University of Padua
Department of Mathematics "Tullio Levi-Civita"

Seminar Activity Report

A Primer on Optimal Mass Transportation and the Supremal Monge Problem

Student
Orlando Ferrari

Supervisor
Prof. Elio Marconi

Version May 2, 2026

To my past self, who always took the easy way out instead of doing it the hard way.

To my present self, who is doing his best to live his dream amidst the chaos caused by the loss of an important part of his life.

To my future self, who claimed he has the intention to take the course "Variational problems and optimal transport" in Paris Dauphine University on his M2 next year.

Abstract

This seminar explores foundational concepts and central existence results in optimal transport theory. We transition from the intuitive but restrictive Monge problem to the Kantorovich relaxation, leveraging direct methods of calculus of variations to guarantee the existence of optimal plans. By constructing the dual problem and applying tools from convex analysis—specifically c-transforms and c-cyclical monotonicity—we establish strong duality. This primal-dual framework allows us to demonstrate the existence and uniqueness of optimal maps for strictly convex costs, culminating in Brenier's Theorem in $\mathbb{R}^d$. Finally, we address stability of our solutions and resolve the supremal (L^∞) optimal transport problem, rigorously proving that a secondary variational problem selects a unique optimal solution induced by a transport map.

Good luck

Preface

This seminar discuss the basic notions for the topic of optimal transport and this note is meant as a support for my presentation. The reason I would like to take this topic, even though it looks elementary, is meant to force myself to study something that I have never done, but it combines all of my previous knowledges, except the part of convex analysis, for which I have merely $\epsilon > 0$ amount of experience. The depth of the proof given for each parts of this seminar depends on my belief as well. One may be surprised I put lots of efforts in the basic notions of convex analysis since it is actually my first time doing them formally (I regretted not taking the course of Nonlinear Analysis).

I admit this report has plenty of flaws such as inconsistency of notations as this is written in a really short time[1] and as in the previous paragraph, it is only a support. I also view this as my more well-written notes than my handwritten one[2], that is the proof of the things I have learnt. This report will follow the contents from the first chapter of [San15], but instead of doing the rigorous part (Section 1.6) on the later part, I decided to complete the rigorous part first along the flow of the exposition in this note. The third chapter of this report contains a particular case for the Monge-Kantarovich problem applied to the L^∞ norm which uses modified approach from the approach done in the second chapter which is meant to fill the last 15 minutes of the presentation.

I am grateful for the supervision of Prof. Elio Marconi and the recommendation of Prof. Marco Alessandro Cirant to contact him. Without them, I would not be enjoying optimal transport and my time writing this report. I am really happy to discuss more of this topic and to do some fixes to the erratas found in this report as well by contacting me through my contacts in my homepage refrainfr.github.io.

Padova
29 April 2026

Orlando Ferrari

[1] I recorded my time on making the first draft on Convex Analysis prerequisites on April 5th-6th, and the rest on April 17th-24th, which is really short. For the study itself, I spent my time from the middle of March to understand all things written here.

[2] I also struggled re-reading my handwritten notes since it was too messy. Additionally, I found plenty of mathematical errors in my self-study notes that I did not realized before and thus desperately fixed it, but I can assume there are plenty erratas hidden inside here

Preface

This seminar discuss the basic notions for the topic of optimal transport and this note is meant as a support for my presentation. The reason I would like to take this topic, even though it looks elementary, is meant to force myself to study something that I have never done, but it combines all of my previous knowledges, except the part of convex analysis, for which I have merely $\epsilon > 0$ amount of experience. The depth of the proof given for each parts of this seminar depends on my belief as well. One may be surprised I put lots of efforts in the basic notions of convex analysis since it is actually my first time doing them formally (I regretted not taking the course of Nonlinear Analysis).

I admit this report has plenty of flaws such as inconsistency of notations as this is written in a really short time[1] and as in the previous paragraph, it is only a support. I also view this as my more well-written notes than my handwritten one[2], that is the proof of the things I have learnt. This report will follow the contents from the first chapter of [San15], but instead of doing the rigorous part (Section 1.6) on the later part, I decided to complete the rigorous part first along the flow of the exposition in this note. The third chapter of this report contains a particular case for the Monge-Kantorovich problem applied to the L^∞ norm which uses modified approach from the approach done in the second chapter which is meant to fill the last 15 minutes of the presentation.

I am grateful for the supervision of Prof. Elio Marconi and the recommendation of Prof. Marco Alessandro Cirant to contact him. Without them, I would not be enjoying optimal transport and my time writing this report. I am really happy to discuss more of this topic and to do some fixes to the erratas found in this report as well by contacting me through my contacts in my homepage refrainfr.github.io.

Padova
29 April 2026 Orlando Ferrari

[1] I recorded my time on making the first draft on Convex Analysis prerequisites on April 5th-6th, and the rest on April 17th-24th, which is really short. For the study itself, I spent my time from the middle of March to understand all things written here.

[2] I also struggled re-reading my handwritten notes since it was too messy. Additionally, I found plenty of mathematical errors in my self-study notes that I did not realized before and thus desperately fixed it, but I can assume there are plenty erratas hidden inside here

Table of Contents

1 Introduction

§1.1 Preliminary Notions

§1.1.1 The Monge-Kantorovich Problem

The starting point of optimal transport is the classical problem by Monge in [Mon81], which started from a very practical civil engineering question: how to transport a pile of dirt (*déblais*) to an excavation or embankment site (*remblais*) while minimizing the total physical effort, or transport cost. In terms of mathematical notation we are familiar with, the problem is stated in the following way:

Given two densities of mass $f,\, g \geq 0$ on $\mathbb{R}^d$, with $\displaystyle\int f(x)\,dx = \int g(y)\,dy = 1$, find a map $T : \mathbb{R}^d \to \mathbb{R}^d$, pushing the first one onto the other, i.e. such that

$$\int_A g(y)dy = \int_{T^{-1}(A)} f(x)\,dx \tag{1.1}$$

for any Borel subset $A \subset \mathbb{R}^d$, and T minimizes the quantity

$$M(T) := \int_{\mathbb{R}^d} |T(x) - x| f(x)\,dx \tag{1.2}$$

among all such maps satisfying the condition.

Going back to the physical world, this means that we have a collection of particles, distributed according to the density f on $\mathbb{R}^d$, that have to be moved so that they form a new distribution whose density is prescribed and is g. The movement has to be chosen as to minimize the average displacement. In the description of Monge, the starting density f represented a distribution of sand that had to be moved to a target configuration g (say, making it into a sand castle or any other shape in another place). Hence. it is why Monge gives the term of these two configurations as *déblais* and *remblais*.

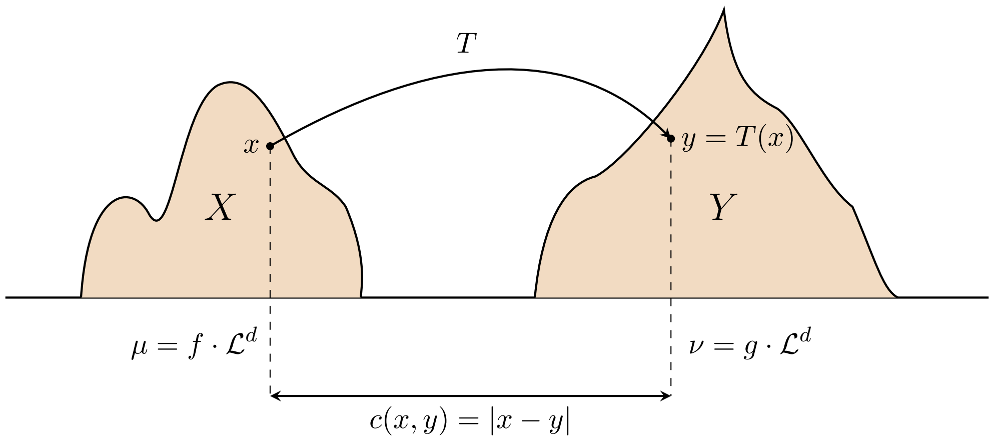

Figure 1.1: Illustration of the pile transportation problem

Indeed, the distribution f and g is associated to probability measures, say μ and ν, respectively, where $\mu,\, \nu \ll \mathcal{L}^d$. However, in the modern framework, we are allowed to consider the distribution of particles for more general measures, and work on spaces broader than $\mathbb{R}^d$.

Before we start on the interesting part, it is convenient for us to familiarize these notions and notice the facts given that will be use for the subsequent parts of this note.

Definition 1.1.1 (Pushforward Measure)
Let (X, σ_X) and (Y, σ_Y) be measurable spaces and $T : X \to Y$ is a Borel measurable function. For any measure $\mu \in \mathcal{M}(X)$, we define the **pushforward** of μ with respect to T, denoted by $T_\#\mu$ as

$$T_\#\mu(A) = \mu(T^{-1}(A)). \qquad \forall A \in \sigma_Y$$

Note that $T_\#\mu \in \mathcal{M}(Y)$.

Remark 1.1.2. Another common notation of pushforward measure is given by $f_*\mu$. However, we will be consistent in using $f_\#\mu$ throughout this notes to follow the spirit of [San15] and [ABS24].

Notice that if μ, $\nu \ll \mathcal{L}^d$ with f and g being the densities of them, respectively, the definition of pushforward is equivalent to (1.1).

After we have our motivation and notion of pushforward, we can read the Monge Problem in its most general version with the most common notation as follows.

Problem 1.1.3 (Monge Problem)
Let X, Y be metric spaces. Given two probability measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ and a **cost function** $c : X \times Y \to [0, +\infty]$, we consider the problem

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

where $\mathcal{P}(X)$ and $\mathcal{P}(Y)$ denotes the space of probability measure in X and Y, respectively. We call such function $T : X \to Y$ that satisfies $T_\#\mu = \nu$ as a **transport map**.

The solution T to (MP) is referred to as an **optimal transport map**.

Indeed, the initial problem Monge proposed of minimization in 1.1 is considered in the case of $c(x, y) = |x - y|$.

The Monge Problem itself is difficult to solve on its own due to its constraint. In particular, the constraint of T itself is not closed under weak convergence as shown in Section 2.4, in particular Example 2.4.1, and so the naïve usage of direct methods of calculus of variations is not feasible. Moreover, there may be no such transport plan as will be shown in Example 2.4.2. A simple example of non-existence is the case $\mu = \delta_{x_0}$ and $\nu = \dfrac{\delta_{y_1} + \delta_{y_2}}{2}$. Even if we have an infimum of (MP), we might not be able to have a solution as what we will have in Example 2.4.4.

Kantorovich in [Kan42] managed to provide an alternative formulation whereas the Monge Problem is "relaxed" (our precise definition of relaxation will be given accordingly in Section 2.5) into a less stricter space as follows.

Problem 1.1.4 (Kantorovich Problem)

Let X, Y be metric spaces. Given two probability measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ and a cost function $c : X \times Y \to [0, +\infty]$, we consider the problem

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

where $\Pi(\mu, \nu) = \{\gamma \in \mathcal{P}(X \times Y) : (\pi_X)_\# \gamma = \mu \text{ and } (\pi_Y)_\# \gamma = \nu\}$ given that π_X and π_Y are the **projections** of elements in $X \times Y$ onto X and Y, respectively. We also call $(\pi_X)_\# \gamma$ and $(\pi_Y)_\# \gamma$ as the **marginals** of γ in X and Y, respectively.

We refer a measure $\gamma \in \Pi(\mu, \nu)$ as a **transport plan** and if γ is a solution (KP) we refer it as an **optimal transport plan**.

On contrary to the Monge problem, $\Pi(\mu, \nu)$ is non-empty since $\mu \otimes \nu \in \Pi(\mu, \nu)$. Before showing explicitly the relation of (MP) and (KP), we first recall the Change of Variable Formula as follows.

Theorem 1.1.5 (Change of Variable Formula)

Let (X, σ_X) and (Y, σ_Y) be measurable spaces and $\mu \in \mathcal{P}(X)$. For any Borel measurable functions $f : X \to Y$ and $\phi : Y \to [0, +\infty]$ we have

$$\int_Y \phi(y)\, df_\# \mu(y) = \int_X (\phi \circ f)(x)\, d\mu(x).$$

It follows that $\psi : Y \to [-\infty, +\infty]$ is $f_\# \mu$-integrable if and only if $\psi \circ f$ is μ-integrable.

Proof. See Proposition 1.7 in Lecture 1 of [ABS24]. The idea is to start with the case of ϕ being a simple function, pass to the case of non-negative Borel measurable function, and finally to the case of any Borel measurable function. ■

With this result, it suggest us to verify the condition of $T_\# \mu = \nu$ for (MP) when we view $T_\# \mu$ and ν a functional over $C_b(Y)$ in the case of Y being a metric space as follows.

Proposition 1.1.6

Let (X, σ_X) be a measurable space, (Y, d_Y) be a metric space, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and $T : X \to Y$ Borel measurable. Then $T_\# \mu = \nu$ holds if and only if

$$\int_Y \phi(y)\, d\nu(y) = \int_X (\phi \circ T)(x)\, d\mu(x)$$

Proof. See Proposition 1.8 in Lecture 1 of [ABS24]. The idea is to start with the fact for any measurable set $B \in \sigma_Y$, where σ_Y is the Borel σ-algebra of (Y, d_Y), we can approximate it by a sequence of monotone non-decreasing functions in $C_b(Y)$. ■

Moreover, with Theorem 1.1.5, we can have the characterization of T being a transport plan if and only if $(\mathrm{id}, T)_{\#}\mu \in \Pi(\mu, \nu)$ since for any $f \in C_b(X \times Y)$ it holds that

$$\int_{X\times Y} f(x, y)\, d(\mathrm{id}, T)_{\#}\mu(x, y) = \int_X f(x, T(x))\, d\mu(x).$$

Additionally, applying $(\mathrm{id}, T)_{\#}\mu$ to (KP) and Theorem 1.1.5 also gives (MP). Hence, we obtained a rough sense of (KP) being a relaxation of (MP). In physical sense, (KP) gives more freedom in the movement of the mass to "spread" to more than one destination than (MP) which restricts the spreading of the mass of a particle to just one destination. Hence, for any $\gamma \in \Pi(\mu, \nu)$, we can view $\gamma(A \times B)$ as the amount of mass moving from A to B.

The benefit of working in (KP) first is the fact that we will able to assert the compactness of $\Pi(\mu, \nu)$ and make a notion of weak convergence for our case and thus we can apply the Direct Methods of Calculus of Variations to find a solution for (KP) in Section 2.1. Hence, our next strategy on solving (MP) is to make the solution of (KP), say γ, is induced by a transport plan T, i.e. $\gamma = (\mathrm{id}, T)_{\#}\mu$ in the particular case of Section 2.3 when $c(x, y) = h(x - y)$ for a strictly convex function h.

§1.1.2 Convex Analysis Tools

Before going to more specific, the reader is invited to recall the following facts from convex analysis before moving forward. This part follows the presentation of [ABS24], but for simplicity, we limit our work on $X = \mathbb{R}^d$ (and thus $X^* = \mathbb{R}^d$) as Ambrosio dealt with the general case of X being a normed space.

Definition 1.1.7 (Convex Function)
A function $f : \mathbb{R}^d \to (-\infty, +\infty]$ is **convex** if and only if for all x, $y \in \mathbb{R}^d$ and $t \in [0, 1]$, we have

$$f((1 - t)x + ty) \leq (1 - t)f(x) + tf(y).$$

Usually, f may be defined on a (convex) set $\Omega \subset X$, but we can extend it nonetheless by defining the value of f to be $+\infty$ outside of Ω, which still preserves convexity of f. If the inequality is instead strict for $t \in (0, 1)$ and $x \neq y$, we have f is **strictly convex**.

Remark 1.1.8. Notice that we can derive a simple stability property of convex functions under the notion of supremum. For any family of convex functions $\{f_\alpha\}$, $f \equiv \sup_\alpha f_\alpha$ is convex since for any α we have

$$f_\alpha((1 - t)x + ty) \leq (1 - t)f_\alpha(x) + tf_\alpha(y) \leq (1 - t)f(x) + tf(y)$$

and thus passing to the supremum of α yields

$$f((1 - t)x + ty) \leq (1 - t)f(x) + tf(y).$$

Remark 1.1.9. By induction, one extend the the convexity (and strict convexity) of f that applies in more general case, that is

$$f\left(\sum_{k=1}^{n} \lambda_k x_k\right) \leq \sum_{k=1}^{n} \lambda_k f(x_k)$$

for any $x_k \in \mathbb{R}^d$ and $\lambda_k \in [0,1]$ such that $\sum_{k=1}^{n} \lambda_k = 1$ since

$$
\begin{aligned}
f\left(\sum_{k=1}^{n} \lambda_k x_k\right) &\leq (1-\lambda_n) f\left(\sum_{k=1}^{n-1} \lambda_k x_k\right) + \lambda_n f(x_n) \\
&\leq (1-\lambda_n) \sum_{k=1}^{n-1} \frac{\lambda_k}{1-\lambda_n} f(x_k) + \lambda_n f(x_n) \\
&= \sum_{k=1}^{n} \lambda_k f(x_k).
\end{aligned}
$$

Definition 1.1.10 (Epigraph)
Let X be a set. The **epigraph** of a function $f : X \to (-\infty, +\infty]$ is the set

$$\operatorname{Epi} f := \{(x, y) \in X \times \mathbb{R} \mid f(x) \leq y\}$$

which is the set of points in $X \times \mathbb{R}$ that lies above the graph of f.

A simple characterization that connects convex function and its epigraph as follows.

Proposition 1.1.11
A function $f : \mathbb{R}^d \to (-\infty, +\infty]$ is convex if and only if $\operatorname{Epi} f$ is convex.

Proof. Let f be a convex function and (x_1, y_1), $(x_2, y_2) \in \operatorname{Epi} f$. For any $t \in [0,1]$ we have

$$f((1-t)x_1 + tx_2) \leq (1-t)f(x_1) + tf(x_2) \leq (1-t)y_1 + ty_2.$$

Hence, $(1-t)(x_1, y_1) + t(x_2, y_2) \in \operatorname{Epi} f$.

Conversely, let $\operatorname{Epi} f$ be convex and to assert that f is convex, it is sufficient to verify the convexity of f holds for points in $\{f < +\infty\}$. Let x_1, $x_2 \in \{f < +\infty\}$. Then $(x_1, f(x_1)), (x_2, f(x_2)) \in \operatorname{Epi} f$ and by convexity of $\operatorname{Epi} f$ we have the desired inequality for any $t \in [0,1]$. ∎

An interesting property of convex function is regarding the fact it is differentiable almost everywhere as it is the corollary after applying Theorem 1.1.15 to the result of the Theorem 1.1.16. We adapt the proof of [ET99] and [AB06] for this fact. In particular, for the case of $\mathbb{R}^d$, we have the following theorem (stated without proof)

Definition 1.1.12 (Relative Interior)
Let $C \subseteq \mathbb{R}^d$. The **relative interior** of C is the set

$$\operatorname{ri}(C) := \{x \in C \mid \exists r > 0 : B_r(x) \cap \operatorname{aff}(C) \subseteq C\}$$

where $\operatorname{aff}(C)$ is the affine hull of C, that is the intersection of all hyperplane that contains C.

Throughout this section, the proof will be given as we don't differentiate between interior and relative interior of C as the main difference is we merely restricting the neighborhood of an element $x \in C$ to be in a hyperplane of co-dimension 1, rather than on the whole space. What makes relative interior better than interior is the fact that it is always non-empty in the case of non-empty closed sets as given by the following theorem, whereas we may have empty interior for a convex set.

Theorem 1.1.13

If C is a non-empty convex subset of $\mathbb{R}^d$, then the relative interior of C is non-empty and is dense in C.

Proof. See Lemma 7.33 of [AB06] ■

Corollary 1.1.14

If $f : \mathbb{R}^d \to (-\infty, +\infty]$ is convex, then $\mathrm{ri}(\{f < +\infty\})$ is non-empty.

Theorem 1.1.15 (Rademacher)

Let U be an open subset of $\mathbb{R}^d$ where U can be taken as $\mathbb{R}^d$ itself. If a function $f : U \to \mathbb{R}^l$ is (locally) Lipschitz, then f is differentiable almost everywhere with respect to the standard Lebesgue measure.

Proof. See Theorem 2 in Chapter 3 of [EG92] ■

Theorem 1.1.16

A convex function $f : \mathbb{R}^d \to (-\infty, +\infty]$ is continuous and locally Lipschitz in the (relative) interior of $\{f < +\infty\}$.

Proof. For simplicity, we shall only prove for the case of interior of $\{f < +\infty\}$.

Step 1. f is locally bounded above in the interior of $\{f < +\infty\}$.

Let x be an interior point of $\{f < +\infty\}$. There exists a neighborhood $B_r(x)$ of x that is contained in $\{f < +\infty\}$. By convexity of $B_r(x)$ and the fact $\mathbb{R}^d$ is of finite dimension, any element $y \in B_r(x)$ can be expressed as a convex combination of $x + e_i$, where $\{e_k\}$ is the standard basis of $\mathbb{R}^d$. Therefore, we can write y as

$$y = \sum_{k=1}^{d} \lambda_k (x + e_k) = x + \sum_{k=1}^{d} \lambda_k e_k$$

for some λ_k such that $\sum_{k=1}^{d} \lambda_k = 1$. Hence, convexity of f implies

$$f(y) \le \sum_{k=1}^{d} \lambda_k f(x + e_k) \le \sum_{k=1}^{d} \lambda_k \max_{1 \le k \le d} |f(x + e_k)| = \max_{1 \le k \le d} |f(x + e_k)| < +\infty.$$

We have proved f is locally bounded above.

Step 2. f is continuous in the interior of $\{f < +\infty\}$.

Let x be an interior point of $\{f < +\infty\}$. As before, we take the neighborhood $B_r(x) = x + B_r(0)$ that is contained in $\{f < +\infty\}$ and by Step 1, f is bounded by $0 \leq a < +\infty$ in $B_r(x)$. Let $\epsilon \in (0,1)$, we have $x + \epsilon B_r(0) \subseteq B_r(x)$. For any $y \in \epsilon B_r(0)$, we have $\frac{y}{\epsilon}, -\frac{y}{\epsilon} \in B_r(0)$. Therefore, $x + \frac{y}{\epsilon}, x - \frac{y}{\epsilon} \in B_r(0)$ and hence

$$f(x+y) \leq (1-\epsilon)f(x) + \epsilon f\left(x + \frac{y}{\epsilon}\right) \implies f(x+y) - f(x) \leq \epsilon(a - f(x))$$
$$f(x) \leq \frac{1}{1+\epsilon} f(x+y) + \frac{\epsilon}{1+\epsilon} f\left(x - \frac{y}{\epsilon}\right) \implies f(x+y) - f(x) \geq -\epsilon(a - f(x))$$

which implies

$$|f(x+y) - f(x)| \leq \epsilon(a - f(x)). \tag{1.3}$$

We have proved f is continuous.

Step 3. f is locally Lipschitz in the interior of $\{f < +\infty\}$.

Let x be an interior point of $\{f < +\infty\}$. By the result of Step 2, there exists $r_0 > 0$ such that $B_{r_0}(x) \subseteq \{f < +\infty\}$ and $-\infty < m \leq f \leq M < \infty$ for some constant M and m in $B_{r_0}(x)$.

Let $r \in (0, r_0)$ and $x_1 \in B_r(x)$. We set

$$g(w) := f(w + x_1) - f(x_1) \quad \forall w \in \mathbb{R}^d$$

where $g(0) = 0$, g is convex and bounded above by $M - m$ in $B_{r-r_0}(0)$. We can apply the argument of Step 2 to arrive in (1.3), that is

$$\forall \epsilon \in [0,1], \forall w \in \epsilon B_{r-r_0}(0) \quad |g(w) - g(0)| = |f(w + x_1) - f(x_1)| \leq \epsilon(M - m). \tag{1.4}$$

Let $x_2 \in B_r(x)$, if $x_2 \in B_{r-r_0}(x_1)$, we can take $w = x_2 - x_1 \in \epsilon B_{r-r_0}(0)$ for $\epsilon = \frac{|x_2 - x_1|}{r - r_0} \in [0,1]$ and we can apply it to inequality (1.4) to have

$$|f(x_2) - f(x_1)| = |g(w)| \leq \frac{M-m}{r-r_0}|x_2 - x_1|. \tag{1.5}$$

If instead $x_2 \notin B_{r-r_0}(x_1)$, we can split the segment $[x_1, x_2]$ into $n+1$ equidistance nodes $v_0 = x_1, \ldots, v_n = x_2$ of distance less than $r - r_0$. Therefore, we can apply inequality (1.5) to get

$$|f(v_k) - f(v_{k-1})| \leq \frac{M-m}{r-r_0}|v_k - v_{k-1}| = \frac{M-m}{r-r_0} \cdot \frac{|x_2 - x_1|}{n} \quad \forall 1 \leq k \leq n. \tag{1.6}$$

and adding all terms to finally get

$$|f(x_2) - f(x_1)| \leq \sum_{k=1}^{n} |f(v_k) - f(v_{k-1})| \leq \frac{M-m}{r-r_0}|x_2 - x_1|.$$

We have proved f is locally Lipschitz. ■

From Theorem 1.1.16, we are ensured of continuity in the interior of $\{f < +\infty\}$. However, in the boundary between $\{f = +\infty\}$ there could be discontinuities. Hence, it is why we typically need at least lower semicontinuity instead.

