particular it is continuous. This implies that the measure f_t is well-defined for any time t, and then it makes sense to say that $\left\{f_t, \overline{E}_t\right\}$ is a *configuration from f^+ to f^-* if $f_0 = f^+$ and $f_1 = f^-$; in particular the measures f_t are somehow a "continuous interpolation" between f_0 and f_1. With an easy generalization of (11.2), we define the cost of a configuration as

$$\mathfrak{C}\left(\left\{f_t, \overline{E}_t\right\}\right) := \int_0^1 \int_M j(x, d\overline{E}_t(x))\; dt \tag{Br}$$

where, written any $\nu \in \mathcal{M}^N(M)$ as $\nu = \overline{\omega}|\nu|$ (a polar decomposition) as discussed in Appendix B, we will always intend

$$j(x,\nu) := j(x,\overline{\omega})\,|\nu| \quad \in \mathcal{M}^+(M):$$

this means that, for any continuous and bounded function φ,

$$\langle j(x,\nu), \varphi \rangle = \int_M \varphi(x)\, j(x,\nu) = \int_M \varphi(x)\, j(x,\overline{\omega})\, d|\nu|$$

Finally, the definition of the *work density* which generalizes (11.3) is

$$\mu := \int_0^1 j(x, d\overline{E}_t(x))\; dt. \tag{12.4}$$

As made for the Monge–Kantorovich problem, we can note that $\mathfrak{C}\left(\left\{f_t, \overline{E}_t\right\}\right) = \langle \mu, 1 \rangle = \mu(M)$.

12.3 The shape optimization problem

As we said in the introduction, the shape optimization problem consists in solving two variational problems; to generalize them, we consider first of all the set $\mathrm{Lip}_j(M)$ of all the functions on M which are $1-$Lipschitz with respect to d (recall that d depends on j). So, instead of (11.4), we will try now to find

$$\sup \int_M u \, df \tag{12.5}$$

among all $u \in \mathrm{Lip}_j(M)$ (recall we write $f = f^+ - f^-$); note that, thanks to the local boundedness of j, the elements of $\mathrm{Lip}_j(M)$ are locally Lipschitz, and so continuous, with respect to d_M; moreover, since (M, d) is bounded by hypothesis *0-a)*, then the elements of $\mathrm{Lip}_j(M)$ are also bounded: thus the integral in (12.5) is well defined for each $u \in \mathrm{Lip}_j(M)$.

On the other hand, the search of the minimizers of (11.5) becomes the search of

$$\inf \int_M j(x, \nu) \tag{SO}$$

among all $\nu \in \mathcal{M}^N(M)$ such that $-\mathrm{div}\ \nu = f$, i.e. $\langle f, \varphi \rangle = \langle \nu, D\varphi \rangle$ for each $\varphi \in \mathrm{C}_b^\infty(M)$. As in the standard setting, each measure $\mu = j(x, \nu)$ where ν solves (SO) is said an *optimal shape.*

Since we refer to (12.5) and (SO) as "dual problems", one can imagine that we will prove these extremals to be equal; we prove now the first inequality, the second one will follow from the proof of Theorem 13.1:

Proposition 12.3. $\sup(12.5) \leq \inf(SO)$.

Proof. Let us take a smooth function $u \in \mathrm{Lip}_j(M)$ and a vectorial measure $\nu \in \mathcal{M}^N(M)$ such that $-\mathrm{div}\ \nu = f$; recalling the definition of the semidistance d and the fact that $u \in \mathrm{Lip}_j(M)$ one can deduce that

$$Du(x) \cdot \overline{\omega}(x) \leq j\left(x, \overline{\omega}(x)\right) \qquad |\nu|-\text{a.e.} \tag{12.6}$$

where we write, as in Appendix B, $\nu = \overline{\omega}|\nu|$. We can then write

$$\int_M u \, df = \int_M Du \cdot d\nu = \int_M Du \cdot \overline{\omega} \, d|\nu| \leq \int_M j\left(x, \overline{\omega}\right) d|\nu| = \int_M j(x, \nu). \tag{12.7}$$

Since the inequality holds for a generic $u \in \mathrm{C}_b^\infty \cap \mathrm{Lip}_j(M)$ and a generic $\nu \in \mathcal{M}^N(M)$ with $-\mathrm{div}\ \nu = f$, it suffices to extend (12.7) to any $u \in \mathrm{Lip}_j(M)$ and then one concludes taking a sup in the left of (12.7) and an inf in the right. A strategy to extend (12.7) to a generic function $u \in \mathrm{Lip}_j(M)$ could be of course to approximate it with smooth ones, but this should be done carefully because also the approximating smooth functions need to be in $\mathrm{Lip}_j(M)$ in order to infer the validity of (12.7) for them. A simpler way is to use the definition and the properties of the tangential derivatives, as given in [14, 15] and generalized in [44]: in fact, using their language, the elements of $\mathrm{Lip}_j(M)$ are easily seen to belong to $\mathrm{W}^{1,2}_{|\nu|,\mathrm{loc}}$, that is a suitable Sobolev space made with respect to the measure $|\nu|$. This implies that, taken any $u \in \mathrm{Lip}_j(M)$, the tangential derivative $D_{|\nu|}u$ is defined $|\nu|-$a.e. and (12.6) can be generalized as

$$D_{|\nu|}u(x) \cdot \overline{\omega}(x) \leq j\left(x, \overline{\omega}(x)\right) \qquad |\nu|-\text{a.e.}\,.$$

Finally, we recall one of the main properties of the notion of tangential derivative: whenever μ is a positive measure, $v \in \mathrm{W}^{1,2}_{\mu,\mathrm{loc}}$ and $\mathrm{div}\ (\varphi\mu) \in \mathcal{M}(M)$, the equality

$$\int_M D_\mu v \cdot \varphi \, d\mu = -\langle \mathrm{div}\ (\varphi\mu), v \rangle$$

holds. Therefore, setting $\varphi = \overline{\omega}$, $\mu = |\nu|$ and $v = u$ we find

$$\int_M D_{|\nu|}u \cdot \overline{\omega} \, d|\nu| = \langle -\mathrm{div}\ \nu, u \rangle = \int_M u \, df$$

and then, as in (12.7), we deduce

$$\int_M u \, df \leq \int_M j\left(x, \overline{\omega}\right) d|\nu| = \int_M j(x, \nu);$$

as already mentioned, this concludes the proof. □

12.4 Relaxation for manifolds of type 2

As we noted in Section 11.2, the completeness of the space M is essential to have existence of solutions for our problems. For example, if M is a ball in $\mathbb{R}^N$ without the center and with the Riemannian distance, x and y are two opposite points and $f^+ = \delta_x$ and $f^- = \delta_y$, then clearly $\inf(MK) = d(x, y)$, but any path between x and y is strictly longer than $d(x, y)$ and then the inf can not be reached; the same reason shows that no configuration is optimal and that also (SO) is not a minimum. Only in (12.5) the maximum is always attained, and in fact we will not need to relax that problem.

Let us then consider a manifold M of type 2: we will relax the problems simply considering them with $\hat{M}$ and $\hat{S}$ in place of M and S. Then a *relaxed transport* between f^+ and f^- is nothing else than a measure $\eta \in \mathcal{M}^+(\hat{S})$ such that $\pi_0\eta = f^+$ and $\pi_1\eta = f^-$; in the same way, a *relaxed configuration* is $\left\{f_t, \overline{E}_t\right\}$, where the f_t and the $\overline{E}_t$ are now scalar and vectorial measures on $\hat{M}$, if the compatibility condition (12.3) holds and $f_0 = f^+$, $f_1 = f^-$; finally, a *relaxed shape* is now the measure $j(x,\nu)$ if $-\mathrm{div}\ \nu = f$ and ν is a vectorial measure on $\hat{M}$. The relaxed problems will be then the search of the relaxed transports, or configurations, or shapes, which are extremals for (MK), (Br) or (SO), where the integrals must be intended over $\hat{M}$ and $\hat{S}$ instead of M and S. Finally, note that the problem to maximize (12.5) need not to be extended: in fact, the 1−Lipschitz functions on M or $\hat{M}$ are exactly the same by construction, and the integral $\int u\, df$ over M or $\hat{M}$ does not change, since f is a measure on M. Moreover, it is easily seen that (12.5) admits a maximum, recalling that $\int df = 0$ and applying Ascoli–Arzelà theorem to a maximizing sequence of functions with 0 mean.

12.5 The extension of problem (11.6)

As we said in the introduction, in the classical setting there is a connection between the shape optimization problem and the PDE equation (11.6), namely that (u,μ) solves (11.6) if and only if u and $\nu = \mu Du$ separately solve (11.4) and (11.5). Since our finslerian context is too general to give a notion of gradient, in this paper we are not interested in a generalization of (11.6); however, in this section we will briefly describe how this can be done in a particular case, which contains most of the interesting situations and generalizes the study of (11.6) done in [12] and [14, 15]. To give a meaningful definition of "gradient", the strict convexity of j is needed: let us then assume, only for this section, that there is a closed Dirichlet region Σ (i.e. $j(x,\cdot) \equiv 0$ for $x \in \Sigma$) such that on $M \setminus \Sigma$ j is continuous and $j(x,\cdot)$ is strictly convex on T_xM; note carefully that, by saying that $j(x,\cdot)$ is strictly convex, we mean that the inequality

$$j(x, tv + (1-t)w) \leq tj(x,v) + (1-t)j(x,w)$$

with $t \in (0,1)$ must be strict unless v and w are parallel (if $v = \lambda w$ with $\lambda \geq 0$ it is an equivalence by the 1−homogeneity). Then, if $u \in W^{1,\infty}_{\mathrm{loc}}(M)$ and $x \notin \Sigma$ is a point where $Du(x)$ is defined, it is $u(x+\varepsilon v) \approx u(x) + \varepsilon\, Du(x)\cdot v$ and $d(x, x+\varepsilon v) \approx \varepsilon j(x,v)$, and then the slope is

$$v \longrightarrow \frac{Du(x)\cdot v}{j(x,v)}. \tag{12.8}$$

On the unit sphere of T_xM this is a continuous function, and then by compactness it admits a maximum $\hat{v}$, which is unique thanks to the strict convexity of j on T_xM. For each $u \in W^{1,\infty}_{\mathrm{loc}}(M)$ and $x \notin \Sigma$ where $Du(x)$ is defined, then, we define the *gradient of u at x* as

$$\nabla_j\, u(x) := \frac{Du(x)\cdot \hat{v}}{j(x,\hat{v})^2}\, \hat{v},$$

which is the vector on the direction of the maximal slope such that $j(x, \nabla_j u(x))$ equals that slope; then it easily follows that $\mathrm{Lip}_j(M)$ are exactly the continuous elements of $W^{1,\infty}_{\mathrm{loc}}(M)$ such that $j\big(x, \nabla_j u(x)\big) \leq 1$ for a.e. $x \in M \setminus \Sigma$. Now we need to generalize this concept with respect to a measure: to do this, suppose we are given a measure μ and let $x \notin \Sigma$ such that $D_\mu u(x)$, the $\mu-$tangential gradient (see [14, 15] and [44]), is defined. Let then $\hat{v}$, as before, be the unique maximum of (12.8) on the unit sphere of T_xM: the *$\mu-$tangential gradient of u at x* is

$$\nabla_{j,\mu} u(x) := \frac{D_\mu u(x) \cdot \hat{v}}{j(x, \hat{v})^2}\, \hat{v},$$

which is clearly defined $\mu-$a.e.; now, we can finally generalize problem (11.6) as the search of a measure $\mu \in \mathcal{M}^+(M)$ and a function $u \in \mathrm{Lip}_j(M)$ such that

$$\begin{cases} -\mathrm{div}\ (\mu \nabla_{j,\mu} u) = f & \text{on } M \setminus \Sigma \\ j(x, \nabla_{j,\mu} u) = 1 & \mu - \text{a.e. on } M \setminus \Sigma. \end{cases} \tag{12.9}$$

This is a good extension of the classical setting, since we can prove the

Proposition 12.4. *There are solutions for problem (12.9); in particular, if u and ν are optimal for (12.5) and (SO), then $(u, j(x, \nu))$ solves (12.9).*

Proof. Let us take u and ν solving (12.5) and (SO) –Theorem 13.8 will prove that such extremals exist– and, defined $\mu = j(x, \nu)$, write $\nu = \theta\mu + \tilde{\nu}$ where $\tilde{\nu} = \nu \llcorner \Sigma$: then $j(x, \theta) \equiv 1$ $\mu-$a.e.; recall now that $D_\mu u(x) \cdot v \leq j(x, v)$ for $\mu-$a.e. x and each v thanks to the $1-$Lipschitz property of u, and that $\langle \tilde{\nu}, Du \rangle = 0$ as one can infer from the continuity of u and the fact that $-\mathrm{div}\ \nu = f \in \mathcal{M}(M)$. Since, as we will note with Corollary 13.2, $\inf(SO) = \sup(12.5)$, we have

$$\begin{aligned} \int_M d\mu = \int_M j(x, \nu) = \inf\ (SO) = \sup\ (12.5) = \int_M u\ df = \langle -\mathrm{div}\ (\theta\mu + \tilde{\nu}), u \rangle \\ = \int_M D_\mu u \cdot \theta\ d\mu \leq \int_M j(x, \theta)\ d\mu = \int_M d\mu : \end{aligned}$$

it follows that $D_\mu u \cdot \theta = j(x, \theta)$, and this implies that $\nabla_{j,\mu} u = \theta$; since problem (12.9) requires the validity of $-\mathrm{div}\ (\mu \nabla_{j,\mu} u) = f$ only on $M \setminus \Sigma$, (u, μ) solves (12.9). $\square$

The converse of the theorem –which is well known to be true in the classical setting– is now only partially true: in fact, if (u, μ) solves (12.9) then u is extremal for (12.5) and $\nu = \mu \nabla_{j,\mu} u$ is such that $\int j(x, \nu) = (SO)$; but $-\mathrm{div}\ \nu = f$ holds only out of Σ and not necessarily on the whole M. However, we can change ν in Σ in such a way that the equation holds on M; moreover, since $j(x, \cdot) \equiv 0$ in Σ, the following fact is true:

Remark 12.5. *The measures μ in the solutions of (12.9) are all and only the optimal shapes.*

12.6 Useful operations with the transports

Here we define three useful operations with transports and we prove their main properties.

The restriction

Let us begin with the restriction: if $\eta \in \mathcal{M}^+(S)$ is a transport between f^+ and f^- and $h \leq f^+$, we define *restriction* of η to h (or, more precisely, restriction on the first component) as the transport

$$\eta_{|h} := (h_{f^+} \circ \pi_0)\, \eta,$$

where $h_{f^+} \in L^1_{f^+}(M)$ is the density of h with respect to f^+, i.e. $h = h_{f^+} f^+$. It is immediate to note that $\pi_0 \eta_{|h} = h$, which means that $\eta_{|h}$ is a transport moving h somewhere, and that if $f^+ = g + h$ then $\eta = \eta_{|g} + \eta_{|h}$ and $\mathfrak{C}(\eta) = \mathfrak{C}(\eta_{|g}) + \mathfrak{C}(\eta_{|h})$. In the same way one can make the restriction on the second component of η to $h \leq f^-$, that we will denote by $\eta^{|h}$.

The composition

The second operation we consider is the *composition* of transports: if η transports f^+ onto g and ν transports g onto f^-, it is clear that we would gladly consider some transport $\nu \circ \eta$ that moves f^+ onto f^-; it is not difficult to write down the definition which corresponds to what we have in mind, that turns out out be

$$\langle \nu \circ \eta, \varphi \rangle := \int_{x \in M} \left(\iint_{(\sigma,\tau) \in S \times S} \varphi(\sigma \circ \tau) d\nu_x(\sigma)\ d\eta_x(\tau) \right) dg(x),$$

where ν_x and η_x denote the disintegrations of ν and η with respect to π_0 and π_1 in the sense of Theorem A.15; note that $\sigma \circ \tau$ is well defined, since one has $\sigma(0) = x$ for ν_x-a.e. σ and $\tau(1) = x = \sigma(0)$ for η_x-a.e. τ. Some straightforward calculations prove the following

Proposition 12.6. *The composition $\nu \circ \eta$ is a transport that moves f^+ onto f^-; moreover $\zeta \circ (\nu \circ \eta) = (\zeta \circ \nu) \circ \eta$; finally the transport density is additive, in the sense that $\mu_{\nu\circ\eta} = \mu_\nu + \mu_\eta$, and then $\mathfrak{C}(\nu \circ \eta) = \mathfrak{C}(\nu) + \mathfrak{C}(\eta)$.*

The simplification

We want to define now the *simplification* of transports: given η_0 which transports $f^+ + \varrho_0$ onto $f^- + \varrho_0$, we would like to build another transport moving f^+ onto f^- in a reasonable way (we mean, doing somehow what η_0 does; anyway, this will be more clearly stated in

Proposition 12.7): to make this, let us consider $\eta_{0|f^+}$, and write it as $\eta_{0,a}+\eta_{0,b}$ where the first part transports mass on f^- and the second on ϱ_0; formally,

$$\eta_{0,a} := \big(\eta_{0|f^+}\big)^{|\pi_1(\eta_{0|f^+})\wedge f^-} \qquad\qquad \eta_{0,b} = \eta_{0|f^+} - \eta_{0,a};$$

on the other hand, we divide $\eta_{0|\varrho_0}$ as $\eta_{0,c}+\eta_{0,d}$ where

$$\eta_{0,c} := \big(\eta_{0|\varrho_0}\big)_{|\pi_1\eta_{0,b}} \qquad\qquad \eta_{0,d} = \eta_{0|\varrho_0} - \eta_{0,c};$$

so we write $\eta_0 = \eta_{0,a}+\eta_{0,b}+\eta_{0,c}+\eta_{0,d}$, and $\eta_{0,a}$ transports part of f^+ onto a part of f^-, $\eta_{0,b}$ transports the rest of f^+ onto a part of ϱ_0, $\eta_{0,c}$ transports this part of ϱ_0 somewhere and finally $\eta_{0,d}$ is the remaining part of the transport; note that one can make the composition $\eta_{0,c}\circ\eta_{0,b}$. Let us then define $\tau_0=0$, $\tau_1=\eta_{0,a}$, $\eta_1=\eta_{0,c}\circ\eta_{0,b}+\eta_{0,d}$, $\varrho_1=\pi_0\eta_{0,d}$ and, finally, $f_0^+=f^+$, $f_0^-=f^-$, $f_1^+=\pi_0\eta_1-\varrho_1$ and $f_1^-=\pi_1\eta_1-\varrho_1$. So, since $\|f^+\|\le\|f_0^+-f_1^+\|+\|f_1^+\|$, recalling that $\tau_0+\tau_1$ and $\eta_{0,b}$ transport $f_0^+-f_1^+$ onto $f_0^--f_1^-$ and f_1^+ onto $\varrho_0-\varrho_1$ respectively, we obtain

$$\|f^+\|_{\mathcal{M}^+(M)} \le \|\tau_0+\tau_1\|_{\mathcal{M}^+(S)} + \|\varrho_0-\varrho_1\|_{\mathcal{M}^+(M)}.$$

But we have now η_1 that transports $f_1^++\varrho_1$ onto $f_1^-+\varrho_1$, and then we can iterate our construction finding η_i, τ_i, ϱ_i, f_i^+ and f_i^-; so $\tau_0+\cdots+\tau_i$ transports $f^+-f_i^+$ onto $f^--f_i^-$, and the last inequality generalizes as

$$\|f^+\|_{\mathcal{M}^+(M)} \le \|\tau_0+\cdots+\tau_i\|_{\mathcal{M}^+(S)} + \|\varrho_{i-1}-\varrho_i\|_{\mathcal{M}^+(M)}; \tag{12.10}$$

now, since $\varrho_0\ge\varrho_1\ge\cdots\ge\varrho_i$ and $\|\tau_0+\cdots+\tau_i\|\le\|f^+\|$, we infer that $\|\varrho_{i-1}-\varrho_i\|\to 0$ and $\tau_0+\cdots+\tau_i\to\tau$; then, thanks to (12.10), it follows that τ transports exactly f^+ onto f^-. This measure τ is what we call *simplification* of η with respect to ϱ_0: we have the

Proposition 12.7. *If τ is the simplification of η, $\mu_\tau\le\mu_\eta$. Recalling that, as noted in Section 12.1, $\mathfrak{C}(\tau)=\|\mu_\tau\|$ and $\mathfrak{C}(\eta)=\|\mu_\eta\|$, it follows $\mathfrak{C}(\tau)\le\mathfrak{C}(\eta)$.*

Proof. The construction made before, together with Proposition 12.6, ensures that $\mu_\eta = \mu_{\tau_0+\cdots+\tau_i}+\mu_{\eta_i}$; then $\mu_{\tau_0+\cdots+\tau_i}$ is an increasing sequence of measures bounded by μ_η, and so $\mu_{\tau_0+\cdots+\tau_i}\to\mu_\infty$ with $\mu_\infty\le\mu_\eta$. We will conclude by proving that $\mu_\tau\le\mu_\infty$: in fact, given $0\le\varphi\in C(M)$, $\langle\mathscr{H}^1_j\llcorner\sigma,\varphi\rangle$ is a l.s.c. function of σ (as we will better discuss in Lemma 13.6), and then the convergence of $\tau_0+\cdots+\tau_i$ to τ and the definition (12.2) of μ imply $\langle\mu_\tau,\varphi\rangle\le\liminf\langle\mu_{\tau_0+\cdots+\tau_i},\varphi\rangle=\langle\mu_\infty,\varphi\rangle$. $\square$

Chapter 13

Main results

Here are the main results of this second Part of the thesis: we begin by proving that on a manifold of type 0 the extremals of the three problems are equal, then we study how is it possible to pass from an object of one of the three problems (that is, a transport, or a configuration, or a measure ν with $-\text{div }\nu = f$) to an object of another one of the problems. Finally we will state and prove our main results, namely that the problems on the manifolds of type 1 and the relaxed problems on the manifolds of type 2 have solutions, and that the corresponding optimal measures (the optimal transport densities, the optimal work densities and the optimal shapes) are the same; finally we will prove the no-crossing property of the transport rays in the strictly convex case.

13.1 Equality of the extremals

Here we generalize the first part of Theorem 11.2:

Theorem 13.1. *If M is a manifold of type 0,* $\inf(MK) = \inf(Br) = \inf(SO)$.

Proof. **Step I.** $\inf(MK) \geq \inf(Br)$.
Given the transport η, we define $f_t := \pi_{t\#}\eta$ and $\langle \overline{E}_t, \varphi\rangle := \int_S \varphi(\sigma(t)) \cdot \sigma'(t) d\eta$: it is easy to check that it is an admissible configuration and that its cost is less than $\mathfrak{C}(\eta)$ by Jensen inequality (we will make the explicit computations in Proposition 13.3).
Step II. $\inf(Br) \geq \inf(SO)$.
Given the configuration $\{f_t, \overline{E}_t\}$, let us define $\nu = -\int_0^1 \overline{E}_t$: from the compatibility condition (12.3) we infer that $-\text{div }\nu = f$; moreover, Jensen inequality immediately implies $\int j(x,\nu) \leq \mathfrak{C}\left(\{f_t, \overline{E}_t\}\right)$.
Step III. $\inf(SO) \geq \inf(MK)$.
We have already noted in Section 12.1 that inf(11.1)=inf(MK); let us then for a moment

consider the "old" transport problem: thanks to Theorem 11.1 and to the boundedness of d, we know that there exists an optimal "old" transport γ (while, as we discussed in the previous section, there could be no optimal "new" transport) and there exists a Kantorovich potential u; so

$$\sup(12.5) \geq \int\limits_M u\ df = \iint\limits_{M\times M} u(x) - u(y)\ d\gamma = \iint\limits_{M\times M} d(x,y)\ d\gamma = \inf(11.1) = \inf(MK),$$

recalling that $\pi_1\gamma = f^+$ and $\pi_2\gamma = f^-$: then this step follows from Proposition 12.3. $\square$

Corollary 13.2. *The statement of Proposition 12.3 can be strengthened by saying that* $\sup(12.5) = \inf(SO)$*: in fact, to prove the previous theorem we showed that* $\sup(12.5) \geq \inf(MK) \geq \inf(Br) \geq \inf(SO)$.

13.2 Passing from a problem to another

Here we will describe how, starting from an object of one of the problems, it is possible to build a "better" object for one of the other problems. Through this section, we assume to deal with a manifold of type 1 (so M, and consequently S, are complete). Let us begin with the first result.

Proposition 13.3. *For each transport η there exists a configuration $\{f_t, \overline{E}_t\}$ such that $\tilde{\mu} \leq \mu$ and then $\mathfrak{C}\left(\{f_t, \overline{E}_t\}\right) \leq \mathfrak{C}(\eta)$, where μ and $\tilde{\mu}$ are the transport density and the work density associated to η and $\{f_t, \overline{E}_t\}$ respectively.*

Proof. First of all, given η we define $\{f_t, \overline{E}_t\}$ as

$$f_t := \pi_{t\#}\eta, \qquad\qquad \langle \overline{E}_t, \varphi\rangle := \int_S \varphi(\sigma(t))\cdot\sigma'(t)\ d\eta(\sigma);$$

$\{f_t, \overline{E}_t\}$ is admissible: in fact, $f_0 = f^+$ and $f_1 = f^-$ because η is a transport; moreover, for any $\varphi \in \mathrm{C}^\infty_b(M)$ it is

$$\begin{aligned}\langle \dot{f}_t, \varphi\rangle &= \frac{d}{dt}\int_M \varphi(x)\ df_t(x) = \frac{d}{dt}\int_S \varphi(\sigma(t))\ d\eta(\sigma) = \int_S D\varphi(\sigma(t))\cdot\sigma'(t)\ d\eta(\sigma)\\ &= \langle \overline{E}_t, D\varphi\rangle = -\langle \operatorname{div}\ \overline{E}_t, \varphi\rangle\end{aligned}$$

and then (12.3) holds. Moreover, let us denote by $\eta = \eta_x \otimes \pi_{t\#}\eta$ the disintegration of η with respect to $\pi_t : S \longrightarrow M$, in the sense of Theorem A.15, and let $h(x) = \int_S \sigma'(t)\ d\eta_x(\sigma)$; then $\overline{E}_t = hf_t$, since

$$\begin{aligned}\langle \overline{E}_t, \varphi\rangle &= \int_S \varphi(\sigma(t))\cdot\sigma'(t)\ d\eta(\sigma) = \int_M \left(\int_S \varphi(\sigma(t))\cdot\sigma'(t)\ d\eta_x(\sigma)\right) d\pi_{t\#}\eta(x)\\ &= \int_M \varphi(x)\cdot\left(\int_S \sigma'(t)\ d\eta_x(\sigma)\right) df_t(x) = \int_M \varphi(x)\cdot h(x)\ df_t(x) = \langle hf_t, \varphi\rangle\end{aligned}$$

(recall that $\sigma(t) = \pi_t(\sigma)$ is equal to x for η_x−a.e. σ). Finally, we can prove $\tilde{\mu} \leq \mu$: in fact, if $0 \leq \varphi \in \mathrm{C}_b(M)$ we have

$$\begin{aligned}
\langle \tilde{\mu}, \varphi \rangle &= \int_0^1 \int_M \varphi(x) j(x, d\overline{E}_t)\ dt = \int_0^1 \int_M \varphi(x) j(x, h(x))\ df_t(x)\ dt \\
&= \int_0^1 \int_M \varphi(x) j \left(x, \int_S \sigma'(t)\ d\eta_x(\sigma) \right) df_t(x)\ dt \\
&\leq \int_0^1 \int_M \varphi(x) \left(\int_S j(x, \sigma'(t)) d\eta_x(\sigma) \right) df_t(x)\ dt \\
&= \int_0^1 \int_S \varphi(\sigma(t)) j\big(\sigma(t), \sigma'(t)\big)\ d\eta(\sigma)\ dt = \int_S \langle \mathscr{H}^1_j \llcorner \sigma, \varphi \rangle\ d\eta(\sigma) = \langle \mu, \varphi \rangle
\end{aligned}$$

recalling (12.4), Jensen inequality, the properties of disintegration described in Theorem A.15, (12.1) and (12.2). This gives $\tilde{\mu} \leq \mu$ and, since $\mathfrak{C}\left(\left\{f_t, \overline{E}_t\right\}\right) = \langle \tilde{\mu}, 1 \rangle$ and $\mathfrak{C}(\eta) = \langle \mu, 1 \rangle$, the proof is completed. □

Remark 13.4. *In the previous proposition $\left\{f_t, \overline{E}_t\right\}$ has been built following exactly the paths used by the measure η, so one could think that $\mathfrak{C}(\eta) = \mathfrak{C}\left(\left\{f_t, \overline{E}_t\right\}\right)$, while the inequality we proved can also hold strictly. The reason of this fact relies essentially on the main difference between the configurations and the transports: in the first case, the mass is moved "effectively", that is on the space M, while in the second case the mass is moved "in theory", that is on the space S. To be more precise, the different paths along which the mass is moved are covered independently by a transport, while they interact when covered by a configuration. And, as the preceding proposition shows, this interaction lowers the cost thanks to the convexity of j; and in case of strict convexity, also the inequality in the preceding proposition can be strict.*

Proposition 13.5. *For each configuration $\left\{f_t, \overline{E}_t\right\}$ there exists ν such that $-\mathrm{div}\ \nu = f$, the shape $j(x, \nu)$ is less than the work density μ and then $\int j(x, \nu) \leq \mathfrak{C}\left(\left\{f_t, \overline{E}_t\right\}\right)$; on the other hand, to each ν such that $-\mathrm{div}\ \nu = f$ we can associate a configuration $\left\{f_t, \overline{E}_t\right\}$ such that $\mu = j(x, \nu)$ and then $\mathfrak{C}\left(\left\{f_t, \overline{E}_t\right\}\right) = \int j(x, \nu)$.*

Proof. Given $\left\{f_t, \overline{E}_t\right\}$, we define $\nu = -\int_0^1 \overline{E}_t\ dt$; the fact that $-\mathrm{div}\ \nu = f$ follows from the compatibility condition (12.3), while $j(x, \nu) \leq \mu$ comes directly from Jensen inequality; then $\int j(x, \nu) \leq \int \mu = \mathfrak{C}\left(\left\{f_t, \overline{E}_t\right\}\right)$.

