CIMAT

Centro de Investigación en Matemáticas, A.C.

THE L^1 THEORY OF OPTIMAL TRANSPORT AND FLOW MINIMIZATION PROBLEMS

T H E S I S

As a Degree Requirement for

Master of Science

with specialty in

Pure Mathematics

Presented by

Emmanuel Rivera Guasco

Advisor:

Dr. Héctor Andrés Chang Lara

Autorización de la versión final

Guanajuato, Gto. 26 de marzo de 2021

Contents

Acknowledgment

Throughout the writing of this dissertation, I have received a great deal of support and assistance.

First, I would like to thank my supervisor, professor Héctor Chang Lara, for his invaluable support during the development of this work. Without his guidance and valuable teachings, it would have been impossible to finish my thesis.

I thank my parents and brothers for their unconditional support, they were always there for me when I felt sad and have always cheered me up when needed. I also thank all my friends in Guanajuato and the CIMAT community for their support and companionship.

Finally, I would like to thank the National Council of Science and Technology (CONACYT) for the scholarship which allowed me to start and conclude successfully my master's degree.

Introduction

Historically, optimal transport theory is originated in the XVIII century thanks to the French mathematician Gaspard Monge, who proposed the problem of finding the most efficient way of moving a sand pile onto another pile with the same volume [18]. Here, efficient means that the transportation has to be chosen so that it minimizes the average displacement. Surprisingly, it turned out be a very difficult problem at that time, and remained untouched until new developments were proposed by the Soviet mathematician Leonid Kantorovich. In 1942, Leonid Kantorovitch proposed a relaxed version of Monge's problem by using a more modern approach which appears more natural from a theoretical perspective. This relaxation gave the possibility to attack Monge's transportation problem and, some decades later, to provide solutions. Roughly speaking, this formulation says: instead of giving for each particle x a specific destination $S(x)$, we will allow the mass of a particle x to be split among several target destinations y. From then on, Monge-Kantorovich type problems have become an important subject of study on probability theory and convex optimization.

More recently, new interests aroused in optimal transport due to Yann Brenier's work on polar decomposition of vector-valued functions ([9], 1987) and its application to determine generalized solutions to the Euler equations ([19], 1989). Also, V. N. Sudakov's work in Monge's original problem ([25], 1979) aroused interest among the mathematical community; however, the solution that he proposed was incomplete. In 2003, this issue was observed and fixed by Luigi Ambrosio in [1]. In the meantime, a different solution of Monge's transport problem was obtained by Lawrence C. Evans and Wilfrid Gangbo in ([16], 1999), via partial differential equations techniques. Nowadays, Optimal Transport has become into a powerful modern theory which has a profound interaction among various fields of mathematics such as partial differential equations [21], gradient flows [4], fluid mechanics [10] and geometry [19]. This also explains why optimal transport theory has found several applications in economics [17], urban planning [11], engineering and statistics.

The aim of this thesis work is to present a rigorous proof of Ambrosio-Sudakov's Theorem, which states that Monge's Transportation problem, with cost $c(x, y) = |x - y|$ admits a unique solution. The proof that we present here is essentially an adaptation made by Filippo Santambrogio in [23] from Ambrosio's original paper [1]. Furthermore, we will study a minimal flow problem introduced by Martin Beckmann in [5], and rigorously prove that L^1 Monge's transport problem and Beckmann's problem are equivalent, via a decomposition theorem due to Smirnov [24].

Now, we give the overall structure of this thesis work: In Chapter 1, we will study some basic notions in optimal transport theory. Section 1.1 presents a brief introduction to Monge-Kantorovich problems, explaining its mathematical formulation. We also study the differences between these two transport problems, and we prove that Kantorovich's problem admits an optimal transport plan. In section 1.2, we establish a strong duality result for Kantorovich's problem and we also prove necessary and sufficient conditions for optimality. Moreover, we show that Kantorovich's dual problem admits a solution over the space of c-concave functions. Finally, we study existence and uniqueness results when the cost is a strictly convex function. In particular, we present Brenier's theorem which states that Kantorovich's problem with quadratic cost admits a unique optimal transport plan which is induced by the gradient of a convex function.

In Chapter 2, we will study the L^1 theory of optimal transportation. Section 2.1 presents some issues of the one-dimensional Monge-Kantorovich transport problem, and shows that, under a convexity assumption on the cost function, there exists a unique solution to Kantorovich's problem and it is induced by the monotone transport map. In Section 2.2, we mainly focus on developing necessary tools and results to prove Ambrosio-Sudakov's Theorem; to accomplish this task, we split this section into two subsections. In the first subsection, we show Kantorovich-Rubinstein's formula and present the secondary variational problem associated to L^1-Kantorovich's problem. The second subsection presents a further study on the geometric properties of transport rays and how a Kantorovich potential interact with them. We also establish Lipschitz regularity results for the directions of all transport rays. Finally, we state and prove Ambrosio-Sudakov's Theorem.

In Chapter 3, we will present Beckmann's minimal flow problem and its connection with the L^1 theory of optimal transportation. In particular, we will exploit its relation with Monge's original problem. Section 3.1 presents a brief remainder about existence and uniqueness of the solutions of the continuity equation. In Section 3.2, we first prove that Beckmann's problem admits a solution which can be built from a solution of L^1-Kantorovich's problem, also we associate to this solution a scalar measure known as transport density. Moreover, we introduce the notion of transport intensity and transport flow to show Smirnov's decomposition theorem; this theorem will allow us to prove that any optimal solution of Beckmann's problem is induced by an optimal transport plan for L^1-Kantorovich's problem, and such a solution does not depend on the choice of the optimal plan, which means that any optimal transport plan gives us the same solution and transport density. Finally, Section 3.3 presents L^p integrability results of the transport density associated to the unique solution of Beckmann's problem.

Chapter 1

Kantorovich and Monge transport problems.

In this chapter, we will present some basic tools and classical results in optimal transport theory needed for this thesis. For a more complete introduction the reader may refer to Cédric Villani's book [26] or Filippo Santambrogio's book [23], which are the references consulted for the development of this chapter.

1.1 A brief introduction to optimal transport theory

Historically, optimal transport theory has its origins in the XVIII century thanks to the French mathematician Gaspard Monge, who proposed the problem of finding the most efficient (economic) way of moving a mass distribution into another one with respect to a given cost function, which represents the needed work to move a unit mass from one place to another [18]. In modern mathematical notation the problem is the following: given two positive densities ϕ, ψ on $\mathbb{R}^3$, with $\int_{\mathbb{R}^3} \phi(x)dx = \int_{\mathbb{R}^3} \psi(y)dy$, the aim is to find a map $T : \mathbb{R}^3 \to \mathbb{R}^3$ such that

$$\int_A \quad (y)dy = \int_{T^{-1}(A)} \phi(x)dx, \tag{1.1.1}$$

for any Borel set $A \subset \mathbb{R}^3$, and

$$\left\{ \int_{\mathbb{R}^3} |T(x) - x| \phi(x)dx : T \text{ satisfies } (1.1.1) \right\}$$

is minimized. In this case, the map T describes the movement and $T(x)$ represents the destination of a unit mass originally located at x. This problem remained unsolved until Sudakov proposed a partial solution in 1979 [25]. Many years later, Ambrosio observed that Sudakov's original proof was incomplete (due to regularity issues) and fixed it in [1]. Nowadays, this problem is not only important from a theoretical point of view, but because of its applications in mathematical economics (see for instance, [17]).

Now, we present the modern formulation of Monge's problem. Let $T : X \to Y$ be a Borel map, with X, Y measure spaces. **The push forward** (or image measure) of a measure μ on X through T is the Borel measure, denoted by $T_\# \mu$, and defined on Y by

$$T_\# \mu(B) = \mu(T^{-1}(B)),$$

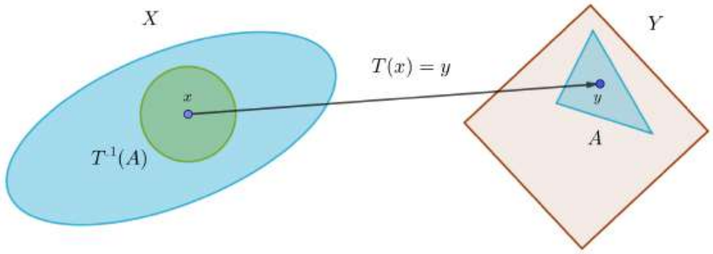

Figure 1.1: A transport map from μ to ν.

for every Borel subset $B \subset Y$. A Borel map $T : X \to Y$ is said to be a **transport map** between μ and ν, if $T_{\#}\mu = \nu$. Let us remark that $T_{\#}\mu$ can equivalently be defined by the change of variables formula:

$$\int_Y f d(T_{\#}\mu)(y) = \int_X f(T(x))d\mu(x), \tag{1.1.2}$$

for every Borel function $f : Y \to \bar{\mathbb{R}}$.

Example 1.1.1. Suppose that $T : \Omega_1 \to \Omega_2$ is a diffeomorphism, and Ω_1, Ω_2 are open subsets of $\mathbb{R}^d$. We also assume that $d\mu = \rho_1 d\mathscr{L}^d, d\nu = \rho_2 d\mathscr{L}^d$ are probability measures on X and Y, respectively. Then, T is a transport map if

$$\int_Y \rho_2(y)dy = \int_X \rho_2(T(x))\mathrm{Det}(\mathrm{DT(x)})dx = \int_X \rho_1(x)dx.$$

Hence, T is a transport map if and only if the following equation holds:

$$\rho_1(x) = \rho_2(T(x))\mathrm{Det}(DT(x)) \text{ almost everywhere.}$$

Problem 1.1.2 (Monge's problem)**.** *Given two probability measures $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$ and a cost function $c : X \times Y \to [0, \infty]$. Consider the following problem which we will call **Monge's transport problem**:*

$$\inf \left\{ M(T) := \int c(x, T(x))d\mu : T_{\#}\mu = \nu \right\}, \tag{MP}$$

Once that we have stated the Monge problem, a natural question arises: do minimizers exist? What we usually do to prove existence is the following: take a minimizing sequence $\{T_n\}_{n\geq 1}$, find a bound on it giving compactness in some topology (here, if the support of ν is compact, then the maps T_n take value in a common bounded set, and so one can get compactness of T_n in the weak-* L^∞ convergence), take a limit $T_n \rightharpoonup T$, and prove that T is a minimizer. This requires semi-continuity of the functional M with respect to this convergence; but we also need that the limit T still satisfies Monge's constraint. In general we can construct a sequence of maps $\{T_n\}_{n\geq 1}$ such that $T_n \rightharpoonup T$, but T does not satisfy the constraint. As an example of this phenomenon, one can consider the sequence $T_n(x) = T(nx)$ where $T : \mathbb{R} \to \mathbb{R}$ is a 1-periodic function such that

$$T(x) = \begin{cases} 1, & \text{if } x \in [0, \frac{1}{2}) \\ -1, & \text{if } x \in [\frac{1}{2}, 1), \end{cases}$$

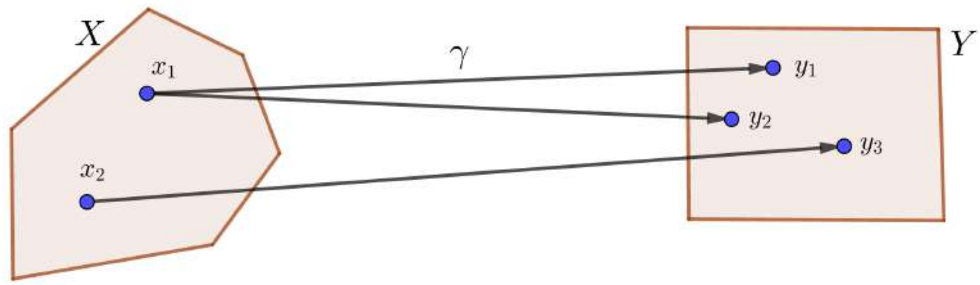

Figure 1.2: A transport plan γ in Kantorovich's problem.

and the measures $\mu = \mathcal{L}\,|_{[0,1]}$ and $\nu = \frac{(\delta_1+\delta_{-1})}{2}$. One may check that $(T_n)_\#\mu = \nu$ for all $n \in \mathbb{N}$; nevertheless, $T_n \rightharpoonup \overline{T} = 0$ and this function satisfies $\overline{T}_\#\mu = \delta_0 \neq \nu$. Therefore, this constraint on T is not closed under weak convergence.

Now, we give an example where an optimal transport does not exist.

Example 1.1.3. Consider the measures $\mu = \delta_a$, $a \in X$, and suppose that ν is not a Dirac measure. In this case there is no transport map, since $T_\#\delta_a = \delta_{T(a)}$ and ν is not of the form δ_b for some $b \in Y$. In general, measures with atoms cannot be sent through a transport map to atomless measures.

Because of these difficulties, we will forget (MP) for a while and pass to the generalization known as Kantorovich's problem. This problem give us an alternative way to describe the displacement of mass. Instead of giving for each x the destination $T(x)$, we give for each pair (x, y) the number of particles going from x to y.

Definition 1.1.4. *Given $\mu \in \mathscr{P}(X)$ and $\nu \in \mathscr{P}(Y)$, we define the set of transport plans as (see figure 1.2):*

$$\Pi(\mu,\nu) = \{\gamma \in \mathscr{P}(X \times Y) : (\pi_x)_\#\gamma = \mu, \quad (\pi_y)_\#\gamma = \nu\}, \tag{1.1.3}$$

where π_x and π_y are the two projections of $X \times Y$ onto X and Y, respectively.

Problem 1.1.5 (Kantorovich's problem)**.** *Let $\mu \in \mathscr{P}(X)$ and $\nu \in \mathscr{P}(Y)$ be probability measures, and a cost function $c : X \times Y \to [0, \infty]$. Consider the following problem which we will call* ***Kantorovich's problem****:*

$$\inf\left\{K(\gamma) := \int_{X\times Y} c d\gamma : \gamma \in \Pi(\mu,\nu)\right\} \tag{KP}$$

The minimizers of this problem are called optimal transport plans between μ and ν.

Example 1.1.6 (The discrete case)**.** Suppose that X and Y are discrete spaces where all points have the same mass:

$$\mu = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}, \qquad \nu = \frac{1}{n}\sum_{i=1}^{n}\delta_{y_j}.$$

Since all points $\{x_i\}_{i=1}^n$ and $\{y_j\}_{j=1}^n$ have the same mass, Monge's problem can be written as a minimization problem over all bijections $T : X \to Y$, that is

$$\inf\left\{\frac{1}{n}\sum_{i=1}^{n} c(x_i, y_{\sigma(i)}) : \sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\} \text{ is a permutation}\right\}.$$

On the other hand, any measure in $\Pi(\mu, \nu)$ can be represented as a double-stochastic $n \times n$ matrix $\pi = (\pi_{ij})$, which means that all the coefficients π_{ij} are non-negative and for all $j \geq 1$, $\sum_i \pi_{ij} = 1$, and for all $i \geq 1$, $\sum_j \pi_{ij} = 1$. Hence, Kantorovich's problem can be rewritten as the following finite dimensional linear optimization problem:

$$\inf\left\{\frac{1}{n}\sum_{ij}\pi_{ij}c(x_i, y_j) : \pi \in \mathscr{B}\right\},$$

where $\mathscr{B}$ is the set of all $n \times n$ double-stochastic matrices. It can be proved that $\mathscr{B}$ is a convex set, and the minimum of the discrete (KP) is attained in one of the extreme points of $\mathscr{B}$. On the other hand, a famous theorem due to Birkhoff (see [26]) states that the set of extremal points of $\mathscr{B}$ is exactly the set of permutation matrices, i.e matrices whose entries are 0 or 1. Therefore, the solution to the discrete (KP) is a permutation matrix, which correspond to a transport map (a bijection from X to Y)

Remark 1.1.7. Thanks to the previous example we observe that, although Monge's problem involves fewer variables than Kantorovich's problem, the existence of minimizers might be harder to prove in the first case (due to the constraint on σ). In fact, we will prove later that, under certain conditions, the infimum is attained in the general case.

Let $T : X \to Y$ be a measurable map, if $\gamma \in \Pi(\mu, \nu)$ is an optimal transport plan of the form $\gamma := (Id, T)_{\#}\mu = \gamma_T$, then T will be called an optimal transport map from μ to ν. Since $(\mathrm{Id}, T)_{\#}^{-1}\mu(C) = \mu\left((\mathrm{Id}, T)^{-1}(C)\right)$ for any γ_T measurable set C and

$$(\mathrm{Id}, T)^{-1}(A \times Y) = \{x \in X : x \in A,\ \ T(x) \in Y\} = A$$
$$(\mathrm{Id}, T)^{-1}(X \times B) = \{x \in X : T(x) \in B\} = T^{-1}(B)$$

for any $A \subset X$ and $B \subset Y$. Then the following lemma is just a straightforward consequence of the change of variable formula (1.1.2).

Lemma 1.1.8. *$\gamma_T \in \Pi(\mu, \nu)$ if and only if T pushes μ onto ν. In such a case, the functional K takes the form $\int_X c(x, T(x))d\mu(x)$.*

From now on, we will understand that X, Y are Polish spaces.

Notation 1.1.9.
- We denote by $\mathscr{M}_+(X)$ the set of finite positive measures on X. Thanks to the Riesz representation theorem we can endow this space with the norm $\|\lambda\| = |\lambda|(X)$, with $|\lambda|$ denotes the total variation measure (see [20]).

- $C_0(X)$ denotes the space of continuous functions vanishing at infinity (for every $\epsilon > 0$ there exist a compact subset $K \subset X$ so that $|f(x)| < \epsilon$ on X/K) endowed with the supremum norm.

- $C_b(X)$ denotes the set of bounded continuous functions on X endowed with the supremum norm. If X is compact, then $C_0(X) = C_b(X) = C(X)$.

Definition 1.1.10. *We will say that μ_n **weakly converges** to μ if and only if for every $\phi \in C_b(X)$ we have that $\int \phi \, d\mu_n \to \int \phi \, d\mu$. We denote by $\mu_n \rightharpoonup \mu$ this type of convergence.*

In a probability space $\mathscr{P}(X)$ it is natural to work in the general setting of Polish spaces. If we require X to be a Polish space then $\mathscr{P}(X)$ is tight, and hence we may apply Prokhorov's theorem to derive several results related with the convergence of subsequences to some probability distribution. We next present an example to illustrate this point.

Example 1.1.11 (Probabilistic interpretation)**.** Let X, Y be two Polish spaces and let $\mu \in \mathscr{P}(X)$ and $\nu \in \mathscr{P}(Y)$. A common problem in probability theory is to minimize the expected value of some function c, over all pairs (X, Y) of random variables such that $\mathrm{Law}(X) = \mu$ and $\mathrm{Law}(Y) = \nu$, in other words

$$\inf \left\{ \mathbb{E}[c(X,Y)] : \mathrm{Law}(X) = \mu, \text{ and } \mathrm{Law}(Y) = \nu \right\},$$

where this expected value depends on the joint law of (X, Y), which is the main known. Transport plans $\gamma \in \Pi(\mu, \nu)$ are all possible joint laws between X and Y. In particular, if X, Y are vector valued random variables with cost $c(X, Y) = |X - Y|^p$, then the problem reads as follows:

$$\inf \left\{ \|X - Y\|_p : \mathrm{Law}(X) = \mu, \text{ and } \mathrm{Law}(Y) = \nu \right\}$$

If $p = 2$, by performing some easy computations it can be shown that the previous problem is reduced to the maximization of the quantity $\mathbb{E}[(X - x_0) \cdot (Y - y_0)]$, where $x_0 = \mathbb{E}(X) = \int x d\mu$ and $y_0 = \mathbb{E}(Y) = \int y d\nu$. This means that we must find the joint law of random variables with maximal covariance.

In Kantorovich's problem we also need to allow the cost function to be lower semi-continuous in order to encompass certain type of applications. Let us give an example of this situation.

Example 1.1.12 (Transport problem with l.s.c cost)**.** Let X be a metric space, and $\mu, \nu \in \mathscr{P}(X)$. Consider the lower semi-continuous cost function

$$c(x,y) = \begin{cases} 0, & \text{if } x = y \\ 1, & \text{if } x \neq y, \end{cases}$$

which is a distance in X. In this particular case, it can be proved that the optimal transportation cost is

$$\min(\mathrm{KP}) = \frac{1}{2}|\mu - \nu|,$$

where the right hand side of the equation is the total variation of $\mu - \nu$. We will postpone the proof of this claim until chapter two (Example 2.2.6).

Finally, we prove that Kantorovich's problem admits a minimizer by applying a direct method in calculus of variations. We will also use Prokhorov's compactness theorem without giving its proof (the proof of this result can be found in [6]).

Theorem 1.1.13. *Let X and Y be Polish spaces, $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$ and $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ a lower semi-continuous function. Then the Kantorovich problem (KP) admits a solution, that is, there is a $\gamma \in \Pi(\mu, \nu)$ that minimizes K.*

Proof. In order to prove compactness of $\Pi(\mu, \nu)$, we will apply Prokhorov's Theorem. First, we show that $\Pi(\mu, \nu)$ is tight (see Definition A.0.6). Indeed, let $\epsilon > 0$; since probability measures are inner regular (because every probability measure is a Radon measure), then we can find compact sets $K_X \subset X$ and $K_Y \subset Y$ such that $\mu(X \setminus K_X) < \frac{\epsilon}{2}$ and $\mu(Y \setminus K_Y) < \frac{\epsilon}{2}$. So, for any $\gamma \in \Pi(\mu, \nu)$ we have:

$$\begin{aligned}\gamma((X \times Y) \setminus (K_X \times K_Y)) &\leq \gamma((X \setminus K_X) \times Y) + \gamma(X \times (Y \setminus K_Y)) \\ &= \mu(X \setminus K_X) + \nu(Y \setminus K_Y) \\ &< \epsilon.\end{aligned}$$

Therefore, $\Pi(\mu, \nu)$ is tight. By Prokhorov's Theorem it follows that for every sequence $\{\gamma_n\}_{n\geq 1}$ there exist $\gamma \in \mathscr{P}(X \times Y)$ and a subsequence $\{\gamma_{n_k}\}_{k\geq 1}$ such that $\gamma_{n_k} \rightharpoonup \gamma$. It remains to show that $\gamma \in \Pi(\mu, \nu)$. Let $\{\gamma_n\}_{n\geq 1} \in \Pi(\mu, \nu)$ such that, for every $\phi \in C_b(X \times Y)$ we have:

$$\int_{X\times Y} \phi d\gamma_n \longrightarrow \int_{X\times Y} \phi d\gamma.$$

Let us fix a point $y_0 \in Y$, then $\phi(x, y_0) \in C_b(X)$ and

$$\int_X \phi(x, y_0) d\mu(x) = \int_{X\times Y} \phi(x, y_0) d\gamma_n(x, y) \longrightarrow \int_{X\times Y} \phi(x, y_0) d\gamma(x, y) = \int_X \phi(x, y_0) d((\pi_x)_\# \gamma)$$

for all $\phi \in C_b(X \times Y)$, then it follows that $(\pi_x)_\# \gamma = \mu$. Similarly $(\pi_y)_\# \gamma = \nu$. Hence, $\gamma \in \Pi(\mu, \nu)$.
On the other hand, since c is lower semi-continuous and bounded from below, then there exists a sequence $\{c_j\}_{j\geq 1}$ of j-Lipschitz functions such that, for every $(x, y) \in X \times Y$ we have $c_j(x, y) \nearrow c(x, y)$. Thus, we can write

$$K(\gamma) = \sup_{j\geq 1} K_j(\gamma),$$

where $K_j(\gamma) = \int c_j d\gamma$ (thanks to the monotone convergence Theorem). Since, each K_j is continuous for the weak convergence and the supremum of continuous functions is lower semi-continuous, then the functional $K : \mathscr{M}_+(X \times Y) \to \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous.

Let $\{\gamma_n\}_{n\geq 1}$ be a minimizing sequence, i.e. $K(\gamma_n) \to \inf_{\gamma \in \Pi(\mu,\nu)} K(\gamma)$. By Weierstrass's Theorem (see Appendix A, Theorem A.0.3) there exists a transport plan $\gamma^* \in \Pi(\mu, \nu)$ such that

$$K(\gamma^*) = \inf \{K(\gamma) : \gamma \in \Pi(\mu, \nu)\}.$$

□

1.2 Kantorovich's duality

Kantorovich's problem (KP) is a linear optimization problem with convex constraints. Thus, we may apply the duality theory for convex problems and try to find the dual problem (DP) for (KP).

Let $\gamma \in \mathscr{M}_+(X \times Y)$. Then, we have

$$\sup_{(\phi,\psi)\in C_b(X)\times C_b(Y)} \left\{ \int_X \phi d\mu + \int_Y \psi d\nu \right\} - \int_{X\times Y} (\phi(x) + \psi(y)) d\gamma = \begin{cases} 0, & \text{if } \gamma \in \Pi(\mu,\nu) \\ +\infty, & \text{otherwise.} \end{cases}$$

Hence, (KP) can be rewritten as follows:

$$\inf_{\gamma\in\Pi(\mu,\nu)} \int_{X\times Y} c d\gamma + \sup_{\phi,\psi} \int_X \phi d\mu + \int_Y \psi d\nu - \int_{X\times Y} (\phi(x) + \psi(y)) d\gamma, \tag{1.2.1}$$

if we assume that it is possible to interchange sup with inf (in other words, if we suppose a mini-max principle holds), then

$$\sup_{\phi,\psi} \int \phi d\mu + \int_Y \psi d\nu + \inf_{\gamma\in\Pi(\mu,\nu)} \int_{X\times Y} c(x,y) - (\phi(x) + \psi(y)) d\gamma. \tag{1.2.2}$$

Notice that one can rewrite the infimum over γ as a constraint on ϕ and ψ:

$$\inf_{\gamma\in\mathscr{M}_+(X\times Y)} \int_{X\times Y} (c - \phi \oplus \psi) d\gamma = \begin{cases} 0, & \text{if } \phi \oplus \psi \leq c \\ -\infty, & \text{otherwise,} \end{cases}$$

where $(\phi \oplus \psi)(x,y) := \phi(x) + \psi(y)$. Indeed, let $\gamma \in \mathscr{M}_+(X \times Y)$. If $\phi \oplus \psi > c$ for some measurable set A with $\gamma(A) > 0$, let $(x_0, y_0) \in A$ and $\epsilon > 0$ such that $\phi(x_0) + \psi(y_0) - c(x_0, y_0) = \epsilon > 0$; let us consider the measure $\gamma_\lambda = \lambda \delta_{(x_0,y_0)}$ for some $\lambda > 0$. Clearly, $\gamma_\lambda \in \mathscr{M}_+(X \times Y)$ and we have that:

$$\int_A (c - (\phi \oplus \psi)) d\gamma \geq \int_A c(x,y) - (\phi(x) + \psi(y)) d\gamma_\lambda = -\lambda\epsilon \to -\infty,$$

when λ goes to ∞.

Before giving a formal formulation of this optimization problem, let us give an informal interpretation of this duality, given by Caffarelli and Villani in [26].

The shipper's problem: Assume for a moment that you are the owner of some bakeries and coffee shops in Venice, and daily need to transport a certain amount of bread to each coffee shop. To accomplish this task, you hire workers who will charge you $c(x,y)$ for transporting a bread tray located at bakery x to the coffee shop located at y. Moreover, suppose that bakeries daily production and the bread needed at each coffee shop are fixed. Now, let us imagine the following situation: while you are trying to figure out the way to minimize your transportation cost, a friend of yours comes over and offers you to transport the bread with his truck, and agrees to set the prices such that

$$\phi(x) + \psi(y) \leq c(x,y),$$

with $\phi(x)$ the price for loading a bread tray located at bakery x, and $\psi(y)$ the price for delivering it at the coffee shop located at y. Surprisingly, Kantorovich's duality tells us that your friend may arrange the prices in such a way that you will pay him exactly the same as you would have spent by doing the transport by yourself.

Problem 1.2.1 (Dual problem)**.** *Let $\mu \in \mathscr{P}(X)$ and $\nu \in \mathscr{P}(Y)$ probability measures, and a cost function $c : X \times Y \to [0, \infty]$ we consider the problem*

$$\sup \left\{ \int_X \phi d\mu + \int_Y \psi d\nu : (\phi, \psi) \in C_b(X) \times C_b(Y),\ \phi \oplus \psi \leq c \right\}. \tag{DP}$$

Remark 1.2.2. One can easily show that sup(DP) $\leq$ min(KP). Indeed, let (ϕ, ψ) be any admissible pair. Then,

$$\int_{X\times Y} c d\gamma \geq \int_{X\times Y} (c - \phi \oplus \psi) d\gamma + \int_X \phi d\mu + \int_Y \psi d\nu \geq \int_X \phi d\mu + \int_Y \psi d\nu,$$

since $c - \phi \oplus \psi \geq 0$. Hence, if we take infimum over all $\gamma \in \Pi(\mu, \nu)$, and supremum with respect to (ϕ, ψ), then the desired inequality holds.

The following definition gives us a natural candidate to be an optimizer for (DP) once one of the functions ϕ or ψ is fixed.

Definition 1.2.3 (c-concavity)**.** *Given a function $f : X \to \overline{\mathbb{R}}$, we define the c-transform $f^c : Y \to \overline{\mathbb{R}}$ (see figure 1.3) by*

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

*We also define the **$\overline{\mathbf{c}}$-transform** of $g : Y \to \overline{\mathbb{R}}$ by*

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

- *A function ψ on X is said to be **c-concave** if there is $\phi : Y \to \overline{\mathbb{R}}$ such that $\psi = \phi^{\overline{c}}$. We denote by $c - conc(X)$ the set of c-concave functions.*
- *A function ψ on Y is said to be **$\overline{\mathbf{c}}$-concave** if there is $\phi : X \to \overline{\mathbb{R}}$ such that $\psi = \phi^c$. We denote by $\overline{c} - conc(Y)$ the set of $\overline{c}$-concave functions.*

Example 1.2.4. Let $c(x, y) = \frac{1}{2}|x - y|^2$, $\lambda > 0$ and consider the function $\chi(x) = \frac{1}{2}(1 - \lambda)|x|^2$ defined on a compact domain $\Omega \subset \mathbb{R}^d$. Now, we compute χ^c. Indeed,

$$\chi^c(y) = \inf_{x\in\mathbb{R}^d} \left\{ \frac{1}{2}|x - y|^2 - \frac{1}{2}(1 - \lambda)|x|^2 \right\} \geq \inf_{x\in\mathbb{R}^d} \left\{ \frac{1}{2}(|x| - |y|)^2 - \frac{1}{2}\lambda|x|^2 \right\}.$$

Then, we have equality when x is a scalar multiple of y and the second infimum is realized whenever $\|x\| = \frac{1}{\lambda}\|y\|$. Hence, if we substitute $x = \frac{1}{\lambda}y$ we get that $\chi^c(y) = -\frac{1-\lambda}{2\lambda}|y|^2$.

Now, we recall the definition of the Legendre transformation (or convex conjugate).

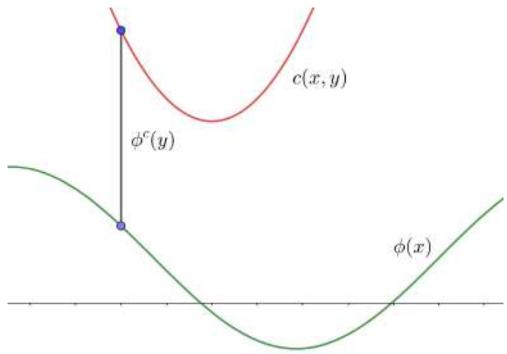

Figure 1.3: Geometric representation of a c-concave function

Definition 1.2.5 (Legendre transform)**.** *Let $(X, \langle\cdot,\cdot\rangle)$ be a Hilbert space. Given a function $f : X \to \overline{\mathbb{R}}$, we define the Legendre transform or convex conjugate $f^* : X \to \overline{\mathbb{R}}$ as follows:*

$$f^*(y) := \sup_{x \in X} \left\{ \langle x, y\rangle - f(x) \right\}.$$

The following proposition gives us a connection between c-concavity and convexity.

Proposition 1.2.6. *Given a function $\chi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, let us define $u_\chi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ through $u_\chi(x) = \frac{1}{2}|x|^2 - \chi(x)$. Then we have $u_{\chi^c} = (u_\chi)^*$. In particular, a function ϕ is c-concave if and only if $x \mapsto \frac{1}{2}|x|^2 - \phi(x)$ is convex and lower semi-continuous.*

Proof. First, we observe that

$$\begin{aligned} u_{\chi^c}(x) = \frac{1}{2}|x|^2 - \chi^c(x) &= \sup_{y \in \mathbb{R}^d} \left\{ \frac{1}{2}|x|^2 - \frac{1}{2}|x-y|^2 + \chi(y) \right\} \\ &= \sup_{y \in \mathbb{R}^d} \left\{ \frac{1}{2}|x|^2 - \frac{1}{2}|x|^2 + x \cdot y - \frac{1}{2}|y|^2 + \chi(y) \right\} \\ &= \sup_{y \in \mathbb{R}^d} \left\{ x \cdot y - \left(\frac{1}{2}|y|^2 - \chi(y) \right) \right\} \\ &= \sup_{y \in \mathbb{R}^d} \left\{ x \cdot y - u_\chi(y) \right\} \\ &= (u_\chi)^*(x). \end{aligned}$$

The second assertion follows from the fact that a function is convex and lower semi-continuous if and only if it is written as the supremum of affine functions. $\square$

We now focus on the existence of an optimal pair (ϕ, ψ) for (DP). Once this result has been proved, we will show that strong duality holds.

The following proposition gives us some useful properties of c-concave functions.

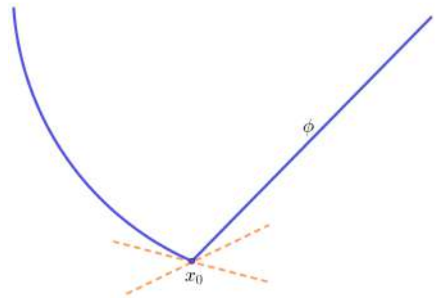

Figure 1.4: Subgradient of a convex function. Geometrically, the elements of $\partial\phi(x_0)$ or subgradients at x_0 are vectors which are perpendicular to the tangent planes on $\phi(x_0)$.

