On the Optimal Transport and its Regularity

Geoffrey Gacemi Ngeru (geoffreygm@aims.ac.za)
African Institute for Mathematical Sciences (AIMS)

Supervised by: Dr. Marc Sedjro
AIMS, South Africa

14^{th} May 2020

Submitted in partial fulfillment of a structured masters degree at AIMS South Africa

Abstract

In this work, we discuss first the Monge's problem and important steps leading to the existence of solutions. Next, we formulate the Brenier theorem which not only solves the Monge's problem for a quadratic cost function but also provides a link to a class of the Monge-Ampere equations. Finally we study the regularity of Brenier solution.

Declaration

I, the undersigned, hereby declare that the work contained in this research project is my original work, and that any work done by others or by myself previously has been acknowledged and referenced accordingly.

Ngeru Geoffrey Gacemi, 14^{th} May 2019.

Contents

1. Introduction

The topic of optimal transport started a long time ago in France around 1781 with a paper by Gaspard Monge [see Monge (1781)]. Monge was interested in moving a pile of sand into a hole in such a way that the work done is economical. In other words, he was interested in minimizing the cost of transporting the pile of sand into the hole.

Over the years, researchers from different fields of Mathematics have taken an interest in the theory of optimal transport as they have come to realise its unexpected connection with their respective fields of research.

To describe the optimal mass transport problem mathematically one models the distributions of the mass within the pile of sand and the hole respectively by probability measures μ (source measure) and ν (target measure). The transport is performed through a transport map T at a cost c so that the cost of transporting a unit mass sitting at point x is given by $c(x, T(x))$. The Monge's problem consists in minimizing the total cost of transporting μ to ν among all the possible transport map T. Figure 1.1 illustrates how the pile of sand is transported to the hole.

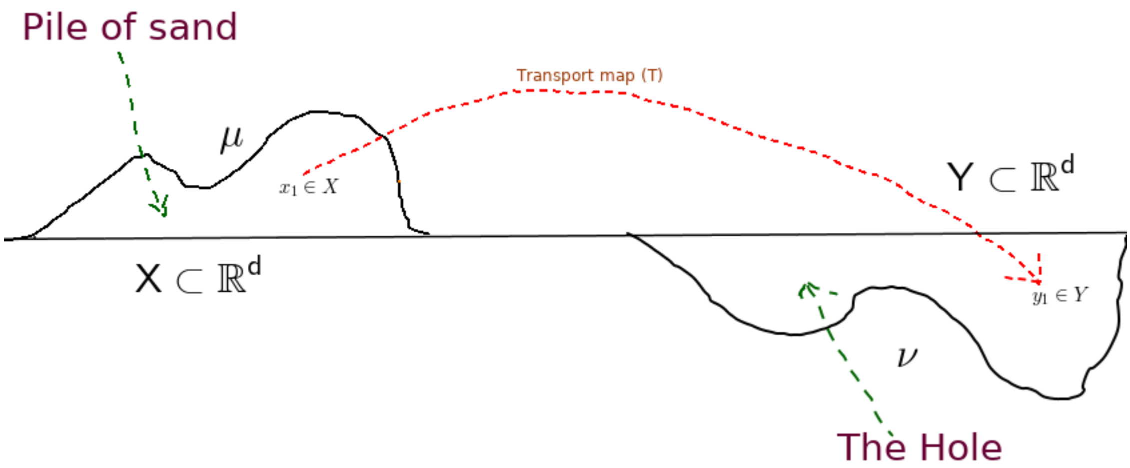

Figure 1.1: Mass Transportation Problem.

The Monge's problem remained a challenge in Calculus of variations until 1987 when Brenier solved the problem [see Brenier (2004)]. However, Brenier solution was obtained under some constraints on the cost function and the source measure, more precisely, the cost function is quadratic and the source measure is absolutely continuous with Lebesgue measure. In this case, the optimal transport is given by the gradient of a convex function.

Brenier also gave an interesting interaction between optimal transports and a class of the Monge-Ampere equations. This interaction shows that the convex potentials that arise from the Brenier solutions solve the Monge-Ampere equations in some weak sense. We note that, Monge-Ampere equations are highly nonlinear and therefore, hard to solve.

In this work, we shall explain the steps taken by Brenier to solve the Monge's problem. Moreover, we study the regularity theory of the convex potential arising in the Brenier theorem. In Caffarelli (1992b) Caffarelli oulined arguments that illustrates the reason why the Brenier solutions is not in general continuously differentiable even if the source and target measures have a smooth densities. Here, we have revisited Caffarelli argument and provided a detailed computations supporting such an argument. The regularity theory for the Brenier solution is important as it contributes, among other things, to the understanding of some problems in Meteorology notably Axisymmetric flows [see Sedjro (2012)] and semigeostrophic equations see [Shutts (1991) and Pisante et al. (2007)].

This work is organized in the following way: In chapter two, we provide some definitions, notation, and recall important theorems. In chapter three, we introduce the Monge's problem, its relaxation the Kantorovich's problem and dual formulation. Chapter three culminates in the formulation of the Brenier theorem. In chapter four, we study the regularity of potential arising in the Brenier theorem.

2. Notation, Definitions and Preliminaries

2.1 Convexity Analysis

In this section, we recall some important concepts in convex analysis. For more details in complex analysis see Rockafellar (1970).

2.1.1 Definition (Convex set)**.** A subset S of $\mathbb{R}^d$ is convex if for every $x, y \in S$ and for any $\lambda \in [0,1]$, then we have

$$\lambda x + (1-\lambda)y \in S.$$

2.1.2 Definition (Convex function)**.** Let X be a convex subset of $\mathbb{R}^d$. The function $f : X \longrightarrow \mathbb{R}$ is convex if for any $x_1, x_2 \in X$ and for any $\lambda \in [0,1]$, then

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2).$$

We say f is strictly convex if for any $(x_1, x_2) \in X$ such that $x_1 \neq x_2$ and any $\lambda \in [0,1]$, we have

$$f(\lambda x_1 + (1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda)f(x_2).$$

2.1.3 Definition (Test of convexity)**.** (i) Assume X is open and convex subset of $\mathbb{R}^d$. If f is twice differentiable then f is a convex on X if and only if

$$D^2 f(x) \geq 0, \tag{2.1.1}$$

for all $x \in X$. Note that $D^2 f(x)$ in equation 2.1.1 is given by the Hessian matrix whose entries are

$$[D^2 f(x)]_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j},$$

$\forall i, j = 1, \dots, n$. Equation 2.1.1 means that the Hessian matrix is positive definite at each point $x \in X$.

(ii) Note that, If f and g are convex on X then their sum $f + g$ is also convex on X.

2.1.4 Definition (Subdifferential of convex function)**.** Let S be a non-empty convex subset $\mathbb{R}^d$ and $u : S \longrightarrow \mathbb{R}$ a convex function. The subdifferential of u at $x_0 \in S$, denoted by, $\partial^0 u(x_0)$ is given by

$$\partial^0 u(x_0) := \left\{ p \in \mathbb{R}^d : u(y) - u(x_0) \geq p.(y - x_0) \quad , \forall y \in S \right\}. \tag{2.1.2}$$

2.1.5 Remark. (i) The subdifferential is always non-empty convex set.

(ii) If u is differentiable at x_0 then the subdifferential of u at x_0 is reduced to a singleton, namely, $Du(x_0)$.

2.2 Measure Theory

Let Σ be a $\sigma-$algebra on $\mathbb{R}^d$. Confer Cannarsa and D'Aprile (2015) for definition of $\sigma-$algebra.

2.2.1 Definition (Definition of a measure)**.** A function $\mu : \Sigma \longrightarrow \mathbb{R}_+$ is called a measure if it satisfies the following properties

$$i) \quad \mu(\emptyset) = 0 \qquad \text{Null empty set} \tag{2.2.1}$$

$$ii) \quad \mu(E) \geq 0 \quad \forall E \in \Sigma \qquad \text{Non-negativity} \tag{2.2.2}$$

$$iii) \quad \mu(\cup_{k=1}^{\infty} E_k) = \sum_{k=1}^{\infty} \mu(E_k) \qquad \text{Countable additivity} \tag{2.2.3}$$

provided that $\{E_k\}$ are disjoint.

2.2.2 Remark. If Σ is a Borel $\sigma-$algebra on $\mathbb{R}^d$ induced by the Euclidean distance, then the measure $\mu : \Sigma \longrightarrow \mathbb{R}_+$ is said to be Borel measure. In addition, if $\mu(\mathbb{R}^d) = 1$ then μ is called a Borel probability measure. We denote $\mathcal{P}(\mathbb{R}^d)$ as a collection of all Borel probability measures on $\mathbb{R}^d$.

2.2.3 Definition (Lebesgue measure)**.** We denote $\mathscr{L}^d$ as a d-dimensional Lebesgue measure on $\mathbb{R}^d$. Confer Cannarsa and D'Aprile (2015) for the definition of Lebesgue measure and its properties.

2.2.4 Definition (Singular measure)**.** Let $\Omega \subset \mathbb{R}^d$. Two measures μ and ν defined on a measurable space (Ω, Σ) are said to be singular if there exists two disjoint sets $A, B \in \Sigma$ whose union is Ω (that is $A \cup B = \Omega$) such that $\mu(B) = \nu(A) = 0$.

2.2.5 Definition (Absolutely continuous measure)**.** Let μ and ν be two measures on a $\sigma-$algebra Σ we say that ν is absolutely continuous with respect to μ if $\nu(A) = 0$ for any $A \in \Sigma$ such that $\mu(A) = 0$.