On the other hand, given ν let us define

$$f_t := (1-t) f^+ + t f^- \qquad\qquad \overline{E}_t := -\nu;$$

the compatibility condition holds since $\mathrm{div}\ \overline{E}_t = f = -df_t/dt$; moreover one has $\mu = \int_0^1 j(x, \overline{E}_t) = j(x, \nu)$; then the thesis follows. □

To finish our comparisons, we should pass from a shape to a transport and this will be the most difficult step: first of all, we need the following

Lemma 13.6. *If $\{\eta_i\}$ is a sequence of retracted transports between f_i^+ and f_i^- with $\mathfrak{C}(\eta_i) \leq H < +\infty$ and $\|\eta_i\|_{\mathcal{M}^+(S)} \leq H$, there exists η such that, up to a subsequence, $\eta_i \xrightarrow[\mathcal{M}^+(S)]{*} \eta$ and $\mu \leq \liminf \mu_i$ (i.e. $\langle \mu, \varphi \rangle \leq \liminf \langle \mu_i, \varphi \rangle$ for any $\varphi \geq 0$), where μ and μ_i are the transport densities relative to η and η_i.*

Proof. Let us define

$$\eta_{i,k} := \begin{cases} \eta_i \llcorner \{\sigma \in S: \ l_M(\sigma) < 1\} & \text{if } k = 0 \\ \eta_i \llcorner \{\sigma \in S: \ 2^{k-1} \leq l_M(\sigma) < 2^k\} & \text{otherwise} \end{cases}$$

and note that, for each $k \in \mathbb{N}$, $\{\eta_{i,k}\}$ is, varying i, a sequence of measures on $S_k = \{\sigma \in S: l_M(\sigma) \leq 2^k\}$, which is easily seen to be complete and totally bounded and then a compact metric space (notice that, also recalling (11.8), (M, d_M) is a complete and bounded Riemannian manifold, thus it is compact): then, since $\|\eta_{i,k}\| \leq \|\eta_i\| \leq H$, we find $\eta_k \in \mathcal{M}^+(S)$ such that, up to a (not relabeled) subsequence, $\eta_{i,k} \xrightarrow[\mathcal{M}^+(S_k)]{*} \eta_k$. Moreover for each $k \geq 1$ we have, recalling (11.8), $\mathfrak{C}(\eta_{i,k}) \geq (2^{k-1} - C)\, \alpha^{-1} \|\eta_{i,k}\|$, and then

$$\|\eta_{i,k}\|_{\mathcal{M}^+(S)} \leq \frac{H\alpha}{2^{k-1} - C}, \qquad \|\eta_k\|_{\mathcal{M}^+(S)} \leq \frac{H\alpha}{2^{k-1} - C},$$

where the second inequality follows by the first one with a limit for $i \to +\infty$: it is then well defined $\eta = \sum \eta_k$ in $\mathcal{M}^+(S)$. Moreover we have $\eta_i \xrightarrow{*} \eta$: in fact, given $\varphi \in C(S)$, we have that $\langle \eta_{i,0} + \cdots + \eta_{i,k}, \varphi \rangle \to \langle \eta_0 + \cdots + \eta_k, \varphi \rangle$ and that the two terms differs from $\langle \eta_i, \varphi \rangle$ and $\langle \eta, \varphi \rangle$ less then $H\alpha \|\varphi\| \sum_{l=k}^{+\infty} \left(2^{l-1} - C\right)^{-1}$. Note that this also implies $f_i^+ \xrightarrow{*} f^+$, $f_i^- \xrightarrow{*} f^-$ and that η is a transport between f^+ and f^-.

To finish the proof, we have to check that $\mu \leq \liminf \mu_i$; first we note that, given $0 \leq \varphi \in C(M)$, $\langle \mathscr{H}^1_j \llcorner \sigma, \varphi \rangle$ is a l.s.c. function of σ: in fact, we know that $\int_0^1 j(\sigma, \sigma')\, dt = \liminf_{\tau \to \sigma} \int_0^1 j(\tau, tau')\, dt$ for each $\sigma \in S$ because j is l.s.c. by hypothesis; then, since φ is continuous and the image of σ compact, we have also

$$\int_0^1 \varphi(\sigma) j(\sigma, \sigma')\ dt = \liminf_{\tau \to \sigma} \int_0^1 \varphi(\tau) j(\tau, \tau')\ dt;$$

then the searched lower semicontinuity of $\langle \mathscr{H}^1_j \llcorner \sigma, \varphi \rangle$ follows directly by the definition (12.1) of $\mathscr{H}^1_j \llcorner \sigma$; since $\eta_i \xrightarrow{*} \eta$, (12.2) ensures that $\langle \mu, \varphi \rangle \leq \liminf \langle \mu_i, \varphi \rangle$. □

Then let us prove the

Proposition 13.7. *For each ν such that $-\mathrm{div}\ \nu = f$ there exists a transport η such that $\mu_\eta \leq j(x, \nu)$ and then $\mathfrak{C}(\eta) \leq \int j(x, \nu)$, where μ_η is the transport density associated to η.*

Proof. **Step I.** *The smooth case.*
Let us first suppose that f and ν are smooth and concentrated on $M \setminus \partial M$; then, following the construction made in Theorem 4.2 of [1] (reminding also the construction in [39]), we prove the thesis: first of all, chosen $\varepsilon > 0$ and writing $\omega_t(x) = tf^-(x) + (1-t)f^+(x) + \varepsilon$, we define the smooth vector field

$$\overline{v}_t(x) := \frac{\nu(x)}{\omega_t(x)},$$

which suggests how to move the mass; in other words, we define the smooth flow $\Psi : [0,1] \times M \longrightarrow M$ as

$$\begin{cases} \Psi(0,x) = x \\ \dfrac{d}{dt}\Psi(t,x) = \overline{v}_t(\Psi(t,x)) \end{cases}$$

and consider the transport $\tilde{\eta} = \xi_{\#}\omega_0$, where $\xi(x)(s) = \Psi(s,x)$. We note then that

$$\Psi(t,\cdot)_{\#}\omega_0 = \omega_t \ : \tag{13.1}$$

in fact, both the measures are easily seen to solve the ODE

$$\begin{cases} \zeta(0,\cdot) = \omega_0 \\ \dot{\zeta} = -\mathrm{div}\,(\overline{v}_t\zeta), \end{cases}$$

and then, all being smooth, equation (13.1) follows from the uniqueness of the solution (see for example [34, 36]). Moreover we have that, also thanks to (13.1),

$$\begin{aligned} \langle \mu_{\tilde{\eta}}, \varphi \rangle &= \int_S \langle \mathscr{H}^1_j \llcorner \sigma, \varphi \rangle \, d\tilde{\eta}(\sigma) = \int_M \langle \mathscr{H}^1_j \llcorner \xi(x), \varphi \rangle \, d\omega_0(x) \\ &= \int_M \int_0^1 \varphi\big(\Psi(t,x)\big) j\big(\Psi(t,x), \overline{v}_t(\Psi(t,x))\big) \, dt \, d\omega_0(x) \\ &= \int_0^1 \int_M \varphi(x) j(x, \overline{v}_t(x)) \, d\omega_t(x) \, dt = \int_0^1 \langle j(x,\nu), \varphi \rangle \, dt = \langle j(x,\nu), \varphi \rangle. \end{aligned}$$

Note now that $\tilde{\eta}$ is a transport that moves $\omega_0 = f^+ + \varepsilon$ onto $\omega_1 = f^- + \varepsilon$, and then the thesis follows from Proposition 12.7 calling η the simplification of $\tilde{\eta}$.

Step II. *The case when j is Lipschitz.*
Let us now suppose that $j(\cdot,\omega)$ is Lipschitz for $|\omega| \leq 1$ uniformly on ω: this means that there exists a constant L such that whenever $|\omega| \leq 1$ the function $j(\cdot,\omega)$ is Lipschitz with constant L (since $j(x,\cdot)$ is $1-$homogeneous, it follows that in general $j(\cdot,\omega)$ is Lipschitz with constant $L|\omega|$). Define now $\nu_{i,\varepsilon}$ and ν_ε as in (B.1), and recall that they are smooth

and supported in $M \setminus \partial M$ as shown in Theorem B.2. Take now $0 \leq \varphi \in \mathrm{C}_b(M)$; fixed i and making for simplicity the calculation in $\mathbb{R}^N$ instead of in U_i, we have

$$\nu_{\varepsilon,i}(x) = \int \rho_{\varepsilon/2}(x-y)\, d\nu_i\big((1-\varepsilon)y\big) = \int \rho_{\varepsilon/2}(x-\tilde{y})\, d\nu_i(y),$$

writing $\tilde{y} = y/(1-\varepsilon)$; set also $\overline{\omega}$ such that $\nu_i = \overline{\omega}|\nu_i|$: since $|\overline{\omega}(y)| = 1$ and thanks to the Lipschitz assumption on j, for each δ there exists $\bar{\varepsilon}$ such that $j(x,\omega(y)) \leq j(y,\omega(y)) + \delta$ if $\varepsilon \leq \bar{\varepsilon}$ and $\rho_{\varepsilon/2}(x-\tilde{y}) > 0$. Then, using the fact that $j(x,\cdot)$ is $1-$homogeneous and Jensen inequality, we have

$$\begin{aligned}
\langle j(x,\nu_{\varepsilon,i}),\varphi\rangle &= \int \varphi(x)\, j(x,\nu_{\varepsilon,i}(x))\, dx = \int \varphi(x)\, j\left(x, \int \rho_{\varepsilon/2}(x-\tilde{y})\, d\nu_i(y)\right) dx \\
&\leq \int \varphi(x) \int j\Big(x, \rho_{\varepsilon/2}(x-\tilde{y})\, d\nu_i(y)\Big)\, dx \\
&= \int\!\!\int \varphi(x)\rho_{\varepsilon/2}(x-\tilde{y}) j\left(x,\overline{\omega}(y)\right)\, d|\nu_i|(y)\, dx \\
&\leq \int\!\!\int \varphi(x)\rho_{\varepsilon/2}(x-\tilde{y}) j\left(y,\overline{\omega}(y)\right)\, d|\nu_i|(y)\, dx + \delta\, \|\varphi\|\, \|\nu_i\| \\
&= \int \varphi_\varepsilon(y) j(y,\nu_i(y)) + \delta\, \|\varphi\|\, \|\nu_i\| = \langle j(x,\nu_i),\varphi_\varepsilon\rangle + \delta\, \|\varphi\|\, \|\nu_i\|,
\end{aligned}$$

writing $\varphi_\varepsilon(y) = \varphi * \rho_{\varepsilon/2}(\tilde{y})$. Since clearly $\langle j(x,\nu_i),\varphi_\varepsilon\rangle \to \langle j(x,\nu_i),\varphi\rangle$, recalling that δ was arbitrary and arguing as in Theorem B.2 to pass from an estimate in a single chart to a global estimate, we deduce $\liminf j(x,\nu_\varepsilon) \leq j(x,\nu)$. Now, using the first Step and denoting $f_\varepsilon := -\mathrm{div}\ \nu_\varepsilon$, to each ν_ε we can associate a transport $\tilde{\eta}_\varepsilon$ which moves f_ε^+ onto f_ε^- with $\mu_{\tilde{\eta}_\varepsilon} \leq j(x,\nu_\varepsilon)$ and moreover, written $\eta_\varepsilon = ret_{\#}\tilde{\eta}_\varepsilon$, thanks to hypothesis *1-c)* of Section 11.2 we know that $\mu_{\eta_\varepsilon} \leq \mu_{\tilde{\eta}_\varepsilon}$.

Since $\liminf j(x,\nu_\varepsilon) \leq j(x,\nu)$ we know that, up to a subsequence, $\mathfrak{C}(\eta_\varepsilon) \leq H$; moreover, $f_\varepsilon \overset{*}{\rightharpoonup} f$ thanks to Theorem B.2 and then $\|\eta_\varepsilon\| = \|\tilde{\eta}_\varepsilon\| = \|f_\varepsilon^+\| \leq \|f_\varepsilon\| \leq H$: so we can apply Lemma 13.6 to the sequence η_ε obtaining η, that is a transport between f^+ and f^-. Now, $\mu_\eta \leq \liminf \mu_{\eta_\varepsilon} \leq \liminf j(x,\nu_\varepsilon) \leq j(x,\nu)$, so the thesis follows.

Step III. *The general case.*
We have now only to remove the Lipschitz assumption on j: first of all note that the construction of the preceding step allows, given ν, to build a transport η which does not depend on j; then, since j is l.s.c., we can choose $j_\varepsilon \nearrow j$ with j_ε Lipschitz and we already know that

$$\mu_\varepsilon := \int_S \mathscr{H}^1_{j_\varepsilon} \llcorner \sigma\, d\eta(\sigma) \leq j_\varepsilon(x,\nu) \leq j(x,\nu):$$

then the thesis will follow by the fact that $\mu_\varepsilon \overset{*}{\rightharpoonup} \mu$; finally, this fact is easily proved by applying twice the Monotone Convergence Theorem (first to an integral in $[0,1]$ and then to an integral in S). $\square$

13.3 Existence of extremals and equivalence between the optimal measures

We prove here the main results of this paper, which generalize Theorem 11.2 for the manifolds of type 1 and 2, in the second case in the relaxed sense. First of all we have the

Theorem 13.8. *If M is a manifold of type 1, there exist extremals for problems (MK), (Br) and (SO). Moreover, the associated optimal transport densities, work densities and shapes are exactly the same.*

Remark 13.9. *As discussed after the statement of Theorem 11.2, in general there is no uniqueness of the optimal transport densities, work densities and shapes (but see Theorem 16.2 for a uniqueness result): however, as the Theorem above claims, each measure $\mu \in \mathcal{M}^+(M)$ is an optimal transport density if and only if it is an optimal work density and if and only if it is an optimal shape.*

Proof. *(of Theorem 13.8):* Let us begin taking an optimal sequence of transports $\{\eta_i\}$ between f^+ and f^-, which we can assume to be retracted since the retraction lowers the costs: thanks to Lemma 13.6, which we can apply since for each i we have $\|\eta_i\| = \|f^+\|$, we get a transport η, and this is also a transport between f^+ and f^- since $\eta_i \overset{*}{\rightharpoonup} \eta$. But η is an optimal transport because $\{\eta_i\}$ was an optimal sequence and $\mathfrak{C}(\eta) = \langle \mu_\eta, 1\rangle \leq \liminf \langle \mu_{\eta_i}, 1\rangle = \liminf \mathfrak{C}(\eta_i)$: so we proved that the Monge–Kantorovich problem has a solution. Then, all we stated immediately follows by Propositions 13.3, 13.5 and 13.7. $\square$

Then we prove that the extension of the problems we defined for the manifolds of type 2 is a "good" relaxation

Theorem 13.10. *The problems defined in Section 12.4 are a standard relaxation in the case when M is a manifold of type 2. This means*

- *the relaxed problems have solutions;*
- *the optimal measures associated to the different problems are the same;*
- *the extremals $\inf(MK)$, $\inf(Br)$ and $\inf(SO)$ of the original problems do not change with the relaxation;*
- *each relaxed solution (i.e. each optimal relaxed transport or configuration or measure ν with $-\mathrm{div}\ \nu = f$) is a weak* limit of non-relaxed transports, or configurations, or measures, such that also the associated measures (transport densities, work densities, shapes) weak* converge.*

Proof. The first fact comes directly from Theorem 13.8, used on the manifold $\hat{M}$ which, by hypothesis *2-a*, is a manifold of type 1. To prove the other asserts, let us first define the functions $\tau_i : \hat{S} \longrightarrow S$ such that $|l(\sigma) - l(\tau_i(\sigma))| \leq 2^{-i}$, $d_S(\sigma, \tau_i(\sigma)) \leq 2^{-i}$ and, if both the extremals of $\sigma \in \hat{S}$ are in M, then they are the same of $\tau_i(\sigma)$: this can be done thanks to hypothesis *2-b*. If we define now $\eta_i := \tau_{i\#}\eta$, it follows that the η_i are non-relaxed transports; moreover, by the definition of τ_i, it holds the inequality $\mathfrak{C}(\eta_i) \leq \mathfrak{C}(\eta) + 2^{-i}\|\eta\|$: this ensures that the extremals do not decrease by relaxation and, since they clearly cannot have been increased (recall that the relaxed items are more than the non-relaxed ones), also the second assert follows. In our construction, we have also that $\eta_i \overset{*}{\rightharpoonup} \eta$ and $\mu_{\eta_i} \overset{*}{\rightharpoonup} \mu_\eta$; to show the last convergence, note that by construction and arguing by lower semicontinuity of $\sigma \to \langle \mathscr{H}^1_j \llcorner \sigma, \varphi \rangle$ as in Proposition 13.6, one obtains $\mu_\eta \leq \liminf \mu_{\eta_i}$; on the other hand $\mathfrak{C}(\eta) = \lim \mathfrak{C}(\eta_i)$ implies $\|\mu_\eta\| = \lim \|\mu_{\eta_i}\|$ and then the stated convergence follows. So the last assert is now proved for the Monge–Kantorovich problem. To prove it for the other two problems it suffices to use again Propositions 13.3, 13.5 and 13.7. $\square$

Note that to prove the above Theorems we have given some constructions that allow, starting from a solution of a problem, to build a solution of one of the other problems preserving the optimal measure.

Remark 13.11. *We can point out that also* $\sup(12.5)$ *is attained, both in the manifolds of type 1 and 2 (since it was the only problem we did not need to relax in section 12.4): in fact, using Theorem 11.1 we can take a Kantorovich potential u associated to an optimal transport and, recalling Step 3 in Theorem 13.1, it follows that u is optimal for (12.5).*

Remark 13.12. *In [30] it was left open the question whether or not each optimal shape on a Riemannian manifold can be obtained by integrating the $\mathscr{H}^1$ measures on a family of geodesics. Thanks to the theorems we proved, we can give an affirmative answer to this question, also in our more general context.*

13.4 No-crossing property of transport rays

In this section, we discuss the generalization of the no-crossing property of the transport rays which holds in the Euclidean case; through this section, M will be a manifold of type 1: then the same results hold on $\hat{M}$ if M is a manifold of type 2.

It is a well known fact that, in the Euclidean case, the transport rays do not intersect: the no-crossing property (4.1) has been of primary importance through the whole first Part of this thesis; recalling the equivalence between transports and Brenier configurations we proved with Propositions 13.3, 13.5 and 13.7 and recalling Remark 12.2, we see that this can be expressed with the equation

$$\int_0^1 \left|\overline{E}_t\right| \; dt = \left| \int_0^1 \overline{E}_t \; dt \right|, \tag{13.2}$$

that holds for each optimal Brenier configuration and can be also found for example in [1]. It is easy to understand that the no-crossing property can not be generalized in our setting without an hypothesis of strict convexity of j: for example, if in $\mathbb{R}^2$ we set $j(x,v) = v_1$, the horizontal component of a vector, and consider the problem to translate horizontally an homogeneous vertical block, it is clear that it can be done in many optimal ways also with a great intersection of rays. More in general, in the first Part of the thesis we saw that the no-crossing property may fail also for a non-strictly convex norm. However, let us note the following

Lemma 13.13. *If $\{f_t, \overline{E}_t\}$ is optimal, then*

$$\int_0^1 j(x, \overline{E}_t) \; dt = j\left(x, \int_0^1 \overline{E}_t \; dt\right). \tag{13.3}$$

Proof. This is an immediate fact; defined $\nu = -\int_0^1 \overline{E}_t$ and calling μ the work density relative to $\{f_t, \overline{E}_t\}$, we have $j(x,\nu) = \mu$: in fact, $j(x,\nu) \leq \mu$ as in Proposition 13.5, and on the other hand the inequality cannot hold strictly since $\{f_t, \overline{E}_t\}$ is optimal. Recalling (12.4), this equality is exactly (13.3) and then the thesis is proved. ☐

Clearly (13.3) is the perfect parallel to (13.2); let us see how this can be read as a no-crossing result in the strictly convex case: first we recall that, being j 1−homogeneous, we call it strictly convex on T_xM if $j\big(tv + (1-t)w\big) \leq tj(x,v) + (1-t)j(x,w)$ for each v, $w \in T_xM$ and $t \in (0,1)$ and the equality holds only if v and w are parallel. Then the hypothesis of strict convexity allows to prove the

Theorem 13.14 (No-crossing property of transport rays). *If $\{f_t, \overline{E}_t\}$ is optimal and j is strictly convex on each T_xM, there are vectors $h(x) \in T_xM$ such that $\overline{E}_t(x)$ is parallel to $h(x)$ for a.e. (x,t) (with respect to the measure $|\overline{E}_t| \otimes dt$).*

Proof. First of all, let us write as usual $\overline{E}_t = \omega_t \, d|\overline{E}_t|$, and define $\alpha := |\overline{E}_t| \otimes dt$, a positive measure on $M \times [0,1]$; moreover, we call π the projection of $M \times [0,1]$ on M and operate a disintegration of α with respect to π, finding the measures α_x on $[0,1]$ –to be precise, the α_x are measures on $M \times [0,1]$ concentrated on the couples (y,t) with $y = x$, and then we consider them as measures on $[0,1]$. Setting now

$$h(x) := \int_0^1 \omega_t(x) \, d\alpha_x(t) \qquad\qquad \xi := \pi_\# \alpha, \tag{13.4}$$

we have $\displaystyle\int_0^1 \overline{E}_t = h\,d\xi$, since for each smooth vector field φ it is

$$\begin{aligned}\langle\int_0^1 \overline{E}_t, \varphi\rangle &= \int_0^1 \int_M \varphi(x)\cdot d\overline{E}_t(x)\,dt = \iint_{M\times[0,1]} \varphi(x)\cdot\omega_t(x)\; d\big(|\overline{E}_t|\otimes dt\big)(x,t)\\ &= \int_M \left(\int_0^1 \varphi(x)\cdot\omega_t(x)\,d\alpha_x(t)\right) d\pi_\#\alpha(x) = \int_M \varphi(x)\cdot h(x)\,d\xi = \langle h\,d\xi, \varphi\rangle,\end{aligned}$$

recalling the properties of disintegration; then we can evaluate

$$\begin{aligned}\int_M j\left(x, \int_0^1 \overline{E}_t\,dt\right) &= \int_M j(x, h(x))\,d\xi(x) = \int_M j\left(x, \int_0^1 \omega_t(x)\,d\alpha_x(t)\right)\,d\xi(x)\\ &\le \int_M \int_0^1 j(x,\omega_t(x))\,d\alpha_x(t)\,d\xi(x) = \int_M \int_0^1 j(x,\omega_t(x))\,d\alpha(x,t)\\ &= \int_0^1 \int_M j(x,\omega_t(x))\,d|\overline{E}_t|(x)\,dt = \int_0^1 \int_M j(x,\overline{E}_t)\,dt,\end{aligned}$$

using Jensen inequality. But thanks to (13.3), the inequality must be an equivalence, and using the strict convexity of j it follows that the unit vectors $\omega_t(x)$ are equal, varying t; to be more precise, for $\xi-$a.e. x the vectors $\omega_t(x)$ equal $h(x)$ for α_x-a.e. t. Recalling that $\overline{E}_t = \omega_t|\overline{E}_t|$ and (13.4), the thesis follows. □

Remark 13.15. *Note that, in presence of a Dirichlet region Σ, the thesis of the theorem cannot be true: since moving in Σ has no cost, there can be intersecting paths also in optimal transports; in fact, if $x \in \Sigma$ then $j(x,\cdot) \equiv 0$ and then the strict convexity fails. However, the assert of the previous Theorem can clearly be strengthened by saying that the different transport rays cannot cross in the points x such that j is strictly convex on T_xM or, equivalently, that for each x such that $j(x,\cdot)$ is strictly convex there exists a vector $h(x)$ on T_xM such that $\overline{E}_t(x)$ is parallel to $h(x)$ for a.e. t (with respect to α_x).*

Part III

Properties of the transport density

Chapter 14

An introductive discussion about the transport density

14.1 Introduction

This third Part of the thesis is entirely devoted to the study of the regularity of the transport density, a measure that we already encountered both in the first and in the second Part. The interest about this measure relies on many elements: among them, we recall the fact that it played a crucial role in the construction of an optimal transport map by Evans and Gangbo (that we briefly discussed in Section 6.4), and the fact that it is a key ingredient in establishing the equivalence between the three problems that we studied in the second Part. Thanks to this equivalence, all the regularity properties that we will prove about the transport density will be also true for the optimal shapes (recall the definitions we gave in Section 11.1 and we generalized in Section 12.3).

In this chapter we will first define and discuss a notion of dimension for positive measures; then we will give some important remarks, discussions and examples about the transport density; finally we will prove some technical geometrical facts. In the next chapters, we will prove sharp estimates about the dimension and the L^p properties of the transport density.