Proposition 1.2.7. *Let X and Y be two nonempty sets, and let $c(x,y) : X \times Y \to \mathbb{R} \cup \{+\infty\}$ be a function. Let $\phi : X \to \mathbb{R} \cup \{+\infty\}$, then:*

- $\phi(x) + \phi^c(y) \leq c(x,y)$
- $\phi^{c\bar{c}c} = \phi^c$.
- $\phi^{c\bar{c}} \geq \phi$. *We have equality $\phi^{c\bar{c}} = \phi$ if and only if ϕ is c-concave. In general, $\phi^{c\bar{c}}$ is the smallest c-concave function larger than ϕ.*

Let $\phi : X \to \overline{\mathbb{R}}$ be a lower semi-continuous function; the **c-superdifferential** is defined as:

$$\partial^c\phi := \{(x,y) \in X \times Y : \phi(x) + \phi^c(y) = c(x,y)\}.$$

Also, we can define $\partial^c\phi(x) := \{y \in Y : (x,y) \in \partial^c\phi\}$.
If ϕ is convex and $c(x,y) = |x-y|^2$, then the definition of $\partial^c\phi(x)$ coincides with the classical definition of **subdifferential**,

$$\partial\phi(x) := \left\{ y \in \mathbb{R}^d : \phi(x) + \phi^*(y) = x \cdot y \right\},$$

where ϕ^* denotes the Legendre transform of ϕ (see figure 1.4).

Remark 1.2.8.

- Let $g_\alpha : X \to \bar{\mathbb{R}}$ be a family of functions such that

$$|g_\alpha(x) - g_\alpha(x')| \leq \omega(d_X(x,x')),$$

with ω a modulus of continuity. Then, $g(x) = \inf_\alpha g_\alpha(x)$ shares the same modulus of continuity.

- If $c : X \times Y \to \mathbb{R}$ is uniformly continuous, then there exists a modulus of continuity $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ with $\omega(0) = 0$ such that $|c(x,y) - c(x',y')| \leq \omega\left(d_X(x,x') + d_Y(y,y')\right)$. Hence, we have

$$|(c(x,y) - \phi(x)) - (c(x,y') - \phi(x))| \leq \omega(d_Y(y,y')),$$

which means that $|\phi^c(y) - \phi^c(y')| \leq \omega(d_Y(y,y'))$ (by the definition of c-transform and the first part of this Remark). Applying a similar reasoning we can show that $\phi^{c\bar{c}}$ also shares the same continuity modulus.

- Given an admissible pair (ϕ, ψ) in the maximization problem (DP), one always can replace it with (ϕ, ϕ^c) and then with $(\phi^{c\bar{c}}, \phi^c)$. Indeed, let us fix $y_0 \in Y$, if $c(x,y) - \phi(x)$ realize its infimum at x_0, then $\phi^c(y_0) + \phi(x_0) = c(x_0, y_0)$ (i.e, $(x_0, y_0) \in \partial^c \phi$). In general, by Proposition 1.2.7 we have that $\phi \oplus \phi^c \leq c$ and therefore (ϕ, ϕ^c) is an admissible pair for (DP). On the other hand, by definition we have

$$\phi^{c\bar{c}}(x) = \inf_{y \in Y} \{c(x,y) - \phi^c(y)\}.$$

Then, $\phi^{c\bar{c}}(x) \leq c(x,y) - \phi^c(y)$ for all $(x,y) \in X \times Y$. Therefore, $(\phi^{c\bar{c}}, \phi^c)$ is an admissible pair as well.

Remark 1.2.9. If (ϕ_n, ψ_n) is a maximizing sequence, then we may find a maximizing sequence of the form $(\phi_n^{c\bar{c}}, \phi_n^c)$. Indeed, let us suppose that

$$\int \phi_n d\mu + \int \psi_n d\nu \longrightarrow \sup(\text{DP}).$$

By the constraints on the dual problem and Proposition 1.2.7 we have that, $\psi_n \leq \phi_n^c$ and $\phi_n^{c\bar{c}} \geq \phi_n$. Then,

$$\int_X \phi_n^{c\bar{c}} d\mu + \int_Y \phi_n^c d\nu \geq \int_X \phi_n d\mu + \int_Y \psi_n d\nu \longrightarrow \sup(\text{DP}).$$

Hence, $\displaystyle\sup_{\phi \in c-\text{conc}(X)} \left\{ \int_X \phi d\mu + \int_Y \phi^c d\nu \right\} = \sup(\text{DP})$ since $(\phi_n^{c\bar{c}}, \phi_n^c)$ also satisfies the constraints of (DP).

Note that the notion of c-transform allows us to improve a maximizing sequence $\{(\phi_n, \psi_n)\}_{n \geq 1}$ by replacing it with the pair $(\phi_n^{c\bar{c}}, \phi_n^c)$, and so we can get a uniform bound. After that, we prove that the sequence is uniformly bounded as well. Hence, by Arzela-Ascoli's Theorem we guarantee the existence of a convergent subsequence to an admissible pair (ϕ, ϕ^c) which is optimal.

Theorem 1.2.10. *Suppose that X and Y are compact metric spaces and $c : X \times Y \to \overline{\mathbb{R}}$ is continuous. Then there exists a solution (ϕ, ψ) to the dual problem (DP) and it has the form $\phi \in$ c-conc(X), $\psi \in$ $\bar{c}$-conc(Y), and $\psi = \phi^c$. In particular*

$$\sup(DP) = \max_{\phi \in c-conc(X)} \left\{ \int_X \phi d\mu + \int_Y \phi^c d\nu \right\}.$$

Proof. Let (ϕ_n, ψ_n) be a maximizing sequence; then by Remarks 1.2.8 and 1.2.9 we can improve the sequence by using c and $\bar{c}$-transforms, so that the new maximizing sequence shares the same modulus of continuity. In order to simplify notation we will denote by (ϕ_n, ψ_n) this new sequence. Then, we can assume a uniform bound on the continuity of these functions since, $|\phi^{c\bar{c}}(x) - \phi^{c\bar{c}}(x')| \leq \omega(d_X(x,x'))$ and $|\phi^c(y) - \phi^c(y')| \leq \omega(d_Y(y,y'))$. So, the sequence is equicontinuous

It only remains to check equiboundedness in order to apply Arzela-Ascoli's Theorem. Since each ϕ_n is continuous on a compact set we can extract its minimum and without loss of generality suppose that $\min_{x \in X} \phi_n(x) = 0$ for all $n \in \mathbb{N}$. Then for every $x \in X$ we have:

$$|\phi_n(x)| = |\phi_n(x) - \phi_n(x_0)| \leq \omega(d_X(x, x_0)) \leq \omega(\operatorname{diam}(X)) = M,$$

for all $n \in \mathbb{N}$. So, $\sup_{n\in\mathbb{N}} |\phi_n(x)| \leq M$. Thus, if $\psi_n = \phi_n^c$, then we also have $\psi_n(y) = \inf_{x\in X} \{c(x, y) - \phi_n(x)\} \in [\min c - M, \max c]$. This shows that $\{\phi_n\}_{n\geq 1}$ and $\{\psi_n\}_{n\geq 1}$ are uniformly bounded. Hence, up to a subsequence we can assume that $\phi_n \to \phi$ and $\psi_n \to \psi$, let us see that $\psi = \phi^c$. Indeed, let $y \in Y$, then

$$\begin{aligned}
\psi(y) = \lim_{n\to\infty} \psi_n(y) &= \lim_{n\to\infty} \phi_n^c(y) \\
&= \lim_{n\to\infty} \inf_{x\in X} \{c(x, y) - \phi_n(x)\} \\
&= \inf_{x\in X} \left\{c(x, y) - \lim_{n\to\infty} \phi_n(x)\right\} \\
&= \inf_{x\in X} \{c(x, y) - \phi(x)\} \\
&= \phi^c(y),
\end{aligned}$$

since $\phi_n \to \phi$ uniformly. By applying the dominated convergence theorem, it is easily seen that

$$\int_X \phi_n d\mu + \int_Y \psi_n d\nu \longrightarrow \int_X \phi d\mu + \int_Y \psi d\nu.$$

Moreover, we have that $\phi(x) + \psi(y) \leq c(x, y)$. This shows that (ϕ, ψ) is an admissible pair for (DP) and that it is optimal (since the original sequence is minimizing). $\square$

Remark 1.2.11. If we assume that strong duality holds, then Theorem 1.2.10 implies

$$\min(\mathrm{KP}) = \max_{\phi\in c-\mathrm{conc}(X)} \left\{\int_X \phi d\mu + \int_Y \phi^c d\nu\right\},$$

which also shows that the minimum of (KP) is a convex function of (μ, ν).

Definition 1.2.12. *The functions ϕ realizing the maximum of (DP) are called* ***Kantorovich potentials*** *for the transport plan between μ and ν.*

Now, we will focus on proving that strong duality holds and characterizing optimal transport plans. To do so, we give some preliminary results without proof (see [23, 26]).

Definition 1.2.13. *On a separable metric space X, the* ***support*** *of a measure γ is defined as the smallest closed set on which γ is concentrated:*

$$spt(\gamma) := \bigcap \{A : A \text{ is closed and } \gamma(X \setminus A) = 0\}.$$

Moreover, we also have the following characterization:

$$spt(\gamma) = \{x \in X : \gamma(B(x, r)) > 0 \ \ \forall r > 0\}.$$

Definition 1.2.14. *Let $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ be a function. We say that a set $\Gamma \subset X \times Y$ is* ***c-cyclically monotone*** *(denoted c-CM) if for every $k \in \mathbb{N}$, every permutation σ of $\{1, 2, \ldots, k\}$ and every finite collection of points $(x_1, y_1), \ldots, (x_k, y_k) \in \Gamma$ we have:*

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

The following theorem is a well known generalization of Rockafeller's Theorem (see Theorem A.0.13).

Theorem 1.2.15. *Let $\Gamma \neq \emptyset$ be a c-CM subset in $X \times Y$ and $c : X \times Y \to \mathbb{R}$. Then, there exists a c-concave function $\phi : X \to \mathbb{R} \cup \{-\infty\}$, with $\phi \neq -\infty$ such that $\Gamma \subset \partial^c \phi$.*

Example 1.2.16. Let $c(x, y) = |x - y|^2$, and $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$ be a c-cyclically monotone set. If σ is a cycle, then for any choice of points $(x_1, y_1), \ldots, (x_T, y_T) \in \Gamma$ we have

$$\sum_{i=1}^{T} x_i \cdot y_i = \sum_{i=1}^{T} |x_i - y_i|^2 \geq \sum_{i=1}^{T} |x_i - y_{i+1}|^2 = \sum_{i=1}^{T} x_i \cdot y_{i+1}, \tag{1.2.3}$$

with $y_{T+1} = y_1$. Every set Γ satisfying (1.2.3) is said to be **cyclically monotone**. Hence, if ϕ is lower semi-continuous and convex, then Theorem 1.2.15 implies that

$$\Gamma \subset \left\{ (x, y) \in \mathbb{R}^d \times \mathbb{R}^d : \phi(x) + \phi^*(y) = x \cdot y \right\},$$

which means that Γ is contained in the graph of $\partial\phi$ (Rockafeller's Theorem).

Although the condition of being a c-CM set may seem strange at first, its continuous counterpart when $c = |x - y|^2$ is clearer. Let $\rho : \{1, 2, \ldots, T, T+1\} \to \{x_1, x_2, \ldots, x_{T+1}\}$ be a discrete closed path, such that $\rho(T + 1) = \rho(0)$. If we imagine that y is a vector field so that $y(x_i) = y_i$ and $y(t) = y(\rho(t))$, then by (1.2.3) we have that

$$\sum_{t=1}^{T} y(t) \cdot (\rho(t) - \rho(t-1)) \geq 0, \quad \sum_{t=1}^{T} y(t-1) \cdot (\rho(t) - \rho(t-1))$$

which means in a continuous setting that $\oint y d\rho \leq 0$ and $\oint y d\rho \geq 0$. Hence, y is a conservative vector field.

Remark 1.2.17. Informally, we might think that c-CM sets provide a criteria to find local minimizers. Indeed, as in Example 1.1.6, let us consider the discrete Kantorovich's problem, and take $A = \{(x_1, y_1), \ldots (x_k, y_k)\} \subset X \times Y$; if A is not c-CM, then it cannot be contained in the support of an optimal transport plan, since we can find a permutation σ such that the optimal transport plan can be improved by sending some mass from x_j to $y_{\sigma(j)}$, for all $j \in \{1, 2, \ldots, k\}$. Moreover if we keep applying this process, we can improve a transport plan until it is supported on a c-CM set (see figure 1.5).

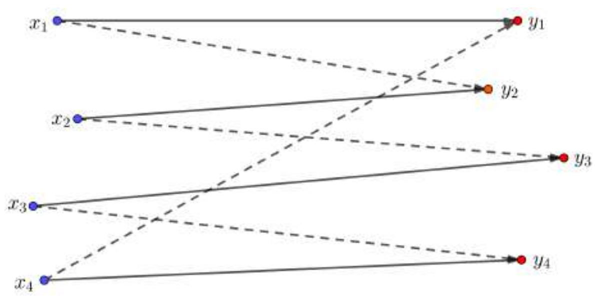

Figure 1.5: Improving the cost by a cycle in Remark 1.2.17.

Remark 1.2.18. In Example 1.1.6 we saw that the solution to the discrete Kantorovich problem is an $n \times n$ permutation matrix, which represents a transport map. Let us assume that $y_i = T(x_i)$, and T is optimal. Then, we have the optimality condition:

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

for every permutation σ. Hence, in the discrete setting, the support of γ is a c-cyclically monotone set.

Now, we shall prove that any optimal transport plan is concentrated on a c-CM set.

Theorem 1.2.19. *If γ is an optimal transport plan for the cost $c : X \times Y \to \mathbb{R}$ and c is continuous, then spt(γ) is a c-CM set.*

Proof. We prove the statement by contradiction: Suppose that $\mathrm{spt}(\gamma)$ is not a c-CM set, then there is a number $N \in \mathbb{N}$, $\{(x_i, y_i)\}_{i=1}^{N} \subset \mathrm{spt}(\gamma)$ and a permutation σ such that

$$\sum_{i=1}^{N} c(x_i, y_i) > \sum_{i=1}^{N} c(x_i, y_{\sigma(i)}).$$

Let $\epsilon < \frac{1}{2N} \sum_{i=1}^{N} (c(x_i, y_i) - c(x_i, y_{\sigma(i)}))$. Since c is continuous there exists $\delta > 0$ such that for all $i \in \{1, \ldots, N\}$ and $(x, y) \in B(x_i, \delta) \times B(y_i, \delta)$ we have

$$c(x, y) > c(x_i, y_i) - \epsilon,$$

and also for all $(x, y) \in B(x_i, \delta) \times B(y_{\sigma(i)}, \delta)$ we have that

$$c(x, y) < c(x_i, y_{\sigma(i)}) + \epsilon.$$

Intuitively, the idea of the proof is to build a new measure $\bar{\gamma}$ which contradicts optimality of γ. Indeed, let $V_i = B(x_i, \delta) \times B(y_i, \delta)$; since $(x_i, y_i) \in \mathrm{spt}(\gamma)$, then $\gamma(V_i) > 0$ for all $i \in \{1, \ldots, N\}$. Now, we define the measures $\gamma_i := \frac{1}{\gamma(V_i)} \gamma \mid_{V_i}$, $\mu_i := (\pi_x)_{\#} \gamma_i$ and $\nu_i := (\pi_y)_{\#} \gamma_i$. For any i we can consider the measure $\bar{\gamma_i} = \mu_i \otimes \nu_{\sigma(i)} \in \Pi(\mu_i, \nu_{\sigma(i)})$ and take $\epsilon_0 < \frac{1}{N} \min_i \gamma(V_i)$. We now define the following measure:

$$\bar{\gamma} := \gamma - \epsilon_0 \sum_{i=1}^{N} \gamma_i + \epsilon_0 \sum_{i=1}^{N} \bar{\gamma}_i.$$

To finish the proof we shall prove that $\bar{\gamma} \in \Pi(\mu, \nu)$ and has lower cost than γ. We fist observe that the conditions $\epsilon_0 \gamma_i = \frac{\epsilon_0}{\gamma(V_i)} \gamma \mid_{V_i}$ and $\frac{\epsilon}{\gamma(V_i)} \leq \frac{1}{N}$ imply that $\epsilon_0 \gamma_i < \frac{1}{N}\gamma$, which means that $\gamma - \sum_{i=1}^N \gamma_i$ is positive and hence $\bar{\gamma}$ is a positive measure. Now, we compute the marginals of $\bar{\gamma}$:

$$(\pi_x)_{\#}\bar{\gamma} = (\pi_x)_{\#}\gamma - \epsilon_0 \sum_{i=1}^N (\pi_x)_{\#}\gamma_i + \epsilon_0 \sum_{i=1}^N (\pi_x)_{\#}\bar{\gamma}_i = \mu - \epsilon_0 \sum_{i=1}^N \mu_i + \epsilon_0 \sum_{i=1}^N \mu_i = \mu,$$

$$(\pi_y)_{\#}\bar{\gamma} = (\pi_y)_{\#}\gamma - \epsilon_0 \sum_{i=1}^N (\pi_y)_{\#}\gamma_i + \epsilon_0 \sum_{i=1}^N (\pi_y)_{\#}\bar{\gamma}_i = \mu - \epsilon_0 \sum_{i=1}^N \nu_i + \epsilon_0 \sum_{i=1}^N \nu_{\sigma(i)} = \nu,$$

since σ is a permutation. Thus, $\bar{\gamma} \in \Pi(\mu, \nu)$. On the other hand, since γ_i is concentrated on V_i and $\bar{\gamma}_i$ on $B(x_i, \delta) \times B(y_{\sigma(i)}, \delta)$, then we have

$$\begin{aligned}
\int_{X\times Y} c d\gamma - \int_{X\times Y} c d\bar{\gamma} &= \epsilon_0 \int_{X\times Y} c d\gamma_i - \epsilon_0 \sum_{i=1}^N c\bar{\gamma}_i \\
&\geq \epsilon_0 \sum_{i=1}^N (c(x_i, y_i) - \epsilon) \int_{X\times Y} d\gamma_i - \epsilon_0 \sum_{i=1}^N (c(x_i, y_{\sigma(i)}) + \epsilon) \int_{X\times Y} d\bar{\gamma}_i \\
&= \epsilon_0 \sum_{i=1}^N (c(x_i, y_i) - \epsilon) - \epsilon_0 \sum_{i=1}^N (c(x_i, y_{\sigma(i)}) + \epsilon) \\
&\epsilon_0 \left(\sum_{i=1}^N c(x_i, y_i) - \sum_{i=1}^N c(x_i, y_{\sigma(i)}) + 2N\epsilon \right) > 0,
\end{aligned}$$

which contradicts the optimality of γ. Therefore, $\mathrm{spt}(\gamma)$ is a c-CM set. □

Now, we present a strong duality result when the cost function is uniformly continuous and bounded.

Theorem 1.2.20. *Suppose that X and Y are Polish spaces and that $c : X \times Y \to \mathbb{R}$ is uniformly continuous and bounded. Then, the problem (DP) admits a solution (ϕ, ϕ^c) and therefore we have* $\max(DP) = \min(KP)$.

Proof. Since c is continuous, then (KP) admits a solution $\gamma \in \Pi(\mu, \nu)$. Also, $\Gamma = \mathrm{spt}(\gamma)$ is a c-CM set thanks to Theorem 1.2.19. Hence, we can apply Theorem 1.2.15 to guarantee the existence of a c-concave function ϕ such that

$$\Gamma \subset \partial^c \phi = \{(x, y) \in \Omega \times \Omega : \phi(x) + \phi^c(y) = c(x, y)\}.$$

Since $\phi^{c\bar{c}} = \phi$ and ϕ^c is $\bar{c}$-concave, then we have that ϕ and ϕ^c are continuous (since c is uniformly continuous). Moreover, we also observe that

$$\phi^c(y) = \inf_{x\in X} \{c(x, y) - \phi(x)\} \leq M_1,$$

with $M_1 > 0$ (since c is bounded). Then, ϕ^c is upper bounded and therefore ϕ is lower bounded. Analogously, from

$$\phi(x) = \inf_{y\in Y} \{c(x, y) - \phi^c(y)\} \leq M_2$$

we obtain that ϕ^c and ϕ are lower and upper bounded, respectively. This proves that ϕ and ϕ^c are bounded as well (just take $M = \max(M_1, M_2)$). Hence, (ϕ, ϕ^c) is an admissible pair for (DP).

On the other hand, since Γ is concentrated on the set where $\phi(x) + \phi^c(y) = c(x, y)$ then

$$\int_X \phi d\mu(x) + \int_Y \phi^c d\nu(y) = \int_{X\times Y} c(x, y) d\gamma(x, y),$$

which implies that

$$\sup(\text{DP}) \geq \int_X \phi d\mu + \int_Y \phi d\nu = \int_{X\times Y} c d\gamma = \min(\text{KP}).$$

Since we already know that $\sup(\text{DP}) \leq \min(\text{KP})$, then the desired equality holds and (ϕ, ϕ^c) is an optimal pair in (DP). $\square$

To prove the duality formula when the cost function is lower semi-continuous and bounded from below, we will use the fact that there exists a sequence $\{c_k\}_{k\geq 1}$ of k-Lipschitz functions such that c_k converge increasingly to c. But first, let us state the following lemma which is a straightforward application of Prokhorov's Theorem.

Lemma 1.2.21. *Suppose that $\{c_k\}_{k\geq 1}$ and c are lower semi-continuous functions bounded from below and that c_k converge increasingly to c. Then*

$$\lim_{k\to\infty} \min \left\{ \int c_k \, d\gamma : \gamma \in \Pi(\mu, \nu) \right\} = \min \left\{ \int c \, d\gamma : \gamma \in \Pi(\mu, \nu) \right\}.$$

Now, we can prove the following duality theorem.

Theorem 1.2.22 (Kantorovich's duality)**.** *If X, Y are Polish spaces and $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous and bounded from below, then the duality formula $\sup(DP) = \min(KP)$ holds.*

Proof. Let us consider a sequence $\{c_k\}_{k\geq 1}$ of k-Lipschitz function such that $c_k \nearrow c$. Then, by Theorem 1.2.19 we have strong duality for each $k \in \mathbb{N}$. Hence,

$$\begin{aligned} \min \left\{ \int c_k d\gamma : \gamma \in \Pi(\mu, \nu) \right\} &= \max_{(\phi,\psi)\in C_b(X)\times C_b(Y)} \left\{ \int \phi d\mu + \int \psi d\nu : \phi \oplus \psi \leq c_k \right\} \\ &\leq \sup_{(\phi,\psi)\in C_b(X)\times C_b(Y)} \left\{ \int \phi d\mu + \int \psi d\nu : \phi \oplus \psi \leq c \right\}, \end{aligned}$$

since every pair (ϕ, ψ) satisfying $\phi(x) + \psi(y) \leq c_k(x, y)$ also satisfies $\phi(x) + \psi(y) \leq c(x, y)$. Letting $k \to \infty$ and applying Lemma 1.2.21 to the left-hand side of the last inequality we get that:

$$\min \left\{ \int c \, d\gamma : \gamma \in \Pi(\mu, \nu) \right\} \leq \sup_{(\phi,\psi)\in C_b(X)\times C_b(Y)} \left\{ \int \phi d\mu + \int \psi d\nu : \phi \oplus \psi \leq c_k \right\}.$$

Therefore, $\sup(\text{DP}) \geq \min(\text{KP})$. Since we know that $\sup(\text{DP}) \leq \min(\text{KP})$, then the duality formula holds. $\square$

Remark 1.2.23. The strong duality formula that we proved for a lower semi-continuous cost differs from the continuous case, since our proof does not guarantee the existence of an optimal pair.

Example 1.2.24. Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be a homothety defined by $T(x) = \lambda x$, with $\lambda > 0$. We will show that for any compactly supported measure $\mu \in \mathscr{P}(\mathbb{R}^d)$, the map T is an optimal transport from μ to $T_{\#}\mu$ with respect to the quadratic cost $c(x, y) = \frac{1}{2}|x-y|^2$. Let us consider the potential $\phi(x) = \frac{1}{2}(1-\lambda)|x|^2$ and compute ϕ^c to show that T is optimal. Indeed, since μ is compactly supported, we can work on a compact domain $\Omega = \max\{1, \lambda\}\mathrm{spt}(\mu) \subset \mathbb{R}^d$ where the potential is bounded. By Example 1.2.4 we have that $\phi^c(y) = -\frac{1-\lambda}{2\lambda}|y|^2$.

By Definition of c-transform we know that $\phi(x) + \phi^c(y) \leq \frac{1}{2}|x-y|^2$. Then, we have

$$\begin{aligned}
\min(\mathrm{KP}) \geq \int_\Omega \phi(x) d\mu(x) + \int_\Omega \phi^c(x) d\nu(x) &= \int_\Omega \frac{1}{2}(1-\lambda)|x|^2 d\mu(x) - \int_\Omega \frac{1-\lambda}{2\lambda}|y|^2 d\mu(x) \\
&= \int_\Omega \frac{(1-\lambda)^2}{2}|x|^2 d\mu(x) \\
&= \int_\Omega \frac{1}{2}|(1-\lambda)x|^2 d\mu(x) \\
&= \int_\Omega c(x, T(x)) d\mu(x).
\end{aligned}$$

Hence, T is an optimal transport map.

Kantorovich's duality allows us to obtain the following theorem taken from [23].

Theorem 1.2.25. *Let c be a lower semi-continuous cost function, $\gamma \in \Pi(\mu, \nu)$ an optimal transport plan. Then, there exist a c-concave function ϕ and a c-CM set Γ such that $spt(\gamma) \subset \Gamma \subset \partial^c \phi$.*

Remark 1.2.26. If we assume in Theorem 1.2.25 that $X \times Y$ is compact and c continuous, then the c-concave function ϕ is the Kantorovich potential associated to the optimal plan γ.

1.2.1 The Kantorovich problem with strictly convex cost

In this section we will consider the case $X = Y = \Omega \subset \mathbb{R}^d$ and the cost c of the form $c(x, y) = h(x - y)$, where h is a strictly convex function. We will assume that Ω is compact for simplicity. We will prove the existence of an optimal transport T and a representation formula for it.

Proposition 1.2.27. *If $c \in C^1(\Omega \times \Omega)$, ϕ is a Kantorovich potential differentiable at x_0, where c is the transport cost from μ to ν, and $(x_0, y_0) \in spt(\gamma)$ with γ an optimal transport plan. Then, $\nabla\phi(x_0) = \nabla_x c(x_0, y_0)$.*

Proof. By the duality formula we have that $\max(\mathrm{DP}) = \min(\mathrm{KP})$ holds and both extremal values are realized (since Ω is compact). Then, if $\gamma \in \Pi(\mu, \nu)$ is optimal for (KP), then there is a Kantorovich potential ϕ such that $\phi(x) + \phi^c(y) = c(x, y)$ on $\mathrm{spt}(\gamma)$. Let us fix $(x_0, y_0) \in$

$\mathrm{spt}(\gamma)$, then $x \mapsto c(x, y_0) - \phi(x)$ is minimal at $x = x_0$, since $\phi^c(y_0) = c(x_0, y_0) - \phi(x_0)$ and $\phi^c(y_0) = \inf_{x \in \Omega} \{c(x, y_0) - \phi(x)\}$. Hence, if ϕ and $c(\cdot, y_0)$ are differentiable at x_0 and $x_0 \notin \partial\Omega$, one gets $\nabla_x(c(x_0, y_0) - \phi(x_0)) = 0$. Therefore, $\nabla\phi(x_0) = \nabla_x c(x_0, y_0)$. $\square$

The aim of this proposition is to deduce from $(x_0, y_0) \in \mathrm{spt}(\gamma)$ that y_0 is indeed uniquely defined from x_0. This would show that γ is concentrated on a graph, that of the map associating y_0 to each x_0. Furthermore, this map turns out to be the optimal transport. Before proving this fact, we state a well known result about differentiability of Lipschitz functions (the proof of this theorem may be found in [3], chapter 3).

Theorem 1.2.28 (Rademacher)**.** *Let $f : \Omega \subset \mathbb{R}^d \to \mathbb{R}$ be a locally Lipschitz function, with Ω an open set. Then, f is differentiable almost everywhere with respect to the Lebesgue measure.*

In the following theorem we will deduce that for every x_0 the point y_0 such that $(x_0, y_0) \in \mathrm{spt}(\gamma)$ is unique, in other words, $\gamma = \gamma_T$ where $T(x_0) = y_0$ by requiring that μ is absolutely continuous with respect to the Lebesgue measure.

Theorem 1.2.29. *Let $\mu, \nu \in \mathscr{P}(\Omega)$, and $\Omega \subset \mathbb{R}^d$ a compact set. Then, there exists an optimal transport plan $\gamma \in \Pi(\mu, \nu)$ for the cost $c(x, y) = h(x - y)$ with h strictly convex. If μ is absolutely continuous with respect to the Lebesgue measure and $\mu(\partial\Omega) = 0$, then the optimal transport plan is unique and of the form $(Id, T)_{\#}\mu$. Moreover, there exists a Kantorovich potential ϕ, and the map T and the potential ϕ are linked by*

$$T(x) = x - (\nabla h)^{-1}(\nabla\phi(x))$$

Proof. Theorems 1.1.13 and 1.2.10 give the existence of an optimal transport plan $\gamma \in \Pi(\mu, \nu)$ and a Kantorovich potential $\phi \in c - \mathrm{conc}(X)$. Let $(x_0, y_0) \in \mathrm{spt}(\gamma)$. If ϕ is differentiable at x_0 and $x_0 \notin \partial\Omega$, then $\nabla\phi(x_0) \in \partial h(x_0 - y_0)$. Indeed, let $z \in \Omega$ such that $y_0 + z \in \Omega$. Since $\nabla\phi(x_0) \in \partial\phi(x_0)$ and $\phi(x_0) + \phi^c(y_0) = h(x_0 - y_0)$, then we have

$$\begin{aligned} h(x_0 - y_0) - \phi^c(y_0) + \nabla\phi(x_0) \cdot (z - (x_0 - y_0)) &= \phi(x_0) + \nabla\phi(x_0) \cdot (z - (x_0 - y_0)) \\ &\leq \phi(y_0 + z). \end{aligned}$$

This last inequality implies that

$$h(x_0 - y_0) + \nabla\phi(x_0) \cdot (z - (x_0 - y_0)) \leq \phi(y_0 + z) + \phi^c(y_0) \leq c(y_0 + z, y_0) = h(z).$$

Hence, $\nabla\phi(x_0) \in \partial h(x_0 - y_0)$. From strict convexity of h we have that ∂h is univalued and injective. Therefore, $\partial h(x_0 - y_0) = \{\nabla\phi(x_0)\}$ and the relation

$$x_0 - y_0 = (\nabla h)^{-1}(\nabla\phi(x_0))$$

holds for every $(x_0, y_0) \in \mathrm{spt}(\gamma)$.

On the other hand, we note that the set of points where h is not differentiable is Lebesgue-negligible by Rademacher's Theorem. Moreover, ϕ is almost everywhere differentiable with

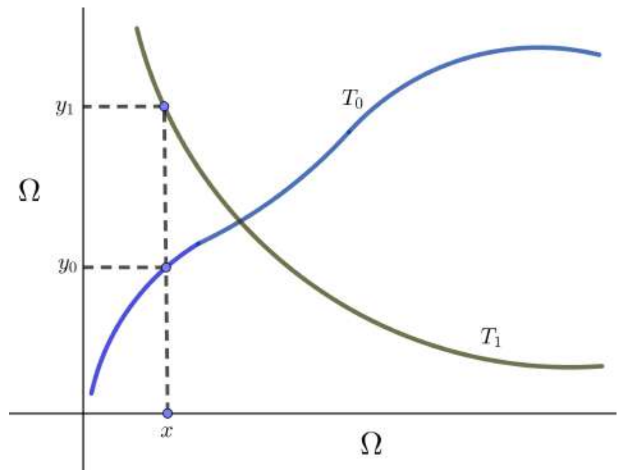

Figure 1.6: Uniqueness of the optimal map

respect to the Lebesgue measure. Indeed, since h is locally Lipschitz and Ω is bounded, then c is Lipschitz on $\Omega \times \Omega$. Also, ϕ shares the same modulus of continuity of c, which means that ϕ is Lipschitz. Hence, Rademacher's Theorem implies that ϕ is differentiable a.e.
From the absolute continuity assumption on μ we have that $\{x \in \Omega : \nabla\phi(x) \text{ does not exist}\}$ and $\partial\Omega$ are μ-negligible as well. Therefore, the map $T(x) = x - (\nabla h)^{-1}(\nabla\phi(x))$ is defined μ almost everywhere.

It remains to show that $\gamma = (\mathrm{Id}, T)_{\#}\mu$. We observe that T is a Borel map on $\mathrm{dom}(\nabla\phi)$ (Since, $\nabla\phi$ is Borel measurable on $\mathrm{dom}(\nabla\phi)$ and $(\nabla\phi)^{-1}$ is continuous). Hence, to complete the proof, it suffices to show that γ and $(\mathrm{Id}, T)_{\#}\mu$ coincides on products $U \times V$ of Borel sets $U, V \subset \Omega$. Let $A := \{(x, y) \in \mathrm{spt}(\gamma) : x \in \mathrm{dom}\nabla\phi(x)\}$. For $(x, y) \in A$ we have $y = T(x)$, then

$$(U \times V) \cap A = \left(\left(U \cap T^{-1}(V)\right) \times \Omega\right) \cap A$$

Clearly, A is Borel measurable since it is the intersection of the Borel measurable set $\mathrm{dom}\nabla\phi \times \Omega$ and the closed set $\mathrm{spt}(\gamma)$. Thus, $\gamma(W \cap A) = \gamma(W)$ for all $W \subset \Omega \times \Omega$. Applying this fact in the last inequality we get that

$$\begin{aligned}
\gamma(U \times V) &= \gamma\left(\left(U \cap T^{-1}(V)\right) \times \Omega\right) \\
&= \mu\left(\left(U \cap T^{-1}(V)\right)\right) \\
&= (\mathrm{Id}, T)_{\#}\mu,
\end{aligned}$$

since $\gamma \in \Pi(\mu, \nu)$. Therefore, we have shown that every optimal transport plan γ is induced by a transport map $T(x) = x - (\nabla h)^{-1}(\nabla\phi(x))$ which pushes μ forward to ν.

To prove uniqueness lets suppose there exist two optimal transport plans $\gamma_1 = (\mathrm{Id}, T_1)_{\#}\mu$ and $\gamma_2 = (\mathrm{Id}, T_2)_{\#}\mu$, and define $\gamma = \frac{1}{2}\gamma_1 + \frac{1}{2}\gamma_2$. We see easily see that γ is also an optimal plan, but it cannot be concentrated on the graph of some measurable function, unless $T_1 = T_2$, μ-a.e (see figure 1.6). $\square$

Let us illustrate how the last theorem works with an example.