For lower semicontinuity, it is better for us to define it in the general case of a topological space for practical means, as we shall see the utilization of this property to solve (KP) later on.

Definition 1.1.17 (Lower Semicontinuity)

Let X be a topological space. A function $f : X \to (-\infty, +\infty]$ is **lower semicontinuous (l.s.c)** if and only if for any $x \in X$ and any sequence $\{x_n\} \subseteq X$ that converges to x it holds that

$$f(x) \leq \liminf_{n\to\infty} f(x_n)$$

Remark 1.1.18. We extend the usage of X being a topological space in this case. Note that the notion of lower semicontinuity is closed under suprema since for any family of lower semicontinuous functions $\{f_\alpha\}$ and any sequence $x_n \to x$, we have

$$f_\alpha(x) \leq \liminf_{n\to\infty} f_\alpha(x_n) \leq \liminf_{n\to\infty} \sup_\alpha f_\alpha(x_n)$$

and we can take the supremum over α to conclude $\sup_\alpha(x) \leq \liminf_{n\to\infty} \sup_\alpha(x_n)$.

We also have a characterization of an l.s.c. with respect to its epigraph as follows.

Proposition 1.1.19

Let X be a topological space. A function $f : X \to (-\infty, +\infty]$ is l.s.c if and only if $\mathrm{Epi} f$ is closed.

Proof. Let f be l.s.c. and $(x, y) \in \overline{\mathrm{Epi} f}$, there exists a sequence $\{(x_n, y_n)\} \subseteq \mathrm{Epi} f$ such that $x_n \to x$ and $y_n \to y$ whereas $f(x_n) \leq y_n$ for all $n \in \mathbb{N}$. By lower semicontinuity, by passing $n \to \infty$ yields

$$f(x) \leq \liminf_{n\to\infty} f(x_n) \leq \lim y_n = y.$$

Hence, $(x, y) \in \mathrm{Epi} f$.

Conversely, let $\mathrm{Epi} f$ be closed and let $x \in X$ and $y < f(x)$ so that $(x, y) \notin \mathrm{Epi} f$. Hence, there exists a neighborhood of (x, y), say $U \times (y - \epsilon, y + \epsilon)$ such that all points in such neighborhood is not contained in $\mathrm{Epi} f$. Moreover, the structure of epigraph gives us $(U \times (-\infty, y + \epsilon)) \cap \mathrm{Epi} f = \varnothing$. Therefore, we have $f(u) \geq y + \epsilon$ for any $u \in U$. Let $\{x_n\} \subseteq X$ converges to x, For n sufficiently large, $x_n \in U$ and therefore,

$$y + \epsilon \leq \liminf_{n\to\infty} f(x_n).$$

We can take $\epsilon \to 0^+$ and then $y \to f(x)^-$ to conclude lower semicontinuity of f. ■

Definition 1.1.20

We denote

$$\Gamma(\mathbb{R}^d) := \{f : \mathbb{R}^d \to (-\infty, +\infty] : f \text{ is convex and l.s.c., } f \not\equiv +\infty\}.$$

We shall define the Legendre-Flenchel transformation for any function as follows.

Definition 1.1.21 (Legendre-Flenchel Transformation)

Given $f : \mathbb{R}^d \to (-\infty, +\infty]$, we define the **Legendre-Flenchel transformation** (or **Legendre-Flenchel conjugate**)

$$f^*(x) := \sup_{y\in\mathbb{R}^d} \{\langle y, x\rangle - f(y)\} \qquad \forall x \in \mathbb{R}^d.$$

Note that the supremum may also be taken with respect to $y \in \{f < +\infty\}$ since elements in $\{f = +\infty\}$ do not contribute the supremum unless $f \equiv +\infty$ which gives $f^* \equiv -\infty$.

Definition 1.1.22 (Affine Function)

A function $f : \mathbb{R}^d \to \mathbb{R}$ is **affine** if and only if there exists $a \in \mathbb{R}^d$ and $b \in \mathbb{R}$ such that $f(x) = \langle a, x\rangle + b$ for all $x \in \mathbb{R}^d$. Moreover, if $b = 0$, f is linear.

Notice that our definition of affine function is sufficient since $\mathbb{R}^d$ with the Euclidean norm is a Hilbert space and thus Riesz Representation Theorem applies. Moreover, by the fact that affine functions are convex and lower semicontinuous, f^* is also convex and lower semicontinuous and hence $f^* \in \Gamma(\mathbb{R}^d)$.

We shall characterize function in $\Gamma(\mathbb{R}^d)$ by affine functions.

Theorem 1.1.23 (Fenchel-Moreau)

For all $f \in \Gamma(\mathbb{R}^d)$, one has $f^{**} = (f^*)^* = f$, namely

$$f(x) = \sup_{y\in\mathbb{R}^d} \{\langle y, x\rangle - f^*(y)\} \qquad \forall x \in \mathbb{R}^d.$$

Proof. Notice that,

$$f^{**}(x) = \sup_{y\in\mathbb{R}^d} \{\langle y, x\rangle - f^*(y)\} \qquad \forall x \in \mathbb{R}^d.$$

We shall prove $f \geq f^{**}$ (the easy part) and $f \leq f^{**}$ (the harder part) as follows.

Step 1. $f \geq f^{**}$.

Let $x \in \mathbb{R}^d$. By the definition of f^*, for any $y \in \mathbb{R}^d$ we have

$$f^*(y) \geq \langle x, y\rangle - f(x) \implies f(x) \geq \langle y, x\rangle - f^*(y).$$

Taking the supremum of all y gives $f(x) \geq f^{**}(x)$.

Step 2. $f \le f^{**}$.

This theorem will utilize the Separation Theorem of as the corollary of Hahn-Banach theorem. For more details, see Chapter 5 of [AB06].

Suppose on the contrary that there exists $x_0 \in \mathbb{R}^d$ such that $f(x_0) > f^{**}(x_0)$. There exists $r_0 \in \mathbb{R}$ such that $f(x_0) > r_0 > f^{**}(x_0)$. Notice that by Proposition 1.1.11 and Proposition 1.1.19 makes Epi$f$ a closed convex set and $(x_0, r_0) \notin$ Epi$f$ whereas $\{(x_0, r_0)\}$ itself is a closed convex set. The separation theorem of Hahn-Banach implies the existence of a real hyperplane H that stricly separates Epi$f$ and $\{(x_0, r_0)\}$, that is

$$\langle a, x\rangle + bt > \alpha > \langle a, x_0\rangle + br_0 \tag{1.7}$$

for any $(x, t) \in$ Epi$f$. Notice that if $(x, t) \in$ Epi$f$, then $(x, t + s) \in$ Epi$f$ for any $s > 0$ as well. Therefore, setting $b < 0$ is impossible since passing $t \to \infty$ breaks (1.7). We have $b \ge 0$ and we shall inspect the two possible cases.

- **Case 1.** $b > 0$. For any $(x, f(x))$ which is on Epi$f$, (1.7) implies

  $$f(x) > \frac{\alpha}{b} + \left\langle -\frac{a}{b}, x\right\rangle \implies -\frac{\alpha}{b} > \left\langle -\frac{a}{b}, x\right\rangle - f(x).$$

  Taking the supremum over $x \in \{f < +\infty\}$ gives $-\frac{\alpha}{b} \ge f^*\left(-\frac{a}{b}\right)$. Therefore,

  $$f^{**}(x_0) \ge \left\langle -\frac{a}{b}, x_0\right\rangle - f^*\left(-\frac{a}{b}\right) \ge \left\langle -\frac{a}{b}, x_0\right\rangle + \frac{\alpha}{b}$$

  which implies $\langle a, x_0\rangle + bf^{**}(x_0) \ge \alpha$. However, since $r_0 > f^{**}(x_0)$, we wil arrive at a contradiciton to (1.7).

- **Case 2.** $b = 0$.

  Inequality (1.7) becomes

  $$\langle a, x\rangle > \alpha > \langle a, x_0\rangle \implies -\alpha > \langle -a, x\rangle \tag{1.8}$$

  for any $x \in \{f < +\infty\}$. Since $f \not\equiv +\infty$, $\{f < +\infty\}$ is non-empty and we can invoke Theorem 1.1.28 (proof below) to assert existence of $y_0 \in \mathbb{R}^d$ such that for some $x \in \mathrm{ri}(\{f < +\infty\})$ we have

  $$f(y) \ge f(x) + \langle y_0, y - x\rangle \quad \forall y \in \mathbb{R}^d$$

  and thus

  $$\langle y_0, y\rangle - f(y) \le \langle y_0, x\rangle - f(x) < \infty.$$

  Taking the supremum over $y \in \mathbb{R}^d$ yields $f^*(y_0) < \infty$.

  For $\lambda > 0$, let $y_\lambda = y_0 - \lambda a$, we can utilize (1.8) to have

  $$f^*(y_\lambda) = \sup_{x \in \{f < +\infty\}} \{\langle y_0, x\rangle - f(x) + \lambda \langle a, x\rangle\} \le f^*(y_0) - \lambda\alpha$$

  and thus

  $$f^{**}(x_0) \ge \langle y_\lambda, x_0\rangle - f^*(y_\lambda) \ge \langle y_0, x_0\rangle - f^*(y_0) + \lambda(\alpha - \langle a, x_0\rangle).$$

  From (1.8), we have $\alpha - \langle a, x_0\rangle > 0$. We can pass $\lambda \to \infty$ to get $f^{**}(x_0) = +\infty$. However, this contradicts the fact that $f^{**}(x_0) < r < +\infty$ which is a contradiction.

From both cases, it is impossible to have such point $x_0 \in \mathbb{R}^d$. Therefore, $f \le f^{**}$. ■

Corollary 1.1.24 (Characterization of Convex and Lower Semicontinuous)
The following statements are equivalent for a function $f : \mathbb{R}^d \to (-\infty, +\infty]$:

1. $f \in \Gamma(\mathbb{R}^d)$
2. f is the supremum of affine functions;
3. there exists $g : \mathbb{R}^d \to [-\infty, +\infty)$ such that $f = g^*$.

Proof. 1. $\implies$ 2. is immediate by Theorem 1.1.23.

2. $\implies$ 1. notice that affine functions are also convex and continuous by definition. Since convexity and lower semicontinuity is closed under supremum, the implication holds true.

1. $\implies$ 3. is immediate by Theorem 1.1.23.

3. $\implies$ 1. is immediate since $g^* \in \Gamma(R^d)$. ■

Definition 1.1.25 (Subdifferential)
For all $f \in \Gamma(\mathbb{R}^d)$, the **subdifferential** $\partial f(x)$ is defined as

$$\partial f(x) := \{p \in \mathbb{R}^d : f(y) \geq f(x) + \langle p, y - x\rangle \;\; \forall y \in \mathbb{R}^d\}$$

A simple property that we got from the definition of subdifferential is that for any $p_1 \in \partial f(x_1)$ and $p_2 \in \partial f(x_2)$ we have

$$f(x_1) \geq f(x_2) + \langle p_2, x_1 - x_2\rangle \geq (f(x_1) + \langle p_1, x_2 - x_1\rangle) + \langle p_2, x_1 - x_2\rangle$$

which implies

$$\langle p_1 - p_2, x_1 - x_2\rangle \geq 0.$$

As what we owe in the proof of 1.1.23, we shall the theorem that we will also use in the subsequent chapter for an important result of existence of solution to (MP).

Proposition 1.1.26
For all $f \in \Gamma(\mathbb{R}^d)$ and any $p \in \mathbb{R}^d$, the following characterization holds:

$$p \in \partial f(x) \iff f(x) + f^*(p) = \langle x, p\rangle \iff x \in \partial f^*(p).$$

Proof. Notice that by the definition of f^*, for any x, $p \in \mathbb{R}^d$ we have

$$f^*(p) \geq \langle p, x\rangle - f(x) \implies f(x) + f^*(p) \geq \langle x, p\rangle$$

and thus, to prove equality is sufficient to prove the reverse inequality. Moreover, Theorem 1.1.23 gives $(f^*)^* = f$ which make it sufficient for us to only prove

$$p \in \partial f(x) \iff f(x) + f^*(p) = \langle x, p\rangle\,.$$

Let $p \in \partial f(x)$, for any $y \in \mathbb{R}^d$ we have

$$f(y) \geq f(x) + \langle p, y - x\rangle \implies \langle p, x\rangle - f(x) \geq \langle p, y\rangle - f(y).$$

Taking the supremum of all $y \in \mathbb{R}^d$ gives us

$$\langle p, x\rangle - f(x) \geq f^*(p) \implies \langle x, p\rangle \geq f(x) + f^*(p).$$

Conversely, suppose we have $f(x) + f^*(p) = \langle x, p\rangle$. For any $y \in \mathbb{R}^d$ we have

$$\begin{aligned} f(y) - f(x) = f(y) - \langle x, p\rangle + f^*(p) &\geq f(y) - \langle x, p\rangle + (\langle y, p\rangle - f(y)) \\ &= \langle p, y - x\rangle . \end{aligned}$$

Therefore, $p \in \partial f(x)$. ■

Remark 1.1.27. In general, the characterization $p \in \partial f(x) \iff f(x) + f^*(p) = \langle x, p\rangle$ holds for any function f since we didn't use the characterization of functions in $\Gamma(\mathbb{R}^d)$ unless in the part of using $f^{**} = f$ to prove the whole assertion holds true.

Theorem 1.1.28

Let $f \in \Gamma(\mathbb{R}^d)$. For $x \in \mathrm{ri}(\{f < \infty\})$, $\partial f(x) \neq \varnothing$. Furthermore, at every point where f is differentiable, then $\partial f = \{\nabla f\}$.

Proof. **Part 1.** For $x \in \mathrm{ri}(\{f < \infty\})$, $\partial f(x) \neq \varnothing$.

From Theorem 1.1.16, f is continuous and locally Lipschitz on the $\mathrm{ri}(\{f < \infty\})$. Because it is continuous at x, the point $(x, f(x))$ lies on the boundary of the epigraph, and the epigraph has a non-empty interior and also convex. We can utilize the Supporting Hyperplane Theorem (See [AB06] for more details) to have a non-zero element $(a, b) \in \mathbb{R}^d \times \mathbb{R}$ such that

$$\langle a, y - x\rangle + b(t - f(x)) \geq 0 \qquad \forall (y, t) \in \mathrm{Epi}f \tag{1.9}$$

As what we did in Step 2 of Theorem 1.1.23, $b < 0$ is impossible since we can pass $t \to \infty$ and the inequality will break.

If $b = 0$, then we have

$$\langle a, y - x\rangle \geq 0 \implies \langle a, y\rangle \geq \langle a, x\rangle \qquad \forall y \in \{f < +\infty\}.$$

A point strictly inside the interior of a convex set cannot be separated from the rest of the set by a hyperplane, unless $a = 0$. Hence, it contradicts the fact that (a, b) is a non-zero element.

Therefore, $b > 0$ and thus we can divide (1.9) by b. Rearranging the terms, we have $-\dfrac{a}{b} \in \partial f(x)$.

Part 2. $\partial f(x) = \{\nabla f(x)\}$ when f is differentiable at x. Let $p \in \partial f(x)$. Let $t > 0$ and $w \in \mathbb{R}^d$. By subgradient inequality we have

$$f(x + tw) - f(x) \geq \langle p, (x + tw) - x\rangle = t\,\langle p, w\rangle \implies \frac{f(x + tw) - f(x)}{t} \geq \langle p, w\rangle .$$

Taking the limit $t \to 0^+$ gives us

$$\langle \nabla f(x), w\rangle \geq \langle p, w\rangle \implies \langle \nabla f(x) - p, w\rangle \geq 0.$$

Since w is taken arbitrary, we have $p = \nabla f(x)$ and thus $\partial f(x) = \{\nabla f(x)\}$. ■

Theorem 1.1.29

Let $f \in \Gamma(\mathbb{R}^d)$. If f is strictly convex, then any element $p \in \mathbb{R}^d$ cannot be contained in more than one subdifferential set $\partial f(x)$.

Proof. Suppose $p \in \partial f(x_1) \cap \partial f(x_2)$ for $x_1 \neq x_2$. It follows that

$$\begin{aligned} f(x_2) &\geq f(x_1) + \langle p, x_2 - x_1 \rangle \\ f(x_1) &\geq f(x_2) - \langle p, x_2 - x_1 \rangle \end{aligned}$$

which implies

$$f(x_2) - f(x_1) = \langle p, x_2 - x_1 \rangle .$$

Let $t \in (0,1)$, set $y = (1-t)x_1 + tx_2 \Leftrightarrow t(x_2 - x_1) = y - x_1$. By strict convexity, we have

$$\begin{aligned} f(y) < (1-t)f(x_1) + tf(x_2) &= (1-t)f(x_1) + t(f(x_1) + \langle p, x_2 - x_1 \rangle) \\ &= f(x_1) + \langle p, y - x_1 \rangle \end{aligned}$$

which contradicts the fact $p \in \partial f(x_1)$ where we should have the reverse inequality instead. ∎

Theorem 1.1.30

Let $f \in \Gamma(\mathbb{R}^d)$. Then f is C^1 if and only if f^* is strictly convex.

Proof. See Chapter 7 of [AB06]. ∎

§1.2 c-Concavity and c-Cyclical Monotonicity

After the hassle of preliminaries, we shall move to the technical parts of Optimal Transport. In here, we adapt the contents of Section 1.6 from [San15] as our reference. Before continuing, recall (KP) as the problem of minimizing

$$\int_{X \times Y} c(x,y) d\gamma(x,y)$$

over all $\gamma \in \Pi(\mu,\nu)$ given $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and the cost function $c : X \times Y \to [0,\infty]$. This section will develop the needed tools to solve (MP) and (KP).

In this part, we relaxed the range of the cost function c to be $[-\infty,+\infty]$ or $(-\infty,+\infty)$. We shall generalize some result from Convex Analysis by the fact that $\langle \cdot,\cdot \rangle$ can be realized as the cost $c(\cdot,\cdot) = -\langle \cdot,\cdot \rangle$ in our relaxed definition.

§1.2.1 c-Transformation and c-Concavity

We start by providing a definition of c-transformation and c-concavity as follows

Definition 1.2.1 (c-Transformation)

Given a function $\varphi : X \to [-\infty, +\infty]$, the **$c$-transform** (or **$c$-conjugate**) is $\varphi^c : Y \to [-\infty, +\infty]$ defined by

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

Notice it has similarity with the Legendre-Flenchel transformation in Definition 1.1.21.

We also define the $\bar{c}$-transform of $\phi : Y \to [-\infty, +\infty]$ by

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

Definition 1.2.2 (c-Concavity)

Moreover, we say a function ψ defined on Y is **$\bar{c}$-concave** if there exists ξ such that $\psi = \chi^c$. Analogously, a function φ defined on X is **c-concave** if there exists ζ such that $\varphi = \zeta^{\bar{c}}$.

We denote by $c - conc(X)$ and $\bar{c} - conc(Y)$ the sets of c- and $\bar{c}$-concave functions, respectively.

When $X = Y$ and c is symmetric, the distinction between c and $\bar{c}$ will be of no use and will be ignored.

Notice from the definition of c-transform that $\varphi \leq \psi$ implies $\psi^c \leq \varphi^c$. In most part, we restrict our cost function c to be real valued. We will have a similar characterization as in Theorem 1.1.23 on the case of $\varphi^{c\bar{c}} = c$.

Proposition 1.2.3

Suppose c is real valued. For any $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ we have $\varphi^{c\bar{c}} \geq \varphi$. Moreover, we have $\varphi^{c\bar{c}} = \varphi$ if and only if φ is c-concave (and analogously $\phi^{\bar{c}c} = \phi$ if and only if ϕ is $\bar{c}$-concave). In general, $\varphi^{c\bar{c}}$ is the smallest c-concave function larget than φ (and analogously $\phi^{\bar{c}c}$ is the smallest $\bar{c}$-concave function larget than ϕ).

Proof. For all $x \in X$ and $y \in Y$, by the definition of c-transform we have

$$c(x, y) - \varphi^{\bar{c}}(y) \geq c(x, y) - (c(x, y) - \varphi(x)) = \varphi(x).$$

Taking the infimum over all $y \in Y$ yields $\varphi^{c\bar{c}}(x) \geq \varphi(x)$.

The fact that $\varphi^{c\bar{c}} = \varphi$ implies φ is c-concave is obvious. Conversely, if φ is c-concave, there exists ζ such that $\varphi = \zeta^{\bar{c}}$. We only need to prove $\varphi^{c\bar{c}} \leq \varphi$ immediately by

$$\varphi = \zeta^{\bar{c}} \implies \varphi^c = \zeta^{\bar{c}c} \geq \zeta \implies \varphi^{c\bar{c}} \leq \zeta^{\bar{c}} = c.$$

Now, let $\psi = \chi^{\bar{c}}$ be any c-concave function larger than φ, then

$$\varphi \leq \psi \implies \psi^c \leq \varphi^c.$$

Notice that $\psi^c = \chi^{\bar{c}c} \geq \chi$. Hence,

$$\chi \leq \varphi^c \implies \varphi^{c\bar{c}} \leq \chi^c = \psi.$$

■

Notice that from the definition of subdifferential of a convex function, in the set

$$\text{Graph}(\partial f) := \{(x,p) : p \in \partial f(x)\} = \{(x,p) : f(x) + f^*(p) = \langle x, p\rangle\},$$

we have the monotonicity property

$$\langle x_1 - x_2, p_1 - p_2\rangle \geq 0$$

for any (x_1, p_1), $(x_2, p_2) \in \text{Graph}(\partial f)$. With this property, we are intrigued to prove the converse, that is monotone set is contained in a subdifferential of a convex function. Unfortunately, this is not true. One counterexample is the image of the function $x \mapsto Rx$ where R is the $90°$ rotation in $\mathbb{R}^2$.

Fortunately, a stronger monotonicity helps us on achieving our goal as follows.

Definition 1.2.4 (Cyclical Monotonicity)
A set $A \subset \mathbb{R}^d \times \mathbb{R}^d$ is **cyclically monotone** if for any $k \in \mathbb{N}$, any permutation $\sigma \in S_k$, and any finite elements (x_1, p_1), ..., $(x_k, p_k) \in A$, we have

$$\sum_{i=1}^{k} \langle x_i, p_i\rangle \geq \sum_{i=1}^{k} \left\langle x_i, p_{\sigma(i)}\right\rangle .$$

Notes 1.2.5. To verify cyclical monotonicity, it is enough to verify it in the case of cyclical permutation of elements (and hence it is the origin of the naming "cyclical"). This happens by the fact that any permutation can be decomposed into disjoint components of cyclical permutation.

With this definition, a remarkable result from Rockafellar gives our desired result as follows.

Theorem 1.2.6 (Rockafellar)
Every cyclically monotone set is contained in the graph of subdifferential of a convex function.

Instead of proving them here, we shall prove it in a more general case in the next part since $\langle\cdot,\cdot\rangle$ can be realized as the cost $c(\cdot,\cdot) = -\langle\cdot,\cdot\rangle$.

§1.2.2 c-Cyclical Monotonicity

We shall start from the translation of cyclical monotonicity used by Rockafellar to c-cyclical monotonicity. In this part, we still consider the relaxed range for cost.

Definition 1.2.7

Let X and Y be metric spaces. Let $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$. A set $\Gamma \subset X \times Y$ is **c-cyclically monotone** (**c-CM**) if for all or any $k \in \mathbb{N}$, any permutation $\sigma \in S_k$, and any finite elements (x_1, p_1), ..., $(x_k, p_k) \in \Gamma$, we have

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

Notes 1.2.8. As per the former cyclical monotonicity, it is enough to verify it in the case of cyclical permutation of elements as well to assert c-cyclical monotonicity. Moreover, the c-cyclical inequality is also equivalent to the condition

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

We shall begin our work on generalizing the result of Theorem 1.2.6.