Throughout this third Part, the ambient space $X = Y$ will be a bounded and convex subset of $\mathbb{R}^N$, whose diameter will be denoted by R, and the cost function will be the Euclidean one, $c(x, y) = |y - x|$ (about the hypothesis of boundedness, see also the considerations we do in Section 14.5); given then two positive measures f^+ and f^- of equal total mass, we define the transport density (or the "optimal transport density" in the language of the second Part) as follows:

Definition 14.1. *Given any optimal transport plan γ between f^+ and f^-, we define* transport density *the measure $\mu \in \mathcal{M}^+(X)$ (depending on γ) defined by*

$$\langle \mu, \varphi \rangle := \iint_{X\times X} \left(\int_X \varphi(z)\, d\mathscr{H}^1_{xy}(z) \right) d\gamma(x,y) \qquad \forall\, \varphi \in \mathrm{C}_b(X), \tag{14.1}$$

where $\mathscr{H}^1_{xy}$ denotes the restriction of the one-dimensional Hausdorff measure $\mathscr{H}^1$ to the closed segment xy.

Let us immediately discuss the meaning of this measure: if γ is a transport plan, from the definition it follows that for any set $A \in \mathscr{B}(X)$ the number $\mu(A)$ is the mean (w.r.t. γ) length of the intersection between A and the closed segments xy. From a more expressive point of view, we can regard $\mu(A)$ as the total work done inside the set A while effectively making the transport corresponding to γ.

If we denote by $l(xy \cap A) = \mathscr{H}^1_{xy}(A)$ the length of the intersection between the closed segment xy and the Borel set $A \subseteq X$, we can derive from (14.1) an expressive formula for the measure of A:

$$\mu(A) = \iint_{X\times X} l(xy \cap A)\, d\gamma(x,y). \tag{14.2}$$

Note that in principle the transport density does not depend uniquely on f^+ and f^- (see for instance Example 16.1), but also on the choice of an optimal transport plan γ: anyway, if at least one between f^+ and f^- is absolutely continuous with respect to the Lebesgue measure $\mathscr{L}$ then the transport density is unique, i.e. any optimal transport plan leads to the same measure; we will prove this fact in Theorem 16.2. However, all the regularity results we will prove will depend exclusively on f^+ and f^-: this means that we will prove that, under certain regularity assumptions on f^+ and f^-, some regularity properties will be true for any transport density, whichever is the optimal transport plan γ from which it comes.

Throughout this third Part, differently from what we did before, we will for simplicity denote by xy, instead of $[\![x,y]\!]$, the closed segment connecting x and y, and its length will be denoted by $\overline{xy}$.

We conclude this section remarking an easy but interesting fact:

Remark 14.2. *For any transport density μ the norm $\|\mu\|$ equals the minimum of the Kantorovich problem: in fact, if μ is defined through the optimal transport plan γ, then by (14.2) it immediately follows that $\|\mu\| = \mu(X) = \iint |y-x|\, d\gamma = \mathfrak{C}(\gamma)$: however, this property is true in a much more general setting, in fact it was already noticed in Section 12.1 in the greater generality of the second Part of the thesis.*

14.2 The dimension of a measure

In this section we introduce a notion of "dimension" for the measures belonging to $\mathcal{M}^+(X)$, then we discuss a couple of possible equivalent definitions and some properties about the behavior of the dimension under the action of Lipschitz maps. First of all, we give the

Definition 14.3. *The* dimension *of* $\mu \in \mathcal{M}^+(X)$ *is defined by*

$$\mathscr{H}-\dim\,\mu := \sup\{k:\ \mu \ll \mathscr{H}^k\}.$$

Clearly, the dimension of any measure on X is a number between 0 and N; moreover, if B is a $k-$dimensional set and $\mu = \mathscr{H}^k \llcorner B$, then $\mathscr{H}-\dim\,\mu = k$; note carefully that, more generally, if μ is made of pieces of different Hausdorff dimensions then $\mathscr{H}-\dim\,\mu$ is the smallest of these: for example, if $x \in X$, A is an open subset of X and $\mu = \delta_x + \mathscr{L} \llcorner A$, then $\mathscr{H}-\dim\,\mu = 0$.

The above described connection between the Hausdorff dimension of a measure and the standard Hausdorff dimension of sets can be better precised remarking that the dimension of a measure is in fact the least of the dimensions of sets of positive measure:

Proposition 14.4. *The following equivalence holds:*

$$\mathscr{H}-\dim\,\mu = \inf\left\{k:\ \exists\, B \subseteq X,\ \mu(B) > 0,\ \mathscr{H}^k(B) < +\infty\right\}. \tag{14.3}$$

Proof. Take a set $B \in X$ such that $\mu(B) > 0$ and $\mathscr{H}^k(B) < +\infty$: then for any $h > k$ one has $\mathscr{H}^h(B) = 0$ and $\mu(B) > 0$, then $\mu \not\ll \mathscr{H}^h$, and an inequality follows.

Concerning the other one, take $k \in \mathbb{R}$ and assume that $\nexists B \subseteq X$ such that $\mathscr{H}^k(B) < +\infty$ and $\mu(B) > 0$: then, in particular, $\mathscr{H}^k(B) = 0 \Rightarrow \mu(B) = 0$, thus $\mu \ll \mathscr{H}^k$ and the other inequality is proved. $\square$

From the proof of last proposition it follows that at least one of the extremals in (14.3) is reached; however, it may happen either that only the first one is reached, or only the second one, or both: examples to show this claim can be easily found wherever the Hausdorff measure is introduced, see for example [41, 40, 70].

In order to give another useful characterization of the dimension, we first need another standard definition:

Definition 14.5. *Given* $\mu \in \mathcal{M}^+(X)$ *and* $x \in X$*, the* $k-$upper density *of* μ *at* x *is defined by*

$$\theta^*_k(\mu, x) := \limsup_{\rho\to 0} \frac{\mu\big(B(x,\rho)\big)}{\omega_k \rho^k},$$

where ω_k is, as usual, the constant used to define the Hausdorff measure (for example, the volume of the unit ball in $\mathbb{R}^k$ if k is an integer).

The number $\theta^*_k(\mu,x)$ can be intuitively thought as a mean $\mathscr{H}^k-$dimensional density around x: to give an example, if $\mu = \mathscr{H}^\alpha \llcorner P$ and P is an $\alpha-$dimensional plane containing x, then

$$\theta^*_k(\mu,x) = \begin{cases} +\infty & \text{if } \alpha < k \\ 1 & \text{if } \alpha = k \\ 0 & \text{if } \alpha > k \end{cases} .$$

According to this definition, one can intuitively imagine a "pointwise dimension" of a measure μ in a point x, which is the supremum of the numbers k such that $\theta^*_k(\mu,x) < +\infty$. The dimension of the measure μ turns out to be the $\mu-$essential infimum of the pointwise dimension, as we will show in a moment. First, we recall some known properties, which are a particular case of Theorem 3.2 in [70]:

Theorem 14.6. *a) If $\theta^*_k(\mu,x) \le t$ on $B \in \mathscr{B}(X)$, then $\mu \llcorner B \le 2^k t\, \mathscr{H}^k \llcorner B$;*
b) if $\theta^*_k(\mu,x) \ge t$ on $B \in \mathscr{B}(X)$, then $\mu(X) \ge t\, \mathscr{H}^k(B)$.

Corollary 14.7. $\theta^*_k(\mu,x) < +\infty$ $\mathscr{H}^k-$a.e. .

Proof. Let $B = \{x \in X : \ \theta^*_k(\mu,x) = +\infty\}$: then in particular for any $t \in \mathbb{R}^+$ one has $\theta^*_k(\mu,x) \ge t$ on B, then by Theorem 14.6 b) one obtains $\mathscr{H}^k(B) \le \mu(X)/t$. Since $\mu(X) < +\infty$, the thesis follows. □

Note carefully that last Corollary ensures the finiteness of $\theta^*_k(\mu,x)$ almost everywhere with respect to $\mathscr{H}^k$, but of course not with respect to μ: if for example $\alpha < N$, $\mu = \mathscr{H}^\alpha \llcorner P$ for an $\alpha-$dimensional plane P and we take $k = N$ (then $\mathscr{H}^k = \mathscr{L}$), one has

$$\theta^*_k(\mu,x) = \begin{cases} +\infty & \text{if } x \in P \\ 0 & \text{otherwise;} \end{cases}$$

therefore, since $\mathscr{L}(P) = 0$, $\theta^*_k(\mu,x) = 0$ a.e. with respect to $\mathscr{H}^k$, according to the Corollary, but $\theta^*_k(\mu,x) = +\infty$ a.e. with respect to μ.

We can now prove what we stated before with the

Lemma 14.8. *Given any measure μ, the following characterization holds:*

$$\mathscr{H}-\dim \mu = \sup \{k : \ \theta^*_k(\mu,x) < +\infty \ for \ \mu - a.e. \ x \in X\}. \tag{14.4}$$

Before the statement of Theorem 14.6, we claimed that the dimension of the measure μ is the $\mu-$essential infimum of the pointwise dimension: note that this assertion is completely equivalent to (14.4).

Proof. *(of Lemma 14.8):* Take $k \in \mathbb{R}$: if $\theta_k^*(\mu, x) < +\infty$ for $\mu-$a.e. x, then we can write $X = \cup X_n$ up to a $\mu-$negligible set, where $X_n = \{x : \theta_k^*(\mu, x) \leq n\}$; take moreover a set $B \in \mathscr{B}(X)$ such that $\mathscr{H}^k(B) = 0$: by Theorem 14.6 a) one obtains

$$\mu(X_n \cap B) \leq 2^k\, n\, \mathscr{H}^k(B) = 0,$$

thus $\mu(B) = 0$: we infer $\mu \ll \mathscr{H}^k$ and then an inequality is proved.

Suppose now $\mu \ll \mathscr{H}^k$: then by Corollary 14.7 we obtain that $\theta_k^*(\mu, x) < +\infty$ for $\mu-$a.e. x, and the other inequality follows. □

The above characterization is quite powerful since it allows to use the properties of the Hausdorff measures in the definition of $\mathscr{H}-\dim$ together with pointwise arguments like those we will need in the following chapters.

We now prove some simple facts about the behavior of the dimension under the action of Lipschitz functions, which we will need in Chapter 15 to show that no nontrivial upper estimates for the dimension of μ can be given.

Lemma 14.9. *If $\mu \in \mathcal{M}^+(X)$ and $\varphi : X \to \mathbb{R}^N$ is Lipschitz, then $\mathscr{H}-\dim\, \varphi_\# \mu \leq \mathscr{H}-\dim\, \mu$. Moreover, if φ is bilipschitz then $\mathscr{H}-\dim\, \varphi_\# \mu = \mathscr{H}-\dim\, \mu$.*

Proof. Using the definition of dimension we just need to prove that $\varphi_\# \mu \ll \mathscr{H}^k \implies \mu \ll \mathscr{H}^k$. Then, let $\varphi_\# \mu \ll \mathscr{H}^k$ and let $A \in \mathscr{B}(X)$ be a set with $\mathscr{H}^k(A) = 0$: we have $\mu(A) \leq \mu(\varphi^{-1}(\varphi(A)) = \varphi_\# \mu(\varphi(A)) = 0$, where the last equality holds because $\mathscr{H}^k(\varphi(A)) = 0$. Finally, if φ is bilipschitz, the same argument gives also the other inequality, since $\varphi_\#^{-1}(\varphi_\# \mu) = \mu$. □

Lemma 14.10. *Let $\mu \in \mathcal{M}^+(X)$ and let $\varphi : X \to \mathbb{R}^N$ be a Lipschitz map with the property that $\mu-$almost all of X can be covered by countably many Borel sets A_n, $n \in \mathbb{N}$, such that φ is bilipschitz on each of the A_n. Then $\mathscr{H}-\dim\, \varphi_\# \mu = \mathscr{H}-\dim\, \mu$.*

Proof. We just need to prove that $\mathscr{H}-\dim\, \varphi_\# \mu \geq \mathscr{H}-\dim\, \mu$, and this inequality will follow if we prove that $\mu \ll \mathscr{H}^k \implies \varphi_\# \mu \ll \mathscr{H}^k$; let then k be such that $\mu \ll \mathscr{H}^k$, and take $A \in \mathscr{B}(\mathbb{R}^N)$ with $\mathscr{H}^k(A) = 0$. Thanks to the previous lemma and to the assumptions, we have

$$\varphi_\# \mu(A) = \mu(\varphi^{-1}(A)) \leq \sum_{n \in \mathbb{N}} \mu(\varphi^{-1}(A) \cap A_n) = 0.$$

□

14.3 An important remark

A very important remark must be done about the writing $f = f^+ - f^-$: it is worth to spend some words on it, because it may lead (sometimes it already did) to some misunderstandings. In our formulation of the problem, f^+ and f^- are two given positive measures of the same total mass –or, in the absolutely continuous case, two positive L^1 functions with the same integral– and we write $f = f^+ - f^-$ only for shortness: this corresponds to a "geometrical" point of view, where one really thinks about some mass to move from a place to another. On the other hand, many people studying mass transportation take directly the measure (or the L^1 function) f with 0 mean, and they define f^+ and f^- as the positive and the negative part of it: this may for instance correspond to a PDE point of view. But of course it is different to consider as the data of the problem f^+ and f^- or f, since different pairs (f^+, f^-) may correspond to the same f: all these pairs are equal up to an additive positive measure, i.e. if $\bar{f}^+$ and $\bar{f}^-$ are the positive and negative part of f the pairs corresponding to f are all those of the type $(\bar{f}^+ + h, \bar{f}^- + h)$ for some positive measure h. Therefore, working with the second point of view is equivalent to working with the first one and with the assumption that $f^+ \wedge f^- = 0$.

We did not do this assumption until now, and we did it on purpose: in fact, the formulation we used was rather more general, and we did not see a reasonable reason to assume that the original position of the mass and the final one intersect only in a negligible set. This hypothesis, or the stronger hypothesis of empty intersection of the supports, is often useful, for example it was assumed in [39], [37], in some results of [47] and it will be also assumed later to show Theorem 16.12; but we would like to avoid it as much as possible, reaching the results in the stronger formulation. For instance, recall that we needed to pay some attention in the whole first Part of the thesis to treat the set of the fixed points, that would not exist with the second point of view.

However, in this third Part we will need to use the second formulation: it will be done whenever we consider as a data f instead of the pair (f^+, f^-), in particular in Chapters 17 and 18. Does it mean that we will prove our estimates only under the assumption that $f^+ \wedge f^- = 0$?

Fortunately, not. In fact, we show now that this assumption can be done without loss of generality to study the properties of the transport densities; more precisely, let us consider any pair (f^+, f^-) and let us denote by $\bar{f}^+$ and $\bar{f}^-$ the positive and the negative part of $f = f^+ - f^-$. We claim the the transport densities relative to (f^+, f^-) and those relative to $(\bar{f}^+, \bar{f}^-)$ are exactly the same, and of course this will imply that all the results about the transport densities one finds working with the second point of view are valid also with the first one. The previous claim immediately follows since the transport densities are exactly those measures μ such that there exists a $1-$Lipschitz function u with the property that

$$\begin{cases} -\text{div }(\mu D_\mu u) = f^+ - f^- & \text{on } X, \\ |D_\mu u| = 1 & \mu - \text{a.e.}\,; \end{cases} \tag{14.5}$$

the notion of the tangential derivative $D_\mu u$ was given in [14, 15] and generalized in [44]; however, it coincides with the usual weak derivative Du in the case when $\mu \ll \mathscr{L}$, that is the case we will consider in Chapters 17 and 18. The above mentioned fact, that μ can solve (14.5) with some u if and only if it is a transport density, is well known, see for example [14, 15, 12]; we can also derive this property, in the much greater generality of the second Part of the thesis: in fact, by Remark 12.5 μ can solve (12.9) –the finslerian generalization of (14.5)– if and only if it is an optimal shape, and by Theorem 13.8 we know that optimal shapes and transport densities are the same.

14.4 Some examples about the summability

In this section we give some examples to show the L^p regularity that one cannot expect from the transport density. First of all, we have the

Example 14.11. In $\mathbb{R}^2$ let $R_1 = [-2,-1]\times[0,1]$ and $R_2 = [1,2]\times[0,1]$. Moreover let $g : [0,1] \to \mathbb{R}$ be an absolutely continuous function, consider $a^+ : R_1 \to \mathbb{R}$ and $a^- : R_2 \to \mathbb{R}$ defined by $a^+(x,y) = a^-(-x,y) = g(y)$ and define $f^+ = a^+\mathscr{H}^2 \llcorner R_1$ and $f^- = a^-\mathscr{H}^2 \llcorner R_2$. Thanks to the symmetry of the problem the horizontal translation of f^+ on f^- is an optimal transport map, then using formula (14.1) we can explicitely write the transport density μ (which is unique and absolutely continuous as we will show with Theorems 16.2 and 16.4) obtaining that it is supported in $[-2,2]\times[0,1]$ and it is given by

$$\mu(x,y) = \begin{cases} g(y) & \text{if } |x| \in [0,1] \\ g(y)(2-x) & \text{if } |x| \in [1,2] \\ 0 & \text{if } |x| > 2 \end{cases}$$

(note that, as made throughout this thesis, we use the same letter to denote a measure absolutely continuous with respect to $\mathscr{L}$ and its density). Consequently, one has $\{p :\ \mu \in L^p(\mathbb{R}^2)\} = \{p :\ f^+ \in L^p(\mathbb{R}^2)\} = \{p :\ f^- \in L^p(\mathbb{R}^2)\} = \{p :\ g \in L^p(\mathbb{R})\}$.

This example shows that if both f^+ and f^- belong to L^p there is no reason to hope that $\mu \in L^q$ for some $q > p$; in the following, we will prove that what is true is that also $\mu \in L^p$, therefore this estimate will be sharp.

We will also study the regularity on μ that can be inferred from the regularity of only one between f^+ and f^-: more precisely, we will show that if one between f^+ and f^- is in L^1, then the same is true for μ.

Moreover, if at least one between f^+ and f^- is in L^p with some $p > 1$, then $\mu \in L^q$ with some $1 < q < p$; we see now with an example that the best $q = q(p)$ for which this last claim is true (which thanks to the remarks above is less than or equal to p) is in fact strictly smaller than p. We show with this example that $f^+ \in L^\infty$ does not imply $\mu \in L^\infty$, and in a completely similar way $f^+ \in L^p$ does not imply $\mu \in L^p$:

Example 14.12. Let f^- be the Dirac mass in 0 and f^+ be the restriction of the Lebesgue measure to an annulus centered in 0 and of unitary volume, then $f^+ \in L^\infty$. Obviously $\gamma = f^+ \otimes f^-$, which brings all the mass of f^+ in 0, is the unique transport plan and then it is optimal. It is then easy, using (14.2), to evaluate $\mu(x)$ for each Lebesgue point x for μ, which is in L^1 as we will show in Theorem 16.4: it follows $\mu(x) \propto |x|^{1-N}$ for x inside the annulus, and then $\mu \notin L^\infty$; more precisely, $\mu \in L^p$ if and only if $p < N/(N-1)$.

14.5 About the boundedness of X: a counterexample

Here we discuss the boundedness property on X: as we said in Section 14.1, the set X in this third Part will be a bounded and convex subset of $\mathbb{R}^N$ (in particular, without loss of generality it will be a bounded ball in Chapter 17); the convexity hypothesis is done, as usual, so that the transport rays are straight lines instead of geodesics. But let us try to discuss for a moment the boundedness hypothesis; it is done for technical reasons: in particular the volume of the cylinders C_ε we will encounter in the geometrical lemmas of Section 14.6 can be bounded by ε^{N-1} if X is bounded, while it is clearly $+\infty$ if $X = \mathbb{R}^N$. Similarly, also the estimates of Chapters 17 and 18 are meaningless without a boundedness hypothesis.

However, we will see in a moment which of the results of this third Part are true also in the case $X = \mathbb{R}^N$ and which are not; all the results we will prove will be of one of the following types:

a) Under some hypotheses about the dimensions of f^+ and f^-, something about the dimension of the transport densities follows;

b) If one between f^+ and f^- is in L^1, then there is a unique transport density;

c) If one between f^+ and f^- is in L^1, $\mu \in L^1$;

d) Under some L^q properties of f^+ and f^-, $\mu \in L^p$.

Let us try to understand which of the above types of claims can be extended to the setting when the ambient space is the whole $\mathbb{R}^N$: the claim b), that is Theorem 16.2, will be directly proved in the general case of $\mathbb{R}^N$, so there is no need of extending it. Concerning the other three types, an idea of proof could be the following one: take an optimal transport plan γ between f^+ and f^- (and assume, as usual, that $\mathfrak{C}(\gamma) < +\infty$); then define γ_R the restriction of γ to $B(0,R) \times B(0,R)$: by Lemma 1.12 we know that γ_R is an optimal transport plan between its marginals, and moreover it is immediate from the definition to deduce that the transport density μ_R associated to γ_R is less than the transport density μ associated to γ and it strongly converges to it when $R \to +\infty$. Does it allow to extend the claims a), c) and d) from the case of X bounded to the general case of $\mathbb{R}^N$? The answer is yes for claims a) and c), no for claim d). In fact, by the definition it is clear the the limit of an increasing sequence of measures of dimension less than $k \in \mathbb{R}$ is still a measure of dimension less than k, so the claim a) can be extended; analogously, the limit of an increasing sequence of measures absolutely continuous with respect to $\mathscr{L}$ is still a measure absolutely continuous with respect to $\mathscr{L}$, therefore also claim c) can be extended recalling that a transport density is in L^1 if and only if it is absolutely continuous with respect to $\mathscr{L}$ thanks to Remark 14.2. On the other hand, the limit of an increasing sequence of L^p functions needs not, in general, to be still in L^p, so our idea of proof does not work for the claim d).

Let us finally see a counterexample showing that in fact all the L^p results we will prove in this third Part for $p > 1$ are valid only in the bounded ambient space X, but not in the whole $\mathbb{R}^N$. In fact, in the following example we have a situation where f^+ and f^- are in each L^p with $1 \leq p \leq +\infty$ but the unique transport density μ is not in any L^p with $p > 1$ (but, according to our argument above, $\mu \in L^1$).

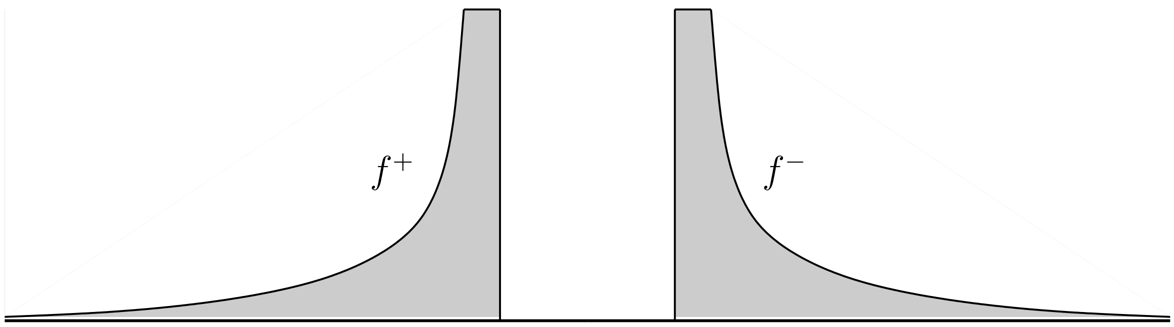


Figure 14.1: Example 14.13

Example 14.13. Let us consider a function $g : [0,1] \to \mathbb{R}^+$ which is in L^2 but not in L^p for any $p > 2$, let $A = \{(x,y) \in \mathbb{R}^2 : 0 \leq y \leq 1,\ 1 \leq -x \leq 1 + g(y)\}$ and define, as in Figure 14.1, $f^+ = \chi_A$ and $f^- = (-\mathrm{Id}, \mathrm{Id})_\# f^+$ (in the Figure, g is decreasing and $g(0) = +\infty$ while $g(1) > 0$). Note that f^+ and f^- belong to L^p for each $1 \leq p \leq +\infty$: in fact, f^+ and f^- are in L^∞ by definition, they are in L^1 because so is g, and then by classical interpolation they are in each L^p space; moreover, the minimum of the Kantorovich problem is easily seen to be finite because an easy calculation shows that $t(-y, x) = (x, y)$ is a transport map and it has a finite cost since $g \in L^2$. In particular, arguing as in the example considered in Section 6.3, it is shown that t is an optimal transport map: then one can compute the value of the transport density μ (which is unique and absolutely continuous thanks to Theorems 16.2 and 16.4), in a completely similar way as done for Example 14.11, and it is found that μ is supported in $\mathbb{R} \times [0, 1]$ and it is given by

$$\mu(x, y) = \begin{cases} g(y) & \text{if } |x| \in [0, 1] \\ g(y) + 1 - |x| & \text{if } |x| \in [1, g(y) + 1] \\ 0 & \text{if } |x| > g(y) + 1. \end{cases}$$

A straightforward calculation ensures now that $\mu \in L^1$ (according to what we proved before) but $\mu \notin L^p$ for any $p > 1$.

14.6 Three geometrical lemmas

In this section we report some technical results (all taken from [31]) which will be the main tools for the proofs in Chapters 15 and 16: through the section, we assume that an optimal transport plan γ is given and we establish some peculiarities of the corresponding transport density μ. A first property which follows directly from the definition of μ is the following:

Proposition 14.14. *μ−a.e. point in X is contained in the interior of some transport ray (i.e. of some segment xy with $(x, y) \in \operatorname{spt}\gamma$).*

Proof. This is a consequence of the definition of μ: let S be the subset of the points of X which are not contained in the interior of a transport ray. For every $(x, y) \in spt(\gamma)$ we have $xy \cap S \subseteq \{x, y\}$ and then $l(xy \cap S) = 0$ which, from (14.2), gives the claimed assertion. □

The next three lemmas consider a point z contained in the interior of some transport ray xy (which happens for μ−a.e. point, thanks to Proposition 14.14) and give estimates about the location of the extreme points of the other rays which meet an open ball centered at z and of sufficiently small radius: in particular in Lemma 14.15 we find a region to which at least one of the extremes belongs, in Lemma 14.16 we find a region to which both the extremes belong (the estimate degenerates when z get closer to x or y), in Lemma 14.17 we find a region to which the first extreme of the ray belong, but under the assumption that z is not contained in the support of f^-. These lemmas will play a key role in the proofs of all the results of Chapters 15 and 16.

To prove these lemmas we will only use the $c-$cyclical monotonicity of the support; more precisely, we will use it only in the case of two pairs of point: recalling that the cost function is the Euclidean distance and writing, as we said before, $\overline{xy}$ to denote $|y-x|$, setting $k=2$ in (3.1) we find the inequality

$$\overline{xy} + \overline{x'y'} \leq \overline{xy'} + \overline{x'y} \tag{14.6}$$

that must be true whenever (x, y) and (x', y') belong to the support of some optimal transport plan, thanks to Proposition 3.2.

The outline of the proofs of the next lemmas is always the same: we will have to prove that the pairs belonging to $\operatorname{spt}\gamma$ fulfill some property. To do that, we will assume that this is not true and we will look for a contradiction with (14.6); in fact, we will take two pairs not fulfilling that property and we will show that (14.6) does not hold. To do that, it will be convenient to reduce the possible configurations of the four points by moving them suitably: but we will only move them in such a way that, if the property (14.6) was true before the movement, it remains true after that. Then, showing that at the end the property (14.6) is not true for the final positions of the points, we will infer that it was not true also for the original points, and the absurd will show our thesis. According to this idea, given (x, y) and (x', y') in $X \times X$ we call *admissible transformations* all the changes of the points such that the quantity

$$\overline{xy} + \overline{x'y'} - \overline{xy'} - \overline{x'y}$$

decreases, which clearly maintain the validity of (14.6): for example, thanks to the triangular inequality it is easily seen that moving x to y or y to x on the line xy is admissible, and the same holds for x', y' and $x'y'$.

