# THE MONGE-KANTOROVICH MASS TRANSPORTATION PROBLEM

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

IN

MATHEMATICS

By

İhsan Demirel

September 2017

THE MONGE-KANTOROVICH MASS TRANSPORTATION PROBLEM
By İhsan Demirel
September 2017

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Ali Süleyman Üstünel(Advisor)

Azer Kerimov

Mine Çağlar

Approved for the Graduate School of Engineering and Science:

Ezhan Karaşan
Director of the Graduate School

# ABSTRACT

# THE MONGE-KANTOROVICH MASS TRANSPORTATION PROBLEM

İhsan Demirel
M.S. in Mathematics
Advisor: Ali Süleyman Üstünel
September 2017

The Monge mass transportation problem was stated by French Mathematician, G. Monge [6]. After that Soviet Mathematician Leonid Kantorovich [4] published a relaxed version of the problem, namely the Monge-Kantorovich mass transportation problem. This paper is concerned with the definitions and relations of problems and the existence, the uniqueness and the characterization of solutions to problems for some specific cases. We will consider three type of function, namely the quadratic, strictly convex and strictly concave cost functions. The Kantorovich Duality and cyclical monotonicity will be main tools to prove results.

*Keywords:* the Monge-Kantorovich mass transportation, Cyclical Monotonicity, The Kantorovich Duality.

# ÖZET

# MONGE-KANTOROVICH KÜTLE TAŞIMA PROBLEMİ

İhsan Demirel
Matematik, Yüksek Lisans
Tez Danışmanı: Ali Süleyman Üstünel
Eylül 2017

Monge kütle taşıma problemi Fransız Matematikçisi G. Monge [6] tarafından belirtildi. Daha sonra Sovyet Matematikçi Leonid Kantorovich [4] sorunun rahatlatılmış versiyonunu yayınladı. Bu makale problemlerin tanımını, ilişkisini ve bazı özel durumlar için çözümün varlığı, tekliği ve karakteristik özellikleri ile ilgili sonuçları içermekte. Biz maliyet fonksiyonu olarak üç farklı türü inceleyeceğiz, sırası ile ikinci dereceden, tekdüze dışbükey ve tekdüze içbükey fonksiyonlar. Sonuçların ispatında Kantorovich eşlekliğini ve periyodik monotonluğu kullanacağız.

*Anahtar sözcükler*: Monge-Kantorovich Kütle Taşıma Problemi, Periyodik Monotonluk, Kantorovich Eşlekliği .

# Acknowledgement

Firstly, I would like to express my sincere gratitude to my advisor Prof. Ali Süleyman Üstünel for the support of my M.S study and related research, for his patience, and knowledge. His guidance helped me in all the time of research and writing of this thesis.
Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Azer Kerimov and Ass.Prof. Mine Çağlar for their comments, encouragement and valuable time.
I would like to thank my family and friends for their support. I would like to specially thank to Elifnur Yazıcı, Fulya Özturhan, Mustafa Kahraman, Müge Fidan, Nazan Günbükü and Oğuzhan Yörük for increasing my motivation and their help.

# Contents

# Chapter 1

# Introduction

In 1781, Gaspard Monge [6] introduced a question about how to minimize the total cost of earth-moving, In other words, finding a transformation map between two measure spaces with same mass which minimizes the total cost. Later this problem has been studied by many mathematicians and Kantorovich [4] relaxed this problem by changing map seeking with measure seeking and he transformed the non-linear Monge problem into the linear problem.

In this paper we will discuss the existence, the uniqueness and the characterization of the solutions to these two problems.

The paper is organized as follows. In Section 2, we introduce Monge and Monge-Kantorovich problems and we discuss some basic examples to understand the relations between these two problems. Then we state Kantorovich Duality theorem which is a very important tool to study Monge-Kantorovich problem. In Section 3, we show that a solution to Monge-Kantorovich problem exists. After that we discuss some important results about the uniqueness and the characterization of solutions for some special cost functions. Firstly we consider the quadratic cost function and state the results which are presented by Brenier, Knott and Smith by using Kantorovich Duality and generalized by McCann [5] by using cyclic monotonicity. Secondly, similar results have given for the cost functions $c(x, y) = h(x-y)$ and $c(x, y) = l(|x-y|)$ where $h$ is strictly convex and $l \geq 0$ is strictly concave. Again, we use the concept of c-cyclical monotonicity. In

Appendix we will give preliminary knowledge about convex analysis. These are necessary to prove our main results.

# Chapter 2

# Monge-Kantorovich problem

## 2.1 Monge problem

Assume that you have some piles of sand and you need to fill up given holes with it. Holes and piles have the same total volume and moving each sand from location $x$ in one of the piles to location $y$ in a hole has a cost. The question is what is the best way to fill up holes in order to minimize the total cost?

Before we state the problem we need to introduce some notations. $\mathcal{M}(\mathbb{R}^n)$ denotes the set of non-negative Borel measure on $\mathbb{R}^n$ with finite total mass and $\mathcal{P}(\mathbb{R}^n)$ denotes set of Borel probability measure.

**Definition 2.1.1.** For a measure $\mu$ on $X$ and a Borel map $T : X \longrightarrow Y$, define $T_{\#}\mu(A) = \mu\left(T^{-1}(A)\right)$ for all sets $A \subset Y$. The map $T_{\#}\mu$ is a measure on $Y$ and called the push-forward of $\mu$ through $T$. Finally we define $\Sigma(\mu, \nu)$ as $\{T : X \to Y; T_{\#}\mu = \nu\}$.

This question is first modeled by French Mathematician Garpast Monge in 1782 [6]. Piles and Holes are modeled by two probability measure $\mu$, $\nu$ defined on some measurable sets $X$ and $Y$. The cost function $c(x, y)$ gives us the cost of moving a sand from location $x$ to location $y$. One can assume that the cost function is measurable, non-negative and can take the value of infinity, i.e., $c$ :

$X \times Y \to \mathbb{R}_+ \cup \{+\infty\}$. Finally, moving the sand is modeled by a measurable function $T : X \to Y$ with property $T_{\#}\mu = \nu$.
Now, we can state the Monge problem.

**Problem 2.1.2** (Monge problem)**.** Let $X, Y$ be measurable spaces and $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$. Monge problem is to find a Transpor map $T^* \in \Sigma(\mu, \nu)$ that minimizes the functional

$$I[T] = \int_X c(x, T(x))d\mu(x)$$

i.e. $I[T^*] = \inf I[T]$, infimum is taken over the set $\Sigma(\mu, \nu)$

One of the the biggest difficulty of finding solution to Monge problem is that we can not split the mass, in other word we can not divide the mass at location $x$ and sent to two different locations in $Y$. To overcome this difficulty Kantorovich introduces new problem, called Monge-Kantorovich problem [4], which allows us to move mass from location $x$ to different locations in $Y$.

## 2.2 Monge-Kantorovich problem

In this problem, we are not looking for a Transport map, but a measure $\pi$ on the product space $X \times Y$ where $d\pi(x, y)$ models the amount of sand we moved from location $x$ to location $y$. On the other hand, if we back to example, all the sand in the pile must be moved to holes and all sand on holes must come from pile. Formal statement for this is

$$\int_Y d\pi(x, y) = d\mu(x) \text{ and } \int_X d\pi(x, y) = d\nu(y).$$

More formally, If $A$ and $B$ are measurable subsets of $X$ and $Y$ respectively, we must have

$$\pi(A \times Y) = \mu(A) \text{ and } \pi(X \times B) = \nu(B),$$

or equivalently for each $\phi \in \mathcal{L}^1(\mu)$ and $\varphi \in \mathcal{L}^1(\nu)$ , we have

$$\int_{X \times Y} (\phi(x) + \varphi(y))\, d\pi(x, y) = \int_X \phi(x)d\mu(x) + \int_Y \varphi(x)d\nu(y).$$

We will denote the set of all measures that satisfy this condition by $\Pi(\mu, \nu)$. We will call each measure $\pi \in \Pi(\mu, \nu)$ as "Transference Plan", and $\mu$ and $\nu$ as the

first and the second marginals of $\pi$.
Next, we will state Monge-Kantorovich problem.

**Problem 2.2.1** (Monge-Kantorovich Problem)**.** Let $X$ and $Y$ be measurable spaces and $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$. Monge-Kantorovich problem is to find a Transference Plan $\pi^* \in \Pi(\mu, \nu)$ that minimizes the functional

$$J[\pi] = \int_{X \times Y} c(x, y) d\pi(x, y)$$

i.e. $J[\pi^*] = \inf J[\pi]$, infimum is taken over the set $\Pi(\mu, \nu)$.

## 2.3 Relations between Monge problem and Kantorovich problem

**Example 2.3.1.** Let $X = Y = [-1, 1]$ , $\mu = \delta_0$ and $\nu = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$, where $\delta_x$ is Dirac measure.
Clearly there is no Transport map such that $T_{\#}\mu = \nu$, i.e., $\Sigma(\mu, \nu) = \emptyset$. Hence, there is no solution to Monge Problem.
However, $\pi^*(A \times B) = \mu(A)\nu(B)$ is almost surely the unique element of $\Pi(\mu, \nu)$. Therefore, it is solution to Monge-Kantorovich problem and

$$\inf J[\pi] = \frac{1}{2}c(0, -1) + \frac{1}{2}c(0, 1).$$

Evidently, although solution to Monge problem does not exist (because of the structure of measures,) solution to Monge-Kantorovich exists since it allows to split masses. As a result, existence of solutions depends on structure of measure.

**Example 2.3.2.** Let $X$ and $Y$ be measurable spaces with measures and cost function such that

$$\mu = \frac{1}{3}\delta_{x_1} + \frac{2}{3}\delta_{x_2}; \qquad \nu = \frac{1}{3}\delta_{y_1} + \frac{2}{3}\delta_{y_2}; \qquad c(x_i, y_j) = c_{ij}.$$

Then $T(x_i) = y_i$ is almost surely the unique element of $\Sigma(\mu, \nu)$, hence it is solution to Monge problem and

$$\inf I(T) = \frac{1}{3}c(1, 1) + \frac{2}{3}c(2, 2).$$

On the other hand, in terms of Monge-Kantorovich problem, the result can be different. Firstly, every element in $\Pi(\mu,\nu)$ must satisfy the following conditions:
i) $J[\pi] = \sum_{i=1}^{2}\sum_{j=1}^{2}\pi_{ij}c_{ij}$
ii) $\pi_{11}+\pi_{12}=\frac{1}{3}$ ; $\pi_{21}+\pi_{22}=\frac{2}{3}$ ; $\pi_{11}+\pi_{21}=\frac{1}{3}$ ; $\pi_{12}+\pi_{22}=\frac{2}{3}$
iii)$0\le\pi_{11}+\le\frac{1}{3}; 0\le\pi_{12}+\le\frac{1}{3}; 0\le\pi_{21}+\le\frac{1}{3}; 0\le\pi_{22}+\le\frac{2}{3}$
This implies that

$$\pi_{12}=\frac{1}{3}-\pi_{11}; \pi_{21}=\frac{1}{3}-\pi_{11}; \pi_{22}=\frac{1}{3}+\pi_{11}.$$

So, the total cost is

$$J[\pi]=\pi_{11}\left(c(1,1)-c(1,2)-c(2,1)+c(2,2)\right)+\frac{1}{3}\left(c(1,2)+c(2,1)+c(2,2)\right).$$

If $(c(1,1)-c(1,2)-c(2,1)+c(2,2))\le 0$, we need to take $\pi_{11}=\frac{1}{3}$ to minimize the total cost and the result will be same as Monge problem;

$$\inf J(\pi)=\frac{1}{3}c(1,1)+\frac{2}{3}c(2,2).$$

If $(c(1,1)-c(1,2)-c(2,1)+c(2,2))\ge 0$, total cost will take minimum value at $\pi_{11}=0$ and equal to;

$$\inf J(\pi)=\frac{1}{3}\left(c(1,2)+c(2,1)+c(2,2)\right).$$

As a result, the solutions can take same or different values depending on the cost function.