Theorem 1.2.9

Let X and Y be metric spaces. If $\Gamma \neq \varnothing$ is c-CM, $c : X \times Y \to \mathbb{R}$. Then there exists a c-concave function $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ where $\varphi \not\equiv -\infty$ such that

$$\Gamma \subset \{(x, y) \in X \times Y : \varphi(x) + \varphi^c(y) = c(x, y)\}.$$

Proof. The main idea of this proof is use the c-CM property of Γ, that is the fact we have

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

provided that $x_{k+1} = x_1$. We shall construct explicitly φ using this idea. We fix $(x_0, y_0) \in \Gamma$ and we define

$$\varphi(x) = \inf \left\{ c(x, y_k) - c(x_k, y_k) + \sum_{i=0}^{k-1} c(x_{i+1}, y_i) - c(x_i, y_i) \right. \\ \left. : k \in \mathbb{N}, (x_i, y_i) \in \Gamma, \forall i = 1, \ldots, k \right\}$$

Note that $\varphi < +\infty$ since $\Gamma \neq \varnothing$. We shall make the candidate of ψ such that $\psi^{\overline{c}} = \varphi$ by considering

$$\varphi^c(y) = c(x_k, y) - \varphi(x_k) \geq c(x_k, y) - \sum_{i=0}^{k-1} c(x_{i+1}, y_i) - c(x_i, y_i).$$

It is reasonable to define ψ as

$$-\psi(y) = \inf \left\{ -c(x_k, y) + \sum_{i=0}^{k-1} c(x_{i+1}, y_i) - c(x_i, y_i) \right. \\ \left. : k \in \mathbb{N}, (x_i, y_i) \in \Gamma, \forall i = 1, \ldots, k, y_k = y \right\}.$$

Notice that if $y \notin \pi_Y(\Gamma)$, then $\psi(y) = -\infty$ and if $y \in \pi_Y(\Gamma)$, then $\psi(y) > -\infty$. With this definition of ψ, for all $x \in \mathbb{R}^d$ we have

$$\psi^{\bar{c}}(x) = \inf_{y\in\mathbb{R}^d} c(x,y) - \psi(y) = \inf_{y\in\pi_Y(\Gamma)} c(x,y) - \psi(y) = \varphi(x).$$

Indeed, φ is c-concave.

Using the c-CM property of Γ, by defining $x_{k+1} = x_0$, we have

$$\varphi(x_0) \geq c(x_0, y_k) - c(x_k, y_k) + \sum_{i=0}^{k-1} c(x_{i+1}, y_i) - c(x_i, y_i) = \sum_{i=0}^{k} c(x_{i+1}, y_i) - c(x_i, y_i) \geq 0.$$

Hence, $\varphi \not\equiv -\infty$.

We shall prove for any $(x, y) \in \Gamma$ we have $\varphi(x) + \varphi^c(y) = c(x, y)$ to conclude our proof. Note that in general we already have $\varphi(x) + \varphi^c(y) \leq c(x, y)$ and we only need to prove the converse inequality. We use the idea $\varphi^c = \psi^{\bar{c}c} \geq \psi$ which make it sufficient to prove

$$\varphi(x) + \psi(y) \geq c(x, y).$$

Let $(x, y) \in \Gamma$ and $\epsilon > 0$, there exists $\overline{y} \in Y$ (in particular, $\overline{y} \in \pi_Y(\Gamma)$) such that

$$c(x, \overline{y}) - \psi(\overline{y}) < \varphi(x) + \epsilon \implies c(x, \bar{y}) - \psi(\bar{y}) + \psi(y) < \varphi(x) + \psi(y) + \epsilon. \tag{1.10}$$

Since $(x, y) \in \Gamma$, for any $k \in \mathbb{N}$, if we take any finite family $(x_1, y_1), \ldots (x_{k+1}, y_{k+1}) \in \Gamma$ such that $(x_{k+1}, y_{k+1}) = (x, y)$ and $y_k = \bar{y}$ in the definition of $-\psi$, we have

$$\begin{aligned} -\psi(y) &\leq -c(x, y) + c(x, \bar{y}) + \left(-c(x_k, \bar{y}) + \sum_{i=0}^{k-1} c(x_{i+1}, y_i) - c(x_i, y_i) \right) \\ &\leq -c(x, y) + c(x, \bar{y}) - \psi(\bar{y}). \end{aligned}$$

We apply this inequality to (1.10) to obtain

$$c(x, y) \leq (x, \bar{y}) - \psi(\bar{y}) + \psi(y) < \varphi(x) + \psi(y) + \epsilon.$$

By passing $\epsilon \to 0^+$ and we can conclude our proof. ■

Remark 1.2.10. Notice that in our proof, we did not assert the measurability and integrability of φ. In the practice of the subsequent chapter, this issue will not trouble our discussion as φ is measurable and integrable for our case.

2 Solving the Monge-Kantorovich Problem

After finishing the technical preliminaries, we can solve the Monge-Kantorovich Problem comfortably. Along the argument, we will also recall additional notions needed to solve the proof for which the reader is assumed to be already familiar with it.

§2.1 Solving the Kantorovich Problem

§2.1.1 Setup for Direct Methods of Calculus of Variations

To make our setup on solving (KP), we should recall the Direct Method of Calculus of Variations as follows.

Theorem 2.1.1 (Direct Method of Calculus of Variations)
Let X be a topological space, $f : X \to \mathbb{R} \cup \{+\infty\}$ such that

1. f is bounded below
2. f is l.s.c.
3. f is coercive, i.e., $\{f \leq c\}$ is (sequentially) compact for any $c \in \mathbb{R}$.

Then there exists $\bar{x} \in X$ such that $f(\bar{x}) = \min_{x \in X} f(x)$.

Proof. By the fact that f is bounded below from 1., $\inf f > -\infty$. Take a minimizing sequence $\{x_n\}$ such that $f(x_n) < \inf f + \dfrac{1}{n}$ for any $n \in \mathbb{N}$. Hence, $\{f(x_n)\}$ is bounded and by 3., up to a subsequence, x_n converges to an element $\bar{x} \in X$. Without the change of notation, from 2. we obtained

$$\inf f \leq f(\bar{x}) \leq \liminf_{n \to \infty} f(x_n) = \inf f.$$

■

We also have another variation of the direct method which we refer to as the Weierstrass criterion as given below.

Theorem 2.1.2 (Weierstrass Critetion)
Let X be a compact topological space and $f : X \to \mathbb{R} \cup \{+\infty\}$ be l.s.c., then there exists $\bar{x} \in X$ such that $f(\bar{x}) = \min_{x \in X} f(x)$.

Proof. Define $l = \inf_{x \in X} f(x) \in [-\infty, +\infty]$. Note that $l = +\infty$ if and only if $f \equiv +\infty$ which means any element can be a minimizer. Suppose $l < +\infty$, there exists a minimizing sequence $\{x_n\}$ such that $f(x_n) \to l$. By compactness of X, there exists $\bar{x} \in X$ such that $x_n \to \bar{x}$ up to subsequence. By l.s.c, we obtained $l \leq f(\bar{x}) \leq \liminf_{n\to\infty} f(x_n) = l$. This also proves that $l \in \mathbb{R}$ and thus f attained the minimum at $\bar{x}$. ■

For ease of notion, we shall define $K_c : \Pi(\mu, \nu) \to [0, +\infty]$ by

$$K_c(\gamma) := \int_{X\times Y} c(x, y)\, d\gamma(x, y)$$

for any $\gamma \in \Pi(\mu, \nu)$, and so, solving (KP) is the same as minimizing K_c. Our setup for Theorem 2.1.2 is only in providing a compact topology for $\Pi(\mu, \nu)$ where K_c is lower semicontinuous.

Since $\Pi(\mu, \nu) \subset \mathcal{P}(X \times Y) \subset \mathcal{M}(X \times Y)$ where $\mathcal{M}(X \times Y)$ is the set of finite signed measures of $X \times Y$, it is natural to make the weak topology of $\mathcal{M}(X \times Y)$ as our candidate topology. For our next setup, we shall recall the property of $\mathcal{M}(X \times Y)$ with some results of functional analysis as follows. We will omit the proofs, but the reader may consult the needed proofs in [AB06].

Theorem 2.1.3 (Banach-Alaoglu)
If X is a normed space, $\{\xi_n\}$ is bounded in X' (X' is the (topological) dual of X), then up to a subseqeunce ξ_n converges to an element $\xi \in X'$.

In practice, we will work in the case of X being a metric space. In particular, throughout this chapter, we work either on $X = \mathbb{R}^d$ or $X = \Omega \subset \mathbb{R}^d$ where Ω is also compact. Additionally, we also have $\mathbb{R}^d$ is separable and locally compact which make us able to apply the Riesz Representation Theorem as follows.

Definition 2.1.4 (Space of Continuous Functions)
Let X be a metric space (X can also be topological). We denote the set

$$C(X) = \{f : X \to \mathbb{R} : f \text{ is continuous}\}$$

to be the **space of continuous function** in X. Moreover, we denote $C_b(X)$ as the **space of bounded continuous function** in X. If the space X is locally compact, we also define $C_0(X)$ as the **space of continuous function vanishing at infinity**, that is for any $\epsilon > 0$, the set $\{|f| \geq \epsilon\}$ is compact.

Note that if X is compact, we have $C(X) = C_0(X) = C_b(X)$. In general, we have the fact that $C_0(X)$ is a closed subspace of $C_b(X)$ with the topology induced by the norm $\|f\| = \sup_{x\in X} |f(x)|$. Moreover, $C_b(X)$ (and also $C_0(X)$) is a Banach space.

With the space of continuous functions already recalled, we can employ the result of the Riesz Representation Theorem. For more details, the reader is invited to consult Chapter 1 of [EG92].

Theorem 2.1.5 (Riesz Representation Theorem)

Let X be separable and locally compact metric space. Then for any $\xi \in (C_0(X))'$, there exists a unique $\lambda \in \mathcal{M}(X)$ such that

$$\langle \xi, \phi \rangle = \int_X \phi \, d\lambda \qquad \forall \phi \in C_0(X).$$

Moreover, we can induce a norm on $\mathcal{M}(X)$ by

$$\|\lambda\| = |\lambda|(X) < \infty$$

where $|\lambda|$ is a positive finite measure defined as

$$|\lambda|(A) := \sup \left\{ \sum_{i=1}^{\infty} |\lambda(A_i)| : A = \bigcup_{i=1}^{\infty} A_i \text{ with } A_i \cap A_j = \varnothing \text{ for } i \neq j \right\}$$

With the Riesz Representation Theorem and our defined norm on $\mathcal{M}(X)$, we have $\mathcal{M}(X) = (C_0(X))'$. We can also invoke another topology with respect to the duality of $C_b(X)$ as follows.

Definition 2.1.6 (Convergences on $\mathcal{M}(X)$)

Let X be separable and locally compact metric space. We define **the weak* topology** of $\mathcal{M}(X)$ with respect to convergence of $\mathcal{M}(X)$ in the duality of $C_0(X)$ with the convergence denoted by $\mu_n \rightharpoonup^* \mu$. We also define **the weak topology** of $\mathcal{M}(X)$ with respect to convergence of $\mathcal{M}(X)$ in the duality of $C_b(X)$ with the convergence denoted by $\mu_n \rightharpoonup \mu$.

Since $C_0(X) \subset C_b(X)$ in general, the weak convergence is stronger. However, in the case of X being a compact space, both convergences coincide.

We also give the following assertion for closedness of $\mathcal{P}(X)$

Proposition 2.1.7

Let X be separable and locally compact metric space. The space of probability measure $\mathcal{P}(X)$ is a closed subspace of $\mathcal{M}(X)$ with respect to the weak topology.

We shall show an important compactness in the weak topology by the following result by Prokhorov which we will utilize for the search of an optimal plan as follows.

Definition 2.1.8 (Tight Sequence)

Let X be a metric space. A sequence $\{\mu_n\}_n \in \mathcal{M}(X)$ is **tight** if for any $\epsilon > 0$, there exists a compact subset $K \subset X$ such that $|\mu|(X \backslash K) < \epsilon$ for all $n \in \mathbb{N}$.

Theorem 2.1.9 (Prokhorov)

Let X be a Polish space, i.e. X is a separable and complete metric space and $\{\mu_n\}_n \subset \mathcal{P}(X)$ be a tight sequence. Then there exists $\mu \in \mathcal{P}(X)$ such that up to subsequence, $\mu_n \rightharpoonup \mu$. Conversely, if $\mu_n \rightharpoonup \mu$ for some $\mu_n \in \mathcal{P}(X)$, then $\{\mu_n\}$ is necessarily tight.

Proof. See Theorem 2.8 of [ABS24]. ■

§2.1.2 Existence of Solution to the Kantorovich Problem

With the results we recalled in Subsection 2.1.1, we can start solving the Kantorovich Problem starting from the simplest case of strong restriction on cost and ambient space in the subsequent theorem. We will improve this result to less stricter space as we move through this subsection.

Theorem 2.1.10 (Existence of Solution, c Continuous)

Let X and Y be compact metric spaces, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and $c : X \times Y \to \mathbb{R}$ be a continuous function. Then (KP) admits a solution.

Proof. We shall prove the weak topology as given in Definition 2.1.6 on $\Pi(\mu, \nu)$ is our desired topology to invoke Theorem 2.1.2.

Step 1. K_c is continuous. It is immediate by the fact that $c \in C(X \times Y) = C_b(X \times Y)$ and the definition of weak convergence.

Step 2. $\Pi(\mu, \nu)$ is compact.

Let $\{\gamma_n\}$ be a sequence in $\Pi(\mu, \nu) \subset \mathcal{P}(X \times Y)$. Obviously, $\|\gamma_n\| = |\gamma_n|(X \times Y) = \gamma(X \times Y) = 1$ for any $n \in \mathbb{N}$. Hence, it is bounded. By Banach-Alaoglu (c.f. Theorem 2.1.3), there exists $\gamma \in \mathcal{M}(X \times Y)$ such that $\gamma_n \rightharpoonup^* \gamma$. Since $X \times Y$ is compact, $C_b(X \times Y) = C_0(X \times Y)$ and thus $\gamma_n \rightharpoonup \gamma$ which implies $\gamma \in \mathcal{P}(X \times Y)$ by the fact that $\mathcal{P}(X \times Y)$ is closed under weak convergence.

We shall show that $\gamma \in \Pi(\mu, \nu)$ by duality. By weak convergence, we can prove $(\pi_X)_\# \gamma_n \rightharpoonup (\pi_X)_\# \gamma$ in general. Let $\phi \in C_b(X) \subset C_b(X \times Y)$, we have

$$\int_X \phi(x)\, d(\pi_X)_\# \gamma_n(x) = \int_{X \times Y} \phi(x)\, d\gamma_n(x, y) \to \int_{X \times Y} \phi(x)\, d\gamma(x, y) = \int_X \phi(x)\, d(\pi_X)_\# \gamma(x).$$

In particular, we infer that $(\pi_X)_\# \gamma = \mu$. Analogously, $(\pi_Y)_\# \gamma_n \rightharpoonup (\pi_Y)_\# \gamma$ and thus $(\pi_Y)_\# \gamma = \nu$. Hence, $\gamma \in \Pi(\mu, \nu)$. ■

We shall improve the result of Theorem 2.1.10 by relaxing the requirement of c to be l.s.c (with some adjustments). Before that, we will recall a result on l.s.c. that will aid us.

Theorem 2.1.11

Let X be a metric space and $f : X \to \mathbb{R} \cup \{+\infty\}$ be bounded below. Then f is l.s.c. if and only if there exists a sequence $\{f_k\}$ of k-Lipschitz functions such that $f_k \uparrow f$ pointwise.

Proof. Since the notion is l.s.c. is closed in suprema, if we have a sequence $\{f_k\}$ of k-Lipschitz function (and thus continuous) such that $f_k \uparrow f$, we have $f = \sup_{k\in\mathbb{N}} f_k$ is lower semicontinuous.

Conversely, let f be l.s.c., we define

$$f_k(x) := \min\left\{k, \inf_{y\in X}(f(y) + kd(x,y))\right\} \qquad \forall x \in X.$$

Note that if $g_k(x) := \inf_{y\in X}(f(y) + kd(x,y))$ is k-Lipschitz, we have f_k k-Lipschitz as well. It is sufficient to prove $g_k(x)$ is k-Lipschitz.

Let x_1, $x_2 \in X$, we have for any $y \in X$ that

$$g_k(x_1) - (f(y) + kd(x_2,y)) \le (f(y) + kd(x_1,y)) - (f(y) + kd(x_2,y)) \le kd(x_1,x_2).$$

Taking the infimum over all $y \in X$, we have

$$g_k(x_1) - g_k(x_2) \le d(x_1,x_2).$$

Swapping x_1 and x_2 yields $|g_k(x_1) - g_k(x_2)| \le kd(x_1,x_2)$ as desired. Hence, g_k and thus f_k is k-Lipschitz.

Let $x \in X$, we shall prove $f_k(x) \uparrow f(x)$. By definition of f_k, we have $f_k(x) \le f_{k+1}(x)$ and

$$f_k(x) \le f(x) + kd(x,x) = f(x)$$

and thus $\sup_{k\in\mathbb{N}} f_k(x) \le f(x)$. Let $s = \sup_{k\in\mathbb{N}} f_k(x)$. Suppose that $s < f(x)$, for any $k \in \mathbb{N}$, we take $y_k \in X$ such that

$$f(y_k) + kd(x,y_k) < f_k(x) + \frac{1}{k}$$

which implies $f(y_k) \le f(x) + 1$ and also

$$d(x,y_k) < \frac{f_k(x) - f(y_k) + \frac{1}{k}}{k} \implies \limsup_{k\to\infty} d(x,y_k) \le \limsup_{k\to\infty} \frac{f_k(x) - f(y_k) + \frac{1}{k}}{k} = 0.$$

Hence, $y_k \to x$. From l.s.c. of f and the fact $f(y_k) < f_k(x) + \frac{1}{k}$,

$$f(x) \le \liminf_{k\to\infty} f(y_k) \le \lim_{k\to\infty} f_k(x) + \frac{1}{k} = s$$

which contradicts the assumption of $s < f(x)$. Therefore, $f_k \uparrow f$ pointwise. ■

Remark 2.1.12. Actually, the seemingly tedious definition of f_k, instead of g_k that is simpler and does not interfere with the proof above, is for further use as need the improved desired sequence of k-Lipschitz functions being bounded.

With Theorem 2.1.11, we can improve Theorem 2.1.10 as follows.

Theorem 2.1.13 (Existence of Solution, c l.s.c. and bounded below)
Let X and Y be compact metric spaces, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ be l.s.c. and bounded below. Then (KP) admits a solution.

Proof. We still use the topology of weak convergence. Borrowing the Step 2 argument of Theorem 2.1.13, we still have $\Pi(\mu,\nu)$ to be compact. We shall show K_c is lower semicontinuous to allow us invoke Theorem 2.1.2.

From Theorem 2.1.11, there exists a sequence of (bounded) k-Lipschitz functions $\{c_k\}$ such that $c_k \uparrow c$ pointwise. Therefore, the Monotone Convergence Theorem ensures $K_{c_k}(\gamma) \uparrow K_c(\gamma)$ for any $\gamma \in \Pi(\mu,\nu)$. Borrowing Step 1 argument of Theorem 2.1.13, each of K_{c_k} is continuous and thus l.s.c., by closedness of l.s.c. under suprema, we have $K_c = \sup\limits_{k\in\mathbb{N}} K_{c_k}$ to be lower semicontinuous. ■

Now, we shall go back to the actual (KP) for which the cost function c has range in $[0,+\infty]$, we shall tweak the assumption of the ambient space for Theorem 2.1.13 to be milder.

Theorem 2.1.14 (Existence of Solution to (KP), X, Y Polish)
Let X and Y be Polish spaces, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and $c : X \times Y \to [0,+\infty]$ be lower semicontinuous. Then (KP) admits a solution.

Proof. We still use the weak topology on $\Pi(\mu,\nu)$. In this topology, K_c is still l.s.c. borrowing the arguments given in Theorem 2.1.13. We shall prove $\Pi(\mu,\nu)$ is still compact by utilizing Prokhorov Theorem (c.f. Theorem 2.1.9), which means we shall prove $\Pi(\mu,\nu)$ is tight and thus we can invoke Theorem 2.1.2.

Since $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, the converse of Prokhorov Theorem applied to a constant measure implies the existence of compact sets $K_\epsilon^X \subset X$ and $K_\epsilon^Y \subset Y$ such that $\mu(X\backslash K_\epsilon^X) < \frac{1}{2}$ and $\nu(Y\backslash K_\epsilon^Y) < \frac{1}{2}$ for any $\epsilon > 0$. Thus, for any sequence $\{\gamma_n\} \subset \Pi(\mu,\nu)$ (and still any $\epsilon > 0$) we have

$$\begin{aligned}
\gamma_n((X\times Y)\backslash(K_\epsilon^X \times K_\epsilon^Y)) &\le \gamma_n((X\backslash K_\epsilon^X)\times Y) + \gamma_n(X\times(Y\backslash K_\epsilon^Y)) \\
&= (\phi_X)_\#\gamma_n(X\backslash K_\epsilon^X) + (\phi_Y)_\#\gamma_n(Y\backslash K_\epsilon^Y) \\
&= \mu(X\backslash K_\epsilon^X) + \nu(Y\backslash K_\epsilon^Y) \\
&< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\end{aligned}$$

Therefore, $\Pi(\mu,\nu)$ is tight and hence compactness in $\Pi(\mu,\nu)$ is obtained. ■

We conclude this section with a miscellaneous result that we can use in the case the cost functions are not continuous (not really used in this notes), which is the following Lemma. Beforehand, it is encouraged to recall the Lusin Theorems.

Theorem 2.1.15 (Lusin Theorems)

Let X and Y be topological spaces such that Y is second countable (i.e., Y admits a countable family of basis), and μ be a finite regular positive measure (i.e., for any Borel set $A \subset X$, $\mu(A) = \sup\{\mu(K) : K \subset A \text{ compact}\} = \inf\{\mu(B) : A \subset B \text{ open}\}$).

(Weak Lusin) If $f : X \to Y$ is measurable, then for every $\epsilon > 0$, there exists a compact set $K_\epsilon \subset X$ such that $\mu(X \backslash K) < \epsilon$ and $f|_{K_\epsilon}$ is continuous.

(Strong Lusin) If $f : X \to \mathbb{R}$ is measurable, then for every $\epsilon > 0$, there exists a compact set $K_\epsilon \subset X$ and continuous function $g : X \to \mathbb{R}$ such that $\mu(X \backslash K) < \epsilon$ and $f = g$ on K_ϵ.

Lemma 2.1.16

Let γ_n, $\gamma \in \Pi(\mu, \nu)$ and $a : X \to \tilde{X}$, $b : Y \to \tilde{Y}$ be measurable maps valued in two separable metric spaces $\tilde{X}$ and $\tilde{Y}$. Let $c : \tilde{X} \times \tilde{Y} \to [0, +\infty)$ be a continuous function with $c(a, b) \le f(a) + g(b)$ with f, g continuous, and $\int_X (f \circ a)(x)\, d\mu(x)$, $\int_Y (g \circ b)(y)\, d\nu(y) < +\infty$. Then

$$\gamma_n \rightharpoonup \gamma \implies \int_{X\times Y} c(a(x), b(y))\, d\gamma_n(x, y) \to \int_{X\times Y} c(a(x), b(x))\, d\gamma(x, y).$$

Proof. Suppose c is bounded, say $0 \le c \le M$, and fix $\delta > 0$. Since a, b is measurable, Weak Lusin Theorem ensures the existence of compact sets $K_\delta^X \subset X$ and $K_\delta^Y \subset Y$ such that $\mu(X \backslash K_\delta^X)$, $\nu(X \backslash K_\delta^Y) < \frac{\delta}{2}$ and $a|_{K_\delta^X}$, $b|_{K_\delta^Y}$ is continuous. Set $K_\delta = K_\delta^X \times K_\delta^Y \subset X \times Y$ and it can be shown that $\gamma_n((X \times Y)\backslash K_\delta)$, $\gamma((X \times Y)\backslash K_\delta) < \delta$. Thus,

$$\begin{aligned}\int_{X\times Y} c(a, b)\, d\gamma_n &= \int_{K_\delta} c(a, b)\, d\gamma_n + \int_{(X\times Y)\backslash K_\delta} c(a, b)\, d\gamma_n \\ &\le \int_{K_\delta} c(a, b)\, d\gamma_n + M\delta.\end{aligned}$$

Note that $c(a, b)|_{K_\delta} \in C_b(K_\delta)$ and by $\gamma_n \rightharpoonup \gamma$ we have

$$\begin{aligned}&\limsup_{n\to\infty} \int_{X\times Y} c(a, b)\, d\gamma_n \le \int_{X\times Y} c(a, b)\, d\gamma + M\delta \\ \overset{\delta\to 0^+}{\Longrightarrow}\ &\limsup_{n\to\infty} \int_{X\times Y} c(a, b)\, d\gamma_n \le \int_{X\times Y} c(a, b)\, d\gamma\end{aligned}$$

Hence, the function $\gamma \mapsto \int_{X\times Y} c(a, b)\, d\gamma$ is upper semicontinous. We can apply analogous argument to $M - c$ to obtain lower semicontinuity for the same function. Hence, $\gamma \mapsto \int_{X\times Y} c(a, b)\, d\gamma$ is continuous under the weak topology.