2.2.6 Theorem (Lebesgue decomposition theorem)**.** *Let μ and ν be $\sigma-$finite measures on a measurable space $(\mathbb{R}^d, \Sigma)$. Then, ν can be uniquely decomposed into $\nu = \nu_c + \nu_s$ where ν_c is absolutely continuous with respect to μ and ν_s and μ are singular.*

2.2.7 Definition (Almost everywhere property)**.** Let $(\mathbb{R}^d, \Sigma, \mu)$ be a measurable space, a property $\mathcal{P}$ is said to $\mu-$hold almost everywhere in $\mathbb{R}^d$ if there exists a set $A \in \Sigma$ such that $\mu(A) = 0$ and property $\mathcal{P}$ holds on $x \in \mathbb{R}^d \setminus A$. For simplicity, we write $\mathcal{P}$ holds μ a.e.

2.2.8 Definition (Monge-Ampere measure)**.** Let $\Omega \subset \mathbb{R}^d$ be a non-empty open convex domain and let u be a convex function on Ω. We define the Monge-Ampere measure associated to μ by the following

$$\mu_u(E) = \mathscr{L}^d(\partial^0 u(E)) \quad \text{for every Borel set} \quad E \subset \Omega, \tag{2.2.4}$$

where

$$\partial^0 u(E) := \cup_{x \in E} \partial u(x). \tag{2.2.5}$$

Let $f : \Omega \times \mathbb{R} \times \mathbb{R}^d \longrightarrow \mathbb{R}_+$ be a function. A convex function $u : \Omega \longrightarrow \mathbb{R}$ is called an Alexandrov solution to the Monge-Ampere equation

$$\det D^2 u = f(x, Du) \quad \text{in} \quad \Omega, \tag{2.2.6}$$

if

$$\mu_u(E) = \int_E f(x, Du)dx \qquad (2.2.7)$$

holds for every Borel set $E \subset \Omega$ and the integral in equation 2.2.7 is well defined. [See in De Philippis and Figalli (2014)].

2.3 Point-Set Topology

We denote $B_r(a)$ as the open ball centred at a with radius r with respect to the euclidean distance on $\mathbb{R}^d$. For simplicity, we shall denote $B_1(0)$ by B_1.

2.3.1 Definition (Opend and closed sets)**.** Let A be a subset of $\mathbb{R}^d$. We say that A is open set on $\mathbb{R}^d$ if it can be writen as a union of open balls on $\mathbb{R}^d$.

Set A is closed if its compliment is open.

2.3.2 Definition (Compact set)**.** Let X be a subset of $\mathbb{R}^d$. We say that X is bounded if there exist an open ball B_r such that $X \subset B_r$ for some positive r. A set that is bounded and closed is called compact set.

2.3.3 Definition (Minkowski sum)**.** Let P and Q be two subsets of $\mathbb{R}^d$. The Minkowski sum of P and Q, denoted by $P \oplus Q$, is the set of points given by:

$$P \oplus Q := \{p + q : p \in P, \quad q \in Q\}.$$

2.4 Functional Spaces

2.4.1 Definition (Lipshitz continous function)**.** Let $S \subset \mathbb{R}^d$ a function $f : S \longrightarrow \mathbb{R}$ is Lipshitz continuous function if there exists a constant $C > 0$ such that

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

for all $x, y \in S$.

2.4.2 Theorem (Radamacher's theorem)**.** *Let Ω be a subset of $\mathbb{R}^d$, if a function $f : \Omega \longrightarrow \mathbb{R}$ is Lipshitz continuous, then f is differentiable almost everywhere in Ω.*

2.4.3 Definition (Distribution)**.** A distribution T is a linear mapping $T : \mathscr{D}(\mathbb{R}^d) \longrightarrow \mathbb{R}$, with the following properties

- $T(\varphi_1 + \varphi_2) = T(\varphi_1) + T(\varphi_2)$, for all $\varphi_1, \varphi_2 \in \mathscr{D}(\mathbb{R}^d)$.
- $T(\lambda\varphi) = \lambda T(\varphi)$, for all $\varphi \in \mathscr{D}(\mathbb{R}^d)$ and $\lambda \in \mathbb{R}$.
- If $\varphi_j \longrightarrow \varphi \in \mathscr{D}(\mathbb{R}^d)$ then $T(\varphi_j) \longrightarrow T(\varphi)$.

Here $\mathscr{D}(\mathbb{R}^d)$ is a topological vector space consisting of $C_c^\infty(\mathbb{R}^d)$ equiped with the topology that corresponds to convergence in the sense of test functions. Note that, we denote the value of T acting on a test function φ by $< T, \varphi >$ and T is positive if $\varphi \geq 0$ implies that $< T, \varphi > \geq 0$. [See Van Dijk (2013)].

The following theorem can be found in Van Dijk (2013).

2.4.4 Theorem. *A positive distribution is a positive measure.*

2.4.5 Remark. Let $f \in L^1_{loc}(\mathbb{R}^d)$. Then, T_f defined by

$$< T_f, \varphi >= \int_{\mathbb{R}^d} f\varphi dx \quad \text{for all} \quad \mathscr{D}(\mathbb{R}^d), \tag{2.4.1}$$

is a distribution on $\mathbb{R}^d$.

2.4.6 Definition (Distributional derivative)**.** Let T be a distribution on $\mathbb{R}^d$. Then, the derivative $\partial_{x_i} T$ is a distribution defined by

$$< \partial_{x_i} T, \varphi >= - < T, \partial_{x_i} \varphi >,$$

for any $1 \leq i \leq d$ and for any test function $\varphi \in \mathscr{D}(\mathbb{R}^d)$. [See Van Dijk (2013)].

3. Optimal Transport

In this section we introduce the theory of optimal mass transport as formulated in the Monge's problem. We will state the Kantorovich's problem and its duality that leads to a fundamental result regarding the Monge's problem, namely, the Brenier theorem.

3.1 Monge's Problem

Let $c : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}_+$ be a continous non-negative cost function, let μ and ν be elements of $\mathcal{P}(\mathbb{R}^d)$. and let $T : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ be a Borel function, we say that T transports μ to ν denoted by $T\#\mu = \nu$ if

$$\nu(B) = \mu(T^{-1}(B)) \tag{3.1.1}$$

holds for all Borel set $B \in \mathbb{R}^d$. Equivalently,

$$\int_{\mathbb{R}^d} (\varphi \circ T) d\mu(x) = \int_{\mathbb{R}^d} \varphi(y) d\nu(y), \tag{3.1.2}$$

for all $\varphi \in L^1(d\nu)$.

The Monge's problem consists in minimising the functional associated with the cost function over all the sets of Borel functions T such that T pushes μ onto ν. The problem is as follows:

$$\text{minimise} \quad I[T] = \int_X c(x, T(x)) d\mu(x), \tag{3.1.3}$$

among all Borel function T such that $T\#\mu = \nu$.

We observe that the Monge's problem is not always well-posed. For instance, if we consider the discrete case where $\mu = \delta_{x_1}$ and $\nu = \frac{1}{2}\delta_{y_1} + \frac{1}{2}\delta_{y_2}$ with y_1 different from y_2, the map T such that $T\#\mu = \nu$ does not exist. Furthermore, the constraint $T\#\mu = \nu$ is in general highly nonlinear. In 1942, Kantorovich proposed a relaxation of Monge's problem [see Kantorovich (1942)].

3.2 Kantorovich's Problem

Before we state the Kantorovich's problem, we define the set of transport plans. Let $\Gamma(\mu, \nu)$ be the set of $\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ such that

$$\gamma[A \times \mathbb{R}^d] = \mu[A] \quad \text{and} \quad \gamma[\mathbb{R}^d \times B] = \nu[B], \tag{3.2.1}$$

holds for any Borel set A and B in $\mathbb{R}^d$.

3.2.1 Remark. (i) All γ that satisfy condition 3.2.1, are said to have a first marginal μ and second marginal ν.

(ii) The condition 3.2.1 is equivalent to

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \varphi(x) d\gamma(x, y) = \int_{\mathbb{R}^d} \varphi(x) d\mu(x) \quad \text{and} \quad \int_{\mathbb{R}^d \times \mathbb{R}^d} \psi(y) d\gamma(x, y) = \int_{\mathbb{R}^d} \psi(y) d\nu(y) \tag{3.2.2}$$

for any $\varphi, \psi \in L^1$.

Next, we state the Kantorovich's problem:

$$\text{minimise} \quad K(\gamma) = \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y) d\gamma(x,y) \tag{3.2.3}$$

over $\gamma \in \Gamma(\mu, \nu)$. What makes this formulation more interesting is the fact that the Kantorovich's problem is linear and easier to solve, unlike Monge's problem. This is because the Kantorovich problem allows for the split of mass. Moreover, the Kanatorovich's problem is well-posed since the tensor product $\mu \otimes \nu$ always lies in $\Gamma(\mu, \nu)$ and $\Gamma(\mu, \nu)$ is compact with respect to narrow topology. The minimizer of problem 3.2.3 is called optimal transport plan.

Note that the solution of the Kantorovich's problem always exists see Thorpe (2018). That is, there exist γ^* such that

$$K[\gamma^*] \leq K[\gamma] \tag{3.2.4}$$

for all $\gamma \in \Gamma(\mu, \nu)$. The Kantorovich's problem is connected to the Monge's problem. To see that, we first prove the following claim.

