MASTERARBEIT | MASTER'S THESIS

Titel | Title

Polar factorization of maps on Riemannian manifolds

verfasst von | submitted by

Fatemeh Montazeri Roudbaraki

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 821
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Mathematik
Betreut von | Supervisor:	Univ.-Prof. Mag. Mag. Dr. Michael Kunzinger

Abstract

In this thesis, we investigate the problem of polar factorization of maps on Riemannian manifolds. Our aim is to factorize a Borel map $f : M \to M$ that never maps positive volume into zero volume on a connected, compact $\mathcal{C}^3$-smooth Riemannian manifold (M, g) with a Riemannian distance $d(x, y)$ into the composition of a Borel map $g(x) = \exp_x[-\nabla\phi(x)]$ and a volume-preserving map $h : M \to M$, where $\phi : M \to \mathbb{R}$ is an infimal convolution $\phi = \phi^{CC}$ with $C(x, y) = \frac{d^2(x,y)}{2}$. Here, $g(x)$ is a solution of the Monge-Kantorovich transportation problem in Riemannian geometry.
For this purpose, we first introduce the notion of the optimal transport problem, which was first introduced by Monge, in Chapter 1. We state and prove Brenier's theorem, which shows the existence and uniqueness of the optimal transport problem in Euclidean space, in Chapter 2. Then, we address our main result. We recall some basic Riemannian geometry in Chapter 3, and then we prove Brenier's theorem on Riemannian manifolds in Chapter 4. Finally, in Chapter 5, we show that the polar factorization in Euclidean space can be generalized to polar factorization on Riemannian manifolds.

Zusammenfassung

In dieser Arbeit untersuchen wir das Problem der Polarisierung von Abbildungen auf Riemannschen Mannigfaltigkeiten. Unser Ziel ist es, eine Borel-Abbildung $f : M \to M$, die nie ein positives Volumen in ein Nullvolumen auf einer zusammenhängenden, kompakten $\mathcal{C}^3$-glatten Riemannschen Mannigfaltigkeit (M, g) mit einer Riemannschen Distanz $d(x, y)$ abbildet, zu faktorisieren in die Zusammensetzung einer Borel-Abbildung $g(x) = \exp_x[-\nabla\phi(x)]$ und einer volumenbewahrenden Abbildung $h : M \to M$. Hierbei ist $\phi : M \to \mathbb{R}$ eine infimale Faltung $\phi = \phi^{CC}$ mit $C(x, y) = \frac{d^2(x,y)}{2}$. Außerdem ist $g(x)$ eine Lösung des Monge-Kantorovich-Transportproblems in der Riemannschen Geometrie.
Hierzu betrachten wir zunächst den Begriff des optimalen Transportproblems, der von Monge eingeführt wurde. In Kapitel 2 beweisen wir Breniers Theorem, das die Existenz und Eindeutigkeit des optimalen Transportproblems im euklidischen Raum zeigt. Dann gehen wir auf unser Hauptergebnis ein. In Kapitel 3 wiederholen wir einige Grundlagen der Riemannschen Geometrie, und in Kapitel 4 beweisen wir Breniers Theorem auf Riemannschen Mannigfaltigkeiten. Schließlich zeigen wir in Kapitel 5, dass die Polarisierung im euklidischen Raum auf Riemannsche Mannigfaltigkeiten verallgemeinert werden kann.

Acknowledgments

First and foremost, I want to thank God for guiding and supporting me throughout my academic journey.
I would like to express my heartfelt thanks to my esteemed supervisor, Univ.-Prof. Mag. Mag. Dr. Michael Kunzinger, whose patience, invaluable knowledge, and experience were generously shared with me throughout the process of this thesis. Without their guidance and encouragement, this work would not have been possible.
I am also deeply grateful to my beloved family, whose unconditional love and support have always been my backbone. In particular, my heartfelt thanks go to my dear parents, who provided me with the best possible conditions for my studies and life, playing a crucial role in my success.
I extend my sincere appreciation to all my friends and classmates, who contributed significantly to the progress of this thesis through their insightful discussions and scientific exchanges.
Finally, I wish to thank all individuals who, in any way, have contributed to the preparation and completion of this thesis.

Contents

1 Introduction

In this chapter, we follow [1], Chapter 1, and [24]. We begin by providing a brief historical overview of Monge and Kantorovich's perspectives on optimal transport. Following that, we revisit fundamental concepts in measure theory that are necessary for the discussion.

1.1 Monge's approach to optimal transport

In 1781, the French mathematician Gaspard Monge introduced the concept of optimal transport for the first time.
If we extract a mass of soil from the ground and transfer it from position A to position B, what might be the cheapest possible way?
To answer this question, one needs to determine the cost of moving a unit of mass from position A to position B. In other words, it is necessary to identify the transportation cost. Obviously, Monge considered the space $\mathbb{R}^3$ and the cost function $C(x, y) := |x - y|$, which represents the Euclidean distance.

1.2 Kantorovich's approach to optimal transport

In 1940, the Russian mathematician Leonid Kantorovich revised Monge's perspective. Now, I would like to illustrate his approach with an example.
Consider m different cafes at positions $(x_i)_{i=1,\dots,m}$ and n different confectioneries at locations $(y_j)_{j=1,\dots,n}$. Assume that the total amount of Sachertorte each cafe needs to buy is equal to the total amount each confectionery sells, i.e., $\sum_{i=1}^{m} \alpha_i = \sum_{j=1}^{n} \beta_j$ and normalize them:

$$\sum_{i=1}^{m} \alpha_i = \sum_{j=1}^{n} \beta_j = 1,$$

where $\alpha_i, \beta_j \geqslant 0$.
In Monge's approach, a unit of mass can be transferred from position A to a unique position B. In contrast, in Kantorovich's approach, each cafe has the option to purchase Sachertorte from different confectioneries, and similarly, each confectionery can sell to multiple cafes. So Kantorovich introduced a new formulation:
He aimed to minimize the total transportation cost, considering the cost per unit from x_i to y_j denoted by $C(x_i, y_j)$. Kantorovich sought matrices $(\gamma_{ij})_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$ where $\gamma_{ij} \geqslant 0$.

These matrices represent the amount of Sachertorte moving from x_i to y_j. He introduced constraints to ensure that the total Sachertorte each cafe needed, i.e., $\alpha_i = \sum_{j=1}^{n} \gamma_{ij}, \quad \forall i$, and the total Sachertorte each confectionery sent, i.e., $\beta_j = \sum_{i=1}^{m} \gamma_{ij}, \quad \forall j$, were equal for all i and j. So, he formulated the expression $\sum_{i,j=1}^{k} \gamma_{ij} C(x_i, y_j)$ with the goal of minimizing the total cost by determining the values of γ_{ij}.

1.3 Preliminaries in measure theory

In this section, we revisit some fundamental concepts of measure theory in order to define the transport map.

Definition 1.3.1 (Push-forward measure)**.** Let X and Y be locally compact, separable, and complete metric spaces. Assume that μ is a probability measure on X, i.e., $\mu \in \mathcal{P}(X)$, and let $H : X \to Y$ be a map. The push-forward measure $H_{\#}\mu \in \mathcal{P}(Y)$ is defined as follows:

$$(H_{\#}\mu)(A) := \mu(H^{-1}(A)) \quad \text{for any } A \ \in \mathcal{B}(Y),$$

where $\mathcal{B}(Y)$ is the class of Borel measurable sets.

Lemma 1.3.1. The push-forward measure $H_{\#}\mu$ is a probability measure on Y.

Proof. See [1] (Lemma 1.2.3, page 3). □

Lemma 1.3.2. Assume that X and Y are locally compact, separable, and complete metric spaces, and let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ be probability measures. Then $H_{\#}\mu = \nu$ if and only if

$$\int_X f(H(x))d\mu(x) = \int_Y f(y)d\nu(y)$$

for any Borel-measurable and bounded function $f : Y \to \mathbb{R}$.

Proof. See [1] (Lemma 1.2.5, page 3). □

Corollary 1.3.1. Let X and Y be as defined in Definition 1.3.1, and consider $H : X \to Y$. Then, for any bounded and Borel-measurable function $g : Y \to \mathbb{R}$, we have:

$$\int_X g \circ H d\mu = \int_Y gd(H_{\#}\mu).$$

Lemma 1.3.3 (Relation between composition and push-forward)**.** Let $H : X \to Y$ and $P : Y \to Z$ be measurable. Then, for $\mu \in \mathcal{P}(X)$:

$$(P \circ H)_{\#}\mu = P_{\#}(H_{\#}\mu).$$

Proof. By Lemma 1.3.2, it is sufficient to demonstrate that for any Borel-measurable and bounded function $f : Z \to \mathbb{R}$:

$$\int_Z f d(P \circ H)_{\#}\mu = \int_Z f dP_{\#}(H_{\#}\mu).$$

We consider $f : Z \to \mathbb{R}$ to be a Borel-measurable and bounded function, and then apply Corollary 1.3.1 and Lemma 1.3.2 to obtain the desired result:

$$\begin{aligned} \int_Z f d(P \circ H)_{\#}\mu &= \int_X f \circ (P \circ H) d\mu = \int_X (f \circ P) \circ H d\mu \\ &= \int_Y f \circ P d(H_{\#}\mu) = \int_Z f dP_{\#}(H_{\#}\mu). \end{aligned}$$

□

1.4 Transport maps

Definition 1.4.1 (Transport map)**.** Let X and Y be locally compact, separable, and complete metric spaces, and let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. If a map $H : X \to Y$ satisfies $H_{\#}\mu = \nu$, then H is called a transport map from μ to ν.

Definition 1.4.2 (Coupling)**.** Assume that X and Y are locally compact, separable, and complete metric spaces, and $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ are probability measures. Then $\gamma \in \mathcal{P}(X \times Y)$ is called a coupling if

$$(\pi_x)_{\#}\gamma = \mu \quad \text{and} \quad (\pi_y)_{\#}\gamma = \nu,$$

where π_x and π_y are projections from $X \times Y$ onto X and Y, respectively.

Equivalently, for any Borel-measurable and bounded functions $f : X \to \mathbb{R}$ and $g : Y \to \mathbb{R}$:

$$\int_{X \times Y} f(x) d\gamma(x, y) = \int_X f(x) d\mu(x),$$

$$\int_{X \times Y} g(y) d\gamma(x, y) = \int_Y g(y) d\nu(y).$$

The set of couplings of μ and ν is denoted by $\mathrm{Coup}(\mu, \nu)$.

Remark 1.4.1 (Any transport map induces a coupling)**.** If $H : X \to Y$ is a transport map, consider the map $Id \times H : X \to X \times Y; x \mapsto (x, H(x))$, and define

$$\gamma_H := (Id \times H)_{\#}\mu.$$

It follows that γ_H is a coupling of μ and ν.