If instead c is unbounded, we truncate c into $c_N = c \wedge N := \min\{N, c\}$ where $c_N \uparrow c$. Lower semicontinuity is still preserved by the closedness of the notion under supremum. For upper semicontinuity, we instead work with the function $\tilde{X} \times \tilde{Y} \ni (\tilde{x}, \tilde{y}) \mapsto f(\tilde{x}) + g(\tilde{y}) - c(\tilde{x}, \tilde{y}) \ge 0$. By truncation argument as before, we have $\gamma \mapsto \int_{X\times Y} (f \circ a) + (g \circ$

$b) - c(a, b)\, d\gamma$ is lower semicontinuous and thus we obtained upper semicontinuity for $\gamma \mapsto \int_{X\times Y} c(a, b)\, d\gamma$. ■

§2.2 Primal and Dual Problems

§2.2.1 Rough Formulation of the Dual Problem

In the previous section, Theorem 2.1.14 ensures the existence of a solution for (KP) under the mild assumption of X, Y being Polish space ($X = Y = \mathbb{R}^d$ is included in this case!) and the cost function c being lower semicontinuous. Now, we are in our way to solve (MP) where we are suggested to prove our optimal γ for (KP) is indeed generated by a transport map. Indeed, if such γ is generated by a transport map, $\min(MP) = \min(KP)$, Later in Section 2.5, under some assumptions, it is always the case that $\inf(MP) = \min(KP)$ which formalized the meaning of (KP) being a relaxation of (MP.)

An idea for tweaking our optimal γ for (KP) is to consider another problem by noticing the fact that (KP) is actually a linear optimization under convex constrains, given by linear equalities or inequalities. Hence, we shall find a dual problem for (KP) and exploit the relations between the two.

First, we shall view of another way to see the criterion for $\gamma \in \Pi(\mu, \nu)$ as follows.

If $\gamma \in \Pi(\mu, \nu)$, then for any $\varphi \in C_b(X)$, $\psi \in C_b(Y)$ we have

$$\int_X \varphi(x)\, d\mu(x) = \int_X \varphi(x)\, d(\pi_X)_\#\gamma(x) = \int_{X\times Y} \varphi(x)\, d\gamma(x, y).$$
$$\int_Y \psi(y)\, d\nu(y) = \int_Y \psi(y)\, d(\pi_Y)_\#\gamma(y) = \int_{X\times Y} \psi(y)\, d\gamma(x, y).$$

which implies

$$\int_X \varphi\, d\mu + \int_Y \psi\, d\nu - \int_{X\times Y} (\varphi(x) + \psi(y))\, d\gamma = 0. \qquad \forall \varphi \in C_b(X), \psi \in C_b(Y).$$

On the other hand, if $\gamma \notin \Pi(\mu, \nu)$, without loss of generality, there exists $\varphi_0 \in C_b(X)$ such that

$$\int_X \varphi_0(x)\, d\mu(x) \neq \int_X \varphi_0(x)\, d(\pi_X)_\#\gamma(x) = \int_{X\times Y} \varphi_0(x)\, d\gamma(x, y).$$

By fixing $\psi \in C_b(Y)$, we can take the scalar multiples of φ_0 that is still bounded to conclude

$$\sup_{\substack{\varphi\in C_b(X)\\ \psi\in C_b(Y)}} \int_X \varphi\, d\mu + \int_Y \psi\, d\nu - \int_{X\times Y} (\varphi(x) + \psi(y))\, d\gamma = +\infty.$$

Hence, in $\Pi(\mu, \nu)$, (KP) is equivalent to

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

and considering the interchange of sup and inf (it is not always possible![1]) to have (KP) as

$$\sup_{\substack{\varphi \in C_b(X) \\ \psi \in C_b(Y)}} \int_X \varphi \, d\mu + \int_Y \psi + \inf_{\gamma \in \Pi(\mu,\nu)} \int_{X \times Y} c - (\varphi(x) + \psi(y)) \, d\gamma.$$

If we come back to the maximization problem over (φ, ψ), for fixed $\varphi \in C_b(X)$, $\psi \in C_b(Y)$, we have

$$\inf_{\substack{\gamma \geq 0 \\ \gamma \in \mathcal{M}(X \times Y)}} \int_{X \times Y} c - (\varphi \oplus \psi) \, d\gamma = \begin{cases} 0 & \text{if } \varphi \oplus \psi \leq c \text{ on } X \times Y \\ -\infty & \text{otherwise} \end{cases}$$

where $(\varphi \oplus \psi)(x, y) := \varphi(x) + \psi(y)$. One can prove the cases for $\varphi \oplus \psi > c$ on at least a point (x_0, y_0) by taking the Dirac delta measure $\gamma = \delta_{(x_0,y_0)}$ and its scalar multiple to have integral tends to $-\infty$. This leads to the formulation of the dual problem.

Problem 2.2.1 (Dual Problem)
Let X and Y be metric spaces. Given $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and the cost function $c : X \times Y \to [0, +\infty)$ we consider the problem

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

Remark 2.2.2. One suggestive interpretation of the dual problem from [ABS24] as follows. Suppose that X is a set of bakeries and Y is a set of cafes, then the problem in the (KP) corresponds to minimizing costs of a consortium between bakeries and cafes. Now suppose that there is a transport company which buys at price $\phi(x)$ from each $x \in X$ and sells at price $\psi(y)$ to each $y \in Y$. To be competitive with the direct agreement between bakeries and cafes, it must hold $\psi(y) - \phi(x) \leq c(x, y)$. Then the profit is $\int \psi(y) \, d\nu - \int \phi(x) \, d\mu$. By changing sign to $+\phi(x)$, the problem for the transport company is the same as the formulation of (DP).

In the formulation of (DP), for any $\gamma \in \Pi(\mu, \nu)$, the condition $\varphi \oplus \psi \leq c$ (and also assumptions from Theorem 2.1.14) implies

$$\int_X \phi \, d\mu + \int_Y \psi \, d\nu = \int_{X \times Y} (\varphi \oplus \psi) \, d\gamma \leq \int_{X \times Y} c \, d\gamma \implies \sup(DP) \leq \min(KP).$$

However, (DP) does not admit a straightforward existence result since the class of admissible functions lack compactness. Nonetheless, we have obtained a satisfactory dual problem.

§2.2.2 Solving the Dual Problem

In this part, we shall find the solution of (DP) in a particular case. We shall recall some result of compactness in the space of continuous functions as follows.

[1] Under additional assumptions requiring concavity in one variable, convexity in the other, and some compactness, it is possible to interchange it with the so-called mini-max theorem

Theorem 2.2.3 (Ascoli-Arzelà)

Let X be a compact metric space and $\mathcal{H} \subset C(X)$ with the supnorm in Definition 2.1.4. Then $\mathcal{H}$ is relatively compact with respect to the uniform convergence (induced by the supnorm) in $C(X)$ if and only if $\mathcal{H}$ satisfies:

1. equiboundedness: there exists $M > 0$ such that for all $f \in \mathcal{H}$, $\|f\| < M$.
2. equicontinuity: for any $\epsilon > 0$, $x_0 \in X$, there exists $\delta > 0$ such that for any $x \in X$ and $f \in \mathcal{H}$, $d(x, x_0) < \delta$ implies $|f(x) - f(x_0)| < \epsilon$.

Proof. See Section 10.1 of [RF10]. ■

Proposition 2.2.4

Let X be a metric space, $\omega : [0, +\infty) \to [0, +\infty]$. If $\{f_\alpha\} \subset \mathbb{R}^X$ satisfies the following estimate

$$|f_\alpha(x) - f_\alpha(x')| \leq \omega(d(x, x')) \qquad \forall x, x' \in X$$

Then $f := \inf_\alpha f_\alpha$ and $g := \sup_\alpha f_\alpha$ also satisfies the same estimate.

Proof. Let x, $x' \in X$ and take any α, then

$$f(x) \leq f_\alpha(x) \leq f_\alpha(x') + \omega(d(x, x')) \implies f(x) - f(x') \leq \omega(d(x, x')).$$

Interchanging the role of x and x' yields $|f(x) - f(x')| \leq \omega(d(x, x'))$. For the case of supremum,

$$f_\alpha(x) \leq f_\alpha(x') + \omega(d(x, x')) \leq g(x) + \omega(d(x, x')) \implies g(x) - g(x') \leq \omega(d(x, x')).$$

Interchanging the role of x and x' yields $|g(x) - g(x')| \leq \omega(d(x, x'))$ as well. ■

In particular, if the function ω in Proposition 2.2.4 satisfies $\lim_{t \to 0^+} \omega(t) = 0$ (which means ω is a modulus of continuity and thus $\{f_\alpha\}$ is equicontinuous), then f and g has the same modulus continuity as the family $\{f_\alpha\}$. We should also recall the following result for modulus of continuity and uniform continuity.

Theorem 2.2.5

Let X be a metric space and $f : X \to \mathbb{R}$. Then f is uniformly continuous if and only if there exists an increasing continuous function $\omega : [0, +\infty) \to [0, +\infty)$ with $\omega(0) = 0$ such that

$$|f_\alpha(x) - f_\alpha(x')| \leq \omega(d(x, x')). \qquad \forall x, x' \in X$$

Such ω is referred to as the modulus of continuity of f.

Let us first restrict the space X and Y to be compact metric spaces and c continuous which implies c is uniformly continuous (and bounded) which implies the existence of a modulus of continuity $\omega : [0, +\infty) \to [0, +\infty)$ such that

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

and take any $\chi : X \to \mathbb{R}$ and let $g_x(y) := c(x, y) - \chi(x)$ for all $x, y \in X$. By the definition of c-transformation, $\chi^C = \inf_{x \in X} g_x$ and notice that

$$|g_x(y_1) - g_x(y_2)| = |c(x, y_1) - c(x, y_2)| \leq \omega(d_Y(y_1, y_2)) \qquad \forall x, y_1, y_2 \in X.$$

Thus, by Proposition 2.2.4 χ^c to also be continuous and has the same modulus of continuity as c.

Going back to (DP), note that for any function φ, $\varphi \oplus \varphi^c \leq c$ by definition. If $\varphi \in C_b(X)$, the pair (φ, φ^c) is admissible to (DP). If (φ, ψ) is a solution to (DP), then (φ, φ^c) is also a solution since for all $y \in Y$

$$\psi(y) \leq c(x, y) - \varphi(x) \quad (\forall x \in X) \implies \psi(y) \leq \varphi^c(y).$$

We can also do the same reasoning to have $(\varphi^{c\bar{c}}, \varphi^c)$ to also be a solution to (DP) (notice in our case, the integrals for (DP) of the three pairs are increasing in each c-transformation while the constraint is still maintained). However, Proposition 1.2.3 implies $\varphi^{c\bar{c}c} = \varphi c$ and $(\varphi^{c\bar{c}}, \varphi^c)$ is our best increment. The reason $(\varphi^{c\bar{c}}, \varphi^c)$ itself is better than (φ, ψ) is the fact that the former pair are c-conjugate of each others and both have the same modulus of continuity while the latter is not necessarily that way.

This consideration yields the following exsistence result

Theorem 2.2.6 (Existence of Solution to (DP))

Let X and Y be compact metric spaces and c is continuous (hence uniformly continuous). Then there exists a solution (φ, ψ) to the problem (DP) and it has the form $\varphi \in c - conc(X)$, $\psi \in \bar{c} - conc(Y)$, and $\psi = \varphi^c$. In particular,

$$\max(DP) = \max_{\varphi \in c - conc(X)} \int_X \varphi \, d\mu + \int_Y \varphi^c \, d\nu.$$

Proof. Note that by Theorem 2.1.10, $\sup(DP) \leq \min(KP) < +\infty$. Let ω be the modulus of continuity of c. Take a maximizing sequence $\{(\varphi_n, \psi_n)\}$ and turn it into $\{(\varphi_n^{c\bar{c}}, \varphi_n^c)\}$ which is still a maximizing sequence as well. For ease of notation, we take our maximizing sequence as $\{(\varphi_n, \varphi_n^c)\}$ where $\varphi_n \in c - conc(X)$ for all $n \in \mathbb{N}$.

Note that all φ_n, φ_n^c shares the same modulus of continuity as ω. Hence, we have equicontinuity and we will be aiming to invoke Ascoli-Arzelà Theorem (c.f. Theorem 2.2.3) where we only lack equiboundedness.

Let $C_n = \min \varphi_n$. Note that $\{(\varphi_n - C_n, \varphi_n^c + C_n)\}$ is still a maximizing sequence and still equicontinuous, but now if we go back to our notation of $\{(\varphi_n, \varphi_n^c)\}$ (for the sake of ease of notation) we will have $\min \varphi_n = 0$ for all $n \in \mathbb{N}$. Moreover, since the maximum and minimum of continuous function is always realized in X, we have

$$\max \varphi_n = \max \varphi_n - \min \varphi_n \leq \omega(\operatorname{diam} X) < +\infty \quad \forall n \in \mathbb{N}$$

since X is compact (note that $\operatorname{diam} X$ is the diameter of X). Hence, we have equiboundedness and we can invoke Ascoli-Arzelà Theorem. Up to subsequence, $\varphi_n \to \varphi \in C_b(X)$. On the other hand, $\{\varphi_n^c\}$ is also equibounded since for any $y \in Y$ we have

$$\min c - \omega(\operatorname{diam} X) \leq c(x, y) - \varphi_n(x) \leq \max c - \min \varphi_n(x) = \max c \quad (\forall x \in X)$$

which implies $-\infty < \min c - \omega(\operatorname{diam} X) \le \varphi_n^c \le \max c < +\infty$. Hence, Ascoli-Arzelà also implies $\varphi_n^c \to \psi \in C_b(X)$ (up to subsequence). By the Uniform Convergence Theorem, we have

$$\int_X \varphi_n \, d\mu + \int_Y \varphi_n^c \, d\nu \to \int_X \varphi \, d\mu + \int_Y \psi \, d\nu = \sup(DP).$$

By uniform convergence as well, $\varphi \oplus \psi \le c$. With this result, we have (φ, ψ) solves (DP) and $(\varphi^{c\bar{c}}, \varphi^c)$ be the pair of c-concave and $\bar{c}$-concave functions that solves (DP). ■

Definition 2.2.7 (Kantorovich Potential)

The functions φ that solve (DP) are called **Kantorovich potentials** for the transport from μ to ν.

Remark 2.2.8. Note that the Kantorovich potential φ is not necessarily c-concave, but coincides with the c-concave function $\varphi^{c\bar{c}}$ μ-a.e. However, it is always possible to choose a c-concave Kantorovich potential as what we did in our previous consideration, and we will do this often.

§2.2.3 Equality Between the Primal and Dual Problems

Now, we have found a realization of (DP), but we merely have $\max(DP) \le \min(KP)$. We shall prove that they are equal with the tools we have prepared in Section 1.2. Before starting, we shall recall the notion of support for a measure.

Definition 2.2.9 (Support of a Measure)

Let X be a separable metric space X, the **support of a measure** γ is defined as the smallest closed set on which γ is concentrated, i.e.,

$$\operatorname{supp}(\gamma) = \bigcap A : A \text{ is closed and } \gamma(X \backslash A) = 0.$$

This is well defined since the intersection may be taken countable by the separability assumption. Moreover, there exists also a characterization:

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

The main idea of this part is fairly straightforward. We will prove an optimal plan γ is a c-CM set in the case of c continuous, and thus

$$\operatorname{supp}(\gamma) \subset \{(x, y) : \varphi(x) + \varphi^c(y) = c(x, y)\}$$

for some $\varphi \in c - conc(X)$. Then, we prove such φ is a Kantorovich potential to conclude $\sup(DP) = \min(KP)$. Afterwards, we improve this result by approximation to have $\sup(DP) = \min(KP)$ holds in the assumption for Theorem 2.1.14, that is the cost c being l.s.c. and bounded below.

Theorem 2.2.10

Let X and Y be a separable metric space, if γ is an optimal transport plan for the cost c in (KP) and c is continuous, then $\operatorname{supp}(\gamma)$ is a c-CM set.

Proof. Suppose on the contrary that there exists $k \in \mathbb{N}$, a permutation $\sigma \in S_k$, and finite family $(x_1, y_1), \ldots, (x_k, y_k) \in \operatorname{supp}(\gamma)$ such that

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

We shall prove an existence of another minimizer that is more optimal. Take $\epsilon > 0$ such that

$$\epsilon < \frac{1}{2k}\left(\sum_{i=1}^{k} c(x_i, y_i) - \sum_{i=1}^{k} c(x_i, y_{\sigma(i)})\right).$$

By continuity of c, there exists $r > 0$ such that for any $i = 1, \ldots, k$ we have

$$\forall (x, y) \in B(x_i, r) \times B(y_i, r),\ c(x_i, y_i) - \epsilon < c(x, y)\epsilon \tag{2.1}$$

$$\forall (x, y) \in B(x_i, r) \times B(y_{\sigma(i)}, r),\ c(x, y) < c(x_i, y_{\sigma(i)}) + \epsilon \tag{2.2}$$

Set $V_i = B(x_i, r) \times B(y_i, r)$ and $W_i = B(x_i, r) \times B(y_{\sigma(i)}, r)$. Since $(x_i, y_i) \in \operatorname{supp}(\gamma)$, $\gamma(V_i) > 0$ for all $i = 1, \ldots, k$. We construct $\gamma_i = \dfrac{\gamma \llcorner V_i}{\gamma(V_i)} \in \mathcal{P}(X \times Y)$ and set $\mu_i = (\pi_X)_{\#}\gamma_i$ and $\nu_i = (\pi_Y)_{\#}\gamma_i$ where μ_i and ν_i are concentrated in $B(x_i, r)$ and $B(y_i, r)$, respectively. Now, we take $\epsilon_0 > 0$ such that $\epsilon_0 < \dfrac{1}{k} \min\limits_{1 \le i \le k} \gamma(V_i)$.

Continuing forward, for any $i = 1, \ldots, k$, we take any $\tilde{\gamma}_i \in \Pi(\mu_i, \nu_{\sigma(i)})$ at will (for instance, $\tilde{\gamma}_i = \mu_i \otimes \nu_{\sigma(i)}$). We have γ_i is concentrated in V_i and $\tilde{\gamma}_i$ is concentrated in W_i. We construct the a candidate for the new optimal plan $\tilde{\gamma}$ as

$$\tilde{\gamma} := \gamma - \epsilon_0 \sum_{i=1}^{k} \gamma_i + \epsilon_0 \sum_{i=1}^{k} \tilde{\gamma}_i$$

which is a signed measure. $\tilde{\gamma}$ is positive since for any borel set $A \subset X \times Y$ we have from our choice of ϵ_0 that

$$\tilde{\gamma} \ge \gamma(A) - \epsilon_0 \sum_{i=1}^{k} \gamma_i(A) > \gamma(A) - \sum_{i=1}^{k} \frac{\gamma(V_i)}{k} = 0.$$

Moreover, $\tilde{\gamma} \in \Pi(\mu, \nu)$ by definition:

$$(\phi_X)_{\#}\tilde{\gamma} = \mu - \epsilon \sum_{i=1}^{k} (\pi_X)_{\#}\gamma_i + \epsilon \sum_{i=1}^{k} (\pi_X)_{\#}\tilde{\gamma}(i) = \mu - \epsilon \sum_{i=1}^{k} \mu_i + \epsilon \sum_{i=1}^{k} \mu_i = \mu,$$

$$(\phi_Y)_{\#}\tilde{\gamma} = \nu - \epsilon \sum_{i=1}^{k} (\pi_Y)_{\#}\gamma_i + \epsilon \sum_{i=1}^{k} (\pi_Y)_{\#}\tilde{\gamma}(i) = \nu - \epsilon \sum_{i=1}^{k} \nu_i + \epsilon \sum_{i=1}^{k} \nu_{\sigma(i)} = \nu.$$

Finally, we shall show $K(\gamma) - K(\tilde{\gamma}) > 0$ to arrive at the contradiction where γ is supposedly to be optimal. From (2.1), (2.2), and choice of ϵ, we have

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

■

Theorem 2.2.11

Let X and Y be Polish spaces. Suppose $c : X \times Y$ is uniformly continuous and bounded, then (DP) admits a solution (φ, φ^c) and we have $\max(DP) = \min(KP)$.

Proof. From Theorem 2.1.14, there exists an optimal plan γ for (KP). From Theorem 2.2.10, $\Gamma := \mathrm{supp}(\gamma)$ is c-CM. Since c is real valued, by Theorem 1.2.9, there exists a c-concave function $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ such that

$$\Gamma \subset \{(x, y) \in X \times Y : \varphi(x) + \varphi^c(y) = c(x, y)\}.$$

Let ω be a modulus of continuity of c and $-M \le c \le M$. Recall the definition of our chosen φ as

$$\varphi(x) = \inf \Bigg\{ c(x, y_k) - c(x_k, y_k) + \sum_{i=0}^{k-1} c(x_{i+1}, y_i) - c(x_i, y_i) \\ : k \in \mathbb{N}, (x_i, y_i) \in \Gamma, \forall i = 1, \ldots, k\}$$

for a fixed $(x_0, y_0) \in \Gamma$ where $\varphi(x_0) \ge 0$ and thus $\varphi^c \le c(x_0, \cdot) - \varphi(x_0) \le M < +\infty$. However, this implies $\varphi \ge -2M > -\infty$. On the other hand, by the definition of φ, we have $\varphi \le c(\cdot, y_0) - c(x_0, y_0) \le 2M < +\infty$ and it implies $\varphi^c \ge -3M > -\infty$. Hence, φ and φ^c is real valued and furthermore bounded. Since φ is c-concave, φ and φ^c is (uniformly) continuous since they have the same modulus of continuity as c. The pair (φ, φ^c) is admissible to (DP) and we have

$$\sup(DP) \le \min(KP) = \int_{X \times Y} c\, d\gamma = \int_{\mathrm{supp}(\gamma)} \varphi \oplus \varphi^c\, d\gamma = \int_X \varphi\, d\mu + \int_Y \varphi^c\, d\nu \le \sup(DP).$$

Therefore, (φ, φ^c) is a solution to (DP) and we have $\max(DP) = \min(KP)$. ■

Notice that so far we have not settled in the case of Theorem 2.1.14 where the minimizer of (KP) can still be realized for the case of the cost function being lower semicontinuous and bounded from below. We shall use the result of Theorem 2.1.11 where any l.s.c. and bounded from below cost c is a limit of a sequence $\{c_k\}$ of bounded k-Lipschitz functions. Beforehand, we need the following lemma.

Lemma 2.2.12

Let X and Y be Polish spaces. Suppose that c_k and c are l.s.c. and bounded from below for any $k \in \mathbb{N}$ and $c_k \uparrow c$. Then

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

Proof. For any $k \in \mathbb{N}$, from Theorem 2.1.14, there exists $\gamma_k \in \Pi(\mu, \nu)$ as the minimizer. Borrowing a fact from the argument of Theorem 2.1.14, $\Pi(\mu, \nu)$ is tight and thus compact. Then up to subsequence, $\gamma_k \rightharpoonup \bar{\gamma} \in \Pi(\mu, \nu)$. Notice by the fact c_k is increasing to c, we have $\min \left\{ \int_{X \times Y} c_k\, d\gamma : \gamma \in \Pi(\mu, \nu) \right\}$ is increasing as well and the supremum is bounded above by $\min\left\{\int_{X\times Y} c\,d\gamma : \gamma\in\Pi(\mu,\nu)\right\}$. Fix $j\in\mathbb{N}$. For any $k\geq j$, we have

$$\begin{aligned}\lim_{k\to\infty}\min\left\{\int_{X\times Y} c_k\,d\gamma : \gamma\in\Pi(\mu,\nu)\right\} = \lim_{k\to\infty}\int_{X\times Y} c_k\,d\gamma_k &\geq \liminf_{k\to\infty}\int_{X\times Y} c_j\,d\gamma_k \\ &\geq \int_{X\times Y} c_j\,d\bar{\gamma}\end{aligned}$$

where the last inequality holds by the lower semicontinuity of the integral cost c_j by arguments we borrowed from Theorem 2.1.13. By Monotone Convergence Theorem, we can pass $j\to\infty$ to get

$$\lim_{k\to\infty}\min\left\{\int_{X\times Y} c_k\,d\gamma : \gamma\in\Pi(\mu,\nu)\right\} = \int_{X\times Y} c\,d\bar{\gamma} \geq \min\left\{\int_{X\times Y} c\,d\gamma : \gamma\in\Pi(\mu,\nu)\right\}.$$

Hence, we obtained our claim, and as the byproduct, $\bar{\gamma}$ is optimal for the limit cost c. ■

We can establish the desired validity and conclude the duality formula

Theorem 2.2.13

Let X and Y be Polish spaces and $c : X\times Y\to\mathbb{R}\cup\{+\infty\}$ is l.s.c. and bounded below. Then the duality formula $\min(KP) = \sup(DP)$ holds.