Example 1.2.30. Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be a homothety defined by $T(x) = \lambda x$, with $\lambda > 0$. We shall prove that T is optimal by using Theorem 1.2.29. Indeed, by Example 1.2.24 we know that $\phi(x) = \frac{1}{2}(1-\lambda)|x|^2$ is a Kantorovich potential. Since μ is compactly supported and $c(x, y)$ is convex then, we may apply Theorem 1.2.29 to get that the optimal transport map T and ϕ are linked by

$$T(x) = x - (\nabla h)^{-1}(\nabla \phi(x)),$$

where $h(z) = \frac{1}{2}|z|^2$. Since, $\nabla\phi(x) = (1-\lambda)x$ and $\nabla h(z) = z$, then

$$T(x) = (\nabla h)^{-1}((1-\lambda)x) = x - (1-\lambda)x = \lambda x.$$

Remark 1.2.31. The regularity of the measure μ is very important in Theorem 1.2.29. If

$$\mu\left(\{x \in \Omega : \nabla\phi(x) \text{ does not exists}\}\right) > 0$$

or $\mu(\partial\Omega) > 0$, then we cannot guarantee the existence of an optimal transport map since T depends on the points where we can differentiate ϕ.

Remark 1.2.32. Since $c(x, y) = |x - y|^p$ with $p > 1$ is strictly convex, then by Theorem 1.2.29 we can guarantee the existence of a unique optimal transport map T such that $T(x) = x - (\nabla h)^{-1}(\nabla\phi(x))$, where ϕ is the corresponding Kantorovich potential.

1.2.2 Brenier's Theorem

In this sub-section we will briefly study the Kantorovich and Monge's problem on the quadratic case, i.e. when the cost function is of the form $c(x, y) = \frac{1}{2}|x - y|^2$. We shall prove in this case a stronger result than Theorem 1.2.29 due to Yann Brenier in [23, 26]. Finally, we will give two well-known applications of Brenier's Theorem to prove the isoperimetric inequality and Brenier's polar factorization Theorem for vector fields.

Let us first prove a compact version of Brenier's Theorem, which turns out to be a direct consequence of Theorem 1.2.29.

Remark 1.2.33. Let $\phi : \mathbb{R}^d \longrightarrow \mathbb{R} \cup \{-\infty\}$ be a c-concave function, and define $v(x) := \frac{|x|^2}{2} - \phi(x)$. Then, $y \in \partial^c\phi(x)$ if and only if $y \in \partial v(x)$. To prove this assertion, it's enough to observe that $y \in \partial^c\phi(x)$ if and only if the following holds

$$\begin{aligned} &\phi(x) = c(x, y) - \phi^c(y), \\ &\phi(z) \leq c(x, y) - \phi^c(y), \;\; \forall z \in \mathbb{R}^d, \end{aligned}$$

for any continuous cost. In the quadratic case, we would have

$$\begin{aligned} \phi(x) - \frac{|x|^2}{2} &= \langle x, -y\rangle + \frac{|y|^2}{2} - \phi^c(y) \\ \phi(z) - \frac{|z|^2}{2} &= \langle z, -y\rangle + \frac{|z|^2}{2} - \phi^c(z) \;\; \forall z \in \mathbb{R}^d. \end{aligned}$$

Therefore,

$$\begin{aligned} y \in \partial^c\phi(x) &\longleftrightarrow \phi(z) - \frac{|z|^2}{2} \leq \phi(x) - \frac{|x|^2}{2} + \langle (z - x), -y\rangle \;\; \forall z \in \mathbb{R}^d \\ &\longleftrightarrow -y \in \partial\left(\phi - \frac{|\cdot|}{2}\right)(x) \\ & y \in \partial v(x). \end{aligned}$$

Theorem 1.2.34 (Brenier)**.** *Let $c(x,y)=\frac{1}{2}|x-y|^2$ and $\mu,\nu\in\mathscr{P}(\Omega)$. If $\mu\ll\mathscr{L}^d$, then there exists a unique optimal transport plan from μ onto ν which is of the form $(Id,\nabla u)_{\#}\mu$, where u is a convex function and ∇u is the unique (up to μ-negligible sets) gradient of a convex function such that $(\nabla u)_{\#}\mu=\nu$. Moreover, if we assume that $\nu\ll\mathscr{L}^d$, then $\gamma=(\nabla u^*,Id)_{\#}\nu$ and ∇u^* pushes forward ν onto μ.*

Proof. Let $\gamma\in\Pi(\mu,\nu)$ be an optimal transport plan. From Theorem 1.2.25 and Proposition 1.2.6 we know that for any Kantorovich potential ϕ (which is c-concave in the compact case) it holds $\operatorname{spt}(\gamma)\subset\partial^c\phi$, and $u(x)=\frac{|x|^2}{2}-\phi(x)$ is convex and lower semicontinuous. Moreover, the previous remark implies that $\partial^c\phi=\partial u$. Since u is a convex function, then u is locally Lipschitz and consequently, ∇u is well-defined μ-a.e. which means that every optimal transport plan is concentrated on its graph. Therefore, $\gamma=(\mathrm{Id},\nabla u)_{\#}\mu$ (it's induced by the gradient of the convex function u).

On the other hand, we note that

$$\{(x,y):u^*(x)+u^{**}(y)=x\cdot y\}=\{(x,y):u^*(x)+u(y)=x\cdot y\},$$

since $u=u^{**}$ whenever u is convex and lower semi-continuous. Thus, $\mathrm{Graph}(\partial u)=\mathrm{Graph}(\partial u^*)$. If γ is optimal between μ and ν, then for every $(x,y)\in\operatorname{spt}(\gamma)$ we have that $y=\nabla u(x)$ is equivalent to $x\in\partial u^*(y)$. But, since u^* is ν-a.e. differentiable then the equation $x=\nabla u^*(y)=\nabla u^*(\nabla u(x))$ holds in $\operatorname{spt}(\gamma)$. Hence, $x=\nabla u^*(\nabla u(x))$ μ-a.e. Similarly, we obtain $y=\nabla u(\nabla u^*(y))$ ν-a.e. Therefore, $\gamma=(\nabla u^*,\mathrm{Id})_{\#}\nu$ and ∇u^* pushes forward ν onto μ. ☐

Remark 1.2.35. We note that the existence of an optimal transport map is true under weaker conditions assumptions on μ. For instance, we can replace the condition μ absolutely continuous with the following: $\mu(A)=0$ for any set $A\subset\mathbb{R}^d$ such that $\mathscr{H}^{d-1}(A)<+\infty$, with $\mathscr{H}^{d-1}$ the $(d-1)$-dimensional Hausdorff measure.

We observe that the previous theorem is valid only if we assume compactness on the domain Ω. In order to give a less restrictive result and to show why the quadratic case deserves special attention, we can adapt our previous analysis (by using some tools from convex analysis) to the case of unbounded domains and prove Brenier's Theorem (see [23, 26]), but first let us sketch an important characterization that we will need to prove this result.

Remark 1.2.36. Suppose that μ,ν are probability measures on $\mathbb{R}^d$ such that $\int_X|x|^2d\mu(x)+\int_Y|y|^2d\nu(y)<+\infty$. Since $|x-y|^2=|x|^2-2(x\cdot y)+|y|^2$, then minimize the Kantorovich functional with quadratic cost is equivalent to maximize $\int x\cdot y\,d\gamma$ (since we can withdraw the parts depending only on x or y in the optimization problem).

Theorem 1.2.37. *Let $\mu,\nu\in\mathscr{P}(\mathbb{R}^d)$ be two probability measures such that*

$$\int_{\mathbb{R}^d}|x|^2d\mu(x),\int_{\mathbb{R}^d}|x|^2d\nu(y)\quad<\infty,$$

and $c(x,y)=\frac{1}{2}|x-y|^2$. Then $\gamma\in\Pi(\mu,\nu)$ is an optimal transport plan if and only if spt(γ) is cyclically monotone.

Proof. According to the Example 1.2.16, we have that every c-cyclically monotone set is necessarily cyclically monotone when the cost is quadratic. Hence, if we assume that $\gamma \in \Pi(\mu,\nu)$ is optimal, then $\mathrm{spt}(\gamma)$ is a cyclically monotone set by Theorem 1.2.19.

On the other hand, let us assume that $\bar{\gamma}$ is cyclically monotone. By Rockafeller's Theorem there exists a proper convex and lower semi-continuous function u such that $\mathrm{spt}(\gamma) \subset \mathrm{Graph}(\partial u)$. Let u^* be the Legendre transform of u, then $x\cdot y \leq u(x)+u^*(y)$ with equality on $\mathrm{spt}(\gamma)$. Moreover, since we have finite second moments it is enough to show that $\bar{\gamma}$ maximizes $\int_{\mathbb{R}^d\times\mathbb{R}^d} x\cdot y d\gamma(x,y)$ over $\Pi(\mu,\nu)$. Now, by integrating the previous inequality we get:

$$\begin{aligned}\int_{\mathbb{R}^d\times\mathbb{R}^d} x\cdot y d\gamma(x,y) &\leq \int_{\mathbb{R}^d\times\mathbb{R}^d} [u(x)+u^*(y)]d\gamma(x,y)\\ &= \int_{\mathbb{R}^d} u(x)d\mu(x) + \int_{\mathbb{R}^d} u^*(y)d\nu(y)\\ &= \int_{\mathbb{R}^d\times\mathbb{R}^d} u(x)d\bar{\gamma}(x,y) + \int_{\mathbb{R}^d\times\mathbb{R}^d} u^*(y)d\bar{\gamma}(x,y)\\ &= \int_{\mathbb{R}^d\times\mathbb{R}^d} [u(x)+u^*(y)]d\bar{\gamma}(x,y)\\ &= \int_{\mathbb{R}^d\times\mathbb{R}^d} x\cdot y d\bar{\gamma}(x,y),\end{aligned}$$

for any $\gamma \in \Pi(\mu,\nu)$. Therefore, $\bar{\gamma}$ is optimal. □

Theorem 1.2.38 (Brenier)**.** *Let $\mu,\nu \in \mathscr{P}(\mathbb{R}^d)$ and $c(x,y) = \frac{1}{2}|x-y|^2$. Suppose that $\int |x|^2 d\mu(x) + \int |y|^2 d\nu(y) < +\infty$, and $\mu \ll \mathscr{L}^d$. Then, there exists a unique optimal transport plan $\gamma_T \in \Pi(\mu,\nu)$, where T can be written as the gradient of a convex function, that is $T = \nabla u$ for some convex function u. Moreover, if we assume that $\nu \ll \mathscr{L}^d$, then $\gamma = (\nabla u^*, Id)_{\#}\nu$ and ∇u^* pushes forward ν onto μ.*

Proof. Let γ be an optimal transport plan. By Theorem 1.2.37 it follows that $\mathrm{spt}(\gamma)$ is cyclically monotone (thanks to the finite second moments' hypothesis), and consequently $\mathrm{spt}(\gamma)$ is concentrated on $\mathrm{Graph}(\partial u)$ for some convex proper and lower semi-continuous function u (due to Rockafeller's Theorem). Since $u \in L^1(\mu)$, then u is finite μ-a.e. Moreover, we deduce that $\mathrm{spt}(\mu) \subset \overline{\{u<+\infty\}}$ (which is a convex set), that is $\mu\left(\overline{\{u<+\infty\}}\right) = 1$. Furthermore, the boundary of the set $\{u<+\infty\}$ is μ-negligible by Rademacher's Theorem. Hence, $\mu\left(\mathrm{Int}(\{u<+\infty\})\right) = 1$. On the other hand, u is differentiable μ-a.e on $\mathrm{Int}(\{u<+\infty\})$ which means that $\partial u(x) = \{\nabla(x)\}$ for μ-a.e point $x \in \mathbb{R}^d$. Then, we have that $y = \nabla u(x)$ for every $(x,y) \in \mathrm{spt}(\gamma)$. Therefore, γ is concentrated on the graph of ∇u, that is $\gamma = (\mathrm{Id}, \nabla u)_{\#}\mu$.

So far we have proven existence, now we will show uniqueness. Indeed, let us assume that $\gamma = (\mathrm{Id},\nabla u_1)_{\#}\mu = (\mathrm{Id},\nabla u_2)_{\#}\mu$ are both optimal. Then, (u_1,u_1^*) is an optimal pair to the dual problem as well as (u_2,u_2^*), which means we have the following equality

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} [u_2(x)+u_2^*(y)]d\gamma(x,y) = \int_{\mathbb{R}^d\times\mathbb{R}^d} [u_1(x)+u_1^*(y)]d\gamma(x,y) = \int_{\mathbb{R}^d\times\mathbb{R}^d} x\cdot y d\gamma(x,y).$$

Since $y = \nabla u_1(x)$ for every $(x,y) \in \mathrm{spt}(\gamma)$, then

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} [u_2(x) + u_2^*(\nabla u_1(x)) - x\cdot\nabla u_1(x)]d\mu(x) = 0.$$

Hence, $u_2(x) + u_2^*(\nabla u_1(x)) = x \cdot \nabla u_1(x)$, and as a consequence $\nabla u_1(x) \in \partial u_2(x)$ for μ-a.e. point $x \in \mathbb{R}^d$. Since u_2 is differentiable μ-a.e we necessarily have that $\nabla u_1 = \nabla u_2$ μ-a.e. This proves uniqueness.

Finally, the idea to prove $\gamma = (\nabla u^*, \mathrm{Id})_{\#}\nu$ is identical to the one given in the compact case. □

Corollary 1.2.39. *Let $X, Y \subset \mathbb{R}^d$ and $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$. Under the assumptions of Brenier's Theorem, we have that ∇u is the unique solution to Monge's transportation problem:*

$$\frac{1}{2}\int_x |x - \nabla u(x)|^2 d\mu(x) = \frac{1}{2}\inf_{T_{\#}\mu=\nu}\left\{\int_X |x - T(x)|^2 d\mu(x)\right\}.$$

Proof. From Section 1.1, we know that $\min(\mathrm{KP}) \leq \inf(\mathrm{MP})$. Let $\gamma \in \Pi(\mu,\nu)$ be an optimal transport plan for c; by Brenier's Theorem there exists a unique optimal transport plan γ_T for (KP), with $T = \nabla u$ some convex function u. Hence, $\gamma_T = \gamma$. Since $T(x) = y$ on $\mathrm{spt}(\gamma)$ we have

$$\begin{aligned}
\inf(\mathrm{MP}) &\leq \frac{1}{2}\int_X |x - \nabla u(x)|^2 d\mu(x) \\
&= \frac{1}{2}\int_{X\times Y} |x - \nabla u(x)|^2 d\gamma(x,y) \\
&= \frac{1}{2}\int_{X\times Y} |x - y|^2 d\gamma(x,y) \\
&= \int_X c(x,y) d\gamma(x,y) \\
&= \min(\mathrm{KP}),
\end{aligned}$$

Therefore, $\min(\mathrm{KP}) = \inf(\mathrm{MP})$. Uniqueness of T comes from uniqueness of γ_T. □

The first application of Brenier's Theorem that we will see is one of the most famous in geometry: the isoperimetric inequality. It establish that the unit ball in $\mathbb{R}^d$ is the surface with lower surface area among all surfaces with the same volume. Before proving this inequality, let us recall a well-known inequality.

Proposition 1.2.40 (Arithmetic vs geometric mean inequality)**.** *Let $x_1, x_2, \ldots, x_n \in$ be any finite collection of non-negative real numbers. Then*

$$\sqrt[n]{x_1 x_2 \cdots x_n} \leq \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Now we prove the isoperimetric inequality using optimal transport tools.

Theorem 1.2.41 (Isoperimetric inequality)**.** *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with piece-wise smooth boundary. If Vol$(B_1(0))$ = Vol(Ω), then $\mathscr{H}^{d-1}(\partial B_1(0)) \leq \mathscr{H}^{d-1}(\partial\Omega)$ (the surface area of Ω is larger than the surface area of $B_1(0)$).*

Proof. Let $f(x) = \chi_\Omega(x)$ and $g(y) = \chi_{B_1}(y)$, then the measures $d\mu = f(x)dx$, $d\nu = g(y)dy$ are absolutely continuous with respect to $\mathscr{L}^d$. If we consider the Kantorovich problem with quadratic cost, then by Brenier's Theorem there exists a unique optimal transport map $T = \nabla u$ for some convex function u. Then, for any $\psi \in C_c^\infty(B_1(0))$ we have

$$\int_\Omega \psi(\nabla u(x)) f(x)dx = \int_{B_1(0)} \psi(y)g(y)dy.$$

Also, we can apply a change of variables $y = \nabla(x)$ to get

$$\int_\Omega \psi(\nabla u(x)) f(x)dx = \int_\Omega \psi(\nabla u(x)) g(\nabla u(x)) |\det(D^2 u(x))| dx.$$

Since, ψ is an arbitrary test function it follows that

$$f(x) = g(\nabla u(x))|\det(D^2u(x))| \text{ a.e.} \quad \text{in } \Omega.$$

Thus, the convex function u satisfies that $|\det(D^2u(x))| = \frac{\chi_\Omega(x)}{\chi_{B_1}(\nabla u(x))} = 1$ almost everywhere in Ω.

On the other hand, we know that the Hessian matrix D^2u is positive semi-definite (since u is convex). Then, all its eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_d$ are non-negative real numbers, which means that $\det(D^2u) = \lambda_1\lambda_2\cdots\lambda_d$. Hence, proposition 1.2.40 implies:

$$1 = \sqrt[d]{\det(D^2u)} = \sqrt[d]{\lambda_1\lambda_2\cdots\lambda_d} \leq \frac{\lambda_1 + \lambda_2 + \cdots \lambda_d}{d} = \frac{1}{d}\text{Tr}(D^2u) = \frac{1}{d}\Delta u.$$

Now, we integrate over Ω the last inequality to obtain

$$\text{Vol}(\Omega) = \int_\Omega 1dx \leq \frac{1}{d}\int_\Omega \Delta u(x)dx.$$

Since ∇u transports Ω onto $B_1(0)$, then $|\nabla u(x)| \leq 1$. By the divergence Theorem A.0.12 we have

$$\begin{aligned}
\text{Vol}(B_1(0)) = \text{Vol}(\Omega) \leq \frac{1}{d}\int_\Omega \Delta u(x)dx = \frac{1}{d}\int_\Omega \nabla\cdot\nabla u(x)dx &= \frac{1}{d}\int_{\partial\Omega} \nabla u(x)\cdot \hat{n}d\mathscr{H}^{d-1}(x) \\
&\leq \frac{1}{d}\int_\Omega |\nabla u(x)||\hat{n}|d\mathscr{H}^{d-1}(x) \\
&\leq \frac{1}{d}\int_\Omega d\mathscr{H}^{d-1}(x) \\
&= \frac{1}{d}\mathscr{H}^{d-1}(\partial\Omega).
\end{aligned}$$

Finally, since $\text{Vol}(B_1(0)) = \frac{1}{d}\mathscr{H}^{d-1}(B_1(0))$, we conclude that $\mathscr{H}^{d-1}(B_1(0)) \leq \mathscr{H}^{d-1}(\partial\Omega)$. □

As a second application of Theorem 1.2.38, we shall prove the Brenier polar decomposition Theorem of vector fields on $\mathbb{R}^d$ (see [9]).

Definition 1.2.42. *Let (Ω, μ) be a measure space. A Borel function $s : \Omega \to \Omega$ is said to be a measure-preserving map if $s_\# \mu = \mu$*

Theorem 1.2.43 (Polar decomposition)**.** *Let $\Omega \subset \mathbb{R}^d$ be an open bounded set and $F : \Omega \to Y \subset \mathbb{R}^d$ a measurable vector field. Consider the rescaled Lebesgue measure $\mathscr{L}^d_\Omega$ on Ω and assume that $\nu = F_\# \mathscr{L}^d_\Omega$ is absolutely continuous with respect to the Lebesgue measure. Then, there exists a unique measure preserving map $s : \Omega \to \Omega$ and ∇u, with $u : \Omega \to \mathbb{R}$ convex, such that $F = \nabla u \circ s$. Moreover, s is the unique solution of*

$$\min_{\xi_\# \mu = \mu} \int |F(x) - \xi(x)|^2 dx.$$

Equivalently, s is the unique solution of

$$\max_{\xi_\# \mu = \mu} \int_\Omega F(x) \cdot \xi(x) dx.$$

Proof. Set $\mu = \mathscr{L}^d_\Omega$ and let ∇u be the optimal transport for the quadratic cost between μ and ν (whose existence is guarantee by Brenier's Theorem). Since ν is also an absolutely continuous measure we have that ∇u^* pushes forward ν onto μ, that is $(\nabla u^*)_\# \nu = \mu$ (where u^* denotes the Legendre transform u). Let $s = \nabla u^* \circ F$, so we have that $s_\# \mu = \mu$, since

$$s_\# \mu = (\nabla u^* \circ F)_\# \mu = (\nabla u^*)_\# (F_\# \mu) = (\nabla u^*)_\# \nu = \mu.$$

Moreover, since $\nabla u \circ \nabla u^* = \mathrm{Id}$ almost everywhere in Ω, then we have

$$\nabla u \circ s = \nabla u \circ (\nabla u^* \circ F) = (\nabla u \circ \nabla u^*) \circ F = \mathrm{Id} \circ F = F,$$

which gives the desired decomposition. Uniqueness of the measure preserving map s comes from the fact that ∇u is unique. Indeed, if $\bar{s}$ is another measure-preserving map such that $\nabla u \circ \bar{s} = F$, then we get $\bar{s} = \mathrm{Id} \circ \bar{s} = \nabla u^* \circ F = s$, which means that s is unique.

On the other hand, since ∇u is an optimal transport map, then we have that

$$\int_{\Omega \times \mathbb{R}^d} y \cdot x d\gamma(x, y) \leq \int_\Omega \nabla u(x) \cdot x dx,$$

for every $\gamma \in \Pi(\mu, \nu)$. Let $\xi : \Omega \to \Omega$ be any measure-preserving map and consider the measure $(\xi, F)_\# \mu \in \Pi(\mu, \nu)$. Then,

$$\begin{aligned} \int_\Omega F(x) \cdot \xi(x) dx = \int_{\Omega \times \mathbb{R}^d} y \cdot x d\left((\xi, F)_\# \mu\right)(x, y) \leq \int_\Omega \nabla u(x) \cdot x dx &= \int_\Omega \nabla u(s(x)) \cdot s(x) dx \\ &= \int_\Omega F(x) \cdot s(x) dx. \end{aligned}$$

Therefore, s is optimal. $\square$

Finally, let us finish this section by giving an example where we can explicitly compute the optimal transport map using Brenier's Theorem.

Example 1.2.44 (Radial transport problem in $\mathbb{R}^2$)**.** Let us consider the quadratic cost $c(x, y) = \frac{1}{2}|x - y|^2$, and the measures $d\mu = f d\mathscr{L}^2$ and $d\nu = g d\mathscr{L}^2$, with $f(x) = \frac{1}{\pi}\chi_{B(0,1)}(x)$ and $g(x) = \frac{1}{8\pi}(4 - |x|^2)$.

We want to find and optimal transport map T between μ and ν. To do that, we first prove the following claim: if f, g are functions satisfying:

(a) $\displaystyle\int_{\mathbb{R}^2} f d\mathscr{L}^2 = \int_{\mathbb{R}^2} g d\mathscr{L}^2$.

(b) The transport condition, $T_{\#} f\mathscr{L}^2 = g\mathscr{L}^2$ is satisfied, for $T : \mathbb{R}^2 \to \mathbb{R}^2$ of the form $T(x) = \rho(|x|)\frac{x}{|x|}$, with $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ a monotone increasing function.

Then, T is an optimal transport map with respect to the quadratic cost.
Since ρ is positive and increasing then, there exist a convex increasing function $\psi : \mathbb{R}_+ \to \mathbb{R}$ such that $\psi' = \rho$. Let us define the function $u : \mathbb{R}^2 \to \mathbb{R}$ by $u(x) := \psi(|x|)$ which is a convex function. Indeed, since ψ is increasing convex, then for all $t \in [0, 1]$ and $x, y \in \mathbb{R}^2$ we have that

$$u((1 - t)x + ty) = \psi(|(1 - t)x + ty|) \leq (1 - t)\psi(|x|) + t\psi(|y|) = (1 - t)u(x) + tu(y).$$

Then, we get that $\nabla u(x) = \psi'(x)\frac{x}{|x|} = \rho(|x|)\frac{x}{|x|} = T(x)$. Hence, Brenier's Theorem implies the desired optimality of T. Now, thanks to our previous claim and the radial symmetry of this problem we only need to find a map $T(x) = \rho(|x|)\frac{x}{|x|}$ satisfying conditions (a) and (b), which is equivalent to finding T satisfying $F_\mu(T(x)) = F_\nu(x)$, with

$$F_\nu(x) = \int_{B(0,|x|)} g d\mathscr{L}^2(t) = \frac{1}{8\pi}\int_0^{2\pi}\int_0^{|x|}(4 - t^2)t dt d\theta = \frac{|x|^2}{2} - \frac{|x|^4}{16}$$
$$F_\mu(x) = \int_{B(0,|x|)} f d\mathscr{L}^2(t) = \frac{1}{\pi}\int_0^{2\pi}\int_0^{|x|} t dt d\theta = |x|^2.$$

Then, we want to solve the equation, $\frac{|\rho(|x|)|^2}{2} - \frac{|\rho(|x|)|^4}{16} = |x|^2$ which is equivalent to solving $|y|^2(|y|^2 - 8) = -16|x|^2$, if we consider $y = \rho(|x|)$. Then, if we simplify this expression and substitute with $z = |y|^2$ we obtain the equation $z^2 - 8z = -16|x|^2$. Write the left hand side as a square and take square root of both sides of the equation to get

$$z = 4\sqrt{1 - |x|^2} + 4 \text{ or } z = -4\sqrt{1 - |x|^2} + 4.$$

Thus, if we substitute back $z = |y|^2$ and take the square root of both sides we have that

$$|y| = \pm\sqrt{4\sqrt{1 - |x|^2} + 4} = \pm 2\sqrt{\sqrt{1 - |x|^2} + 1} \text{ or}$$
$$|y| = \pm\sqrt{-4\sqrt{1 - |x|^2} + 4} = \pm 2\sqrt{1 - \sqrt{1 - |x|^2}},$$

since we are looking for positive functions, then the only two possible candidates are:

$$y = 2\sqrt{\sqrt{1 - |x|^2} + 1} \text{ or } 2\sqrt{1 - \sqrt{1 - |x|^2}}.$$

Therefore, if we take $\rho(|x|) = 2\sqrt{1 - \sqrt{1 - |x|^2}}$, then one obtains the optimal transport map

$$T(x) = 2\sqrt{1 - \sqrt{1 - |x|^2}}\frac{x}{|x|}.$$

Chapter 2

L^1 Optimal Transport Theory

In this chapter, we give a detailed exposition about the Monge's transport problem for the cost $c(x, y) = |x - y|$. We will prove that, under certain hypothesis on the source measure we can guarantee the existence of an optimal transport map. The proof that we present here follows Santambrogio's presentation [23], and is originally due to Ambrosio and Sudakov in [1, 25].

2.1 The one dimensional transport problem

In this section, we look at the one-dimensional transport problem. We will prove that there exists a unique optimal transport plan induced by a transport map which is obtained by monotone rearrangement. We are interested in studying this problem in $\mathbb{R}$ because it turns out that, if we consider the restriction of an optimal plan for the L^1 transport problem over a certain family of segments, then it behaves exactly as the one-dimensional optimal transport map. The results that we include in this section without proof can be found in [23, 27].

Definition 2.1.1. *Given a probability measure $\mu \in \mathscr{P}(\mathbb{R})$ we define its **cumulative distribution function** (CDF) F_μ trough:*

$$F_\mu(x) = \mu((-\infty, x]),$$

which is nondecreasing, right-continuous and continuous at any point where μ has no atom.

Definition 2.1.2. *Given a nondecreasing and right-continuous function $F : \mathbb{R} \to [0, 1]$ its **pseudo-inverse** is the function given by*

$$F^{[-1]}(x) := \inf \{t \in \mathbb{R} : F(t) \geq x\},$$

where the infimum is a minimum as soon as the set is nonempty and bounded from below (otherwise it is $-\infty$), thanks to the right-continuity of F.

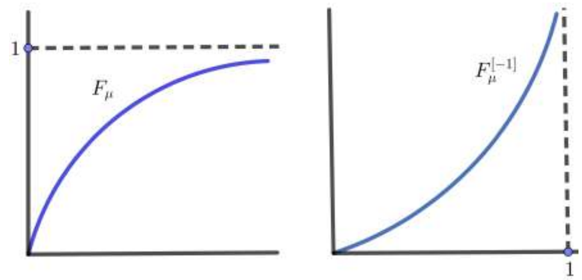

Figure 2.1: Cumulative distribution function F_μ and its pseudo-inverse $F_\mu^{[-1]}$

Note that the following properties hold, thanks to the definition of pseudo-inverse:

- $F^{[-1]}(x) \leq a$ if and only if $F(a) \geq x$.
- $F^{[-1]}(x) > a$ if and only if $F(a) < x$.

Now, we look at some very important properties of the pseudo-inverse.

Theorem 2.1.3. *If $\mu \in \mathscr{P}(\mathbb{R})$ and $F_\mu^{[-1]}$ is the pseudo-inverse of a cumulative distribution function F_μ, then $(F_\mu^{[-1]})_\#(\mathscr{L}^1 \mid_{[0,1]}) = \mu$. Moreover, given $\mu, \nu \in \mathscr{P}(\mathbb{R})$, if we set the measure $\gamma_{mon} := (F_\mu^{[-1]}, F_\nu^{[-1]})_\#(\mathscr{L}^1 \mid_{[0,1]})$, then $\gamma_{mon} \in \Pi(\mu, \nu)$ and*

$$\gamma_{mon}((-\infty, a] \times (-\infty, b])) = \min\{F_\mu(a), F_\nu(b)\}.$$

This plan is known as the co-monotone transport plan between μ and ν.

Example 2.1.4. Let us consider uniform densities on $[a, b]$ and $[c, d]$, respectively; that is, $\mu = \frac{1}{b-a} \cdot \mathscr{L}^1 \mid_{[a,b]}$ and $\nu = \frac{1}{d-c} \cdot \mathscr{L}^1 \mid_{[c,d]}$. Then, we easily compute the cumulative distribution function in each case: $F_\mu(x) = \frac{x-a}{b-a}$ and $F_\nu(y) = \frac{y-c}{d-c}$. Then, the pseudo-inverse for these measures is given by: $F_\mu^{-1}(p) = (1-p)a + pb$ and $F_\nu^{-1}(r) = (1-r)c + rd$. Thus, the co-monotone transport plan is such that

$$\begin{aligned}
\gamma_{\text{mon}}((-\infty, x_0] \times (-\infty, y_0]) &= \mathscr{L}^1 \left(\{p \in [0,1] : (1-p)a + pb \leq x_0,\ (1-p)c + pd \leq y_0\}\right) \\
&= \mathscr{L}^1 \left(\left\{ p \in [0,1] : p \leq \frac{x_0 - a}{b - a},\ p \leq \frac{y_0 - c}{d - c} \right\} \right) \\
&= \min \left\{ \frac{x_0 - a}{b - a}, \frac{y_0 - c}{d - c} \right\}.
\end{aligned}$$

Given $\mu, \nu \in \mathscr{P}(\mathbb{R})$, what we will do next is to build a monotone transport map pushing forward μ onto ν.

Theorem 2.1.5. *Given two measures $\mu, \nu \in \mathscr{P}(\mathbb{R})$, suppose that μ is atomless. Then, there exists a unique (up to μ-negligible sets) nondecreasing map $T_{mon} : \mathbb{R} \to \mathbb{R}$ such that $(T_{mon})_\#\mu = \nu$. This map is defined through $T_{mon}(x) := F_\nu^{[-1]}(F_\mu(x))$ (see figure 2.2).*

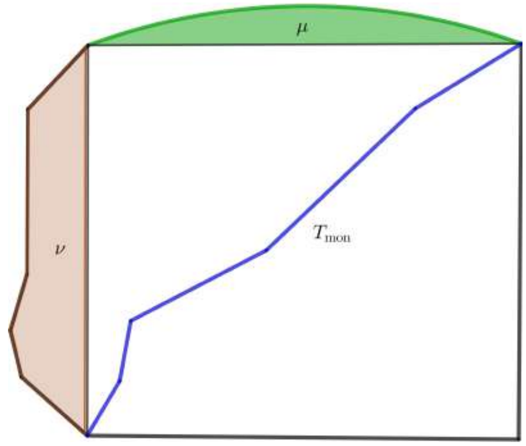

Figure 2.2: Monotone transport map T_{mon}.

The following characterization lemma will be crucial to prove that the monotone map T_{mon} optimizes a whole class of convex transport costs.

Lemma 2.1.6. *Let $\gamma \in \Pi(\nu, \mu)$ be a transport plan between two measures $\mu, \nu \in \mathscr{P}(\mathbb{R})$. Suppose that the following property is satisfied:*

$$(x, x'), (y, y') \in spt(\gamma) \text{ with } x < x' \text{ implies that } y \leq y'.$$

Then, we have $\gamma = \gamma_{mon}$. In particular, there is a unique γ satisfying the previous condition. Moreover, if μ is atomless, then $\gamma = \gamma_{T_{mon}}$ (up to countable sets).

Proof. To verify the first part of the statement, we just need to prove $\gamma((-\infty, a] \times (-\infty, b]) = \min\{F_\mu(a), F_\nu(b)\}$, since the sets $(-\infty, a] \times (-\infty, b]$ generates all the open products $U \times V$.