Proof. Utilizing Lemma 1.3.3 and the definition of γ_H,we have

$$(\pi_x)_\#\gamma_H = (\pi_x)_\#(Id \times H)_\#\mu = (\pi_x \circ (Id \times H))_\#\mu = Id_\#\mu = \mu,$$
$$(\pi_y)_\#\gamma_H = (\pi_y)_\#(Id \times H)_\#\mu = (\pi_y \circ (Id \times H))_\#\mu = H_\#\mu = \nu.$$

By Definition 1.4.2, we conclude that $\gamma_H \in \mathrm{Coup}(\mu, \nu)$. □

2 Optimal transport

In this chapter, we revisit the formulation of Monge's optimal transport problem in terms of transport maps, as well as Kantorovich's optimal transport problem expressed in terms of transport plans. Subsequently, we demonstrate the existence of an optimal transport coupling for the general cost in Kantorovich's problem, explore the concept of Kantorovich duality, and establish the existence of a solution for Monge's problem. This implies that there exists an optimal transport map for suitable functions. To achieve these results, we will rely on fundamental tools from measure theory, which will be introduced in the following section.

2.1 Tools and requirements

In this section, we follow [1], Section 2.1. From now on in this section, X will be assumed to be a compact, separable, and complete metric space. As a model case, consider $X = \mathbb{R}^d$.
Assume that every measure μ is a probability measure on X, denoted by $\mu \in \mathcal{P}(X)$. These assumptions are chosen to illustrate the main ideas of optimal transport theory and to establish that the key theorems, such as the existence of optimal transport plans and Kantorovich duality, hold for arbitrary separable metric spaces.
My goal is to present the 'Prokhorov theorem'. To achieve this, I will define narrow convergence, as it is crucial to ensure that the limit of probability measures remains a probability measure. The following definitions and conditions are necessary:

Remark 2.1.1. If B is a Banach space, B^* denotes its dual space. The following equalities hold, thanks to the Riesz representation theorem [3] (Theorem 7.7):

$$\begin{aligned}\mathcal{F}(X) &:= \{\mu \,:\, \mu \text{ is a finite signed measure}\}\\ &\cong C_c(X)^* := \{\text{continuous functions with compact support}\}^* \\ &\cong C_0(X)^* := \{\text{continuous functions vanishing at } \infty\}^*.\end{aligned} \tag{2.1}$$

The following example illustrates that if X is not compact, then $C_c(X)$ is in general not complete with respect to uniform convergence:

Example 2.1.1. Consider $X = \mathbb{R}$ which is not compact, but is locally compact.
We aim to demonstrate that $C_c(\mathbb{R})$ is not complete with respect to uniform convergence. In other words, I will show the existence of a sequence of functions in $C_c(\mathbb{R})$ whose limit is not in $C_c(\mathbb{R})$.
Define a continuous function satisfying

$$f_n : \mathbb{R} \to [0,1];\ f_n(x) := \begin{cases} 1 & x \in [-n,n], \\ 0 & x \notin [-n-1, n+1]. \end{cases}$$

Now set

$$g_n(x) := \frac{1}{x^2+1} f_n(x).$$

We have:

$$\lim_{n\to\infty} g_n(x) = \lim_{n\to\infty} \frac{1}{x^2+1} f_n(x) = \frac{1}{x^2+1} \lim_{n\to\infty} f_n(x) = \frac{1}{x^2+1}.$$

Thus,

$$g_n(x) \to \frac{1}{x^2+1} =: g(x)$$

as $n \to \infty$.
To demonstrate that $g_n(x)$ forms a Cauchy sequence, we proceed as follows:
Given $\varepsilon > 0$, we choose $N \in \mathbb{N}$ such that $\frac{1}{x^2+1} < \varepsilon$ for all $|x| > N$.
Thus, for all $m, n \geq N$, we consider $\|g_m - g_n\|_\infty = \sup_{x\in\mathbb{R}} |g_m(x) - g_n(x)|$.
Therefore, for all $|x| \leq N$,

$$\begin{aligned} |g_m(x) - g_n(x)| &= \left| \frac{1}{x^2+1} \cdot f_m(x) - \frac{1}{x^2+1} \cdot f_n(x) \right| \\ &= \frac{1}{x^2+1} \cdot |f_m(x) - f_n(x)| \\ &= 0 \\ &< \varepsilon, \end{aligned}$$

and for all $|x| > N$,

$$\begin{aligned} |g_m(x) - g_n(x)| &= \left| \frac{1}{x^2+1} \cdot f_m(x) - \frac{1}{x^2+1} \cdot f_n(x) \right| \\ &\leq \frac{2}{x^2+1} \\ &< \varepsilon. \end{aligned}$$

Consequently, $(g_n)_{n\in\mathbb{N}}$ constitutes a Cauchy sequence.
Since $\lim_{x\to\infty} g(x) = 0$, we conclude that $g \in C_0(\mathbb{R})$, but $g \notin C_c(\mathbb{R})$.
Claim: $(g_n)_{n\in\mathbb{N}} \in C_c(\mathbb{R})$ converges to $g \in C_0(\mathbb{R})$, where $g \notin C_c(\mathbb{R})$:

$$|g_n(x) - g(x)| = \begin{cases} 0, & \text{for } x \in [-n,n], \\ \frac{2}{x^2+1} < \varepsilon, & \text{for } x \notin [-n,n]. \end{cases}$$

Theorem 2.1.1 (Banach-Alaoglu theorem)**.** For any normed space X, the closed unit ball in X^* (dual space of X) is compact with respect to the weak*-topology.
(See [5] , Theorem 5.1 and [6], Theorem 1.)

Now we would like to show why narrow convergence is important. Weak*-convergence doesn't guarantee that the limit of probability measures is a probability measure: Consider $(\mu_n)_{n\geq 1}$ a sequence of probability measures on X. By the definition of probability measure, $(\mu_n)(X) = 1$. Therefore, by Jordan's decomposition theorem (See [11], Theorem 2), the sequence $(\mu_n)_{n\geq 1}$ is uniformly bounded in $\mathcal{F}(X)$. Now, use the Banach-Alaoglu theorem to conclude that there is a subsequence of probability measures $(\mu_{n_k})_{k\geq 1}$ that weakly*-converges to a measure $\mu \in \mathcal{F}(X)$, i.e.,

$$\mu_{n_k} \overset{*}{\rightharpoonup} \mu \in \mathcal{F}(X).$$

Equivalently, $\forall f \in C_c(X)$,

$$\lim_{k\to\infty} \int_X f d\mu_{n_k} = \int_X f d\mu.$$

Since $\mu_n \geq 0$ and $\mu_n \overset{*}{\rightharpoonup} \mu$, we conclude that $\mu \geq 0$ and $\mu \in \mathcal{F}(X)$.
In the following example, we demonstrate that μ is not necessarily a probability measure.

Example 2.1.2. Consider $X = \mathbb{R}$ and the Dirac measure $\mu_n = \delta_n$, $n \in \mathbb{Z}$. As we know, δ_n is a probability measure. Now, we will demonstrate that δ_n converges to 0, which is not a probability measure.
For any $f \in C_c(X)$:

$$\int_X f\, d\mu_n = \int_X f\, d\delta_n = f(n) \to 0 \quad \text{as} \quad n \to \infty,$$

since f has compact support on X.
Hence, $\mu_n \overset{*}{\rightharpoonup} 0$.
To address this issue, we need to introduce a stronger notion of convergence known as narrow convergence.

Definition 2.1.1 (Narrow convergence)**.** If $(\mu_n)_{n\in\mathbb{N}}$ is a sequence of measures and μ is a measure on X, then μ_n converges narrowly to μ if:

$$\int_X f\, d\mu_n \to \int_X f\, d\mu \qquad \forall f \in C_b(X),$$

where $C_b(X)$ denotes the set of all bounded continuous functions on X.
We denote the narrow convergence by $\mu_n \rightharpoonup \mu$.

The following example illustrates the significance of narrow convergence in this context:

Example 2.1.3. Suppose that $X = \mathbb{R}^d$ and $\mu_n := (1-\frac{1}{n})\delta_0+\frac{1}{n}\delta_{x_n}$ for some $x_n \in \mathbb{R}^d$, and let $f \in C_b(\mathbb{R}^d)$. We show that the sequence of probability measures $(\mu_n)_{n\geq 1}$ converges narrowly to a probability measure. We have

$$\begin{aligned}\int_{\mathbb{R}^d} f\,d\mu_n &= \int_{\mathbb{R}^d} f\,d\left[(1-\frac{1}{n})\delta_0 + \frac{1}{n}\delta_{x_n}\right]\\ &= (1-\frac{1}{n})\int_{\mathbb{R}^d} f\,d\delta_0 + \frac{1}{n}\int_{\mathbb{R}^d} f\,d\delta_{x_n}\\ &= (1-\frac{1}{n})f(0) + \frac{1}{n}f(x_n).\end{aligned}$$

Thus

$$\begin{aligned}\lim_{n\to\infty}\int_{\mathbb{R}^d} f\,d\mu_n &= \lim_{n\to\infty}\left[(1-\frac{1}{n})f(0)+\frac{1}{n}f(x_n)\right]\\ &= f(0)\\ &= \int_{\mathbb{R}^d} f\,d\delta_0.\end{aligned}$$

Hence

$$\int_{\mathbb{R}^d} f\,d\mu_n \to \int_{\mathbb{R}^d} f\,d\mu,$$

where $\mu := \delta_0$ is a probability measure.

Now, we need to define tightness, which is necessary to ensure that almost all the mass of the measures remains within a fixed compact set against weak*-convergence.

Definition 2.1.2 (Tight set)**.** A family of probability measures $\mathcal{T} \subset \mathcal{P}(X)$ is said to be tight if for any $\varepsilon > 0$, there exists a compact subset $K_\varepsilon \subset X$ such that:

$$\mu(X \setminus K_\varepsilon) \leq \varepsilon \quad \forall \mu \in \mathcal{T}.$$

Goal: We aim to establish the equivalence between the tightness of a family of probability measures and its compactness concerning the narrow topology.

Theorem 2.1.2 (Prokhorov)**.** A family $\mathcal{K} \subset \mathcal{P}(X)$ is tight if and only if $\mathcal{K}$ is relatively compact with respect to the narrow topology (refer to Theorem 2.1.11 in [1] and Theorem 1.3 in [2]).

Conclusion. The Prokhorov theorem shows that tightness is a necessary and sufficient condition for compactness with respect to narrow convergence.
Now, by the following lemma, we show that if f is a non-negative function and lower semicontinuous, then the map $\mu \mapsto \int f\,d\mu$ is lower semicontinuous under weak*-convergence. Since the narrow topology is stronger than the weak*-topology, it can be concluded that the same result is true under narrow convergence.

Definition 2.1.3 (Lower semicontinuous function)**.** For a given function $f : X \to [0, +\infty]$, if for any $x \in X$,

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

as $x_n \to x$, then f is called a lower semicontinuous function.