Proof. From Theorem 2.1.11, we can take a sequence $\{c_k\}$ of bounded k-Lipschitz functions that converges increasingly to c. Employing Theorem 2.2.13 on each cost c_k, we have

$$\begin{aligned}\min\left\{\int_{X\times Y} c_k\,d\gamma : \gamma\in\Pi(\mu,\nu)\right\} &= \max\left\{\int_X \varphi\,d\mu + \int_Y \psi\,d\mu : \varphi\oplus\psi\leq c_k\right\} \\ &\leq \sup\left\{\int_X \varphi\,d\mu + \int_Y \psi\,d\mu : \varphi\oplus\psi\leq c\right\}\end{aligned}$$

By Lemma 2.2.3, we can pass $k\to\infty$ to conclude

$$\min(KP) = \lim_{k\to\infty}\min\left\{\int_{X\times Y} c_k\,d\gamma : \gamma\in\Pi(\mu,\nu)\right\} \leq \sup(DP).$$

Combining with the general result of $\sup(DP)\leq\min(KP)$, the duality formula $\min(KP) = \sup(DP)$ is concluded. ■

As our last result for this section, we improve the result of Theorem 2.2.10 for the case of c l.s.c. and bounded below.

Theorem 2.2.14

Let X and Y be Polish spaces, $c : X\times Y\to\mathbb{R}\cup\{+\infty\}$ is l.s.c. and bounded below, and γ is an optimal transport plan for c. Then γ is concentrated on a c-CM set Γ (which will not be closed in general).

Proof. From Theorem 2.2.13, the duality formula holds. We can take a sequence of maximizing pairs $\{(\varphi_n, \psi_n)\}$ for (DP). We have

$$\int_{X\times Y} \varphi_n \oplus \psi_n \, d\gamma \to \int_{X\times Y} c \, d\gamma.$$

We also have $\varphi_n \oplus \psi_n \leq c$ which implies the function $f_n(x, y) := c(x, y) - \varphi_n(x) - \psi_n(x, y)$ satisfies $f_n \geq 0$ and from our convergence, we have $f_n \to 0$ in $L^1(X \times Y, \gamma)$. Therefore, up to subsequence, $f_n \to 0$ for γ-a.e. Let Γ a set with $\gamma(\Gamma) = 1$ where the convergence $f_n \to 0$ holds and we also have $\text{supp}(\gamma) \subset \Gamma$. For any $k \in \mathbb{N}$, permutation $\sigma \in S_k$, and finite family (x_1, y_1), ..., $(x_k, y_k) \in \Gamma$ we have

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

which proves Γ is c-CM. ■

Remark 2.2.15. The duality formula that we proved for the case c l.s.c. bounded below in Theorem 2.2.13 differs from the continuous cost that we proved in Theorem 2.2.11 in that there is no guarantee of existence for the dual problem (DP). Actually, we can tweak the condition of admissible functions for (DP) being in L^1 instead of C_b and we have such solution (φ, ψ) by first applying Theorem 2.2.14 and then build a potential φ through Theorem 1.2.9. This works under the assumption that c is real valued and does not depend on its continuity.

The variation of (DP) by tweaking the admissible functions being L^1 will be discussed more in Section 2.3, in particular Subsection 2.3.2.

§2.3 Solving the Monge Problem in the Case $c(x, y) = h(x - y)$ for h strictly convex

In this section, we will be able to solve (MP) in some case of c and the restriction of our domain being $X = Y = \mathbb{R}^d$ or its compact domain. A natural reason to consider h to be strict since in practice we still consider the case when $c(x, y) = |x - y|^p$ for $p \in (1, +\infty)$ where h in that case is strictly convex. For the case $p = +\infty$ we shall consider the approach in Chapter 3 where we our approach cannot be applied. However, the case $p = 1$, in particular, the initial (MP) formulation, unfortunately will not be discussed in this version of note. The reader is invited to refer to Chapter 3 of [San15] for the solution of this case.

§2.3.1 General Case of $c(x, y) = h(x - y)$

We start with the consideration from Theorem 2.2.11 with an additional assumption that our cost c is C^1 (here we have not assumed $c(x, y) = h(x - y)$). Considering our optimal plan γ and Kantorovich potential φ, for a fixed $(x_0, y_0) \in \text{supp}(\gamma)$ we have the function

$$\mathbb{R}^d \ni x \mapsto c(x, y_0) - \varphi(x)$$

achieve its minimum at $x = x_0$ and thus $\nabla_x c(x_0, y_0) = \nabla\varphi(x_0)$ provided that φ is differentiable at x_0. This consideration proved the following proposition

Proposition 2.3.1

If c is C^1, φ is a Kantorovich potential for the cost c, and γ is an optimal transport plan in the transport from μ to ν, and $(x_0, y_0) \in \text{supp}(\gamma)$, then $\nabla\varphi(x_0) = \nabla_x c(x_0, y_0)$ provided that φ is differentiable at x_0.

In particular, the gradients of two Kantorovich potentials coincide on every point $x_0 \in \text{supp}(\mu)$ (for which $(x_0, y_0) \in \text{supp}(\gamma)$ for some $y_0 \in \mathbb{R}^d$) where both potentials are differentiable.

Proof. See above consideration. ■

Moreover, the condition $\nabla\varphi = \nabla_x c$ is particularly useful when x satisfies the "twist" condition in the following definition.

Definition 2.3.2 (Twist Condition)

Let $\Omega \subset \mathbb{R}^d$, we say that $c : \Omega \times \Omega \to \mathbb{R}$ satisfies the **twist condition** whenever $c(x, y)$ is differentiable with respect to x at every point and the map $\Omega \ni y \mapsto \nabla_x c(x_0, y)$ is injective for every x_0.

Note that for more regular domain Ω and smoother c, this condition corresponds to $\det\left(\frac{\partial^2 c}{\partial y_i \partial x_j}\right) \neq 0$.

The goal on having the twist condition is to deduce from $(x_0, y_0) \in \text{supp}(\gamma)$, we have y_0 is uniquely defined from x_0 which shows that γ is concentrated on a graph of a mapping and this map will be the optimal transport. Since the map is constructed using φ and c only, and not γ, this also proves that our optimal γ is unique.

On our case of h is strictly convex and thus it is differentiable almost everywhere with respect to the Lebesgue measure from Theorem 1.1.15, it suggests that in differentiable points of φ at x_0, the equality from Theorem 2.3.1 gives us

$$\{\nabla\varphi(x_0)\} = \partial h(x_0 - y_0) = \{\nabla h(x_0 - y_0)\} \implies y_0 = x_0 - (\nabla h)^{-1}(\nabla\varphi(x_0))$$

where the relation $(\nabla h)^{-1}$ is justified from Theorem 1.1.29 even if h is not differentiable at $x_0 - y_0$ which gives us the suggestion $x \mapsto x - (\nabla h)^{-1}(\nabla\varphi(x))$ as our transport map.

We shall formulate this result formally in the following Theorem

Theorem 2.3.3 (Existence of solution for (MP))

Given μ and $\nu \in \mathcal{P}(\Omega)$ for a compact domain $\Omega \subset \mathbb{R}^d$. There exists an optimal transport plan γ for the cost $c(x, y) = h(x - y)$ with h is strictly convex and real valued. It is unique and of the form $(\text{id}, T)_{\#}\mu$, provided μ is absolutely continuous to the Lebesgue measure and $\partial\Omega$ is μ-negligible. Moreover, there exists a Kantorovich potential μ, and T and the potentials μ are linked by

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

Proof. For ease of readability, we split the proof into three steps.

Step 1. Existence and Regularity of the Kantorovich Potential.
From our assumption, $h : \mathbb{R}^d \to \mathbb{R}$ is strictly convex, so its effective domain is the entire space: $\text{ri}(\{h < +\infty\}) = \mathbb{R}^d$. By Theorem 1.1.16, h is locally Lipschitz. Because Ω is a compact domain, the cost function $c(x, y) = h(x - y)$ is Lipschitz continuous (and thus uniformly continuous) and bounded on $\Omega \times \Omega$.

By Theorem 2.1.10 and Theorem 2.2.11, there exists an optimal transport plan $\gamma \in \Pi(\mu, \nu)$ and a corresponding Kantorovich potential $\varphi \in c\text{-conc}(\Omega)$. By Proposition 2.2.4, because φ is c-concave, it shares the exact same modulus of continuity as c, meaning φ is also Lipschitz continuous.

Since φ is Lipschitz on an open set containing Ω, Rademacher Theorem (c.f. Theorem 1.1.15) guarantees that φ is differentiable almost everywhere with respect to the Lebesgue measure. Because we assumed μ is absolutely continuous with respect to the Lebesgue measure and $\partial\Omega$ is μ-negligible, the set $A \subset \Omega$ where φ is differentiable is a set of full measure for μ (i.e., $\mu(A) = 1$).

Step 2. Uniqueness of the Target Point.
Let $\Gamma = \text{supp}(\gamma)$. We consider the points strictly inside our full-measure set: let $x_0 \in A \cap \pi_X(\Gamma)$ and suppose $(x_0, y_0) \in \Gamma$.

By Proposition 2.3.1, because φ is differentiable at x_0, the gradients must match:

$$\nabla\varphi(x_0) = \nabla h(x_0 - y_0)$$

where equality is in the sense of $\nabla\varphi(x_0) \in \partial h(x_0 - y_0)$.

Because h is strictly convex, Theorem 1.1.29 guarantees that a single subgradient vector cannot belong to the subdifferential of more than one point. Therefore, the vector $\nabla\varphi(x_0)$ uniquely identifies the pre-image $z = x_0 - y_0$. This allows us to rigorously invert the gradient and define a unique target point y_0 for every $x_0 \in A$:

$$y_0 = x_0 - (\nabla h)^{-1}(\nabla\varphi(x_0)) := T(x_0)$$

This proves that for μ-almost every x_0, there is exactly one y_0 such that $(x_0, y_0) \in \Gamma$.

Step 3. Concentration on the graph.
We have established that the support of γ (intersected with $A \times \Omega$) consists exclusively of pairs in the form $(x, T(x))$ in Step 2. Let $\text{Graph}(T) := \{(x, T(x)) : x \in A\}$. Because $\mu(A) = 1$ and $\gamma \in \Pi(\mu, \nu)$ has marginal μ, we are guaranteed that $\gamma(A \times \Omega) = 1$.

Therefore, γ is entirely concentrated on the graph of T:

$$\gamma(\text{Graph}(T)) = 1$$

To prove that $\gamma = (\text{id}, T)_\# \mu$, we test it against any measurable set $E \subset \Omega \times \Omega$. Since γ is concentrated on the graph, the measure of E is simply the measure of its intersection with the graph:

$$\gamma(E) = \gamma(E \cap \text{Graph}(T))$$

A point (x, y) belongs to $E \cap \text{Graph}(T)$ if and only if $y = T(x)$ and $x \in \{x' \in A : (x', T(x')) \in E\}$. Let E_X denote this set of x-coordinates. By the definition of the marginal $(\pi_X)_\#\gamma = \mu$, projecting this graph-restricted set down to X gives:

$$\gamma(E \cap \text{Graph}(T)) = \mu(E_X) = \mu(\{x : (x, T(x)) \in E\})$$

By the definition for pushforward of a measure, this right-hand side is exactly the definition of the pushforward $(\mathrm{id}, T)_{\#}\mu(E)$. Thus, $\gamma = (\mathrm{id}, T)_{\#}\mu$.

Finally, because the transport map $T(x)$ was constructed entirely using the given cost c and the potential φ (without any dependence on the specific choice of optimal γ), any optimal plan must be concentrated on this exact same graph. Therefore, the optimal transport plan γ is absolutely unique. ■

A direct corollary of the theorem is the existence of solution of (MP) for the case $h(x) = |x|^p$.

Corollary 2.3.4

Under the standing assumptions on $\Omega \subset \mathbb{R}^d$ and μ, $\nu \in \mathcal{P}(\Omega)$ as given in Theorem 2.3.3. We have a solution of (MP) for $c(x, y) = |x - y|^p$ for $p \in (1, +\infty)$.

> **Notes 2.3.5.** From now, if a transport plan γ is induced by a transport map T, i.e. $\gamma = (\mathrm{id}, T)_{\#}\mu = \gamma$, then we denote γ as γ_T.

Remark 2.3.6. An additional result can be obtained if we also add the assumption for ν being absolutely continuous to the Lebesgue measure. Then Theorem 2.3.3 also say that there is an optimal map S for the other way around, i.e. $\gamma = (S, \mathrm{id})_{\#}\nu$. In particular, for γ-a.e point (x, y). we have $y = T(x)$ and $x = T(y)$, which means $S(T(x)) = x$ μ-a.e., i.e. x μ-a.e. point x Hence, T is invertible on a set of full measure and its inverse is the optimal map from ν to μ.

§2.3.2 The Quadratic Case in $\mathbb{R}^d$: Brenier's Theorem

In this subsection, we shall consider the special case of

$$c(x, y) = \frac{|x - y|^2}{2} = \frac{|x|^2 + |y|^2}{2} - \langle x, y \rangle .$$

Actually, this case was the result of Brenier in [Bre87] which was independent from our previous approach and it also uses different approach from this note. This form of c suggest connections between our notion of c-concavity and the notion of convexity. Indeed, we have a relation as follows.

Proposition 2.3.7

Given a function $\chi : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$, define $u_\chi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ through

$$u_\chi(x) = \frac{|x|^2}{2} - u(x).$$

Then, $u_{\chi^c} = (u_\chi)^*$. In particular, a function ζ is c-concave if and only if u_ζ is convex and l.s.c.

Proof. For any $x \in \mathbb{R}^d$, we have

$$u_{\chi^c}(x) = \frac{|x|^2}{2} - \inf_{y \in \mathbb{R}^d} \left(\frac{|x - y|^2}{2} - \chi(y) \right) = \sup_{y \in \mathbb{R}^d} \langle x, y \rangle - \left(\frac{|y|^2}{2} - \chi(y) \right) = (u_\chi)^*(x).$$

For the the second passage, if $\zeta = \psi^c$ for some ψ, we have $u_\zeta = u_{\psi^c} = (u_\psi)^*$ hence u_ζ is convex and l.s.c. by Corollary 1.1.24. Conversely, we can also work by Theorem 1.1.23 to have $u_\zeta = ((u_\zeta)^*)^* = (u_{\zeta^c})^* = u_{\zeta^{cc}}$ which implies $\zeta = \zeta^{cc}$ and by Proposition 1.2.3 we have ζ c-concave. ■

As a consequence ot the proposition and the fact that for $h(x) = \dfrac{|x|^2}{2}$ we have $\nabla h(x) = x$ for every $x \in \mathbb{R}^d$, we have a better-looking closed formula of Theorem 2.3.3 as

$$T(x) = x - (\nabla h)^{-1}(\nabla\varphi(x)) = x - \nabla\varphi(x) = \nabla\left(\frac{|x|^2}{2} - \varphi(x)\right) = \nabla u_\varphi(x)$$

where u_φ is a convex since φ is c-concave. This suggest the converse of Theorem 2.3.3 to hold, that is, $T = \nabla u$ for some convex function (with additional assumptions) implies optimality on $\Pi(\mu, \nu)$ where $\nu = T_\#\mu$. This claim is true and will be discussed in Subsection 2.3.3.

Let us try to broaden our approach on solving the quadratic case by letting the domain Ω be unbounded (in this case, we take $\Omega = \mathbb{R}^d$). We should first investigate the problem (KP) the quadratic case. Notice that for any $\gamma \in \Pi(\mu, \nu)$ we have

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{|x-y|^2}{2}\, d\gamma(x,y) = \int_{\mathbb{R}^d} \frac{|x|^2}{2}\, d\mu(x) + \int_{\mathbb{R}^d} \frac{|y|^2}{2}\, d\nu(x) - \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle x, y\rangle\; d\gamma(x,y).$$

Since the term $\displaystyle\int_{\mathbb{R}^d} \frac{|x|^2}{2}\, d\mu(x) + \int_{\mathbb{R}^d} \frac{|y|^2}{2}\, d\nu(x)$ is independent of γ, our minimization problem turns into maximizing $\displaystyle\int_{\mathbb{R}^d\times\mathbb{R}^d} \langle x, y\rangle\; d\gamma(x,y)$. Naturally, we also need the assumptions $\displaystyle\int_{\mathbb{R}^d} |x|^2\, d\mu(x), \int_{\mathbb{R}^d} |y|^2\, d\nu(y) < +\infty$.

Notice that for the quadratic cost also satisfies the assumption of Theorem 2.3.3, hence an optimal plan γ exists and in particular it is optimal for the cost $c'(x,y) = -\langle x, y\rangle$. Theorem 2.2.14 gives us the fact that γ is concentrated in a c'-CM set Γ. In particular, our c'-CM set is actually CM in the sense of convexity. By Rockafellar Theorem (c.f. Theorem 1.2.6) which is a particular case of Theorem 1.2.9, we have

$$\Gamma \subset \{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d : \varphi(x) + \varphi^*(y) = \langle x, y\rangle\}.$$

for some convex function γ. This ease our analysis to the case of Legendre transformation and it also suggest us a candidate of another duality formula

$$\int_\Gamma \langle x, y\rangle\; d\gamma(x,y) = \int_{\pi_X(\Gamma)} \varphi(x)\, d\mu(x) + \int_{\pi_Y(\Gamma)} \varphi^*(y)\, d\nu(y)$$

(assuming integrability for φ and φ^*) in particular, if we go back to our original problem we have

$$\int_\Gamma c(x,y)\, d\gamma(x,y) = \int_{\pi_X(\Gamma)} u_\varphi(x)\, d\mu(x) + \int_{\pi_Y(\Gamma)} u_{\varphi*}(y)\, d\nu(y)$$

Since our domain is unbounded, we cannot have boundedness for u_φ and $u_{\varphi*}$. However, we can make a variance on the restriction of u_φ and $u_{\varphi*}$ to be integrable instead. The consideration gives us the naturality of the following variant of (DP).

Problem 2.3.8 (Dual Problem Variation)

Let $X = Y = \mathbb{R}^d$. Given $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ such that $\int_{\mathbb{R}^d} |x|^2 \, d\mu(x), \int_{\mathbb{R}^d} |y|^2 \, d\nu(y) < +\infty$, and the cost function $c(x,y) = \frac{|x-y|^2}{2}$. Consider the following variant of (DP):

$$\sup \left\{ \int_{\mathbb{R}^d} \phi \, d\mu + \int_{\mathbb{R}^d} \psi \, d\nu : \varphi \in L^1(\mu), \psi \in L^1(\nu) \text{ such that } \varphi \oplus \psi \leq c \right\}. \quad \text{(DP-var)}$$

Following similar argument for Theorem 2.2.11, we have the following result.

Theorem 2.3.9

Under the standing assumptions of (DP-var), then (DP-var) admits a solution (φ, ψ) where u_φ, u_ψ are convex and conjugate with respect to the Legendre transformation. Moreover, we have $\max(DP - var) = \min(KP)$.

Proof. Theorem 2.1.14 ensures the existence of an optimal transport plan γ and by Theorem 2.2.10, there exists a c-concave function φ such that

$$\text{supp}(\gamma) \subset \{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d : \varphi(x) + \varphi^c(y) = c(x,y)\}.$$

Since φ and $\psi = \varphi^c$ are proper c-concave functions, u_φ and u_ψ are proper, convex, and conjugate to each other (in particular, $u_\varphi \in \Gamma(\mathbb{R}^d)$ [sorry for redundant notations!]). Recalling our definition of as given in Theorem 2.2.10, we have φ is an infimum of continuous function (since the quadratic cost is continuous, thankfully) which is an upper semicontinuous function and thus measurable. Analogously, φ^c is measurable as well.

From our considerations, for fixed $(q,p) \in \text{supp}(\gamma)$ we have $u_\varphi(q) + u_{\varphi^c}(p) = u_\varphi(q) + (u_\varphi)^*(p) = c(q,p)$ and by Proposition 1.1.26 we have $p \in \partial u_\varphi(q)$. Therefore, for any $x \in \mathbb{R}^d$ and letting $b = (u_\varphi(q) - \langle p, q \rangle) \in \mathbb{R}$:

$$\begin{aligned} u_\varphi(x) - u_\varphi(q) \geq \langle p, x - q \rangle \implies & u_\varphi(x) \geq \langle p, x \rangle + (u_\varphi(q) - \langle p, q \rangle) \\ \iff & \frac{|x|^2}{2} - \varphi(x) \geq \langle p, x \rangle + b. \\ \iff & \frac{|x|^2}{2} - \langle p, x \rangle - b \geq \varphi(x) \end{aligned}$$

and by the Cauchy-Schwarz inequality we have an upper bound for φ^+ as

$$\varphi^+(x) = \max\{\varphi(x), 0\} \leq \max\left\{ \frac{|x|^2}{2} - \langle p, x \rangle - b, 0 \right\} \leq \left| \frac{|x|^2}{2} - \langle p, x \rangle - b \right| \leq \frac{|x|^2}{2} + |p||x| + |b|.$$

By our assumption of $\int_{\mathbb{R}^d} |x|^2 \, d\mu(x) < \infty$, we also have $\int_{\mathbb{R}^d} |x| \, d\mu(x) < \infty$ by Hölder Inequality. Hence,

$$\int_{\mathbb{R}^d} \varphi^+(x) \, d\mu(x) \leq \int_{\mathbb{R}^d} \frac{|x|^2}{2} + |p||x| + |b| \, d\mu(x) < +\infty$$

and thus $\varphi^+ \in L^1(\mu)$ Analogously, we also have $(\varphi^c)^+ \in L^1(\nu)$. Now we can safely integrate $\varphi \oplus \varphi^c$ with respect to γ since the case $\infty - \infty$, obtaining

$$\int_{\mathbb{R}^d} \varphi \, d\mu + \int_{\mathbb{R}^d} \varphi^c \, d\nu = \int_{\mathbb{R}^d \times \mathbb{R}^d} \varphi \oplus \varphi^c \, d\gamma = \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y) \, d\gamma \geq 0.$$

This forces φ^- and $(\varphi^c)^-$ being integrable and hence $\varphi \in L^1(\mu)$ and $\varphi^c \in L^1(\nu)$ (with $\varphi \oplus \varphi^c \leq c$). The conclusion of $\max(DP - var) = \min(KP)$ follows analogously as we did in Theorem 2.2.11. ■

With the validity of (DP-var), we can use our obtained φ safely to solve Theorem 2.3.3 in the case of unbounded domain following the works of Brenier.

Theorem 2.3.10 (Brenier)
Let $X = Y = \mathbb{R}^d$. Given $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ such that $\int_{\mathbb{R}^d} |x|^2 \, d\mu(x), \int_{\mathbb{R}^d} |y|^2 \, d\nu(y) < +\infty$, and $\mu \ll \mathcal{L}^d$. For the cost function $c(x, y) = \frac{|x - y|^2}{2}$, there exists an unique optimal transport map T that solves (MP) and it is of the form $T = \nabla u$ for a convex function u.

Proof. We simply replace Step 1 on Theorem 2.3.3 and then continuing an analogous result of Step 2 and Step 3 to have our result.

Theorem 2.1.14 ensures the existence of an optimal transport plan γ. Following the strategy on Theorem 2.3.3), our c-concave Kantorovich potential $\varphi \in L^1(\mu)$ for (2.3.9) whose existence is guaranteed by Theorem DP-var and in $\text{supp}(\gamma)$ we have the equality

$$\varphi(x) + \varphi^c = c(x, y) \qquad \forall x, y \in \text{supp}(\gamma).$$

Since $\varphi \in L^1(\mu)$, we have φ finite μ-a.e. and thus u_φ is also finite μ-a.e. Hence, Proposition 1.1.11 implies $\{u_\varphi < +\infty\}$ is a convex set and $\mu(\{u_\varphi < +\infty\}) = 1$. Using the fact that the boundary of a convex set is $\mathcal{L}^d$-negligible. By our assumption on $\mu \ll \mathcal{L}^d$, the interior of $\{u_\varphi < +\infty\}$ is full measure with respect to μ. Following the result of Theorem 1.1.16, we have u_φ is locally Lipschitz in the interior of $\{u_\varphi < +\infty\}$ and thus by Theorem 1.1.15 u_φ is differentiable $\mathcal{L}^d$-a.e. Let the terminating set of this process as A where $\mu(A)$ and u_φ (and thus φ as well) is differentiable in A. Therefore, we have successfully replaced Step 1 of Theorem 2.3.3 and the rest of argument follows analogously. ■

Remark 2.3.11. Actually, we can make the sharpest assumption on μ to just neglect the points where differentiability fails for u convex. As mentioned in p.17 of [San15], this set is $(d-1)$-rectifiable, in particular, it is contained $\mathcal{H}^{d-1}$-a.e. in a countable union of $(d-1)$ surfaces of class C^2 where $\mathcal{H}$ is the Hausdorff measure. Hence, we can replace $\mu \ll \mathcal{L}^d$ with the property that μ neglect such set.

We shall try to see for an example of an optimal transport map in the following example.