Next example is from Villani [10].

**Example 2.3.3.** Let $X, Y$ be discrete spaces with measures

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

Observe that any map $T$ which satisfies $T(x_i)=y_j$ and $T(x_i)\neq T(x_j)$ if $i\neq j$ for each $i,j\in\{1,2,\dots,n\}$, will also satisfy $T\in\Sigma(\mu,\nu)$. Therefore, the solution to Monge problem is

$$\inf I(T)=\sum_{i=1}^{n}c(x_i,y_{\sigma(i)}),$$

where $\sigma$ is a permutation in $S_n$.
On the other hand, any measure $\pi \in \Sigma(\mu, \nu)$ can be repiresented as an $n \times n$ matrix $\{\pi_{ij}\}$ which satisfies

$$\pi_{ij} \geq 0; \qquad \forall j \quad \sum_{i=1}^{n} \pi_{ij} = 1; \qquad \forall i \quad \sum_{j=1}^{n} \pi_{ij} = 1$$

So, solution to Monge-Kantorovich problem is

$$\inf J(\pi) = \inf \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \pi_{ij} c(x_i, y_i); \quad \{\pi_{ij}\} \in B_n \right\},$$

where $B_n = \Big\{ M \in M_n(\mathbb{R}); M_{ij} \geq 0; \quad \forall j \quad \sum_{i=1}^{n} M_{ij} = 1; \quad \forall i \quad \sum_{j=1}^{n} M_{ij} = 1. \Big\}$
By Choquet theorem, this problem has a solution on extremal points[1] of $B_n$.
By Birkhof theorem, extremal points of $B_n$ are permutation matrices[2]. This implies that solutions to the Monge and the Monge-Kantorovich problems exist, but they are not unique, and they are equal, i.e.

$$\inf J(\pi) = \inf I(T) = \sum_{i=1}^{n} c(x_i, y_{\sigma(i)}).$$

The existence, the uniqueness and the equality of solutions to both problems depend on structure of spaces, the cost function and measures $\mu, \nu$. In following section we will introduce same important result by assuming $X = Y = \mathbb{R}^n$, $c(x, y) = |x - y|^p$, $0 < p < \infty$ and measures $\mu$ and $\nu$ have a compact support. Some of results we will show as following:

- For $p > 1$, if $\mu$ and $\nu$ are absolutely continuous, both problems have unique solution and they coincide.
- For $p > 1$, if $\mu$ vanishes at small sets[3], both problem have unique solution and they coincide.
- For $p > 1$, if $\mu$ does not vanish at small sets, although the Monge-Kantorovich problem has a solution, there is no solution to Monge Problem.

[1] Matrix $M$ is called extremal point of $B_n$, if it can't written as nontrivial convex combination of two elements in $B_n$.

[2] $M$ is called a permutation matrix, if it satisfies $M_{ij} \in \{0, 1\}$, $\forall j$ $\sum_{i=1}^{n} M_{ij} = 1$, and $\forall i$ $\sum_{j=1}^{n} M_{ij} = 1$.

[3] Small set means the set which has $n - 1$ Hausdorff dimension

- For $p = 2$, and if one of the first two assumptions hold, then the optimal transport maps are gradient of some convex functions on $\mathbb{R}$.
- For $p = 1$, if $\mu$ and $\nu$ are absolutely continuous, both problems have solutions and they coincide, however there is no unique solution.
- For $p = 1$, if $\mu$ vanishes at small sets, both problems have solutions but they do not necessarily coincide.

## 2.4 Kantorovich Duality

Kantorovich developed a very important tool, called Kantorovich Duality Theorem, which is used for both theories and applications of Monge-Kantorovich problem.
To understand the theorem better, here is a nice interpretation, introduced by Villani[10];

> "Suppose for instance that you are an industrial willing to transfer a huge amount of coal from your mines to your factories. You can hire trucks to do this transportation problem, but you have to pay them $c(x, y)$ for each ton of coal which is transported from place x to place y. Both the amount of coal which you can extract from each mine, and the amount which each factory will receive, are fixed. As you are trying to solve the associated Monge-Kantorovich problem in order to minimize the price you have to pay, another mathematician comes to you and tells you My friend, let me handle this for you: I will ship all your coal with my own trucks and you wont have to care of what goes where. I will just set a price $\phi(x)$ for loading one ton of coal at place $x$, and a price $\psi(y)$ for unloading it at destination $y$. I will set the prices in such a way that your financial interest will be to let me handle all your transportation ! Indeed, you can check very easily that for all $x$ and all $y$, the sum $\phi(x) + \psi(y)$ will always be less that the cost $c(x, y)$ (in order to achieve this goal, I am even ready to give financial compensations for some places, in the form of negative prices !). Of course you accept the deal. Now, what Kantorovichs duality

> tells you is that if this shipper is clever enough, then he can arrange the prices in such a way that you will pay him (almost) as much as you would have been ready to spend by the other method."

Formally:

**Theorem 2.4.1** (Kantorovich Duality Theorem)**.** *Let $X, Y$ be two measure spaces with probability measures $\mu$, $\nu$, and let $c(x,y)$ be a non-negative cost function defined on $X \times Y$; define a set $\Phi_c$ as;*

$$\Phi_c = \left\{(\phi, \psi) \in \mathcal{L}^1(\mu) \times \mathcal{L}^1(\nu); \phi(x) + \psi(y) \leq c(x,y) \ \mu\text{-a.s. } \forall x \in X, \ \text{ and } \nu\text{-a.s. } \forall y \in Y\right\}$$

*and define function $D : \mathcal{L}^1(\mu) \times \mathcal{L}^1(\nu) \to \mathbb{R}$ as;*

$$D(\phi, \psi) = \int_X \phi(x) d\mu(x) + \int_Y \psi(y) d\nu(y).$$

*Then*

$$\inf_{\Pi(\mu,\nu)} J(\pi) = \sup_{\Phi_c} D(\phi, \psi).$$

$\sup_{\Phi_c} D(\phi, \psi)$ *is called dual of* $\inf_{\Pi(\mu,\nu)} J(\pi)$.

Following Villani [10], we will give outline of the proof of this theorem. The key point of proof is use of *Minimax principle* which allows us to change the order of *infimum* and *supremum* under some conditions.

**Theorem 2.4.2** (Minimax principle [3])**.** *Let $X$ be a compact Hausdorff space and $Y$ an arbitrary set (not topologized). Let $f$ be a real-valued function on $X \times Y$ such that, for every $y \in Y$, $f(x,y)$ is lower semi-continuous on $X$. If $f$ is convex on $X$ and concave on $Y$, then*

$$\inf_{x \in X} \sup_{y \in Y} f(x,y) = \sup_{y \in Y} \inf_{x \in X} f(x,y)$$

*Proof of 2.4.1.* Let $M^+(X \times Y)$ denotes the set of all positive borel measures from $X$ to $Y$. Observe that

$$\inf_{\Pi(\mu,\nu)} J(\pi) = \inf_{M^+(X \times Y)} \left[ J(\pi) + \begin{cases} 0 & \pi \in \Pi(\mu,\nu) \\ +\infty & \text{else} \end{cases} \right]$$

$$
\begin{aligned}
&= \inf_{M^+(X\times Y)} \Bigg[ J(\pi) + \sup_{C_b(X)\times C_b(Y)} \Bigg[ \int_X \phi(x)d\mu(x) + \int_Y \psi(y)d\nu(y) \\
&\qquad - \int_{X\times Y} (\phi(y)+\psi(y))\, d\pi(x,y) \Bigg] \\
&= \inf_{M^+(X\times Y)} \sup_{C_b(X)\times C_b(Y)} \Bigg[ J(\pi) + \int_X \phi(x)d\mu(x) + \int_Y \psi(y)d\nu(y) \\
&\qquad - \int_{X\times Y} (\phi(y)+\psi(y))\, d\pi(x,y) \Bigg] \\
&= \sup_{C_b(X)\times C_b(Y)} \inf_{M^+(X\times Y)} \Bigg[ J(\pi) + \int_X \phi(x)d\mu(x) + \int_Y \psi(y)d\nu(y) \\
&\qquad - \int_{X\times Y} (\phi(y)+\psi(y))\, d\pi(x,y) \Bigg] \qquad \text{by 2.4.2} \\
&= \sup_{C_b(X)\times C_b(Y)} \Bigg[ \int_X \phi(x)d\mu(x) + \int_Y \psi(y)d\nu(y) \\
&\qquad - \sup_{M^+(X\times Y)} \Bigg[ \int_{X\times Y} (\phi(y)+\psi(y)-c(x,y))\, d\pi(x,y) \Bigg] \Bigg] \qquad (4.1)
\end{aligned}
$$

Define $f(x,y) = \phi(x) + \psi(y) - c(x,y)$. If $f(x,y) > 0$ for some $(x_0, y_0)$, then by choosing $\pi_\lambda = \lambda\delta_{(x_0,y_0)} \in M^+(X\times Y)$, $\int_{X\times Y} f(x,y)d\pi_\lambda$ goes to $\infty$ as $\lambda$ goes to $\infty$. Hence $\sup \int_{X\times Y} f(x,y)d\pi(x,y) = \infty$. On the other hand, if $f(x) \leq 0$, then supremum attains at $\pi = 0 \in M^+(X\times Y)$ and is equal to 0. Hence

$$
\begin{aligned}
4.1 &= \sup_{C_b(X)\times C_b(Y)} \left[ \int_X \phi(x)d\mu(x) + \int_Y \psi(y)d\nu(y) - \left\{ \begin{array}{ll} 0 & (\phi,\psi)\in\Phi_c \\ +\infty & \text{else} \end{array} \right. \right] \\
&= \sup_{\Phi_c} D(\phi,\psi).
\end{aligned}
$$

□

# Chapter 3

# Solutions to problems

## 3.1 The quadratic cost function

In this section, we let $X = Y = \mathbb{R}^n$ and define the cost function $c(x,y) = \frac{|x-y|^2}{2}$. And finally, we let $\mu$, $\nu$ be Borel Probability measure with finite second moments, i.e.

$$\int_{\mathbb{R}^n} \frac{|x|^2}{2} d\mu(x) + \int_{\mathbb{R}^n} \frac{|y|^2}{2} d\nu(y) = M < \infty. \tag{1.1}$$

### 3.1.1 The existence of solution to Monge-Kantorovich problem

**Lemma 3.1.1.** $\Pi(\mu,\nu)$ *is compact for weak topology of probability measures.*[1]

*Proof.* We already know that $\mu$ and $\nu$ are tight, therefore, for $\epsilon > 0$ there exist compact sets $K_\epsilon, L_\epsilon \subset \mathbb{R}^n$ such that $\mu(K_\epsilon^c) < \epsilon$ and $\nu(L_\epsilon^c) < \epsilon$. Now let $\pi \in \Pi(\mu,\nu)$. Then

$$\pi\left((K_\epsilon \times L_\epsilon)^c\right) \leq \pi\left(\mathbb{R}^n \times L_\epsilon^c\right) + \pi\left(K_\epsilon^c \times \mathbb{R}^n\right) \leq 2\epsilon$$

[1] *The topologoy induced by* $C_b(\mathbb{R}^n_+ \times \mathbb{R}^n_+)$.