Lemma 2.1.1 (Lower semicontinuity of integrals)**.** Assume that $f : X \to [0, +\infty]$ is a lower semicontinuous function, and $\mu_n \overset{*}{\rightharpoonup} \mu$. Then

$$\liminf_{n\to\infty} \int_X f \, d\mu_n \geq \int_X f \, d\mu.$$

Proof. See [1], Lemma 2.1.12. □

2.2 Monge formulation of the optimal transport problem

In this section, we follow [2], Section 1. Let X and Y be compact, separable, and complete metric spaces. We consider probability measures μ and ν on X and Y, respectively, denoted by $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. Additionally, we have a lower semicontinuous cost function $C : X \times Y \to [0, +\infty]$. If $H : X \to Y$ is a transport map which transfers μ to ν, i.e., $H_{\#}\mu = \nu$; then the Monge optimal transport problem is minimizing

$$H \to \int_X C(x, H(x)) \, d\mu(x)$$

among all transport maps H.
The existence of solutions of Monge's problem depends on the choice of cost function. Furthermore, the problem may not be well-posed. Indeed :
(i) acceptable H may not exist: If we consider $\mu = \delta_{x_0}$ where $x_0 \in X$ and $H : X \to Y$; then

$$\int_Y f(y) \, d(H_{\#}\mu)(y) = \int_X f \circ H \, d\mu(x) = f(H(x_0)) \qquad \forall f : Y \to \mathbb{R}.$$

So, we conclude that $H_{\#}\mu = \delta_{H(x_0)}$.
Hence, if $\nu := \delta_{H(x_0)}$; then an acceptable H exists, otherwise H does not exist.
(ii) $H_{\#}\mu = \nu$ is not sequentially closed with respect to any reasonable weak topology:
For example, let $T : \mathbb{R} \to \mathbb{R}$ be defined by

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

be a 1−periodic function and define $T_n(x) := T(nx)$. Assume that μ is the Lebesgue measure on $[0, 1]$, i.e., $\mu = \mathcal{L}|_{[0,1]}$ and $\nu := \frac{\delta_{-1}+\delta_1}{2}$.
Through straightforward computations, $T_{n\#}\mu = \nu$ for any $n \in \mathbb{N}$ and $T_n \to T \equiv 0$ weakly, but $T_{\#}\mu = \delta_0 \neq \nu$.
To address this issue, I will revisit the Kantorovich problem in the next section.

2.3 Kantorovich formulation of the optimal transportation problem

In this section, we follow [1], Chapter 1 and [2], Section 1. Assume that X and Y are compact, separable, and complete metric spaces. Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, and let $C : X \times Y \to [0, +\infty]$ be a lower semicontinuous function.
Kantorovich formulated the transport problem by minimizing

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

where $\gamma \in \mathrm{Coup}(\mu,\nu)$.
This formulation offers several advantages:
(i) $\mathrm{Coup}(\mu,\nu)$ is always nonempty:
Let $\gamma = \mu \otimes \nu$ be the product measure defined by

$$\int F(x,y)\, d\gamma(x,y) = \int F(x,y)\, d\mu(x)\, d\nu(y),$$

for every $F : X \times Y \to \mathbb{R}$. Then $\gamma \in$Coup(μ,ν):

$$\begin{aligned}\int_{X\times Y} f(x)\, d\mu(x)\, d\nu(y) &= \int_Y d\nu(y) \int_X f(x)\, d\mu(x)\\ &= 1 \cdot \int_X f(x)\, d\mu(x)\\ &= \int_X f(x)\, d\mu(x),\end{aligned}$$

$$\begin{aligned}\int_{X\times Y} g(y)\, d\mu(x)\, d\nu(y) &= \int_X d\mu(x) \int_Y g(y)\, d\nu(y)\\ &= 1 \cdot \int_Y g(y)\, d\nu(y)\\ &= \int_Y g(y)\, d\nu(y).\end{aligned}$$

(ii) The infimum of $\{\int_{X\times Y} C(x,y)\, d\gamma(x,y) \mid \gamma \in \mathrm{Coup}(\mu,\nu)\}$ over $\gamma \in \mathrm{Coup}(\mu,\nu)$ always exists under suitable assumptions, cf. Theorem 1.5 of [2].
(iii) Any transport map H induces a coupling $\gamma := (Id \times H)_{\#}\mu$. [See [1], Remark 1.4.5]

Remark 2.3.1. A coupling given by a graph induces a transport map:
Assume that $\gamma \in \mathrm{Coup}(\mu,\nu)$ and suppose there exists $G : X \to Y$ such that $\gamma = (Id \times G)_{\#}\mu$. Then

$$\nu = (\pi_y)_{\#}\gamma = (\pi_y)_{\#}(Id \times G)_{\#}\mu = (\pi_y \circ (Id \times G))_{\#}\mu = G_{\#}\mu.$$

Hence G is a transport map from μ to ν.

2.4 Existence of an optimal coupling

In this section, we follow [1], Section 2.3 and [2], Section 1.4. We want to show that the solution of the Kantorovich problem exists. For this purpose, the following lemma is required:

Lemma 2.4.1. Assume that X and Y are compact, separable, and complete metric spaces. Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. The set of couplings from μ to ν in $\mathcal{P}(X \times Y)$ is tight and closed under narrow convergence.

Proof. We prove this lemma separately in two steps:

Step 1. $\mathrm{Coup}(\mu, \nu)$ is tight:
We need to show that $\forall \varepsilon > 0,\ \exists \tilde{K}_\varepsilon \subset X \times Y$ such that $\gamma(X \times Y \setminus \tilde{K}_\varepsilon) \leq \varepsilon$:

By using Lemma 2.1.9 of [1], we have

$$\begin{aligned}
&\forall \varepsilon > 0, \quad \exists K_\varepsilon \subset X \qquad \text{s.t.} \qquad \mu(X \setminus K_\varepsilon) \leq \frac{\varepsilon}{2}, \\
&\forall \varepsilon > 0, \quad \exists K'_\varepsilon \subset Y \qquad \text{s.t.} \qquad \nu(Y \setminus K'_\varepsilon) \leq \frac{\varepsilon}{2}.
\end{aligned}$$

Now by defining the compact set $\tilde{K}_\varepsilon := K_\varepsilon \times K'_\varepsilon \subset X \times Y, \forall \gamma \in \mathrm{Coup}(\mu, \nu)$, we have

$$\begin{aligned}
\gamma(X \times Y \setminus \tilde{K}_\varepsilon) &= \gamma\left(((X \setminus K_\varepsilon) \times Y) \cup (X \times (Y \setminus K'_\varepsilon))\right) \\
&\leq \gamma\left((X \setminus K_\varepsilon) \times Y\right) + \gamma\left(X \times (Y \setminus K'_\varepsilon)\right) \\
&= \int_{X \times Y} \mathbb{1}_{X \setminus K_\varepsilon}(x)\, d\gamma(x, y) + \int_{X \times Y} \mathbb{1}_{Y \setminus K'_\varepsilon}(y)\, d\gamma(x, y) \\
&= \int_X \mathbb{1}_{X \setminus K_\varepsilon}(x)\, d\mu(x) + \int_Y \mathbb{1}_{Y \setminus K'_\varepsilon}(y)\, d\nu(y) \\
&= \mu(X \setminus K_\varepsilon) + \nu(Y \setminus K'_\varepsilon) \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\
&= \varepsilon. \\
&\Rightarrow \gamma(X \times Y \setminus \tilde{K}_\varepsilon) \leq \varepsilon.
\end{aligned}$$

Hence, $\mathrm{Coup}(\mu, \nu)$ is tight.

Step 2. $\mathrm{Coup}(\mu, \nu)$ is closed under narrow convergence:
Let $(\gamma_n)_{n \geq 1} \subset \mathrm{Coup}(\mu, \nu)$ and assume that $\gamma_n \rightharpoonup \gamma \in \mathcal{P}(X \times Y)$. We show that $\gamma \in \mathrm{Coup}(\mu, \nu)$, i.e., $(\pi_x)_\# \gamma = \mu$ and $(\pi_y)_\# \gamma = \nu$:
Since $\gamma_n \rightharpoonup \gamma$, for any $f \in C_b(X)$, we have

$$\int_X f(x)\, d\mu(x) = \int_{X \times Y} f(x)\, d\gamma_n(x, y) \to \int_{X \times Y} f(x)\, d\gamma(x, y).$$

Hence, $(\pi_x)_{\#}\gamma = \mu$.

Similarly, $(\pi_y)_{\#}\gamma = \nu$.

□

Now, we aim to state and prove the main theorem of this section, demonstrating the existence of an optimal coupling under general assumptions on the cost function.

Theorem 2.4.1 (Existence of an optimal Coupling)**.** For a lower semicontinuous function $C : X \times Y \to [0, +\infty]$ and probability measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, there exists a coupling $\tilde{\gamma} \in \mathrm{Coup}(\mu, \nu)$ that minimizes the Kantorovich problem.

Proof. We show that for

$$C(\mu, \nu) := \inf_{\gamma \in \mathrm{Coup}(\mu,\nu)} \left\{ \int_{X \times Y} C(x, y)\, d\gamma(x, y) | \gamma \in \mathrm{Coup}(\mu, \nu) \right\}$$

the minimizer γ exists:
If $\inf_\gamma \int_{X\times Y} C\, d\gamma = +\infty$, then it is clear that every $\gamma \in \mathrm{Coup}(\mu, \nu)$ is a minimizer. Without loss of generality, assume $\inf_\gamma \int_{X\times Y} C\, d\gamma < +\infty$, and define

$$a := \inf_{\gamma \in \mathrm{Coup}(\mu,\nu)} \int_{X \times Y} C(x, y)\, d\gamma(x, y).$$

Assume that $(\gamma_n)_{n\geq 1}$ is a minimizing sequence in $\mathrm{Coup}(\mu, \nu)$, where

$$\lim_{n\to\infty} \int_{X\times Y} C(x, y)\, d\gamma_n(x, y) = a. \tag{2.2}$$

Since by Lemma 2.4.1, $\mathrm{Coup}(\mu, \nu)$ is tight, and $(\gamma_n)_{n\geq 1} \in \mathrm{Coup}(\mu, \nu)$, by Prokhorov's theorem, there exists a subsequence $(\gamma_{n_\ell})_{\ell\geq 1}$ and $\tilde{\gamma} \in \mathcal{P}(X \times Y)$ such that γ_{n_ℓ} converges narrowly to $\tilde{\gamma}$ (and hence $\gamma_{n_\ell} \overset{*}{\rightharpoonup} \tilde{\gamma}$).
By assumption, C is lower semicontinuous, i.e.,

$$\liminf_{\ell\to\infty} \int_{X\times Y} C\, d\gamma_{n_\ell} \geq \int_{X\times Y} C\, d\tilde{\gamma}. \tag{2.3}$$

Hence:

$$\inf_\gamma \int_{X\times Y} C\, d\gamma = a = \liminf_{\ell\to\infty} \int_{X\times Y} C\, d\gamma_{n_\ell} \geq \int_{X\times Y} C\, d\tilde{\gamma},$$

where (2.2) is used in the first equation and (2.3) is used in the last inequality.
So

$$\inf_\gamma \int_{X\times Y} C\, d\gamma \geq \int_{X\times Y} C\, d\tilde{\gamma}. \tag{2.4}$$

Claim: $\tilde{\gamma} \in \mathrm{Coup}(\mu, \nu)$: By Lemma 2.4.1, $\mathrm{Coup}(\mu, \nu)$ is closed under narrow convergence and since $\gamma_{n_\ell} \rightharpoonup \tilde{\gamma}$ we have $\tilde{\gamma} \in \mathrm{Coup}(\mu, \nu)$. Since $\tilde{\gamma} \in \mathrm{Coup}(\mu, \nu)$, $\tilde{\gamma}$ is a coupling that satisfies

$$\int_{X\times Y} C\, d\tilde{\gamma} \geq a. \tag{2.5}$$

Now the conclusion follows from (2.4) and (2.5), so $\int_{X\times Y} C\,d\tilde{\gamma} = a$. Thus $\tilde{\gamma}$ is a minimizer. $\square$

The previous theorem holds for cost function C bounded from below.
Two important questions are:

1. Is the minimizer γ unique?

2. Is γ induced by a transport map?

To answer these two questions, we consider the following examples.