Let us consider the following sets $A = (-\infty, a] \times (b, +\infty)$ and $B = (a, +\infty) \times (-\infty, b]$. We know by assumption that (x, x'), $(y, y') \in \operatorname{spt}(\gamma)$ with $x < x'$, implies $y \leq y'$. Thus, if $(x_0, y_0) \in \operatorname{spt}(\gamma)$ then

$$\operatorname{spt}(\gamma) \subset \Big\{(a, b) \in \mathbb{R}^2 : a \leq x_0, \;\; b \leq y_0\Big\} \cup \Big\{(a, b) \in \mathbb{R}^2 : a \geq x_0, \;\; b \geq y_0\Big\}$$

Hence, it is not possible to have $\gamma(A)$ and $\gamma(B)$ both positive, otherwise we would have points in the support not satisfying the hypothesis (see figure 2.3). Then, we can write

$$\gamma((-\infty, a] \times (-\infty, b]) = \min\{\gamma((-\infty, a] \times (-\infty, b] \cup A), \gamma((-\infty, a] \times (-\infty, b] \cup B)\}.$$

But,

$$\gamma((-\infty, a] \times (-\infty, b] \cup A) = \gamma((-\infty, a] \times \mathbb{R}) = F_\mu(a)$$
$$\gamma((-\infty, a] \times (-\infty, b] \cup B) = \gamma(\mathbb{R} \times (-\infty, b]) = F_\nu(b).$$

Then, $\gamma((-\infty, a] \times (-\infty, b]) = \min\{F_\mu(a), F_\nu(b)\}$. Therefore, $\gamma = \gamma_{\text{mon}}$. To prove the second statement, we assume that μ is atomless. For any point $x \in \mathbb{R}$, we define the interval I_x as

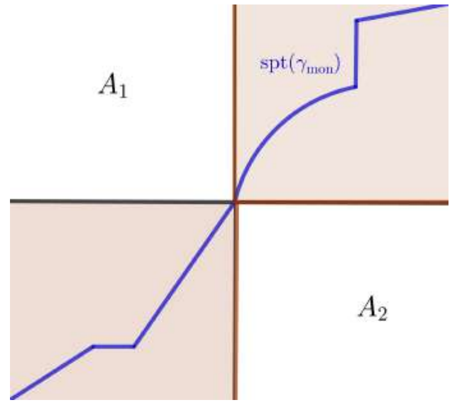

Figure 2.3: $\text{spt}(\gamma) \cap (A_1 \cup A_2) = \emptyset$, in Lemma 2.1.6.

the minimal interval I such that $\text{spt}(\gamma) \cap (\{x\} \times \mathbb{R}) \subset \{x\} \times I$, which might be reduced to a singleton. Notice that the hypothesis on $\text{spt}(\gamma)$ implies that the interior of all these intervals are disjoint. Otherwise there exist $x < x' \in \text{spt}(\mu)$ such that $I_x \cap I_{x'} \neq \emptyset$; let $z \in I_x \cap I_{x'}$, since $(x, z), (x', z) \in \text{spt}(\gamma)$, then we have that $z < z$, which is a contradiction. Thus, there is at most a countable quantity of points such that I_x is not a singleton; then the set

$$\{x \in \text{spt}(\mu) : \text{Card}(\{y : (x, y) \in \text{spt}(\gamma)\}) > 1\}$$

is countable, and hence is μ-negligible (thanks to the fact that μ is atomless). Therefore, for μ-a.e. $x \in \mathbb{R}$ there exists a unique $y = T(x)$ such that $(x, y) \in \text{spt}(\gamma)$, in other words, we can define μ-a.e. a map T such that γ is concentrated on the graph of T. Clearly this map will be monotone (by the monotone condition on $\text{spt}(\gamma)$). Finally, since $T_\# \mu = \nu$ (by Theorem 2.1.5) then we necessarily have that $T = T_{\text{mon}}$(up to μ-negligible sets). □

Now, we are ready to prove that the monotone transport map is optimal when $c(x, y) = h(y - x)$, with h a strictly convex function.

Theorem 2.1.7 (Optimality of the monotone map)**.** *Let $h : \mathbb{R} \to \mathbb{R}_+$ be a strictly convex function and $\mu, \nu \in \mathscr{P}(\mathbb{R})$ be probability measures. Consider the cost function $c(x, y) = h(y - x)$ and suppose that (KP) has finite value. Then, (KP) has a unique solution, which is given by γ_{mon}. If μ is atomless, then this optimal plan is induced by T_{mon}. When h is convex, then the same γ_{mon} is optimal, but no uniqueness is guaranteed. Moreover,*

$$\min(KP) = \int_0^1 h(F_\nu^{[-1]} - F_\mu^{[-1]}) d\mathscr{L}^1.$$

Proof. Let $\gamma \in \Pi(\mu, \nu)$ be an optimal transport plan. Then, by Theorem 1.2.25 we know that $\text{spt}(\gamma) \subset \Gamma$, for some c-CM set Γ. In particular, for every pair $(x, y), (x', y') \in \text{spt}(\gamma)$ we have that:

$$h(y - x) + h(y' - x') \leq h(y' - x) + h(y - x'). \tag{2.1.1}$$

We will use strict convexity of h to prove that inequality 2.1.1 implies a monotone behavior, that is, $x < x'$ implies $y \leq y'$; this will allow us to apply Lemma 2.1.6 and conclude the proof.

In order to prove $y \leq y'$ we argue by contradiction, suppose that $y > y'$ and let $a = y - x$, $b = y' - x'$ and $\delta = x' - x > 0$, then $b + \delta < a$. Moreover, if we let $t = \frac{\delta}{a-b}$, then $t \in (0,1)$ and

$$\begin{aligned} b + \delta &= (1-t)b + ta, \\ a - \delta &= tb + (1-t)a. \end{aligned}$$

From condition (2.1.1) we have that $h(a) + h(b) \leq h(b+\delta) + h(a-\delta)$. But h is strictly convex, so

$$\begin{aligned} h(a) + h(b) &\leq h(b+\delta) + h(a-\delta) \\ &< (1-t)h(b) + th(a) + (1-t)h(a) + th(b) \\ &= h(a) + h(b), \end{aligned}$$

which is a contradiction. Therefore, thanks to the condition $y \leq y'$ and Lemma 2.1.6, the statement follows in the strictly convex case. Now, we generalize the statement when h is convex. It is well known that every convex function can be bounded from below by an affine function $h(x) \geq (ax+b)_+$ (we are taking the positive part since $h \geq 0$). One can check that $f(x) = \sqrt{4 + (ax+b)^2} + \frac{1}{2}(ax+b)$ is strictly convex and satisfies $0 \leq f(x) \leq 1 + h(x)$. Then, for every $\epsilon > 0$ $h_\epsilon = h + \epsilon h$ is strictly convex and satisfies

$$h \leq h_\epsilon \leq (1+\epsilon)h + \epsilon.$$

Let us take the transport cost $c_\epsilon(x,y) := h_\epsilon(y-x)$; in this case we know that γ_{mon} is optimal for $\displaystyle\int_{\mathbb{R}\times\mathbb{R}} c_\epsilon(x,y)d\gamma(x,y)$, and hence:

$$\begin{aligned} \int_{\mathbb{R}\times\mathbb{R}} h(y-x)d\gamma_{\text{mon}}(x,y) \leq \int_{\mathbb{R}\times\mathbb{R}} h_\epsilon(y-x)d\gamma_{\text{mon}}(x,y) &\leq \int_{\mathbb{R}\times\mathbb{R}} h_\epsilon(y-x)d\gamma(x,y) \\ &\leq (1+\epsilon)\int_{\mathbb{R}\times\mathbb{R}} h(y-x)d\gamma(x,y) + \epsilon, \end{aligned}$$

for all $\gamma \in \Pi(\mu,\nu)$. Taking $\epsilon \to 0$, we get that γ_{mon} is also optimal for the cost c. Finally, consider the change of variable $t = F_\mu(x)$. Since $T_{\text{mon}}(x) = F_\nu^{[-1]}(F_\mu(x))$, then it follows

$$\min(\text{KP}) = \int_{\mathbb{R}} h(y-x)d\mu(x) = \int_0^1 h\left(F_\nu^{[-1]}(t) - F_\mu^{[-1]}(t)\right)dt.$$

□

We will give some easy examples where c is convex but not strictly convex and T_{mon} is not the unique optimal transport map.

Example 2.1.8 (Book shifting)**.** Consider the cost function $c(x,y) = |x-y|$, $\mu = \mathscr{L}^1 |_{[0,2]}$ and $\nu = \frac{1}{2}\mathscr{L}^1 |_{[1,3]}$. Then,

$$T_{\text{mon}}(x) = F_\nu^{[-1]}(F_\mu(x)) = \inf\left\{t \in \mathbb{R} : \frac{1}{2}\mathscr{L}^1 |_{[1,3]} ((-\infty,t]) \geq \frac{1}{2}\mathscr{L}^1 |_{[1,2]} (-\infty,x])\right\} = x+1,$$

is the monotone transport form μ onto ν. Its cost is

$$\int |T_{\text{mon}}(x) - x|d\mu = \frac{1}{2}\int_0^2 |x+1-x|dx = 1.$$

On the other hand, the transport map

$$\bar{T}(x) = \begin{cases} x+2 & \text{if } x \leq 1 \\ x & \text{if } x > 1, \end{cases}$$

also satisfies that $\bar{T}_{\#}\mu = \nu$ and $\int |\bar{T}(x) - x| d\mu = \frac{1}{2}\int_0^2 2dx = 1$. However, since $T_{\#}\mu = \nu$ implies $\int Txd\mu = \int yd\nu$, one sees that

$$\int |Tx - x| d\mu \geq \int Txd\mu - \int xd\mu = \int xd\nu - \int xd\mu = \frac{3}{2} - \frac{1}{2} = 1.$$

Hence, an optimal map should give a minimal cost of 1. Therefore, $\bar{T}$ is also optimal.

Example 2.1.9 (Linear cost)**.** Suppose that $c(x,y) = h(x-y)$, with $h : \mathbb{R}^d \to \mathbb{R}$ a linear function and let $\mu, \nu \in \mathscr{P}(\mathbb{R})$ compactly supported measures. Then, any optimal transport plan $\gamma \in \pi(\mu,\nu)$ is optimal and any transport map as well. Indeed, observe that

$$\int_{\mathbb{R}\times\mathbb{R}} h(x-y)d\gamma(x,y) = \int_{\mathbb{R}} h(x)d\mu(x) - \int_{\mathbb{R}} h(y)d\nu(y).$$

Hence, $\int_{\mathbb{R}\times\mathbb{R}} h(x-y)d\gamma(x,y)$ depends only on its marginals, which means that any transport plan γ is optimal.

Example 2.1.10 (Distance costs on the line)**.** Suppose that $c(x,y) = |x-y|$ and that $\mu, \nu \in \mathscr{P}(\mathbb{R})$ are such that $\sup\{\text{spt}(\mu)\} < \inf\{\text{spt}(\nu)\}$. Observe that for every $(x,y) \in \text{spt}(\mu) \times \text{spt}(\nu)$, we have $c(x,y) = |x-y| = y - x$. Thus, any transport plan $\gamma \in \Pi(\mu,\nu)$ is optimal.

2.2 L^1 Monge's transport problem

2.2.1 Secondary variational problem

From now on, we will focus on the original Monge transportation problem, that is, the transport problem with cost $c(x,y) = |x-y|$ and $\mu \ll \mathscr{L}^d$. In general, the strategy to prove this result will be to propose an alternative variational problem such that, any solution to this problem is also a solution to the L^1 Kantorovich's problem, which we know behaves as the monotone transport plan. Then, we prove that such a solution is induced, up to μ-negligible sets, by a transport map if the gradient of the kantorovich potential has Lipschitz regularity.

Before beginning a detailed analysis of this problem, we shall give some preliminary ideas.

Remark 2.2.1. Since any distance is symmetric, we will avoid the distinction between the $\bar{c}$-transform and c-transform.

If c is a distance on X, then we can give the following characterization result about c-concave functions.

Proposition 2.2.2. *If $c : X \times X \to \mathbb{R}$ is a distance, then the function $u : X \to \mathbb{R}$ is c-concave if and only if it is Lipschitz continuous, with $Lip(u) \leq 1$. We will denote by $Lip_1(X)$ the set of such functions. Moreover, for every $u \in Lip_1(X)$, we have $u^c = -u$.*

Proof. Let u be a c-concave function. Then, there exists $\chi : X \to \mathbb{R} \cup \{-\infty\}$ such that

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

Let us assume that $\chi(y) \neq -\infty$. Note that the function $x \mapsto c(x,y) - \chi(y)$ belongs to $\mathrm{Lip}_1(X)$, then we necessarily have $u \in \mathrm{Lip}_1(X)$ (since we are taking the infimum of Lipschitz functions). Conversely, take a function $u \in \mathrm{Lip}_1(X)$. We claim that

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

Indeed, we note that $u(x) \geq \inf_{y \in X} \{c(x,y) + u(y)\}$ (since one can take $y = x$). On the other hand, since $u \in \mathrm{Lip}_1(X)$ then $u(x) - u(y) \leq c(x,y)$, that is, $u(x) \leq c(x,y) + u(y)$ for all $y \in X$. Hence, the desired equality follows and shows that $u = (-u)^c$. $\square$

Notation 2.2.3. Given a cost function $c : X \times Y \to \mathbb{R}$, $\mu \in \mathscr{P}(X)$ and $\nu \in \mathscr{P}(Y)$, let us denote the minimum of (KP) by $\mathscr{J}_c(\mu,\nu)$, that is

$$\mathscr{J}_c(\mu,\nu) := \min\left\{ \int_{X \times X} c(x,y) d\gamma(x,y) : \gamma \in \Pi(\mu,\nu) \right\}.$$

Remark 2.2.4. As a consequence of the duality formula (Theorem 1.2.22) and Proposition 2.2.2, if $c(x,y) = d(x,y)$, with d a metric on X, we have the following duality formula known as the **Kantorovich-Rubinstein formula**:

$$\mathscr{J}_c(\mu,\nu) = \sup\left\{ \int_X u d(\mu - \nu) : u \in \mathrm{Lip}_1(X) \right\}.$$

Now, we will give an example where we show how to find an optimal transport map through the Kantorovich-Rubinstein formula.

Example 2.2.5 (Monge radial problem)**.** Let us consider the same setting as in Example 1.2.44 : we will find the optimal transport between $\mu = f \cdot \mathscr{L}^2$ and $\nu = g \cdot \mathscr{L}^2$, when the cost function is $c(x,y) = |x - y|$ and $f(x) = \frac{1}{\pi}\chi_{B(0,1)}(x)$, $g(x) = \frac{1}{8\pi}(4 - |x|^2)$. Applying the same reasoning of Example 1.2.44 we find the transport map $T(x) = 2\sqrt{1 - \sqrt{1 - |x|^2}}\frac{x}{|x|}$. Now, let us find the cost associated to this T, to do so, we must compute:

$$\int_{\mathbb{R}} |x - T(x)| d\mu(x) = \frac{1}{\pi}\int_0^{2\pi}\int_0^1 \left| t - 2\sqrt{1 - \sqrt{1 - t^2}} \right| dt d\theta = \int_0^1 \left| t - 2\sqrt{1 - \sqrt{1-t^2}} \right| t dt. \tag{2.2.1}$$

To simplify the right hand side of (2.2.1), we use the change of variable $t = \sin(y)$ and $dt = \cos(y)dy$. Then, $\sqrt{1 - t^2} = \sqrt{\cos^2(y)}$. Note that this substitution is invertible over $0 < y < \frac{\pi}{2}$ with inverse $y = \arcsin(t)$. This gives us the bounds $y = \arcsin(0) = 0$ and $y = \arcsin\left(\frac{\pi}{2}\right)$. Thus,

$$\int_{\mathbb{R}} |x - T(x)| d\mu(x) = 2\int_0^{\frac{\pi}{2}} \cos(y) \left| \sin(y) - \sqrt{1 - \sqrt{\cos^2(y)}} \right| dy$$

One can perform the computations to show that, $\int_0^{\frac{\pi}{2}} \cos(y) \left| \sin(y) - \sqrt{1 - \sqrt{\cos^2(y)}} \right| dy = \frac{1}{5}$, which means that

$$\int_{\mathbb{R}} |x - T(x)| d\mu(x) = \frac{2}{5}.$$

Let us prove now that the transport map T is optimal. Indeed, consider the potential $u(x) = -|x|$, which clearly belongs to $\text{Lip}_1(\mathbb{R}^2)$. Then, by the duality formula we have:

$$\begin{aligned}
\mathscr{J}_c(\mu,\nu) &= \max\left\{\int_{\mathbb{R}} u(x)d(\mu-\nu) : u \in \text{Lip}_1(\mathbb{R})\right\} \\
&\geq \int_{\mathbb{R}} -|x|d(\mu-\nu)(x) \\
&= \int_{\mathbb{R}} -|x|d\mu(x) + \int_{\mathbb{R}} |x|d\nu(x) \\
&= \frac{1}{\pi}\int_0^1 -t^2 2\pi dt + \frac{1}{8\pi}\int_0^2 t^2(4-t^2)2\pi dt \\
&= -\frac{2}{3} + \int_0^2 t^2 dt - \frac{1}{4}\int_0^2 t^4 dt \\
&= \frac{2}{5}.
\end{aligned}$$

Since T achieves this lower bound, then it must be optimal.

Finally, we prove the pending claim from Example 1.1.12, that is, $\mathscr{J}_c(\mu,\nu)$ equals to the total variation $|\mu-\nu|$.

Example 2.2.6 (Total variation)**.** Let us consider the lower semi-continuous cost function

$$c(x,y) = \begin{cases} 0, & \text{if } \ x = y \\ 2, & \text{if } \ x \neq y. \end{cases}$$

We easily see that c is a distance on X, and any function $u : X \to \mathbb{R}$ such that $|u(x)| \leq 1$ belongs to $\text{Lip}_1(X)$ with respect to the metric c, since $x \neq y$ implies that, $|u(x) - u(y)| \leq |u(x)| + |u(y)| \leq 2 = c(x,y)$. Let $\mu - \nu = (\mu-\nu)_+ - (\mu-\nu)_-$ be the Jordan decomposition, with $(\mu-\nu)_\pm \in \mathscr{M}(X)$ and $(\mu-\nu)_+ \perp (\mu-\nu)_-$. Since, μ, ν are probability measures we have $(\mu-\nu)_+(X) = (\mu-\nu)_-(X)$. On the other hand, it is well known that

$$\sup_{0\leq u\leq 1}\int_X ud(\mu-\nu) = (\mu-\nu)_+ \ \text{ and } \ \sup_{-1\leq u\leq 0}\int_X ud(\mu-\nu) = (\mu-\nu)_-,$$

if μ,ν are probability measures. Then, we have

$$\sup_{|u|\leq 1}\int_X ud(\mu-\nu) = 2(\mu-\nu)_+(X) = 2(\mu-\nu)_-(X) = |\mu-\nu|.$$

Hence, by the duality formula we get that

$$\begin{aligned}
\mathscr{J}_c(\mu,\nu) = \sup\left\{\int_X vd(\mu-\nu) : v \in \text{Lip}_1(X)\right\} \geq \sup_{0\leq u\leq 1}\int_X ud(\mu-\nu) &= 2\int_X d(\mu-\nu)_+ \\
&= \int_X d|\mu-\nu|.
\end{aligned}$$

For the opposite bound, fix a point $x_0 \neq x$ and let u be a Kantorovich potential. Then

$$\begin{aligned}
\mathscr{J}(\mu,\nu) = \int_X ud(\mu-\nu) = \int_X [u(x)-u(x_0)]d(\mu-\nu) \leq \int_X s(x,x_0)d(\mu-\nu) &\leq 2\int_X d(\mu-\nu)_+ \\
&= \int_X d|\mu-\nu|.
\end{aligned}$$

Therefore, we get that $\mathscr{J}_c(\mu,\nu) = \displaystyle\int_X d|\mu-\nu|$.

As we saw in Examples 2.1.9 and 2.1.10, the optimal transport plan γ whit cost $c(x,y) = |x-y|$ is not unique necessarily unique. To fix this issue, we need to take a special optimizer γ and prove that it is induced by a transport map. For simplicity, we will suppose that X is a domain $\Omega \subset \mathbb{R}^d$.

Definition 2.2.7. *Let us define $O(\mu,\nu)$ as the set of optimal transport plans for the cost $|x-y|$. To simplify notation, we define $K_p(\gamma) := \int_{\Omega\times\Omega} |x-y|^p d\gamma(x,y)$ with $\gamma \in \mathscr{P}(\Omega\times\Omega)$, and m_p its minimal value on $\Pi(\mu,\nu)$. Then,*

$$O(\mu,\nu) := \operatorname{argmin}_{\gamma\in\Pi(\mu,\nu)} K_1(\gamma) = \{\gamma \in \Pi(\mu,\nu) : K_1(\gamma) \leq m_1\}.$$

Note that $O(\mu,\nu)$ is a closed subset (with respect to the weak convergence) of $\Pi(\mu,\nu)$, which is compact. Indeed, let $\{\gamma_n\}_{n\geq 1} \subset O(\mu,\nu)$ a sequence such that $\gamma_n \rightharpoonup \gamma$. This means that $\int_{\Omega\times\Omega} \phi d\gamma_n \to \int_{\Omega\times\Omega} \phi d\gamma$ for all $\phi \in C_b(\Omega\times\Omega)$. In particular, $\int_{\Omega\times\Omega} c d\gamma_n \to \int_{\Omega\times\Omega} c d\gamma$. Since $K_1(\gamma_n) \leq m_1$ for all $n \in \mathbb{N}$ and K_1 is lower semi-continuous, we immediately obtain $K_1(\gamma) \leq m_1$. Thus, $O(\mu,\nu)$ is $*$-closed. Compactness of $O(\mu,\nu)$ follows directly, since $\Pi(\mu,\nu)$ is $*$-compact.

Remark 2.2.8. Let u be a Kantorovich potential for the transport between μ and ν with cost $c(x,y) = |x-y|$. Then, $\gamma \in O(\mu,\nu)$ if and only if

$$\operatorname{spt}(\gamma) \subset \{(x,y) \in \Omega\times\Omega : u(x) - u(y) = |x-y|\}.$$

Indeed, optimality of γ implies

$$\int_{\Omega\times\Omega} (u(x) - u(y)) d\gamma = \int_{\Omega\times\Omega} |x-y| d\gamma,$$

then $u(x) - u(y) = |x-y|$ γ-a.e. On the other hand, if the equality holds on $\operatorname{spt}(\gamma)$ implies that $\int_{\Omega\times\Omega} (u(x)-u(y)) d\gamma = K_1(\gamma)$. Since u is a Kantorovich potential we have

$$m_1 = \max\left\{\int_{\Omega\times\Omega} u d(\mu-\nu) : u \in \operatorname{Lip}_1(\Omega)\right\} = K_1(\gamma).$$

Thus, $\gamma \in O(\mu,\nu)$.

Problem 2.2.9 (Secondary variational)**.** *Keeping the same notation as before, let us define the the **secondary variational** problem as*

$$\inf\{K_2(\gamma) : \gamma \in O(\mu,\nu)\},$$

which has a solution $\bar{\gamma}$ since K_2 is continuous and $O(\mu,\nu)$ is $*$-compact.

Now, the aim is to characterize the transport plan $\bar{\gamma}$ and prove that it is induced by a transport map. By Remark 2.2.8, we know that the condition $\gamma \in O(\mu, \nu)$ can be rewritten as a condition on the support of γ. Then $\bar{\gamma}$ also solves:

$$\min \left\{ \int_{\Omega \times \Omega} c d\gamma : \gamma \in \Pi(\mu, \nu) \right\}, \text{ with } c(x, y) = \begin{cases} |x-y|^2, & \text{if } u(x) - u(y) = |x-y| \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.2.2}$$

Then, Problem 2.2.9 and (2.2.2) are equivalent minimization problems.

Remark 2.2.10. Let $c : \Omega \times \Omega \to \mathbb{R} \cup \{+\infty\}$ be a cost function defined by

$$c(x, y) = \begin{cases} |x-y|^2, & \text{if } (x, y) \in A \\ +\infty, & \text{otherwise,} \end{cases} \tag{2.2.3}$$

with $A \subset \Omega \times \Omega$ a closed set. Since c is lower semi-continuous on $\Omega \times \Omega$, then by theorem 1.2.25 we must have that $\mathrm{spt}(\gamma)$ is concentrated on a c-cyclically monotone set Γ. Hence, any solution for the secondary variational problem $\bar{\gamma}$ is also concentrated on Γ.

The following lemma states that the set Γ from last remark satisfies a certain monotonicity condition.

Lemma 2.2.11. *Suppose that $\Gamma \subset \Omega \times \Omega$ is c-CM for the cost function defined in (2.2.3). Then, $(x_1, y_1), (x_2, y_2) \in \Gamma$ and $(x_1, y_1), (x_2, y_2) \in A$ implies that $(x_1 - x_2) \cdot (y_1 - y_2) \geq 0$ (monotonicity condition).*

Proof. Since Γ is a c-CM set, for any pair $(x_1, y_1), (x_2, y_2) \in \Gamma$ we have that $c(x_1, y_1) + c(x_2, y_2) \leq c(x_1, y_2) + c(x_2, y_1)$. Let us suppose that $(x_1, y_2), (x_2, y_1) \in \mathrm{spt}(\gamma)$. Observe that the previous inequality is equivalent to $|x_1 - y_1|^2 + |x_2 - y_2|^2 \leq |x_1 - y_2|^2 + |x_2 - y_1|^2$. Thus, if we expand the squares we get

$$\begin{aligned} & -2(x_1 \cdot y_1) - 2(x_2 \cdot y_2) \leq -2(x_1 \cdot y_2) - 2(x_2 \cdot y_1) \\ & \Leftrightarrow (x_1 - x_2) \cdot y_2 \leq (x_1 - x_2) \cdot y_1 \\ & \Leftrightarrow (x_1 - x_2) \cdot (y_1 - y_2) \geq 0, \end{aligned}$$

as desired. □

2.2.2 Transport rays and Ambrosio-Sudakov's Theorem

In this section we will study geometric properties of transport rays, and we prove that an optimal plan $\bar{\gamma}$ to the secondary variational problem behaves, on each transport ray, as the monotone transport plan γ_{mon}. We will also study some Lipschitz regularity properties of the Kantorovich potential associated to $\bar{\gamma}$. Finally we present Ambrosio-Sudakov's Theorem, which guarantees the existence of an optimal transport map for Monge's transportation problem.

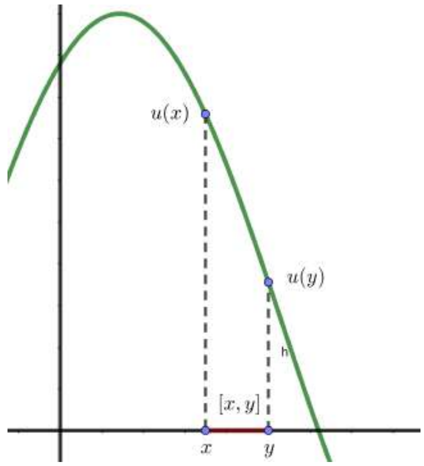

Figure 2.4: If $u : \mathbb{R} \to \mathbb{R}$, then we can easily represent a transport ray $[x, y]$ and its elements $y \in \mathrm{Trans}^{(b+)}(u)$ and $x \in \mathrm{Trans}^{(b-)}(u)$

Definition 2.2.12 (Transport ray)**.** *If a segment $[x, y]$ is maximal with respect to the set inclusion among all segments satisfying the condition $u(x) - u(y) = |x - y|$, we say that $[x, y]$ is a nontrivial **transport ray**. The corresponding open segment $]x, y[$ is called the interior of a transport ray and the points x, y its boundary points. We will call direction of a transport ray the unit vector $\frac{x-y}{|x-y|}$. We now define the following sets:*

- *Let Trans(u) be the union of all nondegenerate transport rays, in other words*

$$Trans(u) := \bigcup \{[x, y] : [x, y] \text{ is a nontrivial transport ray}\}.$$

- *Let Trans$^{(b)}(u)$ be the union of boundary points of all segments in Trans(u),*
- *Let Trans$^{(b+)}(u)$ be the set of upper boundary points of nondegenerate transport rays (those points where u is minimal on the transport ray, say the points y)*
- *Let Trans$^{(b-)}(u)$ be the set of lower boundary points of nondegenerate transport rays (those points where u is maximal on the transport ray, say the points x)*

The following lemma collects some geometric properties of a Kantorovich potential $u \in \mathrm{Lip}_1(\Omega)$.

Lemma 2.2.13 (Differentiability of Kantorovich potentials)**.** *Let $[x, y] \in$ Trans(u) be a transport ray. Then u is affine on $[x, y]$. Moreover, if $z \in]x, y[$ (the open segment), then u is differentiable at z and $\nabla u(z) = e := \frac{x-y}{|x-y|}$.*

Sketch of proof. Geometrically, we can see from figure 2.5 that u must be differentiable on $]x, y[$ since the graph of u is trapped between two cones ($|x - y|$ at x and y). Now, let us

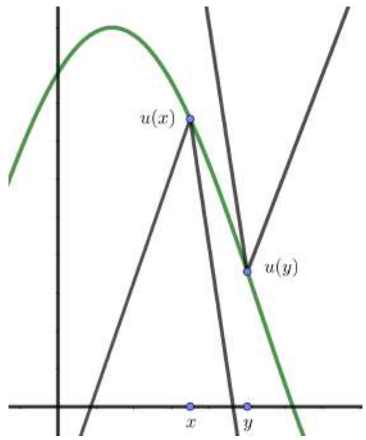

Figure 2.5: Geometric representation of the differentiability of u in Lemma 2.2.13

give a short proof of this fact without technical details. Indeed, we already know that u is an affine function on $[x, y]$, in other words,

$$u(x) - u(x + t(y - x)) = t|y - x|, \quad \forall t \in [0, 1].$$

This implies that, if we set $e = \frac{x-y}{|x-y|}$, then the partial derivative along e is equal to -1 for any $z \in]x, y[$. Let θ be a unit vector such that $e \cdot \theta = 0$, then

$$\begin{aligned} u(z + h\theta) - u(z) &= u(z + h\theta) - u(z + \sqrt{|h|}e) + u(x + \sqrt{|h|}e) - u(z) \\ &= |h\theta - \sqrt{|h|}e| - \sqrt{|h|}|e| \\ &\leq \sqrt{|h|^2 + |h|} - \sqrt{|h|} \\ &= O(|h|^{\frac{3}{2}}) = o(|h|). \end{aligned}$$

Analogously, we can check that $u(z + h\theta) - u(z) \geq o(|h|)$. Hence $u(z + h\theta) - u(z) = o(|h|)$, which means that u is differentiable at z, and $\nabla u(z) = e$. □

Proposition 2.2.14. *(a) Let $[x_1, y_1], [x_2, y_2]$ be two different transport rays. Then, they only can meet at a point z which is a boundary point for both of them, and in such a case, u is not differentiable at z.*

(b) In particular, if one removes the μ-negligible set

$$S(u) := \{z \in \Omega : \nabla u(z) \text{ does not exists}\},$$

the transport rays are disjoint.

Sketch of proof. **(a):** we observe that, if $z \in [x_1, y_1]$ then we must have $e = \nabla u(z)$. Indeed, let $t \ll 1$ such that $z + te \in [x_1, y_1]$ (we may intersect a sufficiently small ball centered at z with $[x_1, y_1]$). By Lemma 2.2.13, we also have

$$t = |z + t - z| = u(z + te) - u(z) = t\nabla u(z) \cdot e + o(t) \Longrightarrow 1 = \nabla u(z) \cdot e + \frac{o(t)}{t},$$

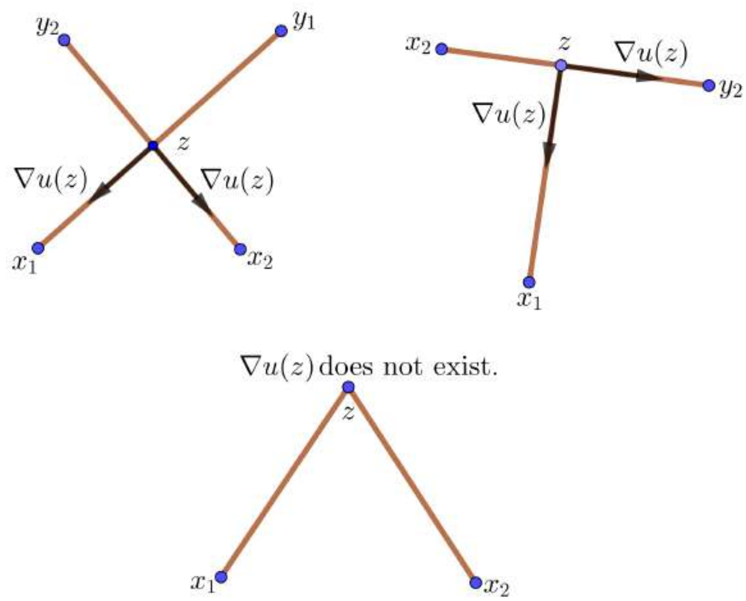

Figure 2.6: $\nabla u(z)$ in Proposition 2.2.14.

which means that $\nabla u(z) \cdot e = 1$. Since $|\nabla u(z)| \leq 1$, then we must have $\nabla u(z) = e$. Instead of giving a proof of the possibilities of $\nabla u(z)$, with $z \in s_{x_1,y_1} \cap s_{x_2,y_2}$, we give a geometric argument in picture 2.6.

(b): This is just a simple consequence of (a) and Rademacher's Theorem. $\square$

Let us fix a transport plan $\bar{\gamma}$ which is optimal for the secondary variational problem, we will try to prove that it is induced by a transport map. To do so, we use the fact that $\bar{\gamma}$ is concentrated on a c-CM set Γ (since the cost function is given as in (2.2.3)) and see how this interacts with transport rays. Thus, we may assume that

$$\Gamma \subset A = \{(x, y) \in \Omega \times \Omega : u(x) - u(y) = |x - y|\} .$$

Let $[x, y]$ be a transport ray, then we can define an **order relation** on $[x, y]$ through:

$$x \leq x' \iff u(x') \leq u(x).$$

Remark 2.2.15. For any transport ray $s \in \operatorname{Trans}(u)$, we have $u(x') - u(y') = |x' - y'|$ whenever $x', y' \in s$ and $x' \leq y'$ (for the order relationship on s).

Finally, the following lemma gives us the relation between an optimal plan $\bar{\gamma}$ of the secondary variational problem and the monotone transport plan. We shall prove that $\bar{\gamma}$ has a monotone behavior on each transport ray as the monotone transport plan that we studied in Section 2.1 (see Lemma 2.1.6).

Lemma 2.2.16. *Suppose that x_1, x_2, y_1 and y_2 are all points in a transport ray $[x, y]$ and $(x_1, y_1), (x_2, y_2) \in \Gamma$. Then, if $x_1 < x_2$, we also have $y_1 \leq y_2$.*

Proof. Since $\Gamma \subset A$, then $|x_j - y_j| = u(x_j) - u(y_j)$ for $j = 1, 2$; so we get that $x_1 \leq y_1$ and $x_2 \leq y_2$. On the other hand, thanks to the order relation on $[x, y]$ and Remark 2.2.15 we have that $(x_1, y_2), (x_2, y_1) \in A$. Also, since Γ is c-CM, then $(x_1 - x_2) \cdot (y_2 - y_1) \geq 0$ (by Lemma 2.2.11), with $x_1 - x_2$ and $y_1 - y_2$ two vectors on the ray $[x, y]$. Since these vectors are parallel to e, then y_1 and y_2 are ordered as x_1 and x_2 are. Hence, $y_1 \leq y_2$. □

Now, we shall prove an important proposition which encompasses all the interaction between Γ and the transport rays that we have observed so far.