Example 2.3.12 (Exercise 3 of [San15])
Find the optimal transport map for the quadratic cost $c(x, y) = |x - y|^2$ between $\mu = f \cdot \mathcal{L}^2$ and $\nu = g \cdot \mathcal{L}^2$ in 2D, where $f(x) = \frac{1}{\pi}\chi_{B(0,1)}(x)$ and $g(x) = \frac{1}{8\pi}(4 - |x|^2)$.

Solution. By the definition of μ, we have $\mu \ll \mathcal{L}^2$ (and also $\nu \ll \mathcal{L}^2$, analogously). First, we must implicitly identify the support of the target measure ν. For $g(x)$ to be a valid non-negative probability density, we must restrict its support to the region where $4 - |x|^2 \geq 0$, which is $B(0, 2)$. We can verify this integrates to 1:

$$\int_{B(0,2)} g(x)\, dx = \int_0^{2\pi} \int_0^2 \frac{1}{8\pi}(4 - r^2) r\, dr\, d\theta = \frac{1}{4}\left[2r^2 - \frac{r^4}{4}\right]_0^2 = \frac{1}{4}(8 - 4) = 1$$

Thus, $g(x) = \dfrac{1}{8\pi}(4 - |x|^2)\chi_{B(0,2)}(x)$.

By Brenier's Theorem, since $\mu \ll \mathcal{L}^2$ and the cost is quadratic, the unique optimal transport map takes the form $T = \nabla u$ for some (strictly) convex scalar potential u. Because both μ and ν are radially symmetric, the uniqueness of the Brenier map guarantees that T must also be radially symmetric. We can try to search for a map of the form:

$$T(x) = R(|x|)\frac{x}{|x|}$$

where $r = |x|$ and $R : [0, 1] \to [0, 2]$ is a strictly increasing radial mapping satisfying $R(0) = 0$ and $R(1) = 2$. The idea of this map is to make a "stretching" map by a factor of $R(|x|)$.

For radially symmetric measures and a strictly increasing radial map, the pushforward condition $T_\#\mu = \nu$ reduces to matching the cumulative mass of balls centered at the origin. For any $r \in [0, 1]$, we must have $\mu(B(0, r)) = \nu(B(0, R(r)))$.

We shall compute the volume of our initial ball $B(0, r)$ as follows.

$$\mu(B(0, r)) = \int_0^{2\pi} \int_0^r \frac{1}{\pi} s\, ds\, d\theta = 2\pi \cdot \frac{1}{\pi}\left[\frac{s^2}{2}\right]_0^r = r^2$$

Next, we compute the mass of the target ball $B(0, R(r))$:

$$\nu(B(0, R)) = \int_0^{2\pi} \int_0^R \frac{1}{8\pi}(4 - s^2) s\, ds\, d\theta = \frac{1}{4}\int_0^R (4s - s^3)\, ds = \frac{1}{4}\left(2R^2 - \frac{R^4}{4}\right) = \frac{R^2}{2} - \frac{R^4}{16}$$

Equating the two masses yields $r^2 = \dfrac{R^2}{2} - \dfrac{R^4}{16}$, which gives us a solution for the equation as

$$R^2 = 4 \pm 4\sqrt{1 - r^2}.$$

Since we want $R(0) = 0$, we choose $R^2 = 4 - 4\sqrt{1 - r^2}$ and thus $R(r) = 2\sqrt{1 - \sqrt{1 - r^2}}$. Indeed, this choice of R ensures $T_\#\mu = \nu$.

To ensure there are no gaps, we verify $T(x) = \nabla u(x)$. Define $u(x) = \displaystyle\int_0^{|x|} R(s)\, ds$. Since $R(s) \geq 0$ and $R'(s) > 0$ for $s \in (0, 1)$, $u(x)$ is a strictly convex, radially symmetric scalar potential, and $\nabla u(x) = R(|x|)\dfrac{x}{|x|} = T(x)$, completely satisfying Brenier's theorem.

Therefore, the unique optimal transport map $T(x)$ is given by:

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

■

Remark 2.3.13. Within this example, we established $T = \nabla u$ where u is strictly convex, whereas Theorem 2.3.10 only guarantees that u is convex. We can deduce this strict convexity because we have $\nu \ll \mathcal{L}^2$. If u were not strictly convex, it would contain flat affine regions, causing its gradient $T = \nabla u$ to be constant over a set of strictly positive Lebesgue measure. This would force T to map a region of positive mass to a single point, creating an atom in the target measure ν. Because $\nu \ll \mathcal{L}^2$, it cannot contain atoms, which completely exclude flat regions for u. Therefore, u must be strictly convex, making T invertible almost everywhere as noted in Remark 2.3.6.

§2.3.3 Sufficient Conditions for Optimality of Transport Maps

In this subsection, we shall consider the "converse" of (MP), that is, if we have a transport map on certain case—in our case, the case of c satisfying the twist condition in Definition 2.3.2 that is more general than the case we considered in Subsection 2.3.1—can we have optimality for such map? The answer is "positive" (we still need additional assumptions) and will be considered under general case following the assumptions we have made beforehand.

Theorem 2.3.14 (Sufficient Condition, Ω Compact)
Let $\Omega \subset \mathbb{R}^d$ be compact and c be a C^1 cost function satisfying the twist condition (c.f. Definition 2.3.2) on $\Omega \times \Omega$. Suppose that $\mu \in \mathcal{P}(\Omega)$ and $\varphi \in c-conc(\Omega)$ are given, that φ is differentiable μ-a.e. and that $\mu(\partial\Omega) = 0$. Suppose that the map T satisfies $\nabla_x c(x, T(x)) = \nabla\varphi(x)$. Then T is optimal for the transport cost c between the measures μ and $\nu := T_{\#}\mu$.

Proof. As what we did in Theorem 2.2.11, we have the functions φ and $\psi := \varphi^c$ are uniformly continuous and bounded since φ and ψ shares the same modulus of continuity as c.

Let A be a μ-full measure set such that φ is differentiable in A, we can neglect $\partial\Omega$ from our assumption. Fix $x_0 \in A$, since φ and c are continuous on a compact set, there exists $y_0 \in \Omega$ such that

$$\varphi(x_0) = \inf_{y\in\Omega} c(x_0, y) - \psi(y) = c(x_0, y_0) - \psi(y_0).$$

and

$$\psi^c(y_0) = \int_{x\in\Omega} c(x, y_0) - \varphi(x) \leq c(x, y_0) - \varphi(x) \quad \forall x \in \Omega$$

where the minimum is realized at $x = x_0$. Therefore, $x \mapsto c(x, y_0) - \varphi(x)$ is minimal at $x = x_0$ and thus

$$\nabla_x c(x_0, y_0) = \nabla\varphi(x_0).$$

The twist condition allows us to define $y_0 := T(x_0)$. Note that T is well-defined in A and thus T is defined μ-a.e. This proves the equality

$$\varphi(x_0) + \psi(T(x_0)) = c(x_0, T(x_0))$$

μ-a.e. Taking $\nu = T_{\#}\mu$, we can integrate with respect to μ and with the Change of Variable formula (c.f. Theorem 1.1.5) to get

$$\int_\Omega c(x, T(x))\, d\mu(x) = \int_\Omega \varphi\, d\mu + \int_\Omega (\psi \circ T)\, d\mu = \int_\Omega \varphi\, d\mu + \int_\Omega \psi\, d\nu.$$

Since pair (φ, ψ) is admissible for (DP) and the first integral equals to the cost of T in (MP) (and thus (KP)), we have

$$(KP) \leq \int_\Omega c(x, T(x))\, d\mu(x) = \int_\Omega \varphi\, d\mu + \int_\Omega \psi\, d\nu \leq (DP) \leq (KP).$$

Hence, we have proved the optimality of T. ■

A proof of similar spirit on the quadratic case concerned the optimality of T when it is the gradient of a convex function, instead of linking T to the Kantorovich potential as what we did in Theorem 2.3.14 by exploiting the characterization of Proposition 1.1.26 to u, although it is similar to that of the property for the Kantorovich potential.

Theorem 2.3.15 (Sufficient Condition, Quadratic Case)
Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ such that $\int_{\mathbb{R}^d} |x|^2\, d\mu(x) < +\infty$ and that $u : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is convex and differentiable μ-a.e. Set $T = \nabla u$ and suppose $\int_{\mathbb{R}^d} |T(x)|^2\, d\mu(x) < +\infty$. Then T is optimal for the transport cost $c(x, y) := \frac{|x-y|^2}{2}$ between the measures μ and $\nu := T_\# \mu$.

Proof. Recall from Proposition 1.1.26 in general case that for any x, $y \in \mathbb{R}^d$ we have

$$u(x) + u^*(y) = \langle x, y\rangle \iff y \in \partial u(x) \qquad u(x) + u * (y) \geq \langle x, y\rangle$$

On the set A of μ-full measure where u is differentiable, Theorem 1.1.28 gives

$$u(x) + u^*(y) = \langle x, y\rangle \iff y = \nabla \varphi(x).$$

Hence, for any $x \in A$, we can define $T(x) = \nabla u(x)$ and by our assumption $|T| \in L^2(\mu)$. Consider the measure $\gamma_T = (\mathrm{id}, T)_\# \mu$. For any $\gamma \in \Pi(\mu, \nu)$ we have

$$\begin{aligned}\int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y\rangle\, d\gamma(x, y) \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} u(x) + u^*(y)\, d\gamma(x, y) &= \int_A u(x) + u^*(T(x))\, d\mu(x) \\ &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y\rangle\, d\gamma_T(x, y)\end{aligned}$$

Hence by the considerations we did in the early part of Subsection 2.3.2, we have

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)\, d\gamma(x, y) \geq \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)\, d\gamma_T(x, y).$$

We have obtained the optimality of T. ■

We conclude this section by passing to a more general criterion that is more satisfactory. The main idea of this result is that we proved that optimal plans are necessarily concentrated on a c-CM set as what we did in Theorem 2.2.14, but the converse is also true: a plan which is concentrated on a c-CM set is optimal, at least in reasonable case such as the cost c is uniformly continuous and bounded. The reason we take this approach is not to concern the measurability (c continuous is enough for measurability) and integrability (in this case, as what we did in Theorem 2.2.10, our functions in the subsequent proof are measurable and bounded) and we will use this result to consider the stability of solutions in the last part of this chapter.

Theorem 2.3.16 (Sufficient Condition, Polish Domains, Uniformly Continuous and Bounded Cost)

Suppose that $\gamma \in \mathcal{P}(X \times Y)$ is given, X and Y are Polish spaces, $c : X \times Y\mathbb{R}$ is uniformly continuous and bounded, and that $\operatorname{supp}(\gamma)$ is c-CM. Then γ is an optimal plan between $\mu = (\pi_X)_\#\gamma$ and $\nu = (\pi_Y)_\#\gamma$ for the cost c.

Proof. Since $\operatorname{supp}(\gamma)$ is c-CM. By Theorem 1.2.9, there exists a c-concave function φ, for which φ and φ^c are uniformly continuous and bounded from our assumption in c, such that

$$\operatorname{supp}(\gamma) \subset \{(x, y) \in X \times Y : \varphi(x) + \varphi^c(y) = c(x, y)\}.$$

Thus, (φ, φ^c) is admissible to (DP) and we have

$$(KP) \leq \int_{X\times Y} c(x, y)\, d\gamma(x, y) = \int_X \varphi\, d\mu + \int_Y \varphi^c\, d\nu \leq (DP) \leq (KP).$$

We have obtained optimality for γ. ■

§2.4 Interesting Counterexamples to the Monge-Kantorovich Problem

In the start of this note, we claimed that the constraint is not closed under (weak) convergence which makes us unable to use Theorem 2.1.2 to find a minimizer.

Example 2.4.1 (Monge Constraint is not Closed, Exercise 1 of [San15])

Define $f_n : [0, 1] \to [0, 1]$ by periodizing $x \mapsto nx - \lfloor nx \rfloor = \{nx\}$ in $[0, 1]$. One can prove that $(f_n)_\#\mathcal{L}\llcorner[0, 1] = \mathcal{L}\llcorner[0, 1]$ for all $n \in \mathbb{N}$ using the Change of Variables formula in Theorem 1.1.5 and $f_n \rightharpoonup \frac{1}{2}$ where $\left(\frac{1}{2}\right)_\# \mathcal{L}\llcorner[0, 1] = \delta_{\frac{1}{2}}\llcorner[0, 1] \neq \mathcal{L}\llcorner[0, 1]$.

Another particular issue with Monge problem is the fact that we may not have a transport map which makes the problem unreasonable to solve. Even though we have mentioned an example in the start of Chapter 1, we can generalize it slightly more However, in Section 2.5, we have a sufficient condition where this issue will not happen.

Example 2.4.2 (Transport Map May Not Exist)

Let X and Y be metric spaces and consider $\mu = \delta_{x_0} \in \mathcal{P}(X)$ for some $x_0 \in X$ and $\nu \in \mathcal{P}(Y)$ is not a Dirac delta measure, then no transport exist since for any map T since $T_\#\delta_{x_0} = \delta_{T(x_0)}$. In particular, one can prove that if $\mu = \delta_{x_0}$, then $\Pi(\mu, \nu) = \{\delta_{x_0} \otimes \nu\}$. Generalizing this result, if μ has atoms and ν has none, for any map T, $T_\#\mu$ has atoms and thus $T_\#\mu \neq \nu$.

This consideration leads to our assumption that our starting mass μ need to be atomless in solving (MP).

In Section 2.3, we are sure that for $c(x, y) = h(x - y)$ strictly convex, we have an unique optimal plan and map. However, what happens if c does not have such property, for example, if $h(x) = |x|$ which is not strictly convex? Uniqueness can be negative as in our following consideration

Example 2.4.3 (Uniqueness May Not Occur for (KP) and (MP) in General)
Let X, $Y \subset \mathbb{R}$ where $\sup X \leq \inf Y$. Consider the cost $c(x, y) = |x - y| = y - x$. For any $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and $\gamma \in \Pi(\mu, \nu)$ we have

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

Therefore, any transport plan γ is optimal. Moreover, if a transport map exists—in particular in the case μ atomless as we will show in Section 2.5—then any transport map is also optimal.

Lastly, we shall consider a case that we may have a solution for (KP), but not (MP), as shown in the following example.

Example 2.4.4 (Solution of (KP) exists, but not (MP))
Let $c(x, y) = |x - y|^2$ (or also $c(x, y) = |x - y|$) and consider $\mu = \mathcal{H}^1 \llcorner A$ and $\nu = \frac{1}{2}\mathcal{H}^1 \llcorner B + \frac{1}{2}\mathcal{H}^1 \llcorner C$, with

$$A = \{[0, y] : y \in [0, 1]\}, \quad B = \{[-1, y] : y \in [0, 1]\}, \quad C = \{[1, y] : y \in [0, 1]\}.$$

Then (KP) admits a solution, but (MP) does not even though we have transport maps.

Solution. Before we start, notice that even though c is induced by a strictly convex function, μ does not neglect the boundary of our domain and hence Theorem 2.3.3 cannot be used. One may also argue that our domain and codomain is different, however this problem can be troubleshooted by enlarging the domain Ω to contain A and $B \cup C$ where μ and ν does not make any issue here.

For the sake of visualization and intuition, this problem is the same as distributing mass from a center into two different locations. Intuitively, we want to say that the cost is minimized when we split the mass evenly into two locations. Indeed, this claim is true and by the optimal transport plan must follow this rule. Hence, we cannot have a transport map that realizes this strategy.

Notice that for any $(x, y) \in A \times (B \cup C)$ we have $c(x, y) \geq 1$ and thus for any $\gamma \in \Pi(\mu, \nu)$ we have

$$\int_{A\times(B\cup C)} c(x, y)\, d\gamma(x, y) \geq 1.$$

Let $e = (1, 0)$. By defining the map $T^{\pm}(x) = x \pm e$, we have $c(x, T^{\pm}(x)) = 1$ and thus for $\gamma = \dfrac{\gamma_{T^+} + \gamma + T^-}{2}$, i.e. we split the mass evenly, we have

$$\int_{A\times(B\cup C)} c(x, y)\, d\gamma(x, y) = 1.$$

Hence, such γ is a solution to (KP). For any $n \in \mathbb{N}$, consider the map $T_n : A \to B \cup C$ by first splitting A into $2n$ equal parts $\{A_k\}_{0\leq k\leq 2n}$ and also splitting B and C into n

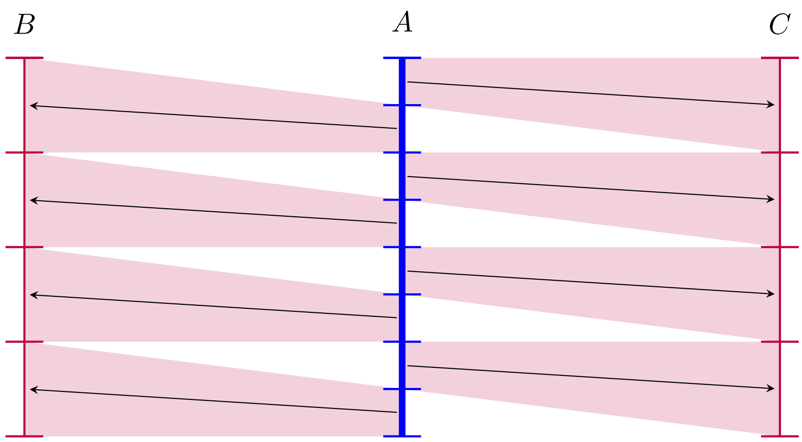

Figure 2.1: visualization of T_n for $n = 4$.

equal parts each $\{B_k\}_{0\leq k\leq n}$, $\{C_k\}_{0\leq k\leq n}$. Then, T_n is defined as a piecewise affine map which sends A_{2i-1} to B_i and A_{2i} to C_i (we don't really care about the boundaries since it is neglectible). One can verify that such T_n is admissible to (MP). Because we are alternating at a scale of $\frac{1}{2n}$, the maximum vertical displacement any particle experiences is strictly bounded by $\frac{1}{n}$. Therefore, the cost for the map T_n is bounded by:

$$\int_A c(x, T_n(x))d\mu = \int_A (1^2 + (\Delta y)^2)d\mu \leq \int_A (1 + \frac{1}{n^2})d\mu = 1 + \frac{1}{n^2}$$

As $n \to \infty$, $M(T_n) \to 1$. Thus, $\inf(\text{MP}) = 1$.

Unfortunately, no optimal plan γ can be induced by a map T. Suppose such T exists. Then

$$\int_A c(x, T(x))\, d\mu(x) = 1 \implies c(x, T(x)) = 1 \quad \text{on } \operatorname{supp}(\mu).$$

Therefore, for any $(0, y) \in \operatorname{supp}(\mu) =: X$, we have $T(0, y) = (\pm 1, y)$, i.e. $T(0, y) = (\pm 1, y)$ for a.e. $y \in [0, 1]$. Let $B_T = T^{-1}(B)$ and $C_T = T^{-1}(C)$ where $\mu(B_T \cup C_T) = 1$ which implies

$$\nu = \frac{1}{2}\mathcal{H}^1 \llcorner B + \frac{1}{2}\mathcal{H}^1 \llcorner C(T(B_T) \cup T(C_T)) = T_\# \mu(T(B_T) \cup T(C_T)) = 1.$$

Since $T(B_T) \subset B$ and $T(C_T) \subset C$, this implies $\mathcal{H}^1 \llcorner B(T(B_T)) = \mathcal{H}^1 \llcorner C(T(C_T)) = 1$.

Let $I \subset [0, 1]$ be any open interval, and define the target segment on the left as $B_I = \{-1\} \times I \subset B$. By the definition of the target measure ν, we have:

$$\nu(B_I) = \frac{1}{2}\mathcal{H}^1(B_I) = \frac{1}{2}|I|$$

By the definition of the pushforward measure, we must also have

$$\nu(B_I) = \mu(T^{-1}(B_I)) = \mathcal{H}^1(B_T \cap I) = \int_I \chi_{B_T}(y)dy$$

Equating the two measure evaluations gives:

$$\int_I \chi_{B_T}(y)dy = \frac{1}{2}|I| \implies \frac{1}{|I|}\int_I \chi_{B_T}(y)dy = \frac{1}{2}$$

Since this holds for any interval $I \subset [0,1]$, we can apply the Lebesgue Differentiation Theorem. By taking a sequence of intervals shrinking to a point $y_0 \in [0,1]$, the theorem guarantees that:

$$\lim_{|I| \to 0} \frac{1}{|I|} \int_I \chi_{B_T}(y) dy = \chi_{B_T}(y_0) \quad \text{for a.e. } y_0 \in [0,1]$$

This forces $\chi_{B_T}(y_0) = \dfrac{1}{2}$ for almost every $y_0 \in [0,1]$. However, by definition, an indicator function can only take values in $\{0,1\}$ which is a contradiction. Thus, no such measurable map T can exist, and the we have proved (MP) admits no solution. ■

§2.5 Why Kantorovich is a relaxation of Monge?

In this section, let us set again $K(\gamma) := \int_{X \times Y} c(x,y)\, d\gamma(x,y)$ and now set $M(T) = \int_X c(x, T(x))\, d\mu(x) = \int_{X \times Y} c(x,y)\, d\gamma_T(x,y) = K(\gamma_T)$. Therefore, (MP) can be rewritten as

$$\min\{J(\gamma) : \gamma \in \Pi(\mu, \nu)\},$$

where

$$J(\gamma) = \begin{cases} K(\gamma) = M(T) & \text{if } \gamma = \gamma_T \\ +\infty & \text{otherwise} \end{cases}$$

our definition of J restricts the minimization problem to the class of plan induced by a transport map. This is useful in order to consider (MP) and (KP) as two problems on the same set of admissible objects, where the only difference is the functional to be minimized, that is J or K.

With our background established, we shall see why Kantorovich decided to replace J with K. In this section, we shall prove $\inf J = \inf K$ for which this case is actually easy if the minimizer γ for K is of the form γ_T since we will arrive at equality. However, as Example 2.4 shows, this is not always the case, even though we still retain $\inf J = \inf K$.

To make this justification possible, we shall finally define the notion of *relaxation.*

Definition 2.5.1 (Relaxation)

Let $F : X \to \mathbb{R} \cup \{+\infty\}$ be a given functional on a metric space X, and suppose that it is bounded from below. We define the **relaxation** of F as the functional $\bar{F} : X \to \mathbb{R} \cup \{+\infty\}$ as

$$\bar{F}(x) = \sup\{G(x) : G : X \to \mathbb{R} \cup \{+\infty\}, \text{ l.s.c. with } G \leq F\}$$

which is the maximal functional among all those $G : X \to \mathbb{R} \cup \{+\infty\}$, which exists since the notion of l.s.c. is closed under suprema.

As a consequence of this definition, we have $\bar{F} \leq F$ which implies $\inf \bar{F} \leq \inf F$ and the constant function $l := \inf F \leq F$ implies $l \leq \inf \bar{F}$. Therefore, $\inf F = \inf \bar{F}$.

Proposition 2.5.2

There exists a representation formula of $\bar{F}$ as

$$\bar{F}(x) = \inf\left\{\liminf_{n\to\infty} F(x_n) : x_n \to x\right\}.$$

Proof. Define $H(x) = \inf\{\liminf_{n\to\infty} F(x_n) : x_n \to x\}$. We shall prove $\bar{F} \le H$ and $H \le \bar{F}$. Let $x \in X$ and take any converging sequence $x_n \to x$, for any $G \le F$ l.s.c. we have

$$G(x) \le \liminf_{n\to\infty} G(x_n) \le \liminf_{n\to\infty} F(x_n) \implies \bar{F}(x) \le H(x).$$

Conversely, we already have $H \le F$ since we can take a constant sequence $x_n = x$. We will show H is l.s.c. Let $x^m \to x$. For each $m \in \mathbb{N}$, there exists $x_n^m \to x^m$. We define $y_k := x_{n(k)}^k$ where $n(k)$ is taken such that $d(x^k, x_{n(k)}^k) < \dfrac{1}{k}$ and $F(x_{n(k)}^k) < H(x^k) + \dfrac{1}{k}$. Hence, $y_k \to x$ and we have

$$H(x) \le \liminf_{k\to\infty} F(y_k) \le \liminf_{k\to\infty} H(x^k).$$

Hence H is l.s.c. and $H \le \bar{F}$. ■

In this section, we claim that under some assumptions, K is actually the relaxation of J in the sense of Definition 2.5.1. In our case, by chance, this relaxation is also continuous, instead of only semi-continuous, and that it coincides with J on $\{J < +\infty\}$.

The assumptions are the following: the domain $\Omega \subset \mathbb{R}^d$ is taken compact, c continuous, μ atomless (i.e. for every $x \in \Omega$, $\mu(\{x\}) = 0$) to avoid Example 2.4.2. We use compactness for the domain for simplicity.

We start to prove our claim by some preliminary results as follows. The proof follows the content of Section 1.5 in [San15] faithfully.

Lemma 2.5.3

Let μ, $\nu \in \mathcal{P}(\mathbb{R})$. If μ is atomless, then there exists at least a transport map T such that $T_\#\mu = \nu$.