Example 2.4.1. The minimizer of the Kantorovich problem is unique.

Proof. Consider $\mu = \delta_{x_0}$ and $\nu = \frac{1}{2}\delta_{y_0} + \frac{1}{2}\delta_{y_1}$. Then $\gamma := \frac{1}{2}\delta_{(x_0,y_0)} + \frac{1}{2}\delta_{(x_0,y_1)}$ is unique, but it is not induced by a transport map. [See [1], Example 2.3.4.] $\square$

Example 2.4.2. Consider $X = Y = \mathbb{R}^2$ and $C(x,y) := |x-y|^2$. Let

$$x_0 := (0,0), x_1 := (1,1), y_0 := (1,0), y_1 := (0,1),$$

be points in $\mathbb{R}^2$. Define

$$\mu := \frac{1}{2}\delta_{x_0} + \frac{1}{2}\delta_{x_1} \quad \text{and} \quad \nu := \frac{1}{2}\delta_{y_0} + \frac{1}{2}\delta_{y_1}.$$

The set of all couplings that send μ to ν is obtained by

$$\gamma_t := t\delta_{(x_0,y_0)} + \left(\frac{1}{2} - t\right)\delta_{(x_0,y_1)} + \left(\frac{1}{2} - t\right)\delta_{(x_1,y_0)} + t\delta_{(x_1,y_1)}, \qquad t \in \left[0, \frac{1}{2}\right].$$

Hence, if $t \in [0, \frac{1}{2}]$, γ_t are optimal:

$$\begin{aligned}\int_{X\times Y} C\,d\gamma_t &= \int_{X\times Y} |x-y|^2\,d\gamma_t \\ &= t|x_0 - y_0|^2 + \left(\frac{1}{2} - t\right)|x_0 - y_1|^2 + \left(\frac{1}{2} - t\right)|x_1 - y_0|^2 + t|x_1 - y_1|^2 \\ &= 1.\end{aligned}$$

[See [1], Example 2.3.5.]

Consequently, for proving uniqueness, additional assumptions are necessary.

2.5 Necessary and sufficient optimal conditions

In this section, we follow [2], Section 1.2 and [1], Section 2.4. In order to understand the structure of optimal plans, we begin by considering the support of a measure. The properties of the support of an optimal coupling are then explored. Subsequently, we define C-cyclical monotonicity and subdifferentiability of convex functions and superdifferentiability of concave functions. Following this, the Rockafellar theorem is presented, illustrating the relationship between cyclically monotone sets and the subdifferential of a function.

Definition 2.5.1 (Support of a measure)**.** Let (X, τ) be a topological space and μ be a measure on $(X, \mathcal{B}(\tau))$. Then its support can be defined by:

$$\operatorname{supp}(\mu) := \{x \in X | \mu(B(x, \varepsilon)) > 0 \quad \text{for any} \quad \varepsilon > 0\}.$$

Definition 2.5.2 (Support of an Optimal Plan)**.** Consider an optimal coupling $\tilde{\gamma} \in \operatorname{Coup}(\mu, \nu)$ with finite cost, i.e.,

$$\int_{X\times Y} C(x, y)\, d\tilde{\gamma}(x, y) = \inf_{\gamma\in\operatorname{Coup}(\mu,\nu)} \int_{X\times Y} C(x, y)\, d\gamma(x, y) < +\infty.$$

We define its support by:

$$\operatorname{supp}(\tilde{\gamma}) := \{(x, y) \in X \times Y \,|\, \tilde{\gamma}(B(x, \varepsilon) \times B(y, \varepsilon)) > 0 \quad \forall\varepsilon > 0\}.$$

Now, for a better understanding of the structure of optimal couplings, consider the case where $X = Y = \mathbb{R}^d$ and $(x_k, y_k) \in \operatorname{supp}(\tilde{\gamma})$ for $k \in \mathbb{N}$. This implies that these pairs (x_k, y_k) can be sent from x to y through the optimal coupling $\tilde{\gamma}$. By considering another transport map that redistributes mass from x_2 to y_1, x_3 to y_2, up to x_{k+1} to y_k (where $x_{k+1} \equiv x_1$), we observe that since $\tilde{\gamma}$ is an optimal coupling, the cost must increase. Formally, this is expressed as:

$$\sum_{k=1}^{d} C(x_k, y_k) \leq \sum_{k=1}^{d} C(x_{k+1}, y_k).$$

This property hols for any pair in $\operatorname{supp}(\tilde{\gamma})$. So now we need to define the following:

Definition 2.5.3 (C-cyclical monotonicity)**.** Let $\mathcal{M} \subset X \times Y$ be a set. $\mathcal{M}$ is called C-cyclically monotone whenever for a finite sequence $(x_k, y_k)_{k=1,\ldots,m}$ in $\mathcal{M}$ we have:

$$\sum_{k=1}^{m} C(x_k, y_k) \leq \sum_{k=1}^{m} C(x_{k+1}, y_k)$$

with $x_{m+1} \equiv x_1$.

As a conclusion to the discussion above, the following theorem demonstrates that optimality implies C-cyclical monotonicity.

Theorem 2.5.1. Assume that $\tilde{\gamma}$ is an optimal coupling, and $C : X \times Y \to \mathbb{R}$ is a continuous cost function. Then $\operatorname{supp}(\tilde{\gamma})$ is C-cyclically monotone. For a proof of this theorem, we refer to [1], Theorem 2.4.3.

Now, we would like to state the Rockafellar theorem. For this purpose, we need to define the concepts of subdifferential and superdifferential. For understanding how this concept is defined, we will consider the following special case:
Suppose that $X = Y = \mathbb{R}^d$ and $C(x, y) := \frac{|x-y|^2}{2}$. Let:

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

For a coupling $\gamma \in \operatorname{Coup}(\mu, \nu)$, we have:

$$\begin{aligned}\int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{|x-y|^2}{2}\, d\gamma(x,y) &= \int_{\mathbb{R}^d\times\mathbb{R}^d} \Big(\frac{|x|^2}{2} - x\cdot y + \frac{|y|^2}{2}\Big)\, d\gamma(x,y) \\ &= \int_{\mathbb{R}^d} \frac{|x|^2}{2}\, d\mu(x) + \int_{\mathbb{R}^d} \frac{|y|^2}{2}\, d\nu(y) + \int_{\mathbb{R}^d\times\mathbb{R}^d} -x\cdot y\, d\gamma(x,y).\end{aligned} \tag{2.6}$$

The coupling γ is optimal for $\frac{|x-y|^2}{2}$ if and only if γ is optimal for the cost function $C(x, y) = -x \cdot y$, since by assumption:

$$\int_{\mathbb{R}^d} \frac{|x|^2}{2}\, d\mu + \int_{\mathbb{R}^d} \frac{|y|^2}{2}\, d\nu < +\infty.$$

Therefore, if $C(x, y) = -x \cdot y$, the C-cyclical monotonicity condition is equivalent to:

$$\sum_{k=1}^{m} \langle y_k, x_{k+1} - x_k\rangle \le 0, \tag{2.7}$$

where $\langle\cdot,\cdot\rangle$ denotes the scalar product on $\mathbb{R}^d$, and $x_{m+1} \equiv x_1$.
Goal. Characterize cyclical monotonicity by the subdifferential of convex functions. We will recall the definition of the subdifferential of a convex function and then discuss the relation between the gradient of a differentiable function and its subdifferential.

Definition 2.5.4 (Subdifferential and Superdifferential)**.** Assume that $\phi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is a convex function. Its subdifferential is defined by:

$$\partial^-\phi(x) := \{y \in \mathbb{R}^d | \phi(z) \ge \phi(x) + \langle y, z - x\rangle \quad \forall z \in \mathbb{R}^d\}.$$

Define

$$\partial^-\phi := \bigcup_{x\in\mathbb{R}^d} \{x\} \times \partial^-\phi(x) \subset \mathbb{R}^d \times \mathbb{R}^d.$$

Similarly, for $\phi : \times\mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ a concave function, its superdifferential is defined by:

$$\partial^+\phi(x) := \{y \in \mathbb{R}^d | \phi(z) \le \phi(x) + \langle y, z - x\rangle \quad \forall z \in \mathbb{R}^d\}.$$

Lemma 2.5.1. Assume that $\phi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is a convex function with a subdifferential at x in $\mathbb{R}^d$. Then:

$$\partial^- \phi(x) = \{\nabla \phi(x)\}.$$

To prove this, refer to [1], Lemma 2.5.2.

Theorem 2.5.2 (Rockafellar)**.** Assume that $\mathcal{M} \subset \mathbb{R}^d \times \mathbb{R}^d$ is a set. $\mathcal{M}$ is cyclically monotone if and only if there is a convex function $\phi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ such that $\mathcal{M} \subset \partial^- \phi$.

Proof. $\Leftarrow$) If a convex function $\phi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ exists such that $\mathcal{M} \subset \partial^- \phi$, then we show that $\mathcal{M}$ is cyclically monotone, i.e., for any finite family of points $(x_i, y_i)_{i=1,\dots,m} \subset \mathcal{M}$:

$$\sum_{i=1}^{m} \langle y_i, x_{i+1} - x_i \rangle \leq 0 \quad \text{with} \quad x_{m+1} \equiv x_1.$$

Assume that $\mathcal{M} \subset \partial^- \phi$ and consider $(x_i, y_i)_{i=1,\dots,m}$ in $\mathcal{M}$. Hence, $y_i \in \partial^- \phi(x_i)$ for any i.
Therefore, by the definition of $\partial^- \phi$, we have:

$$\phi(z) \geq \phi(x_i) + \langle y_i, z - x_i \rangle \qquad \forall z \in \mathbb{R}^d.$$

By setting $z = x_{i+1}$, we get:

$$\phi(x_{i+1}) \geq \phi(x_i) + \langle y_i, x_{i+1} - x_i \rangle.$$

Taking the sum over i implies:

$$\sum_{i=1}^{m} \phi(x_{i+1}) \geq \sum_{i=1}^{m} \phi(x_i) + \sum_{i=1}^{m} \langle y_i, x_{i+1} - x_i \rangle,$$

and since $\sum_{i=1}^{m} \phi(x_{i+1}) = \sum_{i=1}^{m} \phi(x_i)$ and $x_{m+1} \equiv x_1$, we get

$$0 \geq \sum_{i=1}^{m} \langle y_i, x_{i+1} - x_i \rangle,$$

as desired.
$\Rightarrow$) Let $\mathcal{M} \subset \mathbb{R}^d \times \mathbb{R}^d$ be cyclically monotone, i.e., for any family of points $(x_i, y_i)_{i=1,\dots,m}$; $\sum_{i=1}^{m} \langle y_i, x_{i+1} - x_i \rangle \leq 0$ with $x_{m+1} \equiv x_1$.
We want to show the existence of a convex function $\phi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ such that $\mathcal{M} \subset \partial^- \phi$.
To prove this, we construct the function ϕ as follows:
Let $(x_0, y_0) \in \mathcal{M}$ be fixed. Considering ϕ as a convex function such that $\mathcal{M} \subset \partial^- \phi$ and $\phi(x_0) = 0$, we have:

$$(x_0, y_0) \in \partial^- \phi, \text{ i.e., } \phi(z) \geq \phi(x_0) + \langle y_0, z - x_0 \rangle \quad \forall z \in \mathbb{R}^d$$

$$\Rightarrow \quad \phi(z) \geq \langle y_0, z - x_0 \rangle \quad \forall z \in \mathbb{R}^d.$$

Now, setting $z = x$ and using induction on $m \geq 1$ for any sequence $(x_i, y_i)_{i=1,\ldots,m} \subset \mathcal{M}$, we obtain:

$$\phi(x) \geq \langle y_m, x - x_m \rangle + \langle y_{m-1}, x_m - x_{m-1} \rangle + \cdots + \langle y_0, x_1 - x_0 \rangle. \tag{2.8}$$

Now, we can define the minimum function satisfying property (2.8) as:

$$\phi(x) := \sup\{\langle y_m, x - x_m \rangle + \langle y_{m-1}, x_m - x_{m-1} \rangle + \cdots + \langle y_0, x_1 - x_0 \rangle | (x_i, y_i)_{i=1,\ldots,m} \subset \mathcal{M}, m \in \mathbb{N}\}.$$

The function ϕ possesses the following properties:

1. ϕ is convex:
   Since ϕ is the supremum of affine functions.