Claim.

For any Borel function T

$$T\#\mu = \nu \quad \text{is equivalent to} \quad \gamma_T \in \Gamma(\mu, \nu), \tag{3.2.5}$$

with $\gamma_T := (\text{id} \times T)\#\mu$, moreover $K[\gamma_T] = I[T]$.

Proof of the claim. In right of condition 3.2.2 γ_T to belong to $\Gamma(\mu, \nu)$, if and only if

$$\int_{\mathbb{R}^d} [\varphi(x) + \psi \circ T(x)] d\mu(x) = \int_{\mathbb{R}^d} \varphi(x) d\mu(x) + \int_{\mathbb{R}^d} \psi(y) d\nu(y), \tag{3.2.6}$$

holds for $\varphi \in L^1(d\mu)$ and $\psi \in L^1(d\nu)$. Cancelling $\int \varphi d\mu$ on both sides of equation 3.2.6, we have

$$\int_{\mathbb{R}^d} (\varphi \circ T) d\mu(x) = \int_{\mathbb{R}^d} \varphi(y) d\nu(y) \tag{3.2.7}$$

for all $\varphi \in L^1(d\nu)$. Equation 3.2.7 means that

$$T_{\#}\mu = \nu.$$

Furthermore,

$$\begin{aligned} K[\gamma_T] &= \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y) d\gamma_T \\ &= \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y) d(\text{id} \times T)_{\#\mu} \\ &= \int_{\mathbb{R}^d} c \circ (\text{id} \times T) d\mu(x) \\ &= \int_{\mathbb{R}^d} c(x, T(x)) d\mu(x) \\ &= I[T]. \end{aligned}$$

□

Next assume that the minimiser $\gamma*$ is obtained 3.2.4 is induced by a map, that is, $\gamma* = (\text{id} \times T^*)\#\mu$. Then, in line with the claim above we have

$$I[T^*] = K[\gamma*] \leq K[\gamma_T] = I[T] \tag{3.2.8}$$

for all Borel function T such that $T\#\mu = \nu$. As $T^*\#\mu = \nu$ Equation 3.2.8 implies that T^* is the minimizer of the Monge's problem equation 3.1.3

The following example 3.2.2 is one of the cases where the Monge's problem the Kantorovich's problem concide.

3.2.2 Example (Dirac mass)**.** Assume that ν is a dirac mass $\nu = \delta_a$. Then, $\Gamma(\mu, \nu)$ is reduced to a singleton $\{\gamma_0\}$. In this case, $\{\gamma_0\}$ is the trivial optimal transport plan given by $\mu \otimes \delta_a$. Therefore

$$\begin{aligned}
K[\gamma_0] &= K[\mu \otimes \delta_a] &(3.2.9)\\
&= \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d(\mu(x) \otimes \delta_a(y)), &(3.2.10)\\
&= \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\delta_a(y) \otimes d\mu(x), &(3.2.11)\\
&= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} c(x, y) d\delta_a(y) \right) d\mu(x), &(3.2.12)\\
&= \int_{\mathbb{R}^d} c(x, a) d\mu(x). &(3.2.13)
\end{aligned}$$

On the other hand, when $\nu = \delta_a$, the set of Borel function T such that $T\#\mu = \nu$ is reduced to T_0 where $T_0(x) = a$ μ a.e. it follows that T_0 is the optimal map and

$$I(T_0) = \int_{\mathbb{R}^d} c(x, a)) d\mu(x). \tag{3.2.14}$$

Therefore,

$$K[\gamma_0] = I[T_0].$$

The Kanatorovich's problem has an interesting dual formulation which can be found in Villani (2003b).

3.2.3 Theorem (Kantorovich's duality)**.** *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\nu \in \mathcal{P}(\mathbb{R}^d)$, and let $c : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow [0, +\infty)$ be a continuous function.*

Whenever $\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ and $(\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu)$, define

$$K[\gamma] = \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\gamma(x, y) \quad \text{and} \tag{3.2.15}$$

$$J(\varphi, \psi) = \int_{\mathbb{R}^d} \varphi d\mu + \int_{\mathbb{R}^d} \psi d\nu. \tag{3.2.16}$$

Define $\Gamma(\mu, \nu)$ to be the set of all Borel probability measures γ on $\mathbb{R}^d \times \mathbb{R}^d$ such that for all measurable subset $A \subset \mathbb{R}^d$ and $B \subset \mathbb{R}^d$, then

$$\gamma[A \times \mathbb{R}^d] = \mu(A) \quad \text{and} \quad \gamma[\mathbb{R}^d \times B] = \nu(B), \tag{3.2.17}$$

and define ϕ to be the set of all measurable function $(\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu)$ satissfying

$$\varphi(x) + \psi(y) \leq c(x, y), \tag{3.2.18}$$

for $d\mu-$almost all $x \in \mathbb{R}^d$ and $d\nu-$almost all $y \in \mathbb{R}^d$. Then

$$\min_{\gamma \in \Gamma(\mu,\nu)} K(\gamma) = \sup_{(\varphi,\psi)\in\phi} J(\varphi, \psi). \tag{3.2.19}$$

A casual illustration of this duality result is given by Caffarelli, see in Villani (2003b), as follows: consider the shippers problem. Suppose we own a number of coal mines and a number of factories, we wish to transport the coal from mines to factories. The amount each mine produces and each factory requires is fixed (and we assume equal). The cost for you to transport from mine x to factory y is c(x, y). The total optimal cost is the solution to Kantorovich's optimal transport problem. Now a clever shipper comes to you and says they will ship for you and you just pay a price $\varphi(x)$ for loading and $\psi(y)$ for unloading. To make it in your interest, the shipper makes sure that $\varphi(x) + \psi(y) \leq c(x, y)$ that is the cost is no more than what you would have spent transporting the coal yourself. Kantorovich duality tells us that one can find φ and ψ such that this price scheme costs just as much as paying for the cost of transport yourself.

In 1987, Yann Brenier exploited this duality of the Kantorovich's problem to construct the solution of Monge's problem in a specific setting.

3.3 Brenier Theorem

Under some regularity assumption on source measure, the Brenier theorem provides a characterization of the optimal transport in the case where the cost function is quadratic. Furthermore, the Brenier theorem provides a link between optimal transport and partial differential equations. Next, we state the Brenier theorem as found in Villani (2003a).

3.3.1 Theorem (Brenier theorem)**.** *Let μ and ν be two compactly supported probability measures on $\mathbb{R}^d$. If μ is absolutely continous with Lebesgue measure, then:*

(i) there exists a unique solution T to the optimal transport problem with the cost function $c(x, y) = |x - y|^2/2$;

(ii) there exists a convex function $\varphi : \mathbb{R}^d \longrightarrow \mathbb{R}$ such that the optimal map T is given by $T(x) = D\varphi(x)$ for all $\mu-$almost everywhere $x \in \mathbb{R}^d$.

Furthermore, if $\mu(dx) = f(x)dx$ and $\nu(dy) = g(y)dy$, then $D\varphi(x)$ is differentiable μ a.e. and

$$\det(D^2\varphi(x)) = \frac{f(x)}{g(D\varphi(x))}, \tag{3.3.1}$$

for $\mu-$ a.e. and $x \in \mathbb{R}^d$.

3.3.2 Remark. (i) We say that φ is a Brenier solution to the Monge-Ampere equation 3.3.1.

(ii) Here we give a formal proof of equation 3.3.1.

Formal proof. Assume $d\mu(x) = f(x)dx$ and $d\nu(y) = g(y)dy$, then, by the first part of Brenier theorem, there exists a convex function $\varphi : \mathbb{R}^d \longrightarrow \mathbb{R}$ such that the optimal map T is characterised by $T\#\mu = \nu$ where $T(x) = D\varphi(x)$. This implies that

$$\int (\psi \circ D\varphi) d\mu(x) = \int \psi(y) d\nu(y). \tag{3.3.2}$$

for all $\psi \in L^1$. But since $d\mu(x) = f(x)dx$ and $d\nu(y) = g(y)dy$, we have

$$\int \psi(D\varphi(x)) f(x) dx = \int \psi(y) g(y) dy, \tag{3.3.3}$$

for all $\psi \in L^1$. If we assume that $D\varphi$ is smooth and $D\varphi$ is a one-to-one function, meaning that it is strictly convex, then by the change of variables $y = D\varphi(x)$, we have $dy = \det D^2\varphi(x)dx$ and equation 3.3.3 becomes

$$\int \psi(D\varphi(x)) f(x) dx = \int \psi(D\varphi(x)) g(D\varphi(x)) \det D^2\varphi(x) dx \tag{3.3.4}$$

for all $\psi \in L^1$. Since ψ is arbitrary, equation 3.3.4 yields

$$f(x) = g(D\varphi(x)) \det D^2\varphi(x) \tag{3.3.5}$$

for almost $x \in \mathbb{R}^d$. Assume that g in equation 3.3.5 is strictly positive on the support of ν. Since μ is absolutely continuous with respect to Lebesgue measure, then equation 3.3.5 can be written as

$$\det D^2\varphi(x) = \frac{f(x)}{g(D\varphi(x))} \tag{3.3.6}$$

for $\mu-$ a.e. It follows that equation 3.3.6 satisfies the Monge-Ampere equation.

□

The Brenier theorem was an important contribution to understanding the structure of optimal transport, more specifically Brenier showed that the optimal transport is a mapping with a convex potential. However, the Brenier theorem provides little information on the regularity of the convex potential.