So $\Pi(\mu,\nu)$ is tight, hence relatively compact. On the other hand $\Pi(\mu,\nu)$ is clearly closed. These two imply that $\Pi(\mu,\nu)$ is compact. $\square$

**Proposition 3.1.2.** *Monge-Kantorovich problem admits a minimizer,i.e. there exist $\pi' \in \Pi(\mu,\nu)$ such that*

$$J(\pi') = \inf_{\Pi(\mu,\nu)} J(\pi).$$

**Proof:** Let $(\pi_k)_{k\in\mathbb{N}} \subset \Pi(\mu,\nu)$ be minimizing sequence, i.e. $\lim J(\pi_k) = \inf_{\Pi(\mu,\nu)} J(\pi)$ as $k \to \infty$, then it admits a cluster point $\pi' \in \Pi(\mu,\nu.)$ On the other hand, we can find a sequence of nondecreasing bounded continuous functions $(c_l(x,y))_{l\in\mathbb{N}}$ which converges to $c(x,y)$. So, we have

$$\begin{aligned}
J(\pi') &= \lim_{l\to\infty} \int_{X\times Y} c_l(x,y)d\pi'(x,y) && \text{by monotone convergence theorem}\\
&\leq \lim_{l\to\infty} \limsup_{k\to\infty} \int_{X\times Y} c_l(x,y)d\pi_k(x,y) && \text{since } \pi' \text{ is a cluster point}\\
&\leq \limsup_{k\to\infty} \int_{X\times Y} c(x,y)d\pi_k(x,y) && \text{since } c_l(x,y) \leq c(x,y)\\
&= \inf_{\Pi(\mu,\nu)} J(\pi)
\end{aligned}$$

Hence $\pi'$ is an optimal transportation plan for Monge-Kantorovich problem. A more elegant result will be given in following sections.

### 3.1.2 More about Kantorovich Duality

Due to our assumptions at the beginning of the section we will have a more elegant result about Kantorovich Duality. Again we will follow th book of villani [10]

By definition $(\phi,\psi) \in \Phi_c$ if and only if

$$\phi(x) + \psi(y) \leq \frac{|x-y|}{2},$$

and this is true if and only if

$$x.y \leq \left(\frac{|x|^2}{2} - \phi(x)\right) + \left(\frac{|y|^2}{2} - \psi(y)\right).$$

So $(\phi, \psi) \in \Phi_c$ if and only if $(\widetilde{\phi}, \widetilde{\psi}) \in \widetilde{\Phi_c}$ where $\widetilde{\phi}(x) = \frac{|x|^2}{2} - \phi(x)$, $\widetilde{\psi}(y) = \frac{|y|^2}{2} - \psi(y)$ and

$$\widetilde{\Phi_c} = \left\{(\phi, \psi) \in \mathcal{L}(\mu)^1 \times \mathcal{L}(\nu)^1;\ x.y \leq \phi(x) + \psi(y)\right\}. \tag{1.2}$$

So by using 1.1, we have

$$\begin{aligned}\sup_{\Phi_c} D(\phi, \psi) &= \sup_{\widetilde{\Phi}_c} \left(M - \int_{\mathbb{R}^n} \phi(x) d\mu(x) + \int_{\mathbb{R}^n} \psi(x) d\nu(y)\right) \\ &= M - \inf_{\widetilde{\Phi}_c} D(\phi, \psi)\end{aligned}$$

On the other hand

$$\begin{aligned}\inf_{\Pi(\mu,\nu)} J(\pi) &= \inf_{\Pi(\mu,\nu)} \left(\int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{|x-y|^2}{2} d\pi(x,y)\right) \\ &= M - \left(\sup_{\Pi(\mu,\nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} x.y d\pi(x,y)\right)\end{aligned}$$

Finally we can rewrite Duality Theorem as

$$\begin{aligned}\sup_{\Pi(\mu,\nu)} \left(\int_{\mathbb{R}^n \times \mathbb{R}^n} x.y d\pi(x,y)\right) &= \inf_{\widetilde{\Phi}_c} \left(\int_{\mathbb{R}^n} \phi(x) d\mu(x) + \int_{\mathbb{R}^n} \psi(x) d\nu(y)\right) \\ &= \inf_{\widetilde{\Phi}_c} J(\phi, \psi). \end{aligned} \tag{1.3}$$

Now define $\phi^*(y) = \sup_{x \in X} (x.y - \phi(x))$. This transform is called Legendre-Fenchel transform. From 1.3, $(\phi, \psi) \in \widetilde{\Phi}_c$ implies for all $y \in Y$ $\nu$-almost surely

$$\psi(y) \geq \phi^*(y). \tag{1.4}$$

Moreover, it is clear that

$$\phi(x) + \phi^*(y) \geq xy. \tag{1.5}$$

Hence $\mu-$almost surely

$$\phi(x) \geq \sup_{y \in Y} (x.y - \phi^*(y)) = \phi^{**}(x). \tag{1.6}$$

**Lemma 3.1.3.** *Let $\mu$ and $\nu$ be probability measures with second finite moments and they are supported in subsets $X$ and $Y$ of $\mathbb{R}^n$. For each measurable functions $\phi$, $\psi$ with values in $\mathbb{R} \cup \{+\infty\}$, define*

$$\phi^*(y) = \sup_{x \in X} (x.y - \phi(x))$$

$$\psi^*(x) = \sup_{y\in Y}\left(x.y - \psi(y)\right).$$

*Let $\widetilde{\Phi_c}$ be defined by 1.2 and let $(\phi_k, \psi_k)_{k\in\mathbb{N}}$ be a minimizing sequence for $D$ over $\widetilde{\Phi_c}$. Then*
*(i) one can modify $\phi_k$, $\psi_k$ on sets $M_k$, $N_k$ of measure zero such that the equation $x.y \leq \phi_k + \psi_k$ holds for each $x \in X$, $y \in Y$ without changing the value of $D(\phi_k, \psi_k)$.*
*(ii) There exists a sequence of real numbers $(a_k)_{k\in\mathbb{N}}$ such that*

$$(\overline{\phi_k}, \overline{\psi}_k) = (\phi_k^{**} - a_k, \psi_k^* + a_k)$$

*it is still a minimizing sequence for $D$ over $\widetilde{\Phi_c}$ and satisfying followings:*

$$\overline{\phi_k}(x) \geq -\frac{|x|^2}{2}\ \forall x\in X \quad and \quad \overline{\psi}_k(y) \geq -\frac{|y|^2}{2}\ \forall y \in Y$$

$$\liminf_{k\to\infty}\ \inf_{x\in X}\overline{\phi_k}(x) + \frac{|x|^2}{2} \leq \inf_{\widetilde{\Phi}_c} D(\phi,\psi) + M$$

$$\liminf_{k\to\infty}\ \inf_{y\in Y}\overline{\psi}_k(y) + \frac{|y|^2}{2} \leq \inf_{\widetilde{\Phi}_c} D(\phi,\psi) + M$$

*(iii) If $X = Y = \mathbb{R}^n$, then*

$$\inf_{\widetilde{\Phi}_c} D(\phi,\psi) = \inf_{\mathcal{L}^1(\mu)} D(\phi^{**},\phi^*)$$

*Proof. (iii)* From 1.5 if we show that $\phi^* \in \mathcal{L}^1(\nu)$, we will have $(\phi, \phi^*) \in \widetilde{\Phi_c}$. Firstly $(\phi,\psi) \in \widetilde{\Phi_c}$ implies that $\psi(y) \geq xy - \phi(x)$, hence $\psi(y) \geq \phi^*(y)$ On the other hand there exist $(x_0, b_0) \in \mathbb{R}^n \times \mathbb{R}$ such that $\phi^*(y) \geq x_o y - b_0 \geq \frac{-1}{2}\left(|x_0|^2 + |y|^2 + b_0\right)$ combining these two, we get $|\phi^*(y)| \leq \max\left(|\psi(y)|, \frac{1}{2}\left(|x_0|^2 + |y|^2 + |b_0|\right)\right)$ for each $y \in Y$ $\nu$-almost surely. Hence, letting $A = \{y \in \mathbb{R} : |\psi(y)| \leq \frac{1}{2}(|x_0|^2 + |y|^2 + |b_0|)\}$

$$\begin{aligned}\int_{\mathbb{R}^n}\phi^*(y)d\nu(y) &\leq \int_A \frac{1}{2}\left(|x_0|^2 + |y|^2 + |b_0|\right)d\nu(y) + \int_{A^c}\psi(y)d\nu(y)\\ &\leq \int_{\mathbb{R}^n}\frac{1}{2}\left(|x_0|^2 + |y|^2 + |b_0|\right)d\nu(y) + \int_{\mathbb{R}^n}\psi(y)d\nu(y) < \infty\end{aligned}$$

So $(\phi,\phi^*) \in \widetilde{\Phi_c}$, and for each $(\phi,\psi) \in \widetilde{\Phi_c}$, we have $D(\phi,\phi^*) \leq D(\phi,\psi)$ since $\psi(y) \geq \phi^*(y)$. By the same reason, $(\phi^{**},\phi^*) \in \widetilde{\Phi_c}$ and $D(\phi^{**},\phi^*) \leq D(\phi,\psi)$. Therefore

$$\inf_{\widetilde{\Phi}_c} D(\phi,\psi) = \inf_{\mathcal{L}^1(\mu)} D(\phi^{**},\phi^*)$$

*(i)* By definition of $\widetilde{\Phi_c}$, $(\phi_k, \psi_k) \in \widetilde{\Phi_c}$ implies that there exist sets $M_k$, and $N_k$ with zero measure such that $x.y \leq \phi_k + \psi_k$ holds for all $(x, y) \in M_k^c \times N_k^c$. Next we change values of $\phi_k$, and $\psi_k$ to be $+\infty$ on $M_k$, and $N_k$ respectively. We still have $(\phi_k, \psi_k) \in \widetilde{\Phi_c}$ but value of $D$ did not change since we worked on sets with zero measure.