2. $\phi(x_0) \geq 0$:
   By choosing $m = 1$, we have $(x_1, y_1) = (x_0, y_0)$. Therefore:

   $$\phi(x) \geq \langle y_0, x - x_0 \rangle,$$

   and since $x = x_0$ we have:

   $$\phi(x_0) \geq \langle y_0, x_0 - x_0 \rangle.$$

   Hence

   $$\phi(x_0) \geq 0. \tag{2.9}$$

3. $\phi(x) \not\equiv +\infty$:
   We know that $\mathcal{M}$ is cyclically monotone. For any $(x_i, y_i)_{i=1,\ldots,m} \subset \mathcal{M}$ we have

   $$\sum_{i=1}^{m} \langle y_i, x_{i+1} - x_i \rangle \leq 0.$$

   Thus

   $$\langle y_m, x_{m+1} - x_m \rangle + \cdots + \langle y_1, x_2 - x_1 \rangle \leq 0,$$

   hence

   $$\langle y_m, x_0 - x_m \rangle + \cdots + \langle y_0, x_1 - x_0 \rangle \leq 0.$$

   Therefore

   $$\phi(x_0) = \sup_{m \in \mathbb{N}} \{\langle y_m, x_0 - x_m \rangle + \cdots + \langle y_0, x_1 - x_0 \rangle\} \leq 0.$$

   Thus:

   $$\phi(x_0) \leq 0. \tag{2.10}$$

   By (2.9) and (2.10), we conclude that $\phi(x_0) = 0$, thus $\phi(x) \not\equiv +\infty$. Therefore we have constructed a convex function $\phi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$.

Claim. $\mathcal{M} \subset \partial^- \phi$:
To demonstrate this, we will establish that for any $(x, y) \in \mathcal{M}$, it holds that $(x, y) \in \partial^- \phi$, which implies $\phi(z) \geq \phi(x) + \langle y, z - x \rangle$.
Assume that $(x, y) \in \mathcal{M}$ and $\phi(x) > a$. Now by using the definition of ϕ, there is a natural number m and a sequence $(x_i, y_i)_{i=1,\ldots,m}$ such that

$$\langle y_m, x - x_m \rangle + \cdots + \langle y_0, x_1 - x_0 \rangle \geq a. \tag{2.11}$$

(Since

$$\phi(x) := \sup_{m \geq 1} \{\langle y_m, x - x_m \rangle + \cdots + \langle y_0, x_1 - x_0 \rangle | (x_i, y_i)_{i=1,\ldots,m} \subset \mathcal{M}\}$$

and $a < \phi(x)$, and since

$$a < \sup_{m \geq 1} \{\langle y_m, x - x_m \rangle + \cdots + \langle y_0, x_1 - x_0 \rangle | (x_i, y_i)_{i=1,\ldots,m} \subset \mathcal{M}\},$$

thus

$$a \leq \langle y_m, x - x_m \rangle + \cdots + \langle y_0, x_1 - x_0 \rangle.)$$

Set $(x_{m+1}, y_{m+1}) = (x, y)$ and consider the sequence $(x_i, y_i)_{i=1,\ldots,m+1}$. This new sequence satisfies the definition of ϕ. So by using (2.11), we conclude that for any $z \in \mathbb{R}^d$:

$$\phi(z) \geq \langle y_{m+1}, z - x_{m+1} \rangle + \langle y_m, x_{m+1} - x_m \rangle + \cdots + \langle y_0, x_1 - x_0 \rangle.$$

Since $y = y_{m+1}$ and $x = x_{m+1}$; then

$$\phi(z) \geq \langle y, z - x \rangle + \langle y_m, x - x_m \rangle + \cdots + \langle y_0, x_1 - x_0 \rangle,$$

and since

$$\langle y_m, x - x_m \rangle + \cdots + \langle y_0, x_1 - x_0 \rangle \geq a,$$

by (2.11) we get $\phi(z) \geq \langle y, z - x \rangle + a$.
If $a \to \phi(x)$, then $\phi(z) \geq \langle y, z - x \rangle + \phi(x)$ for all $z \in \mathbb{R}^d$, i.e., $y \in \partial^- \phi(x)$ or $(x, y) \in \partial^- \phi$. ☐

2.6 Kantorovich Duality

In this section, we follow [1], subsection 2.5.2, and [2], section 1.3. Our objective is to derive a dual problem for the Kantorovich Problem using the Legendre transform. We begin by defining the Legendre transform and subsequently review its essential properties. Finally, we state and prove the Kantorovich duality theorem.
Having previously established the existence of an optimal coupling for nonnegative cost functions, we now need to introduce additional assumptions to demonstrate the existence of an optimal coupling for the cost function $C(x, y) = -x \cdot y$.
The cost function $C(x, y) = -x \cdot y$ is equivalent to $\overline{C}(x, y) := \frac{|x-y|^2}{2}$ if

$$\int \frac{|x|^2}{2}\, d\mu + \int \frac{|y|^2}{2}\, d\nu < +\infty.$$

Since $\gamma := \mu \otimes \nu \in \mathrm{Coup}(\mu, \nu)$ and

$$\begin{aligned}
&\int_{\mathbb{R}^d \times \mathbb{R}^d} \overline{C}(x, y)\, d(\mu \otimes \nu)(x, y) \\
&= \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{|x-y|^2}{2}\, d(\mu \otimes \nu)(x, y) \\
&= \int_{\mathbb{R}^d \times \mathbb{R}^d} \left(\frac{|x|^2}{2} - x \cdot y + \frac{|y|^2}{2} \right) d(\mu \otimes \nu)(x, y) \\
&\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} (|x|^2 + |y|^2)\, d(\mu \otimes \nu)(x, y) \\
&= \int_{\mathbb{R}^d} |x|^2\, d\mu(x) + \int_{\mathbb{R}^d} |y|^2\, d\nu(y) < +\infty,
\end{aligned}$$

so we conclude that $\inf_{\gamma \in \mathrm{Coup}(\mu,\nu)} \int_{X \times Y} \overline{C}\, d\gamma < +\infty$.
Therefore, Theorem 2.4.1. holds for the cost function $\overline{C}$; i.e., an optimal coupling for $\overline{C}$ exists, and this coupling is also optimal for C.

Definition 2.6.1 (Legendre Transform)**.** Assume that $\phi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is convex and $\phi \not\equiv +\infty$. Then its Legendre transform is defined by:

$$\begin{aligned}
&\phi^* : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}; \\
&\phi^*(y) := \sup_{x \in \mathbb{R}^d} \{x \cdot y - \phi(x)\}.
\end{aligned}$$

Proposition 2.6.1. For a given convex function $\phi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, the following properties hold:

(i) $\phi(x) + \phi^*(y) \geq x \cdot y \quad \forall x, y \in \mathbb{R}^d$;

(ii) $\phi(x) + \phi^*(y) = x \cdot y$ if and only if $y \in \partial^- \phi(x)$.

Proof. (i) For $x \in \mathbb{R}^d$, by definition $\phi^*(y) = \sup_{x \in \mathbb{R}^d}\{x \cdot y - \phi(x)\}$, we have:

$$\phi^*(y) \geq x \cdot y - \phi(x).$$

Hence $\phi(x) + \phi^*(y) \geq x \cdot y$ for all $x, y \in \mathbb{R}^d$.

(ii) $\Rightarrow$) If $\phi(x) + \phi^*(y) = x \cdot y$, then $y \in \partial^- \phi(x)$:
We show that

$$\phi(z) \geq \phi(x) + \langle y, z - x \rangle \quad \text{for all } z \in \mathbb{R}^d.$$

By assumption: $\phi(x) + \phi^*(y) = x \cdot y$. Thus,

$$x \cdot y - \phi(x) = \phi^*(y) \tag{2.12}$$

By (i), we have

$$\phi(z) + \phi^*(y) \geq z \cdot y \quad \text{for all } z \in \mathbb{R}^d. \tag{2.13}$$

Thus,

$$\phi^*(y) \geq z \cdot y - \phi(z).$$

Now, by (2.12) and (2.13):

$$x \cdot y - \phi(x) \geq z \cdot y - \phi(z).$$

Thus,

$$\phi(z) \geq \phi(x) + \langle y, z - x\rangle \quad \text{for all } z \in \mathbb{R}^d.$$

$\Leftarrow$) Assume that $y \in \partial^- \phi(x)$. We show

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

Since $y \in \partial^- \phi(x)$, we have:

$$\begin{aligned} & \phi(z) \geq \phi(x) + \langle y, z - x\rangle \quad \text{for all } z \in \mathbb{R}^d \\ \Leftrightarrow \quad & \phi(z) \geq \phi(x) + z \cdot y - x \cdot y \\ \Rightarrow \quad & x \cdot y - \phi(x) \geq z \cdot y - \phi(z). \end{aligned}$$

By taking the supremum over $z \in \mathbb{R}^d$:

$$\sup_{z \in \mathbb{R}^d} \{x \cdot y - \phi(x)\} \geq \sup_{z \in \mathbb{R}^d} \{z \cdot y - \phi(z)\}.$$

Thus:

$$\begin{aligned} & x \cdot y - \phi(x) \geq \sup_{z \in \mathbb{R}^d} \{z \cdot y - \phi(z)\} \\ \Rightarrow \quad & x \cdot y - \phi(x) \geq \phi^*(y) \\ \Rightarrow \quad & \phi(x) + \phi^*(y) \leq x \cdot y. \end{aligned} \tag{2.14}$$