In the next proposition, we give an explicit solution to the Monge-Ampere equation

$$g(D\varphi(x)) det D^2\varphi(x) = f(x) \tag{3.3.7}$$

for a specific choice of g and f. This proposition will set the stage for the study of the regularity of Brenier solution to the Monge-Ampere equation.

Set

$$D^+ := \{(x_1, x_2) \in B_1(0) : x_1 > 0\} \quad \text{and} \quad D^- := \{(x_1, x_2) \in B_1(0) : x_1 \leq 0\}. \tag{3.3.8}$$

Next consider the set D^0 defined by

$$D^0 = \left(D^+ + (1,0)\right) \cup \left(D^{(-)} - (1,0)\right) \tag{3.3.9}$$

3.3.3 Proposition. Let μ and ν be two probability measures with density f and g respectively, where $f = \frac{1}{\pi}\mathcal{X}_{B_1}$ and $g = \frac{1}{\pi}\mathcal{X}_{D^0}$. Set $\psi_0(x_1, x_2) = |x_1| + \frac{1}{2}(x_1^2 + x_2^2)$ for all $(x_1, x_2) \in \mathbb{R}^2$. Then, the following hold:

(i) ψ_0 is convex.

(ii) $D\psi_0(x_1, x_2) = (x_1 + sign(x_1), x_2)$ μ a.e. for $(x_1, x_2) \in \mathbb{R}^2$ and $D\psi_0 \# \mu = \nu$.

(iii) ψ_0 solves equation 3.3.7 μ a.e. and $D\psi_0(x_1, x_2) \in D^0$ for almost every $(x_1, x_2) \in B_1$.

Proof. Set $\psi_0 = f_0 + f_1$ where $f_0(x_1, x_2) = |x_1|$ and $f_1(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2)$ for all $(x_1, x_2) \in \mathbb{R}^2$.

Convexity of f_0:

Let (x_1, x_2) and $(y_1, y_2) \in \mathbb{R}^2$, and $\lambda \in [0, 1]$. We have

$$\begin{aligned} f_0\left(\lambda(x_1, x_2) + (1-\lambda)(y_1, y_2)\right) &= |\lambda x_1 + (1-\lambda) y_1| && (3.3.10)\\ &\leq \lambda |x_1| + (1-\lambda)|y_1| && (3.3.11)\\ &\leq \lambda f_0(x_1, x_2) + (1-\lambda) f_0(y_1, y_2). && (3.3.12) \end{aligned}$$

Inequality 3.3.11 is obtained by triangular inequality and the fact that λ and $1-\lambda$ are always positive. Since (x_1, x_2), (y_1, y_2) and λ are arbitrary we conclude that f_0 is a convex function.

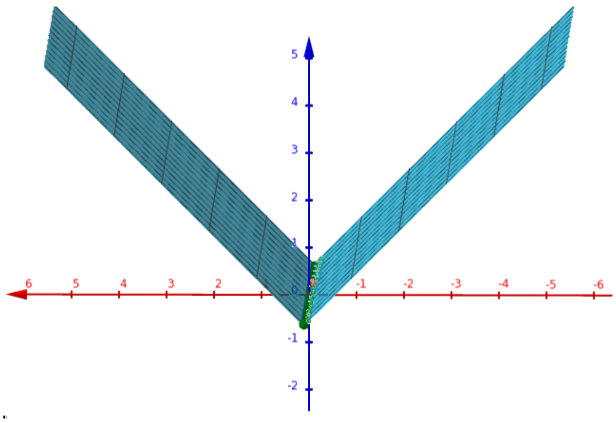

Figure 3.1: Graphical representation of f_0

Convexity of f_1.

Since f_1 is twice differetiable on $\mathbb{R}^2$, the Hessain matrix of f_1 is given by

$$D^2 f_1 = \begin{bmatrix} \frac{\partial^2 f_1}{\partial x_1^2} & \frac{\partial^2 f_1}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f_1}{\partial x_2 \partial x_1} & \frac{\partial^2 f_1}{\partial x_2^2} \end{bmatrix}. \tag{3.3.13}$$

Note that,

$$\frac{\partial^2 f_1}{\partial x_1^2} = 1, \qquad \frac{\partial^2 f_1}{\partial x_2^2} = 1, \qquad \frac{\partial^2 f_1}{\partial x_1 \partial x_2} = 0 \quad \text{and} \quad \frac{\partial^2 f_1}{\partial x_2 \partial x_1} = 0. \tag{3.3.14}$$

Combining equation3.3.13 and 3.3.14, we get:

$$D^2 f_1 = I. \tag{3.3.15}$$

Since the Hessian matrix is symmetric and positive definite, f_1 is a convex function.

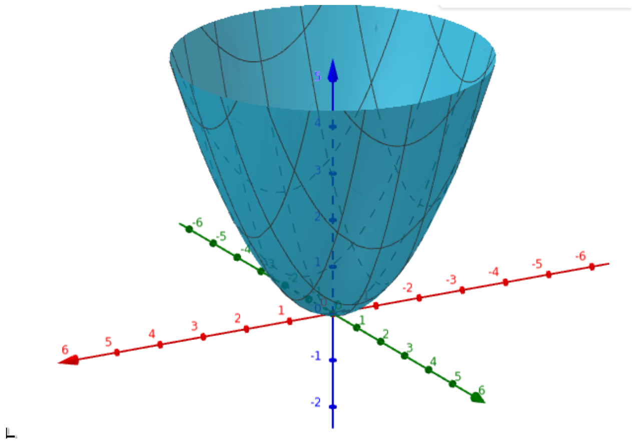

Figure 3.2: Graphical representation of f_1

As f_0 and f_1 are convex, we conclude that ψ_0 is convex which proves (i).

Observe that the function ψ_0 can be written as

$$\psi_0(x_1, x_2) = \begin{cases} x_1 + \frac{1}{2}(x_1^2 + x_2^2) & \text{when} \quad x_1 > 0, \\ \frac{1}{2}x_2^2 & \text{when} \quad x_1 = 0, \\ -x_1 + \frac{1}{2}(x_1^2 + x_2^2) & \text{when} \quad x_1 < 0, \end{cases} \tag{3.3.16}$$

for all $(x_1, x_2) \in \mathbb{R}^2$. Note that ψ_0 is differentiable everywhere except at points (x_1, x_2) such that $x_1 = 0$. At each point (x_1, x_2) of differentiablity,

$$D\psi_0(x_1, x_2) = \begin{cases} (x_1 + 1, x_2) & \text{when} \quad x_1 > 0, \\ (x_1 - 1, x_2) & \text{when} \quad x_1 < 0. \end{cases} \tag{3.3.17}$$

Observe that the set of points at which ψ_0 is not differentiable is $\mathscr{L}^2$-negligible. Since μ is absolutely continuous with Lebesgue measure, this set is also $\mu-$negligible. It follows from equation 3.3.17 that

$$D\psi_0(x_1, x_2) = (x_1 + sign(x_1), x_2) \qquad \mu \text{ a.e.}$$

Next, we show that $D\psi_0 \# \mu = \nu$. Let $\varphi \in L^1(d\nu)$ then, we have:

$$\begin{aligned}
\int_{\mathbb{R}^d} \varphi \circ D\psi_0 d\mu &= \frac{1}{\pi} \int_{B_1} \varphi(x_1 + sign(x_1), x_2) d\mu \\
&= \frac{1}{\pi} \int_{B_1} \varphi(x_1 + sign(x_1), x_2) dx_1 dx_2 \\
&= \frac{1}{\pi} \int_{D^-} \varphi(x_1 - 1, x_2) dx_1 dx_2 + \frac{1}{\pi} \int_{D^+} \varphi(x_1 + 1, x_2) dx_1 dx_2 \\
&= \frac{1}{\pi} \int_{(D^- + (-1,0)} \varphi(x_1, x_2) dx_1 dx_2 + \frac{1}{\pi} \int_{D^+ + (1,0)} \varphi(x_1, x_2) dx_1 dx_2 \\
&= \frac{1}{\pi} \int_{\left(D^{(-)} + (-1,0) \cup D^+ + (1,0)\right)} \varphi(x_1, x_2) dx_1 dx_2 \\
&= \frac{1}{\pi} \int_{D^0} \varphi(x_1, x_2) dx_1 dx_2 \\
&= \frac{1}{\pi} \int_{\mathbb{R}^d} \varphi d\nu.
\end{aligned}$$

As φ is arbitrary in the above equations implies $D\psi_0$ transports μ onto ν. This result proves (ii).

The fact that $D\psi_0 \# \mu = \nu$ implies that ψ_0 solves equation 3.3.7 μ a.e. by the Brenier theorem. Next we prove that $D\psi_0(x_1, x_2) \in D^0$ for almost every $x_1, x_2 \in B_1$. Let $(x_1, x_2) \in B_1$ such that $x_1 > 0$. It follows that

$$D\psi_0(x_1, x_2) = (x_1, x_2) + (1, 0) \in D^+ + (1, 0) \subset D^0. \tag{3.3.18}$$

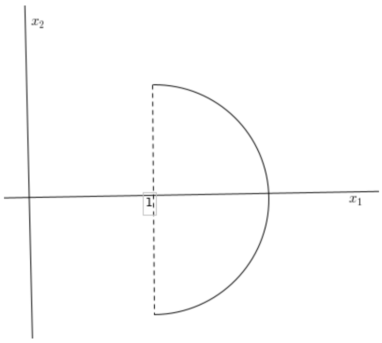

Figure 3.3: $D^+ + (1, 0)$

Similarly, let $(x_1, x_2) \in B_1$ such that $x_1 < 0$. We have:

$$D\psi_0(x_1, x_2) = (x_1, x_2) - (1, 0) \in D^- - (1, 0) \subset D^0. \tag{3.3.19}$$

Note that, the set of points $(x_1, x_2) \in B_1$ where $x_1 = 0$ is $\mathscr{L}^2$-negligible. Therefore $D\psi_0(x_1, x_2) \in D^0$ for almost every $x_1, x_2 \in B_1$. This results proves (iii).