Proposition 2.2.17. *The optimal transport plan $\bar{\gamma}$ is concentrated on a set Γ with the following properties:*

(a) If $(x, y) \in \Gamma$, then

- *either $x \in S(u)$, which is Lebesgue-negligible,*
- *or $x \notin Trans(u)$, which means that x does not belong to a nondegenerate transport ray. In this case, we necessarily have that $y = x$.*
- *or $x \in Trans^{(b+)}(u) \setminus S(u)$. In such a case, we must have $y = x$ since x is contained in a unique transport ray s and it cannot have other images $y \in s$, due to the order relation on s.*
- *or $x \in Trans(u) \setminus (Trans^{(b+)}(u) \cup S(u))$. Here, we observe that $y \in s$, with s the unique transport ray s which contains the point x.*

(b) On each transport ray s, $\Gamma \cap (s \times s)$ is contained in the graph of a monotone increasing multivalued function.

(c) On each transport ray s, the set

$$N_s = \Big\{ x \in s \setminus Trans^{(b+)}(u) : Card(\{y : (x, y) \in \Gamma\}) > 1 \Big\}$$

is countable.

Proof. **(a):** This statement follows from the properties of transport rays and Γ studied in Proposition 2.2.14, Lemma 2.2.15 and Lemma 2.2.16.

(b), (c): Notice that if we argue as in Lemma 2.1.6, then the following holds. If s is a segment and $\Gamma' \subset (s \times s)$ is such that: $(x_1, y_1), (x_2, y_2) \in \Gamma'$, with $x_1 < x_2$ implies, $y_1 \leq y_2$ for the order relation on s. Then, Γ' is contained in the graph of a monotone increasing multi-valued function such that, every point might be sent to either a point or a segment. From Lemma 2.1.6, we also observe that the interiors of these segments are disjoint, and there is at most a countable number of points where the image is not a singleton. Thus, $\{x \in s : \text{Card}(\{y : (x, y) \in \Gamma'\}) > 1\}$ is countable. This means that, up to countable sets, Γ' is contained in the graph of a monotone single-valued map T_s.

Hence, by Lemma 2.2.16 we can apply the previous reasoning to $\Gamma \cap (s \times s)$, with s a transport ray. Then, $\Gamma \cap (s \times s)$ is concentrated in the graph of a monotone increasing multi-valued function, which is a single-valued map T_s (up to a countable) and each set N_s is at most countable. □

Remark 2.2.18. Note that the set $\Gamma \setminus \bigcup_s \Gamma \cap (s \times s)$, with s a transport ray, is μ-negligible. Hence, if $\mu(\cup_s N_s) = 0$, by Proposition 2.2.17 we would have that the optimal transport plan $\bar{\gamma}$ is induced by a transport map T (up to μ-negligible sets) as long as $\mu \ll \mathscr{L}^d$.

Remark 2.2.19. If we assume that $\mu \ll \mathscr{L}^d$, then all the measures $\{\mu_s\}$ given by the disintegration of μ along the transport rays s are atomless. Now, we consider the measures $\{\nu_s\}$ given by the disintegration of ν over the rays s. We know that on each transport ray s, $\bar{\gamma}$ is induced by a monotone map T_s (up to countable sets) such that $T_{s\#}\mu_s = \nu_s$ (thanks to Proposition 2.2.17). Thus, roughly speaking, the transport map T is obtained by gluing all the monotone maps (in a measurable way) on every transport ray.

Now, it is important to note that with the last remark a very natural question arises: under which circumstances $\mu(\cup_s N_s) = 0$? In order to answer this question, let us introduce a property for negligibility which will guarantee that $\mu(\cup_s N_s) = 0$, and therefore $\bar{\gamma}$ will be induced by a transport map.

Definition 2.2.20. *We say that **property** N holds for a given Kantorovich potential u, if for every subset $B \subset \Omega$ such that*

- *$B \subset \mathit{Trans}(u) \setminus \mathit{Trans}^{(b+)}(u)$,*
- *$B \cap s$ is at most countable for every transport ray s,*

we have that $\mathscr{L}^d(B) = 0$.

The following lemma states that property N holds, if ∇u has Lipschitz regularity on each transport ray.

Lemma 2.2.21. *Property N holds if ∇u is Lipschitz continuous. Moreover, if there exists a countable family of sets $\{E_h\}_{h=1}^{\infty}$ such that $\nabla u\,|_{E_h}$ is Lipschitz continuous and*

$$\mathscr{L}^d(\mathit{Trans}(u) \setminus \cup_{h=1}^{\infty} E_h) = 0.$$

Then, property N also holds.

Proof. First, we assume that ∇u is Lipschitz. Let $\{Y_q\}_{q\in\mathbb{Q}}$ be the collection of all hyperplanes parallel to the first $d-1$ coordinate axes and with rational entries on the last coordinate ($Y_q = \{(x_1, \ldots, x_{d-1}, q) : x_j \in \mathbb{R},\ \forall j \in \{1, 2, \ldots, d-1\}\}$). Consider a set $B \subset \Omega$ which satisfies the hypothesis in Definition 2.2.20. Since $B \subset \mathrm{Trans}(u) \setminus \mathrm{Trans}^{(b+)}(u)$, then the points of B belong to nondegenerate transport rays (those with positive length). Thus, for every $z \in B_1 = B \setminus \{x \in s : s \text{ is parallel to each } Y_q\}$ there is a transport ray s such that $z \in s$ and $s \cap Y_q \neq \emptyset$ for some $q \in \mathbb{Q}$, this means that every point of B_1 belongs to a transport ray that meets at least one hyperplane at exactly one point. Analogously, if we consider the collection of all hyperplanes $\{H_q\}_{q\in\mathbb{Q}}$ parallel to the last $d-1$ coordinate axes with rational coefficient on the first coordinate, then we have similar conditions for the set $B_2 = B \setminus \{x \in s : s \text{ is parallel to each } H_q\}$. Thus, without loss of generality, we can suppose that every point of B belongs to a transport ray that meets at least one hyperplane Y_q at exactly

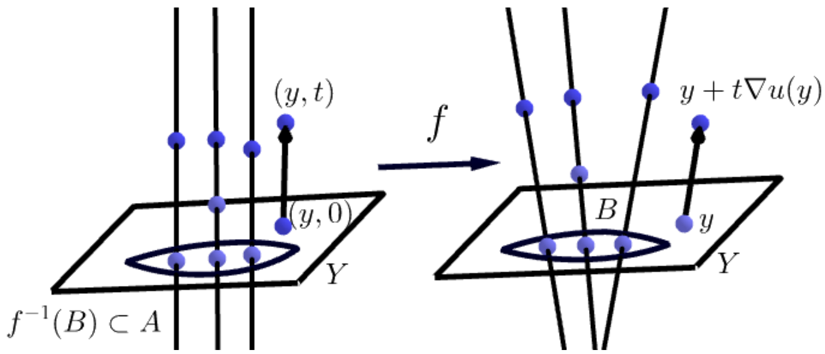

Figure 2.7: Rectification map f in Lemma 2.2.21

one interior point (since $B = B_1 \cup B_2$, and if we prove $\mathscr{L}^d(B_j) = 0$ for $j = 1, 2$, then we would have that $\mathscr{L}^d(B) = 0$).

On the other hand, since $B \cap s$ is at most countable for every transport ray s and the collection $\{Y_q\}$ is countable as well then, up to countable unions, we can suppose that $B \subset S_Y$, where S_Y is the collection of transport rays all meeting the same hyperplane Y. Moreover, by Proposition 2.2.14 we also may assume that B does not contain boundary points of two different transport rays.

Now, let us consider the hyperplane Y (the one which satisfies $B \subset S_Y$) and let us define $f : Y \times \mathbb{R} \to \mathbb{R}^d$ given by $f(y,t) = y + t\nabla u(y)$. This map is well defined on the set

$$A := \{(y,t) \in Y \times \mathbb{R} : y \in \mathrm{Int}(s),\ \text{for some } s \in S_Y\},$$

since u is differentiable at y, and $y + t\nabla u(y)$ belongs to the interior of the same transport ray (see figure 2.7). We also have that $f\mid_A$ is injective. Indeed, if we suppose that $y + t\nabla u(y) = f(y,t) = f(y',t') = y' + t'\nabla u(y')$, then two different transport rays cross at this points, but this is only possible if $y = y'$ (since y, y' cannot be boundary points). Thus, $\nabla u(y)$ and $\nabla(y')$ must coincide as well. Therefore, $(y,t) = (y',t')$. Since B is contained in the image of f, then $f : B' = f^{-1}(B) \to B$ is a bijection. We also have that f is Lipschitz continuous. Indeed, by our hypothesis ∇u is Lipschitz, then

$$\begin{aligned}|f(y,t) - f(y',t')| &= |y + t\nabla u(y) - (y' + t'\nabla u(y'))| \\ &\le |y - y'| + |t\nabla u(y) - t'\nabla u(y')| \\ &\le C(|y - y'| + |t - t'|).\end{aligned}$$

Note that B' is a subset of $Y \times \mathbb{R}$ containing at most countably many points on every line $\{y\} \times \mathbb{R}$. Then, we may apply Fubini's Theorem to get

$$\mathscr{L}^d(B') = \int_{\mathbb{R}^d} \chi_{B'}(x)\,d\mathscr{L}^d(x) = \int_{\mathbb{R}} \mathscr{H}^{d-1}(B')\,d\mathscr{L}(y) = 0,$$

which implies that $\mathscr{L}^d(B) = \mathscr{L}^d(f(B')) \le \mathrm{Lip}(f)^d \mathscr{L}^d(B') = 0$. Thus, $\mathscr{L}^d(B) = 0$. This shows that property N holds.

It is clear now that property N is also true when $\nabla u\mid_{E_h}$ is Lipschitz continuous, since one may apply the same arguments on each set $B \cap E_h$ to obtain that $\mathscr{L}^d(B \cap E_h) = 0$, then we deduce

$$\mathscr{L}^d(B) = \mathscr{L}^d\left(\cup_{h=1}^{\infty} B \cap E_h\right) = 0.$$

$\square$

Definition 2.2.22. *A function $f : \Omega \to \mathbb{R}^d$ is said to be* ***countably Lipschitz*** *if there exists a countable family of sets $\{E_h\}_{h=1}^{\infty}$, such that $f\mid_{E_h}$ is Lipschitz continuous and*

$$\mathscr{L}^d(\Omega \setminus \cup_{h=1}^{\infty} E_h) = 0.$$

Remark 2.2.23. Thanks to Lemma 2.2.21, if we are able to prove that ∇u is countably Lipschitz, then we would have that property N holds for our Kantorovich potential u, and therefore $\cup_s N_s$ is μ-negligible.

Now, our goal is to prove that ∇u is countably Lipschitz on $\text{Trans}(u)$. The following theorem from real analysis is a well-known result that allows us to extend Lipschitz functions defined in any subset of $\mathbb{R}^d$.

Theorem 2.2.24 (Kirszbraun)**.** *If $\Omega \subset \mathbb{R}^d$, then every L-Lipschitz function $f : \Omega \to \mathbb{R}$ can be extended to an L-Lipschitz function $\bar{f} : \mathbb{R}^d \to \mathbb{R}$.*

A function $f : \Omega \to \mathbb{R}$ is said to be **λ-convex** if $x \mapsto f(x) - \frac{\lambda}{2}|x|^2$ is convex and λ-concave if $x \mapsto f(x) + \frac{\lambda}{2}|x|^2$ is concave. **λ-convex** functions for $\lambda > 0$ are strictly convex, and for $\lambda < 0$, they just have second derivative (where it exists) which are bounded from below. Function which are λ-convex or λ-concave for some values λ are also called semi-convex or semi-concave functions.

Now we want to prove that ∇u is countably Lipschitz. But first, we will show that u coincides with some λ-concave function on a sequence of sets covering everything but the points in $\text{Trans}(u)^{(b+)}$. We do that via the following result.

Theorem 2.2.25. *There exist some sequence $\{E_h\}_{h\geq 1}$ such that:*

(a) $\cup_{h=1}^{\infty} E_h = \mathit{Trans}(u) \setminus \mathit{Trans}^{(b+)}(u)$.

(b) On each E_h the function u is the restriction of a λ-concave function, for a value $\lambda(h)$.

Proof. We first prove (a): Let us define the sets:

$$E_h := \left\{ x \in \text{Trans}(u) : \exists\ z \in \text{Trans}(u) \text{ such that, } u(x) - u(z) = |x - z| > \frac{1}{h} \right\},$$

that is, E_h is the set of points in the transport rays which are at least at distance $\frac{1}{h}$ from the upper boundary of its respective ray (see figure 2.8). We note that $E_h \subset \text{Trans}(u) \setminus \text{Trans}(u)^{(b+)}$, since E_h does not contain boundary points of the rays. Then, $\cup_{h=1}^{\infty} E_h \subset \text{Trans}(u) \setminus \text{Trans}(u)^{(b+)}$. We easily obtain the other contention. Indeed, if one takes $x \in \text{Trans}(u) \setminus \text{Trans}(u)^{(b+)}$, then $x \in s_{a,b}$ for some transport ray $s_{a,b}$. In this way u is differentiable at x, and $z = x + t\nabla u(x) \in s_{a,b}$ for a sufficiently small t and $u(x) - u(z) = |x - z| = t|\nabla u(x)| = t$. Hence, if we set $h \in \mathbb{N}$ such that $t > \frac{1}{h}$, then $x \in E_h$. We have shown then

$$\cup_{h=1}^{\infty} E_h = \text{Trans}(u) \setminus \text{Trans}^{(b+)}(u)$$

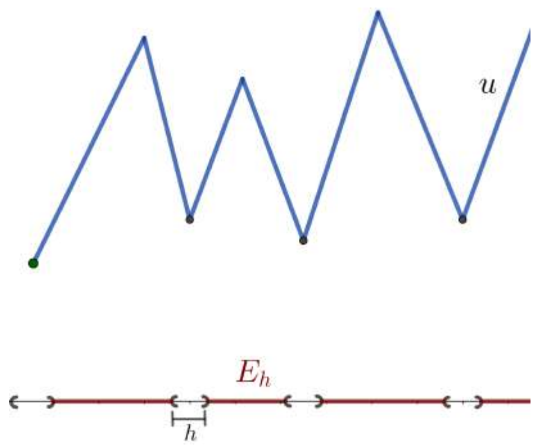

Figure 2.8: The sets E_h defined in Theorem

(b): Let $c_h : \mathbb{R}^d \to \mathbb{R}$ be a fixed function such that:

- $c_h \in C^2(\mathbb{R}^d)$, $|\nabla c_h(z)| \leq 1$,
- $c_h(z) \geq |z|$ for all $z \in \mathbb{R}^d$, and $c_h(z) = |z|$ for all $z \notin B\left(0, \frac{1}{h}\right)$.

Let us set $\lambda_h := -\|D^2 c_h\|_{L^\infty}$. We note that, if $x \in E_h$ then the following inequality holds:

$$u(x) = \inf_{y \in \mathbb{R}^d} \{|x-y| + u(y)\} \leq u_h(x) := \inf_{y \in \mathbb{R}^d} \{c_h(x-y) + u(y)\},$$

since we have the condition $|z| \leq c_h(z)$. Moreover,

$$u_h(x) \leq \inf_{y \notin B\left(0, \frac{1}{h}\right)} \{|x-y| + u(y)\} = u(x),$$

where the inequality is justified thanks to the fact that $c_h(z) = |z|$ for all $z \notin B\left(0, \frac{1}{h}\right)$, and the equality holds by definition of E_h. Thus, we have that $u(x) = u_h(x)$ for all $x \in E_h$.

Now, we shall prove that u_h is λ_h-concave. Indeed, notice that $c_h - \frac{\lambda_h}{2}$ is concave since the determinant of $D^2\left(c_h(x) - \frac{\|D^2 c_h\|_{L^\infty}}{2}|x|^2\right)$ turns out to be negative. Then, we have

$$\begin{aligned} u_h(x) + \frac{\lambda_h}{2}|x|^2 &= \inf_{y \in \mathbb{R}^d} \left\{c_h(x-y) + u(y) + \frac{\lambda_h}{2}|x|^2\right\} \\ &= \inf_{y \in \mathbb{R}^d} \left\{c_h(x-y) + \frac{\lambda_h}{2}|x-y|^2 + \lambda_h(x \cdot y) - \frac{\lambda_h}{2}|y|^2 + u(y)\right\}. \end{aligned}$$

Thus, we have written $u_h(x)$ as the infimum of concave functions, since $c_h(x-y) + \frac{\lambda_h}{2}|x-y|^2$ is concave, $x \cdot y$ is linear and $\frac{\lambda_h}{2}|y|^2 + u(y)$ is constant with respect to x. Therefore, $u_h(x)$ is λ_h-concave. □

The previous theorem allowed us to replace the function u with some functions u_h which share the same regularity of a convex function. However, this is not enough, since convex functions in general are not differentiable. A countable decomposition is needed and the following theorem give us what we need (see [2], Theorem 5.34).

Theorem 2.2.26. *If f is a convex function, then ∇f is countably Lipschitz.*

Now, we are ready to prove that ∇u has the desired property.

Proposition 2.2.27. *If u is a Kantorovich potential, then $\nabla u : \mathrm{Trans}(u) \to \mathbb{R}^d$ is countably Lipschitz.*

Proof. By Theorem 2.2.25 we have that, $\cup_{h=1}^{\infty} E_h = \mathrm{Trans}(u) \setminus \mathrm{Trans}^{(b+)}(u)$ and $u \mid_{E_h}$ is λ-concave for some value λ. Note that the countable Lipschitz regularity of Theorem 2.2.27 is also true for the gradient of λ-concave functions. This implies that each ∇u_h is countably Lipschitz and, by countable unions, ∇u is countably Lipschitz in $\mathrm{Trans}(u) \setminus \mathrm{Trans}^{(b+)}(u)$. $\square$

Finally, we obtain the main result of this chapter.

Theorem 2.2.28 (Ambrosio-Sudakov)**.** *Let $\mu, \nu \in \mathscr{P}(\Omega)$, such that $\mu \ll \mathscr{L}^d$. Then, the secondary variational problem admits a unique solution $\bar{\gamma} \in \Pi(\mu, \nu)$, which is induced by a transport map T, monotone nondecreasing on every transport ray.*

Proof. **Existence:** Let $\bar{\gamma} \in \Pi(\mu, \nu)$ be a solution of the secondary variational problem, and let u be the Kantorovich potential associated to $\bar{\gamma}$. Note that Proposition 2.2.26, together with Lemma 2.2.21 guarantees that Property N holds, which means that $\mu(\cup_{s=1}^{\infty} N_s) = 0$, since the set $\cup_{h=1}^{\infty} N_s$, with

$$N_s = \left\{ x \in s \setminus \mathrm{Trans}^{(b+)}(u) : \mathrm{Card}(\{y : (x, y) \in \Gamma\}) > 1 \right\}$$

satisfies the conditions from Definition 2.2.20 and $\mu \ll \mathscr{L}^d$. Hence, Proposition 2.2.17 can be applied to get that $\bar{\gamma} = \gamma_T$, where T is a transport map which is monotone nondecreasing on each transport ray. Notice that we still need to prove that T is Borel measurable; since this can be done by applying a measurability criterion in the setting of disintegration of measures given in [1], we will not deal with this issue here.

Uniqueness: Suppose that two different transport plans γ_{T_1} and γ_{T_2} optimize the secondary variational problem, then the measure $\gamma = \frac{1}{2}\gamma_{T_1} + \frac{1}{2}\gamma_{T_2}$ also solves the problem. Hence, there exists a transport map T such that $\gamma = \gamma_T$. But this is impossible unless $T_1 = T_2$ μ-a.e. Therefore, $\gamma_{T_1} = \gamma_{T_2}$ (up to μ-negligible sets). $\square$

The optimal transport plan $\bar{\gamma}$ in Theorem 2.2.28 will be called **ray-monotone transport plan**, and the transport map which corresponds to $\bar{\gamma}$ is known as the **ray-monotone transport map**.

Corollary 2.2.29. *If $\mu \ll \mathscr{L}^d$, then the ray-monotone transport map is optimal for the L^1 Monge's transport problem.*

Proof. By Ambrosio-Sudakov's Theorem, there is an optimal solution $\bar{\gamma}$ for the secondary variational problem, which is induced by the ray-monotone transport map T. Hence, since $\bar{\gamma} = (\mathrm{Id}, T)_{\#}\mu$ we must have

$$\inf(\mathrm{MP}) = \min K_1(\bar{\gamma}) = \int_{\Omega\times\Omega} |x-y| d\bar{\gamma}(x,y) = \int_{\Omega} |x - T(x)|\, d\mu(x).$$

$\square$

Chapter 3

Beckmann's Problem and Transport Density

In this chapter, we present Beckmann's problem and its connection with the L^1 Monge's transport problem. The aim of this chapter is to prove that both problems are equivalent, and study regularity issues of the transport density associated to the solution of Beckmann's problem. The proofs that we present here essentially follow Santambrogio's presentation [21, 22, 23].

3.1 The continuity equation

Eulerian formalism: Describes for every point x and every time t what are the velocity, the density, and the flow rate (in intensity and direction) of particles x at times t. We can classify these models into static ones and dynamic ones.

- In a dynamical model, one usually works with two variables, the density $\rho_t(x)$ and the velocity $v_t(x)$. If we suppose that v_t is Lipschitz continuous and bounded, then the position of the particle initially located at x will be given by the solution of the ODE:

$$\begin{cases} y_x'(t) = v_t(y_x(t)), \\ y_x(0) = x. \end{cases} \tag{3.1.1}$$

  Let us now define the map $Y_t(x) = y_x(t)$ and consider an initial density $\rho_0 \ll \mathscr{L}^d$. Then, we can define a family of absolutely continuous densities $\{\rho_t\}_t$ as $\rho_t := (Y_t)_\#(\rho_0)$ (understanding that ρ_t is a measure). We will see later that ρ_t along with v_t satisfy the equation $\partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0$.

- Static models, might be thought as a time average of some dynamical model. For instance, suppose that a certain fluid passes through a pipe. If some fluid is injected through one side of the pipe according to a density ρ_1 and then exits through the other side with density ρ_2. In a static framework, the vector field w standing for flows connecting ρ_1 and ρ_2 is expected to satisfy, $\nabla \cdot \mathrm{w} = \rho_1 - \rho_2$.

Let $\{\rho_t\}_t$ be a family of densities with no flux boundary conditions (Neumann conditions) $\rho_t v_t \cdot n = 0$, with n a normal interior vector. The following equation is known as the **continuity equation**:

$$\partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0, \tag{3.1.2}$$

which is formally thought in the distributional sense.

From now on, we will suppose that $\Omega \subset \mathbb{R}^d$ is a bounded domain, or $\Omega = \mathbb{R}^d$. The following definition give us a notion of solution for the continuity equation in a distributional sense.

Definition 3.1.1. *We say that a family of pairs* $\{(\rho_t, v_t) : \rho_t \in \mathscr{P}(\Omega),\ v_t$ *a vector fields*$\}$ *with* $v_t \in L^1(\rho_t; \mathbb{R}^d)$ *and* $\int_0^T \|v_t\|_{L^1(\rho_t)} = \int_0^T \int_\Omega |v_t| d\rho_t dt < \infty$ *solves the continuity equation 3.1.2 on* $(0,T)$ *in the distributional sense, if for any bounded and Lipschitz test function* $\phi \in C_c^1((0,T) \times \overline{\Omega})$ *we have that*

$$\int_0^T \int_\Omega \left[(\partial_t \phi) d\rho_t dt + \nabla \phi \cdot v_t\right] d\rho_t dt = 0$$

- Note that in the last definition we require the support of ϕ to be far form $t = 0, 1$ but not from $\partial\Omega$, when Ω is bounded; also Ω is usually supposed to be close, but we will write $\overline{\Omega}$ to stress the fact that we include its boundary.
- If $\Omega \neq \mathbb{R}^d$, this definition includes no-flux boundary conditions on $\partial\Omega$, i.e., $\rho_t v_t \cdot n = 0$.

Remark 3.1.2. If we want to impose the initial and final measures, we say that (ρ_t, v_t) solves equation 3.1.2 in the distributional sense, with initial data ρ_0 and ρ_T, if for every test function $\phi \in C_c^1([0,T] \times \overline{\Omega})$ (now we do not require that the support to be far from $t = 0, 1$), we have

$$\int_0^T \int_\Omega (\partial_t \phi) d\rho_t dt + \int_0^T \int_\Omega \nabla \phi \cdot v_t d\rho_t dt = \int_\Omega \phi(T, x) d\rho_T(x) - \int_\Omega \phi(0, x) d\rho_0(x).$$

We also define a weak solution of the continuity equation as follows.

Definition 3.1.3. *We say that* $\{(\rho_t, v_t)\}_t$ *solves the continuity equation in the weak sense if for any test function* $\psi \in C_c^1(\overline{\Omega})$*, the map* $t \mapsto \int_\Omega \psi d\rho_t$ *is absolutely continuous in* t *and, for a.e.* t*, we have that*

$$\frac{d}{dt} \int_\Omega \psi d\rho_t = \int_\Omega \nabla \psi \cdot v_t d\rho_t.$$

Remark 3.1.4.

- If the same conditions from definition 3.1.3 hold, the map $t \mapsto \rho_t$ is automatically continuous for the weak convergence (take $\psi = 1$), and if we impose the initial data ρ_0 and ρ_1, then, $\rho_t \rightharpoonup \rho_0$ and $\rho_t \rightharpoonup \rho_1$, with $t \to 0$ and $t \to 1$, respectively.
- These two notions of solution are equivalent: It can be shown that every weak solution is a distributional solution and every distributional solution admits a representative (another family $\bar{\rho}_t = \rho_t$, for a.e. t) which is weakly continuous and is a weak solution.

A usual way to produce a solution to the continuity equation is by using a flow (3.1.1). Let us assume that v_t is Lipschitz continuous and uniformly bounded, then by existence and uniqueness theorem of ODE's there is a family of densities $\{\rho_t\}_t$ induced by the flow

(3.1.1), such that $\rho_t := (Y_t)_{\#}\rho_0$. Let us suppose that $\mathrm{spt}(\rho_t) \subset \Omega$ (which is satisfied if ρ_0 is concentrated on Ω and v_t satisfies appropriate Neumann boundary conditions). Now, we will show that $(\rho_t, v_t)_t$ is a weak solution. Indeed, let $\psi : \Omega \to \mathbb{R}$ be a test function, then:

$$\begin{aligned}
\frac{d}{dt}\int_\Omega \psi d\rho_t &= \frac{d}{dt}\int_\Omega \psi(y_x(t))d\rho_0(x) \qquad\qquad (\rho_t = (Y_t)_{\#}\rho_0)\\
&= \int_\Omega \nabla\psi(y_x(t)) \cdot y_x'(t)d\rho_0(x)\\
&= \int_\Omega \nabla\psi(y_x(t)) \cdot v_t(y_x(t))d\rho_0(x)\\
&= \int_\Omega \nabla\psi(y) \cdot v_t(y)d\rho_t(y),
\end{aligned}$$

which proves that we have, $\partial_t\rho_t + \nabla \cdot (\psi v_t) = 0$, in the weak sense.

Remark 3.1.5. Suppose that ρ is Lipschitz continuous in (t, x), that v is Lipschitz in x, and the continuity equation $\partial_t\rho_t + \nabla \cdot (\rho_t v_t) = 0$ is satisfied in the weak sense. Then, from Rademacher's Theorem it can be inferred that the continuity equation also holds in the a.e. sense.

From now on, we will understand a solution to the continuity equation either as a weak solution or solution in the distributional sense. The following result states that the solution of the continuity equation is also unique under certain assumptions on the vector fields v_t (see [23], chapter 4).

Theorem 3.1.6. *Suppose that $v_t : \Omega \to \mathbb{R}^d$ is Lipschitz continuous in x, uniformly in t, and uniformly bounded, and consider its flow Y_t. Also, we suppose that for every $x \in \Omega$ and every $t \in [0, T]$, we have $Y_t(x) \in \Omega$ (which is clear for $\Omega = \mathbb{R}^d$ and require suitable Neumann conditions on v_t otherwise). Then, for every probability measure $\rho_0 \in \mathscr{P}(\Omega)$, the measures $\rho_t = (Y_t)_{\#}\rho_0$ solve the continuity equation 3.1.2 with initial data ρ_0. Moreover, every solution of the same equation with $\rho_t \ll \mathscr{L}^d$ for every t is necessarily obtained as $\rho_t = (Y_t)_{\#}\rho_0$. In particular, the continuity equation admits a unique solution.*

3.2 Beckmann's problem

In this section, we will show that Beckmann's problem admits a unique solution. Moreover, we shall prove a decomposition theorem which will allow us to deduce that the solution of Beckmann's problem is induced by an optimal transport plan of Monge's original problem [22, 23].

In 1952 Martin Beckmann introduced the following flow minimization problem in [5]:

$$\min\left\{\int_\Omega \mathscr{H}(\mathrm{w})dx : \nabla \cdot \mathrm{w} = \mu - \nu\right\},$$

where $\mathscr{H}(z) = H(|z|)$ and $H(t) = g(t)t$, with H a super-linear function. Intuitively, The super-linearity condition on H means that the congestion effect becomes stronger as the traffic increases.

This problem has been wildly studied not only because its applications in economics and urban traffic congestion [11, 12], but because it turns out to provide a natural way to pass from Lagrangian to Eulerian frameworks and back. In this section, we will only discuss the problem when $H(z) = |z|$, which is a limit case in the class of costs $H(z) = |z|^p$, $1 < p < \infty$. The reason to deal with this case lies in its relation with Monge's transportation problem that we discussed in Chapter 2.

Let us consider the following minimization problem

$$\min\left\{\int_\Omega |\,\mathrm{w}(x)|dx : \mathrm{w} : \Omega \to \mathbb{R}^d,\ \nabla\cdot\mathrm{w} = \mu - \nu\right\},$$

where the divergence condition is to be read in the weak sense, with no-flux boundary conditions, in other words, $-\int_\Omega \nabla\phi\cdot d\,\mathrm{w} = \int_\Omega \phi d(\mu-\nu)$ for any test function $\phi \in C^1(\overline{\Omega})$.

Remark 3.2.1. We observe that there may not exist an L^1 vector field minimizing the L^1 norm under this constraints. Indeed, if we think in a direct method in calculus of variations, we take a sequence $\{\mathrm{w}_n\}_{n\geq 1}$ and we would like to get a converging subsequence. If we could do so, it would be easy to prove that w satisfies the divergence constraints, since $\mathrm{w}_n \rightharpoonup \mathrm{w}$(weakly in L^1) and the relation $-\int_\Omega \nabla\phi\cdot\mathrm{w}_n\,dx = \int_\Omega \phi d(\mu-\nu)$ implies that $\nabla\cdot\mathrm{w} = \mu-\nu$. However, the condition $\int_\Omega |\,\mathrm{w}(x)|dx < \infty$ is not enough to guarantee the existence of a weakly convergent subsequence, for instance, the sequence $\mathrm{w}_n(x) = n\chi_{\left(0,\frac{1}{n}\right)}$ is bounded in $L^1([0,1])$, but does not has any weakly convergent subsequence.

To avoid these well-posedness difficulties, we will work on a more natural setting, the framework of vector measures (see [6] or Appendix A to see relevant definitions and notation).

Notation 3.2.2. We denote by $\mathscr{M}^d(\Omega)$ the space of finite vector measures on Ω endowed with the total variation norm, i.e.

$$\|\lambda\| := |\lambda|(\Omega) = \sup\left\{\int_\Omega \xi\cdot d\lambda : \|\xi\|_{L^\infty} \leq 1\right\}.$$

We will say that $\lambda_n \rightharpoonup \lambda$ (weakly converge) if and only if $\int_\Omega \xi\cdot d\lambda_n \to \int_\Omega \xi\cdot d\lambda$ for every $\xi \in C_b(\Omega, \mathbb{R}^n)$. Also, we will denote by

$$\mathscr{M}^d_{\mathrm{div}}(\Omega) := \left\{\mathrm{w} \in \mathscr{M}^d(\Omega) : \nabla\cdot\mathrm{w} \text{ is a scalar measure}\right\}.$$

Problem 3.2.3 (Beckmann's problem). *Let $\Omega \subset \mathbb{R}^d$ be a compact and convex domain. The following flow minimization problem is known as **Beckmann's problem:***

$$\min\left\{|\,\mathrm{w}\,|(\Omega) : \mathrm{w} \in \mathscr{M}^d_{div}(\Omega),\ \nabla\cdot\mathrm{w} = \mu-\nu\right\}, \tag{BP}$$

with divergence imposed in the weak sense, and no-flux boundary conditions.

Now, we are ready to prove the following theorem which guarantee that (BP) admits a solution and min(BP) = min(KP).

Theorem 3.2.4. *Suppose that $\Omega \subset \mathbb{R}^d$ is compact and a convex domain. Then, Beckmann's problem admits a solution. Moreover,* $\min(BP) = \min(KP)$ *and a solution to (BP) can be built from a solution for (KP).*

Proof. Let us start by proving that $\min(\text{KP}) \leq \min(\text{BP})$. Indeed, let $\phi \in C^1(\overline{\Omega})$ such that $|\nabla\phi| \leq 1$. Then, if one takes any $\mathrm{w} \in \mathscr{M}^d_{\mathrm{div}}(\Omega)$ with $\nabla \cdot \mathrm{w} = \mu - \nu$, we have

$$|\,\mathrm{w}\,|(\Omega) = \int_\Omega d|\,\mathrm{w}\,| \geq \int_\Omega (-\nabla\phi) d|\,\mathrm{w}\,| = -\int_\Omega \nabla\phi \cdot d\,\mathrm{w} = \int_\Omega \phi d(\mu - \nu). \tag{3.2.1}$$

Now, let u be a Kantorovich potential such that $\displaystyle\int_\Omega u d(\mu - \nu) = \max(\text{DP}) = \min(\text{KP})$ and define the sequence $\phi_k = \eta_k * u$, with $\eta_k = k^d \eta(kx)$ the standard mollifier function such that $\text{spt}(\eta_k) \subset B\left(0, \frac{1}{k}\right)$. By some well known properties of mollifier function we get that $\phi_k \to u$ uniformly (since we are assuming that Ω is compact). Moreover, the condition $u \in \text{Lip}_1(\Omega)$ implies that $\{\phi_k\}_{k\geq 1} \subset \text{Lip}_1(\Omega) \cap C^1(\Omega)$ and $|\nabla\phi_k| \leq 1$ (by the mean value theorem). Hence, inequality 3.2.1 holds for $\{\phi_k\}_{k\geq 1}$, then we get

$$\int_\Omega d|\,\mathrm{w}\,| \geq -\lim_{k\to\infty} \int_\Omega \nabla\phi_k \cdot d\,\mathrm{w} = \lim_{k\to\infty} \int_\Omega \phi_k d(\mu - \nu) = \int_\Omega u d(\mu - \nu) = \min(\text{KP}),$$

for any admissible w. Thus, $\min(\text{BP}) \geq \min(\text{KP})$.