By (i),

$$\phi(x) + \phi^*(y) \geq x \cdot y. \tag{2.15}$$

By (2.14) and (2.15), we obtain:

$$\phi(x) + \phi^*(y) = x \cdot y \quad \text{if } y \in \partial^- \phi(x).$$

□

Theorem 2.6.1 (Kantorovich Duality)**.** Let

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

If the functions $\Phi, \Psi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ are measurable and $\gamma \in \mathrm{Coup}(\mu, \nu)$, then

$$\min_{\gamma \in \mathrm{Coup}(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} -x \cdot y \, d\gamma(x, y) = \max_{\Phi(x)+\Psi(y) \geq x \cdot y} \left\{ \int_{\mathbb{R}^d} -\Phi(x) \, d\mu(x) + \int_{\mathbb{R}^d} -\Psi(y) \, d\nu(y) \right\}.$$

Furthermore, if we set $\Psi = \Phi^*$, then

$$\min_{\gamma \in \mathrm{Coup}(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} -x \cdot y \, d\gamma(x, y) = \max_{\Phi \text{ convex}} \left\{ \int_{\mathbb{R}^d} -\Phi(x) \, d\mu(x) + \int_{\mathbb{R}^d} -\Phi^*(y) \, d\nu(y) \right\}.$$

Proof. Consider $\Phi, \Psi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ with $\Phi(x) + \Psi(y) \geq x \cdot y$ for all $x, y \in \mathbb{R}^d$.
Now, we integrate $\Phi(x) + \Psi(y) \geq x \cdot y$ with respect to an arbitrary coupling $\gamma \in \mathrm{Coup}(\mu, \nu)$. Hence:

$$\begin{aligned} -x \cdot y &\geq -\Phi(x) + (-\Psi(y)) \\ &\Rightarrow \int_{\mathbb{R}^d \times \mathbb{R}^d} -x \cdot y \, d\gamma(x, y) \\ &\geq \int_{\mathbb{R}^d \times \mathbb{R}^d} -\Phi(x) \, d\gamma(x, y) + \int_{\mathbb{R}^d \times \mathbb{R}^d} -\Psi(y) \, d\gamma(x, y) \\ &= \int_{\mathbb{R}^d} -\Phi(x) \, d\mu(x) + \int_{\mathbb{R}^d} -\Psi(y) \, d\nu(y), \end{aligned}$$

where we will prove the last equality later, namely in $(_*)$.
Therefore,

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} -x \cdot y \, d\gamma(x, y) \geq \int_{\mathbb{R}^d} -\Phi(x) \, d\mu(x) + \int_{\mathbb{R}^d} -\Psi(y) \, d\nu(y). \tag{2.16}$$

As we see in (2.16), $\int_{\mathbb{R}^d \times \mathbb{R}^d} -x \cdot y \, d\gamma(x, y)$ is independent of Φ and Ψ, and $\int_{\mathbb{R}^d} -\Phi \, d\mu + \int_{\mathbb{R}^d} -\Psi \, d\nu$ is independent of γ. Hence, by utilizing Proposition 2.6.1, we have:

$$\begin{aligned} \inf_{\gamma \in \mathrm{Coup}(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} -x \cdot y \, d\gamma(x, y) &\geq \sup_{\Phi(x)+\Psi(y) \geq x \cdot y} \left\{ \int_{\mathbb{R}^d} -\Phi(x) \, d\mu(x) + \int_{\mathbb{R}^d} -\Psi(y) \, d\nu(y) \right\} \\ &\geq \sup_{\Phi \text{ convex}} \left\{ \int_{\mathbb{R}^d} -\Phi(x) \, d\mu(x) + \int_{\mathbb{R}^d} -\Phi^*(y) \, d\nu(y) \right\}. \end{aligned}$$

Thus,

$$\inf_{\gamma \in \mathrm{Coup}(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} -x \cdot y \, d\gamma(x, y) \geq \sup_{\Phi \text{ convex}} \left\{ \int_{\mathbb{R}^d} -\Phi(x) \, d\mu(x) + \int_{\mathbb{R}^d} -\Phi^*(y) \, d\nu(y) \right\}. \tag{2.17}$$

If we consider an optimal coupling $\tilde{\gamma} \in \mathrm{Coup}(\mu, \nu)$, by using Theorem 2.5.1, $\mathrm{supp}(\tilde{\gamma})$ is cyclically monotone.
Now, by Rockafellar's theorem, there exists a convex map $\Phi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ such that $\mathrm{supp}(\tilde{\gamma}) \subset \partial^- \Phi$, i.e., $y \in \partial^- \Phi(x) \;\; \forall (x, y) \in \mathrm{supp}(\tilde{\gamma})$.

So, by Proposition 2.6.1, for $\tilde{\gamma}$-a.e. (x,y):

$$\Phi(x) + \Phi^*(y) = x \cdot y.$$

Hence,

$$\begin{aligned}\int_{\mathbb{R}^d \times \mathbb{R}^d} -x \cdot y \, d\tilde{\gamma}(x,y) &= \int_{\mathbb{R}^d} -\Phi(x) \, d\tilde{\gamma}(x,y) + \int_{\mathbb{R}^d} -\Phi^*(y) \, d\tilde{\gamma}(x,y) \\ &= \int_{\mathbb{R}^d} -\Phi(x) \, d\mu(x) + \int_{\mathbb{R}^d} -\Phi^*(y) \, d\nu(y).\end{aligned}$$

Thus,

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} -x \cdot y \, d\tilde{\gamma}(x,y) = \int_{\mathbb{R}^d} -\Phi(x) \, d\mu(x) + \int_{\mathbb{R}^d} -\Phi^*(y) \, d\nu(y).$$

Since $\tilde{\gamma}$ is optimal, in inequality (2.16) we have equality.
Now we prove $(_*)$:
Φ is measurable. To show that Φ is integrable, we want to use Fubini's theorem. Also, we will show

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \Phi(x) \, d\gamma(x,y) = \int_{\mathbb{R}^d} \Phi(x) \, d\mu(x).$$

Claim: Φ is integrable.
We have $x \cdot y \leq \Phi(x) + \Psi(y)$ and also, by the fact that $(x+y)^2 \geq 0$, we have

$$x \cdot y \geq -\frac{|x|^2}{2} - \frac{|y|^2}{2},$$

therefore

$$\Phi(x) + \Psi(y) \geq x \cdot y \geq -\frac{|x|^2}{2} - \frac{|y|^2}{2}.$$

Thus,

$$\Phi(x) + \frac{|x|^2}{2} + \Psi(y) + \frac{|y|^2}{2} \geq 0 \quad \forall x, y \in \mathbb{R}^d.$$

Now we choose $x_0 \in \mathbb{R}^d$ such that $\Phi(x_0) < +\infty$ and also $y_0 \in \mathbb{R}^d$ with $\Psi(y_0) < +\infty$. If $\Phi = \Psi \equiv +\infty$, then (2.16) is true.
Thus,

$$\begin{aligned}\phi(x) &:= \Phi(x) + \frac{|x|^2}{2} \geq -C := -\Psi(y_0) - \frac{|y_0|^2}{2} \quad \forall x, \\ \psi(y) &:= \Psi(y) + \frac{|y|^2}{2} \geq -C' := -\Phi(x_0) - \frac{|x_0|^2}{2} \quad \forall y.\end{aligned}$$

We conclude

$$\phi(x) + C \geq 0 \quad \text{and} \quad \psi(y) + C' \geq 0.$$

Hence, they can be monotonically approximated with Borel and bounded functions:

$$\phi_n := \max\{\phi + C, n\} \quad \text{and} \quad \psi_n := \min\{\psi + C', n\} \quad \forall n \in \mathbb{N}.$$

Apply the definition of coupling to ϕ_n and ψ_n and let $n \to \infty$. By the monotone convergence theorem, we have:

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} (\phi(x) + C)\, d\gamma(x, y) = \int_{\mathbb{R}^d} (\phi(x) + C)\, d\mu(x),$$
$$\int_{\mathbb{R}^d \times \mathbb{R}^d} (\psi(y) + C')\, d\gamma(x, y) = \int_{\mathbb{R}^d} (\psi(y) + C')\, d\nu(y).$$

Since $\int |x|^2\, d\gamma = \int |x|^2\, d\mu < +\infty$ and $\int |y|^2\, d\gamma = \int |y|^2\, d\nu < +\infty$, we obtain:

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} (\phi(x) + C)\, d\gamma(x, y) - \int_{\mathbb{R}^d \times \mathbb{R}^d} \left(\frac{|x|^2}{2} + C \right) d\gamma(x, y)$$
$$= \int_{\mathbb{R}^d} (\phi(x) + C)\, d\mu(x) - \int_{\mathbb{R}^d} \left(\frac{|x|^2}{2} + C \right) d\mu(x).$$

Hence,

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \left(\phi(x) - \frac{|x|^2}{2} \right) d\gamma(x, y) = \int_{\mathbb{R}^d} \left(\phi(x) - \frac{|x|^2}{2} \right) d\mu(x).$$

Therefore, by the definition of $\phi(x)$, we get:

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \Phi(x)\, d\gamma(x, y) = \int_{\mathbb{R}^d} \Phi(x)\, d\mu(x);$$

as desired.
Similarly,

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \Psi(y)\, d\gamma(x, y) = \int_{\mathbb{R}^d} \Psi(y)\, d\nu(x).$$

□

Remark 2.6.1. Given $\phi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ convex and $\tilde{\gamma} \in \mathrm{Coup}(\mu, \nu)$. If $\mathrm{supp}(\tilde{\gamma}) \subset \partial^- \phi$, then $\tilde{\gamma}$ is optimal.

Proof. Assume that $\mathrm{supp}(\tilde{\gamma}) \subset \partial^- \phi$ where ϕ is convex. Take $(x, y) \in \mathrm{supp}(\tilde{\gamma})$, i.e., $y \in \partial^- \phi(x)$. Thus, $\phi(z) \geq \phi(x) + \langle y, z - x \rangle$, and then $\phi(z) - \phi(x) \geq y \cdot (z - x)$.
Hence, by Proposition 2.6.1, (ii), we have:

$$\phi(x) + \phi^*(y) = x \cdot y \quad \text{if} \quad y \in \partial^- \phi.$$

Then:

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} -x \cdot y\, d\tilde{\gamma} = \int_{\mathbb{R}^d} -\phi\, d\mu + \int_{\mathbb{R}^d} -\phi^*\, d\nu. \tag{2.18}$$

Since by (2.16), we have:

$$\int_{\mathbb{R}^d} -\phi(x)\, d\mu(x) + \int_{\mathbb{R}^d} -\phi^*(y)\, d\nu(y) \leq \inf_{\gamma \in \mathrm{Coup}(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} -x \cdot y\, d\gamma(x, y),$$

together with (2.18), we get:

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} -x\cdot y\, d\tilde{\gamma} = \int_{\mathbb{R}^d} -\phi\, d\mu + \int_{\mathbb{R}^d} -\phi^*\, d\nu \le \inf_{\gamma} \int_{\mathbb{R}^d\times\mathbb{R}^d} -x\cdot y\, d\gamma.$$

Thus

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} -x\cdot y\, d\tilde{\gamma} \le \inf_{\gamma\in\mathrm{Coup}(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} -x\cdot y\, d\gamma.$$

Therefore, $\tilde{\gamma}$ is optimal. □

Now, by Remark 2.6.1 and Theorem 2.5.1 and Theorem 2.5.2, we conclude:

Corollary 2.6.1. For the given cost function $C(x,y) = \frac{|x-y|^2}{2}$ or equivalently $C'(x,y) = -x\cdot y$, the following statements are equivalent:

(i) $\tilde{\gamma}$ is optimal;

(ii) there exists a convex function $\phi : \mathbb{R}^d \to \mathbb{R}\cup\{+\infty\}$ such that $\mathrm{supp}(\tilde{\gamma}) \subset \partial^-\phi$;

(iii) $\mathrm{supp}(\tilde{\gamma})$ is cyclically monotone.