Proof. As a consequence of the consideration in Subsection 2.3.2. By the fact that every convex function ψ in $\mathbb{R}$ has at most countable non-differentiability points as a consequence of the fact that if ψ is convex, then the intervals $(\psi_l'(x), \psi_r'(x))$, where ψ_l' and ψ_r' denote the left and right derivatives (it is always defined, but there may be jumps), are all non-empty and disjoint when x ranges among non-differentiability points,i.e., ψ_l' and ψ_r' are monotonely increasing. Therefore, any such ψ is differentiable μ-a.e. since μ has no atoms.

Moreover, the non-differentiable points on Theorem 2.3.10 can be neglected (again, the assumption $\mu \ll \mathcal{L}^d$ is actually too strong) and we can safely use the theorem to have a $T = \psi'$ (it is defined μ-a.e. and actually optimal for the quadratic cost) for some convex function ψ where such T is monotonely increasing by our consideration and such T is a transport map between μ to ν. ■

Lemma 2.5.4

There exists a Borel map $\sigma_d : \mathbb{R}^d \to \mathbb{R}$ which is injective, its image is a Borel subset of $\mathbb{R}$, and its inverse map is Borel measurable as well.

Proof. First, note that it is sufficient to prove it for $d = 2$, since then we can proceed by induction $\sigma_d(x_1, \dots, x_d) = \sigma_2(x_1, \sigma_{d-1}(x_2, \dots, x_d))$.

Then, note also it is enough to define such a map on $(0,1)^2$ since we have the measurable bijection $(0,1)^2 \ni (x,y) \mapsto \left(\frac{1}{2} + \frac{\arctan x}{\pi}, \frac{1}{2} + \frac{\arctan y}{\pi}\right)$.

Then, consider the map which associates to the pair (x,y) where $x = 0.x_1x_2x_3\dots$ and $y = 0.y_1y_2y_3\dots$ (in decimal) to the point $0.x_1y_1x_2y_2x_3y_3\dots$. To avoid ambiguities, we decide that no decimal notation is allowed to end with a periodic 9 (i.e., 0.3472999... has to be wirtten as 0.3473). This is also why the image of this map will not be the whole interval, since the points like $0.39393939\dots$ are not obtained through this map. But this set of points is actually Borel measurable.

To check that our map is Borel measurable, as well as its inverse, since the pre-image of every interval defined by prescribing the first $2k$ digits of a number in $\mathbb{R}$ is justra recultangle in $\mathbb{R}^2$, that is the product of two intervals defined by prescribing the first k digits of every component. These particular intervals being a base for the Borel tribe proves the needed measurability. ∎

Corollary 2.5.5

If μ, ν are two probability measures on $\mathbb{R}^d$ and μ is atomless, then there exists at least a transport map T such that $T_\# \mu = \nu$.

Proof. This is simply obtained by using Lemma 2.5.3 to obtain a transport map T from $(\sigma_d)_\# \mu$ to $(\sigma_d)_\# \nu$ which are probability measures in $\mathbb{R}$ and then composing T into $\sigma_d \circ T \circ (\sigma_d)^{-1}$. ∎

One more lemma for our setup

Lemma 2.5.6

Consider on a compact metric space X, endowed with a probability $\rho \in \mathcal{P}(X)$, a sequence of partitions G_n, each G_n being a family of disjoint subsets C_i^n such that $\bigcup_{i \in I_n} C_i^n = X$ for every $n \in \mathbb{N}$. Suppose that $\text{size}(G_n) := \max_i \text{diam}(C_i^n)$ tends to 0 as $n \to \infty$ and consider a sequence of probability measures ρ_n on X such that, for every n and $i \in I_n$, we have $\rho_n(C_i^n) = \rho(C_i^n)$. Then, $\rho_n \rightharpoonup \rho$.

Proof. Let $m_i^n := \rho_n(C_i^n) = \rho(C_i^n)$. For any $\phi \in C(X)$ with ω as its modulus of continuity, we note that

$$\left|\int_X \phi\, d\rho_n - \int_X \phi\, d\rho\right| \leq \sum_{i\in I_n} \left|\int_{C_i^n} \phi\, d\rho_n - \int_{C_i^n} \phi\, d\rho\right| \leq \sum_{i\in I_n} \omega(\operatorname{diam} C_i^n)\cdot m_i^n$$
$$\leq \operatorname{size}(G_n) \to 0.$$

The inequality is justified by the fact that

$$c_{min}\cdot m_i^n \leq \int_{C_i^n} \phi\, d\rho_n \leq c_{max}\cdot m_i^n \quad \text{and} \quad c_{min}\cdot m_i^n \leq \int_{C_i^n} \phi\, d\rho \leq c_{max}\cdot m_i^n$$

implies

$$\left|\int_{C_i^n} \phi\, d\rho_n - \int_{C_i^n} \phi\, d\rho\right| \leq |c_{max} - c_{min}|\cdot m_i^n \leq \omega(\operatorname{diam} C_i^n)\cdot m_i^n.$$

where c_{max} and c_{min} are the maximum and minimum value of ϕ at C_i^n, respectively. This proves $\displaystyle\int_X \phi\, d\rho_n \to \int_X \phi\, d\rho$ and thus $\rho_n \rightharpoonup \rho$. ■

We can now prove the following theorem for which the proof is borrowed from Theorem 1.32 of [San15].

Theorem 2.5.7

On a compact subset $\Omega \subset \mathbb{R}^d$, the set of plans γ_T induced by a transport is dense in the set of plans $\Pi(\mu,\nu)$ whenever μ is atomless.

Proof. Fix n and consider any partition of Ω into sets K_i^n of diameter smaller than $\frac{1}{2n}$ (for instance, small cubes). The sets $C_{i,j}^n := K_i^n \times K_j^n$ make a partition of $\Omega\times\Omega$ with size smaller than $\frac{1}{n}$.

Let us now take any measure $\gamma \in \Pi(\mu,\nu)$. Thanks to Lemma 2.5.6, we only need to build a transport T sending μ to ν, such that γ_T gives the same mass as γ to each one of the sets $C_{i,j}^n$. To do this, define the columns $Col_i^n := K_i^n \times \Omega$ and denote by γ_i^n the restriction of γ on Col_i^n. Its marginal will be denoted by μ_i^n and ν_i^n.

Consider now, for each i, a transport map T_i^n sending μ_i^n to ν_i^n. It exists thanks to Corollary 2.5.5, since for each (i,n), we have $\mu_i^n \leq \mu$, which makes these measures atomless. Since the μ_i^n's are concentrated on disjoint sets, by "gluing" the transports T_i^n, we get a transport T sending μ to ν (using $\sum_i \mu_i^n = \mu$ and $\sum_i \nu_i^n = \nu$).

It is enough to check that γ_T gives the same mass as γ to every $C_{i,j}^n$ to conclude by Lemma 2.5.6, but it is easy to prove. Indeed, this mass equals that of $\gamma_{T_i^n}$ and $\gamma_{T_i^n}(C_{i,j}^n) = \mu_i^n(\{x : x \in K_i^n, T_i^n(x) \in K_j^n\}) = \mu_i^n(\{x : T_i^n(x) \in K_j^n\}) = \nu_i^n(K_j^n) = \gamma(K_i^n \times K_j^n)$. ■

Therefore, the relaxation result is just a consequence.

Theorem 2.5.8

On a compact subset $\Omega \subset \mathbb{R}^d$ and $c : \Omega\times\Omega \to \mathbb{R}$ is continuous, K is the relaxation of J. In particular, $\inf J = \min K$, and hence Monge and Kantorovich problems have the same infimum.

Proof. Notice that K is continuous borrowing our argument from Theorem 2.1.10 and thus l.s.c. By the definition of J itself, $K \leq J$. Hence, $K \leq \bar{J}$. To prove $K = \bar{J}$, for any γ, Theorem 2.5.8 ensures existence of a sequence of transport plans T_n such that $\gamma_{T_n} \rightharpoonup \gamma$. Therefore, $J(T_n) = K(\gamma_{T_n}) \to K(\gamma)$ by the characterization in Proposition 2.5.2. Therefore, $\bar{J} = K$.

Moreover, $\min K$ is realized at some γ and since K is the relaxation of J, we have $\min K = \inf J$. ■

§2.6 Stability of Solutions

To close this chapter, we offer some simple stability results that we only consider in the compact domain case. For more general case, it is out of the scope of this note (and also [San15]). Beforehand, we shall mention results of Hausdorff convergence in a compact metric space. Note that these facts will be needed only for few passages in our first stability result.

Definition 2.6.1 (Hausdorff Distance)

Let X be a compact metric space. We define the **Hausdorff distance** on pair of compact subsets of X by setting

$$d_H(A, B) := \max\{\max d(x, A) : x \in B, \max\{d(x, B) : x \in A\}\}.$$

This metric is equivalent to the other two metrics:

1. $d_H(A, B) = \max\{|d(x, A) - d(x, B)| : x \in X\}$,
2. $d_H(A, B) = \inf\{\epsilon > 0 : A \subset B_\epsilon, B \subset A_\epsilon\}$, where A_ϵ and B_ϵ stand for the ϵ-neighborhodd of A and B, respectively. That is, the "enlargement" of A and B with "scale" $1 + \epsilon$.

Theorem 2.6.2 (Blaschke)

The function d_H as defined in Definition 2.6.1 is indeed a metric: it is positive and symmetric, it only vanishes if the two sets coincide, and it satisfies the triangle inequality. With this metric, the set of compact subsets of X becomes a compact metric space itself.

Proof. See [AT03]. ■

Proposition 2.6.3

If $d_H(A_n, A) \to 0$ and μ_n is a sequence of positive measures such that $\text{supp}(\mu_n) \subset A_n$ with $\mu_n \rightharpoonup \mu$, then $\text{supp}(\mu) \subset A$.

Proof. For any $n \in \mathbb{N}$, the function $X \ni x \mapsto d_H(x, A_n)$ is uniformly continuous and bounded. Moreover, by the fact μ_n is concentrated in A_n we have

$$\int_X d(x, A_n)\, d\mu_n(x) = 0.$$

For any $x \in X$,

$$|d(x, A_n) - d(x, A)| < Cd_H(A_n, A)$$

where $c > 0$ is a constant since we use the equivalences between d_Hs in Definition 2.6.1. Therefore, $d(\cdot, A_n) \to d(\cdot, A)$ uniformly. Let $\epsilon > 0$, for large enough n we have $|d(x, A_n) - d(x, A)| < \epsilon$ and by the fact $\mu_n \rightharpoonup \mu$, we have

$$\begin{aligned}
&\left| \int_X d(x, A_n)\, d\mu_n(x) - \int_X d(x, A)\, d\mu(x) \right| \\
&\le \left| \int_X d(x, A_n)\, d\mu_n(x) - \int_X d(x, A)\, d\mu_n(x) \right| + \left| \int_X d(x, A)\, d\mu_n(x) - \int_X d(x, A)\, d\mu(x) \right| \\
&\le \int_X |d(x, A_n) - d(x, A)|\, d\mu_n(x) + \left| \int_X d(x, A)\, d\mu_n(x) - \int_X d(x, A)\, d\mu(x) \right| \\
&\le \epsilon + \left| \int_X d(x, A)\, d\mu_n(x) - \int_X d(x, A)\, d\mu(x) \right| \to \epsilon + 0.
\end{aligned}$$

Hence, $0 = \int_X d(x, A_n)\, d\mu_n(x) \to \int_X d(x, A)\, d\mu(x)$. Therefore, $\int_X d(x, A)\, d\mu(x) = 0$ which implies $\operatorname{supp}(\mu) \subset A$. ∎

With this result we can work on the stability of solution as follows.

Theorem 2.6.4 (Stability of Optimal Transport Plans)
Suppose that X and Y are compact metric spaces and $c : X \times Y \to \mathbb{R}$ is continuous. Suppose that $\gamma_n \in \mathcal{P}(X \times Y)$ is a sequence of transport plan which are optimal for the cost c between their own marginals $\mu_n := (\pi_X)_\# \gamma_n$ and $\nu_n := (\pi_Y)_\# \gamma_n$, and suppose $\gamma_n \rightharpoonup \gamma$; then $\mu_n \rightharpoonup \mu := (\pi_X)_\# \gamma$, $\nu_n \rightharpoonup \nu := (\pi_Y)_\# \gamma$, and γ is optimal in the transport between μ and ν.

Proof. Let $\Gamma_n := \operatorname{supp}(\gamma_n)$. Then, up to subsequence $\Gamma_n \to \Gamma$ by Theorem 2.6.2. By Theorem 2.2.10, each Γ_n is c-CM. By Hausdorff convergence, for any finite family (x_1, y_1), $\ldots$, (x_k, y_k), there exists (x_1^n, y_1^n), $\ldots$, $(x_k^n, y_k^n) \in \Gamma_n$ such that, for each $i = 1, \ldots, k$, we have $(x_i^n, y_i^n) \to (x_i, y_i)$. Hence, by the fact Γ_n is c-CM we have

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

and since c is continuous we can pass $n \to \infty$ to get

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

Therefore, Γ is also c-CM. By Proposition 2.6.3, we have $\operatorname{supp}(\gamma) \subset \Gamma$ and by Theorem 2.3.16 we concluded the optimality of γ. ∎

As as result of the Theorem 2.6.4, we can have stability between the masses as well. For simplicity of notation, for a given cost $c : X \times Y \to \mathbb{R}$ and $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, we define

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

Theorem 2.6.5 (Stability of Optimal Transport Masses)
Suppose that X and Y are compact metric spaces and that $c : X \times Y \to \mathbb{R}$ is continuous. Suppose that $\mu_n \in \mathcal{P}(X)$ and $\nu_n \in \mathcal{P}(Y)$ are two sequences of probability measures, with $\mu_n \rightharpoonup \mu$ and $\nu_n \rightharpoonup \nu$. Then we have $\mathcal{J}_c(\mu_n, \nu_n) \to \mathcal{J}_c(\mu, \nu)$.

Proof. By Theorem 2.1.10, there exsits γ_n optimal from μ_n to ν_n. Since $X \times Y$ is compact, $\mathcal{P}(X \times Y)$ is necessarily compact. Thus, up to subsequence, $\gamma_n \rightharpoonup \gamma \in \mathcal{P}(X \times Y)$. In particular, $(\pi_X)_\# \gamma = \mu$ and $(\pi_Y)_\# \gamma = \nu$. Hence, Theorem 2.6.6 implies γ is optimal for cost c. Since c continuous, from weak convergence we have

$$\mathcal{J}_c(\mu_n, \nu_n) = \int_{X\times Y} c(x, y)\, d\gamma_n(x, y) \to \int_{X\times Y} c(x, y)\, d\gamma(x, y) = \mathcal{J}_c(\mu, \nu).$$

■

Finally, we conclude the result of stability within the Kantorovich potentials.

Theorem 2.6.6 (Stability of Kantorovich Potentials)
Suppose that X and Y are compact metric spaces and that $c : X \times Y \to \mathbb{R}$ is continuous. Suppose that $\mu_n \in \mathcal{P}(X)$ and $\nu_n \in \mathcal{P}(Y)$ are two sequences of probability measures, with $\mu_n \rightharpoonup \mu$ and $\nu_n \rightharpoonup \nu$. Let (μ_n, ψ_n) be a pair of c-concave Kantorovich potentials for the cost c in the transport from μ_n to ν_n for any $n \in \mathbb{N}$. Then, up to subsequences, we have $\varphi_n \to \varphi$, $\psi_n \to \psi$, where the convergence is uniform and (φ, ψ) is a pair of Kantorovich potentials for μ and ν.

Proof. Since $X \times Y$ is compact, c is uniformly bounded and continuous and thus any c-concave and $\bar{c}$-concave functions are bounded and have the same modulus of continuity as c. By the same argument as what we did in Theorem 2.2.6, up to translating by a constant, we can apply Ascoli-Arzelà (c.f. Theorem 2.2.3) to get $\varphi_n \to \tilde{\varphi}$ and $\psi_n \to \tilde{\psi}$ up to subsequence uniformly. We also have $\tilde{\varphi} \oplus \tilde{\psi} \leq c$ which implies the pair $(\tilde{varphi}, \tilde{psi})$ is admissible to (DP). Moreover, we have

$$\int_X \varphi_n(x)\, d\mu_n(x) \to \int_X \tilde{\varphi}(x)\, d\mu(x) \quad \text{and} \quad \int_Y \psi_n(y)\, d\nu_n(y) \to \int_Y \tilde{\psi}(y)\, d\nu(y)$$

by applying same estimate method as what we did in the proof of Proposition 2.6.3. Therefore, by duality result in Theorem 2.2.11 we have

$$\mathcal{J}_c(\mu_n, \nu_n) = \int_X \varphi_n(x)\, d\mu_n(x) + \int_Y \psi_n(y) \to \int_X \tilde{\varphi}(x)\, d\mu(x) + \int_Y \tilde{\psi}(y)\, d\nu(y).$$

However, Theorem 2.6.5 gives $\mathcal{J}_c(\mu_n, \nu_n) \to \mathcal{J}_c(\mu, \nu)$. Therefore $(\tilde{\varphi}, \tilde{\psi})$ is a pair of Kantorovich potentials for μ and ν.

Note that the convergence $\mathcal{J}_c(\mu_n, \nu_n) \to \int_X \tilde{\varphi}(x)\, d\mu(x) + \int_Y \tilde{\psi}(y)\, d\nu(y)$ is true in the full sequence if we have uniqueness of the Kantorovich potentials at the limit. ■

3 Solving the Supremal Case, L^∞

We dedicate this chapter to consider a case for which the approach given in Chapter 2 is not possible. Instead of minimizing

$$K_p(\gamma) = \int |x-y|^p \, d\gamma$$

we want to minimize the maximal displacement, i.e., its L^∞ norm. This problem has been first adressed in [CDPJ08] and we will present a proof taken from Section 3.2 of [San15] which in essence is highly geometrical and measure theoretic.

Let us define

$$\begin{aligned} K_\infty(\gamma) := \|x-y\|_{L^\infty(\gamma)} &= \inf\{m \in \mathbb{R} : |x-y| \le m \text{ for } \gamma\text{-a.e. } (x,y)\} \\ &= \max\{|x-y| : (x,y) \in \operatorname{supp}(\gamma)\}. \end{aligned}$$

where the last equality between L^∞ and a maximum on the support is justified by the continuity of the function $|x-y|$.

In this chapter, we limit our domain $\Omega \subset \mathbb{R}^d$ to be compact. We start our consideration problem by recalling the general result of $\|f\|_{L^p(\gamma)} \to \|f\|_{L^\infty(\gamma)}$ to imply the existence of a minimizer.

Lemma 3.0.1

Let $\Omega \subset \mathbb{R}^d$ be a compact domain. For every $\gamma \in \mathcal{P}(\Omega\times\Omega)$, we have $K_p(\gamma)^{\frac{1}{p}} \uparrow K_\infty(\gamma)$. In particular, $K_\infty(\gamma) = \sup_{p\ge 1} K_p(\gamma)^{\frac{1}{p}}$ and K_∞ is l.s.c. for the weak convergence in $\mathcal{P}(\Omega\times\Omega)$. Thus, it admits a minimizer over $\Pi(\mu,\nu)$, which is compact.

Proof. Let $c : \Omega\times\Omega \to [0,+\infty]$ defined as $c(x,y) = |x-y|$. For non-triviality, we assume $K_\infty(\gamma) > 0$ since the case equal 0 implies $K_p(\gamma) = 0$ and it is trivial.

Since Ω is bounded, c is bounded and thus $c \in L^p$ for $1 \le p \le +\infty$. We shall prove $(K_p(\gamma))^{\frac{1}{p}} \uparrow K_\infty(\gamma)$.

Notice that for any $p \in [1,+\infty]$, $K_p(\gamma)^{\frac{1}{p}} \le K_\infty(\gamma) < \infty$. Then, by Hölder Inequality, for any $q > p$ we have

$$\int c^p d\gamma \le \left(\int c^q \, d\gamma\right)^{\frac{p}{q}} \left(\int 1^{\frac{q}{q-p}} \, d\gamma\right)^{1-\frac{p}{q}} = \left(\int c^q \, d\gamma\right)^{\frac{p}{q}} \implies K_p(\gamma)^{\frac{1}{p}} \le K_q(\gamma)^{\frac{1}{q}}.$$

This implies $K_p(\gamma)^{\frac{1}{p}}$converges increasingly to a number $M < +\infty$. Let $L = K_\infty(\gamma) = \|c\|_{L^\infty(\gamma)} > 0$ from our initial assumption. For any small $\epsilon > 0$, there exists a set A with $\gamma(A) > 0$ such that $c(x,y) > L - \epsilon$. Therefore, for any $p \in [1,+\infty)$ we have

$$(L-\epsilon)^p \cdot \gamma(A) = \int_A (L-\epsilon)^p \, d\gamma < \int c^p \, d\gamma = K_p(\gamma) \implies (L-\epsilon)\cdot\gamma(A)^{\frac{1}{p}} < K_p(\gamma)^{\frac{1}{p}}.$$

Taking $p \to +\infty$ implies $L - \epsilon \le M$ and thus $L = M$ and we have proved $K_p(\gamma)^{\frac{1}{p}} \uparrow K_\infty(\gamma)$.

Since c is continuous, we can use the argument of Theorem 2.1.10 to say $K_p^{\frac{1}{p}}$ is continuous and thus K_∞ is l.s.c. Therefore, Theorem 2.1.2 ensures the existence of a minimizer over $\Pi(\mu, \nu)$. ■

Now, we shall analyze the set of minimizers of K_∞, denoted as

$$O_\infty(\mu, \nu) = \text{argmin}_{\gamma \in \Pi(\mu,\nu)} K(\gamma)$$

which is a set. Then, we will prove that we can take a particular $\bar(\gamma) \in O_\infty(\mu, \nu)$ that can be induced by a transport map.

Let $L = \min\limits_{\gamma \in \Pi(\mu,\nu)} K_\infty(\gamma)$. Then we can deduce

$$\gamma \in O_\infty(\mu, \nu) \iff \text{supp}(\gamma) \subset \{(x, y) : |x - y| \le L\}.$$

Note that, if $L = 0$, $\text{supp}(\gamma)$ is concentrated on the graph of the identity function and therefore $\gamma = (\text{id}, \text{id})_\mu$ which also implies $\mu = \nu$. Hence, throughout this chapter we shall assume $L > 0$ and $\mu \neq \nu$.

Note that, by the lower semicontinuity of K_∞, we have $O_\infty(\mu, \nu)$ is compact. Thus, we can consider the secondary problem

$$\min\{K_2(\gamma) : \gamma \in O_\infty(\mu, \nu)\}$$

which admits a minimizer $\bar{\gamma} \in O_\infty(\mu, \nu)$ by Theorem 2.1.2 and we will prove $\bar{\gamma}$ is induced by a map T which would solve

$$\min\{\|T - \text{id}\|_{L^\infty(\mu)} : T_\# \mu = \nu\}.$$

Since $\bar{\gamma}$ also enjoys the property $\text{supp}(\gamma) \subset \{(x, y) : |x - y| \le L\}$, we can say that $\bar{\gamma}$ also solves

$$\min\left\{\int_{\Omega \times \Omega} c(x, y)\, d\gamma(x, y) : \gamma \in \Pi(\mu, \nu)\right\}$$

where $c : \Omega \to [0, +\infty]$ is defined as

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

where c is l.s.c. and bounded below. Thus, with Theorem 2.2.14, we have the fact that $\bar{\gamma}$ is concentrated on a c-CM set $\Gamma \subset \Omega \times \Omega$. We can intersect Γ with $\text{supp}(\gamma)$ to have a c-CM set $\Gamma \subset \{(x, y) : |x - y| \le L\}$. In particular, by c-CM, Γ has the property

$$(\forall (x_1, y_1), (x_2, y_2) \in \Gamma) \quad |x_1 - y_2|, |x_2 - y_1| \le L \implies \langle x_1 - x_2, y_1 - y_2 \rangle \ge 0. \tag{3.1}$$

We should recall the notion of Lebesgue points as follows.

Definition 3.0.2 (Lebesgue Points)
Let $E \subset \mathbb{R}^d$ be measurable, we define the **set of Lebesgue points** of E, denoted by $\text{Leb}(E)$, as the set of $x \in \mathbb{R}^d$ such that

$$1 = \lim_{r \to 0^+} \frac{|E \cap B(x, r)|}{|B(x, r)|}.$$

Notice that by the Lebesgue-Besicovich Differentiation Theorem (c.f. Theorem 1 in Section 1.7 of [EG92]), we have

$$\fint_{B(x,r)} \chi_E(y)\, dy = \chi_E(x)$$

for a.e. x. Therefore, $|E \backslash \operatorname{Leb}(E)| + |\operatorname{Leb}(E) \backslash E| = 0$. Moreover, if $\operatorname{Leb}(E)$ is non-empty, we necessarily have $|E| > 0$ since $|E| = 0$ will contradict the definition of $x \in \operatorname{Leb}(E)$.