Lemma 14.15. *If z is an interior point of a transport ray xy, ε is sufficiently small and another ray $x'y'$ of length less than $\alpha\overline{xy}$ (with $\alpha \geq 1$) intersects $B(z, \varepsilon)$, then either x' or y' belongs to the cylinder C_ε with axis xy and radius $6\alpha\varepsilon$.*

Proof. Let us assume $\varepsilon \ll 1$ and that there exists a transport ray $x'y'$ of length less than $\alpha\overline{xy}$ that intersects $B(z, \varepsilon)$ but neither x' nor y' is in the cilinder of axis xy and radius $6\alpha\varepsilon$: in order to prove the lemma we must show that this leads to a contradiction.

By symmetry, we can assume that $0 < a \leq b \leq R$, where $a = \overline{xz}$, $b = \overline{xy}$ and R is the diameter of X: let us move x' closer to y' moving it on the line $x'y'$ (which is admissible) until the distance between x' and the line extending xy is $5\alpha\varepsilon$. We fix now coordinates $\{c_1, \ldots, c_N\}$ such that z is the origin, the segment xy is on the first axis (with $c_1(x) < 0,\ c_1(y) > 0$), x' is on the plane $\{c_3 = \cdots = c_N = 0\}$ with $c_2(x') < 0$ (and then $c_2(x') = -5\alpha\varepsilon$) and y' is in the space $\{c_4 = \cdots = c_N = 0\}$ with $c_3(y') \geq 0$. Since the distance between y' and the line xy is greater then $6\alpha\varepsilon$, $x'y'$ intersects $B(z,\varepsilon)$ and $c_3(x') = c_3(z) = 0$, from Pitagora's theorem it follows that $c_2(y') > 5\alpha\varepsilon$: bring then y' closer to x' (moving it on the line $x'y'$) until $c_2(y') = 5\alpha\varepsilon$. The situation is illustrated in figure 14.2.

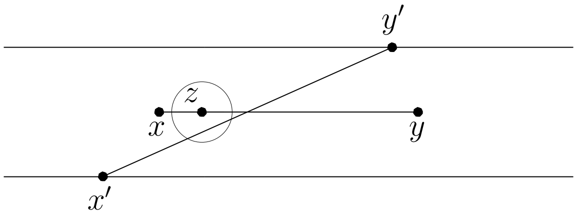


Figure 14.2: Geometry of Lemma 14.15

Let us define now $\delta = c_1(y')$ and $l = c_1(y') - c_1(x')$. We can assume $l \geq 0$: otherwise, applying to x' and y' the simmetry across the hyperplane $\{c_1 = \lambda\}$ with $\lambda = 0,\ c_1(y')$ or $c_1(x')$ respectively if $c_1(y') \leq 0 \leq c_1(x')$, $0 \leq c_1(y') \leq c_1(x')$ or $c_1(y') \leq c_1(x') \leq 0$ changes the sign of l and is an admissible transformation, because $\overline{xy}$ and $\overline{x'y'}$ remain equal while $\overline{xy'}$ and $\overline{x'y}$ increase. Moreover we have $l \gg \varepsilon$: otherwise, using (14.6) and the fact that it would be $\overline{x'y'} \geq \sqrt{l^2 + (10\alpha\varepsilon)^2}$ and $\overline{xy'} + \overline{x'y} = a + b + l + (5\alpha\varepsilon)^2(1/2a + 1/2b) + o(\varepsilon^2)$, one easily finds $l \geq 4ab/(a+b) \gg \varepsilon$. We have then $l \gg \varepsilon$ and, as a consequence, $\delta \gg \varepsilon$ (recall that $x'y'$ intersects $B(z,\varepsilon)$): the coordinates of the points are now

$$x \equiv (-a, 0, 0) \qquad\qquad y \equiv (b, 0, 0)$$
$$x' \equiv (\delta - l, -5\alpha\varepsilon, 0) \qquad\qquad y' \equiv (\delta, 5\alpha\varepsilon, h\varepsilon),$$

writing only the first three coordinates because all the points are in the space $\{c_4 = \cdots = c_N = 0\}$. The facts that $x'y'$ intersects $B(z,\varepsilon)$, $c_3(x') = c_3(z) = 0$, $b \geq a > 0$ and $l \leq \overline{x'y'} \leq \alpha\overline{xy}$ imply, together with trivial similitude arguments,

$$0 \leq h < 2.5, \qquad \delta + \varepsilon \geq \frac{5\alpha - 1}{10\alpha}\, l, \qquad \delta - \varepsilon \leq \frac{5\alpha + 1}{10\alpha}\, l, \qquad b \geq \frac{\overline{xy}}{2} \geq \frac{l}{2\alpha}. \tag{14.7}$$

Let us write now the lengths of the segments:

$$\begin{aligned}
\overline{xy} &= a+b, & \overline{x'y} &= b+l-\delta+\frac{25\alpha^2+o(1)}{2(b+l-\delta)}\,\varepsilon^2\\
\overline{x'y'} &= l+\frac{100\alpha^2+h^2+o(1)}{2l}\,\varepsilon^2, & \overline{xy'} &= a+\delta+\frac{25\alpha^2+h^2+o(1)}{2(a+\delta)}\,\varepsilon^2.
\end{aligned} \tag{14.8}$$

The inequality (14.6) gives $\overline{xy}+\overline{x'y'} \leq \overline{xy'}+\overline{x'y}$ and this, thanks to (14.8) and using (14.7), implies

$$\begin{aligned}
\frac{100\alpha^2+h^2}{2l} &\leq \frac{25\alpha^2}{2(b+l-\delta)}+\frac{25\alpha^2+h^2}{2(a+\delta)}+o(\varepsilon^2) \Longrightarrow\\
\frac{100\alpha^2+h^2}{l} &\leq \frac{25\alpha^2}{l/(2\alpha)+l-((5\alpha+1)/10\alpha)l}+\frac{25\alpha^2+h^2}{((5\alpha-1)/10\alpha)l} \Longrightarrow\\
750\alpha^3-400\alpha^2 &\leq h^2\left(25\alpha^2+25\alpha+4\right) \Longrightarrow\\
750\alpha^3-400\alpha^2 &\leq 6.25\left(25\alpha^2+25\alpha+4\right),
\end{aligned}$$

which is easily seen to be false for each $\alpha \geq 1$. □

Lemma 14.16. *If z is an interior point of a transport ray xy, $a = \min\{\overline{xz}, \overline{zy}\}$, ε is sufficiently small and another transport ray $x'y'$ intersects $B(z,\varepsilon)$, then both x' and y' belong to the cylinder C_ε with axis xy and radius $5R\varepsilon/a$, where R is the diameter of X.*

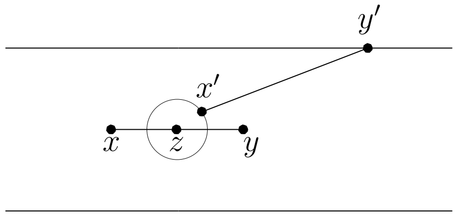


Figure 14.3: Geometry of Lemma 14.16

Proof. Possibly moving x closer to y or y closer to x on the line xy, which we know to be an admissible transformation, we can assume that $\overline{xy} = 2a$ and z is the middle point of xy; let us also assume that y' does not belong to the cylinder and show that this leads to a contradiction. Since y' does not belong to the ball $B(z,\varepsilon)$ but $x'y'$ intersects the ball, we can move x' in the direction of y' until it crosses $B(z,\varepsilon)$ and exits from it, and this is admissible.

We can assume now that x, y, x' and y' are on a plane: in fact, moving y' to the point of least distance from x' remaining on the $(N-2)-$dimensional sphere of the points whose distances from x and y are fixed is admissible (because in this way $\overline{xy}$, $\overline{xy'}$ and $\overline{x'y}$ remain fixed, while $\overline{x'y'}$ decreases) and causes all the points to be on a same plane without changing the fact that $y' \notin C_\varepsilon$. Finally, we can move y' closer to x' until it reaches C_ε, and the situation is illustrated in figure 14.3. If we fix coordinates in the obvious way (as we did in the previous lemma), the points are

$$x \equiv (-a, 0) \qquad\qquad y \equiv (a, 0)$$
$$x' \equiv (\delta\varepsilon, h\varepsilon) \qquad\qquad y' \equiv (\delta\varepsilon + l, 5R\varepsilon/a)$$

and we know that

$$a \leq R \qquad l \leq R \qquad -1 \leq \delta \leq 1 \qquad 0 \leq h \leq 1: \tag{14.9}$$

writing the lengths of the segments and using (14.6) and (14.9) exactly as we did in the proof of Lemma 14.15, we find a contradiction, and then the desired inequality. □

Lemma 14.17. *If z is an interior point of a transport ray xy, $dist(z, spt(f^-)) = \delta > 0$, ε is sufficiently small and another ray $x'y'$ intersects $B(z, \varepsilon)$, then x' belongs to the cylinder C_ε of axis xy and radius $6R\varepsilon/\delta$, where R is the diameter of X.*

Proof. Let us begin with the admissible transformations of bringing y and y' closer to x and x' (on the lines xy and $x'y'$) until they are at a distance δ from z, which is possible if $\varepsilon \leq \delta$. If x' belongs to $B(z, \varepsilon)$ we have nothing to prove, otherwise $\overline{x'y'} \geq \delta$ –recall that $\overline{zy'} = \delta$– and then, applying Lemma 14.15, we know that either x' or y' belong to the cylinder of axis xy and ray $6\varepsilon\overline{x'y'}/\overline{xy} \leq 6R\varepsilon/\delta$. If x' belongs to that cylinder there is nothing to prove, then let us assume that y' belongs to it and move x closer to y until $x \equiv z$. Finally we fix a system of coordinates such that z is the origin, $y \equiv (\delta, 0, \dots, 0)$, y' belongs to the plane $\{c_3 = \cdots c_N = 0\}$ with $c_2(y') \leq 0$ and x' belongs to the space $\{c_4 = \cdots c_N = 0\}$. Since $c_3(y') = c_3(z) = 0$ and $x'y'$ intersects $B(z, \varepsilon)$, it follows by similitude that $|c_3(x')| \leq R\varepsilon/(\delta - \varepsilon) \leq 2R\varepsilon/\delta$ if ε is sufficiently small. Moreover the fact that $c_2(y') \leq 0$ implies, by a similar argument, that $c_2(x') \geq -2R\varepsilon/\delta$: it will be then sufficient to prove that $c_2(x') \leq 5R\varepsilon/\delta$.

To this aim let us note that y' belongs to C_ε and $\overline{zy'} = \delta$, and then if $\varepsilon \ll 1$ we have $|c_1(y')| \approx \delta$; then we can assume that $c_1(y') > 0$: in fact, if $c_1(y') \leq 0$ then $c_1(y') \approx -\delta$ and then either $c_1(x') \leq 0$, and then easily $c_2(x') < 5R\varepsilon/\delta$ if $\varepsilon \ll 1$, or $c_1(x') > 0$, and then applying to x' and y' the simmetry across the hyperplane $\{c_1 = 0\}$ is admissible and leads us to a situation in which $c_1(y') > 0$.

If $c_2(y') \geq -3\varepsilon$ a similitude argument similar to those we already used implies $c_2(x') \leq 5R\varepsilon/\delta$; if $c_2(y') < -3\varepsilon$, let us consider the situation, illustrated by figure 14.4: denoting

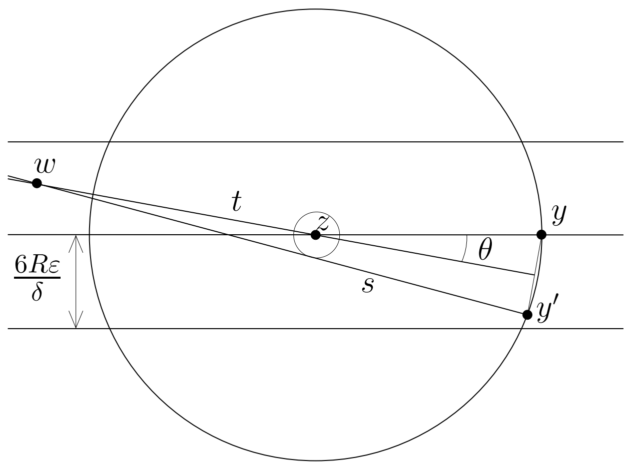


Figure 14.4: Geometry of Lemma 14.17

by x_2 the projection of x' on the plane $\{c_3 = \cdots = c_N = 0\}$, since $x'y'$ intersects $B(z, \varepsilon)$, the point x_2 is over the line s, and since $\overline{x'y'} \leq \overline{x'y}$ (from (14.6) recalling that $\overline{xy} = \overline{xy'}$) it is under the line t, the axis of the segment yy': then $c_2(x') = c_2(x_2) \leq c_2(w)$. With elementary calculations one can then write the equations of the lines s and t and using the fact that, thanks to our assumptions, $-c_2(y') = \delta \sin(2\theta) \in [3\varepsilon, 6R\varepsilon/\delta]$, we find $c_2(w) \leq 4\varepsilon \leq 4R\varepsilon/\delta$ if $\varepsilon \ll 1$, that concludes the proof. $\square$

Chapter 15

Sharp dimensional estimate via geometrical arguments

In this chapter we provide a lower estimate for the dimension of the transport density, and we will show with an example that it is sharp and that no nontrivial upper estimate can be given. The estimate is taken from [31], and it appeared first in [61].

We begin with the lower estimate, which follows by a careful analysis from Lemma 14.16.

Theorem 15.1. *Given any transport density μ, the following estimate holds:*

$$\mathscr{H}-\dim\ \mu \geq \max\{\mathscr{H}-\dim\ f^+,\ \mathscr{H}-\dim\ f^-,\ 1\}.$$

Notice that in general there can be many different transport densities (see for instance Example 16.1); the preceding estimate is true for each of them.

Proof. (of Theorem 15.1): Fix an optimal transport plan γ between f^+ and f^- and consider the associated transport density μ. We also assume that $k = \mathscr{H}-\dim\ f^+ \geq \mathscr{H}-\dim\ f^-$, that is possible by the symmetry of the problem. We have then to prove that $\mathscr{H}-\dim\ \mu \geq k \vee 1$.

Step I. $\mathscr{H}-\dim\ \mu \geq 1$.

From (14.2) it follows that for any $x \in X$ and $\rho > 0$ we have $\mu(B(x,\rho)) \leq 2\rho\|\gamma\|$ which implies, from definition 14.5, that $\theta_1^*(\mu, x) \leq \|\gamma\|$ for any x. Then using Proposition 14.8 we have the claim.

Step II. *Proof with hypothesis (15.1).*

Assume for the moment that the following hypothesis holds:

$$\exists K:\ \forall x \in X,\ \forall \rho > 0\ \ f^+(B(x,\rho)) \leq K\rho^k; \tag{15.1}$$

we will get rid of it in the next step.

If z is contained in the interior of some transport ray xy and $a = \min\{\overline{xz},\ \overline{zy}\}$, let C_ρ be the cylinder or axis xy and radius $5R\rho/a$, which can be covered by $M(z)/\rho$ balls of radius ρ: this, thanks to (15.1), ensures that

$$f^+(C_\rho) \le \frac{M(z)}{\rho}\ K\rho^k \le M(z)K\rho^{k-1}.$$

Then, if ρ is sufficiently small, using Lemma 14.16 we have

$$\begin{aligned}\mu(B(z,\rho)) &= \iint_{C_\rho\times C_\rho} l(xy\cap B(z,\rho))\ d\gamma(x,y) \le 2\rho \iint_{C_\rho\times C_\rho} d\gamma(x,y)\\ &\le 2\rho f^+(C_\rho) \le 2M(z)K\rho^k.\end{aligned}$$

We then infer that $\theta_k^*(\mu,x) < +\infty$ for every $z\in X$ which is contained in the interior of some transport ray, and then from Lemma 14.8 and Proposition 14.14 it follows that $\mathscr{H}-\dim\ \mu \ge k$. The theorem is then proved under the additional assumption (15.1).

Step III. *General case.*

Fix $\varepsilon > 0$ and define for any $m\in\mathbb{N}$

$$X_{m,\varepsilon} = \{x\in X : \forall\rho>0,\ f^+(B(x,\rho)) \le m\rho^{k-\varepsilon}\}.$$

Since $k-\varepsilon < \mathscr{H}-\dim\ f^+$, we have $\theta_{k-\varepsilon}^*(f^+,x) = 0$ f^+-a.e., and then

$$\sup_{\rho>0}\ \frac{f^+(B(x,\rho))}{\rho^{k-\varepsilon}} < +\infty \qquad \text{for } f^+\text{−a.e. } x:$$

then, varying m, $X_{m,\varepsilon}$ is an increasing sequence of subsets of X that fill it up to an f^+-negligible set and then, as $m\to+\infty$, $f^+ \llcorner X_{m,\varepsilon} \longrightarrow f^+$ in the strong convergence of $\mathcal{M}^+(X)$. Defining $\gamma_{m,\varepsilon} = \gamma\llcorner(X_{m,\varepsilon}\times X)$ and recalling that $\pi_1\gamma = f^+$ we infer that $\gamma_{m,\varepsilon} \longrightarrow \gamma$. Let us define now $\mu_{m,\varepsilon}$ from $\gamma_{m,\varepsilon}$ by (14.1). It is clear from the definition that $\mu_{m,\varepsilon} \longrightarrow \mu$, and thanks to Lemma 1.12 we can apply the result of Step II to $X_{m,\varepsilon}$ to obtain $\mathscr{H}-\dim\ \mu_{m,\varepsilon} \ge k-\varepsilon$: this, thanks to the definition of dimension and to the convergence of $\mu_{m,\varepsilon}$, implies $\mathscr{H}-\dim\ \mu \ge k-\varepsilon$. Finally the generality of ε completes the proof. □

Remark 15.2. *Note that the preceding Theorem is true when the ambient space is the whole $\mathbb{R}^N$, instead of only a convex and bounded subset X of it; in fact, we did not use in the proof the boundedness of the ambient space. Alternatively, one can recall the argument of Section 14.5.*

We give now some examples to show that the lower estimate given above is sharp and that no upper estimate can be given:

Example 15.3. Let C_1 and C_2 be two disjoint segments of length l lying on the same line, and let f^+ and f^- be the restrictions of the measure $\mathscr{H}^1$ to C_1 and C_2 respectively. Clearly an optimal transport map is given by the translation, then it is simple to calculate μ, which is the measure on the line which density is l between the segments, goes to 0 linearly on the segments and is 0 out, and which is then one–dimensional.

Example 15.4. Let us consider again Example 2.8: it is easy to notice that, calling μ the transport density related to the unique optimal transport plan, $\mathscr{H}-\dim\,\mu = N$, the maximum possible, while the dimensions of f^+ and f^- are both 0, the least possible.

The lack of control in the previous example is due to the fact that the pointwise dimensions of f^+ and f^- are not constant. However, it is known that, even if the pointwise dimensions are constant, it is not possible to give an upper estimate of $\mathscr{H}-\dim\,(f^+\otimes f^-)$ in terms of $\mathscr{H}-\dim\,f^+$ and $\mathscr{H}-\dim\,f^-$ (on this topic see: 3.2.23, 2.10.45, 2.10.29 of [41] and Theorem 5.11 of [40]). In the next example we show how to build a situation where a transport density μ has dimension equal to $\mathscr{H}-\dim\,(f^+\otimes f^-)+1$: putting together the above cited facts and the construction below, one can infer that it is not possible to give a non trivial estimate from above of $\mathscr{H}-\dim\,\mu$ even if the pointwise dimensions of f^+ and f^- are constant.

Example 15.5. Let $p,\ k$ be two positive integers such that $p+1+k\leq N$, S be the unit sphere of the subspace of $\mathbb{R}^N$ given by the first $(p+1)$ coordinates and B be the unit ball of the subspace given by the coordinates from the $(p+2)-th$ to the $(p+1+k)-th$. Let f^+ and f^- be two probability measures on S and B respectively. By the symmetry of the problem it follows that any transport plan is optimal. If we take, for example, $\gamma = f^+\otimes f^-$, then we remark that μ is the push forward of $f^+\otimes f^-\otimes\mathscr{H}^1 \llcorner [0,1]$ through the map $\varphi : S\times B\times[0,1]\to\mathbb{R}^N$ defined by $(x,y,t)\to tx+(1-t)y$. Applying Lemma 14.10 with $A_n = S\times B\times(1/n, 1-1/n)$, we obtain

$$\mathscr{H}-\dim\,\mu = \mathscr{H}-\dim\,(f^+\otimes f^-\otimes\mathscr{H}^1 \llcorner [0,1]) = \mathscr{H}-\dim\,(f^+\otimes f^-)+1.$$

Chapter 16

L^p estimates via geometrical arguments

In this chapter we begin the study of the summability properties of the transport density, that will be our interest in the remaining of the thesis. As noticed in Section 14.4, it does not seem reasonable to expect that $\mu \in L^1$ if neither f^+ nor f^- are in L^1. On the other hand, if at least one between f^+ and f^- belongs to L^1, one can prove not only that also any transport density μ is in L^1, but in addition that there is in fact only one transport density. We will first establish this uniqueness result, already known (see for example [1] or [42]), then we will discuss some estimates that can be proved via the geometrical arguments developed in Section 14.6, already appeared in [31].

16.1 Uniqueness of the transport density

The definition (14.1) of the transport density depends on the choice of an optimal transport plan γ, then a priori the transport density is not uniquely determined by the data f^+ and f^-. The fact that the uniqueness does not hold in general can be noticed with the following

Example 16.1. Let us consider again Example 4.17: we already noticed that the optimal transport plans are

$$\gamma_\lambda = \lambda\big(\delta_{x_+} \otimes \delta_{y_-} + \delta_{z_+} \otimes \delta_{w_-}\big) + (1-\lambda)\big(\delta_{x_+} \otimes \delta_{w_-} + \delta_{z_+} \otimes \delta_{y_-}\big)$$

with $0 \le \lambda \le 1$: an immediate calculation shows that the transport density μ_λ corresponding to γ_λ is given by

$$\mu_\lambda = \lambda\big(\mathscr{H}^1 \llcorner x_+y_- + \mathscr{H}^1 \llcorner z_+w_-\big) + (1-\lambda)\big(\mathscr{H}^1 \llcorner x_+w_- + \mathscr{H}^1 \llcorner z_+y_-\big),$$

so different optimal transport plans γ are associated to different transport densities μ.

However, the uniqueness of the transport density is true if f^+ is absolutely continuous with respect to the Lebesgue measure $\mathscr{L}$ (or, by symmetry, if so is f^-). We point out that this result is true in the general situation when the ambient space is the whole $\mathbb{R}^N$, instead of only a bounded convex subset of it (recall the discussion made in Section 14.5). We will need also the hypothesis, often used in the first Part of this thesis, that the first order moments are finite, that is clearly weaker than the boundedness assumption.

Theorem 16.2. *Let f^+ and f^- be two positive measures in $\mathbb{R}^N$ with the same total mass and finite first order moments, and with $f^+ \ll \mathscr{L}$: then the transport density associated to f^+ and f^- is unique, that is, any optimal transport plan between f^+ and f^- defines the same transport density. The same result is true if f^- (in place of f^+) is absolutely continuous with respect to $\mathscr{L}$.*

Proof. **Step I.** *The 1−dimensional case.*
Let us begin considering the case when $N = 1$; we prove that the transport density associated to any transport plan $\gamma \in \mathfrak{M}$ is the same –recall that the set $\mathfrak{M}$ is the set of the transport plans moving all the mass "from left to right", defined by (4.3) in Section 4.2.

We take then two measures f^+ and f^- in $\mathbb{R}$, with finite first order moments and with $f^+ \ll \mathscr{L}$, and we also let γ be a transport plan contained in $\mathfrak{M}$ and μ the transport density associated to γ. Fixed $a \in \mathbb{R}$, we want to evaluate $\mu(A)$ where $A =]\!]-\infty, a[\![$. To do it we notice that, by the fact that $\gamma \in \mathfrak{M}$, for any $(x, y) \in \operatorname{spt}\gamma$ it is

$$l(xy \cap A) = \begin{cases} 0 & \text{if } x \geq a; \\ y - x & \text{if } y < a; \\ a - x & \text{if } x < a \leq y. \end{cases}$$

Therefore (recall that we write $f = f^+ - f^-$) one has

$$\begin{aligned} \mu(A) &= \iint_{\{y<a\}} y - x \, d\gamma(x,y) + \iint_{\{x<a\leq y\}} a - x \, d\gamma(x,y) \\ &= \iint_{\{y<a\}} y \, d\gamma(x,y) - \iint_{\{x<a\}} x \, d\gamma(x,y) + a \cdot \big(f^+(A) - f^-(A)\big) \\ &= -\int_A z \, df(z) \; + a \cdot f(A) = \int_A a - z \, df(z). \end{aligned} \tag{16.1}$$

This implies that $\mu(A)$ depends only on f^+ and f^-, but not on the choice of the plan γ (provided it moves all the mass from left to right); since this is true for the set $A =]\!]-\infty, a[\![$ with a generic $a \in \mathbb{R}$, the uniqueness of the measure μ follows.

Note that (16.1) implies also that the transport density μ is absolutely continuous (this fact is true in general, not only with dimension 1: compare with Theorem 16.4); in particular, writing $\mu = h\mathscr{L}$, again from (16.1) one obtains

$$h(z) = f\big(\rrbracket - \infty, z\llbracket\big).$$

Step II. *Definition of f_C^+, f_C^- and μ_C.*
We consider now the general $N-$dimensional case of the measures f^+ and f^- in $\mathbb{R}^N$: let u be a Kantorovich potential for f^+ and f^- according to Theorem 3.28 (recall also Remark 3.21), and let $\Gamma = \partial^c u$, so that a plan $\gamma \in \Pi(f^+, f^-)$ is an optimal transport plan if and only if it is concentrated in Γ. The function u is $1-$Lipschitz, since the cost function c is the Euclidean one, and then $f^+(\mathrm{D}_\Gamma) = 0$ according to Corollary 6.15. We can also assume, as in Lemma 4.15, that $\mathrm{F}_\Gamma = \emptyset$, since all the optimal transport plans leave fixed the mass in F_Γ and the transport density is defined only by the part of γ relative to non-fixed points (if $x = y$ then $l(xy \cap A) = 0$ for any set A). Arguing then exactly as in Lemma 4.18 (as we did quite often in Chapter 4), we can write $\nu = {r_\Gamma}_\# f^+$ and

$$f^+ = f_C^+ \otimes \nu, \qquad\qquad f^- = f_C^- \otimes \nu, \qquad\qquad \gamma = \gamma_C \otimes \nu,$$

and we find that the measures ν, f_C^+ and f_C^- do not depend on γ but only on f^+ and f^-; moreover, each γ_C is an optimal transport plan between f_C^+ and f_C^-, and it moves the mass "from left to right" by definition, since γ is concentrated in $\Gamma = \partial^c u$. Therefore by Step I the transport density relative to any plan γ_C is μ_C, a measure depending only on f^+ and f^-.
Step III. *Conclusion.*
The thesis will clearly follow by Step II if we prove that $\mu = \mu_C \otimes \nu$, since the measures μ_C and ν are independent of γ. In fact, given any $A \in \mathscr{B}(\mathbb{R}^N)$ and recalling the conclusions of last step, one has

$$\begin{aligned}\mu(A) = \iint_{\mathbb{R}^N \times \mathbb{R}^N} l(xy \cap A)\, d\gamma(x,y) &= \int_{\mathcal{S}_c} \left(\iint_{\mathbb{R}^N \times \mathbb{R}^N} l(xy \cap A)\, d\gamma_C(x,y) \right) d\nu(C) \\ &= \int_{\mathcal{S}_c} \mu_C(A)\, d\nu(C) = \big(\mu_C \otimes \nu\big)(A).\end{aligned}$$

The proof is then completed with the hypothesis that $f^+ \ll \mathscr{L}$. Since the transport plans between f^+ and f^- remain the same swapping f^+ and f^- (differently from what happens with the transport maps, recall the discussion after Definition 1.8), the uniqueness claim is true also under the hypothesis that f^-, in place of f^+, is absolutely continuous with respect to $\mathscr{L}$. $\square$

Remark 16.3. *This Theorem was proved in [1] in the case when f^+ and f^- are compactly supported in $\mathbb{R}^N$: our proof is essentially the same as there, just slightly modified. This result is obtained also in [42] with the same compactness hypotesis and with the further assumption that both f^+ and f^- are absolutely continuous with respect to $\mathscr{L}$.*

16.2 Sharp L^1 estimate

In this section we investigate the absolute continuity of the transport density μ with respect to the Lebesgue measure: in particular, we will prove that if either f^+ or f^- are absolutely continuous then so is μ (well defined by Theorem 16.2). Let us point out that, for a measure, the property of being absolutely continuous is a bit stronger than having dimension N, which is obviously the maximal possible dimension in X; moreover, from Theorem 15.1 we already know that if f^+ has dimension N then also the dimension of μ is N. Our result is then a little step forward and it represents the connection between the study of the dimension and that of the summability of μ. This result is shown making strongly use of the boundedness of X, but it is still true if the ambient space is the whole $\mathbb{R}^N$, as we will show in Remark 16.5.