Remark 2.6.2. If a certain transport map is optimal, then statements (i)-(iii) from the previous corollary are useful [See [1], Remark 2.5.9].

Now, we want to state the Kantorovich duality in the general case, i.e., for general cost functions. For this purpose, we need a suitable definition for the notion of convexity. Therefore, proofs follow the same structure as before.
Now we define C-convexity and C-concavity.
We can define a convex function as the supremum of affine functions:

$$\phi : \mathbb{R}^d \to \mathbb{R}\cup\{+\infty\}$$

is convex if it is defined by

$$\phi(x) := \sup_{y\in\mathbb{R}^d}\{x\cdot y + \alpha(y)\}$$

for some values $\{\alpha(y)\}_{y\in\mathbb{R}^d}$ where $\alpha(y)\in\mathbb{R}\cup\{-\infty\}$.
To define C-convex functions and C-subdifferentiability, we recall that we had $C(x,y) = -x\cdot y$. Hence $x\cdot y = -C(x,y)$.

Definition 2.6.2 (C-convex and C-concave function)**.** Assume that X and Y are metric spaces and $C : X\times Y \to \mathbb{R}$ is a cost function. If $\phi : X \to \mathbb{R}\cup\{+\infty\}$, then it is called C-convex if:

$$\phi(x) := \sup_{y\in Y}\{-C(x,y) + \alpha(y)\},$$

for some $\{\alpha(y)\}_{y\in Y} \subset \mathbb{R}\cup\{-\infty\}$.
If $\psi : Y \to \mathbb{R}\cup\{-\infty\}$, then it is called C-concave if:

$$\psi(y) := \inf_{x\in X}\{C(x,y) - \beta(x)\},$$

for some $\{\beta(x)\}_{x\in X} \subset \mathbb{R}\cup\{+\infty\}$.

Definition 2.6.3 (C-subdifferential and C-superdifferential)**.** Let X and Y be metric spaces and $C : X \times Y \to \mathbb{R}$ be the cost function. Assume that $\phi : X \to \mathbb{R} \cup \{+\infty\}$ is a C-convex function. Its C-subdifferential is defined by:

$$\partial_c\phi(x) := \{y \in Y | \phi(z) \geq -C(z, y) + C(x, y) + \phi(x) \quad \forall z \in X\}.$$
$$\partial_c\phi := \bigcup_{x \in X} \{x\} \times \partial_c\phi(x) \subset X \times Y.$$

Also for any C-concave function $\psi : X \to \mathbb{R} \cup \{-\infty\}$, its C-superdifferential is defined by:

$$\partial^c\psi(x) := \{y \in Y | \psi(z) \leq -C(z, y) + C(x, y) + \psi(x) \quad \forall z \in X\}.$$
$$\partial^c\psi := \bigcup_{x \in X} \{x\} \times \partial^c\psi(x) \subset X \times Y.$$

Theorem 2.6.2. Assume that X and Y are metric spaces. Consider a subset $\mathcal{M}$ of $X \times Y$. Then $\mathcal{M}$ is C-cyclically monotone if and only if there is a C-convex function ϕ such that $\mathcal{M} \subset \partial_c\phi$.

Proof. The proof of this theorem is similar to the proof of the Rockafellar theorem, except we replace $-x \cdot y$ with $C(x, y)$.
$\Rightarrow$) If $\mathcal{M}$ is C-cyclically monotone, then there exists a C-convex function ϕ such that $\mathcal{M} \subset \partial_c\phi$:
Since $\mathcal{M}$ is C-cyclically monotone in $X \times Y$, for any finite family of pairs $(x_i, y_i)_{i=1,\ldots,m} \subset \mathcal{M}$, we have:

$$-C(x_{m+1}, y_m) + \cdots + C(x_0, y_0) \leq 0.$$

It suffices to define:

$$\phi(x) := \sup_{m \in \mathbb{N}} \{-C(x, y_m) + C(x_m, y_m) - C(x_m, y_{m-1}) + \cdots + C(x_0, y_0) | (x_i, y_i)_{i=1,\ldots,m} \subset \mathcal{M}\}.$$

Now, by applying the proof of the Rockafellar theorem [See [4], 77-78], we obtain the result.
$\Leftarrow$) Assume that ϕ is a C-convex function such that $\mathcal{M} \subset \partial_c\phi$.We show that $\mathcal{M}$ is C-cyclically monotone.
Consider a finite family of points $(x_i, y_i)_{i=1,\ldots,m} \subset \mathcal{M} \subset \partial_c\phi$, meaning $y_i \in \partial_c\phi(x_i)$ for any $i = 1, ..., m$. By the definition of $\partial_c\phi(x)$:

$$\phi(z) \geq \phi(x_i) - C(z, y_i) + C(x_i, y_i) \quad \forall z \in X.$$

Set $z = x_{i+1}$ and $x_{m+1} = x_1$; then:

$$\phi(x_{i+1}) \geq \phi(x_i) - C(x_{i+1}, y_i) + C(x_i, y_i).$$

Summing over $i = 1, \dots, m$, we get:

$$\sum_{i=1}^{m} \phi(x_{i+1}) \geq \sum_{i=1}^{m} \phi(x_i) + \sum_{i=1}^{m} \left(-C(x_{i+1}, y_i) + C(x_i, y_i)\right).$$

Since $x_{m+1} = x_1$ and $\phi(x_{m+1}) = \phi(x_1)$, we have:

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

hence

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

Therefore, $\mathcal{M}$ is C-cyclically monotone. □

Now, we would like to present a general Kantorovich duality. First, we will define the C-Legendre transform.

Definition 2.6.4 (C-Legendre transform)**.** Assume that X and Y are locally compact, separable, and complete metric spaces. Let $\phi : X \to \mathbb{R} \cup \{+\infty\}$ be a C-convex function. Its C-Legendre transform is defined by:

$$\begin{aligned} &\phi_C : Y \to \mathbb{R} \cup \{+\infty\}, \\ &\phi_C(y) := \sup_{x \in X} \{-C(x, y) - \phi(x)\}. \end{aligned}$$

If $\phi : X \to \mathbb{R} \cup \{-\infty\}$ is a C-concave function, then its C-transform is defined by:

$$\begin{aligned} &\phi^C : Y \to \mathbb{R} \cup \{-\infty\}, \\ &\phi^C(y) := \inf_{x \in X} \{C(x, y) - \phi(x)\}. \end{aligned}$$

Proposition 2.6.2 (Relation between ϕ and ϕ_C)**.** For any given C-convex function $\phi : X \to \mathbb{R} \cup \{+\infty\}$, the following statements hold:

(i) $\phi(x) + \phi_C(y) + C(x, y) \geq 0$ for all $x \in X$ and $y \in Y$,

(ii) $\phi(x) + \phi_C(y) + C(x, y) = 0$ if and only if $y \in \partial_c \phi(x)$.

Proof. See [1], Proposition 2.6.4. □

Theorem 2.6.3 (General case of Kantorovich duality)**.** Assume that X and Y are locally compact, separable, and complete metric spaces, and consider a continuous cost function $C \in \mathcal{C}(X \times Y)$ which is bounded from below. If

$$\inf_{\gamma \in \mathrm{Coup}(\mu, \nu)} \int_{X \times Y} C \, d\gamma < +\infty,$$

then:

$$\min_{\gamma \in \mathrm{Coup}(\mu,\nu)} \int_{X \times Y} C\, d\gamma = \max_{\phi(x)+\psi(y)+C(x,y) \geq 0} \left\{ \int_X -\phi\, d\mu + \int_Y -\psi\, d\nu \right\}.$$

If we replace ψ by ϕ_C, then we have:

$$\min_{\gamma \in \mathrm{Coup}(\mu,\nu)} \int_{X \times Y} C\, d\gamma = \max_{\phi \text{ is } C\text{-convex}} \left\{ \int_X -\phi\, d\mu + \int_Y -\phi_C\, d\nu \right\}.$$

Proof. Same as the proof of Theorem 2.6.3. [See [1], Theorem 2.6.5 and [4], 78-79]. □

Now, we demonstrate the applicability of this theory to a special case where the cost function is the distance function in a metric space X.

Proposition 2.6.3. Assume that (X, d) is a metric space and $Y = X$. If $C(x, y) := d(x, y)$, then

(i) $\phi : X \to \mathbb{R} \cup \{+\infty\}$ is a C-convex function if and only if it is 1-Lipschitz continuous, i.e.,

$$|\phi(x) - \phi(y)| \leq d(x, y) \quad \forall x, y \in X.$$

(ii) For a C-convex function $\phi : X \to \mathbb{R} \cup \{+\infty\}$, if $y \in \partial_c \phi(x)$, then

$$\phi(y) - \phi(x) = d(x, y).$$

(iii) For a C-convex function ϕ, it holds that $\phi_C = -\phi$.

(iv) If μ and ν are probability measures in X with finite cost function, i.e.,

$$\inf_{\gamma \in \mathrm{Coup}(\mu,\nu)} \int_{X \times Y} d(x, y)\, d\gamma(x, y) < +\infty,$$

then

$$\min_{\gamma \in \mathrm{Coup}(\mu,\nu)} \int_{X \times Y} d(x, y)\, d\gamma(x, y) = \max_{\phi \text{ 1-Lip}} \left\{ \int_X \phi\, d\nu - \int_X \phi\, d\mu \right\}.$$

Proof. Proof. See [1], Proposition 2.6.6, 49-51. □

As a conclusion of this section, we state the following corollary:

Corollary 2.6.2. Assume that X and Y are locally compact, separable, and complete metric spaces. Let C be a continuous function in $X \times Y$ and bounded from below such that $\inf_{\gamma \in \mathrm{Coup}(\mu,\nu)} \int_{X \times Y} C\, d\gamma < +\infty$. Then for any coupling $\tilde{\gamma} \in \mathrm{Coup}(\mu, \nu)$, the following statements are equivalent:

(i) $\tilde{\gamma}$ is optimal;

(ii) $\mathrm{supp}(\tilde{\gamma})$ is C-cyclically monotone;

(iii) There is a C-convex function $\phi : X \to \mathbb{R} \cup \{+\infty\}$ such that

$$\mathrm{supp}(\tilde{\gamma}) \subset \partial_c \phi.$$

2.7 Existence and uniqueness of optimal transport maps

In this section, we state and prove Brenier's theorem in a special case where $X = Y = \mathbb{R}^d$ and the cost function $C(x,y) = \frac{|x-y|^2}{2}$. We aim to establish the existence and uniqueness of the optimal transport map under these conditions. Then, utilizing the proof of Brenier's theorem, we extend our analysis to the general case. We will demonstrate the existence of an optimal transport map that is unique, even in this broader context.