Going back to our next (and last) refinement, we shall neglect non-Lebesgue points in Γ and canonically we will add the assumption $\mu \ll \mathcal{L}^d$ which also implies we will have maps T that transport μ to ν by Corollary 2.5.5.

Lemma 3.0.3

The plan $\bar{\gamma}$ is concentrated on a c-CM set $\bar{\Gamma}$ such that for any pair $(x_0, y_0) \in \bar{\Gamma}$, any errors $\epsilon, \delta > 0$, any unit vector ξ (as a direction), and every sufficiently small $r > 0$, there are points

$$x \in (B(x_0, r) \backslash B(x_0, r/2)) \cap C(x_0, \xi, \delta),\;\; y \in B(y_0, \epsilon)$$

such that $(x, y) \in \bar{\Gamma}$, where $C(x_0, \xi, \delta)$ is the convex cone

$$C(x_0, \xi, \delta) := \left\{ x : \left\langle \frac{x - x_0}{|x - x_0|}, \xi \right\rangle > 1 - \delta \right\}.$$

Proof. The visualization of the property of the desired $(x, y) \in \bar{\Gamma}$ as follows.

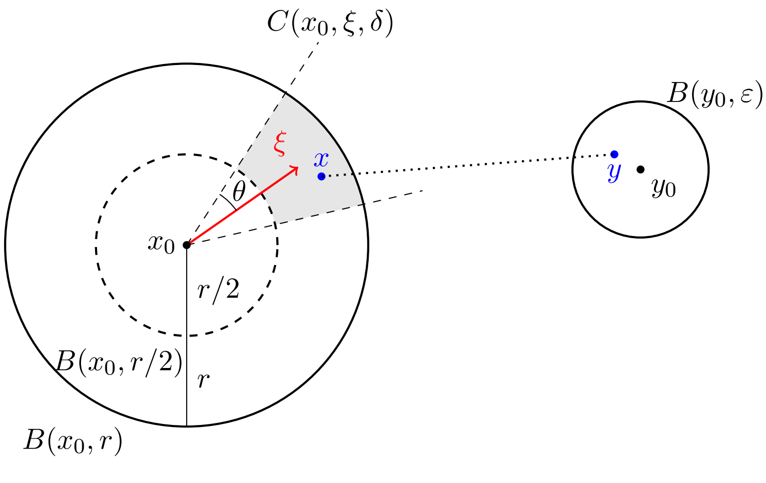

Figure 3.1: Visualization of the property of chosen $(x, y) \in \bar{\Gamma}$.

Let $\{B_i\}$ be a countable topological basis of $\mathbb{R}^d$ (we can take it as collection of balls of center in $\mathbb{Q}^d$ and of rational length). For all $i \in \mathbb{N}$, define

$$A_i := \pi_X(\Gamma \cap (\Omega \times B_i)) \subset \Omega$$

that is the set of all $x \in \Omega$ such that there exists $y \in B_i$ such that $(x, y) \in \Gamma$. Furthermore, set $N_i = A_i \backslash (\operatorname{Leb}(A_i))$ and we have $|N_i| = 0$ which implies N_i is μ-neglectible. Moreover,

$\mu\left(\bigcup N_i\right) = 0$. Hence, we can now consider

$$\bar{\Gamma} = \Gamma \backslash \left(\left(\bigcup N_i\right) \times \Omega\right)$$

that is the set of all $(x, y) \in \Gamma$ such that x is a Lebesgue point of some A_i, and $\mu(\bar{\Gamma}) = 1$. We shall prove $\bar{\Gamma}$ satisfies the property as suggested by Figure 3.1.

Let $(x_0, y_0) \in \bar{\Gamma}$, ϵ, $\delta > 0$, and any unit vector ξ. Using the back $\{B_i\}$ is a basis, there exists some $i \in \mathbb{N}$ such that $y_0 \in B_i \subset B(y_0, \epsilon)$. Therefore, $x_0 \in A_i$, and by the definition of $\bar{\Gamma}$, $x_0 \in \text{Leb}(A_i)$, and as an implication $|A_i| > 0$.

For all $r > 0$ define $S_r := (B(x_0, r) \backslash B(x_0, r/2)) \cap C(x_0, \xi, \delta)$. Note that by the fact S_r is a sector of $B(x_0, r)$ and the volume scale appropriately with the volume of $B(x_0, r)$, we have $|S_r| = C \cdot |B(x, r)|$ where $0 < C < 1$ and $C = C(\delta)$.

We claim there exists $r_0 > 0$ such that for any $0 < r < r_0$, $A_i \cap S_r \neq \varnothing$. Suppose on the contrary that we have a sequence $r_n \to 0^+$ such that $A_i \cap S_r = \varnothing$, then $A_i \cap B(x_0, r) \subset B(x_0, r) \backslash S_r$ where we will have

$$\begin{aligned} &|A_i \cap B(x_0, r_n)| \leq |B(x_0, r_n) \backslash S_{r_n}| = |B(x_0, r_n)| - |S_{r_n}| = (1 - C)|B(x_0, r_n)| \\ &\implies \frac{|A_i \cap B(x_0, r_n)|}{|B(x_0, r_n)|} \leq 1 - C \implies \limsup_{n \to \infty} \frac{|A_i \cap B(x_0, r_n)|}{|B(x_0, r_n)|} \leq 1 - C < 1 \end{aligned}$$

which contradicts the fact that $x_0 \in \text{Leb}(A_i)$. Hence, our claim is true.

In particular, we can also deduce $(A_i \backslash N_i) \cap S_r \neq \varnothing$ for small enough $r > 0$. Notice that $|A_i \cap S_r| > 0$ since

$$\begin{aligned} &|A_i \cap S_r| = |A_i \cap B(x_0, r)| - |A_i \cap (B(x_0, r) \backslash S_r)| \geq |A_i \cap B(x_0, r)| - (1 - C)|B(x_0, r)| \\ &\implies \frac{|A_i \cap S_r|}{|B(x_0, r)|} \geq \frac{|A_i \cap B(x_0, r)|}{|B(x_0, r)|} - 1 + C \implies \liminf_{r \to 0+} \frac{|A_i \cap S_r|}{|B(x_0, r)|} \geq 1 - 1 + C = C > 0. \end{aligned}$$

Then, by the fact $|(A_i \backslash N_i) \cap S_r| = |A_i \cap S_r| > 0$, we obtained $\text{Leb}(A_i) \cap S_r \neq \varnothing$ and we can take $x \in (A_i \backslash N_i) \cap S_r \subset A_i$. Due to the definition of A_i, we also have $y \in B_i \subset B(y_0, \epsilon)$ such that $(x, y) \in \bar{\Gamma}$. ■

Now, with this property we can assert the "concentration" of $\bar{\Gamma}$ with this last lemma.

Lemma 3.0.4
For any (x_0, y_0), $(x_0, z_0) \in \bar{\Gamma}$ we have $y_0 = z_0$.

Proof. Suppose $y_0 \neq z_0$. Without loss of generality, $|x_0 - z_0| \leq |x_0 - y_0|$. We can assume $x_0 \neq y_0$ since equality will imply $y_0 = z_0$.

We shall take a pair $(x, y) \in \bar{\Gamma}$ that from an appropriate direction of unit vector ξ that satisfies the property of Lemma 3.0.3 with respect to the point (x_0, y_0) such that $|x - z_0|$, $|x_0 - y| \leq L$ as suggested by Figure 3.2. The reason we would like to consider this result is by property of $\bar{\Gamma}$ that satisfies 3.1 (since $\bar{\Gamma} \subset \Gamma$) to deduce

$$\langle x - x_0, y - z_0 \rangle \geq 0 \tag{3.2}$$

and we can use this to arrive at a contradiction, by means of a suitable choice of ξ. Indeed, the direction of $x - x_0$ is almost that of ξ (up to an error of $O(\sqrt{\delta})$) and the direction of $y - z_0$ is almost that of $\zeta := \dfrac{y_0 - z_0}{|y_0 - z_0|}$ (up to an error $O(\epsilon)$). If we choose ξ such that $\langle \xi, y_0 - z_0 \rangle < 0$, this means that for small enough δ and ϵ, we would get a contradiction which is what Figure 3.2 means to us.

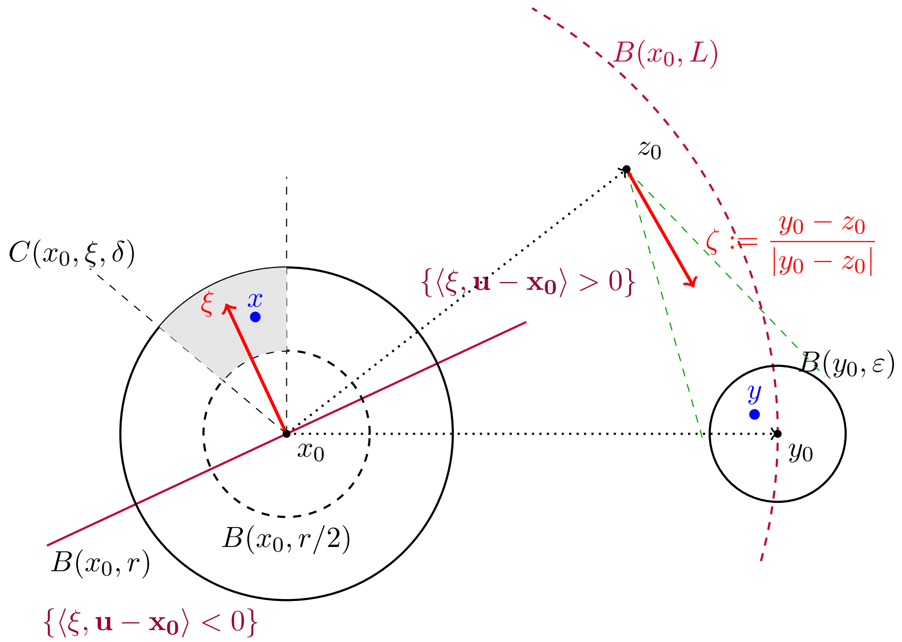

Figure 3.2: Visualization of the taken $(x, y) \in \Gamma$ from direction ξ.

We will create a pathway to a formal proof from our idea in the previous paragraphs. We first determine the precise error term of inner product with respect to ϵ and δ as sharp as we can.

- For the direction ζ, since $y \in B(y_0, \epsilon)$, we can write $y = y_0 + e_\epsilon$, where $|e_\epsilon| \leq \epsilon$. Substituting this into our vector:

$$y - z_0 = (y_0 + e_\epsilon) - z_0 = (y_0 - z_0) + e_\epsilon.$$

Since $y_0 - z_0 = |y_0 - z_0|\zeta$, we get $y - z_0 = |y_0 - z_0|\zeta + e_\epsilon$. Whenever we take an inner product of any element v with respect to e_ϵ, we have

$$|\langle e_\epsilon, v \rangle| \leq |v|\epsilon. \tag{3.3}$$

Therefore, $\langle e_\epsilon, v \rangle = O(\epsilon)$.

- For the direction ξ, we can write $v := x - x_0$ and so $v = |v|\xi + e_\delta$. Whenever we take an inner product of e_δ with any w, we have

$$\langle e_\delta, w \rangle = \langle v, w \rangle - |v|(\langle \xi, w \rangle) = \langle (v - |v|\xi), w \rangle .$$

By Cauchy-Schwarz inequality, we have

$$|\langle (v - |v|\xi), w \rangle| \leq |v - |v|\xi||w|.$$

We shall compute $|v - |v|\xi|$ by expanding the square norm

$$|v - |v|\xi|^2 = 2|v|^2 - 2|v| \langle v, \xi \rangle \tag{3.4}$$

Now, we shall finally use the definition of our cone $C(x_0, \xi, \delta)$, which states $\langle v, \xi \rangle > (1-\delta)|v|$ and thus $|v| - \langle v, \xi \rangle < \delta |v|$. We can substitute this upper bound into (3.4) to get

$$|v - |v|\xi|^2 < 2|v|(\delta|v|) = 2\delta|v|^2$$

Taking the square root of both sides and plugging to the upper bound of $|\langle e_\delta, w \rangle|$ gives

$$|\langle e_\delta, w \rangle| \le |w| r \sqrt{2\delta}. \tag{3.5}$$

Therefore, $\langle e_\delta, w \rangle = O(\sqrt{\delta})$ for fixed r.

Let us first go back to our idea of contradiction. First, we have (3.2). However, if we choose ξ such ξ such that $\langle \xi, y_0 - z_0 \rangle < 0 \iff \langle \xi, \zeta \rangle < 0$, we have

$$\begin{aligned}\langle x - x_0, y - z_0 \rangle &= \langle \xi |x - x_0| + e_\delta, |y_0 - z_0|\zeta + e_\epsilon \rangle \\ &= |x - x_0||y_0 - z_0| \langle \xi, \zeta \rangle + |y_0 - z_0| \langle e_\delta, \zeta \rangle + |x - x_0| \langle \xi, e_\epsilon \rangle + \langle e_\delta, e_\epsilon \rangle\end{aligned}$$

By letting $\alpha = |y_0 - z_0| \langle \xi, \zeta \rangle < 0$, from approximations (3.3) and (3.5) and the fact $|x - x_0| \le r$, we can take an upper bound

$$\begin{aligned}\langle x - x_0, y - z_0 \rangle &\le |x - x_0|\alpha + |y_0 - z_0| \cdot |x - x_0|\sqrt{2\delta} + |x - x_0| \cdot \epsilon + |x - x_0|\epsilon\sqrt{2\delta} \\ &\le r(\alpha + |y_0 - z_0|\sqrt{2\delta} + \epsilon + \epsilon\sqrt{2\delta}).\end{aligned}$$

We can take ϵ and δ small enough such that the term $\alpha + |y_0 - z_0|\sqrt{2\delta} + \epsilon + \epsilon\sqrt{2\delta}$ is still negative. Hence, $\langle x - x_0, y - z_0 \rangle < 0$ which contradicts the fact that $\langle x - x_0, y - z_0 \rangle \ge 0$ if our assumption is satisfied.

We only need to guarantee $|x - z_0|, |x_0 - y| \le L$ to settle our proof. As a precaution to the reader, we will do a seemingly intimidating expansions to later use approximations (3.5) and (3.3).

First, we compute $|x_0 - y|$:

$$|x_0 - y|^2 = |x_0 - x|^2 + |x - y|^2 + 2\langle x_0 - x, x - y \rangle .$$

Within this sum, we have

$$\begin{aligned}&|x_0 - x|^2 \le r^2, \\ &|x - y|^2 \le L^2, \\ &2\langle x_0 - x, x - y \rangle = 2\langle -|x - x_0|\xi - e_\delta, (x_0 - y_0) + (x - x_0) - (y - y_0) \rangle .\end{aligned}$$

Analogously, for what concerns $|x - z_0|$, we have

$$|x - z_0|^2 = |x_0 - z_0|^2 + |x - x_0|^2 + 2\langle x - x_0, x_0 - z_0 \rangle ,$$

and the three terms satisfy

$$\begin{aligned}&|x_0 - z_0|^2 \le L^2, \\ &|x - x_0| \le r^2, \\ &2\langle x - x_0, x_0 - z_0 \rangle = 2(\langle |x - x_0|\xi, x_0 - z_0 \rangle + \langle e_\delta, x_0 - z_0 \rangle).\end{aligned}$$

We shall consider three cases that depends on $|x_0 - y_0|$ and $|x_0 - z_0|$, starting from the easiest to the hardest case to assert what we need to guarantee and then the proof is done.

Case 1. $|x_0 - y_0| < L$ (which implies $|x_0 - z_0| < L$ from assumption of y_0 and z_0).

We have from approximations (3.3) and (3.5) that

$$\begin{aligned}|x - x_0|^2 + 2\langle x - x_0, x_0 - z_0\rangle &< r^2 + 2(r|\xi||x_0 - z_0| + |x_0 - z_0|r\sqrt{2\delta}) \\ &= L^2 + r^2 + 2|x_0 - z_0|(1 + \sqrt{2\delta})r.\end{aligned}$$

and actually can fix any δ and set r small enough such that $r^2 + 2|x_0 - z_0|(1 + \sqrt{2\delta})r < L^2 - |x_0 - z_0|$ to have $|x - z_0| < L$.

On the other hand, we can choose r (can be smaller than previously chosen r) and ϵ small enough (e.g. such that $r + \epsilon < L - |x_0 - y_0|$), to obtain $|x - y| < L$ and thus we can simply use Cauchy-Schwarz in the two other terms to get

$$|x_0 - x|^2 + 2\langle x_0 - x, x - y\rangle < r^2 + 2\langle x_0 - x, x - y\rangle \leq r^2 + 2Lr$$

We can choose a possibly even smaller r such that $r^2 + 2Lr < L^2 - |x - y|^2$ to have $|x_0 - y| < L$.

Thus, we can have our contradiction by taking ξ such that $\langle \xi, y_0 - z_0\rangle < 0$ (e.g. $\xi = -\zeta$) after all of our determined r, ϵ, δ.

Case 2. $|x_0 - y_0| = L$ and $|x_0 - z_0| < L$.

Notice that we can choose r dan δ small enough as in Case 1 to have $|x - z_0| < L$. First, we shall take any ξ such that $\langle \xi, y_0 - x_0\rangle < 0$ instead (this can be adjusted later on) We shall consider on first two subcases for the case of $|x_0 - y|$ for our choice of ξ.

- We cannot have any direction ξ that also satisfies $\langle \xi, x_0 - z_0\rangle < 0$. This only occurs when $y_0 \in (x_0, z_0)$, i.e., z_0 is in the segment connecting x_0 and z_0. We abort our choice of ξ such that. Hence, we can immediately have any choise of ξ that immediately imply $\langle \xi, y_0 - z_0\rangle < 0$. One can take $\xi = -\dfrac{y_0 - x_0}{|y_0 - x_0|}$ for example.
- We can have a direction ξ that satisfy $\langle \xi, x_0 - z_0\rangle < 0$. With this choice of ξ, we have $\langle \xi, y_0 - z_0\rangle < 0$.

Regardless from the choice of ξ, we are ensured we can have ξ that satisfies $\langle \xi, y_0 - z_0\rangle < 0$ and $\langle \xi, y_0 - x_0\rangle < 0$.

Fix $\alpha = \langle \xi, x_0 - y_0\rangle > 0$. We shall consider the term $2\langle x_0 - x, x - y\rangle$ from $|x_0 - y|^2$. We expand $2\langle x_0 - x, x - y\rangle$ and use the fact $|x - x_0| \geq \dfrac{r}{2}$ to have an upper bound by (3.5) and Cauchy-Schwarz in-between

$$\begin{aligned}2\langle x_0 - x, x - y\rangle = -2|x - x_0|\langle \xi, x_0 - y_0\rangle - 2|x - x_0|\langle \xi, (x - x_0) - (y - y_0)\rangle \\ -2\langle e_\delta, (x_0 - y_0) + (x - x_0) - (y - y_0)\rangle \\ \leq -2\alpha r + 2r(r + \epsilon) + 2r(|x_0 - y_0| + r + \epsilon)\sqrt{2\delta}.\end{aligned}$$

Therefore,

$$|x_0 - x|^2 + 2\langle x_0 - x, x - y\rangle \leq (3 + 2\sqrt{2\delta})r^2 + (2\epsilon(1 + \sqrt{2\delta}) + 2|x_0 - y_0|\sqrt{2\delta} - \alpha)r$$

We can take ϵ and δ small enough such that the coefficient of r is negative. Hence, for small enough r, we have $|x_0 - x|^2 + 2\langle x_0 - x, x - y\rangle < 0$ and thus $|x_0 - y| < L$.

In conclusion, we can take an appropriate ξ first, and then taking r, δ, ϵ small enough to have our contradiction of this case.

Case 3. $|x_0 - y_0| = |x_0 - z_0| = L$.

Since we assumed $y_0 \neq z_0$, the three points cannot be colinear. Therefore, there exists ξ such that $\langle \xi, y_0 - x_0 \rangle < 0$ and $\langle \xi, x_0 - z_0 \rangle < 0$. With a small enough choice of r, δ, ϵ as in Case 2, we have $|x_0 - y| < L$.

Now, we shall check for $|x - z_0|^2$ instead. Let $\beta = -\langle \xi, x_0 - z_0 \rangle > 0$. Using the fact $|x - x_0| \geq \frac{r}{2}$ and approximate (3.5), we have

$$|x - x_0|^2 + 2\langle x - x_0, x_0 - z_0 \rangle \leq r^2 - r\beta + 2r|x_0 - z_0|\sqrt{2\delta} = r^2 + (2|x_0 - z_0|\sqrt{2\delta} - \beta)r.$$

As we did beforehand, we can possible choose even smaller δ such that $2|x_0 - z_0|\sqrt{2\delta} < \beta$ to have negative coefficient on the term r and an even smaller r to finally ensured of $|x - x_0|^2 + 2\langle x - x_0, x_0 - z_0 \rangle < 0$. Therefore, $|x - z_0| < L$ which implies the same contradiction as the other two cases.

We conclude we always have a contradiction regarding the property of $\bar{\Gamma}$ to the assumption $y_0 \neq z_0$. Hence, $y_0 = z_0$. ■

With Lemma 3.0.4, it is now easy to conclude this chapter.

Theorem 3.0.5

The secondary variational problem

$$\min\{K_2(\gamma) : \gamma \in O_\infty(\mu, \nu)\}$$

admits a unique solution $\bar{\gamma}$, it is induced by a transport map T, and such a map is an optimal transport for the problem

$$\min\{\|T - \mathrm{id}\|_{T^\infty(\mu)} : T_\# \mu = \nu\}.$$

Proof. From our considerations until Lemma 3.0.4, the optimal $\bar{\gamma}$ is concentrated on a set $\bar{\Gamma}$ such that $\bar{\Gamma}$ is contained in a graph since for any $x_0 \in \pi_X(\bar{\Gamma})$, there is no more than one possible point y_0 such that $(x_0, y_0) \in \bar{\Gamma}$. Let us consider such a point y_0 as $T(x_0)$. Therefore, $\bar{\gamma} = \gamma_T$. Note that this choice of T that is defined on $\pi_X(\bar{\Gamma})$ is measurable since $\mathrm{Graph}(T) = \bar{\Gamma}$ is measurable from our construction.

To prove uniqueness, we first mark the fact that any solution γ to the secondary problem is then generated by a map T. Suppose there exists two solutions γ_{T_1} and γ_{T_2}. Consider another optimal plan $\gamma' = \frac{\gamma_{T_1} + \gamma_{T_2}}{2}$. From our fact, $\gamma' = \gamma_{T'}$ for some T'. Let $\Gamma_{T'} = \mathrm{graph}(T')$, we have $\gamma'(\Gamma_{T'})$ and thus $\gamma_{T_1}(\Gamma_{T'}) = \gamma_{T_2}(\Gamma_{T'}) = 1$. By this fact, $T_1 = T'$ and $T_2 = T'$ μ-a.e. since γ_{T_1} and γ_{T_2} is concentrated in a graph as well and we just proved uniqueness. ■

Remark 3.0.6. We can also apply our more general approach in proving uniqueness to the proof of Theorem 2.3.3 following the same argument we used in this theorem.

Bibliography

[AB06] C. D. Aliprantis and K. C. Border, *Infinite dimensional analysis: A hitchhiker's guide*, 3rd ed., Springer Science & Business Media, Berlin; London, 2006.

[ABS24] L. Ambrosio, E. Brué, and D. Semola, *Lectures on optimal transport*, 2nd ed., UNITEXT, Springer, Cham, 2024.

[AT03] L. Ambrosio and P. Tilli, *Topics on analysis in metric spaces*, Oxford University Press, 2003.

[Bre87] Y. Brenier, *Décomposition polaire et réarrangement monotone des champs de vecteurs*, C. R. Acad. Sci. Paris Sér. I Math. **305** (1987), 805–808.

[CDPJ08] T. Champion, L. De Pascale, and P. Juutinen, *The ∞-wasserstein distance: Local solutions and existence of optimal transport maps*, SIAM Journal on Mathematical Analysis **40** (2008), no. 1, 1–20.

[EG92] L.C. Evans and R.F. Gariepy, *Measure theory and fine properties of functions*, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1992.

[ET99] I. Ekeland and R. Témam, *Convex analysis and variational problems*, Society for Industrial and Applied Mathematics, 1999.

[Kan42] L.V. Kantorovich, *On the translocation of masses*, Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS (N.S.) (1942), 199–201.

[Mon81] G. Monge, *Mémoire sur la théorie des déblais et des remblais*, Histoire de l'Académie Royale des Sciences de Paris (1781), 666–704.

[RF10] H. L. Royden and P. M. Fitzpatrick, *Real analysis*, 4th ed., Pearson, Boston, MA, 2010.

[San15] F. Santambrogio, *Optimal transport for applied mathematicians*, Progress in Nonlinear Differential Equations and Their Applications, vol. 87, Birkhäuser/Springer, Cham, 2015.