Theorem 16.4. *If at least one between f^+ and f^- belongs to L^1, then the same is true for the transport density μ. In particular,*

$$\|\mu\|_{L^1} \leq R\,\|f^+\|_{L^1}. \tag{16.2}$$

Before giving the proof, we point out the importance of this result: this summability claim is a meaningful fact about the transport density, and it is probably the most interesting of the L^p estimates, even though the one with the easiest proof. This result was first proved in [61] (see also [31, 1]); an independent proof has been given in [42], where both the measures f^+ and f^- are supposed to belong to L^1, but the result is shown in the more general case of a quasi-euclidean norm (see Chapter 8).

Proof. *(of Theorem 16.4):*
Step I. *The case when $f^+ \in L^\infty$.*
We first assume that $f^+ \in L^\infty$: then, if z is contained in the interior of some transport ray, a simple application of Lemma 14.16 (similar to the one made in second step of Theorem 15.1) gives a constant $K(z)$ such that $\mu(B(z,\varepsilon)) \leq K(z)\|f^+\|_{L^\infty}\varepsilon^N$ for ε sufficiently small, and then $\theta^*_N(\mu,z) < +\infty$. Since this holds for each z contained in the interior of some transport ray, from Proposition 14.14 we have

$$\theta^*_N(\mu,z) < +\infty \quad \text{for } \mu\text{--a.e. } z. \tag{16.3}$$

If now $\mathscr{L}(B) = 0$ and we define $B_m = \{z \in B : \theta^*_N(\mu, z) \leq m\}$, part a) of Theorem 14.6 implies

$$\mu(B_m) \leq 2^N m \mathscr{H}^N(B_m) \leq 2^N m \mathscr{H}^N(B) = 0,$$

which together with (16.3) ensures $\mu(B) = 0$. Then $\mu \ll \mathscr{L}$ and so, since the total mass of μ is bounded by $\mathfrak{C}(\gamma) < +\infty$, $\mu \in L^1$.

Step II. *General case.*

Let $X_m = \{x \in X : f^+(x) \leq m\}$, $\gamma_m = \gamma \llcorner (X_m \times X)$ and let μ_m be defined from γ_m using (14.1). Thanks to Lemma 1.12 and to the first step we can infer that $\mu_m \in L^1$. Arguing as in the proof of Theorem 15.1, $\gamma_m \longrightarrow \gamma$ in $\mathcal{M}^+(X \times X)$, and then $\mu_m \longrightarrow \mu$ in $\mathcal{M}^+(X)$, which gives $\mu \in L^1$ in the general case too.

Step III. *The estimate (16.2).*

This estimate is really trivial: since we proved that $\mu \in L^1$, from (14.2) it immediately follows that

$$\|\mu\|_{L^1} = \mu(X) \leq R\|\gamma\| = R\|f^+\|_{L^1},$$

thus (16.2) is established. $\square$

We remark that, in the sequel, we will always consider situations in which at least one between f^+ and f^- is absolutely continuous: then, we will always denote by μ the transport density, which is a well defined L^1 function thanks to the above result and Theorem 16.2. Note also that the estimate we found is sharp thanks to the remarks of Section 14.4.

Remark 16.5. *The above result is true also if we replace the convex and bounded set X by the whole $\mathbb{R}^N$: in fact, as discussed and shown in Section 14.5, it is true in general that if $f^+ \in L^1$ then $\mu \in L^1$ (under the usual hypothesis that the minimum of the Kantorovich problem is finite). But the inequality (16.2) cannot be extended, as one can understand with easy examples –for instance, one can find situations with $\|f^+\|_{L^1}$ fixed and $\|\mu\|_{L^1}$ arbitrarily large using Example 14.13 with suitable choices of the function g.*

16.3 L^p estimates with summability of one of the measures

In this section we will find estimates for the summability of μ under the hypothesis that f^+ (or f^-) belongs to L^q for some $q > 1$. Theorem 16.4 and the boundedness of X show that $\mu \in L^1$, but we can also prove that μ belongs to L^p for some $p > 1$. Let us begin with the case $q = +\infty$, the general result will then follow. In the sequel we will denote by α' the conjugate exponent of α, i.e. $1/\alpha + 1/\alpha' = 1$.

Lemma 16.6. *If $f^+ \in L^\infty$ then $\mu \in L^p$ for all $p < (2N)'$. More precisely*

$$\|\mu\|_{L^p} \leq K \|f^+\|_{L^\infty}^{1/p'} \|f^+\|_{L^1}^{1/p},$$

where K depends only on R, N and p.

Proof. Given $(x, y) \in spt(\gamma)$, let y_σ be the point on the segment xy such that $\overline{yy_\sigma} = \sigma$, or $y_\sigma = x$ if $\overline{xy} < \sigma$. Let us define the measure μ_r as follows:

$$\langle \mu_r, \varphi \rangle := \iint_{X \times X} \left(\int_X \varphi(z) \, d\mathscr{H}^1_{y_r y_{2r}}(z) \right) d\gamma(x, y). \tag{16.4}$$

Thanks to (14.1), it is clear that $\mu = \mu_R + \mu_{R/2} + \mu_{R/4} + \ldots$, where R is the diameter of X. From Theorem 16.4 we know that $\mu \in L^1$, and then $\mu_r \in L^1$ because $\mu_r \ll \mu$. It is useful to write the measure μ_r of a set A, which from (16.4) is

$$\mu_r(A) = \iint_{X \times X} l(y_r y_{2r} \cap A) \, d\gamma(x, y),$$

where $l(xy \cap A)$ is understood as in (14.2).

Step I. *A particular case.*

Let us fix $r > 0$ and begin with a very particular case: we assume that $spt(f^-) \subseteq Q$, where Q is an hypercube of side λ such that $\lambda\sqrt{N} \leq r/2$ (note that $\lambda\sqrt{N}$ is the diameter of Q). From the definition, it follows that $spt(\mu_r)$ is contained in an annulus S with radii $r - \lambda\sqrt{N}/2$ and $2r + \lambda\sqrt{N}/2$ centered at the center of the hypercube, furthermore $l(y_r y_{2r} \cap S) \leq r$ for all (x, y) and then

$$\|\mu_r\|_{L^1} = \mu_r(S) \leq r f^-(Q). \tag{16.5}$$

On the other hand, the hypothesis $\lambda\sqrt{N} \leq r/2$ ensures that $dist(S, Q) \geq r/2$ and then from Lemma 14.17 (arguing as in the proofs of Theorems 15.1 and 16.4) it follows that

$$\|\mu_r\|_{L^\infty} \leq 2\,\omega_{N-1} \left(\frac{6R}{r/2} \right)^{N-1} R \, \|f^+\|_{L^\infty} = \frac{C\|f^+\|_{L^\infty}}{r^{N-1}}. \tag{16.6}$$

It is well known that if $\varphi \in L^1 \cap L^\infty$ then $\varphi \in L^p$ and

$$\|\varphi\|_{L^p} \leq \|\varphi\|_{L^1}^{1/p} \|\varphi\|_{L^\infty}^{1/p'},$$

then from (16.5) and (16.6) we infer that

$$\|\mu_r\|_{L^p} \leq \big(r f^-(Q) \big)^{1/p} \left(\frac{C\|f^+\|_{L^\infty}}{r^{N-1}} \right)^{1/p'} \leq \frac{K \|f^+\|_{L^\infty}^{1/p'} f^-(Q)^{1/p}}{r^{N-1-N/p}}.$$

Step II. *The general case.*
Let us cover $spt(f^-)$ with $M(r)$ disjoint hypercubes Q_i with sides as in Step I and define $\gamma_i = \gamma \llcorner (X \times Q_i)$. From Lemma 1.12 it follows that γ_i is an optimal transport plan from $f_i^+ = \pi_1 \gamma_i$ to $f_i^- = \pi_2 \gamma_i = f^- \llcorner Q_i$. Define $\mu_{r,i}$ from γ_i using (16.4): from the first step and the fact that $\|f_i^+\|_{L^\infty} \le \|f^+\|_{L^\infty}$ we obtain that $\mu_r = \mu_{r,1} + \mu_{r,2} + \cdots + \mu_{r,M(r)}$ and

$$\|\mu_{r,i}\|_{L^p} \le \frac{K\ \|f^+\|_\infty^{1/p'}\ f^-(Q_i)^{1/p}}{r^{N-1-N/p}}.$$

Applying the inequality between the arithmetic mean and the p-th mean, integrating, extracting the p-th root and using the fact that $f^+(X) = f^-(X)$, we obtain that

$$\|\mu_r\|_{L^p} \le M(r)^{1/p'}\ \frac{K\ \|f^+\|_\infty^{1/p'}}{r^{N-1-N/p}}\ f^+(X)^{1/p}.$$

We recall $\mu = \mu_R + \mu_{R/2} + \mu_{R/4} + \ldots$ and we remark that, up to a geometric constant, $M(r) \le r^{-N}$, then it suffices to show that

$$\sum_{i\in\mathbb{N}} \left(\frac{R}{2^i}\right)^{-N+1+N/p-N/p'} < +\infty; \tag{16.7}$$

a simple calculation ensures that (16.7) holds if and only if $p < (2N)'$, and this concludes the proof. $\square$

We derive now the general result:

Theorem 16.7. *If $f^+ \in L^q$ then $\mu \in L^p$ for all $p < (2N)' \wedge (1 + (q-1)/2)$.*

Proof. Let $p < (2N)' \wedge (1 + (q-1)/2)$ and let us write $X = \bigcup_i X_i$ up to a f^+-negligible set, where $X_i = \{x : i-1 \le f^+(x) < i\}$. As in the previous lemma we can define $\gamma_i = \gamma \llcorner (X_i \times X)$ and μ_i from γ_i using (14.1). Observe that, thanks to Lemma 1.12, γ_i is an optimal transport plan. As $p < (2N)'$, we can then apply Lemma 16.6 obtaining

$$\|\mu_i\|_{L^p} \le K i^{1/p'}\ f^+(X_i)^{1/p} \le K i^{1/p'} (i \cdot |X_i|)^{1/p} \le K i |X_i|^{1/p}.$$

Denoted $\lambda_i = |X_i|$, it is sufficient to prove that $\sum_i i \lambda_i^{1/p} < +\infty$. Let us define

$$\rho_i := \begin{cases} \lambda_i & \text{if } \lambda_i \ge i^{p'(1-q)} \\ 0 & \text{if } \lambda_i < i^{p'(1-q)} \end{cases} \qquad \sigma_i := \begin{cases} 0 & \text{if } \lambda_i \ge i^{p'(1-q)} \\ \lambda_i & \text{if } \lambda_i < i^{p'(1-q)} \end{cases} :$$

it is clear that $\sum_i i\lambda_i^{1/p} = \sum_i i\rho_i^{1/p} + \sum_i i\sigma_i^{1/p}$. Since $p < 1 + (q-1)/2$, the fact that $\sigma_i < i^{p'(1-q)}$ implies, by a simple calculation, that $\sum_i i\sigma_i^{1/p} < +\infty$. Moreover $f^+ \in L^q$ easily implies $\sum_i i^q \lambda_i < +\infty$ and then, since by definition $i\rho_i^{1/p} \le i^q \lambda_i$, $\sum i\rho_i^{1/p} < +\infty$, this gives the assert. $\square$

Remark 16.8. *We do not know if the preceding Theorem gives a sharp estimate or not; in fact, it is presently not known which is the best constant $p = p(q)$ such that $f^+ \in L^q \Longrightarrow \mu \in L^p$ (except for $q = 1$, since we know by Theorem 16.4 that $p(1) = 1$). However, the constant $p(q)$ is not much greater than the one we just found, as one can see for instance from Example 14.12.*

16.4 Sharp L^∞ estimate

From now on, we interest ourselves in the situation when both f^+ and f^- belong to a same space L^q; in this section we consider the case when $q = +\infty$ and we give a strong estimate, namely that if both f^+ and f^- are bounded, then so is μ.

Theorem 16.9. *If both f^+ and f^- are in L^∞, then $\mu \in L^\infty$. More precisely,*

$$\|\mu\|_{L^\infty} \leq 2R\,\omega_{N-1} 12^{N-1}\big(\|f^+\|_{L^\infty} + \|f^-\|_{L^\infty}\big).$$

Proof. Let us define $\Lambda_i = \{(x,y) \in X \times X :\ R/2^{i+1} < \overline{xy} \leq R/2^i\}$, where R is the diameter of X, and $\mu_i,\ i \in \mathbb{N}$, by

$$\mu_i(A) := \iint_{\Lambda_i} l(xy \cap A)\ d\gamma(x,y). \tag{16.8}$$

Recalling (14.2) it is clear that $\mu = \sum_{i\in\mathbb{N}} \mu_i$. We already know that $\mu \in L^1$, and then $\mu_i \in L^1$ for each $i \in \mathbb{N}$: let us give then the last definition,

$$X_i := \{z \in X : \mu_i(z) > 0,\ \nexists (x,y) \in \Lambda_i \cap spt(\gamma)\ s.t.\ z \textit{ is in the interior of } xy\}\,.$$

From (16.8) it follows that $\mu_i(X_i) = 0$, because for each $(x,y) \in \Lambda_i$ we have $xy \cap X_i \subseteq \{x,y\}$ and then $l(xy \cap X_i) = 0$. Then $\mathscr{L}(X_i) = 0$, and this implies $\mathscr{L}(\bigcup_i X_i) = 0$ and then $\mu(\bigcup_i X_i) = 0$.

Let $z \notin \bigcup_i X_i$ be a Lebesgue point for each of the μ_i such that $\mu(z) = \sum_i \mu_i(z) > 0$, and let j be the least integer such that $\mu_j(z) > 0$. Since $z \notin X_j$, there exists a transport ray uv such that z is in the interior of uv and $R/2^{j+1} < \overline{uv} \leq R/2^j$, and then if $\overline{xy} \leq R/2^j$ we have $\overline{xy} \leq 2\overline{uv}$. Let us then denote by C_r the cylinder of axis uv and radius $12r$: thanks to Lemma 14.15, if $r \ll 1$

$$\begin{aligned}\sum_{i=j}^{\infty} \mu_i(B(z,r)) &= \iint_{\overline{xy} \leq R/2^j} l(B(z,r) \cap xy)\ d\gamma(x,y) \\ &\leq 2r\left(\iint_{C_r \times X} d\gamma(x,y) + \iint_{X \times C_r} d\gamma(x,y)\right) \leq 2r\left(f^+(C_r) + f^-(C_r)\right).\end{aligned}$$

Since the volume of the cone C_r is less than $R\,\omega_{N-1}(12r)^{N-1}$,

$$\sum_{i=j}^{\infty} \mu_i(B(z,r)) \leq 2rR\,\omega_{N-1}(12r)^{N-1}\left(\|f^+\|_{L^\infty} + \|f^-\|_{L^\infty}\right),$$

which gives

$$\mu(z) = \sum_{i=j}^{\infty} \mu_i(z) \leq 2R\,\omega_{N-1}12^{N-1}\left(\|f^+\|_{L^\infty} + \|f^-\|_{L^\infty}\right):$$

since this inequality holds for $\mu-$a.e. point z, this completes the proof. $\square$

16.5 L^p estimates

After the results of the preceding sections, a natural question that arises is what happens when both f^+ and f^- are in L^q for $1 < q < +\infty$: we already know, by Example 14.11, that μ needs not to belong to any L^p with $p > q$, but we can expect some result stronger than Theorem 16.7, due to the summability of both the measures f^+ and f^-. In particular, one can ask whether also μ is in L^q: as far as we saw by now, this is true only in the case when $q = 1, +\infty$. In fact, this conjecture happens to be true, and we will prove it in the next two chapters, respectively via a PDE argument and an interpolation argument. In this section, we see how our geometrical arguments can show that if f^+ and f^- are in L^q then $\mu \in L^p$ for any $p < q$, then only the implication $L^q \Rightarrow L^q$ remains to be proved; moreover, we will also show this sharp implication under the assumption that the supports of f^+ and f^- are disjoint (since they are closed and bounded, this means that there is a strictly positive distance between them).

We start with the following technical lemma:

Lemma 16.10. *If* $g = \sum_{i\in X} |g_i|$, $g_i \in L^\infty \cap L^1$, $X \subseteq \mathbb{Z}$ *and* $1 \leq p < +\infty$, *then*

$$\int g^p dx \leq \sum_{i\in X} \frac{\|g_i\|_{L^1}}{\|g_i\|_{L^\infty}} \left(\sum_{j\leq i} \|g_i\|_{L^\infty}\right)^p.$$

Proof. Let us first assume that X is finite. Then, in the subset of $(L^\infty \cap L^1)^X$ for which $\|g_i\|_{L^1}$ and $\|g_i\|_{L^\infty}$ are fixed, the maximum of $\int g^p$ is attained in the case $g_i = \|g_i\|_{L^\infty}\chi_{A_i}$, where $\mathscr{L}(A_i)\|g_i\|_{L^\infty} = \|g_i\|_{L^1}$ and the sets A_i intersect as much as possible. The formula is then straightforward if X is composed by two elements, and an easy induction argument implies the validity in general.

Since all the terms are positive, a passage to the limit gives the assert also for X infinite. $\square$

We prove now the

Theorem 16.11. *If both $f^+,\ f^- \in L^q$, then $\mu \in L^p$ for $1 \le p < q$.*

Proof. Let us define $X_i^+,\ X_i^-,\ \mu_{ij}$ and λ_{ij} with $i, j \in \mathbb{N}$ as follows

$$
\begin{aligned}
X_0^{\pm} &:= \left\{x \in X:\ f^{\pm}(x) \le 1\right\}, & X_i^{\pm} &:= \left\{x \in X:\ 2^{i-1} < f^{\pm}(x) \le 2^i\right\},\\
\mu_{ij} &:= \iint_{X_i^+ \times X_j^-} l(xy \cap A)\ d\gamma(x,y), & \lambda_{ij} &:= \gamma\left(X_i^+ \times X_j^-\right):
\end{aligned}
$$

then $\mu = \sum_{i,j} \mu_{ij}$, it is easy to give the bound

$$
\|f^{\pm}\|_{L^q}^q \ge \sum_{i \ge 1} 2^{(i-1)q} \mathscr{L}\left(X_i^{\pm}\right) \ge \sum_{i \ge 1} \frac{2^{i(q-1)} f^{\pm}(X_i^{\pm})}{2^q} \tag{16.9}
$$

and we have also $\sum_j \lambda_{ij} = f^+(X_i^+)$ and $\sum_i \lambda_{ij} = f^-(X_j^-)$. From the definition it follows that $\|\mu_{ij}\|_{L^1} \le R\lambda_{ij}$ and, since $\gamma \llcorner (X_i \times X_j)$ is an optimal transport plan for the problem given by his marginals, thanks to Theorem 16.9 we have $\|\mu_{ij}\|_{L^\infty} \le K(2^i + 2^j)$. Thanks to these estimates, to give an upper bound for the L^p norm of μ we can make use of Lemma 16.10 with $X = \mathbb{N}$ after given an isomorphism between $\mathbb{N}$ and $\mathbb{N} \times \mathbb{N}$, or equivalently an order to $\mathbb{N} \times \mathbb{N}$. We do it by setting $(m,n) < (i,j)$ if $m \vee n < i \vee j$, or if $m \vee n = i \vee j$ and $m \wedge n < i \wedge j$, or if $m = j$, $n = i$ and $m < n$. Then Lemma 16.10 gives

$$
\|\mu\|_{L^p}^p \le \sum_{i,j} \frac{R\lambda_{ij}}{K\left(2^i + 2^j\right)} \Bigg(\sum_{(m,n) \le (i,j)} K\left(2^m + 2^N\right) \Bigg)^p. \tag{16.10}
$$

But if $i \ge j$ then

$$
\sum_{(m,n) \le (i,j)} 2^m + 2^N \le 2(i+1)\left(1 + 2 + 4 + \dots + 2^i\right) \le (i+1)\ 2^{i+2},
$$

and then from (16.10) we have

$$
\begin{aligned}
\|\mu\|_{L^p}^p &\le RK^{p-1} \Bigg(\sum_{i \ge j} \frac{\lambda_{ij}\left((i+1)\ 2^{i+2}\right)^p}{2^i} + \sum_{i \le j} \frac{\lambda_{ij}\left((j+1)\ 2^{j+2}\right)^p}{2^j} \Bigg)\\
&\le 4^p RK^{p-1} \Bigg(\sum_{i \ge j} \lambda_{ij}\ (i+1)^p\ 2^{i(p-1)} + \sum_{i \le j} \lambda_{ij}\ (j+1)^p\ 2^{j(p-1)} \Bigg)\\
&\le \widetilde{K} \Bigg(\sum_i \Big((i+1)^p\ 2^{i(p-1)} \sum_j \lambda_{ij} \Big) + \sum_j \Big((j+1)^p\ 2^{j(p-1)} \sum_i \lambda_{ij} \Big) \Bigg)\\
&\le \widetilde{K} \Bigg(\sum_i (i+1)^p\ 2^{i(p-1)} f^+(X_i^+) + \sum_j (j+1)^p\ 2^{j(p-1)} f^-(X_j^-) \Bigg).
\end{aligned}
$$

By (16.9) and the fact that $p < q$, this proves the assert. $\square$

The sharp $L^q \Rightarrow L^q$ result can be obtained with our geometrical arguments under the additional assumption that the supports of f^+ and f^- are disjoint. This hypothesis of disjointness plays an important role in many other situations and it is assumed, for instance, in [39], [37] and some claims of [47].

Theorem 16.12. *If both f^+ and f^- are in L^p and $dist\big(spt(f^+), spt(f^-)\big) = \delta > 0$, then $\mu \in L^p$. More precisely,*

$$\|\mu\|_{L^p} \leq \frac{K\big(\|f^+\|_{L^p} + \|f^-\|_{L^p}\big)}{\delta^{(N-1)/p'}},$$

where K depends only on R, N and p.

Proof. Let us define $B^\pm = \{x \in X : d(x, spt(f^\pm) \geq \delta/2\}$: since δ is the distance between the supports of f^+ and f^-, $X = B^+ \cup B^-$, then we will study the summability of $\mu^\pm = \mu \llcorner B^\pm$. Let us first define, for $i \in \mathbb{Z}$, $X_i = \{x : 2^i < f^+(x) \leq 2^{i+1}\}$, $f_i^+ = f^+ \llcorner X_i$, $\gamma_i = \gamma \llcorner (X_i \times X)$, $\tilde{\mu}_i$ from γ_i using (14.1) and $\mu_i = \tilde{\mu}_i \llcorner B^-$: it is clear that $\mu^- = \sum_{i\in\mathbb{Z}} \mu_i$. Moreover thanks to the definition of f_i^+ we have $\|f_i^+\|_{L^1} \leq 2^{i+1}\mathscr{L}(X_i)$ and $\|f_i^+\|_{L^p}^p \geq 2^{ip}\mathscr{L}(X_i)$, and then $\|f_i^+\|_{L^p}^p \geq 2^{ip-i-1}\|f_i^+\|_{L^1}$; since $f^+ = \sum_{i\in\mathbb{Z}} f_i^+$ and the supports of f_i^+ are disjointed, it follows

$$\|f^+\|_{L^p}^p \geq \frac{1}{2}\sum_{i\in\mathbb{Z}} 2^{i(p-1)}\|f_i^+\|_{L^1}. \tag{16.11}$$

Thanks to the definition of μ_i, the L^1-norm of μ_i is less then $R\|f_i^+\|_{L^1}$; moreover, using Lemma 14.17 we know that if $x \in B^-$ and ε is sufficiently small, then

$$\mu_i(B(x,\varepsilon)) \leq 2\varepsilon f_i^+(C_\varepsilon) \leq 2\varepsilon 2^{i+1} R\omega_{N-1}\left(\frac{6R\varepsilon}{\delta/2}\right)^{N-1}:$$

summarizing, we have found the following estimates for μ_i

$$\|\mu_i\|_{L^1} \leq R\|f_i^+\|_{L^1} \qquad\qquad \|\mu_i\|_{L^\infty} \leq \frac{\omega_{N-1}6^{N-1}R^N 2^{i+2}}{\big(\delta/2\big)^{N-1}}.$$

Thanks to Lemma 16.10 with $X = \mathbb{Z}$, this gives

$$\begin{aligned}
\int (\mu^-)^p &\leq \sum_{i\in\mathbb{Z}} \frac{\|\mu_i\|_{L^1}}{\|\mu_i\|_{L^\infty}}\left(\sum_{j=-\infty}^{i}\|\mu_i\|_{L^\infty}\right)^p \\
&\leq R^{1+N(p-1)}\big(\delta/2\big)^{(N-1)(1-p)}\omega_{N-1}^{p-1}6^{(N-1)(p-1)}\sum_{i\in\mathbb{Z}} 2^{(i+3)p-(i+2)}\|f_i^+\|_{L^1} \\
&\leq R^{1+N(p-1)}\big(\delta/2\big)^{(N-1)(1-p)}\omega_{N-1}^{p-1}6^{(N-1)(p-1)}2^{3p-1}\|f^+\|_{L^p}^p,
\end{aligned}$$

using also (16.11). To estimate the L^p norm of μ^+, clearly, we can make exactly the same calculation above changing all the "+" with "−" and vice versa and then, since $\mu \leq \mu^+ + \mu^-$, we can conclude the proof with the estimate

$$\|\mu\|_{L^p} \leq \frac{2^{3-1/p} R^{1+(N-1)/p'} \omega_{N-1}^{1/p'} 12^{(N-1)/p'}}{\delta^{(N-1)/p'}} \Big(\|f^+\|_{L^p} + \|f^-\|_{L^p} \Big).$$

□

We didn't find a geometrical proof of the fact that Theorem 16.11 is true even with $p = q$ without additional hypotheses; however, in the next two chapters we will prove this fact with two different strategies.

Chapter 17

Sharp L^p estimates for $p \geq 2$ via PDE

In this chapter we see a PDE method, presented in [33], which allows to determine that the transport density is in L^p if so are f^+ and f^-, for each $p \geq 2$. In the next chapter we will see that an interpolation argument shows that the same claim is true also for $1 < p < 2$.

17.1 Approximation

For simplicity, only for this chapter we consider the case –of course not restrictive– when $X = B(0, R)$, the open ball centered in $0 \in \mathbb{R}^N$ and with radius R. We also, as usual, denote $f = f^+ - f^-$, which is an L^1 function with $\int_X f\,dx = 0$.