□

4. Regularity of Convex Potential Arising in the Brenier Theorem.

Caffarelli has developed a regularity theory for Alexandrov solution to the Monge-Ampere equations see in Caffarelli (1990b),Caffarelli (1990a),Caffarelli (1991). However, this theory does not extend in general to Brenier solutions for even smooth densities. In this section, we illustrate this fact and provide a key condition necessary for obtaining regularity for the Brenier solution to the Monge-Ampere equation following Caffarelli (1992b).

As in proposition 3.3.3, we define

$$\psi_0 = f_0 + f_1, \tag{4.0.1}$$

where

$$f_0(x_1, x_2) = |x_1| \quad \text{and} \quad f_1(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2), \tag{4.0.2}$$

for all $x_1, x_2 \in \mathbb{R}^2$. We set out to find the Monge-Ampere measure associated to ψ_0. In order, to do that we first compute subdifferential of ψ_0 in the proposition 4.0.1 below.

4.0.1 Proposition. ψ_0 is convex and subdifferential of ψ_0 is given by

$$\partial^0 \psi_0(a, b) = \begin{cases} \{(1 + a, b)\} & \text{for} \quad a > 0, \\ \{(s, b) : -1 \leq s \leq 1\} & \text{for} \quad a = 0, \\ \{(-1 + a, b)\} & \text{for} \quad a < 0, \end{cases} \tag{4.0.3}$$

for all $a, b \in \mathbb{R}^2$.

Proof. Note that,

$$f_0(x_1, x_2) = \begin{cases} x_1 & \text{when} \quad x_1 \geq 0 \quad \text{and} \quad x_2 \in \mathbb{R}^2, \\ -x_1 & \text{when} \quad x_1 < 0 \quad \text{and} \quad x_2 \in \mathbb{R}^2. \end{cases} \tag{4.0.4}$$

Let $(a, b) \in \mathbb{R}^2$ such that a is strictly positive. Note that f_0 is differentiable in the neighborhood of (a, b). As a result, the subdifferential of f_0 at is given by

$$\partial^0 f_0(a, b) = \{Df_0(a, b)\} \tag{4.0.5}$$
$$= \{(1, 0)\}. \tag{4.0.6}$$

Assume now that a is strictly negative, note that f_0 is differentiable in the neighborhood of (a, b), therefore, the subdifferential of f_0 is given by

$$\partial^0 f_0(a, b) = \{Df_0(a, b)\} \tag{4.0.7}$$
$$= \{(-1, 0)\}. \tag{4.0.8}$$

Assume that $a = 0$. Then, we obtain $\partial^0 f_0(0, b)$ as follows:

Let $A_1(b), A_2(b) \in \partial^0 f_0(0,b)$. Then, we have

$$< (A_1(b), A_2(b)), (x_1, x_2 - b) > \quad \leq f(x_1, x_2) - f(0,b), \tag{4.0.9}$$

for any $(x_1, x_2) \in \mathbb{R}^2$. Using inequality 4.0.9, we have

$$A_1(b)x_1 + A_2(b)(x_2 - b) \quad \leq |x_1| - |0|. \tag{4.0.10}$$

for any $(x_1, x_2) \in \mathbb{R}^2$. We set $x_1 = 0$ in 4.0.10 and obtain

$$A_2(b)(x_2 - b) \leq 0 \tag{4.0.11}$$

for all $x_2 \in \mathbb{R}$. Inequality 4.0.11 holds in particular when $x_2 = b + 1$, and we then have

$$A_2(b) \leq 0. \tag{4.0.12}$$

In similar way, inequality 4.0.11 holds in particular when $x_2 = b - 1$, and we then have

$$-A_2(b) \leq 0. \tag{4.0.13}$$

By combining inequality 4.0.12 and inequality 4.0.13 we have:

$$A_2(b) = 0. \tag{4.0.14}$$

Set $x_2 = b$, in inequality 4.0.10, then, we have

$$A_1(b)x_1 \leq |x_1|, \tag{4.0.15}$$

for $x_1 \in \mathbb{R}$. Taking $x_1 = 1$ in equation 4.0.15, we get

$$A_1(b) \leq 1. \tag{4.0.16}$$

Taking $x_1 = -1$ in equation 4.0.15, we have

$$-A_1(b) \leq |-1| = 1. \tag{4.0.17}$$

Combining inequality 4.0.16 and inequality 4.0.17, we get:

$$-1 \leq A_1(b) \leq 1. \tag{4.0.18}$$

Therefore, we conclude that

$$\partial^0 f(0,b) \subset [-1,1] \times \{0\}. \tag{4.0.19}$$

Conversely, for equality to hold we need to show that

$$[-1,1] \times \{0\} \subset \partial^0 f_0(0,b). \tag{4.0.20}$$

Let $A \in [-1,1]$ and $(x_1, x_2) \in \mathbb{R}^2$. Then, we have the following estimate

$$\begin{aligned} Ax_1 + 0(x_2 - b) &\leq |A||x_1| && (4.0.21)\\ &\leq |x_1| - |0| && (4.0.22)\\ &\leq f_0(x_1, x_2) - f_0(0,b). && (4.0.23) \end{aligned}$$

As $A \in [-1, 1]$, we have $|A| \leq 1$. It follows that 4.0.21 become 4.0.22. But since x_1 is arbitrary, equation 4.0.23 implies that

$$(A, 0) \in \partial^0 f_0(0, b). \tag{4.0.24}$$

Therefore,

$$[-1, 1] \times \{0\} \subset \partial^0 f_0(0, b). \tag{4.0.25}$$

We conclude that

$$\partial^0 f_0(0, b) \quad = [-1, 1] \times \{0\}\,. \tag{4.0.26}$$

In conclusion, using equations 4.0.6,4.0.8 and 4.0.26, we obtain

$$\partial^0 f_0(a, b) = \begin{cases} \{(1, 0)\} & \text{if} \quad a > 0, b \in \mathbb{R}^2, \\ [-1, 1] \times \{0\} & \text{if} \quad a = 0, b \in \mathbb{R}^2, \\ \{(-1, 0)\} & \text{if} \quad a > 0, b \in \mathbb{R}^2. \end{cases} \tag{4.0.27}$$

Graphically, Figure 4.1 can represent one subgradient of f_0 at $(0, b)$ with a gradient vector $(0, 0)$ as

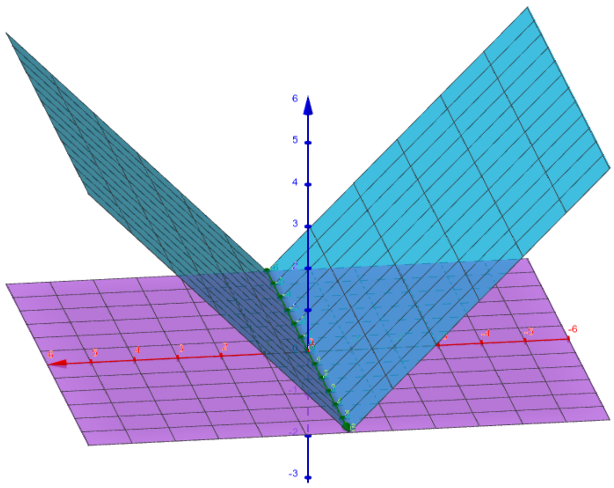

Figure 4.1: subgradient of f_0 at $(0, b)$

Recall that

$$f_1(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2). \tag{4.0.28}$$

Let $(a, b) \in \mathbb{R}^2$, note that f_1 is differentiable in the neighborhood of (a, b) as a result the subdifferential of f_1 is given by

$$\partial^0 f_1(a, b) = Df_1(a, b), \tag{4.0.29}$$

$$= \{(a, b)\}\,. \tag{4.0.30}$$

We conclude that,

$$\partial^0 f_1(a,b) = \{(a,b)\} \quad \text{for} \quad (a,b) \in \mathbb{R}^2. \tag{4.0.31}$$

Graphically, Figure 4.2 represents one subgradient of f_1 at $(0,b)$ with a gradient vector $(0,b)$ as

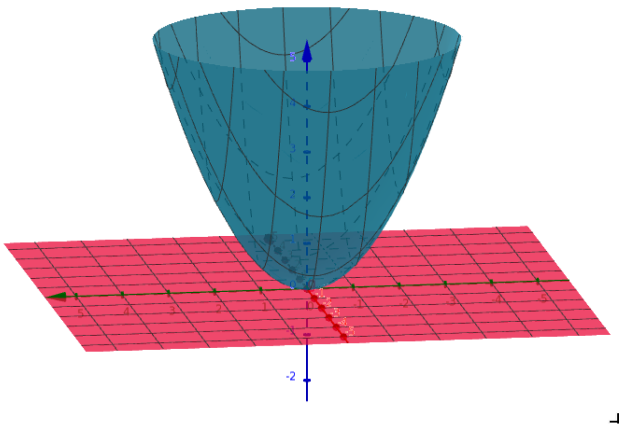

Figure 4.2: subgradient of f_1 at $(0,b)$

Since ψ_0 is convex see proposition 3.3.3, the subddifferential of ψ_0 at point (a,b) is given by

$$\partial^0 \psi_0(a,b) = \partial^0 f_0(a,b) \oplus \partial^0 f_1(a,b), \tag{4.0.32}$$

where $\oplus$ is the Minkowski sum define in 2.3.3.