*(ii)* We know that $\psi_k$ is not identically $+\infty$ and by part *(i)* for each $x$, $y$ we have $\phi_k(x) \geq x.y - \psi_k(y)$. Hence there exist $y_0 \in Y$ and $b_0 \in \mathbb{R}$ such that $\phi_k(x) \geq x.y_0 - b_0$. for all $x$. Hence

$$\phi_k^*(y_o) = \sup_{x \in X} (x.y_0 - \phi_k(x)) \leq -b_0,$$

and $\phi_k^*$ is a proper function. Moreover, since $\phi_k$ is proper, there exist $x_0$ and $d_0$ such that for all $y$ $\phi_k^*(y) \geq x_0.y - d_0$. So,

$$a_k = \inf_{y \in Y} \left( \frac{|y|^2}{2} + \phi_k^*(y) \right)$$

is finite. Now define:

$$\overline{\phi_k} = \phi_k^{**} + a_k$$
$$\overline{\psi}_k = \phi_k^* - a_k$$

Clearly $\forall y \in Y$, we have

$$\overline{\psi}_k(y) \geq -\frac{|y|^2}{2} \tag{1.7}$$

On the other hand, for all $x \in X$

$$\begin{aligned}
\overline{\phi_k}(x) + \frac{|x|^2}{2} &= \sup_{y \in Y} \left( x.y - \phi_k^*(y) + \frac{|x|^2}{2} \right) + a_k \\
&\geq \sup_{y \in Y} \left( -\phi_k^*(y) - \frac{|y|^2}{2} \right) + a_k \\
&= -\inf_{y \in Y} \left( \phi_k^*(y) + \frac{|y|^2}{2} \right) + a_k = 0
\end{aligned} \tag{1.8}$$

So we have done with lower bounds. Now we will deal with upper bounds. By 1.4, 1.6 and the definition of function $D$ we have

$$D(\overline{\phi_k}, \overline{\psi}_k) = D(\phi_k^{**} + a_k, \phi_k^* - a_k) = D(\phi_k^{**}, \phi_k^*) \leq D(\phi_k, \psi_k) < \infty. \tag{1.9}$$

So,

$$0 \leq \int_X \left( \overline{\phi_k}(x) + \frac{|x|^2}{2} \right) d\mu + \int_Y \left( \overline{\psi}_k(y) + \frac{|y|^2}{2} \right) d\nu$$

$$
\begin{aligned}
&\leq \int_X \left(\phi_k(x) + \frac{|x|^2}{2}\right) d\mu + \int_Y \left(\psi_k(y) + \frac{|y|^2}{2}\right) d\nu \\
&= D(\phi_k, \psi_k) + M < \infty
\end{aligned}
\tag{1.10}
$$

1.7 and 1.8 imply first and 1.9 implies the second inequality.
Hence $(\overline{\phi_k}, \overline{\psi}) \in \mathcal{L}(\mu)^1 \times \mathcal{L}(\nu)^1$, and by their definition, $(\overline{\phi}, \overline{\psi}) \in \widetilde{\Phi}_c$. So, if we take limit of both sides of 1.10, we see that $(\overline{\phi_k}, \overline{\psi})$ is a minimizing sequence for $D$ over $\widetilde{\Phi_c}$. Moreover,

$$
\begin{aligned}
D(\overline{\phi_k}, \overline{\psi}_k) + M &= \int_X \left(\overline{\phi_k}(x) + \frac{|x|^2}{2}\right) d\mu + \int_Y \left(\overline{\psi}_k(y) + \frac{|y|^2}{2}\right) d\nu \\
&\geq \inf_X \left(\overline{\phi_k}(x) + \frac{|x|^2}{2}\right) + \inf_Y \left(\overline{\psi}_k(y) + \frac{|y|^2}{2}\right)
\end{aligned}
$$

Since terms on the right side are nonnegative we have

$$
\begin{aligned}
D(\overline{\phi_k}, \overline{\psi}_k) + M &\geq \inf_X \left(\overline{\phi_k}(x) + \frac{|x|^2}{2}\right) \\
D(\overline{\phi_k}, \overline{\psi}_k) + M &\geq \inf_Y \left(\overline{\psi}_k(y) + \frac{|y|^2}{2}\right)
\end{aligned}
$$

Since $(\overline{\phi_k}, \overline{\psi}_k)$ is minimizing sequence for $D$ over $\widetilde{\Phi_c}$, we have

$$
\begin{aligned}
\inf_{\widetilde{\Phi}_c} D(\phi, \psi) + M &\geq \liminf_{k\to\infty} \inf_X \left(\overline{\phi_k}(x) + \frac{|x|^2}{2}\right) \\
\inf_{\widetilde{\Phi}_c} D(\phi, \psi) + M &\geq \liminf_{k\to\infty} \inf_Y \left(\overline{\psi}_k(y) + \frac{|y|^2}{2}\right).
\end{aligned}
$$

□

**Lemma 3.1.4.** *Let $\mu$, $\nu$ be probability measure with second finite moments. Then there exist a pair $\phi, \phi^*$ of semi-continuous convex function such that*

$$
\inf_{\widetilde{\Phi}_c} D(\phi, \psi) = D(\phi, \phi^*)
$$

*Proof.* **(Convex Settings)** Assume that $\mu$ and $\nu$ are supported in subsets $X$ and $Y$ of $\mathbb{R}^n_+$. Let $(\phi_k, \psi_k)_{k\in\mathbb{N}}$ be a minimizing sequence for $D$ over $\widetilde{\Phi_c}$, and $(\overline{\phi_k}, \overline{\psi})$ be sequence that satisfies conditions in lemma 3.1.3. By definition, we know that $\overline{\phi_k}$ is uniformly lipstick, i.e.

$$
\|\overline{\psi}_k\|_{Lip(X)} \leq \sup_Y |y|.
$$

Moreover by lemma 3.1.3, there sexist $x_k \in X$

$$-\sup_X \frac{|x|^2}{2} \le \overline{\psi}_k(x_k) \le \sup_X \frac{|x|^2}{2} + \inf_{\widetilde{\Phi}_c} D + M + 1.$$

Since $\overline{\phi_k}$ is uniformly lipstick, we decide that for all $x \in X$

$$\begin{aligned}\overline{\psi}_k(x) &\le \overline{\phi_k}(x_k) + \sup_X |x|^2 \\ &\le \sup_X \frac{3|x|^2}{2} + \inf_{\widetilde{\Psi}_c} D + M + 1.\end{aligned}$$

And

$$\begin{aligned}\overline{\psi}_k(x_k) - \sup_X |x|^2 &\le \overline{\psi}_k(x) \\ -\sup_X \frac{3|x|^2}{2} &\le \overline{\psi}_k(x).\end{aligned}$$

Since $X$ is compact, $\overline{\psi}_k$ is uniformly bounded. Similarly $\overline{\phi_k}$ is uniformly bounded. Moreover $(\overline{\phi_k})_{k\in\mathbb{N}}$, $(\overline{\psi}_k)_{k\in\mathbb{N}}$ are both sequences of equicontinuous (since uniformly Lipschitz) functions defined on compact sets of a separable metric space since they are uniformly Lipschitz. Therefore we can apply Arzela-Ascoli theorem. Hence, there exist subsequences $(\overline{\phi}_{k_n})$, $(\overline{\psi}_{k_n})$ that converge uniformly to $\overline{\phi} \in C_b(X)$, $\overline{\psi} \in C_b(Y)$ respectively. These convergences also holds in $\mathcal{L}_1$. Therefore,

$$\inf_{\widetilde{\Phi}_c} D \le D(\overline{\phi}^{**}, \overline{\phi}^{*}) \le D(\overline{\phi}, \overline{\psi}) \le \lim_{n\to\infty} D(\overline{\phi}_{k_n}, \overline{\psi}_{k_n}) \le \lim_{n\to\infty} D(\phi_{k_n}, \psi_{k_n}) \le \inf_{\widetilde{\Phi}_c} D$$

This completes the proof of lemma in compact settings. For general case please see the references. □

### 3.1.3 Theorems for quadratic case

In this section we will prove main theorems for Quadric cost function. In order to do this, we need to use some convex analysis. Therefore, if you are not familiar with convex analysis, please see the appendix for some necessary background information.

**Theorem 3.1.5** (Knott-Smith criterion)**.** *Let $\mu$, $\nu$ be probability measure with second finite moments, and define the cost function $c(x, y) = \frac{|x-y|^2}{2}$. The measure*

*$\pi \in \Pi(\mu,\nu)$ is optimal transportation plan if and only if there exists a convex lower-semi continuous function $\phi$ such that*

$$supp(\pi) \subset \partial\phi,$$

*in other words;*

$$y \in \partial\phi(x),\ \pi \text{ almost surely.}$$

*Moreover, we have*

$$\inf_{\widetilde{\Phi}_c} D(\phi,\psi) = D(\phi,\phi^*)$$

*Proof.* Assume that $\pi \in \Pi(\mu,\nu)$ is an optimal transportation plan. By lemma 3.1.4, there exists convex lower semi-continuous function $\phi$ such that $(\phi,\phi^*) \in \widetilde{\Phi_c}$ and minimize $D$ over $\widetilde{\Phi_c}$. So we have

$$\begin{aligned}\int_{\mathbb{R}^n\times\mathbb{R}^n} x.y\ d\pi(x,y) &= \int_{\mathbb{R}^n} \phi\ d\mu(x) + \int_{\mathbb{R}^n} \phi^*\ d\nu(y) && \text{by 1.3}\\ &= \int_{\mathbb{R}^n\times\mathbb{R}^n} (\phi+\phi^*)\ d\pi(x,y) && \text{since } \pi \in \Pi(\mu,\nu)\end{aligned}$$

Hence we have

$$\int_{\mathbb{R}^n\times\mathbb{R}^n} (\phi+\phi^*-x.y)\ d\pi(x,y) = 0$$

Since $(\phi,\phi^*) \in \widetilde{\Phi_c}$, integrand is non-negative. Hence $\phi+\phi^* = x.y$ $\pi$-almost surely. This is equivalent to, by A.0.6

$$\text{supp}(\pi) \subset \partial\phi,$$

or

$$y \in \partial\phi(x),\ \pi \text{ almost surely.}$$

On the contrary, assume that $\pi \in \Pi(\mu,\nu)$ and there exists a convex lower-semi continuous function $\phi$ such that

$$\text{supp}(\pi) \subset \partial\phi.$$

By the following each above steps in reverse order, we have

$$\int_{\mathbb{R}^n\times\mathbb{R}^n} x.y\ d\pi(x,y) = \int_{\mathbb{R}^n} \phi\ d\mu(x) + \int_{\mathbb{R}^n} \phi^*\ d\nu(y) by \qquad \text{by A.0.6}$$

$$\geq \inf_{\widetilde{\Psi_c}} D \qquad\qquad \text{since } (\phi, \phi^*) \in \widetilde{\Phi_c}$$

$$= \sup_{\Pi(\mu,\nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} x.y \; d\pi(x,y) \qquad\qquad \text{by 1.3}$$

$$\geq \int_{\mathbb{R}^n \times \mathbb{R}^n} x.y \; d\pi(x,y)$$

So there is equality for everywhere, hence $\pi$ is optimal transportation plan. □

**Theorem 3.1.6** (Breniers theorem)**.** *Let $\mu$, and $\nu$ be probability measures with second finite moments, and define the cost function $c(x,y) = \frac{|x-y|^2}{2}$. If $\mu$ vanish on small sets, then there exists a unique optimal transportation plan $\pi \in \Pi(\mu,\nu)$ such that*

$$\pi = (id \times \bigtriangledown\phi)_{\#}\, \mu,$$

*where $\bigtriangledown\phi$ is unique gradient of convex function such that $\bigtriangledown\phi_{\#}\mu = \nu$.*

*Proof.* By 3.1.2 and 3.1.5, we already know that there exist optimal transportation plan $\pi \in \Pi(\mu,\nu)$, and a convex lower-semi continuous function $\phi$ such that $\mathrm{spt}(\pi) \subset \partial\phi$. Since $\phi$ is integrable, it is $\mu$-almost surely finite. Hence $\mu\,(\mathrm{Dom}\,(\phi)) = 1$. Moreover, the boundary of $\mathrm{Dom}\,(\phi)$ is a small set, so by assumption it has zero $\mu$ measure. Hence $\mu\,(\mathrm{int}\,(\mathrm{Dom}\,(\phi))) = 1$. Moreover $\phi$ is differentiable everywhere except small set by A.0.7. By A.0.3, we know that differentiability at $x$ implies unique sub-gradient at $x$, which is $\bigtriangledown\phi(x)$. As a result we have $\partial\phi(x) = \{\nabla\phi(x)\}$ for all $x \in X$ $\mu$ almost surely. Therefore, we have $\pi$-almost surely

$$(x,y) \in \mathrm{spt}(\pi) \Rightarrow (x,y) = (x, \nabla\phi(x))$$

Which implies

$$d\pi(x,y) = \mu(x).\delta_{(x-\nabla\phi(x))}.$$

Equivalently $\pi' = (id \times \bigtriangledown\phi)_{\#}\, \mu$, and $\bigtriangledown\phi_{\#}\mu = \nu$.