Denote by u_k the solution of the nonlinear boundary–value problem

$$\begin{cases} -\mathrm{div}\,(\sigma_k Du_k) = f & \text{in } X \\ u_k = 0 & \text{on } \partial X, \end{cases} \tag{17.1}$$

where we write

$$\sigma_k := e^{\frac{k}{2}\left(|Du_k|^2 - 1\right)}. \tag{17.2}$$

Observe that u_k is the unique minimizer of the functional

$$F_k(v) := \int_X \frac{1}{k} e^{\frac{k}{2}(|Dv|^2-1)} - f v\,dx$$

in $W_0^{1,k}$. This approximation is suggested by the recent paper [38]. Regularity theory (see [59]) implies that u_k is smooth, provided f is.

We want to study the limits of u_k and σ_k as $k \to +\infty$, and begin with some uniform bounds.

Lemma 17.1. *The sequence $\{u_k\}_{k\in\mathbb{N}}$ is bounded in $W_0^{1,q}(X)$, for each $1 \leq p < +\infty$.*

Proof. Observe first that $x \le e^{\frac{x^2-1}{2}}$ for $x \ge 0$, and therefore that $|Du_k| \le \sigma_k^{1/k}$. Recalling then (17.1) and (17.2), we deduce for $k > n$ that

$$\int_X |Du_k|^{k+2}\, dx \le \int_X |Du_k|^2 \sigma_k \, dx = \int_X f u_k \, dx \le C\|u_k\|_{L^\infty} \le C\|Du_k\|_{L^k}.$$

Note that $\|Du_k\|_{L^k}^k \le \|Du_k\|_{L^{k+2}}^{k+2} + C$. Hence $\|Du_k\|_{L^k}^k \le C + C\|Du_k\|_{L^k}$, and so $\|Du_k\|_{L^k} \le C$. We deduce for each $k > p$ that

$$\|Du_k\|_{L^p} \le \|Du_k\|_{L^k} \, \|1\|_{L^{\frac{kp}{k-p}}} \le C.$$

□

We next indentify the Γ–limit of problem (17.1) –subject to the constraint (17.2)– as $k \to +\infty$; about the $\Gamma-$convergence, recall the brief introduction we gave in Section 8.2, for a more precise reference see [28]. We define

$$F(v) := \begin{cases} -\int_X f v \, dx & \text{if } v \in \mathrm{Lip}(X),\ |Dv| \le 1 \text{ a.e. };\\ +\infty & \text{otherwise.} \end{cases} \tag{17.3}$$

Proposition 17.2. *$F_k \xrightarrow{\Gamma} F$ with respect to the uniform convergence of functions.*

Proof. **Step I.** *Reduction to $E_k \xrightarrow{\Gamma} E$.*
Since the mapping $u \mapsto \langle f, u\rangle = \int_X f u \, dx$ is linear, it is enough to prove

$$E_k(v) := \frac{1}{k}\int_X e^{\frac{k}{2}(|Dv|^2-1)} dx \xrightarrow{\Gamma} E(v),$$

for

$$E(v) := \begin{cases} 0 & \text{if } v \in \mathrm{Lip}(X),\ |Dv| \le 1 \text{ a.e. };\\ +\infty & \text{otherwise.} \end{cases}$$

Step II. *Proof of the $\Gamma-$convergence.*
If $E(v) < +\infty$, we clearly have

$$E(v) = 0 = \lim_{k\to+\infty} E_k(v).$$

Suppose now that $v_k \to v$ uniformly, and $\limsup_{k\to+\infty} E_k(v_k) \le C < +\infty$. Fix an integer m and let $k > m$. Since $e^{\frac{x^2-1}{2}} \ge x$, we have for each open set $V \subseteq X$ that

$$\begin{aligned} \left(\int_V |Dv_k|^m \, dx\right)^{1/m} &\le |V|^{1/m-1/k} \left(\int_V |Dv_k|^k dx\right)^{1/k} \\ &\le |V|^{1/m-1/k}\, k^{1/k}\, E_k(v_k)^{1/k} \le |V|^{1/m-1/k}\, k^{1/k}\, C^{1/k}. \end{aligned}$$

Passing to limits in k and recalling the lower semicontinuity of the L^m norm of the gradient, we discover

$$\left(\int_V |Dv|^m dx\right)^{1/m} \leq |V|^{1/m}.$$

This inequality, valid for all V as above, implies that Dv is in L^∞, with $|Dv| \leq 1$ almost everywhere. Consequently,

$$E(v) = 0 \leq \liminf_{k\to\infty} E_k(v_k).$$

$\square$

We can prove now the following approximation result:

Proposition 17.3. *Take a sequence $\mathbf{G}_k = \sigma_k Du_k$, where σ_k and u_k are defined via (17.1) and (17.2) for some bounded sequence $f_k \in L^p(X)$, where $1 < p < +\infty$; suppose also that we have the uniform bound*

$$\sup_{k\in\mathbb{N}} \|\mathbf{G}_k\|_{L^p(X,\mathbb{R}^N)} < +\infty \tag{17.4}$$

and that

$$\begin{cases} f_k \rightharpoonup f & \text{weakly in } L^p(X); \\ \mathbf{G}_k \rightharpoonup \mathbf{G} & \text{weakly in } L^p(X,\mathbb{R}^N); \\ u_k \longrightarrow u & \text{uniformly.} \end{cases} \tag{17.5}$$

Then there exists a positive function $a \in L^p$ such that

$$\begin{cases} \mathbf{G} = aDu; \\ |Du| = 1 \quad a.e. \text{ on } \{a > 0\}; \\ \sigma_k \rightharpoonup a \quad \text{weakly in } L^p(X). \end{cases} \tag{17.6}$$

In particular, it follows that $a = |\mathbf{G}|$ and that $-\mathrm{div}\ \mathbf{G} = f$

Proof. **Step I.** *The estimate (17.7).*
Let us fix $0 < \lambda < 1$ and calculate:

$$\begin{aligned} \int_X |\mathbf{G}|\,dx &\leq \liminf_{k\to\infty} \int_X |\mathbf{G}_k|\,dx = \liminf_{k\to\infty}\left(\int_{\{|Du_k|^2>1-\lambda\}} |\mathbf{G}_k|\,dx + \int_{\{|Du_k|^2\leq 1-\lambda\}} |\mathbf{G}_k|\,dx\right) \\ &\leq \liminf_{k\to\infty}\left(\frac{1}{\sqrt{1-\lambda}}\int_X |\mathbf{G}_k||Du_k|\,dx + \int_X e^{-\frac{k}{2}\lambda}\sqrt{1-\lambda}\,dx\right). \end{aligned}$$

When k goes to infinity, the last integral goes to 0. Notice also that

$$\int_X |\mathbf{G}_k||Du_k|\,dx = \int_X \sigma_k |Du_k|^2\,dx = \int_X f_k u_k\,dx.$$

Therefore

$$\sqrt{1-\lambda}\int_X |\mathbf{G}|\,dx \le \liminf_{k\to\infty}\int_X f_k u_k\,dx = \int_X fu\,dx = \int_X \mathbf{G}\cdot Du\,dx$$

for each $0<\lambda<1$, and consequently

$$\int_X |\mathbf{G}|\,dx \le \int_X \mathbf{G}\cdot Du\,dx. \tag{17.7}$$

Step II. *First two parts of (17.6).*

Reasoning now as in the proof of Proposition 17.2, we fix an integer m and let $k>m$. Then for each open set $V\subseteq X$

$$\begin{aligned}\left(\int_V |Dv_k|^m\ dx\right)^{1/m} &\le |V|^{1/m-1/k+1}\left(\int_V |Dv_k|^{k+1}dx\right)^{1/k+1}\\ &\le |V|^{1/m-1/k+1}||\mathbf{G}_k||_{L^1}^{1/k+1} \le |V|^{1/m-1/k+1}C^{1/k+1}.\end{aligned}$$

Pass to limits in k to find

$$\left(\int_V |Dv|^m dx\right)^{1/m} \le |V|^{1/m},$$

and therefore $|Du|\le 1$ almost everywhere. The existence of a unique $a\in L^p$ such that the first two assertions of (17.6) hold now follows from (17.7).

Step III. *Third part of (17.6): beginning.*

To show also that $\sigma_k \rightharpoonup a$, let us fix $\psi\in C_0^\infty$ and prove

$$\int_X \sigma_k\psi\,dx \to \int_X a\psi\,dx.$$

We write

$$\int_X \sigma_k\psi\,dx = \int_X \sigma_k|Du_k|^2\psi\,dx + \int_X \sigma_k(1-|Du_k|^2)\psi\,dx =: A_1+A_2.$$

Notice now that

$$\begin{aligned}A_1 &= \int_X \psi\,\mathbf{G}_k\cdot Du_k\,dx = \int_X \mathbf{G}_k\cdot D(u_k\psi)\,dx - \int_X u_k\,\mathbf{G}_k\cdot D\psi\,dx\\ &= \int_X f_k u_k\psi\,dx - \int_X u_k\mathbf{G}_k\cdot D\psi\,dx.\end{aligned}$$

This expression converges as $k\to+\infty$ to

$$\begin{aligned}\int_X fu\psi\,dx - \int_X u\,\mathbf{G}\cdot D\psi\,dx &= \int_X \mathbf{G}\cdot D(u\psi)\,dx - \int_X u\,\mathbf{G}\cdot D\psi\,dx\\ &= \int_X \psi\,\mathbf{G}\cdot Du\,dx = \int_X \psi a|Du|^2\,dx = \int_X a\psi\,dx.\end{aligned}$$

Step IV. *Third part of (17.6): conclusion.*
It remains to show that $A_2 \to 0$. If we write $\varphi_k := |Du_k|^2 - 1$, then

$$|A_2| \le \|\psi\|_{L^\infty} \int_X e^{\frac{k}{2}\varphi_k} |\varphi_k| \, dx.$$

Since $xe^{-\frac{x}{2}} \le 1$ for each $x > 0$, we have

$$\int_{\{\varphi_k<0\}} e^{\frac{k}{2}\varphi_k} |\varphi_k| \, dx = \frac{1}{k} \int_X k|\varphi_k| e^{-\frac{k|\varphi_k|}{2}} \, dx \le \frac{|X|}{k} \to 0.$$

Finally, since $p > 1$ there exists a constant $c_p > 0$ such that

$$\frac{e^{x(p-1)}}{x} \ge c_p > 0$$

for all $x > 0$. Consequently,

$$\begin{aligned}
\int_{\{\varphi_k>0\}} e^{\frac{k}{2}\varphi_k} |\varphi_k| \, dx &= \frac{2}{k} \int_{\{\varphi_k>0\}} e^{\frac{k}{2}\varphi_k} \, \frac{k}{2}\varphi_k \, dx \\
&\le \frac{2}{c_p k} \int_{\{\varphi_k>0\}} e^{\frac{pk}{2}\varphi_k} \, dx = \frac{2}{c_p k} \int_X \sigma_k^p \, dx \le \frac{2C^p}{c_p k} \to 0.
\end{aligned}$$

This completes the proof that $A_2 \to 0$. □

17.2 Sharp L^p estimate

In this section we derive our integral estimate and we infer the summability of the transport density.

Proposition 17.4. *Assume that $2 \le p < +\infty$ and that $f \in C^\infty(\overline{X})$. Then there exist a constant K, depending only on $\|f\|_{L^\infty}$, and a constant C, depending only on p and on X, such that*

$$\int_X \sigma_k^p |Du_k|^p \, dx \le C \left(\int_X |f|^p \, dx + 1 \right) \tag{17.8}$$

for all $k \ge K$.

Proof. **Step I.** *Some introductive calculations.*
For notational simplicity we hereafter write σ and u in place of σ_k and u_k. We note that, since f is smooth, the same is true for u and σ. Observe also the bound $|u| \le C$. We record for later reference these consequences of (17.2):

$$|Du|_i = \frac{\sigma_i}{k\sigma|Du|}, \qquad\qquad u_i u_{ij} = \frac{\sigma_j}{k\sigma}. \tag{17.9}$$

We now multiply the PDE (17.1) by $\sigma^{p-1}|Du|^{p+1}u$ and integrate by parts, to find

$$\int_X \sigma Du \cdot D\left(\sigma^{p-1}|Du|^{p+1}u\right)\,dx = \int_X \sigma^{p-1}|Du|^{p+1}uf\,dx. \tag{17.10}$$

The right hand term in (17.10) is less than or equal to

$$C\int_X \sigma^{p-1}|Du|^{p+1}|f|\,dx \le \frac{1}{2}\int_{\left\{|f|\le\frac{\sigma|Du|}{2C}\right\}} \sigma^p|Du|^{p+2}\,dx + 2^{p-1}\,C^p\int_{\left\{|f|>\frac{\sigma|Du|}{2C}\right\}} |Du|^2|f|^p\,dx.$$

But if $\sigma|Du| < 2C|f|$, then obviously $\sigma|Du| \le 2C\|f\|_{L^\infty}$. Recalling (17.2), we see that this implies $|Du| \le 2$ provided $k \ge K$, for some constant K depending only upon $\|f\|_{L^\infty}$. Therefore

$$\int_X \sigma^{p-1}|Du|^{p+1}uf\,dx \le \frac{1}{2}\int_X \sigma^p|Du|^{p+2}\,dx + C\int_X |f|^p\,dx. \tag{17.11}$$

Step II. *The first main estimate.*
We use (17.9) to evaluate the left hand term in (17.10):

$$\begin{aligned}\int_X \sigma Du \cdot D\left(\sigma^{p-1}|Du|^{p+1}u\right)\,dx &= \int_X \sigma^p|Du|^{p+3}\,dx\\ +(p-1)\int_X \sigma^{p-1}|Du|^{p+1}u\,D\sigma\cdot Du\,dx &+ (p+1)\int_X \sigma^p u|Du|^p Du\cdot(D|Du|)\,dx\\ &= \int_X \sigma^p|Du|^{p+3}\,dx\\ +(p-1)\int_X \sigma^{p-1}|Du|^{p+1}u\,D\sigma\cdot Du\,dx &+ \frac{p+1}{k}\int_X \sigma^{p-1}u|Du|^{p-1}Du\cdot D\sigma\,dx.\end{aligned} \tag{17.12}$$

But $\sigma \ge 1$ only if $|Du| \ge 1$; and hence

$$\int_X \sigma^p|Du|^{p+2}\,dx \le \int_X \sigma^p|Du|^{p+3}\,dx + C, \tag{17.13}$$

since X is bounded.

Combining (17.13), (17.12), (17.10) and (17.11), we deduce the inequality

$$\begin{aligned}\int_X \sigma^p|Du|^{p+2}\,dx &\le C + \frac{1}{2}\int_X \sigma^p|Du|^{p+2}\,dx + C\int_X |f|^p\,dx\\ &\quad + C\int_X \sigma^{p-1}|Du|^{p+1}|D\sigma\cdot Du|\,dx + \frac{C}{k}\int_X \sigma^{p-1}|Du|^{p-1}|D\sigma\cdot Du|\,dx.\end{aligned}$$

Arguing as before (this means dividing the integrals in the set where $|D\sigma\cdot Du| \le \varepsilon\,\sigma|Du|/C$ and in the rest of X), we see that therefore

$$\begin{aligned}\int_X \sigma^p|Du|^{p+2}\,dx &\le C + C\int_X |f|^p\,dx + \varepsilon\int_X \sigma^p|Du|^{p+2}\,dx + \frac{\varepsilon}{k}\int_X \sigma^p|Du|^p\,dx\\ &\quad + \frac{C^2}{\varepsilon}\int_X \sigma^{p-2}|Du|^p|D\sigma\cdot Du|^2\,dx + \frac{C^2}{k\varepsilon}\int_X \sigma^{p-2}|Du|^{p-2}|D\sigma\cdot Du|^2\,dx\end{aligned}$$

for any $\varepsilon > 0$. Since $\int_X \sigma^p |Du|^p\,dx \le \int_X \sigma^p |Du|^{p+2}\,dx + C$, this implies our first main estimate:

$$\begin{aligned}\int_X \sigma^p |Du|^{p+2}\,dx \le C + C\int_X |f|^p\,dx + C\int_X \sigma^{p-2}|Du|^p |D\sigma\cdot Du|^2\,dx \\ + \frac{C}{k}\int_X \sigma^{p-2}|Du|^{p-2}|D\sigma\cdot Du|^2\,dx.\end{aligned} \tag{17.14}$$

Step III. *The second main estimate.*

The last two terms in (17.14) involving $D\sigma\cdot Du$ are dangerous, since $D\sigma$ is of order k: we need another estimate to control them.

Let us therefore continue by multiplying the PDE (17.1) by $-\mathrm{div}\ (\sigma^{p-1}|Du|^p Du)$ and thereby deriving the identity

$$\int_X \mathrm{div}\ (\sigma Du)\,\mathrm{div}\ (\sigma^{p-1}|Du|^p Du)\,dx = -\int_X f\mathrm{div}\ (\sigma^{p-1}|Du|^p Du)\,dx. \tag{17.15}$$

The term on the right equals

$$\begin{aligned}\int_X f\,\sigma^{p-2}|Du|^p\,(-\mathrm{div}\ (\sigma Du))\,dx - \int_X f\sigma Du\cdot D\left(\sigma^{p-2}|Du|^p\right)\,dx = \int_X |f|^2\sigma^{p-2}|Du|^p\,dx \\ -(p-2)\int_X f\sigma^{p-2}|Du|^p Du\cdot D\sigma\,dx - p\int_X f\sigma^{p-1}|Du|^{p-1}Du\cdot (D|Du|)\,dx.\end{aligned}$$

We again recall (17.9) and deduce

$$\begin{aligned}-\int_X f\mathrm{div}\ (\sigma^{p-1}|Du|^p Du)\,dx \le \int_X |f|^2\sigma^{p-2}|Du|^p\,dx \\ +(p-2)\int_X |f|\sigma^{p-2}|Du|^p|Du\cdot D\sigma|\,dx + \frac{q}{k}\int_X |f|\sigma^{p-2}|Du|^{p-2}|Du\cdot D\sigma|\,dx.\end{aligned} \tag{17.16}$$

The left hand term of (17.15) is

$$\begin{aligned}A := {}&-\int_X \sigma u_i(\sigma^{p-1}|Du|^p u_j)_{ij}\,dx + \int_{\partial X}\sigma u_i\nu^i(\sigma^{p-1}|Du|^p u_j)_j\,d\mathscr{H}^{N-1}\\ ={}&\int_X (\sigma u_i)_j(\sigma^{p-1}|Du|^p u_j)_i\,dx \\ &+\int_{\partial X}\sigma u_i\nu^i(\sigma^{p-1}|Du|^p u_j)_j - \sigma u_i\nu^j(\sigma^{p-1}|Du|^p u_j)_i\,d\mathscr{H}^{N-1},\end{aligned} \tag{17.17}$$

where $\nu = (\nu^1,\dots,\nu^N)$ is the unit outer normal to ∂X. If we call B the boundary integral in last equation, it is

$$\begin{aligned}B = {}&(p-1)\int_{\partial X}\sigma^{p-1}|Du|^p\big(u_i\nu^i u_j\sigma_j - u_i\nu^j\sigma_i u_j\big)\,d\mathscr{H}^{N-1}\\ &+\int_{\partial X}\sigma^p\big(u_i\nu^i u_j(|Du|^p)_j - u_i\nu^j u_j(|Du|^p)_i\big)\,d\mathscr{H}^{N-1}\\ &+\int_{\partial X}\sigma^q|Du|^p\big(u_i\nu^i u_{jj} - u_i\nu^j u_{ij}\big)\,d\mathscr{H}^{N-1} = B_1+B_2+B_3.\end{aligned}$$

It is clear, by symmetry, that $B_2 = 0$; it is also immediate to infer $B_1 = 0$ using (17.9) to evaluate σ_i and σ_j: we are then reduced to consider $B = B_3$.

To do that, consider a point $x_0 \in \partial X$; without loss, we can take $x_0 = (0, \dots, R)$: then $\nu = (0, \dots, 1)$ and $Du = (0, \dots, u_n)$, since $u = 0$ on ∂X. Consequently, one has

$$B_3 = \sigma^q |Du|^p \big(\Delta u - u_{nn}\big)\, u_n.$$

Lastly, write $x' = (x_1, \dots, x_{n-1})$ and observe that $u(x', \sqrt{R^2 - |x'|^2}) \equiv 0$ for small x'. We differentiate this identity twice and set $x' = 0$, to compute $\Delta u - u_{nn} = \frac{n-1}{R}\, u_n$ at x_0. Hence

$$B = B_3 = \frac{n-1}{R} \int_{\partial X} \sigma^q |Du|^{p+2}\, d\mathscr{H}^{N-1} \geq 0.$$

Finally, by (17.17) one obtains

$$\begin{aligned}
A &= \int_X (\sigma u_i)_i \left(\sigma^{p-1}|Du|^p u_j\right)_j dx \geq \int_X (\sigma u_i)_j \left(\sigma^{p-1}|Du|^p u_j\right)_i dx \\
&= \int_X (\sigma u_{ij} + \sigma_j u_i) \Big(\sigma^{p-1}|Du|^p u_{ij} + (p-1)\sigma^{p-2}|Du|^p \sigma_i u_j + \frac{q}{k}\sigma^{p-2}|Du|^{p-2}\sigma_i u_j\Big)\, dx \\
&= \int_X \sigma^p |Du|^p |D^2 u|^2 + \frac{q}{k}\sigma^{p-2}|Du|^p |D\sigma|^2 + (p-1)\sigma^{p-2}|Du|^p |D\sigma \cdot Du|^2 + \\
&\quad + \frac{q}{k^2}\, \sigma^{p-2}|Du|^{p-2}|D\sigma|^2 + \frac{q}{k}\, \sigma^{p-2}|Du|^{p-2}|D\sigma \cdot Du|^2\, dx.
\end{aligned}$$

The first, the second and the fourth terms in the last expression are positive, and so we deduce

$$\begin{aligned}
(p-1)\int_X \sigma^{p-2}|Du|^p |D\sigma \cdot Du|^2\, dx + \frac{q}{k}\int_X \sigma^{p-2}|Du|^{p-2}|D\sigma \cdot Du|^2\, dx& \\
\leq \int_X \mathrm{div}\,(\sigma Du)\mathrm{div}\,(\sigma^{p-1}|Du|^p Du)\, dx.&
\end{aligned} \tag{17.18}$$

Collecting (17.18), (17.15) and (17.16), we find

$$\begin{aligned}
\int_X \sigma^{p-2}|Du|^p |D\sigma \cdot Du|^2\, dx + \frac{1}{k}\int_X \sigma^{p-2}|Du|^{p-2}|D\sigma \cdot Du|^2\, dx& \\
\leq C\int_X |f|^2 \sigma^{p-2}|Du|^p\, dx + C\int_X |f|\sigma^{p-2}|Du|^p |Du \cdot D\sigma|\, dx& \\
+ \frac{C}{k}\int_X |f|\sigma^{p-2}|Du|^{p-2}|Du \cdot D\sigma|\, dx.&
\end{aligned} \tag{17.19}$$

Take $\varepsilon > 0$ to be a small constant, which will be fixed later on. Then

$$\begin{aligned}
\int_X f^2 \sigma^{p-2}|Du|^p\, dx &\leq \varepsilon \int_X \sigma^p |Du|^{p+2}\, dx + C\int_{\{|f| > \sigma|Du|\sqrt{\varepsilon}\}} |f|^p |Du|^2\, dx \\
&\leq \varepsilon \int_X \sigma^p |Du|^{p+2}\, dx + C\int_X |f|^p\, dx,
\end{aligned} \tag{17.20}$$

since $|Du| \le 2$ wherever $|f| > \sigma|Du|\sqrt{\varepsilon}$, provided $k \ge K$ and K is a large constant depending only on $\|f\|_{L^\infty}$. Recalling (17.20), we can likewise estimate for each $\delta > 0$ that

$$\begin{aligned}\int_X |f|\sigma^{p-2}|Du|^p|D\sigma\cdot Du|\,dx &\le \delta\int_X \sigma^{p-2}|Du|^p|D\sigma\cdot Du|^2\,dx + C\int_X f^2\sigma^{p-2}|Du|^p\,dx\\ &\le \delta\int_X \sigma^{p-2}|Du|^p|D\sigma\cdot Du|^2\,dx + C\left(\varepsilon\int_X \sigma^p|Du|^{p+2}\,dx + \int_X |f|^p\,dx\right). \quad (17.21)\end{aligned}$$

Similarly,

$$\begin{aligned}\int_X |f|\sigma^{p-2}|Du|^{p-2}|Du\cdot D\sigma|dx &\le \delta\int_X \sigma^{p-2}|Du|^{p-2}|D\sigma\cdot Du|^2dx + C\int_X f^2\sigma^{p-2}|Du|^{p-2}dx\\ &\le \delta\int_X \sigma^{p-2}|Du|^{p-2}|D\sigma\cdot Du|^2\,dx + C\left(\varepsilon\int_X \sigma^p|Du|^q\,dx + C\int_X |f|^p\,dx\right).\end{aligned}$$

Since $\sigma \ge 1$ only if $|Du| \ge 1$, we have

$$\int_X \sigma^q|Du|^p\,dx \le \int_X \sigma^q|Du|^{p+2}\,dx + C,$$

and therefore

$$\begin{aligned}\int_X |f|\sigma^{p-2}|Du|^{p-2}|Du\cdot D\sigma|\,dx &\le \delta\int_X \sigma^{p-2}|Du|^{p-2}|D\sigma\cdot Du|^2\,dx\\ &\quad + C\left(\varepsilon\int_X \sigma^p|Du|^{p+2}\,dx + \int_X |f|^p\,dx + 1\right).\end{aligned} \quad (17.22)$$

Taking $\delta > 0$ small, we then derive from (17.19), (17.20), (17.21) and (17.22) our second main estimate

$$\begin{aligned}\int_X \sigma^{p-2}|Du|^p|D\sigma\cdot Du|^2\,dx + \frac{1}{k}\int_X \sigma^{p-2}|Du|^{p-2}|D\sigma\cdot Du|^2\,dx&\\ \le \varepsilon\int_X \sigma^p|Du|^{p+2}\,dx + C\int_X |f|^p + C.&\end{aligned} \quad (17.23)$$

Step IV. *Conclusion.*
Putting together inequalities (17.14) and (17.23) and fixing $\varepsilon > 0$ small, we finally discover

$$\int_X \sigma^p|Du|^{p+2}\,dx \le C + C\int_X |f|^p\,dx.$$

As $\sigma \ge 1$ only if $|Du| \ge 1$, estimate (17.8) follows. □

One can now understand how we hope to prove the L^p estimate that we have in mind: the idea is to apply Proposition 17.3 to find a function $a \in L^p$ and a $1-$Lipschitz function u such that $-\text{div}\,(aDu) = f$; we will infer that $\mu = a\mathscr{L}$ is the transport density associated to f^+ and f^-, and therefore the searched L^p bound for the transport density; however, since Proposition 17.4 concerns only smooth functions, we will need an approximation method: in fact, the smooth functions are also in L^∞, then we already know that the associated transport densities are in L^∞, so the result for the smooth functions would not be anything new.