We will show the reverse inequality and how to construct an optimal w from an optimal transport plan γ for (KP). Let $\gamma \in \Pi(\mu, \nu)$ be an optimal transport plan, now we construct the vector measure $\mathbf{w}_{[\gamma]}$, which is induced by γ:

$$\langle \mathrm{w}_{[\gamma]}, \xi \rangle := \int_{\Omega\times\Omega} \int_0^1 \omega'_{x,y}(t) \cdot \xi(\omega_{x,y}(t)) dt d\gamma(x, y), \tag{3.2.2}$$

for every $\xi \in C(\Omega; \mathbb{R}^d)$, and $\omega_{x,y}$ being a parametrization of the segment $[x.y] \subset \Omega$ (since Ω is convex). Without loss of generality we will fix $\omega_{x,y}(t) = (1-t)x + ty$. It can be shown that $\mathrm{w}_{[\gamma]}$ satisfies the divergence constraint. Indeed, if one takes $\xi = -\nabla\phi$, with $\phi \in C^1(\overline{\Omega})$ then

$$\begin{aligned}
-\int_\Omega \nabla\phi \cdot d\,\mathrm{w}_{[\gamma]} = \langle \mathrm{w}_{[\gamma]}, -\nabla\phi \rangle &= \int_{\Omega\times\Omega} \int_0^1 \omega'_{x,y}(t) \cdot \nabla\phi(\omega_{x,y}(t)) dt d\gamma(x, y) \\
&= -\int_{\Omega\times\Omega} \int_0^1 \frac{d}{dt}(\phi(\omega_{x,y}(t)) dt d\gamma(x, y) \\
&= \int_{\Omega\times\Omega} (\phi(x) - \phi(y)) d\gamma(x, y) \\
&= \int_\Omega \phi d(\mu - \nu),
\end{aligned}$$

and hence $\mathrm{w}_{[\gamma]} \in \mathscr{M}^d_{\mathrm{div}}(\Omega)$. Now we estimate the mass of $\mathrm{w}_{[\gamma]}$. Let us define the **transport density $\boldsymbol{\sigma_\gamma}$** through the duality:

$$\langle \sigma_\gamma, \phi \rangle := \int_{\Omega\times\Omega} \int_0^1 |\omega'_{x,y}(t)| \phi(\omega_{x,y}(t)) dt d\gamma(x, y), \tag{3.2.3}$$

for all *phi* $\in C(\Omega;\mathbb{R})$. We will prove that $|\,\mathrm{w}_{[\gamma]}\,| = \sigma_\gamma$. Indeed, let u be the Kantorovich potential associated to γ, by Lemma 2.2.13 and Proposition 2.2.17 u is differentiable on each transport ray, and for every $t \in (0,1)$ and $x, y \in \mathrm{spt}(\gamma)$ we have that $\omega_{x,y}(t)$ is in the interior of the transport ray $[x, y]$ (if $x \neq y$), and the following is valid

$$\omega'_{x,y}(t) = -|x-y|\frac{x-y}{|x-y|} = -|x-y|\nabla u(\omega_{x,y}(t)).$$

If we define the function $\phi_t(x,y) = \omega_{x,y}(t) = (1-t)x + ty \in \Omega$, then for every $\xi \in C(\Omega;\mathbb{R}^d)$ we can write:

$$\begin{aligned}\int_\Omega \xi \cdot d\,\mathrm{w}_{[\gamma]} = \langle \mathrm{w}_{[\gamma]}, \xi\rangle &= \int_{\Omega\times\Omega}\int_0^1 -|x-y|\nabla u(\omega_{x,y}(t))\cdot\xi(\omega_{x,y}(t))dtd\gamma(x,y)\\ &= -\int_0^1 dt\int_{\Omega\times\Omega}\nabla u(\omega_{x,y}(t))\cdot\xi(\omega_{x,y}(t))|x-y|d\gamma(x,y)\\ &= -\int_0^1 dt\int_{\Omega\times\Omega}\nabla u(z)\cdot\xi(z)d((\phi_t)_\#(c\gamma)) \qquad ,\end{aligned}$$

where $c\gamma$ is the measure on $\Omega\times\Omega$ with density $c(x,y) = |x-y|$ with respect to γ. On the other hand, performing a similar computation we get that

$$\int_\Omega \phi d\sigma_\gamma = \langle\sigma_\gamma, \phi\rangle = \int_0^1 dt\int_{\Omega\times\Omega}\phi(z)d((\pi_t)_\#(c\gamma)).$$

Hence, if we take $\phi = -\xi\cdot\nabla u$ we get $\int_\Omega \xi\cdot d\,\mathrm{w}_{[\gamma]} = -\int_\Omega \xi\cdot\nabla u d\sigma_\gamma$ for every $\xi \in C_0(\Omega,\mathbb{R}^d)$, which shows that $\mathrm{w}_{[\gamma]} = -\nabla u\sigma_\gamma$ and $|\nabla u| = 1$ σ_γ-a.e. This gives the density of $\mathrm{w}_{[\gamma]}$ with respect to σ_γ, and shows that $|\,\mathrm{w}_{[\gamma]}\,| = \sigma_\gamma$. We also note that the mass of σ_γ is given by

$$\begin{aligned}\int_\Omega d\sigma_\gamma = \int_{\Omega\times\Omega}\int_0^1|\omega'_{x,y}(t)|dtd\gamma(x,y) &= \int_{\Omega\times\Omega}\int_0^1|x-y||\nabla u(\omega_{x,y}(t))|dtd\gamma(x,y)\\ &= \int_{\Omega\times\Omega}\int_0^1|x-y|dtd\gamma(x,y)\\ &= \int_{\Omega\times\Omega}|x-y|d\gamma(x,y)\\ &= \min(\mathrm{KP}).\end{aligned}$$

Thus, by equation 3.2.4 we get

$$\min(\mathrm{KP}) \leq \min(\mathrm{BP}) \leq |\,\mathrm{w}_{[\gamma]}\,|(\Omega) = \sigma_\gamma(\Omega) = \min(\mathrm{KP}). \tag{3.2.4}$$

This proves the optimality of $\mathrm{w}_{[\gamma]}$. Therefore, a minimizer for the (BP) exists and $\min(\mathrm{BP}) = \min(\mathrm{KP})$ (such a minimizer has been built from a solution of (KP)). □

Remark 3.2.5. Let us consider the transport density σ_γ from Theorem 3.2.4. If we look at the action of σ_γ on sets, then we have for every Borel set A,

$$\sigma_\gamma(A) = \int_{\Omega\times\Omega}\mathscr{H}^1(A\cap[x,y])d\gamma(x,y),$$

where $[x, y]$ is the segment joining x and y.

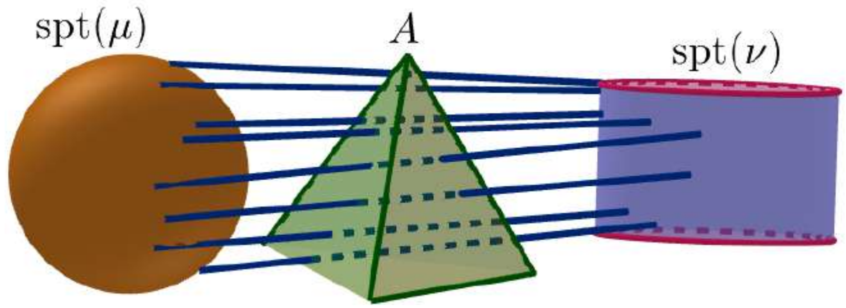

Figure 3.1: $\sigma_\gamma(A)$ represents the average of the length of all segments $A \cap [x, y]$ according to γ.

Remark 3.2.6. The transport density $\sigma = \sigma_\gamma$ that we defined in (3.2.3) satisfies the so-called Monge-Kantorovich system:

$$\begin{cases} \nabla \cdot (\sigma \nabla u) = \mu - \nu & \text{in } \Omega, \\ |\nabla u| \leq 1 & \text{in } \Omega, \\ |\nabla u| = 1 & \sigma - \text{a.e.} \end{cases}$$

We note that this system is formally solved if the product $\nabla u \cdot \sigma$ makes sense, in other words, if $\sigma \in L^1$. To overcome this difficulty, in the last section we will study L^p regularity of the transport density σ.

In the following example, we will find the transport density σ between two concrete measures in $\mathbb{R}^2$.

Example 3.2.7. Let $\mu = f \cdot \mathscr{L}^2$ and $\nu = g \cdot \mathscr{L}^2$, with $f = \frac{1}{\pi}\chi_{B(0,1)}$ and $g = \frac{1}{4\pi}\chi_{B(0,2)}$. By applying the same reasoning as in Example 2.2.5 and Theorem 2.2.28, we find a radial optimal transport plan T which is the unique solution of Kantorovich's transport problem. Hence, σ must be a radial function. Let $u(x) = -|x|$ be its associated Kantorovich potential; thanks to Remark 3.2.6 we know that σ and u must satisfy

$$\begin{aligned} &-\nabla \cdot (\sigma \nabla(|x|)) = -\nabla \cdot \left(\sigma \frac{x}{|x|}\right) = \frac{1}{\pi}\chi_{B(0,1)}(x) - \frac{1}{4\pi}\chi_{B(0,2)}(x) \\ &\Longrightarrow -\nabla\sigma \cdot \frac{x}{|x|} - \sigma \nabla\left(\frac{x}{|x|}\right) = \frac{1}{\pi}\chi_{B(0,1)}(x) - \frac{1}{4\pi}\chi_{B(0,2)}(x) \\ &\Longrightarrow \sigma' \frac{x \cdot x}{|x|} + \sigma \frac{2-1}{|x|} = \frac{1}{4\pi}\chi_{B(0,2)}(x) - \frac{1}{\pi}\chi_{B(0,1)}(x) \\ &\Longrightarrow \sigma' + \sigma \frac{1}{|x|} = \frac{1}{4\pi}\chi_{B(0,2)}(x) - \frac{1}{\pi}\chi_{B(0,1)}(x). \end{aligned}$$

To find the transport density we need to solve the ODE

$$\sigma'(r) + \sigma \frac{1}{r} = \begin{cases} -\frac{3}{4\pi}, & \text{if } r \leq 1, \\ \frac{1}{4\pi}, & \text{if } 1 \leq r \leq 2, \end{cases}$$

with $\sigma(0) = 0$. Indeed, if we multiply both sides of the equation $\sigma'(r) + \sigma(r)\frac{1}{r} = -\frac{3}{4\pi}$ by $\exp \int \frac{1}{r} dr = r$, it can be obtained

$$(r\sigma(r))' = -\frac{3r}{8\pi} \Longrightarrow r\sigma(r) = -\frac{3}{4\pi}\int r dr = -\frac{3r^2}{8\pi} + c_1 \Longrightarrow \sigma(r) = -\frac{3}{8\pi}r + \frac{c_1}{r},$$

which means that $\sigma(r) = -\frac{3r}{8\pi}$, if $x \in B(0,1)$. Applying analogous computations, we obtain that $\sigma(r) = \frac{r}{2\pi} + \frac{c_2}{r}$, if $x \in B(0,2) \setminus B(0,1)$. Finally, we need both possible solutions to coincide if $|x| = 1$, that is, $-\frac{3}{8\pi} = \frac{1}{2\pi} + c_2$. Therefore, the transport density is given by

$$\sigma(|x|) = \begin{cases} -\frac{3}{8\pi}|x|, & \text{if } x \in B(0,1), \\ \frac{1}{2\pi}|x| - \frac{7}{8\pi|x|}, & \text{if } x \in B(0,2) \setminus B(0,1). \end{cases}$$

In order to complete the analysis of this section, we will analyze the one dimensional (BP). With this in mind, we suppose that $\Omega = [a,b] \subset \mathbb{R}$.

First, we note that the condition $\nabla \cdot \mathrm{w} = \mu - \nu$ is stronger if $d = 1$. Indeed, we note that in $\mathbb{R}$ there exist just one partial derivative for the vector field w, this implies that the condition $\nabla \cdot \mathrm{w} = \mu - \nu$ along with the Neumann boundary conditions means that

$$\mathrm{w}(x) = \mathrm{w}(a) + \int_a^x \mathrm{w}'(t)dt = \int_a^x \mathrm{w}'(t)dt = \int_a^x d(\mu - \nu)(t) = F_\mu(x) - F_\nu(x). \tag{3.2.5}$$

Thus, $\mathrm{w}(b) = 0$ thanks to the fact $F_\mu(b) = F_\nu(b)$. On the other hand, we recall the following connections between measures and functions with bounded variation in the one dimensional case.

Well known Facts:

- $\mathrm{BV}([a,b]) \subset L^\infty([a,b])$.
- Every monotone increasing function on a compact interval is the cumulative distribution function of unique positive measure
- Every function in $\mathrm{BV}([a,b])$ is the cumulative distribution of a unique signed measure.

Remark 3.2.8. Thanks to the fact that w' is a measure, we have $\mathrm{w} \in \mathrm{BV}([a,b])$, and hence we can say that $\mathscr{M}^1_{\mathrm{div}}([a,b]) = \mathrm{BV}([a,b])$. Therefore we have the following facts:

- Since $\mathrm{w} \in \mathrm{BV}([a,b])$, then w belong to every L^p space, including the case $p = \infty$.
- If $\mu, \nu \in L^p([a,b])$, then $\mathrm{w} \in \mathrm{W}^{1,p}([a,b])$ thanks to the fact that $\mathrm{w}' \in L^p([a,b])$.
- By (3.2.5) it follows:

$$\begin{aligned} \min(\mathrm{BP}) = \{|\,\mathrm{w}\,|([a,b]) : \mathrm{w} \in \mathrm{BV}([a,b]),\ \nabla \cdot \mathrm{w} = \mu - \nu\} &= \int_a^b |F_\mu - F_\nu| d\mathscr{L}^1 \\ &= \|F_\mu - F_\nu\|_{L^1}([a,b]). \end{aligned}$$

- In this case, the transport density σ is given by $\sigma = |\,\mathrm{w}\,|$ and $\sigma \in \mathrm{BV}([a.b])$, since the absolute value of a BV function is also a BV function.

3.2.1 Smirnov's decomposition

In this section, we first introduce some objects that generalize $\mathrm{w}_{[\gamma]}$, σ_γ and we prove the Smirnov's decomposition theorem. This theorem will be useful to characterize an optimal vector measure w coming from an optimal transport plan. Let us now introduce some notation. Let $\Omega \subset \mathbb{R}^d$ be a compact, connected, with no empty interior set. We define the set

$$\mathscr{C} := \{\omega : [0,1] \to \Omega : \omega \text{ is an absolutely continuous curve}\}.$$

Given $\omega \in \mathscr{C}$ and $\phi \in C(\Omega)$, let us set the quantity

$$L_\phi(\omega) := \int_0^1 \phi(\omega(t))|\omega'(t)|dt, \tag{3.2.6}$$

which represents the length of the curve, weighted with some function ϕ. When we take $\phi = 1$, $L_1(\omega)$ is the usual length of ω, let us denote it by $L(\omega)$.

Remark 3.2.9. Note that these quantities are well defined, since absolutely continuous curves have distributional derivative $\omega' \in L^1([0,1])$, and $\omega(t_1) = \omega(t_0) + \int_{t_0}^{t_1} \omega'(t)dt$ for every $t_0 < t_1$, so ω is differentiable a.e.

Definition 3.2.10. *If $\mathcal{Q} \in \mathscr{P}(\mathscr{C})$ is a probability measure such that $\int_{\mathscr{C}} L(\omega)d\mathcal{Q}(\omega) < +\infty$. We will called $\mathcal{Q}$ a **traffic plans**.*

Example 3.2.11. If we consider the flow given by

$$\begin{cases} y_x'(t) = v_t(y_x(t)), \\ y_x(0) = x, \end{cases} \tag{3.2.7}$$

with v_t a velocity vector field. Then we can define $Y : \Omega \to \mathscr{C}$ as $Y(x) = y_x$. This map allow us to transport the measure $\mu \in \mathscr{P}(\Omega)$ to another measure $\mathcal{Q} = Y_{\#}\mu$. Then, $\mathcal{Q}$ is a traffic plan.

Remark 3.2.12. We endow the space $\mathscr{C}$ with the uniform convergence metric. Then, Arzela-Ascoli theorem guarantees that the set $\{\omega \in \mathscr{C} : \mathrm{Lip}(\omega) \leq \alpha\}$ is compact for every $\alpha > 0$. Please be aware that the space $(\mathscr{C}, \|\cdot\|_\infty)$ is not complete (since the uniform limit of absolutely continuous curves does not necessary belong to $\mathscr{C}$).

We will associate two measures on Ω related to a traffic plan $\mathcal{Q}$, the traffic intensity and the traffic flow.

Definition 3.2.13. *The scalar measure, called **traffic intensity** and denoted by $i_{\mathcal{Q}} \in \mathscr{M}_+(\Omega)$ is defined by*

$$\int_\Omega \phi di_{\mathcal{Q}} := \int_{\mathscr{C}} \left(\int_0^1 \phi(\omega(t))|\omega'(t)|dt \right) d\mathcal{Q}(\omega) = \int_{\mathscr{C}} L_\phi(\omega)d\mathcal{Q}(w),$$

for all $\phi \in C(\Omega; \mathbb{R}_+)$. The interpretation of this measure is the following: for a connected set $A \subset \Omega$, the quantity $i_{\mathcal{Q}}(A)$ represents the total accumulated traffic in A induced by $\mathcal{Q}$.

We also associate a vector measure $\mathrm{w}_{\mathcal{Q}}$ with any traffic plan $\mathcal{Q} \in \mathscr{P}(\mathscr{C})$ through

$$\int_\Omega \xi \cdot d\,\mathrm{w}_{\mathcal{Q}} := \int_{\mathscr{C}} \left(\int_0^1 \xi(\omega(t)) \cdot \omega'(t)dt \right) d\mathcal{Q}(\omega),$$

*for all $\xi \in C(\Omega; \mathbb{R}^d)$. We will call $\mathrm{w}_{\mathcal{Q}}$ **traffic flow** induced by $\mathcal{Q}$.*

Example 3.2.14. Let us consider $Y(x) = y_x$, induced by the flow (3.1.1). If we set $\mu = \delta_{x_0}$ for some $x_0 \in \Omega$. Then, we have that $\mathcal{Q} = \delta_{y_{x_0}}$ and

$$\int_\Omega di_{\mathcal{Q}} = \int_\Omega \int_0^1 |v_t(y_x(t))| dt d(\delta_{y_{x_0}}) = \int_0^1 |v_t(y_{x_0}(t))| dt.$$

Then, $i_{\mathcal{Q}} = \mathscr{H}^1 \, |_{y_{x_0}([0,1])}$. Also,

$$\int_\Omega 1 \cdot d\,\mathrm{w}_{\mathcal{Q}} = \int_\Omega \int_0^1 1 \cdot \frac{v_t(y_x(t))}{|v_t(y_x(t))|} dt d(\delta_{y_{x_0}}) = \int_0^1 1 \cdot \frac{v_t(y_{x_0}(t))}{|v_t(y_{x_0}(t))|} dt.$$

Hence, $\mathrm{w}_{\mathcal{Q}} = \frac{v_t(y_{x_0}(t))}{|v_t(y_{x_0}(t))|} \mathscr{H}^1 \, |_{y_{x_0}([0,1])}$.

Example 3.2.15. Consider the transport plan $\gamma = \sum_{i=1}^n r_i \delta_{(x_i, y_i)}$, with $r_i \geq 0$ and $\sum_{i=1}^n r_i = 1$. In this case, the traffic plan is given by $\mathcal{Q} = \sum_{i=1}^n r_i \delta_{\omega_{x_i, y_i}}$, where ω_{x_i, y_i} is the segment joining x_i with y_i. Following the same ideas from Example 3.2.14 we can compute the traffic intensity (which is in this case σ_γ), and the traffic flow $\mathrm{w}_{\mathcal{Q}}$:

$$\sigma_\gamma = \sum_{i=1}^n r_i \mathscr{H}^1 \, |_{\omega_{x_i, y_i}} \ ([0,1])$$

$$\mathrm{w}_{\mathcal{Q}} = \sum_{i=1}^n r_i \frac{y_i - x_i}{|x_i - y_i|} \mathscr{H}^1 \, |_{\omega_{x_i, y_i}} \ ([0,1])$$

Definition 3.2.16. *Let $e_t : \mathscr{C} \to \mathbb{R}^d$ be the evaluation map at time t. We will say that a traffic plan $\mathcal{Q}$ is* ***admissible*** *if $(e_0)_\# \mathcal{Q} = \mu$ and $(e_1)_\# \mathcal{Q} = \nu$, where $\mu, \nu \in \mathscr{P}(\Omega)$.*

Remark 3.2.17. If we take a gradient field $\xi = \nabla \phi$ in the definition of traffic flow, then we have

$$\begin{aligned}\int_\Omega \nabla \phi \cdot d\,\mathrm{w}_{\mathcal{Q}} = \int_{\mathscr{C}} \left(\int_0^1 \nabla \phi(\omega(t)) \cdot \omega'(t) \right) d\mathcal{Q}(\omega) &= \int_{\mathscr{C}} [\phi(\omega(1)) - \phi(\omega(0))] d\mathcal{Q}(\omega) \\ &= \int_\Omega \phi d((e_1)_\# \mathcal{Q} - (e_0)_\# \mathcal{Q}),\end{aligned}$$

where e_t denotes the evaluation map, $e_t(\omega) = \omega(t)$. This means that $\nabla \cdot \mathrm{w}_{\mathcal{Q}} = \mu - \nu$ in the distributional sense, and is endowed with no-flux conditions. Hence, $\mathrm{w}_{\mathcal{Q}}$ is an admissible flow connecting μ and ν.

The aim of this section is to prove Smirnov's decomposition Theorem following Santambrogio's approach [22]. This Theorem will allow us to decompose a vector measure into a flow induced by a measure on paths and a vector field in $\mathrm{Ker}(\nabla \cdot)$. With this goal in mind, we first need to prove some preliminary results.

Remark 3.2.18. If we take $\xi \in C(\Omega; \mathbb{R}^d)$, then by Cauchy-Schwarz inequality we have

$$\int_\Omega \xi \cdot d\,\mathrm{w}_{\mathcal{Q}} \leq \int_{\mathscr{C}} \left(\int_0^1 |\xi(\omega(t))| |\omega'(t)| dt \right) d\mathcal{Q}(\omega) = \int_\Omega |\xi| di_{\mathcal{Q}}.$$

Thus, $|\,\mathrm{w}_{\mathcal{Q}}\,| \leq i_{\mathcal{Q}}$, where $|\,\mathrm{w}_{\mathcal{Q}}\,|$ is the total variation measure. It can be shown, that in general, equality does not holds (since the curves of $\mathcal{Q}$ could produce some cancellations).

The following proposition give us some useful properties of the traffic intensity and traffic flow.

Proposition 3.2.19. *(a) If $T : \mathscr{C} \to \mathscr{C}$ is a map such that for every ω, the curve $T(\omega)$ is a re-parametrization in time of ω, then $\mathrm{w}_{T_{\#}\mathcal{Q}} = \mathrm{w}_{\mathcal{Q}}$ and $i_{T_{\#}\mathcal{Q}} = i_{\mathcal{Q}}$ (both are invariant under re-parametrization.*

(b) For every $\mathcal{Q}$, we have that $\displaystyle\int_{\mathscr{C}} L(\omega) d\mathcal{Q}(\omega) = i_{\mathcal{Q}}(\Omega)$.

(c) If $\mathcal{Q}_n \rightharpoonup \mathcal{Q}$ and $i_{\mathcal{Q}_n} \rightharpoonup i$, then $i \geq i_{\mathcal{Q}}$.

(d) If $\mathcal{Q}_n \rightharpoonup \mathcal{Q}$, $\mathrm{w}_{\mathcal{Q}_n} \rightharpoonup \mathrm{w}$ and $i_{\mathcal{Q}_n} \rightharpoonup i$, then

$$\| \mathrm{w} - \mathrm{w}_{\mathcal{Q}} \| \leq i(\Omega) - i_{\mathcal{Q}}(\Omega).$$

In particular, if $\mathcal{Q}_n \rightharpoonup \mathcal{Q}$ and $i_{\mathcal{Q}_n} \rightharpoonup i_{\mathcal{Q}}$, then $\mathrm{w}_{\mathcal{Q}_n} \rightharpoonup \mathrm{w}_{\mathcal{Q}}$

Proof. **(a):** The proof of this part comes form the invariance of both, $L_\phi(\omega) = L_\phi(T(\omega))$ and $\displaystyle\int_0^1 \xi(\omega(t)) \cdot \omega'(t) dt = \int_0^1 \xi(T(\omega(t))) \cdot T(\omega(t))' dt$.
(b): It follows directly from the definition of $i_{\mathcal{Q}}$ and by testing with $\phi = 1$.

(c): To check the inequality $i \geq i_{\mathcal{Q}}$, we first prove a particular case. Let us fix a test function $\phi \in C(\Omega)$ such that $\phi \geq \epsilon_0 > 0$. By definition of traffic intensity we can write

$$\int_\Omega \phi di_{\mathcal{Q}_n} = \int_{\mathscr{C}} \left(\int_0^1 \phi(\omega(t)) |\omega'(t)| dt \right) d\mathcal{Q}_n(\omega).$$

Then, the function $L_\phi(\omega) = \int_0^1 \phi(\omega(t))|\omega'(t)|dt$ is positive and lower semi-continuous. Indeed, let $\{\omega_n\}_{n\geq 1} \subset \mathscr{C}$ a sequence such that $\omega_n \to \omega$. Let $\left\{\omega_{n_j}\right\}_{j\geq 1}$ be a subsequence such that $L_\phi(\omega_{n_j}) \to \liminf\limits_{n\to\infty} L_\phi(\omega_n) < +\infty$. Then, we can assume that $\displaystyle\int_0^1 |\omega'_{n_j}| dt$ is bounded for every $j \in \mathbb{N}$, since

$$L_{\epsilon_0}(\omega_{n_j}) = \epsilon_0 \int_0^1 |\omega'_{n_j}| dt \leq L_\phi(\omega_{n_j}) < +\infty.$$

Then, there exists a subsequence of $\left\{\omega_{n_j}\right\}_{j\geq 1}$ weakly converging to ω' as a measure. To avoid confusion, let us assume that $\omega'_n \rightharpoonup \omega'$ weakly (as measures). Then, up to a subsequence, there exist a measure $f \in \mathscr{M}_+([0,1])$ such that $f \geq |\omega'|$ and $|\omega'_n| \rightharpoonup f$, since the variation measure satisfies $|\omega'| \leq \liminf\limits_{n\to\infty} |\omega'_n|$. Moreover, $\phi(\omega_n(t)) \to \phi(\omega(t))$ uniformly, which gives

$$\int_0^1 \phi(\omega_n(t))|\omega'_n(t)|dt \to \int_0^1 \phi(\omega(t)) df(t) \geq \int_0^1 \phi(\omega(t))|\omega'(t)|dt.$$

Thus, $\mathcal{Q}_n \rightharpoonup \mathcal{Q}$ and $i_{\mathcal{Q}_n} \rightharpoonup i$ implies that

$$\int_\Omega \phi di = \lim_{n\to\infty} \int_\Omega \phi di_{\mathcal{Q}_n} = \liminf_{n\to\infty} \int_{\mathscr{C}} L_\phi(\omega) d\mathcal{Q}_n(\omega) \geq \int_{\mathscr{C}} L_\phi(\omega) d\mathcal{Q}(\omega) = \int_\Omega \phi di_{\mathcal{Q}}. \quad (3.2.8)$$

Now, if we take an arbitrary $\phi \in C(\Omega)$ then we can add a constant $\epsilon_0 > 0$ and apply the same reasoning of equation (3.2.1) to get that

$$\int_\Omega (\phi + \epsilon_0) di = \liminf_{n\to\infty} \int_{\mathscr{C}} L_\phi(\omega) d\mathcal{Q}_n(\omega) + \epsilon_0 i(\Omega) \geq \int_\Omega (\phi + \epsilon_0) di_{\mathcal{Q}}.$$

Since i is a finite measure, we let $\epsilon_0 \to 0$ to obtain that L_ϕ is lower semi-continuous and $i_{\mathcal{Q}} \leq i$.

(d): Let us fix a vector field $\xi \in C(\Omega; \mathbb{R}^d)$ and a number $\lambda > 1$. Then we have:

$$\begin{aligned}
\int_\Omega \xi \cdot d\,\mathrm{w}_{\mathcal{Q}_n} &= \int_{\mathscr{C}} \left(\int_0^1 \xi(\omega(t)) \cdot \omega'(t) dt \right) d\mathcal{Q}_n(\omega) \\
&= \int_{\mathscr{C}} \left(\int_0^1 \xi(\omega(t)) \cdot \omega'(t) + \lambda \|\xi\|_{L^\infty} L(\omega) - \lambda \|\xi\|_{L^\infty} L(\omega) dt \right) d\mathcal{Q}_n(\omega) \qquad (3.2.9) \\
&= \int_{\mathscr{C}} \left(\int_0^1 \xi(\omega(t)) \cdot \omega'(t) + \lambda \|\xi\|_{L^\infty} L(\omega) dt \right) d\mathcal{Q}_n(\omega) - \lambda \|\xi\|_{L^\infty} i_{\mathcal{Q}_n}(\Omega).
\end{aligned}$$

We also note that the function

$$\omega \mapsto \int_0^1 (\xi(\omega(t)) \cdot \omega'(t) + \lambda \|\xi\|_{L^\infty} L(\omega)) dt \geq (\lambda - 1) \|\xi\|_{L^\infty} L(\omega),$$

is lower semi-continuous with respect to ω. Indeed, by similar arguments if we take $\omega \to \omega$ uniformly (as we did before), we can suppose that $L(\omega_n)$ is bounded and then obtain that $\omega_n' \rightharpoonup \omega'$ weakly (up to a subsequence). Then, $\displaystyle\int_0^1 \xi(\omega_n(t)) \cdot \omega_n' dt \to \int_0^1 \xi(\omega) \cdot \omega'(t) dt$ and

$$\int_0^1 (\xi(\omega(t)) \cdot \omega'(t) + \lambda \|\xi\|_{L^\infty} L(\omega)) dt \leq \liminf_{n\to\infty} \int_0^1 (\xi(\omega_n(t)) \cdot \omega_n'(t) + \lambda \|\xi\|_{L^\infty} L(\omega_n)) dt,$$

since $L(\omega)$ is lower semi-continuous. This means that, if we pass to the limit in equation (3.2.13) we get

$$\begin{aligned}
\int_\Omega \xi \cdot d\,\mathrm{w} &= \lim_{n\to\infty} \int_\Omega \xi \cdot d\,\mathrm{w}_{\mathcal{Q}_n} \\
&\geq \int_{\mathscr{C}} \left(\int_0^1 \xi(\omega(t)) \cdot \omega'(t) + \lambda \|\xi\|_{L^\infty} L(\omega) \right) d\mathcal{Q}(\omega) - \lambda \|\xi\|_{L^\infty} i(\Omega) \\
&= \int_\Omega \xi \cdot d\,\mathrm{w}_{\mathcal{Q}} + \lambda \|\xi\|_{L^\infty} (i_{\mathcal{Q}}(\Omega) - i(\Omega)).
\end{aligned}$$

Analogously, if we replace ξ with $-\xi$ we get, $\displaystyle\int_\Omega \xi \cdot d\,\mathrm{w} \leq \int_\Omega \xi \cdot d\,\mathrm{w}_{\mathcal{Q}} + \lambda \|\xi\|_{L^\infty} (i(\Omega) - i_{\mathcal{Q}}(\Omega))$. Thus, letting $\lambda \to 1$ we have the following estimate

$$\begin{aligned}
&\left| \int_\Omega \xi \cdot d\,\mathrm{w} - \int_\Omega \xi \cdot d\,\mathrm{w}_{\mathcal{Q}} \right| \leq \|\xi\|_{L^\infty} (i(\Omega) - i_{\mathcal{Q}}(\Omega)) \\
\Longrightarrow &\sup_{\|\xi\|_{L^\infty} \leq 1} \left| \int_\Omega \xi \cdot d\,\mathrm{w} - \int_\Omega \xi \cdot d\,\mathrm{w}_{\mathcal{Q}} \right| \leq i(\Omega) - i_{\mathcal{Q}}(\Omega)
\end{aligned}$$

Therefore, $\|\,\mathrm{w} - \mathrm{w}_{\mathcal{Q}}\,\| \leq i(\Omega) - i_{\mathcal{Q}}(\Omega)$. Finally, the last property follows from the previous work. Indeed, may can assume that there exists a subsequence such that $\mathrm{w}_{\mathcal{Q}_{n_k}} \rightharpoonup \mathrm{w}$ for some vector measure w (since $\mathcal{Q}_n \rightharpoonup \mathcal{Q}$, and then one may suppose that, up to a subsequence, $\mathrm{w}_{\mathcal{Q}_n} \rightharpoonup \mathrm{w}$), and $i = i_{\mathcal{Q}}$ implies that $\|\,\mathrm{w} - \mathrm{w}_{\mathcal{Q}}\,\| \leq i(\Omega) - i_{\mathcal{Q}}(\Omega) = 0$. Thus, $\mathrm{w}_{\mathcal{Q}_{n_k}} \rightharpoonup \mathrm{w}_{\mathcal{Q}}$, which also implies the full convergence of the sequence. $\square$

Now, we will show an approximation lemma which will be useful to prove Smirnov's decomposition theorem.