Next we will prove uniqueness part. Let $\pi$ and $\pi_2$ be two optimal transportation plan such that $\pi = (id \times \bigtriangledown\phi)_{\#}\, \mu$, and $\pi_2' = (id \times \bigtriangledown\phi_2)_{\#}\mu$. By 3.1.5, both $(\phi, \phi^*)$ and $(\phi_2, \phi_2^*)$ are solution to dual problem, i.e.

$$\inf_{\widetilde{\Phi}_c} D(\phi,\phi^*) = D(\phi,\phi^*) = \int_{\mathbb{R}^n} \phi(x) d\mu(x) + \int_{\mathbb{R}^n} \phi^*(x) d\nu(y)$$

$$= D(\phi_2,\phi_2^*) = \int_{\mathbb{R}^n} \phi_2(x) d\mu(x) + \int_{\mathbb{R}^n} \phi_2^*(y) d\nu(y)$$

$$= \int_{\mathbb{R}^n \times \mathbb{R}^n} x.y \; d\pi(x,y) \qquad \text{by 1.3}$$

Since $\pi = (id \times \triangledown\phi)_{\#}\,\mu,$, we have

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} (\phi_2(x) + \phi_2^*(\triangledown\phi(x))) \; d\mu(x) = \int_{\mathbb{R}^n \times \mathbb{R}^n} x. \triangledown \phi(x) \; d\mu(x)$$

Hence

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} (\phi_2(x) + \phi_2^*(\triangledown\phi(x)) - x. \triangledown \phi(x)) \; d\mu(x) = 0$$

Since integrand is finite, we have $\phi_2(x) + \phi_2^*(\triangledown\phi(x)) - x. \triangledown \phi(x) = 0$ almost surely. So A.0.6 implies $\triangledown\phi(x) \in \partial\phi_2(x)$, and A.0.3 implies $\triangledown\phi(x) = \triangledown\phi_2(x)$ $\mu$-almost surely . □

### 3.1.4 Cyclical monotonicity

Until now we have studied the quadratic cost function under the assumption of finite second order moments. In order to receive more general result for the quadratic cost, we need to introduce a geometrical idea, namely the concept of cyclical monotonicity.

**Definition 3.1.7** (Cyclically monotone ). A subset $\Gamma \in \mathbb{R}^n \times \mathbb{R}^n$ is called Cyclically monotone if it satisfies the condition; for all $m \geq 1$ and for all $(x_1, y_1), \ldots, (x_m, y_m) \in \Gamma$, we have

$$\sum_{i=1}^{m} |x_i - y_i|^2 \leq \sum_{i=1}^{m} |x_i - y_{i-1}|^2$$

where $y_0 = y_m$ or equivalently

$$\sum_{i=1}^{m} y_i.(x_{i+1} - x_i) \leq 0.$$

The concept of cyclical monotonicity will allow us to prove of one way of Knott-Smith optimality criterion without assumption of finite second moments. This will be enough to prove generalized Brenier's theorem.

**Theorem 3.1.8** (Rockafellar [7]). *A nonempty subset $\Gamma \subset \mathbb{R}^n \times \mathbb{R}^n$ is cyclically monotone if and only if there exists a proper convex function $\phi$ on $\mathbb{R}^n$ such that $\Gamma \subset \partial\phi$.*

*Proof.* First, assume that $\Gamma$ is cyclically monotone and let $(x_0, y_0)$ be any element in $\Gamma$. Now define

$$\phi(x) = \sup\{y_m.(x - x_m) + y_{m-1}.(x_m - x_{m-1}) + \cdots + y_0.(x_1 - x_0); m \in \mathbb{N}, \\ (x_i, y_i) \in \Gamma \, \forall\, 1 \leq i \leq m\}$$

Note first that $\phi$ is a supremum of a nonempty collection of affine functions, hence it is a lower semi-continuous convex function. Moreover, $\phi(x_0) \leq 0$ since $\Gamma$ is cyclically monotone, which means $\phi$ is proper. Next, we will show that $\Gamma$ is subset of $\partial\phi$, i.e., if $(x, y) \in \Gamma$, then we have

$$\phi(z) \geq \phi(x) + y.(z - x), \ \forall z \in \mathbb{R}^n.$$

First, let $a$ be any number less than $\phi(x)$. By definition of $\phi$, there exist $m \in \mathbb{N}$ and $(x_1, y_1) \ldots (x_m, y_m)$ such that

$$a \leq y_m.(x - x_m) + y_{m-1}.(x_m - x_{m-1}) + \cdots + y_0.(x_1 - x_0).$$

Thus,

$$\begin{aligned} a + y.(z - x) &\leq y.(z - x) + y_m.(x - x_m) + y_{m-1}.(x_m - x_{m-1}) + \cdots + y_0.(x_1 - x_0) \\ &\leq \phi(z). \end{aligned}$$

Last inequality come from the defition of $\phi$. Since we choose $a$ arbitrarily, we have

$$\phi(z) \geq a + y.(z - x), \ \forall z \in \mathbb{R}^n, \ \forall a < \phi(x).$$

This implies

$$\phi(z) \geq \phi(x) + y.(z - x), \ \forall z \in \mathbb{R}^n.$$

and $\Gamma$ is a subset of $\partial\phi$.
Conversely, assume $\Gamma \subset \partial\phi$ where $\phi$ is a proper convex function on $\mathbb{R}^n$. We will show that $\partial\phi$ is cyclically monotone subset of $\mathbb{R}^n \times \mathbb{R}^n$. Then our claim will follow. Let $(x_1, y_1), \ldots, (x_m, y_m) \in \partial\phi$. By the definition of subdifferential we have $\forall z \in \mathbb{R}^n$.

$$\phi(z) \geq \phi(x_1) + y_1.(z - x_1)$$

Hence,

$$\phi(x_2) \geq \phi(x_1) + y_1.(x_2 - x_1)$$
$$\phi(x_3) \geq \phi(x_2) + y_2.(x_3 - x_2)$$
$$\vdots$$
$$\phi(x_1) \geq \phi(x_m) + y_m.(x_1 - x_m)$$

Adding up these inequalities, we will observe

$$\sum_{i=1}^{m} y_i.(x_{i+1} - x_i) \leq 0.$$

This is just the definition of cyclically monotone set. Therefore, $\partial\phi$ is cyclically monotone. In particular, $\Gamma$ is cyclically monotone since, it is subset of a cyclically monotone set. $\square$

**Theorem 3.1.9.** *Let $\mu$ and $\nu$ be probability measures on $\mathbb{R}^n$, and define the cost function $c(x,y) = |x-y|^2$. If $\pi \in \Pi(\mu,\nu)$ is optimal transportation plan then support of $\pi'$ is cyclically monotone.*

**Remark 3.1.10.** Please note that this theorem is not a characterization of optimal transportation plan since converse of it is still an open problem.

From 3.1.8 and 3.1.9 we can decude the following result.

**Theorem 3.1.11.** *Let $\mu$ and $\nu$ be probability measure on $\mathbb{R}^n$, and define the cost function $c(x,y) = \frac{|x-y|^2}{2}$. If $\pi \in \Pi(\mu,\nu)$ is optimal transportation plan then the support of $\pi$ is supported in the sub-differential of a proper lower semi-continuous convex function.*

This is one direction of Knott-Smith optimality criterion but we did not need any assumptions of finiteness of the second moments.

**Theorem 3.1.12** (Breniers theorem, extended version)**.** *Let $\mu$ and $\nu$ be probability measure and define the cost function $c(x,y) = |x-y|^2$. If $\mu$ vanish on small sets, then there exists unique optimal transportation plan $\pi \in \Pi(\mu,\nu)$ such that*

$$\pi = (id \times \nabla\phi)_{\#}\,\mu,$$

*where $\nabla\phi$ is unique gradient of convex function such that $\nabla\phi_{\#}\mu = \nu$.*

*Proof.* Recall that we only use one direction of Knott-Smith criterion 3.1.5, which is what we just proved without the assumption of finiteness of the second moments, in the existence part of Brenier's theorem 3.1.6. Hence, the proof of existence part of Brenier's theorem 3.1.6 also holds for this theorem.
However, for the uniqueness part of Brenier's theorem, we used 3.1.4 which requires the assumption of finiteness of the second moments. Hence, we need help of Aleksandrovs lemma A.0.8 in order to prove uniqueness
Now, let $\pi$ and $\pi_2$ be two optimal transportation plans such that $\pi = (id \times \bigtriangledown\phi)_{\#}\,\mu$, and $\pi_2 = (id \times \bigtriangledown\phi_2)_{\#}\mu$ and $\bigtriangledown\phi \neq \bigtriangledown\phi_2$ $\mu-$almost surely. Then, there exists $x_0 \in \operatorname{spt}(\mu)$ such that $\bigtriangledown\phi(x_0) \neq \bigtriangledown\phi_2(x_0)$. Adding some constant number to $\phi$ and $\phi_2$, we can ensure that $\phi(x_0) = \phi_2(x_0) = 0$, and this will not affect the gradients. By A.0.9 , there exists a small neighborhood $U$ of $x_0$ such that $U \cap \{\phi = \phi_2\}$ has $d-1$ Hausdorff dimension, and it has zero $\mu$ measure by assumption. On the other hand, $U$ has a positive measure since it contains $x_0$. Hence, it intersects $\{\phi > \phi_2\}$ or $\{\phi < \phi_2\}$ with positive $\mu$ measure. Exchanging $\{\phi, \phi_2\}$ if necessary, we can say that $U$ intersects $V = \{\phi > \phi_2\}$ with positive $\mu$ measure.

By lemma A.0.8, we know that $x_0$ lies in a positive distance away from $Z = \nabla\phi_2{}^{-1}\left(\partial\phi(V)\right)$. So, we can find some neighborhood $U_2$ of $x_0$ such that $U_2 \cap Z = \emptyset$. Let $T = U \cap U_2$.
Now, since $V \cap \mathrm{Dom}\phi \subseteq Z$ and $1 = \nu(\mathbb{R}^n) = \nabla\phi_{\#}\mu(\mathbb{R}^n) = \mu(\nabla\phi^{-1}(\mathbb{R}^n)) = \mu(\mathrm{Dom}\nabla\phi)$, we have

$$\mu(V) \leq \mu(Z).$$

On the other hand, we know $Z \subseteq V$ from A.0.8, hence $\mu(Z) \leq \mu(V)$. Moreover $T \cap Z = \emptyset$ and $\mu(T \cap V) > 0$, hence

$$\mu(Z) < \mu(V).$$

We observed

$$\mu(Z) < \mu(V) \leq \mu(Z).$$

This contradiction implies uniqusness.

□

## 3.2 Result for other cost functions

In this section, we will focus on two special cost function which are

$$c(x,y) = h(x-y), \qquad h \text{ is strictly convex on } \mathbb{R}^n,$$
$$c(x,y) = l(|x-y|), \qquad l \text{ is strictly concave on } \mathbb{R}^+.$$

This section corresponds to the work of Gangbo and McCann [11]. The main idea is to use generalized notions of Convex Analysis so, again if you are not familiar with convex analysis please see the appendix before starting.