Theorem 17.5. *Take $2 \leq p < +\infty$ and $f \in L^p(X)$: then the (unique) transport density μ associated to f^+ and f^- belongs to $L^p(X)$. Furthermore, there is a constant C, depending only upon p and X, such that*

$$\|\mu\|_{L^p} \leq C\left(\|f\|_{L^p} + 1\right). \tag{17.24}$$

Proof. **Step I.** *The case when f is smooth.*
Let us first assume f is smooth; we can then apply Proposition 17.3 with $f_k = f$ for any k: the uniform bound (17.4) holds by Proposition 17.4, while the hypothesis (17.5) can always be achieved up to a subsequence when (17.4) holds; in fact, by the $\Gamma-$convergence proved with Proposition 17.2, it is always true, up to a subsequence, that the sequence $\{u_k\}$ uniformly converges to some u which minimizes problem $F(v)$ defined in (17.3). We have then a $1-$Lipschitz function u and a positive function $a \in L^p$ such that (17.6) holds. In particular, if we define $\mu = a\mathscr{L}$, by Proposiyion 17.4 we find a constant C depending only on p and X such that (17.24) holds. We must then only show that μ is the transport density (unique by Theorem 16.2) associated to f^+ and f^-. To do that, we recall that μ and u are a measure and a $1-$Lipschitz function on X with the property that

$$\begin{cases} -\text{div}\,(\mu Du) = f = f^+ - f^- & \text{on } X \\ |Du| = 1 & \mu - \text{a.e.}, \end{cases}$$

and it is well know that, given any such pair, the measure μ happens to be a transport density (recall also the discussion we made in Section 14.3). We can also immediately show this fact recalling the results we obtained in the second Part of the thesis: by Remark 12.5 we know that the measures μ solving the above equation together with some $1-$Lipschitz function u are all and only the optimal shapes, and by Theorem 13.8 the optimal shapes are all and only the (optimal) transport densities (we proved these facts in a much greater generality!). It follows that the measure $\mu = a\mathscr{L}$ is a transport density, thus the unique one by Theorem 16.2. Thus the claim and (17.24) are proved under the assumption that f is smooth.
Step II. *The general case.*
In the general case, let us first define

$$f_j := f * \rho_{1/j},$$

the convolution of f with a standard mollifier. Moreover, for each integer j we define $\sigma_{k,j}$ and $u_{k,j}$ the solutions of (17.1) and (17.2) associated to f_j.

According to (17.8), we can derive the estimate

$$\int_X \sigma_{k,j}^p |Du_{k,j}|^p \, dx \le C \left(\int_X |f_j|^p \, dx + 1 \right) \le C \left(\int_X |f|^p \, dx + 1 \right)$$

for all k greater than or equal to some constant $K = k(j)$ depending only on the L^∞ norm of f_j. Now it suffices to define

$$\sigma_j := \sigma_{k(j),j}, \qquad u_j := u_{k(j),j}, \qquad \mathbf{G}_j := \sigma_j Du_j$$

and to apply Proposition 17.3 exactly as in the first step, again noticing that the measure $\mu = a\mathscr{L}$ is the transport density associated to f^+ and f^-. $\square$

Chapter 18

Sharp L^p estimates via interpolation

In this chapter we present the result of [32], namely the fact that the L^p estimate about the transport density is true also with any $1 < p < 2$, the only case left open after Theorem 17.5. Our claim is the following

Theorem 18.1. *Take any $1 \leq p \leq +\infty$; then there is a constant C_p, depending only on the ambient space X, with the property that whenever $f = f^+ - f^- \in L^p$ then also the (unique) transport density μ is in L^p and the following estimate holds:*

$$\|\mu\|_{L^p} \leq C_p \|f\|_{L^p}. \tag{18.1}$$

To prove this Theorem we will make use of the following known interpolation result, whose proof can be found for instance in [49]:

Theorem 18.2 (Marcinkievicz). *If $T : L^1 \to L^1$ is a linear mapping such that, for two suitable constants M_r and M_q with $1 \leq r < q \leq +\infty$,*

$$\|T(g)\|_{L^r} \leq M_r \|g\|_{L^r} \qquad \text{and} \qquad \|T(g)\|_{L^q} \leq M_q \|g\|_{L^q}, \tag{18.2}$$

then it is also true that for any $s \in (r, q)$

$$\|T(g)\|_{L^s} \leq C\, M_r^{\frac{r(q-s)}{s(q-r)}}\, M_q^{\frac{q(s-r)}{s(q-r)}}\, \|g\|_{L^s}, \tag{18.3}$$

where C is a geometric constant depending only on X.

We will make use of the previous interpolation result with $r = 1$ and $q = +\infty$, thanks to the

Remark 18.3. *We already know that the claim of Theorem 18.1 is true for $p = 1$ and $p = +\infty$: Theorem 16.4 and Theorem 16.9 show this fact with $C_1 = R$ and $C_\infty = 4R\,\omega_{N-1} 12^{N-1}$.*

Note carefully that the Marcinkievicz interpolation result and the above remark do not give in a trivial way the general L^p estimate for all $1 < p < +\infty$: in fact, the map associating to any function f with $\int f = 0$ the corresponding transport density σ is absolutely not linear nor sublinear (this would suffice since the interpolation result is true also in the case of sublinear maps).

To prove our result, let us then fix a constant $1 < p < +\infty$ and a function $f \in L^p$ with $\int f = 0$; we fix also an optimal transport map t from f^+ and f^-, which exists by Theorem 6.17 –recall that in this third Part we deal only with the Euclidean cost function and with a bounded ambient space, so the finiteness of the first order moments is always true. By Remark 6.18 we can also assume that t is invertible and that t^{-1} is an optimal transport map from f^- to f^+.

Given any function $g \in L^1$, there are of course two uniquely determined measurable functions λ and ν from X to itself supported respectively in $\operatorname{spt}(f^+)$ and $X \setminus \operatorname{spt}(f^+)$ such that

$$g = \lambda f^+ + \nu; \tag{18.4}$$

we can then define the operator $T : L^1 \longrightarrow \mathcal{M}(X)$ to which we will apply later the Marcinkievicz theorem: given any $g \in L^1$ and following the notations of (18.4), we define

$$T(g) := \int_X \lambda(x) f^+(x) \mathscr{H}^1_{xt(x)}\, dx, \tag{18.5}$$

that means that for any $\varphi \in C_0(X)$ we set

$$\langle T(g), \varphi \rangle = \int_{x \in X} \left(\int_{z \in X} \varphi(z)\, d\mathscr{H}^1_{xt(x)}(z) \right) \lambda(x)\, f^+(x)\, dx.$$

Notice that $T(g)$ is a priori only a measure, not a L^1 function, and that the definition of T depends on the function f we fixed.

We define now μ_1, μ_2 and f_1, f_2 as follows:

$$\begin{array}{ll} \mu_1 := \int_X \lambda^+(x) f^+(x) \mathscr{H}^1_{xt(x)}\, dx & \qquad f_1 := \lambda^+ f^+ - (\lambda^+ \circ t^{-1}) f^- \\ \mu_2 := \int_X \lambda^-(x) f^+(x) \mathscr{H}^1_{xt(x)}\, dx & \qquad f_2 := \lambda^- f^+ - (\lambda^- \circ t^{-1}) f^-; \end{array} \tag{18.6}$$

note that also these definitions depend on f and g, and that $T(g) = \mu_1 + \mu_2$. First we prove the

Lemma 18.4. *The function $t : X \longrightarrow X$ is defined f_i^+–a.e. and it is an optimal transport map from f_i^+ to f_i^- for $i = 1, 2$; moreover, each μ_i is the transport density associated to f_i.*

Proof. We recall that a transport map is optimal if and only if its graph is $c-$cyclically monotone, as it follows from Theorem 3.22. Then, since t is an optimal transport map from f^+ to f^-, the graph of t is $c-$cyclically monotone: it follows that given any pair of functions h^+ and h^- with the same total mass and such that $h^+ \ll f^+$ (so that t is defined h^+-a.e.), t is an optimal transport map from h^+ to h^- if and only if it is simply a transport map from h^+ to h^-. Then to prove the first part of the assert it is enough to check that $t_{\#}f_i^+ = f_i^-$ for $i = 1, 2$, which is a straightforward consequence of the fact that $t_{\#}f^+ = f^-$ and of the properties of the push-forward. Finally, the fact that each μ_i is the transport density associated to f_i follows comparing (18.6) with (14.1). □

We can then prove the following

Lemma 18.5. $T : L^1 \longrightarrow L^1$ *is a linear operator.*

Proof. The fact that T is linear follows immediately from the definition (18.5), since $g \to \lambda$ is linear according to (18.4); moreover $T(g)$ is a L^1 function (recall that a priori we knew it only to be a measure) since it is the sum of the two transport densities μ_i thanks to the preceding Lemma, and thanks to Theorem 16.4 each of these densities is in L^1 since so is each f_i –recall (18.6) and the fact that $g \in L^1$. □

Since we want to apply Theorem 18.2 to T, we check now the validity of (18.2).

Lemma 18.6. *The inequalities (18.2) hold for T with $r = 1$ and $q = +\infty$; in particular, $M_1 = 2\,C_1$, $M_\infty = 2\,C_\infty$, where the C_i's are the constants as in Remark 18.3.*

Proof. In view of Lemma 18.4 and Theorem 16.4, μ_1 is the transport density relative to f_1 and then $\|\mu_1\|_{L^1} \leq C_1\|f_1\|_{L^1}$; but since t maps f_1^+ to f_1^-, then $\|f_1^+\|_{L^1} = \|f_1^-\|_{L^1}$ and we infer

$$\|\mu_1\|_{L^1} \leq 2\,C_1\|f_1^+\|_{L^1}.$$

In the same way we deduce also $\|\mu_2\|_{L^1} \leq 2\,C_1\|f_2^-\|_{L^1}$. We use now the fact that the supports of the functions f_1^+ and f_2^+ are essentially disjoint, that is clear from (18.6): therefore, also recalling (18.4), we have

$$\begin{aligned}
\|T(g)\|_{L^1} &= \|\mu_1 + \mu_2\|_{L^1} \leq \|\mu_1\|_{L^1} + \|\mu_2\|_{L^1} \leq 2\,C_1\left(\|f_1^+\|_{L^1} + \|f_2^+\|_{L^1}\right) \\
&= 2\,C_1\|f_1^+ + f_2^+\|_{L^1} = 2\,C_1\|\lambda^+ f^+ + \lambda^- f^+\|_{L^1} \\
&= 2\,C_1\|\lambda f^+\|_{L^1} \leq 2\,C_1\|g\|_{L^1},
\end{aligned}$$

which gives the first estimate.

On the other hand, to show the L^∞ inequality we just note that thanks to (18.4) and (18.6) it is $\|f_i\|_{L^\infty} \leq \|g\|_{L^\infty}$ for $i = 1, 2$. Since $T(g) = \mu_1 + \mu_2$, from Lemma 18.4 and Remark 18.3 we infer

$$\|T(g)\|_{L^\infty} \leq 2\, C_\infty \|g\|_{L^\infty},$$

and then also the L^∞ inequality follows. □

It is now trivial to conclude with the strong L^p estimate:

Proof. (of Theorem 18.1): Thanks to Lemma 18.5 and Lemma 18.6, we can apply Theorem 18.2 to find that (18.1) holds, just noticing that $T(f)$ obviously equals the transport density μ. Recall now that we fixed at the beginning a function $f \in L^p$; but the constant C_p, obtained through (18.3), does not depend on the choice of f, then the estimate (18.1) is true, with the same constant, for any function $f \in L^p$. □

Appendix A

Some results about measures on Polish spaces

Here we recall some general results we have been using throughout the thesis, but whose proof has been omitted.

A.1 Polish spaces and main Theorems

We give here some results about Polish spaces; first of all, we need the definition:

Definition A.1. *A topological space X is said a Polish space if it is metrizable with a distance making it separable and complete.*

Recall that a space is said metrizable if there is some distance on it inducing the same topology; note also that the completeness, but not the separability, depends on the distance. In this paper, we always consider each Polish space endowed with the Borel $\sigma-$algebra, denoted by $\mathscr{B}(X)$.

Given a Polish space X, we consider the space $\mathcal{M}^+(X)$ of the finite positive measures on it, that is defined, as usual, as the set of the countably additive functions from $\mathscr{B}(X)$ to $\mathbb{R}^+$; it is well known that $\mathcal{M}^+(X)$ is a (strongly) closed subset of $\big(C_b(X)\big)'$, the dual space of the space $C_b(X)$ of the continuous and bounded real functions on X, endowed with the norm

$$\|u\|_{C_b(X)} := \sup\big\{|u(x)|,\ x \in X\big\};$$

we denote by $\langle\cdot,\cdot\rangle$ the duality between $C_b(X)'$ and $C_b(X)$, and also –in particular– the duality between $\mathcal{M}^+(X)$ and $C_b(X)$. Even though $\mathcal{M}^+(X)$ is not a linear space because we deal only with positive measures, we consider on it the norm induced by the inclusion in $C_b(X)'$, that is

$$\|\mu\|_{\mathcal{M}^+(X)} := \sup\left\{\langle\mu,u\rangle,\ \|u\|_{C_b(X)} = 1\right\} = \mu(X).$$

The subspace of $C_b(X)'$ made by the increasing functionals, i.e. all those $\mu \in C_b(X)'$ such that $\langle\mu,u\rangle \geq 0$ whenever $u \geq 0$, is denoted by $C_b(X)'_+$: of course, there is the inclusion $\mathcal{M}^+(X) \subseteq C_b(X)'_+$.

We can also consider on $\mathcal{M}^+(X)$ the weak* convergence induced by $C_b(X)$: we say that a sequence $\mu_n \in \mathcal{M}^+(X)$ weak* converge to μ, and we write $\mu_n \stackrel{*}{\rightharpoonup} \mu$, if

$$\langle\mu_n,u\rangle \longrightarrow \langle\mu,u\rangle \qquad \text{for each } u \in C_b(X).$$

We give now a couple of other definitions:

Definition A.2. *Given any positive measure $\eta \in \mathcal{M}^+(X)$, we define the set $\mathscr{B}_\eta(X)$ of the* measurable sets w.r.t. η *as the smallest $\sigma-$algebra containing both the set $\mathscr{B}(X)$ of all Borel sets and the set of all the negligible (w.r.t. η) subsets of X. Moreover, we say that a set $B \subseteq X$ is* universally measurable *if $B \in \mathscr{B}_\eta(X)$ for any positive measure η.*

Definition A.3. *Given any positive measure $\mu \in \mathcal{M}^+(X)$ we define* support of μ, $\operatorname{spt}\mu$, *the smallest closed set $K \subseteq X$ such that*

$$\mu(X \setminus K) = 0.$$

More in general, one says that the measure μ is concentrated *in Γ if*

$$\mu(X \setminus \Gamma) = 0.$$

For any measure μ, $\operatorname{spt}\mu$ is a well-defined closed set of full measure; in particular, it is the smallest closed set of full measure; on the other hand, there are in general many sets in which μ is concentrated. However, it is often very useful to select some set of full measure for μ with particular properties, which may also differ from the support. In particular, a measure can be concentrated in a set much smaller than its support: if for example $X = \mathbb{R}$ and $\mu = \sum_i 2^{-i}\,\delta_{q_i}$ where $\mathbb{Q} = \{q_i\}$ is the set of the rationals, then μ is concentrated in the countable set $\mathbb{Q}$ but its support is the whole $\mathbb{R}$.

A first result on Polish spaces is given by the following

Theorem A.4 (Ulam). *Given a Polish space X and $\mu \in \mathcal{M}^+(X)$, there is a $\sigma-$compact set μ in which μ is concentrated.*

This result is of primary importance throughout the paper, and can be found for example in the book of Dellacherie and Meyer [27]; one can refer to the same book for the proof of another very important Theorem, which is also often used in this thesis:

Theorem A.5 (Prokhorov). *A bounded set $\{\mu_i,\, i \in I\}$ in $\mathcal{M}^+(X)$ is weak* sequentially relatively compact if and only if for any ε there is a compact set K_ε such that*

$$\mu_i(X \setminus K_\varepsilon) \leq \varepsilon \qquad \text{for any } i \in I.$$

As we mentioned before, any measure is an element of $C_b(X)'_+$; however, it is fundamental to know when the opposite is also true: in fact, through the thesis we define very often a measure μ by specifying the value of $\langle \mu, u \rangle$ for any $u \in C_b(X)$. It is then sufficient to check that $\langle \mu, \cdot \rangle$ is linear, continuous and increasing to derive that μ belongs to $C_b(X)'_+$, and this usually follows trivially from the definition; but we need also to establish that μ is a measure, that is not straightforward since the inclusion $\mathcal{M}^+(X) \subseteq C_b(X)'_+$ is in general strict for Polish spaces. We collect now some results which give necessary and sufficient conditions for an element of $C_b(X)'_+$ to be in $\mathcal{M}^+(X)$. It is easy to notice that these results cover all the situations we deal with throughout this thesis: therefore, we are allowed to define all the measures we use by means of their duality with $C_b(X)$ and we know, without need of checking it, that they are effectively measures.

To this aim, the most important result one can use is Daniell's Theorem:

Theorem A.6 (Daniell). *Let $\mu \in C_b(X)'_+$: then $\mu \in \mathcal{M}^+(X)$ if and only if for every sequence $\{h_n\} \subseteq C_b(X)$ such that $h_n \searrow 0$ (i.e. h_n is pointwise decreasing to 0) one has $\langle \mu, h_n \rangle \longrightarrow 0$.*

Note that, in the hypotheses above, $\langle \mu, h_n \rangle$ is a real positive decreasing sequence, then $\langle \mu, h_n \rangle \longrightarrow l$ for some $l \geq 0$; the Theorem claims that μ is a measure if and only if this limit l is never strictly positive. Once again, one can find the proof of a more general assert on the book [27]. We remark now some easy consequences of Daniell's Theorem.

Proposition A.7. *Let $\mu \in C_b(X)'_+$: then $\mu \in \mathcal{M}^+(X)$ if and only if for each $\varepsilon > 0$ there is a compact set K such that $\langle \mu, u \rangle \leq \varepsilon \|u\|_{C_b(X)}$ whenever $u \in C_b(X)$, $u \geq 0$ on X, $u = 0$ on K.*

Proof. If $\mu \in \mathcal{M}^+(X)$ then the stated property immediately follows from Ulam's Theorem. On the other hand, assume that the property holds and take a sequence $h_n \searrow 0$ as in Daniell's Theorem: to show $\mu \in \mathcal{M}^+(X)$ it is sufficient to check that $\langle \mu, h_n \rangle \to 0$; fix then $\varepsilon > 0$ and consider the corresponding compact set K.

Since K is compact, there is an integer m such that $h_m \leq \varepsilon$ on K; let then U be an open neighborhood of K such that $h_m < 2\varepsilon$ on U. Notice that this is possible since h_m is continuous, and that the open set U depends on ε, on K and on m. Take now a partition of the unity $\{\varphi_1,\, \varphi_2\}$ associated to the open sets U and $X \setminus K$, i.e. $\varphi_i : X \to [0, 1]$ is a continuous function for $i = 1, 2$, $\varphi_1 + \varphi_2 \equiv 1$, and φ_1 (resp. φ_2) is positive only inside the open set U (resp. $X \setminus K$). We recall that this is always possible, since X is metrizable: for example, if we denote by $d_1(x)$ and $d_2(x)$ the distance of x from $X \setminus U$ and K respectively, it suffices to note that d_1 and d_2 are continuous, that $d_1 + d_2$ is everywhere strictly positive and finally to define

$$\varphi_1(x) := \frac{d_1(x)}{d_1(x) + d_2(x)}, \qquad \varphi_2(x) := \frac{d_2(x)}{d_1(x) + d_2(x)}.$$

Now we write $h_m = g_1 + g_2$, where $g_i = \varphi_i h_m$: then we note that $g_1 \leq h_m$ on U and $g_1 = 0$ out of U: consequently $\|g_1\|_{C_b(X)} \leq 2\varepsilon$. On the other hand, $g_2 = 0$ on K and, of course, $\|g_2\|_{C_b(X)} \leq \|h_m\| \leq \|h_1\|$, since the sequence $\{h_n\}$ is decreasing. Then, thanks to the hypothesis, we can evaluate

$$\langle \mu, h_m \rangle = \langle \mu, g_1 \rangle + \langle \mu, g_2 \rangle \leq 2\varepsilon \, \|\mu\|_{C_b(X)'_+} + \varepsilon \|h_1\|_{C_b(X)'_+} \leq \varepsilon \big(2\|\mu\| + \|h_1\| \big).$$

Since the sequence $\{h_n\}$ is decreasing and by the generality of ε we deduce $\langle \mu, h_n \rangle \to 0$ and then the thesis follows by Daniell's Theorem. $\square$

We can derive a very useful corollary:

Corollary A.8. *Assume that $\gamma \in C_b(X \times Y)'_+$ and that $\pi_1\gamma \in \mathcal{M}^+(X)$ and $\pi_2\gamma \in \mathcal{M}^+(Y)$. Then $\gamma \in \mathcal{M}^+(X \times Y)$.*

Proof. By Ulam's Theorem (or by Proposition A.7), given $\varepsilon > 0$ there are two compact sets $K_1 \subseteq X$ and $K_2 \subseteq Y$ such that $\langle \pi_1\gamma, v \rangle \leq \varepsilon \|v\|$ (resp. $\langle \pi_2\gamma, w \rangle \leq \varepsilon \|w\|$) whenever $v \in C_b(X)$, $v \geq 0$ on X, $v = 0$ on K_1 (resp. $w \in C_b(Y)$, $w \geq 0$ on Y, $w = 0$ on K_2). We consider then the compact set $K = K_1 \times K_2$ in $X \times Y$: given any $u \in C_b(X \times Y)$ such that $u \geq 0$ on X and $u = 0$ on K, and given $\delta > 0$, we can find an open neighborhood $U_1 \times U_2$ of K such that $0 \leq u \leq \delta$ in $U_1 \times U_2$. Moreover, arguing as in Proposition A.7, we can find two functions $v : X \to [0, \|u\|]$ and $w : Y \to [0, \|u\|]$ such that $v = 0$ on K_1, $w = 0$ on K_2, $v = \|u\|$ out of U_1 and $w = \|u\|$ out of U_2. Since of course $v(x) + w(y) \geq u(x, y)$ whenever $(x, y) \notin U_1 \times U_2$ and $\langle \gamma, v(x) + w(y) \rangle = \langle \pi_1\gamma, v \rangle + \langle \pi_2\gamma, w \rangle$, arguing as in Proposition A.7 one derives

$$\langle \gamma, u \rangle \leq 2\,\varepsilon \|u\| + \delta \|\gamma\|;$$

by the generality of δ, the thesis follows from Proposition A.7. $\square$

The corollary above is really important in mass transportation problem: in fact, it clearly implies in particular that any continuous and linear functional on $X \times Y$ whose marginals are f^+ and f^- is in fact a measure, then it is a transport plan.

Let us briefly discuss now the problem to understand whether or not the inclusion $\mathcal{M}^+(X) \subseteq C_b(X)'_+$ is in fact an equivalence: first of all, we have an immediate positive result:

Lemma A.9. *If X is compact, then $\mathcal{M}^+(X) = C_b(X)'_+$.*

Proof. This trivially follows from Daniell's Theorem, since a sequence pointwise decreasing to 0 in a compact space is easily seen to be converging uniformly to 0. □

However, the result above cannot be extended, as the following example shows:

Example A.10. Take $X = \mathbb{R}$, define $u_0 \equiv 1$ and for any integer $n \geq 1$ let $u_n : \mathbb{R} \to [0,1]$ be a continuous function such that $u_n \equiv 1$ on $[-n, n]$, $u_n \equiv 0$ out of $(-(n+1), n+1)$. Define D as the span of $\{u_n,\ n \geq 0\}$, that is a subspace of $C_b(X)$. Let moreover $\tilde{\mu}$ be the linear functional on D given by $\langle \tilde{\mu}, u_0 \rangle = 1$ and $\langle \tilde{\mu}, u_n \rangle = 0$ for any $n \geq 1$, which is easily checked to be continuous and of unit norm. By Hahn-Banach Theorem, there is an element μ of $C_b(X)'_+$ extending $\tilde{\mu}$. Since the functions $h_n = u_0 - u_n$ are as in the claim of Daniell's Theorem but $\langle \mu, h_n \rangle = 1$ for any n, μ is not a measure, then $\mu \in C_b(X)'_+ \setminus \mathcal{M}^+(X)$.

The example above can be clearly extended to cover all the non-compact Polish spaces, therefore in Lemma A.9 also the opposite implication is valid.

Remark A.11. *Recall that Riesz Theorem claims that $\mathcal{M}^+(\mathbb{R})$ is exactly the set of the increasing linear functionals on $C_c(\mathbb{R})$. In general, using $C_c(X)$ in place of $C_b(X)$ to define a duality on $\mathcal{M}^+(X)$ has the advantage that Riesz Theorem –i.e. $\mathcal{M}^+(X) = C_c(X)'_+$– is true in a wide generality: for example, making use of Daniell's Theorem, one can quite easily show that this is true whenever X is a countable union of open sets $\{U_n\}$ with compact closures $\{K_n\}$ such that $K_n \subseteq U_{n+1}$ for any integer n; in particular, Riesz Theorem is true whenever X is a locally compact Polish space. Notice in particular that, since $C_c(X) \subseteq C_b(X)$, whenever Riesz Theorem holds for X one has that for any $\mu \in C_b(X)'_+$ there is a $\tilde{\mu} \in \mathcal{M}^+(X)$ that equals μ on each bounded subset of X; more precisely, $\langle \mu, u \rangle = \langle \tilde{\mu}, u \rangle$ for any $u \in C_c(X)$.*

However, in this thesis we prefer to use the duality with $C_b(X)$, since otherwise there would be many disadvantages due to the fact that we do not deal with compactly supported measures.

We conclude this section by stating two other important results about Polish spaces, which can be found for example in the book of Castaing and Valadier [25] (in a more general version).

Theorem A.12 (Measurable selection). *Let X and Y be two Polish spaces, and let η be a measure on X; if $\Gamma \subseteq X \times Y$ is $\mathscr{B}_\eta(X) \otimes \mathscr{B}(Y)-$measurable and the projection of Γ on X is a set of full $\eta-$measure, then there exists a measurable selection of Γ, i.e. a $\eta-$measurable function $\sigma : X \to Y$ such that for $\eta-$a.e. $x \in X$ one has $(x, \sigma(x)) \in \Gamma$.*

Theorem A.13 (Projection). *Let X and Y be two Polish spaces, and let $\Gamma \subseteq X \times Y$ be a Borel set; then the projection of Γ on X is universally measurable.*

A.2 Push-forward of measures

Given a measurable function $\varphi : X \to Y$, it is defined a linear and mass-preserving function $\varphi_\# : \mathcal{M}^+(X) \to \mathcal{M}^+(Y)$ (by mass-preserving we mean that $\|\mu\|_{\mathcal{M}^+(X)} = \|\varphi_\#\mu\|_{\mathcal{M}^+(Y)}$), according to the following definition:

$$\varphi_\#\mu(B) := \mu\big(\varphi^{-1}(B)\big) \qquad \forall\, A \in \mathscr{B}(Y).$$

It is easily noticed that $\varphi_\#\mu \in \mathcal{M}^+(Y)$, and in particular for any $\alpha \in C_b(Y)$

$$\langle \varphi_\#\mu, \alpha \rangle = \int_Y \alpha(y)\, d\varphi_\#\mu(y) = \int_X \alpha\big(\varphi(x)\big)\, d\mu(x).$$

An immediate but very useful consequence of the definition is that for any $\varphi : X \to Y$ and $\psi : Y \to Z$ it is

$$(\psi \circ \varphi)_\#\mu = \psi_\#(\varphi_\#\mu). \tag{A.1}$$

A.3 Tensor product of measures

Given two measures $\mu \in \mathcal{M}^+(X)$ and $\nu \in \mathcal{M}^+(Y)$, we define the *tensor product* of μ and ν, $\mu \otimes \nu \in \mathcal{M}^+(X \times Y)$, as the unique measure on $X \times Y$ satisfying

$$\mu \otimes \nu(A \times B) := \mu(A) \cdot \nu(B) \qquad \forall\, A \in \mathscr{B}(X),\ B \in \mathscr{B}(Y).$$

The fact that there is a measure satisfying the above property follows from Fubini-Tonelli Theorem, while the unicity is clear because the smallest $\sigma-$algebra containing all the sets $A \times B$ with $A \in \mathscr{B}(X)$, $B \in \mathscr{B}(Y)$ is the whole $\mathscr{B}(X \times Y)$. In particular, the marginals of $\mu \otimes \nu$ are

$$\pi_1(\mu \otimes \nu) = \nu(Y)\mu, \qquad\qquad \pi_2(\nu \otimes \mu) = \mu(X)\nu,$$

where we write, with a slight abuse, π_1 and π_2 in place of $\pi_{1\#}$ and $\pi_{2\#}$.