Therefore we conclude that,

$$\partial^0 \psi_0(a,b) = \begin{cases} \{(1+a,b)\} & \text{if} \quad a>0, b\in\mathbb{R}^2, \\ \{s,b : -1 \le s < 1\} & \text{if} \quad a=0, b\in\mathbb{R}^2, \\ \{(a-1,b)\} & \text{if} \quad a<0 b\in\mathbb{R}^2. \end{cases} \tag{4.0.33}$$

□

We compute the distributional derivative of ψ_0 relying solely on definition 2.4.6. The singularity of the Hessain ψ_0 appears in the following proposition 4.0.2

4.0.2 Proposition. The second order distributional derivative of ψ_0 is given by

$$[D^2\psi_0] = I\mathscr{L}^2 + [D^2\psi_0]_s, \tag{4.0.34}$$

where I is the identity matrix and $[D^2\psi_0]_s = \begin{bmatrix} 2\delta_0 \otimes \mathscr{L} & 0 \\ 0 & 0 \end{bmatrix}$.

Proof. Observe that $\psi_0 \in L^1_{loc}(\Omega)$. To compute the distributional derivative of ψ_0. Let $\phi \in C_c^\infty(\Omega)$ which means that ϕ is infinitely differentiable and the support of ϕ is contained in bounded set $\Omega \subset \mathbb{R}^2$

Assume that $\Omega = [-A,A] \times [-A,A]$ with $A>0$. Define

$$\Omega^+ = \{(x_1,x_2) \in \Omega : x_1 \ge 0, \} \quad \text{and} \quad \Omega^- = \{(x_1,x_2) \in \Omega : x_1 < 0.\}$$

and the boundary of Ω is defined by

$$\partial\Omega+ = b_1^+ \cup b_2^+ \cup b_3^+ \cup b_4^+ \quad \text{and} \quad \partial\Omega- = b_5^- \cup b_6^- \cup b_7^- \cup b_4^+. \tag{4.0.35}$$

As shown in Figure 4.3.

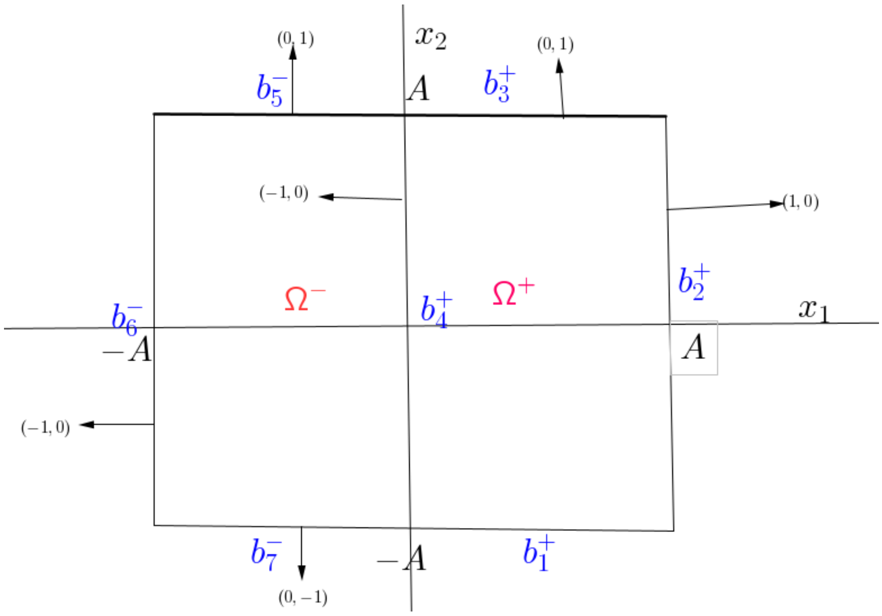

Figure 4.3: Bounded domain Ω

We can rewrite ψ_0 as follows:

$$\psi_0(x_1, x_2) = \begin{cases} x_1 + \frac{1}{2}({x_1}^2 + {x_2}^2) & \text{when} \quad x_1 \geq 0, \\ -x_1 + \frac{1}{2}({x_1}^2 + {x_2}^2) & \text{when} \quad x_1 < 0. \end{cases} \tag{4.0.36}$$

By definition 2.4.6, we can find the distribution derivative of ψ_0 as follows:

$$\begin{aligned} < \partial_{x_1} T_{\psi_0} \phi > &= - < T_{\psi_0} \partial_{x_1} \phi > \\ &= - \int_{\mathbb{R}^2} \psi_0 \partial_{x_1} \phi dx_1 dx_2 \\ &= - \int_{\Omega+} \psi_0 \partial_{x_1} \phi dx_1 dx_2 - \int_{\Omega-} \psi_0 \partial_{x_1} \phi dx_1 dx_2 \\ &= -A - B, \end{aligned}$$

where A and B are give by

$$\begin{aligned} A &= \int_{\Omega+} (x_1 + \frac{1}{2}(x_1^2 + x_2^2)) \partial_{x_1} \phi dx_1 dx_2, \\ B &= \int_{\Omega-} (-x_1 + \frac{1}{2}(x_1^2 + x_2^2)) \partial_{x_1} \phi dx_1 dx_2. \end{aligned}$$

Let us compute A using integration by part. We have:

$$A = \int_{\partial\Omega+} (x_1 + \frac{1}{2}(x_1^2 + x_2^2)))\phi n_{x_1} d\Gamma - \int_{\Omega+} \partial_{x_1}(x_1 + \frac{1}{2}(x_1^2 + x_2^2))\phi dx_1 dx_2 \tag{4.0.37}$$

$$= \int_{b_4^+} (x_1 + \frac{1}{2}(x_1^2 + x_2^2)))\phi n_{x_1} d\Gamma - \int_{\Omega+} (1 + x_1)\phi dx_1 dx_2 \tag{4.0.38}$$

$$= \int_{-A}^{A} (0 + \frac{1}{2}(0 + x_2^2)))\phi(0, x_2)(-1) dx_2 - \int_{\Omega+} (1 + x_1)\phi dx_1 dx_2 \tag{4.0.39}$$

$$= \int_{-A}^{A} -\frac{x_2^2}{2}\phi(0, x_2) dx_2 - \int_{\Omega+} (1 + x_1)\phi dx_1 dx_2 \tag{4.0.40}$$

We obtain equation 4.0.38 since $\partial\Omega+ = b_1^+ \cup b_2^+ \cup b_3^+ \cup b_4^+$ and $\phi = 0$ on b_1^+, b_2^+, b_3^+ and 0n b_4^+ is such that $n_{x_1} = -1$, $x_1 = 0$ and x_2 ranges from A to $-A$.

In the same way we compute B using integration by parts as follows.

$$B = \int_{\partial\Omega-} (-x_1 + \frac{1}{2}(x_1^2 + x_2^2)))\phi n_{x_1} d\Gamma - \int_{\Omega-} \partial_{x_1}(-x_1 + \frac{1}{2}(x_1^2 + x_2^2))\phi dx_1 dx_2 \tag{4.0.41}$$

$$= \int_{-A}^{A} \frac{x_2^2}{2}\phi(0, x_2)(1) dx_2 - \int_{\Omega-} (-1 + x_1)\phi dx_1 dx_2 \tag{4.0.42}$$

We obtain equation 4.0.42 since $\partial\Omega- = b_5^- \cup b_6^- \cup b_7^- \cup b_4^+$ and $\phi = 0$ on b_5^-, b_6^-, b_7^- and 0n b_4^+ is such that $n_{x_1} = 1$, $x_1 = 0$ and x_2 ranges from A to $-A$..

Combining equation 4.0.40 and 4.0.42, we have:

$$\begin{aligned}
-A - B = &- \int_{\mathbb{R}} -\frac{x_2^2}{2}\phi(0, x_2) dx_2 + \int_{\Omega+} (1 + x_1)\phi dx_1 dx_2 - \int_{\mathbb{R}} \frac{x_2^2}{2}\phi(0, x_2)(1) dx_2 \\
&+ \int_{\Omega-} (-1 + x_1)\phi dx_1 dx_2 \\
&= \int_{\Omega+} (1 + x_1)\phi dx_1 dx_2 + \int_{\Omega-} (-1 + x_1)\phi dx_1 dx_2 \\
&= \int_{\Omega+} sign(x_1) + x_1\phi dx_1 dx_2 + \int_{\Omega-} sign(x_1) + x_1\phi dx_1 dx_2 \\
&= \int_{\Omega} sign(x_1) + x_1\phi dx_1 dx_2 \\
&= \int_{\mathbb{R}^2} sign((x_1) + x_1)\phi dx_1 dx_2.
\end{aligned}$$

Therefore,

$$< \partial_{x_1} T_{\psi_0}, \phi > = \int_{\mathbb{R}^2} sign((x_1) + x_1)\phi dx_1 dx_2, \tag{4.0.43}$$

$$\partial_{x_1}\psi_0 = x_1 + sign(x_1), \tag{4.0.44}$$

in the sense of distributions.