Lemma 3.2.20. *Let $\mu, \nu \in \mathscr{P}(\Omega)$ be two probability measures on a smooth compact domain Ω, and a vector measure $\mathrm{w} \in \mathscr{M}^d_{div}$ satisfying $\nabla \cdot \mathrm{w} = \mu - \nu$ in the distributional sense (with no-flux boundary conditions). If $\{\Omega_\epsilon\}$ is a sequence of smooth compact domains such that, $\Omega \subset Int(\Omega_\epsilon)$ and $\Omega_\epsilon \to \Omega$ in the Hausdorff topology, then there exist a family of vector fields $\{\mathrm{w}^\epsilon\}_\epsilon \subset C^\infty(\Omega_\epsilon)$ with $\mathrm{w}^\epsilon \cdot n_{\Omega\epsilon} = 0$, two families of densities $\{\mu^\epsilon\}_\epsilon$ and $\{\nu^\epsilon\}_\epsilon$ in $C^\infty(\Omega_\epsilon)$, bounded from below by positive constants $k_\epsilon > 0$, such that*

$$\nabla \cdot \mathrm{w}^\epsilon = \mu^\epsilon - \nu^\epsilon \text{ and } \int_{\Omega_\epsilon} \mu^\epsilon = \int_{\Omega_\epsilon} \nu^\epsilon = 1,$$

and $\mathrm{w}^\epsilon \rightharpoonup \mathrm{w}$, $\mu^\epsilon \rightharpoonup \mu$, $\nu^\epsilon \rightharpoonup \nu$ and $|\,\mathrm{w}^\epsilon\,| \rightharpoonup |\,\mathrm{w}\,|$.

Proof. We first extend the measures μ, ν and w equals to zero out of Ω (that is, $\mu(\Omega^c) = \nu(\Omega^c) = \mathrm{w}(\Omega^c) = 0$) and take convolution in $\mathbb{R}^d$ with a Gaussian kernel $\eta_\epsilon(x) = \frac{1}{\sqrt{\epsilon}} \exp\left(-\frac{\pi|x|^2}{\epsilon}\right)$, then we get the measures $\bar{\mu}^\epsilon = \mu * \eta_\epsilon$, $\bar{\nu}^\epsilon = \nu * \eta_\epsilon$ and $\bar{\mathrm{w}}^\epsilon = \mathrm{w} * \eta_\epsilon$, where

$$\mu * \eta_\epsilon(x) := \int_{\mathbb{R}^d} \eta_\epsilon(x-y) d\mu(y), \quad \nu * \eta_\epsilon(x) := \int_{\mathbb{R}^d} \eta_\epsilon(x-y) d\nu(y),$$
$$\mathrm{w} * \eta_\epsilon(x) := \int_{\mathbb{R}^d} \eta_\epsilon(x-y)(1 \cdot d\,\mathrm{w}(y)).$$

These measures satisfy the constraint $\nabla \cdot \bar{\omega}^\epsilon = \bar{\mu}^\epsilon - \bar{\nu}^\epsilon$, since

$$\begin{aligned} -\int_\Omega \nabla\phi \cdot d\bar{\omega}^\epsilon = -\int_\Omega \nabla\phi \cdot d(\mathrm{w} * \eta_\epsilon) &= -\int_{\mathbb{R}^d}\int_\Omega \eta_\epsilon(x-y)(\nabla\phi(x) \cdot d\,\mathrm{w})dy \\ &= \int_{\mathbb{R}^d}\int_\Omega \eta_\epsilon(x-y)\phi(x) d(\mu-\nu)dy \\ &= \int_\Omega \phi d(\bar{\mu}^\epsilon - \bar{\nu}^\epsilon), \end{aligned}$$

for every test function. Since $\eta_\epsilon > 0$, we also have that $\bar{\mu}^\epsilon$ and $\bar{\nu}^\epsilon$ are positive densities. Using these densities, we will do a regularization on their support and boundary behavior, since their support is not necessarily Ω_ϵ and $\bar{\mathrm{w}}^\epsilon \cdot n_{\Omega_\epsilon}$ is not zero. Indeed, let us set

$$\int_{\Omega_\epsilon} d\bar{\mu}^\epsilon = 1 - a_\epsilon \text{ and } \int_{\Omega_\epsilon} d\bar{\nu}^\epsilon = 1 - b_\epsilon,$$

with $a_\epsilon, b_\epsilon \to 0$ as $\epsilon \to 0$ (since $\bar{\mu}^\epsilon \rightharpoonup \mu$ and $\bar{\nu}^\epsilon \rightharpoonup \nu$). On the other hand, since $t_\epsilon = \mathrm{dist}(\Omega, \partial\Omega_\epsilon) \to 0$ (thanks to the Hausdorff convergence), if we set $t_\epsilon = \epsilon^{\frac{1}{3}}$ then it follows that $\frac{t_\epsilon}{\epsilon} \to +\infty$, as $\epsilon \to 0$ and $\|\eta_\epsilon\|_{L^\infty(B(0,t_\epsilon)^c)} \to 0$, as ϵ goes to zero. Hence, $|\,\mathrm{w}_\epsilon \cdot n_{\Omega'}| \le c_\epsilon$, with $c_\epsilon \to 0$.

Let us consider u_ϵ the solution to the problem

$$\begin{cases} \Delta u_\epsilon = \frac{a_\epsilon - b_\epsilon}{\mathscr{L}^d(\Omega_\epsilon)}, & \text{if } \Omega_\epsilon \\ \frac{\partial u_\epsilon}{\partial n} = -\bar{\mathrm{w}}^\epsilon \cdot n_{\Omega_\epsilon}, & \text{, on } \partial\Omega_\epsilon \\ \int_{\Omega_\epsilon} u_\epsilon = 0, \end{cases} \tag{3.2.10}$$

and the vector field ∇u_ϵ. We observe that the problem (3.2.10) admits a solution, since

$$-\int_{\partial\Omega_\epsilon} \bar{\mathrm{w}}^\epsilon \cdot n_{\Omega_\epsilon} = \int_{\Omega_\epsilon} (\bar{\mu}^\epsilon - \bar{\nu}^\epsilon) = a_\epsilon - b_\epsilon.$$

Now, we use integration by parts to get that

$$\begin{aligned}
\int_{\Omega_\epsilon} |\nabla u_\epsilon|^2 d\mathscr{L}^d &= -\int_{\Omega_\epsilon} u_\epsilon \Delta u_\epsilon d\mathscr{L}^d - \int_{\partial\Omega_\epsilon} u_\epsilon(\bar{\mathrm{w}}^\epsilon \cdot n_{\Omega_\epsilon}) d\mathscr{H}^{d-1} \\
&= -\int_{\Omega_\epsilon} u_\epsilon \left(\frac{a_\epsilon - b_\epsilon}{\mathscr{L}^d(\Omega_\epsilon)}\right) d\mathscr{L}^d - \int_{\partial\Omega_\epsilon} u_\epsilon(\bar{\mathrm{w}}^\epsilon \cdot n_{\Omega_\epsilon}) d\mathscr{H}^{d-1} \\
&\le c_\epsilon \int_{\partial\Omega_\epsilon} |u_\epsilon| d\mathscr{H}^{d-1} + (a_\epsilon + b_\epsilon)\mathscr{L}^d(\Omega_\epsilon) \int_{\Omega_\epsilon} |u_\epsilon|^2 d\mathscr{L}^d \\
&\le c_\epsilon \mathscr{H}^{d-1}(\partial\Omega_\epsilon) \int_{\partial\Omega_\epsilon} |u_\epsilon|^2 d\mathscr{H}^{d-1} + (a_\epsilon + b_\epsilon)\mathscr{L}^d(\Omega_\epsilon) \int_{\Omega_\epsilon} |u_\epsilon|^2 d\mathscr{L}^d \\
&\le C(c_\epsilon + a_\epsilon + b_\epsilon)\|\nabla u_\epsilon\|_{L^2(\Omega_\epsilon)},
\end{aligned}$$

thanks to Poincare's inequality and (3.2.10). This shows that $\|\nabla u_\epsilon\|_{L^2(\Omega_\epsilon)} \to 0$ (since C does not depend on ϵ), and then Holder inequality implies $\|\nabla u_\epsilon\|_{L^1(\Omega_\epsilon)} \to 0$. Also we observe that, $|\bar{\mathrm{w}}^\epsilon| \rightharpoonup |\,\mathrm{w}\,|$ since $\bar{\mathrm{w}}^\epsilon \rightharpoonup \mathrm{w}$ and $\bar{\mathrm{w}}^\epsilon$ is a convolution. Hence, the following measures:

$$\mu^\epsilon = \bar{\mu}^\epsilon \mid_{\Omega_\epsilon} + \frac{a_\epsilon}{\mathscr{L}^d(\Omega_\epsilon)}; \quad \nu^\epsilon = \bar{\nu}^\epsilon \mid_{\Omega_\epsilon} + \frac{b_\epsilon}{\mathscr{L}^d(\Omega_\epsilon)}; \quad \mathrm{w}^\epsilon = \bar{\mathrm{w}}^\epsilon \mid_{\Omega_\epsilon} + \nabla u_\epsilon$$

satisfy all the required conditions. Therefore, we have built an approximating sequence $\{(\mathrm{w}^\epsilon, \mu^\epsilon, \nu^\epsilon)\}_\epsilon$ such that, $\mathrm{w}^\epsilon, \mu^\epsilon, \nu^\epsilon \in C^\infty(\Omega_\epsilon)$, $\mathrm{w}^\epsilon \cdot n_{\Omega'_\epsilon} = 0$, $\nabla \cdot \mathrm{w}_\epsilon = \mu^\epsilon - \nu^\epsilon$ and $(\mathrm{w}^\epsilon, \mu^\epsilon, \nu^\epsilon) \rightharpoonup (\mu, \nu, \mathrm{w})$, as desired. □

Dacorogna-Moser transport: Let us present a very useful construction due to Dacorogna and Moser. Let $\mathrm{w} : \Omega \to \mathbb{R}^d$ be a Lipschitz vector field such that $\mathrm{w} \cdot n = 0$ on $\partial\Omega$ and $\nabla \cdot \mathrm{w} = f_0 - f_1$, with f_0 and f_1 two probability densities which are Lipschitz continuous and bounded from below. Then we can define the vector field $v_t(x)$ as follows:

$$v_t(x) = \frac{\mathrm{w}(x)}{f_t(x)}, \text{ where } f_t = (1-t)f_0 + tf_1 \text{ (see figure 3.2)},$$

and consider the Cauchy problem

$$\begin{cases} y'_x(t) = v_t(y_x(t)) \\ y_x(0) = x. \end{cases}$$

We define a map $Y : \Omega \to \mathscr{C}$ given by $Y(x) = y_x(\cdot)$. Then, consider the measure $\mathcal{Q} := Y_\# f_0$ and $\rho_t := (e_t)_\# \mathcal{Q} := (Y_t)_\# f_0$. Thanks to section 3.1, we already know that $\{\rho_t\}_t$ solves the continuity equation $\partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0$. On the other hand, since $\partial_t f_t = f_1 - f_0$ and $\nabla \cdot (f_t v_t) = \nabla \cdot \mathrm{w} = f_0 - f_1$, we must have that $\{f_t\}_t$ also solves the continuity equation. By Theorem 3.1.6, since $\rho_0 = f_0$ we deduce that $\rho_t = f_t$ (recall all the ingredients are Lipschitz continuous). In particular, $x \mapsto y_x(1)$ is a transport map from f_0 to f_1.

In the following lemma we will show some useful properties of the traffic intensity and traffic flow of the associated measure in Dacorogna-Moser construction.

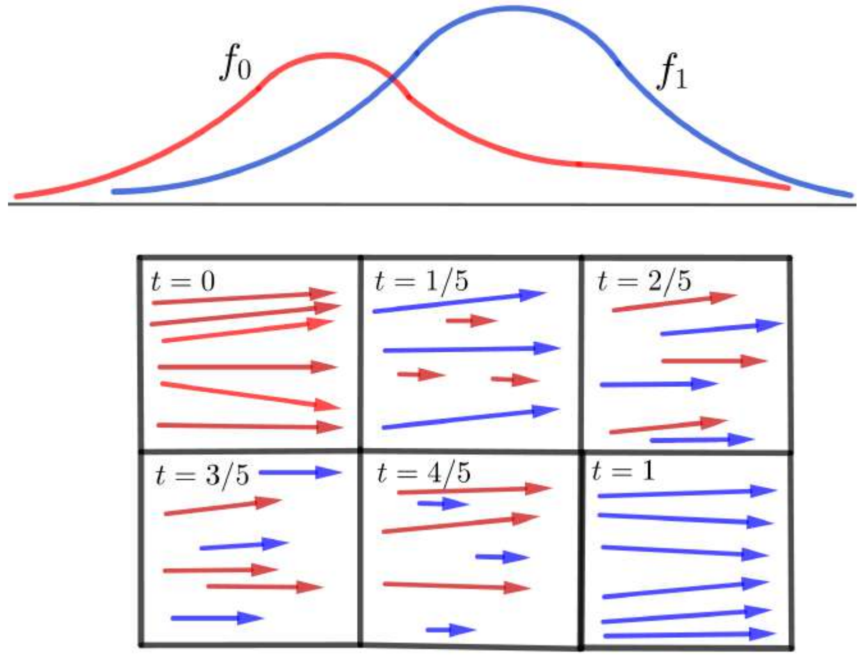

Figure 3.2: $v_t(x)$ in Dacorogna-Moser construction.

Lemma 3.2.21. *If $\mathcal{Q}$ is the associated measure in Dacorogna-Moser construction, then* $\mathrm{w} = \mathrm{w}_Q$ *and* $|\,\mathrm{w}_{\mathcal{Q}}\,| = i_{\mathcal{Q}}$.

Proof. Let $\phi \in C(\Omega; \mathbb{R})$ be a test function, then

$$\int_\Omega \phi di_{\mathcal{Q}} = \int_\Omega \int_0^1 \phi(y_x(t))|y_x'(t)|dtd(Y_\# f_0) = \int_\Omega \int_0^1 \phi(y_x(t))|v_t(y_x(t))|dtf_0(x)dx$$
$$= \int_0^1 \int_\Omega \phi(y)|v_t(y)|f_t(y)dydt$$
$$= \int_\Omega \phi(y)|\,\mathrm{w}(y)|dy,$$

so we have that $i_{\mathcal{Q}} = |\,\mathrm{w}\,|$. On the other hand, let $\xi \in C(\Omega; \mathbb{R}^d)$ be a vector field. Then,

$$\int_\Omega \xi \cdot d\,\mathrm{w}_{\mathcal{Q}} = \int_\Omega \int_0^1 \xi(y_x(t)) \cdot v_t(y_x(t))dtf_0(x)$$
$$= \int_0^1 \int_\Omega \xi(y) \cdot v_t(y)f_t(y)dydt$$
$$= \int_\Omega \xi \cdot \mathrm{w}(y),$$

since $v_t(y) = \frac{\mathrm{w}(y)}{f_t(y)}$. Thus, $\mathrm{w}_{\mathcal{Q}} = \mathrm{w}$. In this case, we also have $|\,\mathrm{w}_Q\,| = i_{\mathcal{Q}} = |\,\mathrm{w}\,|$ due to the fact that all curves share the same direction at every given point, as a consequence of the Existence and Uniqueness ODE's theorem and hence no cancellation is possible. $\square$

Finally, we are ready to prove Smirnov's decomposition theorem. The statement that we present here is a particular version of a more general decomposition theorem due to Smirnov [24].

Theorem 3.2.22 (Smirnov's decomposition)**.** *For every finite vector measure* $\mathrm{w} \in \mathscr{M}^d_{div}(\Omega)$ *and* $\mu, \nu \in \mathscr{P}(\Omega)$ *with* $\nabla \cdot \mathrm{w} = \mu - \nu$*, there exists a traffic plan* $\mathcal{Q} \in \mathscr{P}(\mathscr{C})$ *with* $(e_0)_\# \mathcal{Q} = \mu$ *and* $(e_1)_\# \mathcal{Q} = \nu$ *such that* $|\,\mathrm{w}_\mathcal{Q}\,| = i_\mathcal{Q} \leq |\,\mathrm{w}\,|$*, and*

$$\|\,\mathrm{w} - \mathrm{w}_\mathcal{Q}\,\| + \|\,\mathrm{w}_\mathcal{Q}\,\| = \|\,\mathrm{w} - \mathrm{w}_\mathcal{Q}\,\| + i_\mathcal{Q}(\Omega) = \|\,\mathrm{w}\,\|.$$

In particular, $|\,\mathrm{w}_\mathcal{Q}\,| \neq |\,\mathrm{w}\,|$ *unless* $\mathrm{w}_\mathcal{Q} = \mathrm{w}$.

Remark 3.2.23.

- Note that the statement $\|\,\mathrm{w} - \mathrm{w}_\mathcal{Q}\,\| + \|\,\mathrm{w}_\mathcal{Q}\,\| = \|\,\mathrm{w}\,\|$ is equivalent to the measure decomposition $\mathrm{w} = \mathrm{w}_\mathcal{Q} + (\mathrm{w} - \mathrm{w}_\mathcal{Q})$, with $|\,\mathrm{w}_\mathcal{Q}\,| \perp |\,\mathrm{w} - \mathrm{w}_\mathcal{Q}\,|$ (by uniqueness of the Jordan decomposition).
- We will say that a vector measure w is a **cycle**, if $\nabla \cdot \mathrm{w} = 0$, i.e. a vector measure in $\mathrm{Ker}(\nabla\cdot)$. The decomposition given in Theorem 3.2.22 give us the decomposition of any vector measure into a cycle $\mathrm{w} - \mathrm{w}_\mathcal{Q}$ and a flow $\mathrm{w}_\mathcal{Q}$ induced by a measure on paths, such that $\|\,\mathrm{w}\,\| = \|\,\mathrm{w} - \mathrm{w}_\mathcal{Q}\,\| + \|\,\mathrm{w}_\mathcal{Q}\,\|$

Proof of theorem 3.2.19. Let $\{\Omega_\epsilon\}$ be a sequence of smooth compact domains such that, $\Omega \subset \mathrm{Int}(\Omega_\epsilon)$ and $\Omega_\epsilon \to \Omega$ in the Hausdorff topology. Then, by Lemma 3.2.20 there exists a sequence $\{(\mathrm{w}^\epsilon, \mu^\epsilon, \nu^\epsilon)\}_\epsilon$ with $\mathrm{w}^\epsilon, \mu^\epsilon, \nu^\epsilon \in C^\infty(\Omega_\epsilon)$ such that, $\mathrm{w}^\epsilon \cdot n_{\Omega_\epsilon} = 0$, $\nabla \cdot \mathrm{w}_\epsilon = \mu^\epsilon - \nu^\epsilon$ (in the distributional sense), $(\mathrm{w}^\epsilon, \mu^\epsilon, \nu^\epsilon) \rightharpoonup (\mathrm{w}, \mu, \nu)$ and $|\,\mathrm{w}^\epsilon\,| \rightharpoonup |\,\mathrm{w}\,|$. If we apply Dacorogna-Moser's construction to the vector fields w^ϵ, we obtain a sequence of measures $\{\mathcal{Q}_\epsilon\}_\epsilon$ such that $\mathcal{Q}_\epsilon = Y_\# \mu^\epsilon$. Let us suppose that these are probability measures on $\mathscr{C} := \mathrm{AC}(\Omega')$, with $\Omega \subset \Omega_\epsilon \subset \Omega'$ and such that each measure is concentrated on curves valued in Ω_ϵ.

By properties of the traffic intensity and traffic flow associated to $\mathcal{Q}_\epsilon$ (Lemma 3.2.21) we have that $i_{\mathcal{Q}_\epsilon} = |\,\mathrm{w}^\epsilon\,|$ and $\mathrm{w}_{\mathcal{Q}_\epsilon} = \mathrm{w}^\epsilon$, also the invariance under reparametrization (Proposition 3.2.19) allows us to reparametrize by constant speed the curves on which $\mathcal{Q}_\epsilon$ is supported, without changing traffic intensities and traffic flows. Then, we will use curves ω such that $L(\omega) = \mathrm{Lip}(\omega)$. On the other hand, we note that $\displaystyle\int_\mathscr{C} \mathrm{Lip}(\omega) d\mathcal{Q}_\epsilon(\omega)$ is bounded, since

$$\begin{aligned}
\lim_{\epsilon \to 0^+} \int_\mathscr{C} \mathrm{Lip}(\omega) d\mathcal{Q}_\epsilon(\omega) = \lim_{\epsilon \to 0^+} \int_\mathscr{C} L(\omega) d\mathcal{Q}_\epsilon &= \int_{\Omega'} d i_{\mathcal{Q}_\epsilon} \\
&= \lim_{\epsilon \to 0^+} \int_{\Omega'} d|\,\mathrm{w}^\epsilon\,| \\
&= \int_{\Omega'} d|\,\mathrm{w}\,| = |\,\mathrm{w}\,|(\Omega') < +\infty.
\end{aligned}$$

Thus, the sequence $\{\mathcal{Q}_\epsilon\}_\epsilon$ is tight, thanks to the fact that the sets $\{\omega \in \mathscr{C} : \mathrm{Lip}(\omega) \leq L\}$ are compact for every $L \in \mathbb{R}_+$. By Prokhorov's Theorem we may assume that, up to a subsequence, $\mathcal{Q}_\epsilon \rightharpoonup \mathcal{Q}$, where this measure is concentrated on curves valued on Ω (since $\Omega_\epsilon \to \Omega$ in the Hausdorff topology). From the Dacorogna-Moser construction we know that $\mathcal{Q}_\epsilon$ were constructed so that $(e_0)_\# \mathcal{Q}_\epsilon = \mu^\epsilon$ and $(e_1)_\# \mathcal{Q}_\epsilon = \nu^\epsilon$, thus letting $\epsilon \to 0^+$ it follows that $(e_0)_\# \mathcal{Q} = \mu$ and $(e_1)_\# \mathcal{Q} = \nu$.

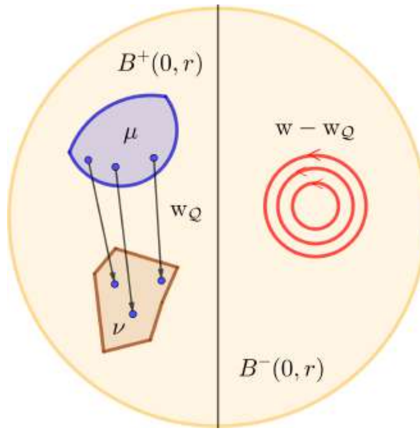

Figure 3.3: The Smirnov's decomposition of w in Example 3.2.24

Moreover, since $i_{\mathcal{Q}_\epsilon} = |\,\mathrm{w}^\epsilon\,| \rightharpoonup |\,\mathrm{w}\,|$ and $\mathrm{w}^\epsilon \rightharpoonup \mathrm{w}$, then Proposition 3.2.19 implies $|\,\mathrm{w}_{\mathcal{Q}}\,| \leq i_{\mathcal{Q}} \leq |\,\mathrm{w}\,|$ and

$$\|\,\mathrm{w}-\mathrm{w}_{\mathcal{Q}}\,\| \leq |\,\mathrm{w}\,|(\Omega) - i_{\mathcal{Q}}(\Omega) \leq \|\,\mathrm{w}\,\| - i_{\mathcal{Q}}(\Omega).$$

Thus, we have that

$$\|\,\mathrm{w}-\mathrm{w}_{\mathcal{Q}}\,\| + \|w_{\mathcal{Q}}\| \leq \|\,\mathrm{w}-\mathrm{w}_{\mathcal{Q}}\,\| + i_{\mathcal{Q}}(\Omega) \leq \|\,\mathrm{w}\,\|. \tag{3.2.11}$$

Let $\xi \in C(\Omega;\mathbb{R})$ be a vector field such that $\|\xi\|_{L^{+\infty}} \leq 1$. Then,

$$\left|\int_\Omega \xi \cdot d\,\mathrm{w}\right| \leq \left|\int_\Omega \xi \cdot d\,\mathrm{w} - \int_\Omega \xi \cdot d\,\mathrm{w}_{\mathcal{Q}}\right| + \left|\int_\Omega \xi \cdot d\,\mathrm{w}_{\mathcal{Q}}\right|$$
$$\Longrightarrow \sup_{\|\xi\|_{L^\infty}\leq 1} \left|\int_\Omega \xi \cdot d\,\mathrm{w}\right| \leq \sup_{\|\xi\|_{L^\infty}\leq 1} \left|\int_\Omega \xi \cdot d\,\mathrm{w} - \int_\Omega \xi \cdot d\,\mathrm{w}_{\mathcal{Q}}\right| + \sup_{\|\xi\|_{L^\infty}\leq 1} \left|\int_\Omega \xi \cdot d\,\mathrm{w}_{\mathcal{Q}}\right|,$$

which means that $\|\,\mathrm{w}\,\| \leq \|\,\mathrm{w}-\mathrm{w}_{\mathcal{Q}}\,\| + \|\,\mathrm{w}_{\mathcal{Q}}\,\|$ is satisfied. Hence, we conclude that $\|\,\mathrm{w}\,\| = \|\,\mathrm{w}-\mathrm{w}_{\mathcal{Q}}\,\| + \|\,\mathrm{w}_{\mathcal{Q}}\,\|$ by combining last inequality and (3.2.11). □

The following example shows what happens to a cycle in Dacorogna-Moser's construction.

Example 3.2.24. Let $B(0,r) \subset \mathbb{R}^2$ with $r > 0$, and let $B^-(0,r), B^+(0,r)$ be the half circles with respect to the vertical axis. Also, we assume that $\mathrm{spt}(\mu) \cup \mathrm{spt}(\nu) \subset B^+(0,r)$ and a cycle of w is contained in $B^-(0,r)$ (see figure 3.3).

In this case, we note that the approximating sequence $\{\mu^\epsilon\}$ and $\{\nu^\epsilon\}$ may have positive mass in $B^-(0,r)$, an such mass goes to zero as $\epsilon \to 0^+$. Also, we observe that the vector field from Dacorogna-Moser's construction satisfies

$$v_t^\epsilon = \frac{\mathrm{w}^\epsilon}{(1-t)\mu^\epsilon + \nu^\epsilon t} \Longrightarrow \mathrm{w}^\epsilon = v_t^\epsilon((1-t)\mu^\epsilon + \nu^\epsilon t). \tag{3.2.12}$$

Hence, the curves satisfying (3.2.12) will follow the cycle $\mathrm{w}-\mathrm{w}_Q$, passing many times on each point of this cycle, which means that the flow $\mathrm{w}^\epsilon\,|_{B^-(0,r)}$ is obtained from a small mass which passes many times through the cycle. Thus, since $\mathcal{Q}_\epsilon \rightharpoonup \mathcal{Q}$ we must have that $\mathcal{Q}$ is concentrated on curves staying in $B^+(0,r)$ and $\mathrm{w}_{\mathcal{Q}}\,|_{B^-(0,r)} = 0$.

Remark 3.2.25. In general, if cycles are located in regions where μ and ν have positive measure, we cannot apply the same reasoning from Example 3.2.24.

3.2.2 Characterization and uniqueness of optimal vector measures.

Now, we are ready to prove the following characterization theorem which states that an optimal vector measure w in Beckmann's problem always comes from an optimal transport plan γ, and also that all the optimal transport plans γ give us the same $\mathrm{w}_{[\gamma]}$ and σ_γ, if $\mu \ll \mathscr{L}^d$. Hence, in the end, $\mathrm{w}_{[\gamma]}$ does not depend on the choice of γ.

Theorem 3.2.26 (Characterization and uniqueness of the optimal w)**.** *Let* w *be an optimal vector measure for (BP). Then, there is an optimal transport plan $\gamma \in \Pi(\mu,\nu)$ such that $\mathrm{w} = \mathrm{w}_{[\gamma]}$. Moreover, If $\mu \ll \mathscr{L}^d$, then the vector field $\mathrm{w}_{[\gamma]}$ does not depend on the choice of the optimal transport plan γ.*

Proof. By Theorem 3.2.22 (Smirnov's decomposition), there exists a measure $\mathcal{Q} \in \mathscr{P}(\mathscr{C})$ with $(e_0)_\#\mathcal{Q} = \mu$ and $(e_1)_\#\mathcal{Q} = \nu$ such that $|\,\mathrm{w}_{\mathcal{Q}}\,| = i_{\mathcal{Q}} \leq |\,\mathrm{w}\,|$. Since w is optimal, then the equality $|\,\mathrm{w}\,| = |\,\mathrm{w}_{\mathcal{Q}}\,| = i_{\mathcal{Q}}$ must hold. Moreover, we have from Smirnov's decomposition:

$$\|\,\mathrm{w}\,\|_{\mathcal{Q}} = \|\,\mathrm{w} - \mathrm{w}_{\mathcal{Q}}\,\| + \|\,\mathrm{w}_{\mathcal{Q}}\,\| = \|\,\mathrm{w}\,\|,$$

since $\|\,\mathrm{w} - \mathrm{w}_{\mathcal{Q}}\,\| = 0$. In this case, $\mathrm{w} = \mathrm{w}_{\mathcal{Q}}$. Since $\mathrm{w}_{\mathcal{Q}}$ and $i_{\mathcal{Q}}$ are invariant under reparametrizations, we can assume that $\mathcal{Q}$ is concentrated on curves which are parametrized by constant speed. Let $\mathcal{S} = \{\omega_{x,y} \in \mathscr{C} : \omega_{x,y}(t) = (1-t)x + ty\}$ and define the map $S : \Omega \times \Omega \to \mathcal{S}$ by $S(x,y) = \omega_{x,y}$. Since the definition of w_γ is just a particular case of the definition of $\mathrm{w}_{\mathcal{Q}}$, when we take $\mathcal{Q} = S_\#\gamma$, then the first part of statement is proved if we show that $\mathcal{Q} = S_\#\gamma$, with γ and optimal transport plan. Indeed, by optimality of w we have:

$$\begin{aligned}
\min(\mathrm{BP}) = |\,\mathrm{w}\,|(\Omega) = i_{\mathcal{Q}}(\Omega) &= \int_{\mathscr{C}} L(\omega)d\mathcal{Q}(\omega) \\
&= \int_{\mathscr{C}} \left(\int_0^1 |\omega'(t)|dt\right) d\mathcal{Q}(\omega) \\
&\geq \int_{\mathscr{C}} |\omega(0) - \omega(1)| d\mathcal{Q}(\omega) \\
&= \int_{\Omega\times\Omega} |x-y| d((e_0,e_1)_\#\mathcal{Q})(x,y) \\
&\geq \min(\mathrm{KP}) = \min(\mathrm{BP}),
\end{aligned}$$

since Theorem 3.2.4 guarantees $\min(\mathrm{BP}) = \min(\mathrm{KP})$. Thus, we have the equality

$$|\,\mathrm{w}\,|(\Omega) = \int_{\mathscr{C}} |\omega(0)-\omega(1)|d\mathcal{Q}(\omega) = \int_{\Omega\times\Omega} |x-y|d((e_0,e_1)_\#\mathcal{Q})(x,y), \tag{3.2.13}$$

which means that $\mathcal{Q}$ must be concentrated on curves such that $L(\omega) = |\omega(0)-\omega(1)|$, in other words, $\mathcal{Q}$ is concentrated in $\mathcal{S}$. Also, the measure $\gamma = (e_0,e_1)_\#\mathcal{Q}$ belongs to $\Pi(\mu,\nu)$ and is optimal for the Kantorovich problem, thanks to equation (3.2.13). Therefore $\mathcal{Q} = S_\#\gamma$, and in such a case, we get $\mathrm{w} = \mathrm{w}_{[\gamma]}$, which is the first part of the statement.

Now, we prove the second part of the theorem. Let $u \in \mathrm{Lip}_1(\Omega)$ be a Kantorovich potential for an optimal transport $\gamma \in \Pi(\mu,\nu)$; since u does not depend on γ, then it determines a partition into transport rays, let us denote by $\mathscr{R}$ such a partition. Note that, by Proposition 2.2.14, the points of Ω where ∇u does not exist, are the points which belong to several transport rays and hence are Lebesgue negligible thanks to Rademacher's theorem; we denote by S the set of such points. Since $\mu \ll \mathscr{L}^d$, then $\mu(S) = 0$ and γ is concentrated in the set $(\pi_x)^{-1}(S^c)$. Let $R : \Omega \times \Omega \to \mathscr{R}$ be a map sending each pair (x, y) into the transport ray containing x, which is well defined γ-a.e. and is a Borel map. Then, we can disintegrate the measure γ along to the transport rays containing the point x according to R: there exist a family of measures $\{\gamma^r\}_{r\in\mathscr{R}}$ such that, for every test function $\phi \in C(\Omega \times \Omega)$ we have that

$$\int_{\Omega\times\Omega} \phi d\gamma = \int_{\mathscr{R}^{-1}(r)} \int_{\Omega\times\Omega} \phi d\gamma^r d\lambda,$$

which we will denote by $\gamma = \gamma^r \otimes \lambda$, with $\lambda = R_\# \gamma$ and r the variable related to each transport ray, every γ^r is a probability measure concentrated in $R^{-1}(r)$. Also, we have that for a.e. $r \in \mathscr{R}$, the transport plans $\gamma^r \in \Pi(\mu^r, \nu^r)$ are optimal.

On the other hand, the vector measure $\mathrm{w}_{[\gamma]}$ can also be obtained through the disintegration of γ, that is, $\mathrm{w}_{[\gamma]} = \mathrm{w}_{\gamma^r} \otimes \lambda$, where w_{γ^r} is the vector measure induced by γ^r, such that it is optimal in (BP). Hence, if want to prove that $\mathrm{w}_{[\gamma]}$ does not depend on γ, we just need show that each λ and each w_{γ^r} only depends on the marginals μ, ν. It is clear that λ does not depend on γ (since λ is the push-forward measure of a map which just depends on x and hence only depends on μ). Also, we note that w_{γ^r} only depends on the marginal measures of γ^r, since each w_{γ^r} is optimal in a one dimensional Beckmann's problem, and according to Remark 3.2.8 w_{γ^r} only depends on the marginal measures.

We claim now that $(\pi_x)_\# \gamma^r$ and $(\pi_y)_\# \gamma^r$ do not depend on γ. Indeed, note that the claim easily follows for the first marginal thanks to the fact that $(\pi_x)_\# \gamma^r$ must coincides with the disintegration of μ according to R, that is, $(\pi_x)_\# \gamma^r = \mu^r$. For the marginal $\nu^r := (\pi_y)_\# \gamma^r$, we have to perform a further analysis due to the condition $\nu(S) = 0$ does not necessarily holds. Let us decompose this marginal into:

$$(\pi_y)_\# \gamma^r = (\pi_y)_\# (\gamma^r \mid_{\Omega\times S}) + (\pi_y)_\# (\gamma^r \mid_{\Omega\times S^c}). \tag{3.2.14}$$

We observe that the second term on the right hand side of (3.2.14) is equal to $\nu \mid_{S^c}$ (again by uniqueness of the disintegration), and then $(\pi_y)_\# (\gamma^r \mid_{\Omega\times S^c})$ does not depend on γ since $\nu \mid_{S^c}$ depends only on the set S. Now, it just remain to prove the claim for $\nu \mid_S = (\pi_y)_\# (\gamma^r \mid_{\Omega\times S})$. By definition of the set S, we know that this measure is concentrated on those points where different transport rays may intersect, that is, $\nu \mid_S$ is concentrated on the two end points of the transport rays r. Then, $\nu \mid_{S^c}$ is an atomic measure which is composed by at most two Dirac masses.