### 3.2.1 c-cyclical monotonicity

**Definition 3.2.1** (c-cyclically monotone)**.** A subset $\Gamma \in \mathbb{R}^n \times \mathbb{R}^n$ is called c-cyclical monotone if it satisfies the condition: for all $m \geq 1$ and for all $(x_1, y_1), \ldots, (x_m, y_m) \in \Gamma$ and permutation $\sigma$ on m-letters, we have

$$\sum_{i=1}^{m} c(x_i, y_i) \leq \sum_{i=1}^{m} c(x_{\sigma(i)}, y_i) \tag{2.1}$$

or equivalently

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

where $x_0 = x_m$.

**Theorem 3.2.2** (Ruschendorf [8])**.** *A nonempty subset $\Gamma \subset \mathbb{R}^n \times \mathbb{R}^n$ is c-cyclically monotone if and only if there exists a proper c-concave function $\phi$ on $\mathbb{R}^n$ such that $\Gamma \subset \partial^c \phi$.*

*Proof.* Assume $\Gamma \subset \partial^c \phi$ where $\phi$ is a proper convex function on $\mathbb{R}^n$. We will show that $\partial^c \phi$ is cyclically monotone subset of $\mathbb{R}^n \times \mathbb{R}^n$. Then our claim will follow. Let $(x_1, y_1), \ldots, (x_m, y_m) \in \partial^c \phi$. By the definition of c-superdifferential we have

$$\phi(x_2) - \phi(x_1) \geq c(x_2, y_1) - c(x_1, y_1)$$
$$\phi(x_3) - \phi(x_2) \geq c(x_3, y_2) - c(x_2, y_2)$$

$$\vdots$$

$$\phi(x_1) - \phi(x_m)) \geq c(x_1, y_m) - c(x_m, y_m)$$

Adding up these inequalities, we will observe

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

This is just the definition of cyclically monotone Therefore $\partial\phi$ is cyclically monotone. In particular, $\Gamma$ is cyclically monotone since it is subset of a cyclically monotone set.
Conversely, assume that $\Gamma$ is cyclically monotone and let $(x_0, y_0)$ be any element in $\Gamma$. Now define

$$\phi(x) = \inf\{c(x, y_m) - c(x_m, y_m) + \cdots + c(x_2, y_1) - c(x_1, y_1); m \in \mathbb{N},\ (x_i, y_i) \in \Gamma\}$$

Note first that $\phi$ is a infimum of a nonempty collection of functions of type $c(., y) - b$, hence it is a c-concave function. Moreover, $\phi(x_0) \leq 0$ since $\Gamma$ is cyclically monotone, which means $\phi$ is proper. Next we will show that $\Gamma$ is subset of $\partial^c\phi$. First let $a$ be any number greater than $\phi(x)$. By definition of $\phi$, there exist $m \in \mathbb{N}$ and $(x_1, y_1) \ldots (x_m, y_m)$ such that

$$a \geq c(x, y_m) - c(x_m, y_m) + \cdots + c(x_2, y_1) - c(x_1, y_1).$$

Thus

$$\begin{aligned} a + c(z, y) - c(z, x) &\geq c(z, y) - c(z, x) + c(x, y_m) - c(x_m, y_m) + \cdots + c(x_2, y_1) - c(x_1, y_1) \\ &\geq \phi(z). \end{aligned}$$

Last inequality come from the definition of $\phi$. Since we choose $a$ arbitrarily, we have

$$\phi(z) \leq \phi(x) + c(z, y) - c(z, x),\ \forall z \in \mathbb{R}^n,$$

and $\Gamma$ is a subset of $\partial^c\phi$. □

**Theorem 3.2.3.** *Let $\mu$, $\nu$ be probability measure on $\mathbb{R}^n$, and define continuous cost function $c(x, y) \geq 0$. If $\pi \in \Pi(\mu, \nu)$ is optimal transportation plan then support of $\pi$ is c-cyclically monotone.*

*Proof.* Suppose that $\mathrm{spt}(\pi)$ is not cyclically monotone. Which means we can find an integer $m$ and permutation $\sigma$ such that

$$\sum_{i=1}^{m} c(x_i, y_i) > \sum_{i=1}^{m} c(x_{\sigma(i)}, y_i)$$

for some $(x_1, y_1), \ldots, (x_m, y_m) \in \mathrm{spt}(\pi)$.

Now consider the function

$$f(x_1, \ldots, x_m, y_1, \ldots, y_m) = \sum_{i=1}^{m} c(x_{\sigma(i)}, y_i) - \sum_{i=1}^{m} c(x_i, y_i).$$

This function is continuous and takes negative value at $(x_1, y_1), \ldots, (x_m, y_m) \in \mathrm{spt}(\pi)$. Hence there exist compact neighborhoods $U_j \subset \mathbb{R}^n$ of $x_j$ and $V_j \subset \mathbb{R}^n$ of $y_j$ such that $f(u_1, \ldots, u_m, v_1, \ldots, v_m) < 0$ if $u_j \in U_j$ and $v_j \in V_j$. On the other hand, $\lambda = \inf_j \pi(U_j \times V_j) > 0$ since $(x_j, y_j) \in \mathrm{spt}(\pi)$.

Next, define a measure $\pi_i$ as $\pi_i(A) = \lambda^{-1}\pi(U_j \times V_j \cap A)$ for all $A \subset \mathbb{R}^n \times \mathbb{R}^n$, and let $\mu_i$ and $\nu_i$ be the first and the second marginals of $\pi_i$. Our goal is to build better measure than $\pi$ in the transport problem. Now define

$$\pi' = \pi - \frac{\lambda}{m}\sum_{i=1}^{m} \pi_i + \frac{\lambda}{m}\sum_{i=1}^{m} \pi_i'$$

where $\pi_i' = \mu_{\sigma(i)} \otimes \nu_i$.

The measure $\pi'$ is positive and in $\Pi(\mu, \nu)$. Moreover

$$J(\pi') - J(\pi) = \frac{\lambda}{m}\sum_{i=1}^{m} \int c \, d(\pi_i' - \pi_i) < 0$$

by the condition on the function $f$. This contradicts to the the fact that $\pi$ is optimal. $\square$

### 3.2.2 Strictly convex cost

In this subsection we will consider the cost function which has a form $c(x, y) = h(x - y)$ where the function satisfies the following conditions

**(H1)** $h : \mathbb{R}^d \to [0, \infty)$is strictly convex,

**(H2)** $\lim \frac{h(x)}{|x|} = +\infty$ as $|x| \to +\infty$,

**(H3)** For given $r < \infty$ and angle $\theta \in (0, \pi)$: whenever $p \in \mathbb{R}^n$ is far enough from the origin, there exists a direction $z \in \mathbb{R}^n$ such that, on the truncated cone $K$ with angle $\frac{\theta}{2}$, vertex $p$, and direction z, defined by

$$K(p, z, \theta) = \{z \in \mathbb{R}^n : |x - p||z|cos(\theta/2) \leq (z, x - p) \leq r|z|\},$$

$h$ consume its maximum at $p$.

We need these conditions to ensure the invertibility of $\nabla h$ on whole $\mathbb{R}^n$. Although we will not prove, it will turn out that $(\nabla h)^{-1} = \nabla h^*$ where $*$ is the usual Legendre transform.
We can state main theorem of this subsection.

**Theorem 3.2.4** (Breniers theorem for strictly convex costs)**.** *Let $\mu$ and $\nu$ be probability measures and define the cost function $c(x, y) = h(x - y)$ where $h$ is satisfying (H1)-(H3). If $\mu$ is absolutely continuous with respect to Lebesgue measure, then there exists a unique optimal transportation plan $\pi \in \Pi(\mu, \nu)$ in the form*

$$\pi = (id \times s)_{\#}\, \mu,$$

*where $s(x) = x - (\nabla h)^{-1} \circ \nabla\phi$ such that $s_{\#}\mu = \nu$. and it is uniquely determined with a c-concave function $\phi$ $\mu$-almost surely.*

We will divide the proof of this theorem in order to make it easy to read.

**Theorem 3.2.5.** *Let $\mu$, $\nu$ be probability measure and define the cost function $c(x, y) = h(x - y)$ where $h$ is satisfying (H1)-(H3). Suppose $\pi \in \Pi(\mu, \nu)$ has support in $\partial^c\phi$ for some c-concave function $\phi$. If $\phi$ is differentiable $\mu$-almost surely, then $s_{\#}\mu = \nu$ and $\pi = (id \times s)_{\#}\, \mu$ where $s(x) = x - \nabla h^* \circ \nabla\phi$.*

*Proof.* Note that by A.0.16 part $(i)$ $s(x)$ is a Borel map and the domain of $s$ is $\mathrm{Dom}\nabla\phi$ where $\phi$ is differentiable. So, $\mathrm{Dom}\nabla\phi$ is a Borel set and by the assumption $\mu(\mathrm{Dom}\nabla\phi) = 1$.
First, we will show that $\pi = (id \times s)_{\#}\, \mu$. Let $U, V$ be a Borel sets in $\mathbb{R}^n$ and define

$$S = \{(x, y) \in \partial^c\phi : x \in \mathrm{Dom}\nabla\phi\}$$

Observe that $S = \partial^c\phi \cap \text{Dom}\nabla\phi \times \mathbb{R}^n$. Since $\partial^c\phi$ is closed a set and $\text{Dom}\nabla\phi \times \mathbb{R}^n$ is a Borel set, $S$ is a Borel set and $\pi(S) = 1$ since $\pi(\partial^c\phi) = \pi(\text{Dom}\nabla\phi \times \mathbb{R}^n) = 1$. Moreover, by A.0.16 part $(ii)$ for $(x, y) \in S$ we have $y = s(x)$, hence

$$(U \times V) \cap S = \left(\left(U \cap s^{-1}(V)\right) \times \mathbb{R}^n\right) \cap S$$

This implies

$$\begin{aligned}
\pi(U \times V) &= \pi((U \times V) \cap S) \\
&= \pi\left(\left(\left(U \cap s^{-1}(V)\right) \times \mathbb{R}^n\right) \cap S\right) \\
&= \pi\left(\left(U \cap s^{-1}(V)\right) \times \mathbb{R}^n\right) \\
&= \mu\left(U \cap s^{-1}(V)\right) \\
&= (id \times s)_{\#}\, \mu(U \times V).
\end{aligned}$$

Since, the semi-algebra of products $U \times V$ generates all borel sets in $\mathbb{R}^n \times \mathbb{R}^n$, we have $\pi = (id \times s)_{\#}\, \mu$, from which $s_{\#}\mu = \nu$ follows. $\square$

**Theorem 3.2.6.** *Let $\mu$, $\nu$ be probability measure and define the cost function $c(x, y) = h(x - y)$ where $h$ is satisfying (H1)-(H3). Suppose $\mu$ is absolutely continuous with respect to Lebesgue. If a map $s(x)$, which has form form $s(x) = x - \nabla h^* \circ \nabla\phi(x)$, pushes $\mu$ forward to $\nu$, then it is $\mu-$almost surely unique.*

*Proof.* Assume that in addition to $s(x)$ there exists $t(x)$ such that $t(x) = x - \nabla h^* \circ \nabla\psi(x)$, for some c-concave $\psi$ such that $t_{\#}\mu = s_{\#}\mu = \nu$ but that $t = s$ $\mu-$almost surely does not hold. Then, there exists $x_0 \in \mathbb{R}^n$ at which

**(k1)** $\phi$ and $\psi$ are differentiable but $t(x_0) \neq s(x_0)$, and

**(k2)** $x_0$ is Lebesgue point for $d\mu(x) = f(x)dx$ with $f(x_0) > 0$ where $f$ is the Radon-Nikodym derivative of $\mu$ respect to Lebesgue

Adding some constant, if necessary, to $\phi$ and $\psi$ we can have $\phi(x_o) = \psi(x_0) = 0$ and this change does not effect $t$ and $s$. Define $U = \{x \in \text{int}(\text{Dom}\phi) : \phi(x) > \psi(x)\}$ and $V = \partial^c\psi(U)$.