Let us compute $\partial_{x_2}\psi_0$,

$$< \partial x2 T_{\psi_0}, \phi > = - < T_{\psi_0, \partial_{x_2}\phi} > \tag{4.0.45}$$

$$= - \int_{\Omega} \psi_0 \partial_{x_2}\phi dx_1 dx_2. \tag{4.0.46}$$

Using integration by parts, we have:

$$-\int_{\Omega} \psi_0 \partial_{x_2} \phi dx_1 dx_2 = -\int_{\partial\Omega} \psi_0 \phi n_{x_2} d\Gamma + \int_{\Omega} \partial_{x_2} \psi_0 \phi dx_1 dx_2. \quad (4.0.47)$$

But $\phi = 0$ on the boundary of Ω, we get

$$-\int_{\Omega} \psi_0 \partial_{x_2} \phi dx_1 dx_2 = \int_{\Omega} \partial_{x_2} (|x_1| + \frac{1}{2}(x_1^2 + x_2^2)) \phi dx_1 dx_2 \quad (4.0.48)$$

$$= \int_{\Omega} x_2 \phi dx_1 dx_2 \quad (4.0.49)$$

$$(4.0.50)$$

Therefore,

$$\partial x_2 \psi_0 = x_2, \quad (4.0.51)$$

in the sense of distributions. Combine both equation 4.0.44 and 4.0.51, we have:

$$D\psi_0 = x_1 + sign(x_1), x_2. \quad (4.0.52)$$

n the sense of distributions. Next, we compute the distributional second order derivative of ψ_0.

$$D^2 \psi_0 = D(D\psi_0) \quad (4.0.53)$$

$$= D(x_1 + sign(x_1), x_2). \quad (4.0.54)$$

Set $h_1(x_1, x_2) = x_1 + sign(x_1)$ and $h_2(x_1, x_2) = x_2$ for all $(x_1, x_2) \in \mathbb{R}^2$. We start by computing the weak derivative of h_1 with respect to x_1, as follows

$$< \partial x_1 T_{h_1}, \phi > = - < T_{h_1}, \partial x_1 \phi > \quad (4.0.55)$$

$$= -\int_{\mathbb{R}^2} h_1 \partial x_1 \phi dx_1 dx_2 \quad (4.0.56)$$

$$= -C - D, \quad (4.0.57)$$

where C and D are as follows

$$C = \int_{\Omega+} (1 + x_1) \partial x_1 \phi dx_1 dx_2 \quad (4.0.58)$$

$$D = \int_{\Omega-} (-1 + x_1) \partial x_1 \phi dx_1 dx_2. \quad (4.0.59)$$

Computing C uisng integration by parts we have;

$$C = \int_{\partial\Omega+} (1 + x_1) \phi n_{x_1} d\Gamma - \int_{\Omega+} \partial x_1 (1 + x_1) \phi dx_1 dx_2 \quad (4.0.60)$$

$$= \int_{\mathbb{R}} -\phi(0, x_2) dx_2 - \int_{\Omega+} \phi dx_1 dx_2, \quad (4.0.61)$$

Equation 4.0.60 becomes 4.0.61 Since $\partial\Omega^+ = b_4^+$ $n_{x_1} = -1$ and $x_1 = 0$. Similarly computing D using integration by parts, we have

$$D = \int_{\partial\Omega-} (-1 + x_1) \phi n_{x_1} d\Gamma - \int_{\Omega-} \partial x_1 (-1 + x_1) \phi dx_1 dx_2 \quad (4.0.62)$$

$$= \int_{\mathbb{R}} -\phi(0, x_2) dx_2 - \int_{\Omega-} \phi dx_1 dx_2. \quad (4.0.63)$$

since $\partial\Omega- = b_4^+$ $n_{x_1} = -1$ and $x_1 = 0$, 4.0.62 becomes 4.0.63.

Subsituting, equations 4.0.61 and 4.0.63 into equation 4.0.57 we have:

$$\begin{aligned}-C - D &= \int_{\mathbb{R}} \phi(0, x_2)dx_2 + \int_{\Omega+} \phi dx_1 dx_2 + \int_{\mathbb{R}} \phi(0, x_2)dx_2 + \int_{\Omega-} \phi dx_1 dx_2 \\ &= 2\int_{\mathbb{R}} \phi(0, s)ds + \int_{\Omega} \phi dx_1 dx_2 \\ &= 2\int_{\mathbb{R}} \left[\int_{\mathbb{R}} \phi(y, s)d\delta_0(y)\right] ds + \int_{\Omega} \phi dx_1 dx_2 \\ &= 2\int_{\mathbb{R}\times\mathbb{R}} \phi(y, s)d\delta_0(y) \otimes \mathscr{L}^1 + \int_{\Omega} \phi dx_1 dx_2.\end{aligned}$$

Therefore, we have

$$\begin{aligned}< \partial_{x_1} T_{h_1}, \phi > &=< 2\delta_0 \otimes \mathscr{L}^1, \phi > + < 1, \phi > && (4.0.64)\\ &=< 2\delta_0 \otimes \mathscr{L}^1 + T_1, \phi >, && (4.0.65)\end{aligned}$$

hence

$$\partial_{x_1} h_1 = 2\delta_0 \otimes \mathscr{L}^1 + 1, \quad (4.0.66)$$

in the sense of distribution.

Computing the weak derivative of h_1 with respect to x_2 as follows.

$$\begin{aligned}< \partial_{x_2} T_{h_1}, \phi > &= - < T_{h_1}, \partial x_2 \phi > && (4.0.67)\\ &= -\int_{\Omega} h_1 \partial_{x_2} \phi dx_1 dx_2 && (4.0.68)\\ &= -E - G && (4.0.69)\end{aligned}$$

where E and G

$$\begin{aligned}E &= \int_{\Omega+} (1 + x_1)\partial x_2 \phi dx_1 dx_2 && (4.0.70)\\ G &= \int_{\Omega-} (-1 + x_1)\partial x_2 \phi dx_1 dx_2. && (4.0.71)\end{aligned}$$

Computing E by integration by part we have

$$\begin{aligned}E &= -\int_{\partial\Omega+} (1 + x_1)n_{x_2} \phi d\Gamma + \int_{\Omega} \partial_{x_2}(1 + x_1)\phi dx_1 dx_2 && (4.0.72)\\ &= \int_{\Omega} \partial_{x_2}(1 + x_1)\phi dx_1 dx_2 && (4.0.73)\\ &= 0 && (4.0.74)\end{aligned}$$

we obtain equation 4.0.73 since $\phi = 0$ on $\partial\Omega+$. Computing D using integration by part we have

$$\begin{aligned}D &= -\int_{\partial\Omega+} (-1 + x_1)n_{x_2} \phi d\Gamma + \int_{\Omega} \partial_{x_2}(1 + x_1)\phi dx_1 dx_2 && (4.0.75)\\ &= \int_{\Omega} \partial_{x_2}(-1 + x_1)\phi dx_1 dx_2 && (4.0.76)\\ &= 0 && (4.0.77)\end{aligned}$$

we obtain equation 4.0.76 since $\phi = 0$ on $\partial\Omega-$. Since $-E - D = 0$, therefore

$$\partial_{x_2} h_1 = 0. \tag{4.0.78}$$

in distributional sense.

Computing the distributional derivative of h_2 with respect to x_1, we have

$$
\begin{aligned}
< \partial_{x_1} T_{h_2}, \phi > &= - < T_{h_2}, \partial x_1 \phi > && (4.0.79)\\
&= - \int_\Omega h_2 \partial_{x_1} \phi dx_1 dx_2 && (4.0.80)\\
&= - \int_{\partial\Omega} (x_2) n_{x_1} \phi d\Gamma + \int_\Omega \partial_{x_1}(x_2) \phi dx_1 dx_2 && (4.0.81)\\
&= \int_\Omega \partial_{x_1}(x_2) \phi dx_1 dx_2 && (4.0.82)\\
&= 0 && (4.0.83)
\end{aligned}
$$

we obtain equation 4.0.82 since on $\partial\Omega$ $\phi = 0$

$$\partial_{x_1} h_2 = 0 \tag{4.0.84}$$

in distribution sense. Computing the distributional derivative of h_2 with respect to x_2, we have

$$
\begin{aligned}
< \partial_{x_2} T_{h_2}, \phi > &= - < T_{h_2}, \partial x_2 \phi >\\
&= - \int_\Omega h_2 \partial_{x_2} \phi dx_1 dx_2\\
&= - \int_{\partial\Omega} x_2 n_{x_2} \phi d\Gamma + \int_\Omega \partial_{x_2} x_2 \phi dx_1 dx_2\\
&= \int_\Omega \partial_{x_2}(x_2) \phi dx_1 dx_2\\
&= \int_\Omega \phi dx_1 dx_2\\
&=< T_v, \phi >,
\end{aligned}
$$

where $v(x_1, x_2) = 1$, hence

$$\partial_{x_2} h_2 = 1, \tag{4.0.85}$$

in distribution sense. Using equations 4.0.66,4.0.78,4.0.84 and 4.0.85, $D^2\psi_0$ is given by the following Hessian matrix

$$[D^2\psi_0] = I\mathscr{L}^2 + [D^2\psi_0]_s, \tag{4.0.86}$$

where I is the identity matrix and $[D^2\psi_0]_s = \begin{bmatrix} 2\delta_0 \otimes \mathscr{L} & 0 \\ 0 & 0 \end{bmatrix}$. □