An interesting particular case is when μ and ν are probability measure; so, the projections of $\mu \otimes \nu$ are precisely μ and ν: this is particularly useful in the study of the mass transportation, since if $\|f^+\| = \|f^-\| = 1$ then $f^+ \otimes f^-$ is always a transport plan between f^+ and f^-, and it is the trivial transport plan which moves each part of the mass f^+ of X in Y proportionally to f^-. However, it is often useful to have a fast symbol to denote this trivial transport plan also when the measures f^+ and f^- are not probability measures. Therefore, whenever the measures μ and ν have the property that $\|\mu\| = \|\nu\|$, we define the *rescaled tensor product* of μ and ν as

$$\mu \boxtimes \nu(A \times B) = \frac{\mu(A) \cdot \nu(B)}{\|\mu\|} = \frac{\mu(A) \cdot \nu(B)}{\|\nu\|} \qquad \forall\, A \in \mathscr{B}\,(X),\ \forall\, B \in \mathscr{B}\,(Y). \tag{A.2}$$

Note that

$$\mu \boxtimes \nu = \frac{\mu \otimes \nu}{\|\mu\|} = \frac{\mu \otimes \nu}{\|\nu\|}.$$

A.4 Measure valued maps

Here we briefly introduce a couple of tools concerning the measure-valued maps. First of all, we give the

Definition A.14. *A map $\tau : X \to \mathcal{M}(Y)$ is said a Borel measure-valued map (resp. $\mu-$measurable measure-valued map, being $\mu \in \mathcal{M}(X)$) if for any Borel set $B \subseteq Y$ the function $x \to \tau(x)(B)$ is Borel (resp. $\mu-$measurable). Equivalently, τ is said to be Borel (resp. $\mu-$measurable) if for any bounded Borel $\varphi : X \times Y \to \mathbb{R}$ the function*

$$x \to \int_Y \varphi(x, y)\, d\tau(x)(y)$$

is Borel (resp. $\mu-$measurable).

Thanks to the notion above, we can generalize the idea of tensor product between two measures: consider a measure $\nu \in \mathcal{M}(Y)$ and a $\nu-$measurable measure valued map $y \to \gamma_y$ with $\gamma_y \in \mathcal{M}(X)$. We define the *tensor product* between $\{\gamma_y\}$ and ν to be the measure $\gamma \in \mathcal{M}(X)$ defined by the formula

$$\langle \gamma, \varphi \rangle := \int_Y \langle \gamma_y, \varphi \rangle \, d\nu(y)$$

for any $\varphi \in C_b(X)$. We always denote this measure by

$$\gamma = \gamma_x \otimes \nu.$$

A.5 Disintegration Theorem

This theorem allows to decompose a measure γ on the space X with respect to a Borel function $\alpha : X \to Y$, where this "decomposition" is intended as a tensor product between suitable probability measures and the push-forward of γ. The proof of this Theorem can be found, for instance, in [3] or [27].

Theorem A.15 (Disintegration). *Assume $\alpha : X \to Y$ is a given Borel map and $\gamma \in \mathcal{M}^+(X)$ is a given measure, and define $\mu \in \mathcal{M}^+(Y)$ as $\mu = \alpha_\# \gamma$. Then there exists a $\mu-$measurable measure valued function $y \to \gamma_y$ such that γ_y is a probability measure on X for any y and*

(i) $\gamma = \gamma_y \otimes \mu$;

(ii) γ_y *is concentrated in* $\{x : \alpha(x) = y\}$ *for* $\mu-$*a.e.* y.

Moreover, the measures γ_y are uniquely determined by (i) and (ii) for $\mu-$a.e. y.

We state and prove now an useful consequence of the Theorem above.

Lemma A.16. *The operations of disintegration and of composition commute. In formula, if*

$$\gamma = \gamma_y \otimes \alpha_\# \gamma$$

is the decomposition of $\gamma \in \mathcal{M}^+(X)$ with respect to some $\alpha : X \to Y$ and a function $\beta : X \to Z$ is given, then

$$\beta_\# \gamma = \beta_\# \gamma_y \otimes \alpha_\# \gamma. \tag{A.3}$$

In particular, if $\alpha = \delta \circ \beta$ for some $\delta : Z \to Y$ and

$$\beta_\# \gamma = \mu_y \otimes \delta_\# \bigl(\beta_\# \gamma\bigr) \tag{A.4}$$

is the disintegration of $\beta_\# \gamma$ with respect of δ, for a.e. $y \in Y$ it is

$$\mu_y = \beta_\# \gamma_y. \tag{A.5}$$

Proof. The first part is really easy: for any $\varphi \in C_c(Z)$, recalling the properties of the push-forward it is

$$\begin{aligned}
\langle \beta_\# \gamma, \varphi \rangle &= \int_X \varphi(\beta(x))\, d\gamma(x) = \int_Y \Bigl(\int_X \varphi(\beta(x))\, d\gamma_y(x) \Bigr)\, d\nu(y) \\
&= \int_Y \Bigl(\int_Z \varphi(z)\, d\beta_\# \gamma_y(z) \Bigr)\, d\nu(y) = \langle \beta_\# \gamma_y \otimes \nu, \varphi \rangle,
\end{aligned}$$

thus the claim follows.

Concerning the second part, which is usually the application we have in mind, by the properties of disintegration (A.3) becomes

$$\beta_\#\gamma = \beta_\#\gamma_y \otimes \big(\delta_\#\beta_\#\gamma\big);$$

recall now the disintegration (A.4): according to Theorem A.15, for a.e. $y \in Y$ μ_y is concentrated in the set $\{z \in Z : \delta(z) = y\}$; analogously, γ_y is concentrated in the set $\{x \in X : \alpha(x) = y\} = \{x \in X : \delta(\beta(x)) = y\}$. But then $\beta_\#\gamma_y$ is concentrated in

$$\beta\Big(\big\{x \in X : \delta(\beta(x)) = y\big\}\Big) \subseteq \big\{z \in Z : \delta(z) = y\big\};$$

moreover, $\alpha_\#\gamma = \delta_\#(\beta_\#\gamma)$ by (A.1); then, by the uniqueness part of Theorem A.15, we infer the validity of (A.5) and then also the second claim is achieved. □

A.6 The relationship between measurable and Borel maps

We state here two well known fundamental results about the Borel property of maps; one can find the proof of the first one for instance in [70], the proof of the second one in [67].

Theorem A.17. *Given a measure $\lambda \in \mathcal{M}^+(X)$ and a Borel set $A \subseteq X$, there is a $\sigma-$compact subset of A of measure $\lambda(A)$.*

Theorem A.18. *Let $g : X \to Y$ a measurable function, and $\lambda \in \mathcal{M}^+(X)$. Then, there exists $\bar{g} : X \to Y$ such that $\bar{g}$ is Borel and $\{x : \bar{g}(x) \neq g(x)\}$ is $\lambda-$negligible.*

A.7 Borel property when glueing maps

Here we consider an important result which allows to "glue" different maps preserving the Borel property, and that is a fundamental result to build optimal transport maps working on the different maximal transport rays.

Lemma A.19. *Let $\pi : X \to Y$ be a Borel map and $\mu = \mu_y \otimes \pi_\#\mu \in \mathcal{M}^+(X)$ according to Theorem A.15. Let moreover $g : X \to Z$ be a map such that g is μ_y-measurable for $\pi_\#\mu-$a.e. y and that $y \to (\mathrm{Id}, g)_\#\mu_y$ is a $\pi_\#\mu-$measurable measure-valued map from Y to $\mathcal{M}^+(X \times Z)$. Then g is $\mu-$essentially a Borel map, i.e. there exists a Borel function $h : X \to Z$ such that $g = h$ μ_y-a.e. for $\pi_\#\mu-$a.e. $y \in Y$.*

The proof of this statement can be found in [2]; note that we did not ask g to be $\mu-$measurable: if this is true, then the thesis can be rewritten as $g = h$ $\mu-$a.e..

Appendix B

Vectorial measures on Manifolds

Let M be an $N-$dimensional connected Riemannian manifold, possibly with boundary; the space $\mathcal{M}^N(M)$ of the vectorial measures on the manifold M can be defined as the dual space of the continuous and bounded sections of the tangent bundle TM. Given a measure $\nu \in \mathcal{M}^N(M)$, one would clearly find it useful (as in the Euclidean case) to write $\nu = \omega|\nu|$ for some unitary vector field ω and a scalar measure $|\nu| \in \mathcal{M}^+(M)$: by writing $\nu = \omega|\nu|$ we mean that

$$\langle \nu, \varphi \rangle = \langle |\nu|, \varphi \cdot \omega \rangle = \int_M \varphi(x) \cdot \omega(x)\, dx$$

for any continuous section φ of the tangent bundle TM. It turns out that (exactly as in the Euclidean case) this decomposition is possible and the correct definition for $|\nu| \in \mathcal{M}^+(M)$ has to be

$$\langle |\nu|, h \rangle := \sup \{ \langle \nu, \omega h \rangle,\ |\omega| \leq 1 \} \qquad \forall\, h \in C_b(M),\ h \geq 0.$$

We define then L^1_ν as the completion of $C_b(M)$ endowed with the norm $\|\varphi\|_{L^1_\nu} := \langle |\nu|, \varphi \rangle$; finally, one can easily prove the following generalization of Riesz's Theorem, either using it on the charts of an atlas or directly by intrinsic duality arguments:

Theorem B.1. *There exists a $|\nu|-$measurable unit vector field $\overline{\omega}$ such that $\nu = \overline{\omega}|\nu|$.*

Given a smooth and bounded function $u \in C^\infty_b(M)$, it is well defined the gradient of u at any point x, that we denote by Du: this is a (smooth) section of the tangent bundle TM, since the vectors and the covectors are isomorphic thanks to the Riemannian structure on M. Therefore, if $\nu \in \mathcal{M}^N(M)$ is a given vectorial measure, it is well defined the *divergence* of nu, a linear functional on $C^\infty_b(M)$ given by

$$\langle \text{div } \nu, u \rangle := -\langle \nu, Du \rangle \qquad \forall u \in C^\infty_b(M).$$

We could want to approximate a measure $\nu \in \mathcal{M}^N(M)$ with regular measures ν_ε supported in $M \setminus \partial M$; to do this, let us first fix an atlas for M: so we have a locally finite open covering $\{U_i\}$ of M and the corresponding functions $\psi_i : U_i \to \mathbb{R}^N$, where each ψ_i is a diffeomorphism between U_i and the image $\psi_i(U_i)$ and that image is the open unitary hypercube $(0,1)^N$ if $U_i \cap \partial M = \emptyset$, while it is $(0,1)^{N-1} \times [0,1)$ if $U_i \cap \partial M \neq \emptyset$: moreover, in the last case it is $\psi_i^{-1}\big((0,1)^{N-1} \times \{0\}\big) = U_i \cap \partial M$. Let us define now for any $0 < \varepsilon < 1$ the function $\tau_\varepsilon : [0,1]^N \to [0,1]$ as follows: $\tau_\varepsilon(x) = 1$ if $x_N \geq \varepsilon$ (x_N is simply the $N-$th coordinate of x), and $\tau_\varepsilon(x) = 0$ otherwise. Denote moreover by $\{\varphi_i\}$ a partition of the unity corresponding to the covering $\{U_i\}$ and by ρ_ε the standard convolution kernel. Finally, being $\varepsilon \to \delta(\varepsilon, i)$ an increasing function to be determined later with $\delta(\varepsilon, i) \leq \varepsilon/2$ and $\lim_{\varepsilon \to 0} \delta(\varepsilon, i) = 0$, we set

$$\begin{cases} \nu_i := & \varphi_i \, \nu; \\ \nu_{\varepsilon,i} := & {\psi_i^{-1}}_{\#}\big(\tau_\varepsilon(\psi_{i\#}\nu_i) * \rho_{\delta(\varepsilon,i)}\big); \\ \nu_\varepsilon = & \sum_i \nu_{\varepsilon,i}. \end{cases} \tag{B.1}$$

This definition is what we needed, since we have the

Theorem B.2. *$\nu_\varepsilon \in C^\infty(M)$, $\operatorname{spt} \nu_\varepsilon \cap \partial M = \emptyset$ and $\nu_\varepsilon \xrightarrow[\mathcal{M}^N(M)]{*} \nu$. In addition, if $-\operatorname{div} \nu = f$ belongs to $\mathcal{M}(M)$, then $f_\varepsilon := -\operatorname{div}\, \nu_\varepsilon \xrightarrow[\mathcal{M}(M)]{*} f$.*

Proof. The smoothness assertion is obvious, since ν_ε is a locally finite sum of smooth functions supported out of ∂M. To prove the weak* convergence of ν_ε to ν, let ω be a continuous field of vectors of norm less than 1: a standard computation, starting from the definition of the convolution and using Fubini's Theorem, shows that $\langle \nu_{\varepsilon,i}, \omega \rangle \leq \langle |\nu_i|, 1 \rangle$, so the generality of ω ensures that $\langle |\nu_{\varepsilon,i}|, 1 \rangle \leq \langle |\nu_i|, 1 \rangle$. Now, from the definition it follows that $\nu_i = |\nu_i| \, \overline{\omega}$ (i.e. the ν_i are all parallel to ν, since the φ_i are scalar functions), and then

$$\sum_i \langle |\nu_i|, 1 \rangle = \sum_i \langle \nu_i, \overline{\omega} \rangle = \langle \nu, \overline{\omega} \rangle < +\infty.$$

Given $\sigma > 0$, it follows the existence of a finite set I such that

$$\sum_{i \notin I} \langle |\nu_i|, 1 \rangle \leq \sigma. \tag{B.2}$$

Given now a continuous vector field h, since $\langle \nu, h \rangle = \sum_i \langle \nu_i, h \rangle$ we can assume, possibly passing to a greater finite set I, that

$$\left| \langle \nu, h \rangle - \sum_{i \in I} \langle \nu_i, h \rangle \right| \leq \sigma. \tag{B.3}$$

On the other hand, thanks to (B.2) we have

$$\sum_{i\notin I}\langle \nu_{\varepsilon,i}, h\rangle \leq \|h\|_{C_b(M)} \sum_{i\notin I}\langle |\nu_{\varepsilon,i}|, 1\rangle \leq \|h\| \sum_{i\notin I}\langle |\nu_i|, 1\rangle \leq \sigma\|h\|. \tag{B.4}$$

Recalling (B.1) it is clear that, provided $\delta(\varepsilon, i)$ is sufficiently small, one has $\nu_{\varepsilon,i} \xrightharpoonup[\varepsilon\to 0]{*} \nu_i$: since I is finite, if $\varepsilon \ll 1$ we derive

$$\Big|\sum_{i\in I}\langle \nu_{\varepsilon,i} - \nu_i, h\rangle\Big| \leq \sigma. \tag{B.5}$$

Putting together (B.3), (B.4) and (B.5), we infer

$$\Big|\langle \nu, h\rangle - \langle \nu_\varepsilon, h\rangle\Big| \leq \sigma\left(2 + \|h\|_{C_b(M)}\right),$$

and from the generality of h and $\sigma > 0$ it follows $\nu_\varepsilon \xrightharpoonup{*} \nu$.

Concerning the last assert, from the convergence $\nu_\varepsilon \xrightharpoonup{*} \nu$ one immediately derives $\langle f_\varepsilon, u\rangle \to \langle f, u\rangle$ for any smooth function u: to show $f_\varepsilon \xrightharpoonup{*} f$ is then sufficient to check that $\|f_\varepsilon\|$ is bounded. To this aim, define $\tilde\nu_{\varepsilon,i} = {\psi_i^{-1}}_{\#}\big(\tau_\varepsilon(\psi_{i\#}\nu_i)\big)$: it is clear that $\tilde\nu_{\varepsilon,i}$ strongly converges to ν_i, but one can also show that div $\tilde\nu_{\varepsilon,i}$ weak* converges to div ν_i. As a consequence, $\|\tilde f_\varepsilon\|$ is bounded, where $\tilde f_\varepsilon = \sum_i -\mathrm{div}\ \tilde\nu_{\varepsilon,i}$. But one has that

$$-\mathrm{div}\ \left({\psi_i^{-1}}_{\#}\big(\tau_\varepsilon(\psi_{i\#}\nu_i) * \rho_\delta\big)\right)$$

strongly converges to $\tilde f_{\varepsilon,i} = -\mathrm{div}\ \tilde\nu_{\varepsilon,i}$ when $\delta \to 0$: as a consequence, recalling (B.1) it follows that, if $\delta(\varepsilon)$ is sufficiently small, from the boundedness of $\|\tilde f_\varepsilon\|$ one derives that of $\|f_\varepsilon\|$. And, as noticed before, the boundedness of $\|f_\varepsilon\|$ gives our desired convergence. $\square$

Bibliography

[1] L. Ambrosio, Lecture Notes on Optimal Transport Problems, in "Mathematical Aspects of Evolving Interfaces", Lecture Notes in Mathematics, LNM **1812**, Springer (2003), 1–52.

[2] L.Ambrosio & A.Pratelli, Existence and stability results in the L^1 theory of optimal transportation, in "Optimal Transportation and Applications", Lecture Notes in Mathematics, LNM **1813**, Springer (2003), 123–160.

[3] L. Ambrosio, N. Fusco & D. Pallara, Functions of Bounded Variation and Free Discontinuity Problems, Oxford University Press (2000).

[4] L. Ambrosio, B. Kirchheim & A. Pratelli, Existence of optimal transport maps for crystalline norms, to appear on Duke Math. J..

[5] G. Anzellotti & S. Baldo, Asymptotic development by Γ-convergence, Appl. Math. Optim., **27** (1993), 105–123.

[6] G. Anzellotti, S. Baldo & D. Percivale, Dimension reduction in variational problems, asymptotic developments in Γ-convergence and thin structures in elasticity, Asymptotic Anal., **9** (1994), 61–100.

[7] P. Appel, Le probleme geometrique des déblais et remblais, Memor. des Sciences Math., Acad. de Sciences de Paris, Gauthier Villars, **27** (1928), 1–34.

[8] V. Bangert, Minimal measures and minimizing closed normal one-currents, Geom. Funct. Anal., **9** (1999), 413–427.

[9] D. Bao, S.S. Chern & Z. Shen, An introduction to Riemann-Finsler geometry, Springer (2000).

[10] J.-D. Benamou & Y. Brenier, A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem, Numer. Math., **84** (2000), 375–393.

[11] J.D. Benamou & Y. Brenier, Mixed L^2-Wasserstein optimal mapping between prescribed density functions, J. Optim. Theory Appl., **111** no. 2 (2001), 255–271.

[12] G. Bouchitté & G. Buttazzo, Characterization of optimal shapes and masses through Monge–Kantorovich Equation, J. Eur. Math. Soc., **3** (2001), 139–168.

[13] G. Bouchitté, G. Buttazzo & L. De Pascale, A p-Laplacian approximation for some mass optimization problems, J.O.T.A., **118** (2003), 1–25.

[14] G. Bouchitté, G. Buttazzo & P. Seppecher, Shape Optimization Solutions via Monge–Kantorovich Equation, C. R. Acad. Sci. Paris, **324–I** (1997), 1185–1191.

[15] G. Bouchitté, G. Buttazzo & P. Seppecher, Energies with respect to a measure and applications to low dimensional structures, Calc. Var., **5** (1997), 37–54.

[16] Y. Brenier, Décomposition polaire et réarrangement monotone des champs de vecteurs, C. R. Acad. Sci. Paris Sér. I Math., **305** no. 19 (1987), 805–808.

[17] Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions, Comm. Pure Appl. Math., **44** no. 4 (1991), 375–417.

[18] Y. Brenier, A Homogenized Model for Vortex Sheets, Arch. Rat. Mech. Anal., **138** (1997), 319–353.

[19] Y. Brenier, A Monge-Kantorovich approach to the Maxwell equations, Hyperbolic problems: theory, numerics, applications, Vol. I, II (Magdeburg, 2000), 179–186, Internat. Ser. Numer. Math., 140, 141, Birkhäuser, Basel, 2001.

[20] Y. Brenier, Extended Monge–Kantorovich theory, in "Optimal Transportation and Applications", Lecture Notes in Mathematics, LNM **1813**, Springer (2003), 91–121.

[21] Y. Brenier & W. Gangbo, L^p approximation of maps by diffeomorphisms, Calc. Var. **16** (2003), 147–164.

[22] H. Brezis, Analyse fonctionnelle: théorie et application, Collection mathématiques appliquées pour la maîtrise, Masson, Paris (1983).

[23] L. Caffarelli, Allocation maps with general cost functions, Lecture Notes in Pure and Appl. Math., **177** (1996), 29–35.

[24] L. Caffarelli, M. Feldman & R.J. McCann, Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs, J. Amer. Math. Soc., **15** (2002), 1–26.

[25] C. Castaing & M. Valadier, Convex analysis and measurable multifunctions, Lecture Notes in Mathematics, **580**, Springer (1977).

[26] B. Dacorogna & J. Moser, On a partial differential equation involving the Jacobian determinant, Ann. Ist. H. Poincaré, **7** (1990), 1–26.

[27] C. Dellacherie & P.A. Meyer, Probabilities and potential, North-Holland Publ. Co. (1978).

[28] G. Dal Maso, An Introduction to Γ-Convergence, Birkhäuser (1993).

[29] R.O. Davies, On accessibility of plane sets and differentiation of functions of two real variables, Proc. Cambridge Philos. Soc., **48** (1952), 215–232.

[30] L. De Pascale, On a constrained shape optimization problem, Communications in Applied Analysis, Vol. **7** n. 4 (2003).

[31] L. De Pascale & A. Pratelli, Regularity properties for Monge transport density and for solutions of some shape optimization problem, Calc. Var., **14** (2002), 249–274.

[32] L. De Pascale & A. Pratelli, Sharp summability for Monge Transport density via interpolation, to appear on COCV.

[33] L. De Pascale, L.C. Evans & A. Pratelli, Integral Estimates for Transport Densities, to appear on Bulletin of the London Mathematical Society.

[34] R.J. Di Perna & P.L. Lions, Ordinary differential equations, transport theory and Sobolev spaces, Invent. Math., **98** no. 3 (1989), 511–547.

[35] I. Ekeland & R. Temam, Analyse Convexe et Problèmes Variationnels, Dunod Gauthier–Villars (1976).

[36] L.C. Evans, Partial Differential Equations, Graduate Studies in Mathematics, Vol. **19**, American Mathematical Society (1998).

[37] L.C. Evans, Partial Differential Equations and Monge–Kantorovich Mass Transfer, Current Developments in Mathematics (1997), 65–126.

[38] L.C. Evans, Some new PDE methods for weak KAM theory, to appear in Calc. Var. P.D.E. .

[39] L.C. Evans & W. Gangbo, Differential Equations Methods for the Monge–Kantorovich Mass Transfer Problem, Memoirs of the A.M.S., Vol. **137**, Number 653, (1999).

[40] K.J. Falconer, The geometry of fractal sets, Cambridge University Press (1985).

[41] H. Federer, Geometric measure theory, Springer (1969).

[42] M. Feldman & R.J. McCann, Uniqueness and transport density in Monge's mass transportation problem, Calc. Var. P.D.E. **15** (2002), 81–113.

[43] M. Feldman & R.J. McCann, Monge's transport problem on a Riemannian manifold, Trans. Amer. Math. Soc. **354** (2002), 1667–1697.

[44] I. Fragalà & C. Mantegazza, On some notions of tangent space to a measure, Proc. Roy. Soc. Edinburgh, **129A** (1999), 331–342.

[45] W. Gangbo, An elementary proof of the polar factorization of vector-valued functions, Arch. Rat. Mat. Anal., **128** (1994), 381–399.

[46] W. Gangbo, The Monge mass transfer problem and its applications, Contemp. Math., **226** (1999), 79–104.

[47] W. Gangbo & R.J. McCann, The geometry of optimal transportation, Acta Math., **177** (1996), 113–161.

[48] W. Gangbo & R.J. McCann, Shape recognition via Wasserstein distance, Quart. Appl. Math., **58** no. 4 (2000), 705–737.

[49] M. Giaquinta, Introduction to regularity theory for nonlinear elliptic systems, Birkhäuser Verlag (1993).

[50] S. Hohloch, Optimale Massebewegung im Monge–Kantorovich–Transportproblem, Degree thesis, University of Freiburg, Germany, 2002.

[51] M. Iri, Theory of flows in continua as approximation to flows in networks, in "Survey of Math Programming", North-Holland (1979).

[52] U. Janfalk, On certain problems concerning the $p-$Laplace operator, Linköping Studies in Science and Technology, Dissertation # 326, Linköping University, Sweden (1993).

[53] L.V. Kantorovich, On the transfer of masses, Dokl. Akad. Nauk. SSSR, **37** (1942), 227–229.

[54] L.V. Kantorovich, On a problem of Monge, Uspekhi Mat. Nauk., **3** (1948), 225–226.

[55] H.G. Kellerer, Duality Theorems for Marginal Problems, Z. Wahrscheinlichkeitstheorie verw. Gebiete, **67** (1984), 399-432.

[56] M. Knott & C. Smith, On the optimal mapping of distributions, J. Optim. Theory Appl., **43** no. 1 (1984), 39–49.

[57] M. Knott & C. Smith, On Hoeffding–Frechet bounds and cyclice monotone relations, J. Multivariate Anal., **40** (1992), 328–334.

[58] D.G. Larman, A compact set of disjoint line segments in $\mathbf{R}^3$ whose end set has positive measure, Mathematika, **18** (1971), 112–125.

[59] P. Marcellini, General growth conditions and regularity, in "Variational methods for discontinuous structures", edited by Serapioni, Raul et al., Progress in Nonlinear Differential Equations and Applications #25, Birkhauser (1996), 111-118.

[60] G. Monge, Memoire sur la Theorie des Déblais et des Remblais, Hist. de l'Acad. des Sciences de Paris (1781).

[61] A. Pratelli, Regolarità delle misure ottimali nel problema di trasporto di Monge–Kantorovich, Degree thesis, University of Pisa, Italy, 2000.

[62] A. Pratelli, Equivalence between some definitions for the optimal mass transport problem and for the transport density on manifolds, to appear on Ann. Mat. Pura Appl. .

[63] A. Pratelli, On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation, forthcoming.

[64] S.T. Rachev & L. Rüschendorf, Mass Transportation Problems, Springer–Verlag (1998).

[65] T. Rockafellar, Characterization of the subdifferential of convex functions, Pacific Journal of Math **17** (1966), 497–510.

[66] H.L. Royden, Real Analysis (2nd edition), Macmillan, (1968).

[67] W. Rudin, Real and Complex Analysis, McGraw-Hill, (1970).

[68] L. Rüschendorf, The Wasserstein distance and approximation theorems, Z. Wahrsch. Verw. Gebiete, **70** no. 1 (1985), 117–129.

[69] L. Rüschendorf, On c−optimal random variables, Statistics & Probability Letters, **27** (1996), 267–270.

[70] L. Simon, Lectures on Geometric Measure Theory, Proceedings of the Centre for Mathematical Analysis, Australian National University, Volume **3** (1983).

[71] G. Strang, Maximal flow through a domain, Math. Programming, **26** (1983), 123–143.

[72] V.N. Sudakov, Geometric Problems in the Theory of Infinite–Dimensional Probability Distributions, Proc. of the Steklov Institute of Mathematics, **141** (1979).

[73] N.S. Trudinger & X.J. Wang, On the Monge mass transfer problem, Calc. Var. P.D.E., **13** (2001), 19–31.

[74] C. Villani, Topics in mass transportation, forthcoming book by AMS.