On the other hand, let us recall that the condition $u(x) - u(y) = |x - y|$ on $\mathrm{spt}(\gamma)$, implies that the transport must follow a unique direction on each transport ray r. Now, we claim that the boundary point where u is maximal cannot have any mass of ν. In order to prove this claim we argue by contradiction: if we suppose that the measure ν has positive mass at the beginning of a transport ray, then we have an atom for the measure μ as well. By hypothesis we know that, $\mu \ll \mathscr{L}^d$ and property N holds (by Theorem 2.2.21), this means that the set of rays r such that μ^r has an atom is negligible. Thus, the source measure cannot have an atom at the beginning of any transport ray, which is a contradiction. Hence, $\nu^r \mid_S$ is a single Dirac mass.

Finally, since $\nu^r \mid_S$ is Dirac mass and $\nu^r \mid_S (r) + \nu^r \mid_{S^c} = \mu^r(r)$ then, $\nu^r \mid_S (r) = 1 - \nu^r \mid_{S^c} (r)$ (thanks to the fact that the total mass of μ^r must be transport to the single boundary point where $\nu^r \mid_S$ is concentrated), which only depends on μ and ν.

Hence, we have proved that w_γ does not depend on γ, and therefore w neither depend on γ. $\square$

Remark 3.2.27. If $\mu \ll \mathscr{L}^d$, we saw in Theorem 3.2.26 that any optimal w is vector measure of the form $\mathrm{w}_{[\gamma]}$, and does not depend on the optimal γ. Hence, (BP) admits a unique solution.

3.3 Integrability of the transport density

We have proved some results concerning existence and uniqueness of an optimal vector measure $\mathrm{w} \in \mathscr{M}^d_{\mathrm{div}}$ for (BP). In this case, we do not have any additional information about the transport density σ, except that is a measure. In this section, we will investigate whether the transport density have additional integrability properties assuming more regularity on the marginal measures of an optimal transport plan. We will prove that the transport density is absolutely continuous and belongs to L^p. The proofs that we present here follows essentially Santambrogio's research work [21], and originally proved through different techniques in [13].

Before proving this integrability theorems, we give some useful results.

Proposition 3.3.1. *Suppose that X and Y are compact metric spaces and that $c : X \times Y \to \mathbb{R}$ is continuous. Let us suppose that $\{\gamma_n\}_{n\geq 1} \subset \mathscr{P}(X \times Y)$ is a sequence which are optimal between their own marginals $\mu_n = (\pi_x)_\#\gamma_n$ and $\nu_n = (\pi_y)_\#\gamma_n$, and suppose that $\gamma_n \rightharpoonup \gamma$. Then, $\mu_n \rightharpoonup \mu = (\pi_x)_\#\gamma$ and $\nu_n \rightharpoonup \nu = (\pi_y)_\#\gamma$, and γ is an optimal transport plan between μ and ν.*

Proof. Let $\Gamma_n = \mathrm{spt}(\gamma_n)$. Then, up to a subsequence, we can suppose that $\Gamma_n \to \Gamma$ in the Hausdorff topology. By optimality of γ_n, we have that each Γ_n is a c-CM set. Now, we claim that the Hausdorff limit of c-CM sets is a c-CM set. Indeed, let us fix $(x_1, y_1), \ldots, (x_k, y_k) \in \Gamma$, then there are points $(x_1^n, y_1^n), \cdots, (x_k^n, y_k^n) \in \Gamma^n$ such that, for each $i = 1, \ldots, k$ we have that $(x_i^n, y_i^n) \to (x_i, y_i) \in \Gamma$.

The cyclical monotonicity of the sets Γ_n implies $\displaystyle\sum_{i=1}^k c(x_i^n, y_i^n) \leq \sum_{i=1}^k c(x_i^n, y_{\sigma(i)})$, and then we have that

$$\lim_{n\to\infty} \sum_{i=1}^k c(x_i^n, y_i^n) = \sum_{i=1}^k c(x_i, y_i) \leq \lim_{n\to\infty} \sum_{i=1}^k c(x_i^n, y_{\sigma(i)}) = \sum_{i=1}^k c(x_i, y_{\sigma(i)}),$$

with σ a permutation. Hence, Γ is a c-CM set. Finally, since $\gamma_n \rightharpoonup \gamma$ and $\Gamma_n \to \Gamma$ we may apply Proposition A.0.5 to obtain that $\mathrm{spt}(\gamma) \subset \Gamma$. Moreover, since c is continuous and X is compact, then γ is an optimal transport plan between of μ onto ν. □

In the following lemma, we shall prove that discrete measures are dense for the weak topology in $\mathscr{P}(\Omega)$. In particular, this result will allows us to approximate the marginal measure ν of an optimal transport plan by atomic measures in Theorem 3.3.5.

Lemma 3.3.2. *Let $\Omega \subset \mathbb{R}^d$ be a compact subset and $\mu \in \mathscr{P}(\Omega)$. Then, the set of Dirac measures of the form $\sum_{j=1}^n c_j\delta_{x_j}$ is dense for the weak convergence.*

Proof. Let $\epsilon > 0$, then there exist $N \in \mathbb{N}$ and point $x_1, x_2, \ldots, x_N \in \Omega$ such that $\Omega \subset \cup_{i=1}^{N} B(x_i, \epsilon)$ (by compactness). Let us define a partition of Ω as follows: take $K_1 = B(x_1, \epsilon)$, and $K_i = B(x_i, \epsilon) \setminus \cup_{r=1}^{i-1} K_r$, for $i = 2, \ldots, N$ and $K_1 = B(x_1, \epsilon)$. We also define the sequence $\mu^\epsilon = \sum_{i=1}^{N} \mu(K_i)\delta_{x_i}$.
We shall prove that $\mu^\epsilon \rightharpoonup \mu$ as $\epsilon \to 0$. Indeed, let $\phi \in C(\Omega)$. Since Ω is compact, then ϕ admits a modulus of continuity ω, that is, $|\phi(x) - \phi(y)| \leq \omega(|x - y|)$ and ω an increasing function such that $\lim_{tto0} \omega(t) = 0$. Since $\mathrm{diam}(K_i) \leq \epsilon$, then we get

$$\left| \int_\Omega \phi d\mu - \int_\Omega \phi d\mu^\epsilon \right| \leq \sum_{i=1}^{N} \int_{K_i} |\phi(x) - \phi(x_i)| d\mu(x) \leq \sum_{i=1}^{N} \int_{K_i} \omega(\mathrm{diam}(K_i)) \leq \omega(\epsilon),$$

which means that μ^ϵ weakly converge to μ as $\epsilon \to 0$. $\square$

Definition 3.3.3. *Let $\mu, \nu \in \mathscr{P}(\Omega)$ be two probability measures, $\gamma \in \Pi(\mu, \nu)$ and $\pi_t(x, y) = (1-t)x + ty$. We define the **standard interpolation measure** μ_t as, $\mu_t = (\pi_t)_\# \gamma$.*

Remark 3.3.4. From now on, we will assume that $d > 1$, since in the one dimensional case ($d = 1$) the transport density $\sigma \in BV(\Omega)$, which means that σ is essentially bounded (see Remark 3.2.8).

From now on, we assume that $\Omega \subset \mathbb{R}^d$ is a compact and convex domain, and $\mu, \nu \in \mathscr{P}(\Omega)$. The following theorem states L^1 regularity of the transport density σ.

Theorem 3.3.5. *Let $\Omega \subset \mathbb{R}^d$ is a convex domain, and $\mu \ll \mathscr{L}^d$. Let σ be the transport density associated with the transport plan of μ onto ν. Then, $\sigma \ll \mathscr{L}^d$.*

Proof. Let $\gamma \in \pi(\mu, \nu)$ be an optimal transport plan, and take $\sigma = \sigma_\gamma$ (the transport density). Consider $\mu_t := (\pi_t)_\# \gamma$, the standard interpolation between the two measures $\mu_0 = \mu$ and $\mu_1 = \nu$. In the previous section we saw that σ can be written as

$$\sigma = \int_0^1 (\pi_t)_\# (c\gamma) dt,$$

where c is the cost function $c(x, y) = |x - y|$. Since Ω is compact, then for all $\phi \in C(\Omega)$ we have

$$\langle \sigma, \phi \rangle = \int_0^1 \int_\Omega \phi(z) (\pi_t)_\# (c\gamma) dt \leq \|\phi\|_{L^\infty} \int_0^1 \mu_t dt.$$

Thus, there is a constant $C > 0$ such that

$$\sigma \leq C \int_0^1 \mu_t dt. \tag{3.3.1}$$

Hence, it is sufficient to prove that almost every measure μ_t is absolutely continuous to get the desired result. Indeed, if $\mathscr{L}^d(A) = 0$ and $\mu_t \ll \mathscr{L}^d$ then, $\sigma(A) \leq C \int_0^1 \mu_t(A) dt = 0$, which means that $\sigma \ll \mathscr{L}^d$. Now, we will show that $\mu_t \ll \mathscr{L}^d$ for $0 < t < 1$, to do so, we will

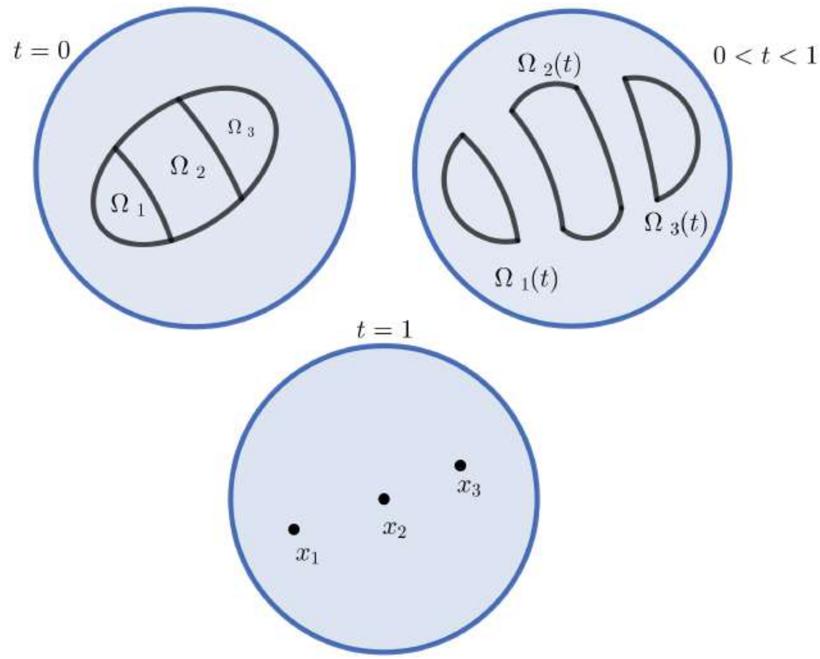

Figure 3.4: The sets $\Omega_i(t)$ of Theorem 3.3.5

prove first that the statement holds in the discrete case, and then we apply an approximation argument (by means of Lemma 3.3.2 and Proposition 3.3.1) to show the general case.

Let ν be a finitely atomic measure, with $\{x_i\}_{i=1}^N$ its atoms. Since $\mu \ll \mathscr{L}^d$, then γ is induced by a transport map T (the ray-monotone transport map), which is composed by N homotheties. Now, we want to quantify this absolute continuity in order to apply an approximation argument.

Let A be any Borel set, then there exists $\delta(\epsilon) > 0$ such that, $\mathscr{L}^d(A) < \delta$ implies $\mu(A) < \epsilon$. Note that the domain Ω is the disjoint union of a finite number of sets $\Omega_i = T^{-1}(\{x_i\})$; we will call $\Omega_i(t)$ the images of Ω_i through the map $x \mapsto (1-t)x + tT(x)$ (see figure 3.4). We claim that these sets are essentially disjoint. Indeed, if $z \in \Omega_i(t) \cap \Omega_j(t)$ with $i \neq j$, then two transport rays, $[x_i', x_i]$ and $[x_j', x_j]$, cross at z with $x_i' \in \Omega \in \Omega_i$ and $x_j' \in \Omega_j$. Since Kantorovich's potential is not differentiable at points where transport rays meet, then we must have that the five points x_i', x_j', z, x_i, x_j are collinear. But this implies that $z \in l_{x_i,x_j}$, where l_{x_i,x_j} is one of the lines connecting two atoms x_i and x_j. But we only have finitely many of these lines, then the set of such lines is $\mathscr{L}^d$-negligible. Hence, $\mathscr{L}^d(\Omega_i(t) \cap \Omega_j(t)) = 0$ for $i \neq j$ for $d > 1$. On the other hand, we also have

$$\mu_t(A) = \sum_{i=1}^N \mu_t(A \cap \Omega_i(t)) = \sum_{i=1}^N \mu\left(\frac{A \cap \Omega_i(t) - tx_i}{1-t}\right) = \mu\left(\bigcup_{i=1}^N \frac{A \cap \Omega_i(t) - tx_i}{1-t}\right),$$

since $\Omega_i(t)$ are essentially disjoint and $\gamma(\pi_t^{-1}(A \cap \Omega_i(t))) = \gamma\left(\pi_x^{-1}\left(\frac{A\cap\Omega_i(t)-tx_i}{1-t}\right)\right)$. Also, for every $i \in \{1, 2, \ldots, N\}$ we have

$$\mathscr{L}^d\left(\frac{A \cap \Omega_i(t) - tx_i}{1-t}\right) = \frac{1}{(1-t)^d}\mathscr{L}^d(A \cap \Omega_i(t)),$$

then

$$\mathscr{L}^d\left(\bigcup_{i=1}^{N}\frac{A\cap\Omega_i(t)-tx_i}{1-t}\right)\leq\frac{1}{(1-t)^d}\sum_{i=1}^{N}\mathscr{L}^d(A\cap\Omega_i(t))=\frac{1}{(1-t)^d}\mathscr{L}^d(A).$$

Hence, if we suppose that $\mathscr{L}^d(A)<(1-t)^d\delta(\epsilon)$ we get $\mathscr{L}^d\left(\bigcup_{i=1}^{N}\frac{A\cap\Omega_i(t)-tx_i}{1-t}\right)<\delta(\epsilon)$, and then $\mu_t(A)<\epsilon$.

We now show the general case. Indeed, since Ω is compact, then we can apply Lemma 3.3.2 to get a sequence of atomic measures $\{\nu^n\}_{n\geq 1}$, such that $\nu^n\rightharpoonup\nu$; let $\gamma^n\in\Pi(\mu,\nu^n)$ be the corresponding optimal transport plan. By Proposition 3.3.1, we also have that $\gamma^n\rightharpoonup\gamma\in\Pi(\mu,\nu)$ and $\mu_t^n\rightharpoonup\mu_t$, with $\mu_t=(\pi_t)_\#\gamma$. Let us now prove that the interpolation measures μ_t are absolutely continuous for all $t\in(0,t)$. As we did before, take a Borel set A such that $\mathscr{L}^d(A)<(1-t)^d\delta(\epsilon)$. Since $\mathscr{L}^d$ is a regular measure, then there is an open set B such that $A\subset B$ and $\mathscr{L}^d(B)<(1-t)^d\delta(\epsilon)$, which implies $\mu_t^n(B)<\epsilon$. Thus, by weak convergence and Theorem A.0.4 (lower semi-continuity on open sets), we have

$$\mu_t(A)\leq\mu_t(B)\leq\liminf_{n\to\infty}\mu_t^n(B)\leq\epsilon.$$

Therefore, $\mu_t\ll\mathscr{L}^d$ and then $\sigma\ll\mathscr{L}^d$ as well. □

Note: Form now on, we will understand $\|\mu\|_{L^p}$ as $\|f\|_{L^p(\Omega)}$, when $\mu=f\cdot\mathscr{L}^d$.

Theorem 3.3.6. *Let us suppose that $\mu=f\cdot\mathscr{L}^d$, with $f\in L^p(\Omega)$, where Ω is compact and convex. Then,*

(a) if $p<\bar{d}:=\frac{d}{d-1}$, the unique transport density σ associated with a transport plan of μ onto ν belongs to $L^p(\Omega)$;

(b) if $p\geq\bar{d}$, then $\sigma\in L^p(\Omega)$ for all $q<\bar{d}$.

Proof. We prove first **(a):** applying the same ideas from Theorem 3.3.5, we just need to prove that each measure μ_t with $t\in[0,1)$ belongs to L^p, and after that, we perform some estimations on their L^p norm. Let $\mu,\nu\in\mathscr{P}(\Omega)$ be probability measures, both belonging to $L^p(\Omega)$, and σ the unique transport density associated with the transport of μ onto ν. By (3.3.1) and Minkowski's inequality we have

$$\begin{aligned}\|\sigma\|_{L^p}=\left(\int_\Omega f(x)^p dx\right)^{\frac{1}{p}}&\leq\left[\int_\Omega\left(c_1\int_0^1 f_t(x)^p dt\right)^p dx\right]^{\frac{1}{p}}\\&\leq C\int_0^1\left[\int_\Omega f_t(x)^p dx\right]^{\frac{1}{p}}dt\\&=C\int_0^1\|\mu_t\|_{L^p}dt,\end{aligned}$$

with $\mu_t=f_t\cdot\mathscr{L}^d$. As in the previous theorem, we consider first the discrete case and then we use an approximation argument. Let us recall that each measure is absolutely continuous and its density must coincides on each set $\Omega_i(t)$ with the density of a homothetic image of μ

on Ω_i, with homothetic ratio $(1-t)$, that is, $f_t = (1-t)^{-d} f$. Then, we have

$$\begin{aligned}\int_\Omega f_t(x)^p dx = \sum_{i=1}^N \int_{\Omega_i(t)} f_t(x)^p dx &= \sum_{i=1}^N \int_{\Omega_i} \left(\frac{f(x)}{(1-t)^d}\right)^p (1-t)^d dx \\ &= (1-t)^{d(1-p)} \sum_{i=1}^N \int_{\Omega_i} f(x)^p dx \\ &= (1-t)^{d(1-p)} \int_\Omega f(x)^p dx. \end{aligned} \tag{3.3.2}$$

Thus, we obtain $\|\mu_t\|_{L^p} = (1-t)^{-\frac{d}{p'}}$, with $p' = \frac{p}{p-1}$.

We now prove that this inequality is also true for the general case. If ν is not atomic, we use the same approximation of Theorem 3.3.5, which means that we may approximate ν with atomic measures $\{\nu^n\}_{n\geq 1}$ to obtain that $\gamma^n \rightharpoonup \gamma$ and $\mu_t^n \rightharpoonup \mu_t$, where $\gamma^n \in \Pi(\mu, \nu^n)$ and $\gamma \in \Pi(\mu, \nu)$ are the associated optimal plans. Thus, by lower semi-continuity with respect to the weak convergence and equation (3.3.2) it follows that

$$\|\mu_t\|_{L^p} = \int_\Omega f_t(x)^p dx \leq \liminf_{n\to\infty} \int_\Omega f_t^n(x)^p dx = \liminf_{n\to\infty} \|\mu_t^n\|_{L^p} \leq (1-t)^{\frac{-d}{p'}} \|\mu\|_{L^p}.$$

Hence,

$$\|\sigma\|_{L^p} \leq C \int_0^1 \|\mu_t\|_{L^p} dt \leq C\|\mu\|_{L^p} \int_0^1 (1-t)^{-\frac{d}{p'}} dt.$$

Since the last integral is finite whenever $p' = \frac{p}{p-1} > d \iff p(d-1) > -d \iff p(d-1) < d$, which is equivalent to $p < \frac{d}{d-1} = \bar{d}$, then desired result follows.

To prove **(b):** we assume $p \geq \bar{d} = \frac{d}{d-1}$. Since the densities μ and μ_t belong to L^p, then $\mu_t, \mu \in L^q$ for all $q < p$. Thus, we can repeat estimations from (a) to get

$$\|\sigma\|_{L^q} \leq C \int_0^1 \|\mu_t\|_{L^q} dt \leq C\|\mu\|_{L^q} \int_0^1 (1-t)^{-\frac{d}{q'}} dt,$$

with $q' = \frac{q}{q-1}$ and $q < p$. Therefore, $\sigma \in L^q$ for all $q < p$. $\square$

The following example shows the sharpness of the bound on p that we set in Theorem 3.3.6.

Example 3.3.7. We give an example where, even if $\mu \in L^\infty$, the transport density σ does not belong to $L^{\bar{d}}$.

Consider $\mu = f \cdot \mathscr{L}^d$ with $f = \chi_A$ (the characteristic function on A) and the annulus

$$A = B(0,R) \setminus B\left(0, \frac{R}{2}\right),$$

with R chosen such that $\int f(x)dx = 1$, and take $\nu = \delta_0$ (see figure 3.5). Since $c(x,y) = |x-y|$, then $\frac{R}{2}\gamma \leq c\gamma \leq R\gamma$, which implies that the integrability of $\sigma = \displaystyle\int_0^1 (\pi_t)_\#(c\gamma)dt$ is the same as that for $\displaystyle\int_0^1 (\pi_t)_\#(\gamma)dt = \int_0^1 \mu_t dt$.

On the other hand, we saw that $\mu_t = f_t \cdot \mathscr{L}^d$ and its density must coincide on each set $A_i(t)$ with the density of a homothetic image of μ on A, with ratio $(1-t)$, then we have $f_t = (1-t)^{-d}\chi_{(1-t)A}$ (since we need $\int f_t(x)dx = 1$). Then,

$$\int_0^1 f_t(x)dt = \int_0^1 \frac{1}{(1-t)^d}\chi_{(1-t)A}(x)dt = \int_{\frac{1}{R}|x|}^{\frac{2}{R}|x|} \frac{1}{s^d} ds = \frac{2}{R}|x|^{1-d} - \frac{1}{R}|x|^{1-d} = \frac{1}{R}|x|^{1-d}$$

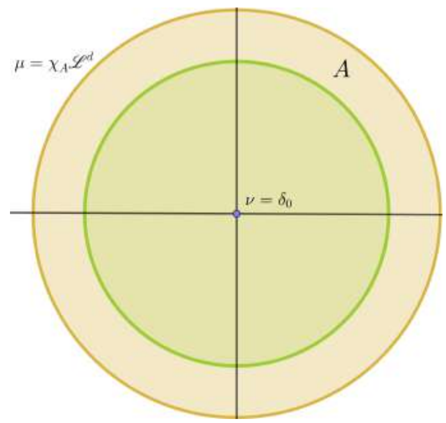

Figure 3.5: μ and ν in Example 3.3.7.

for $|x| \leq \frac{R}{2}$, thanks to the change of variable $s = 1 - t$. Thus, the function $|x|^{1-d}$ belongs to L^p in a neighborhood of zero if $(1-d)p + d > 0 \Leftrightarrow (1-d)p + (d-1) > -1 \Leftrightarrow p - dp + d - 1 > -1 \Leftrightarrow dp - d - p + 1 = (d-1)(p-1) < 1$, which means that this function belong to L^p only when $p < \bar{d}$.

Appendix A

Preliminary results

We established some definitions and results (meanly from measure theory) that were used throughout this work.

Theorem A.0.1 (Banach-Alaoglu). *Let X a Banach space which is separable and let $\{\phi_n\}_{n\geq 1}$ a bounded sequence in X^*(the dual space of X), then there is a subsequence $\{\phi_{n_k}\}_{k\geq 1}$ such that ϕ_{n_k} converges weakly to some $\phi \in X$.*

We denote by $\mathscr{M}(X)$ the set of finite measures on X, to such measures, we associate its total variation measure $|\lambda| \in \mathscr{M}_+(X)$ through:

$$|\lambda|(A) := \sup\left\{\sum_{i=1}^{\infty} |\lambda(A_i)| : A = \cup_{i=1}^{\infty} A_i, \;\; A_i \cap A_j = \emptyset \,, i \neq j\right\}$$

Theorem A.0.2 (Riesz representation Theorem). *Suppose that X is a separable and locally compact metric space, and consider $C_0(X)$. Then for all $\psi \in C_0(X)^*$ there exist a unique $\lambda \in \mathscr{M}(X)$ such that $\langle \psi, \phi \rangle = \int_X \phi \, d\lambda$ for every $\phi \in C_0(X)$. Moreover, $C_0(X)^*$ is isometrically isomorphic to $\mathscr{M}(X)$ endow with the norm $\|\lambda\| = |\lambda|(X)$.*

Theorem A.0.3 (Weierstrass). *If $f : X \to \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous and X is compact, then there exists $\bar{x} \in X$ such that $f(\bar{x}) = \inf\{f(x) : x \in X\}$.*

Theorem A.0.4. *Let $\{\mu_n\}_{n\geq 1}$ be a sequence of Borel probability measures and $\mu \in \mathscr{P}(\Omega)$. Then the following conditions are equivalent:*

- $\mu_n \rightharpoonup \mu$.
- *For every close set F, one has $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$.*
- *For every open set U, one has $\mu(U) \leq \liminf_{n\to\infty} \mu_n(U)$.*

Proposition A.0.5. *Let X be a compact metric space, and d_H the Hausdorff distance among compact subsets of X. If $d_H(A_n, A) \to 0$, and $\{\mu_n\}_{n\geq 1}$ is a sequence of positive measures such that spt$(\mu_n) \subset A_n$ and $\mu_n \rightharpoonup \mu$, then spt$(\mu) \subset A$.*

Definition A.0.6. *A sequence $\{\mu_n\}_{n\geq 1}$ of probability measures is said to be* ***tight*** *if for every $\epsilon > 0$, there is a compact subset $K \subset X$ so that $\mu_n(X \setminus K) < \epsilon$ for every $n \in \mathbb{N}$*

The following theorem gives us a characterization of sequential compactness in $\mathscr{M}(X)$ with respect to the narrow topology.

Theorem A.0.7 (Prokhorov)**.** *Let $\{\mu_n\}_{n\geq 1}$ be a tight sequence of probability measures over a Polish space (a metrizable, complete and separable space) X. Then, there exist $\mu \in \mathscr{P}(X)$ and a subsequence $\{\mu_{n_k}\}_{k\geq 1}$ such that $\mu_{n_k} \rightharpoonup \mu$ (converge in duality with $C_b(X)$). Conversely, every sequence which satisfies $\mu_n \rightharpoonup \mu$ is necessarily tight.*

We now introduce the notion of disintegration of measures, which is widely use in the context of optimal transportation theory [6].

Definition A.0.8. *Let X be a measure space endow with a Borel measure μ and a map $f : X \to Y$, with Y a topological space. We say that a family $(\mu_y)_{y\in Y}$ is a **disintegration of** μ **according to** f, if every μ_y is concentrated on $f^{-1}(\{y\})$, and for every test function $\phi \in C(X)$, the map $y \mapsto_X \phi d\mu_y$ is Borel measurable and*

$$\int_X \phi d\mu = \int_Y d\nu(y) \int_X \phi d\mu_y,$$

where $\nu = f_{\#}\mu$. The disintegration of μ is usually denoted by $\mu := \mu_y \otimes \nu$.

Theorem A.0.9 (Disintegration)**.** *Let X, Y be two Polish spaces, and let $f : X \to Y$ be a Borel map, and $\mu\mathscr{M}(Y)$ a positive finite measure; let us consider $\nu = f_{\#}\mu$. Then, there exists a disintegration $(\mu_y)_{y\in Y}$ of μ according to f.*

Now, we give a quick reminder about vector measures.

Definition A.0.10. *A finite **vector measure** $\lambda : \mathscr{B}(\Omega) \to \mathbb{R}^d$ is map associating to any Borel subset $A \subset \Omega$ a vector $\lambda(A) \in \mathbb{R}^d$ such that for every disjoint countable union $A = \cup_{i=1}^{\infty} A_i$, with $A_i \cap A_j = \emptyset$ for $i \neq j$, we have that*

$$\sum_{i=1}^{\infty} |\lambda(A_i)| < +\infty, \text{ and } \lambda(A) = \sum_{i=1}^{\infty} \lambda(A_i),$$

where the second sum means a sum over each coordinate in the vector measure. The integral of a Borel function $\xi : \Omega \to \mathbb{R}^d$ with respect to λ is well defined if $|\xi| \in L^1(\Omega, |\lambda|)$, and is denoted $\int_\Omega \xi \cdot d\lambda$ and can be compute as $\sum_{i=1}^{d} \int_\Omega \xi_i d\lambda_i$.

Proposition A.0.11 (Polar decomposition)**.** *for every $\lambda \in \mathscr{M}^d(\Omega)$ there exists a Borel function $u : \Omega \to \mathbb{R}^d$ such that $\lambda = u|\lambda|$ and $|u| = 1$ $|\lambda|$-a.e. In particular, we have $\int_\Omega \xi \cdot d\lambda = \int_\Omega (\xi \cdot u) d|\lambda|$. This is also know as the polar decomposition of λ.*

Theorem A.0.12 (Divergence Theorem)**.** *Let $\Omega \subset \mathbb{R}^d$ be a compact domain with piece-wise smooth boundary. If F is a C^1 vector field defined on a neighborhood of Ω, then*

$$\int_\Omega \nabla \cdot F dx = \int_{\partial\Omega} F \cdot \hat{n} d\mathscr{H}^{d-1}(x).$$

Now, we state a well-known theorem in convex analysis due to Rockafellar.

Theorem A.0.13 (Rockafellar)**.** *Let $\Gamma \subseteq \mathbb{R}^n \times \mathbb{R}^n$. Then, the following statement is equivalent*

- *There exists a a proper lower semi-continuous and convex function $\varphi : \mathbb{R}^n \to \mathbb{R}$ such that Γ is contained in the graph of $\partial\varphi$, namely if $(x, y) \in \Gamma$ then $y \in \partial\varphi(x)$.*
- *Γ is cyclically-monotone.*

Bibliography

[1] L. Ambrosio, Lecture notes on optimal transport problems, in Mathematical Aspects of Evolving Interfaces. Lecture Notes in Mathematics (1812) (Springer, New York, 2003), pp. 1–52 iii, 1, 27, 45

[2] L. Ambrosio, N. Fusco, D. Pallara, Functions of Bounded Variation and Free Discontinuity Problems (Oxford University Press, Oxford, 2000) 44

[3] L. Ambrosio, P. Tilli, Topics on Analysis in Metric Spaces. Oxford Lecture Series in Mathematics and its Applications, vol. 25 (Oxford University Press, Oxford, 2004) 18

[4] L. Ambrosio, N. Gigli, G. Savaré, Gradient Flows in Metric Spaces and in the Spaces of Probability Measures. Lectures in Mathematics, ETH Zurich (Birkhäuser, Basel, 2005) iii

[5] M. Beckmann, A continuous model of transportation. Econometrica 20, 643–660 (1952) iii, 49

[6] P. Billingsley, Convergence of Probability Measures, Wiley Series in Probability and Statistics (2013) 6, 50, 73

[7] A. Braides, Γ-Convergence for Beginners (Oxford University Press, Oxford, 2002)

[8] Y. Brenier, The least action principle and the related concept of generalized flows for incompressible perfect fluids. J. Am. Mat. Soc. 2, 225–255 (1989) iii

[9] Y. Brenier, Décomposition polaire et réarrangement monotone des champs de vecteurs. (French) C. R. Acad. Sci. Paris Sér. I Math. 305(19), 805-808 (1987) iii, 24

[10] Y. Brenier, The least action principle and the related concept of generalized flows for incompressible perfect fluids. J. Am. Mat. Soc. 2, 225-255 (1989) iii

[11] G. Buttazzo, F. Santambrogio, A mass transportation model for the optimal planning of an Urban region. SIAM Rev. 51(3), 593–610 (2009) iii, 49

[12] G. Carlier, F. Santambrogio, A variational model for urban planning with traffic congestion. ESAIM Control Optim. Calc. Var. 11(4), 595–613 (2005) 49

[13] L. De Pascale, A. Pratelli, Regularity properties for monge transport density and for solutions of some shape optimization problem. Calc. Var. Part. Differ. Equat. 14(3), 249–274 (2002) 66

[14] EVANS, L. C. Current developments in mathematics. Partial differential equations and monge-kantorovich mass transfer, journal, p. 65–126, 1997.

[15] L.C. Evans, Partial Differential Equations. Graduate Studies in Mathematics (American Mathematical Society, Providence, 2010)

[16] L.C. Evans, W. Gangbo, Differential equations methods for the Monge-Kantorovich mass transfer problem. Mem. Am. Math. Soc. 137(653) (1999) iii

[17] V. Levin, Optimal solutions of the Monge problem. In: Kusuoka S., Maruyama T. (eds) Advances in Mathematical Economics. Advances in Mathematical Economics, vol 6. Springer, Tokyo (2004). iii, 1

[18] G. Monge, Mémoire sur la théorie des déblais et des remblais. Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année, 666–704, (1781) iii, 1

[19] J. Lott, C. Villani, Ricci curvature for metric-measure spaces via optimal transport. Ann. Math. 169, 903–991 (2009) iii

[20] W. Rudin, Real and Complex Analysis, 3rd edn. (McGraw-Hill, New York, 1986) 4

[21] F. Santambrogio, Absolute continuity and summability of transport densities: simpler proofs and new estimates. Calc. Var. Par. Differ. Equat. 36(3), 343–354 (2009) iii, 47, 66

[22] F. Santambrogio, A Dacorogna-Moser approach to flow decomposition and minimal flow problems. ESAIM: Proc. Surv. (SMAI 2013) 45, 265–174 (2014) 47, 49, 56

[23] F. Santambrogio, Optimal Transport for Applied Mathematicians, in Progress in Nonlinear Differential Equations and Their Applications 87, Birkhäuser Basel (2015) iii, 1, 12, 17, 20, 21, 27, 47, 49

[24] S.K. Smirnov, Decomposition of solenoidal vector charges into elementary solenoids, and the structure of normal one-dimensional flows. Algebra i Analiz 5(4), 206–238 (1993). Later translated into English in St. Petersburg Math. J. 5(4), 841–867 (1994) iii, 62

[25] V.N. Sudakov, Geometric problems in the theory of infinite-dimensional probability distributions. Cover to cover translation of Trudy Mat. Inst. Steklov 141 (1976). Proc. Steklov Inst. Math. 2(i-v), 1-178 (1979) iii, 1, 27

[26] C. Villani, Topics in Optimal Transportation. Graduate Studies in Mathematics (American Mathematical Society, Providence, 2003) 1, 4, 7, 12, 20, 21

[27] C. Villani, Optimal Transport: Old and New (Springer, New York, 2008)