Note that $V$ is a Borel set since $U$ is open and $\partial^c\psi$ is closed. Moreover, $\phi$ is continuous [2] on $U$ and $\psi$ is upper semi-continuous.

[2] We will not explain the reason because an extensive background is needed, but one can check [11] theorem 3.3 for the answer.

Observe that $t$ is defined $\mu$-almost surely for all $x \in U$ and A.0.16 part $(ii)$ implies $\{t(x)\} = \partial^c\psi(x) \subset V$, or $U \subseteq t^{-1}(V)$. Hence

$$\mu(U) \leq \mu(t^{-1}(V)) \tag{2.2}$$

We have $s^{-1}(V) \subset \partial^c\phi^{-1}(V)$ again by A.0.16 part $(ii)$. Let $K = \{\phi > \psi, \}$ then by using A.0.17

$$\begin{aligned} U = K \cap \mathrm{int}(\mathrm{Dom}\phi) &\supseteq \partial^c\phi^{-1}\left(\partial^c\psi(K)\right) \cap \mathrm{int}(\mathrm{Dom}\phi) \\ &\supseteq \partial^c\phi^{-1}\left(V\right) \cap \mathrm{int}(\mathrm{Dom}\phi) \supseteq s^{-1}(V) \cap \mathrm{int}(\mathrm{Dom}\phi) = s^{-1}(V), \end{aligned}$$

hence

$$\mu(t^{-1}(V)) \leq \mu(U) \tag{2.3}$$

But we still need string inequality. Remember that by A.0.16 part $(ii)$ we have $\partial^c\phi(x_0) = \{s(x_0)\} \neq \{t(x_0)\} = \partial^c\psi(x_0)$. So $\partial^c\phi(x_0)$ and $\partial^c\psi(x_0)$ are disjoint. Which implies $x_0$ lies in a positive distance away from $s^{-1}(V) \subset \partial^c\phi^{-1}(V)$ and we can find a neighborhood $\Omega$ of $x_0$ which is disjoint from $s^{-1}(V)$.
Now translate $\mu$, $\phi$ and $\psi$ so that $x_0 = 0$. Now, consider the cone

$$C = \left\{ x : x.(\nabla\phi(x_0) - \nabla\psi(x_0)) \geq \frac{1}{2}.|x| \right\}.$$

The differentiability of $\phi$ and $\psi$ at $x_0 = 0$ implies

$$\phi(x) - \psi(x) = x.(\nabla\phi(x_0) - \nabla\psi(x_0)) + o(|x|).$$

Thus for very small $x \in C$ we have $x \in U$. Moreover since $x_0$ is Lebesgue point, the average value of $f(x)$ over $C \cap B_r(x_0)$ converges $f(x_0) > 0$ as $r$ goes to 0. Hence for small $r$ this set has positive $\mu$ measure and lies in $U$ and $\Omega$. So $\mu(U \cap \Omega) > 0$. Combining this with 2.2 and 2.3 we have

$$\mu(t^{-1}(V)) < \mu(U) \leq \mu(t^{-1}(V))$$

This contradiction yields that there can not be such a $t(x)$. $\square$

*Proof of 3.2.4.* By 3.1.2 we know that there exists an optimal transportation plan $\pi \in \Pi(\mu, \nu)$. By 3.2.3 support of $\pi$ is c-cyclically monotone. By 3.2.2 there exists a proper c-concave function $\phi$ on $\mathbb{R}^n$ such that $\mathrm{spt}(\pi) \subset \partial^c\phi$. Observe that the function $f(x, y) = x$ on $\mathbb{R}^n \times \mathbb{R}^n$ pushes $\pi$ forward to $\mu$. Hence,

$$\mu(f(\partial^c\phi)) = f_{\#}\pi(f(\partial^c\phi)) = \pi(\partial^c\phi) \geq \pi(\mathrm{spt}(\pi)) = 1$$

So, $f(\partial^c\phi)$ has full $\mu$ measure. On the other hand by A.0.16 part $(iii)$ and $(iv)$ we have $f(\partial^c\phi) \subseteq \text{Dom}\phi$, and the set $\text{Dom}\phi \backslash \text{Dom}\nabla\phi$ has a Lebesgue measure zero. Combining these with absolute continuity assumption, we see that $\phi$ is differentiable $\mu$- almost surely on $\mathbb{R}^n$. So we can apply 3.2.5 which says $s_{\#}\mu = \nu$ and $\pi = (id \times s)_{\#}\, \mu$ where $s(x) = x - \nabla h^* \circ \nabla\phi$. From 3.2.6 Such map $s$ is unique, hence $\pi$ is unique. $\square$

### 3.2.3 Strictly concave cost

For last cost function, the invertibility of $\nabla h$ had important role for the solution. We have this invertibility thanks to the convexity of $h$. There is another case that ensures the invertibility of $\nabla h$, when $c(x, y) = h(x - y) = l(|x - y|)$ where $l$ is strictly concave. Note that here we do not consider the usual concave function, since every non-negative concave function is constant. By strictly concave, we mean the functions $l : \mathbb{R}_+ \to \mathbb{R}_+$ is strictly concave with $l(0) = 0$.
For this cost function, we again have almost the same result as for convex case.

**Theorem 3.2.7** (Breniers theorem,for strictly concave costps). *Let $l : [0, \infty) \to [0, \infty)$ and define $c(x, y) = h(x - y) = l(|x - y|)$. Let $\mu$ and $\nu$ be Borel probability measures on $\mathbb{R}^n$ and define $\mu_0 = [\mu - \nu]_+$ and $\nu_0 = [\nu - \mu]_+$. If $\mu_0$ vabishes on support of $v_0$ and on small sets, then*

*(i) there exists c-concave function $\phi : \mathbb{R}^n \to \mathbb{R}$ on support of $\nu_o$ such that a map $s(x) = id - (\nabla h)^{-1} \circ \nabla\phi$ pushes $\mu_0$ to $\nu_0$ and it is $\mu_0$-alomost surely unique,*

*(ii) there is unique optimal measure $\pi \in \Pi(\mu, \nu)$,*

*(iii) the restriction of $\pi$ to diagonal is given by $\pi_d = id \times id_{\#}(\mu - \mu_o)$,*

*(iv) the off-diagonal part of $\pi = \pi_d + \pi_o$ is given by $\pi_o = id \times s_{\#}\mu_o$.*

Basicly, we have same result as for convex case if $\mu$ and $\nu$ do not share a mass, i.e., they are singular to each other. Otherwise we send shared mass with identity map and other mass with map $s$, and in this case we again have a unique optimal plan but Monge problem does not have a solution in this case. We do not give a proof but one can see [11] for the proof. It is very similar to convex case.

# Bibliography


[1] A. D. ALEKSANDOROV. Almost everywhere existence of the second differential of a convex function and some properties of convex functions. *Leningrad Univ. Ann.*, 37:3–35, 1939.

[2] Leonard D. Berkovitz. *Convexity and Optimization in Rn.* Wiley-Interscience, 2001.

[3] Ky Fan. Minimax theorems. *Proceedings of the National Academy of Sciences of the United States of America*, 39(1):42–47, 1953.

[4] L. Kantorovitch. On the translocation of masses. *Management Science*, 5(1):1–4, 1958.

[5] Robert J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80(2):309–323, 1995.

[6] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *In Histoire de l'Académie Royale des Sciences de Paris*, pages 666–704, 1781.

[7] R. T. Rockafellar. Characterization of the subdifferentials of convex functions. *Pacific Journal Of Mathematics*, 17(3):497–510, 1966.

[8] Ludger Ruschendorf. On c-optimal random variables. *Statistics & Probability Letters*, 27(3):267–270, 1996.

[9] J. Thomas. Robust pricing of options & optimal transportation. Master's thesis, Oxford University, United Kingdom, 2013.

[10] C. Villani. *Topics in optimal transportation*, volume 58 of *volume of Graduate Studies in Mathematics*. AMS, Providence, RI,, 2003.

[11] WilfridGangbo and Robert J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.

# Appendix A

# Preliminaries on convex analysis

### A.0.1 Convexity

**Definition A.0.1.** Let $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, but not identically $+\infty$.
$\phi$ is called proper convex function if:

$$\forall x, y \in \mathbb{R}^n, t \in [0,1], \quad \phi\left(tx + (1-t)y\right) \leq t\phi(x) + (1-t)\phi(y)$$

$\phi$ is called strickly convex function if:

$$\forall x \neq y \in \mathbb{R}^n, t \in (0,1), \quad \phi\left(tx + (1-t)y\right) < t\phi(x) + (1-t)\phi(y)$$

We denote the set of points where $\phi$ is finite by $Dom(\phi)$. Note that the boundary of $Dom(\phi)$ is small set since $\phi$ is convex.

**Definition A.0.2.** The subdifferential of a convex function $\phi$ on $\mathbb{R}^n$ is a subset $\partial\phi \subseteq \mathbb{R}^n \times \mathbb{R}^n$ of pairs (x,y) which satisfies

$$\phi(z) \geq \phi(x) + < y, z - x >, \ \forall z$$

The subgradients of $\phi$ at $x$ will form a closed and convex set $\partial\phi(x) = \{y; (x,y) \in \partial\phi\}$

**Theorem A.0.3** ([2] p.106)**.** *If $D$ is convex subset of $R^n$ and $f : D \to R^m$ is a convex function, then $f$ is differentiable at $x \in int(D)$ if and only if $f$ has unique subgradient at $x$.Moreover, this unique element is $\bigtriangledown f(x)$.*

**Remark A.0.4.** From A.0.7 and A.0.3, $\bigtriangledown\phi(x)$ exists almost everywhere.

**Definition A.0.5.** For proper function $\phi : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$, we define its convex conjuge function (or Legendre transform) by

$$\phi^*(y) = \sup_{x\in\mathbb{R}^n} (x.y - \phi(x))$$

Clearly we have

$$\forall x, y \in \mathbb{R}^n, \;\; x.y \leq \phi(x) + \phi^*(y).$$

Reason for why convex conjuge function have an important role is that we use it fot the characterization of the subdifferential of a convex function. This characterization will be crucial for the proof of main theorems.