4.0.3 Proposition. The Monge-Ampere measure associated to ψ_0 is given by

$$\mu_{\psi_0} = \mathscr{L}^2 + 2\mathscr{L}^1_{|\Gamma_0} \tag{4.0.87}$$

Proof. Let E be a Borel set such that $E = E_+ \cup E_0 \cup E_-$, where

$$\begin{aligned} E_+ &:= \{(x_1, x_2) \in E \quad : \quad x_1 > 0\}, \\ E_0 &:= \{(x_1, x_2) \in E \quad : \quad x_1 = 0\}, \\ E_+ &:= \{(x_1, x_2) \in E \quad : \quad x_1 < 0\}. \end{aligned}$$

See in Figure 4.4 below. Note that

$$\mathscr{L}^2(\{E_0\}) = 0. \tag{4.0.88}$$

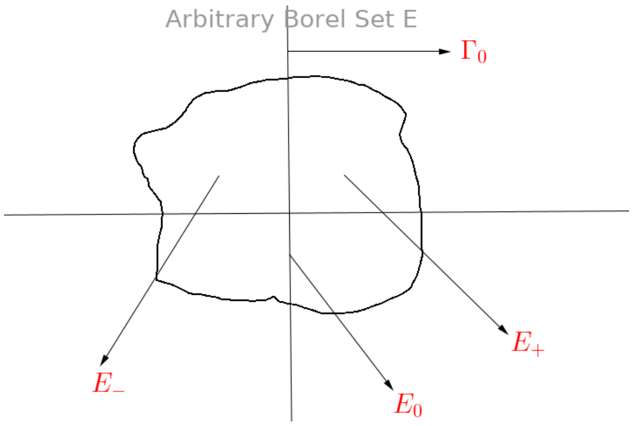

Figure 4.4: Arbitrary Borel set E

Using definition 2.2.8, Monge-Ampere measure is given by

$$\partial^0\psi_0(E) = \cup_{(a,b)\in E}\partial^0\psi_0(a,b) \tag{4.0.89}$$
$$= \cup_{(a,b)\in E_+}\partial\psi_0(a,b) \cup_{(a,b)\in E_-} \partial\psi_0(a,b) \cup_{(a,b)\in E_0} \partial\psi_0(0,b). \tag{4.0.90}$$

By lemma 4.0.1

$$\partial^0\psi_0(E) = \cup_{(a,b)\in E_+}(1+a,b) \cup_{(a,b)\in E_-} (a-1,b) \cup_{(a,b)\in E_0} (\cup_{s\in[-1,1]}(s,b)) \tag{4.0.91}$$
$$= \{(E_+) + (1,0)\} \cup \{(E_-) + (-1,0)\} \cup \{([-1,1]) \times (E \cap \Gamma_0)\}, \tag{4.0.92}$$

where Γ_0 represents the y-axis. Monge-Ampere measure is given by

$$\mu_{\psi_0}(E \quad) = \mathscr{L}^2(\{(E_+) + (1,0)\} + \{(E_-) + (-1,0)\} + \{([-1,1]) \times (E \cap \Gamma_0)\}) \tag{4.0.93}$$

Since the coordinates x_1, x_2 in $\{E_+ + (1,0)\}$, $\{E_- - (1,0)\}$ and $\{E_0\}$ are respectively given by $x_1 > 1$, $x_1 < -1$ and $x_1 \in [-1,1]$, we then above set are disjoint. Therefore,

$$\mu_{\psi_0}(E) = \mathscr{L}^2(\{E_+ + (1,0)\}) + \mathscr{L}^2(\{E_- + (-1,0)\}) + \mathscr{L}^2(\{\{([-1,1]) \times (E \cap \Gamma_0)\}) \tag{4.0.94}$$

Since Lebesgue measure is invariant by translation and $\mathscr{L}^2\{E_0\} = 0$, we have equation 4.0.94 reduces to

$$\mu_{\psi_0}(E) = \mathscr{L}^2\{E\} + \mathscr{L}^1\{[-1,1]\} \times \mathscr{L}^1\{E \cap \Gamma_0\} \tag{4.0.95}$$

We can rewrite equation 4.0.95 by restricting Γ_0 on $\{E\}$ as

$$\mu_{\psi_0}(E) = \mathscr{L}^2\{E\} + \mathscr{L}^1\{[-1,1]\} \times \mathscr{L}^1_{|\Gamma_0}\{E\} \tag{4.0.96}$$

Therefore, the Monge-Ampere measure will be given by;

$$\mu_{\psi_0}(E) = \mathscr{L}^2\{E\} + 2\mathscr{L}^1_{|\Gamma_0}\{E\}, \tag{4.0.97}$$

for any arbitrary borel set E. Hence,

$$\mu_{\psi_0} = \mathscr{L}^2 + 2\mathscr{L}^1_{|\Gamma_0}. \tag{4.0.98}$$

□

4.0.4 Remark. (i) Recall that ψ_0 is the Brenier solution for the Monge-Ampere equation 3.3.7 with $f = \frac{1}{\pi}\mathcal{X}_{B_1}$ and $g = \frac{1}{\pi}\mathcal{X}_{D^0}$, see proposition 3.3.3. However, proposition 4.0.3 implies that ψ_0 is not an Alexandrov solution to Monge-Ampere equation 3.3.7, since μ_{ψ_0} has a mass on lower-dimensional space Γ_0.

(ii) One may think the singularity of μ_{ψ_0} arises because of the support target measure D^0 is disconnected. Brenier showed that even when the target measure is connected, one may still have singularity, see Figalli (2010).

We now state a necessary condition for the regularity of Brenier solution to the Monge-Ampere equation. Caffarelli has shown that when the support of the target is convex then the Brenier solution coincides with Alexandrov solution, see theorem 4.0.5. Subsequently, the regularity theory developed in Caffarelli (1990b),Caffarelli (1990a),Caffarelli (1991) applies.

The following theorem 4.0.5 gives a summary of the discussion above and can be found in Villani (2003b).

4.0.5 Theorem. *(Alexandrov solution to the Monge-Ampere equation for convex target) Let μ and ν be two probability measures on $\mathbb{R}^d$, absolutely continuous with respect to Lebesgue measure. Let f and g be their respective densiities, and let X and Y be their respective supports. Let ψ_0 be a convex function such that $D\psi_0\#\mu = \nu$. Assume that Y is convex and g is positive almost everywhere on Y. Then the Monge-Ampere measure associated with ψ_0 has no singular part on X. In particular ψ is an Alexandrov solution to the Monge-Ampere equation.*

Acknowledgements

I would like to acknowledge AIMS and its funders for supporting this work, my supervisor, Dr. Marc Sedjro from AIMS South Africa, my tutor Dr. Faraniaina, friends and family for their support.

References

Brenier, Y. Optimal transportation theory and geometric partial differential equations. *Applied Mathematics Entering the 21st Century", SIAM, Philadelphia*, pages 37–47, 2004.

Caffarelli, L. A. Interior w2, p estimates for solutions of the Monge-Ampere equation. *Annals of Mathematics*, pages 135–150, 1990a.

Caffarelli, L. A. A localization property of viscosity solutions to the Monge-Ampere equation and their strict convexity. *Annals of mathematics*, 131(1):129–134, 1990b.

Caffarelli, L. A. Some regularity properties of solutions of Monge-Ampere equation. *Communications on Pure and Applied Mathematics*, 44(8-9):965–969, 1991.

Caffarelli, L. A. Boundary regularity of maps with convex potentials. *Communications on Pure and Applied Mathematics*, 45(9):1141–1151, 1992a.

Caffarelli, L. A. The regularity of mappings with a convex potential. *Journal of the American Mathematical Society*, 5(1):99–104, 1992b.

Cannarsa, P. and D'Aprile, T. *Introduction to measure theory and functional analysis*, volume 89. Springer, 2015.

De Philippis, G. and Figalli, A. The monge–ampère equation and its link to optimal transportation. *Bulletin of the American Mathematical Society*, 51(4):527–580, 2014.

Evans, L. C. and Gangbo, W. *Differential equations methods for the Monge-Kantorovich mass transfer problem*. Number 653. American Mathematical Society., 1999.

Figalli, A. Regularity properties of optimal maps between nonconvex domains in the plane. *Communications in Partial Differential Equations*, 35(3):465–479, 2010.

Kantorovich, L. On the transfer of masses dokl. *Acad. Nauk. USSR*, 37:7–8, 1942.

Monge, G. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*, 1781.

Pisante, G., Cullen, M., and Gangbo, W. The semigeostrophic equations discretized in reference and dual variables. 2007.

Rockafellar, R. T. *Convex analysis*. Number 28. Princeton university press, 1970.

Sedjro, M. M. *On the almost axisymmetric flows with forcing terms*. PhD thesis, Georgia Institute of Technology, 2012.

Shutts, G. Some exact solutions to the semi-geostrophic equations for uniform potential vorticity flows. *Geophysical & Astrophysical Fluid Dynamics*, 57(1-4):99–114, 1991.

Thorpe, M. *Introduction to Optimal Transport*. University of Cambridge, 2018.

Van Dijk, G. *Distribution theory*. De Gruyter Graduate Lectures, 2013.

Villani, C. *Optimal Transport.* American Mathematical Society., 2003a.

Villani, C. *Topics in optimal transportation*. Number 58. American Mathematical Society., 2003b.