**Proposition A.0.6.** *Let $\phi$ be proper lower semi-continous convex function on $\mathbb{R}^n$. Then, for all $x, y \in \mathbb{R}^n$,*

$$x.y = \phi(x) + \phi^*(y) \Longleftrightarrow y \in \phi(x) \Longleftrightarrow x \in \phi^*(y)$$

*Proof.*

$$\begin{aligned} x.y = \phi(x) + \phi^*(y) &\Longleftrightarrow x.y \geq \phi(x) + \phi^*(y) \\ &\Longleftrightarrow \forall z \in \mathbb{R}^n, \;\; x.y \geq \phi(x) + y.z - \phi(z) \\ &\qquad\qquad\qquad\qquad \text{by definition of conjuge function} \\ &\Longleftrightarrow \forall z \in \mathbb{R}^n, \;\; \phi(z) \geq \phi(x) + < y, z - x > \;\; \Longleftrightarrow y \in \phi(x) \\ &\qquad\qquad\qquad\qquad \text{by definition of subdifferential} \end{aligned}$$

□

**Theorem A.0.7** (Alexandrov theorem [1])**.** *If $U$ is open subset of $R^n$ and $f : U \to R^m$ is a convex function, then $f$ has second derivative almost everywhere.*

**Lemma A.0.8** (Aleksandrovs lemma)**.** *Let $\phi$ and $\bar{\phi}$ be two convex functions such that $\phi(x_0) = \bar{\phi}(x_0) = 0$ but $\nabla\phi(x_0) \neq \nabla\bar{\phi}(x_0) = 0$. Let $V = \{\phi > \bar{\phi}\}$ and $Z = \nabla\bar{\phi}^{-1}(\partial\phi(V))$. Then $Z \subseteq V$ and $Z$ lies positive distance away from $x_0$.*

*Proof.* Let $x \in Z$, then there exists $y \in \nabla\phi(V)$ such that $y = \nabla\bar{\phi}(x)$. Since $y \in \partial\phi(V)$, there exists $m \in V$ such that $y \in \partial\phi(m)$. So, for all $z \in \mathbb{R}^n$ we have

$$\phi(z) \geq \bar{\phi}(m) + < y, z - m >$$

$$\bar{\phi}(m) \geq \bar{\phi}(x) + < y, m - x >$$

Since $\phi(m) > \bar{\phi}(m)$, combining these inequalities we have

$$\phi(z) > \bar{\phi}(x) + < y, z - x > . \tag{0.1}$$

Taking $z = x$ shows $\phi(x) > \bar{\phi}(x)$, hence $x \in V$.
Now, assume that $x_0$ lies in closer of $Z$. Then, we can find a sequence $x_n \in Z$ converges to $x_0$. Similarly, there exists a sequence $m_n \in V$ such that $\nabla\bar{\phi}(x_n) \in \partial\phi(m_n)$. For all $z$ $\bar{\phi}(z) \geq \bar{\phi}(x_0) + < \nabla\bar{\phi}(x_0), z - x_0 >$ and $\nabla\bar{\phi}(x_0) = \bar{\phi}(x_0) = 0$ implies that $\bar{\phi} \geq 0$. We have also $\nabla\bar{\phi}(x_n) \longrightarrow 0$ by continuity of $\partial\phi$. Moreover, $\nabla\bar{\phi}(x_0) \neq 0$ implies that $\phi(z_0) < 0$ for some $z_0$ near $x_0$. By using 0.1, we observe

$$\begin{aligned} 0 > \phi(z_0) &> \bar{\phi}(x_v) + < \nabla\bar{\phi}(x_n), z_0 - x_n > \\ &\geq < \nabla\bar{\phi}(x_n), z_0 - x_n > && \text{since } \bar{\phi} \geq 0 \\ &\geq -|\nabla\bar{\phi}(x_n|.|z_0 - x_n| \longrightarrow 0 \end{aligned}$$

since $x_n \longrightarrow x_0$ and $\nabla\bar{\phi}(x_n) \longrightarrow 0$. We obtained a contradiction. Hence, $x_0$ can not lie in closer of $Z$ □

**Theorem A.0.9** (An implicit function theorem)**.** *Let $\phi$ and $\bar{\phi}$ be two convex functions such that $\phi(x_0) = \bar{\phi}(x_0) = 0$ but $\nabla\phi(x_0) \neq \nabla\bar{\phi}(x_0) = 0$. There exists small neighbourhood $U$ of $x_0$ such that $U \cap \{\phi = \bar{\phi}\}$ has $d-1$ Hausdorff dimension.*

This is actually a corollary of An implicit function theorem. One can find this theorem and its proof in [5].

### A.0.2 Generalized convexity and concavity

**Definition A.0.10** (Generalized convexity)**.** Let $X$ and $Y$ be two non-empty set and let c(x,y) be a function on $X \times Y$ with values on $\mathbb{R}^n \cup \{-\infty\}$.
$\phi$ is called *c-convex function* if there exists a proper function $\psi : Y \longrightarrow \mathbb{R} \cup \{\infty\}$ such that

$$\phi(x) = \sup_{y \in Y} \{c(x, y) - \psi(y)\}$$

We denote the set $\{x \in X : \phi(x) \neq \infty\}$ by $Dom(\phi)$.

**Definition A.0.11** (c-subdifferential)**.** The *c-subdifferential* of a function $\phi$ on $\mathbb{R}^n$ is a subset $\partial_c\phi \subseteq \mathbb{R}^n \times \mathbb{R}^n$ of pairs $(x, y)$ which satisfies

$$\phi(z) \geq \phi(x) + c(z, y) - c(x, y), \ \forall z \in X$$

The c-subdifferential of $\phi$ at $x$ is $\partial_c\phi(x) = \{y \in Y : (x, y) \in \partial_c\phi\}$

**Definition A.0.12** (Generalized concavity)**.** Let $X$ and $Y$ be two non-empty set and let $c(x, y)$ be a function on $X \times Y$ with values on $\mathbb{R}^n \cup \{+\infty\}$.
$\phi$ on $X$ is called *c-concave function* if there exist a proper function $\psi : Y \longrightarrow \mathbb{R} \cup \{-\infty\}$ such that

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

We denote the set $\{x \in X : \phi(x) \neq -\infty\}$ by $Dom(\phi)$.

**Definition A.0.13** (c-superdifferential)**.** The *c-superdifferential* of a function $\phi$ on $\mathbb{R}^n$ is a subset $\partial^c\phi \subseteq \mathbb{R}^n \times \mathbb{R}^n$ of pairs $(x, y)$ which satisfies

$$\phi(z) \leq \phi(x) + c(z, y) - c(x, y), \ \forall z \in X$$

The c-subdifferential of $\phi$ at $x$ is $\partial^c\phi(x) = \{y \in Y : (x, y) \in \partial^c\phi\}$

**Definition A.0.14** (c-transforms)**.** Let $X$ and $Y$ be two non-empty sets.
If $c(x, y)$ be a function on $X \times Y$ with values on $\mathbb{R} \cup \{-\infty\}$ and $\phi$ is proper function on $X$ with values in $\mathbb{R} \cup \{+\infty\}$, we define its c-transform by

$$\phi_c(y) = \sup_{x \in X} \{c(x, y) - \phi(x)\}$$

If $c(x, y)$ be a function on $X \times Y$ with values on $\mathbb{R} \cup \{+\infty\}$ and $\phi$ is proper function on $X$ with values in $\mathbb{R} \cup \{-\infty\}$, we define its c-transform by

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

**Proposition A.0.15** (Fenchel-Young inequality)**.**
*(c-convex case)*
*If $c(x, y)$ be function on $X \times Y$ with values on $\mathbb{R}^n \cup \{+\infty\}$ and $\phi$ is proper function on $X$ with values in $\mathbb{R}^n \cup \{-\infty\}$, then we have*

*(i)* $\forall (x, y) \in X \times Y$, $\phi(x) + \phi_c(y) \geq c(x, y)$

*(ii)* $\forall x \in X,\ \phi(x) \geq \phi_{cc}(x)$

*(iii)* $\phi(x) + \phi_c(y) = c(x, y)$ *if and only if* $y \in \partial_c \phi(x)$

*In particular,* $\phi$ *is c-convex if and only if* $\phi = \phi_{cc}$
*(c-concave case)*
*If* $c(x, y)$ *be function on* $X \times Y$ *with values on* $\mathbb{R}^n \cup \{-\infty\}$ *and* $\phi$ *is proper function on* $X$ *with values in* $\mathbb{R}^n \cup \{+\infty\}$*, then we have*

*(i)* $\forall (x, y) \in X \times Y,\ \phi(x) + \phi^c(y) \leq c(x, y)$

*(ii)* $\forall x \in X,\ \phi(x) \leq \phi^{cc}(x)$

*(iii)* $\phi(x) + \phi^c(y) = c(x, y)$ *if and only if* $y \in \partial^c \phi(x)$

*In particular,* $\phi$ *is c-concave if and only if* $\phi = \phi^{cc}$

*Proof.* We will give a proof for just c-convex case, since the proof of other case is almost the same. $(i)$ and $(ii)$ is clear by definition of c-transdorm. For $(iii)$ observe that

$$\begin{aligned}
\phi(x) + \phi^c(y) \leq c(x, y) &\iff \phi(x) + \sup_{z \in X} \{c(z, y) - \phi(z)\} \leq c(x, y) \\
&\iff \forall z \in X,\ \phi(x) + c(z, y) - \phi(z) \leq c(x, y) \\
&\iff \forall z \in X,\ \phi(x) + c(z, y) - c(x, y) \leq \phi(z) \\
&\iff y \in \partial_c \phi(x)
\end{aligned}$$

Combining this result with $(i)$, we have $\phi(x) + \phi^c(y) = c(x, y) \iff y \in \partial^c \phi(x)$ if $\phi$ is c-convex then, there exists a proper $\psi$ such that $\phi(x) = \sup_y \{c(x, y) - \psi(y)\}$. Observe that

$$\begin{aligned}
f_c(y) &= \sup_{z \in X} \{c(z, y) - \sup_{w \in Y} \{c(z, w) - \psi(w)\}\} \\
&\leq \sup_{z \in X} \{c(z, y) - c(z, y) - \psi(y)\} \\
&= \psi(y)
\end{aligned}$$

and this implies

$$\begin{aligned}\phi_{cc}(x) &= \sup_{y\in Y}\{c(x,y)-\phi_c(y)\}\\ &\geq \sup_{y\in Y}\{c(x,y)-\psi(y)\}\\ &= \phi(x)\end{aligned}$$

By part $(ii)$ we $\phi_{cc}(x) \leq \phi(x)$, so we have equality.
On the other hand if $\phi_{cc}(x) = \phi(x)$, $\phi$ is c-convex by definition. □

**Proposition A.0.16.** *Fix the function $c(x,y) = h(x-y)$ where $h$ is stringly convex satisfiying (H1)-(H3) and a c-concave $\phi$ on $\mathbb{R}^n$. Let Dom$\phi$ and Dom$\nabla\phi$ denote the sers on which $\phi$ is finite and $\phi$ is differentiable respectively. Then we have*

*(i)* $s(x) = x - \nabla h^* \circ \nabla\phi$ *defines a Borel map from Dom$\nabla\phi$ to $\mathbb{R}^n$,*

*(ii)* $\partial^c\phi(x) = \{s(x)\}$ *for all $x \in$ Dom$\nabla\phi$,*

*(iii)* $\partial^c\phi(x) = \emptyset$ *unless $x \in$ Dom$\phi$,*

*(iv) the set Dom$\phi\backslash$Dom$\nabla\phi$ has Lebesgue measure zero.*

*Proof.* see [11] thorem 3.4. □

**Lemma A.0.17** (Generalized Aleksandrovs lemma)**.** *Fix the function $c(x,y) = h(x-y)$ where $h$ is stringly convex satisfiying (H1)-(H3). Suppose $\phi$ is c-concave on $\mathbb{R}^n$ and continuous at $x_0 \in \mathbb{R}^n$ with $\phi(x_0) = \bar{\phi}(x_0) = 0$. Define $V = \{\phi > \bar{\phi}\}$ and $Z = \partial^c\phi^{-1}\left(\partial^c\bar{\phi}(V)\right)$. Then $Z \subseteq V$ and $Z$ lies in a positive distance away from $x_0$.*

*Proof.* The proof is very similar to Aleksandrovs lemma A.0.8. □