Theorem 2.7.1 (Brenier's theorem)**.** Assume that $X = Y = \mathbb{R}^d$ and $C(x,y) = \frac{|x-y|^2}{2}$ (or equivalently $C(x,y) = -x \cdot y$). Let

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

and suppose that μ is absolutely continuous with respect to the Lebesgue measure (i.e., $\mu \ll \ dx$). Then there exists an optimal coupling $\tilde{\gamma}$ which is unique.
Furthermore, for some convex function ϕ, we have

$$\tilde{\gamma} = (Id \times H)_{\#}\mu \qquad \text{and} \qquad H = \nabla\phi.$$

Proof. Step 1. Existence of optimal couping $\tilde{\gamma}$:
$C(x,y) = \frac{|x-y|^2}{2}$ is a continuous and nonnegative cost function, and the product measure $\mu \otimes \nu$ is a coupling. Furthermore, we observe that

$$\begin{aligned} \frac{|x-y|^2}{2} &= \frac{|x|^2}{2} + \frac{|y|^2}{2} - x \cdot y \leq \frac{|x|^2}{2} + \frac{|y|^2}{2} \\ \Rightarrow &|x-y|^2 \leq |x|^2 + |y|^2 \leq 2(|x|^2 + |y|^2). \end{aligned} \tag{2.19}$$

Hence, we get

$$\begin{aligned} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \, d(\mu \otimes \nu) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} (|x|^2 + |y|^2 - 2x \cdot y) \, d(\mu \otimes \nu) \\ &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} (|x|^2 + |y|^2) \, d(\mu \otimes \nu) \\ &\leq 2 \int_{\mathbb{R}^d \times \mathbb{R}^d} (|x|^2 + |y|^2) \, d(\mu \otimes \nu) \\ &= 2 \int_{\mathbb{R}^d} |x|^2 \, d\mu(x) + 2 \int_{\mathbb{R}^d} |y|^2 \, d\nu(y) \\ &< +\infty, \end{aligned}$$

where in the second inequality, we used (2.19) and the last inequality is true by our assumption.
Therefore, we get

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \, d(\mu \otimes \nu) < +\infty.$$

Hence, by Theorem 2.4.1, there exists a nontrivial optimal transport plan $\tilde{\gamma}$.
Step 2. ϕ is differentiable μ-a.e.
Since $\tilde{\gamma}$ is optimal, by using Corollary 2.6.1, we conclude that

$$\operatorname{supp}(\tilde{\gamma}) \subset \partial^- \phi \quad \text{for some convex function } \phi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}.$$

By Proposition 2.6.1, we have:

$$\phi(x) + \phi^*(y) = x \cdot y \quad \text{on} \quad \partial^- \phi$$

where

$$\phi^*(y) := \sup_{z \in \mathbb{R}^d} \{z \cdot y - \phi(z)\}.$$

Since $\operatorname{supp}(\tilde{\gamma}) \subset \partial^- \phi$, thus:

$$\phi(x) + \phi^*(y) = x \cdot y \quad \text{on} \quad \operatorname{supp}(\tilde{\gamma}). \tag{2.20}$$

Now we state that $(\phi(x), \phi^*(y))$ is finite for $\tilde{\gamma}$-almost everywhere (x, y), i.e., since $\int_{\mathbb{R}^d} |x|^2 \, d\mu + \int_{\mathbb{R}^d} |y|^2 \, d\nu < +\infty$. Thus,

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} x \cdot y \, d\tilde{\gamma}(x, y) < +\infty.$$

By (2.20), we have:

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} (\phi(x) + \phi^*(y)) \, d\tilde{\gamma}(x, y) < +\infty.$$

Then,

$$\int_{\mathbb{R}^d} \phi(x) \, d\mu(x) + \int_{\mathbb{R}^d} \phi^*(y) \, d\nu(y) < +\infty.$$

Hence, we conclude that $\phi(x)$ is finite at $\mu - a.e.$ point x.
Since $\mu \ll dx$ and by Alexandrov's theorem (see [8], Theorem 14.25), we know that convex functions are differentiable almost everywhere on the region where they are finite. Thus, since $\phi(x)$ is finite at $\mu - a.e.$ point x , ϕ is differentiable $\mu - a.e.$
Step 3. $\tilde{\gamma} = (Id \times H)_\# \mu$ and $H = \nabla \phi$.
We want to show that for any $f \in \mathcal{C}_b(\mathbb{R}^d \times \mathbb{R}^d)$,

$$\int f(x, y) \, d\tilde{\gamma}(x, y) = \int f(x, y) \, d((Id \times \nabla \phi)_\# \mu)(x, y).$$

Assume that $A \subset \mathbb{R}^d$ is a subset with $\mu(A) = 0$ and that ϕ is differentiable at $x \in \mathbb{R}^d \setminus A$.
Consider $\tilde{x} \in \mathbb{R}^d \setminus A$ and assume that $(\tilde{x}, \tilde{y}) \in \operatorname{supp}(\tilde{\gamma}) \subset \partial^- \phi$. Thus, $\tilde{y} \in \partial^- \phi(\tilde{x})$.
Since $\tilde{y} \in \partial^- \phi(\tilde{x})$ and ϕ is differentiable at $\tilde{x}$, by Lemma 2.5.1, which states that $\partial^- \phi(x) = \{\nabla \phi(x)\}$, we get:

$$\tilde{y} = \nabla \phi(\tilde{x}),$$

and therefore

$$H = \nabla \phi.$$

Thus,

$$\mathrm{supp}(\tilde{\gamma}) \cap [(\mathbb{R}^d \setminus A) \times \mathbb{R}] \subset \mathrm{graph}(\nabla\phi).$$

Since $\tilde{\gamma}(A \times \mathbb{R}^d) = \mu(A) = 0$, we have

$$(x, y) = (x, \nabla\phi(x)) \qquad \tilde{\gamma} - a.e. \tag{2.21}$$

Then for any $f \in \mathcal{C}_b(\mathbb{R}^d \times \mathbb{R}^d)$:

$$\begin{aligned}\int_{\mathbb{R}^d\times\mathbb{R}^d} f(x,y)\, d\tilde{\gamma}(x,y) &= \int_{\mathbb{R}^d\times\mathbb{R}^d} f(x, \nabla\phi(x))\, d\tilde{\gamma}(x,y) \\ &= \int_{\mathbb{R}^d} f(x, \nabla\phi(x))\, d\mu(x) \\ &= \int_{\mathbb{R}^d\times\mathbb{R}^d} f(x,y)\, d((Id \times \nabla\phi)_{\#}\mu)(x,y),\end{aligned}$$

where we used (2.21) in the first equality.
Therefore, we get $\tilde{\gamma} = (Id \times \nabla\phi)_{\#}\mu$ as desired.
Step 4. $\tilde{\gamma}$ is unique.
Assume that $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ are optimal couplings. We show that $\frac{\tilde{\gamma}_1+\tilde{\gamma}_2}{2}$ is an optimal coupling. We know that the problem is linear with convex constraints. Thus,

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2\, d\Big(\frac{\tilde{\gamma}_1+\tilde{\gamma}_2}{2}\Big) = \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2\, d\tilde{\gamma}_1 + \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2\, d\tilde{\gamma}_2.$$

Now consider $g \in \mathcal{C}_b(\mathbb{R}^d)$. Then

$$\begin{aligned}\int_{\mathbb{R}^d\times\mathbb{R}^d} g(x)\, d\Big(\frac{\tilde{\gamma}_1+\tilde{\gamma}_2}{2}\Big)(x,y) &= \frac{1}{2}\int_{\mathbb{R}^d} g(x)\, d\tilde{\gamma}_1(x,y) + \frac{1}{2}\int_{\mathbb{R}^d} g(x)\, d\tilde{\gamma}_2(x,y) \\ &= \frac{1}{2}\int_{\mathbb{R}^d} g(x)\, d\mu(x) + \frac{1}{2}\int_{\mathbb{R}^d} g(x)\, d\mu(x) \\ &= \int_{\mathbb{R}^d} g(x)\, d\mu(x).\end{aligned}$$

Thus,

$$(\pi_x)_{\#}\Big(\frac{\tilde{\gamma}_1+\tilde{\gamma}_2}{2}\Big) = \mu.$$

Similarly,

$$(\pi_y)_{\#}\Big(\frac{\tilde{\gamma}_1+\tilde{\gamma}_2}{2}\Big) = \nu.$$

Therefore, $\frac{\tilde{\gamma}_1+\tilde{\gamma}_2}{2}$ is an optimal coupling.
Now we apply step 2 and step 3 to $\tilde{\gamma}_1$; therefore, there exists a convex function ϕ_1 such that by step 2, it is differentiable $\mu - a.e.$, and by step 3, $\tilde{\gamma}_1 = (Id \times \nabla\phi_1)_{\#}\mu$. Hence, by (2.21), we have:

$$(x, y) = (x, \nabla\phi_1(x)) \quad \tilde{\gamma}_1 - a.e. \tag{2.22}$$

By repeating the process above for $\tilde{\gamma}_2$ and $\frac{\tilde{\gamma}_1+\tilde{\gamma}_2}{2}$, we obtain convex functions ϕ_2 and $\tilde{\phi}$ such that:

$$\tilde{\gamma}_2 = (Id \times \nabla\phi_2)_{\#}\mu.$$

Then by (2.21):

$$(x,y) = (x, \nabla\phi_2(x)) \quad \tilde{\gamma}_2 - a.e. \tag{2.23}$$

and

$$\frac{\tilde{\gamma}_1+\tilde{\gamma}_2}{2} = (Id \times \nabla\tilde{\phi})_{\#}\mu,$$

we use (2.21) to get:

$$(x,y) = (x, \nabla\tilde{\phi}) \quad \frac{\tilde{\gamma}_1+\tilde{\gamma}_2}{2} - a.e. \tag{2.24}$$

Therefore by (2.24),

$$(x,y) = (x, \nabla\tilde{\phi}) \quad \text{for } \tilde{\gamma}_1\text{-almost every } (x,y). \tag{2.25}$$

Now, by the fact that (2.22), (2.23), (2.24), and (2.25) don't depend on y, by using (2.22) and (2.25), we obtain:

$$(x,y) = (x, \nabla\tilde{\phi}(x)) \qquad \mu - a.e.$$
$$(x,y) = (x, \nabla\phi_1(x)) \qquad \mu - a.e.$$

Therefore,

$$\nabla\tilde{\phi}(x) = \nabla\phi_1(x) \quad \mu - a.e., \quad \forall x \in \mathbb{R}^d. \tag{2.26}$$

Similarly,

$$\nabla\phi_2(x) = \nabla\tilde{\phi}(x) \quad \mu - a.e., \quad \forall x \in \mathbb{R}^d. \tag{2.27}$$

Hence, (2.26) and (2.27) imply:

$$\nabla\phi_1(x) = \nabla\phi_2(x) \quad \mu - a.e.$$

Thus, $\tilde{\gamma}_1 = \tilde{\gamma}_2$, i.e., $\tilde{\gamma}$ is unique. $\square$

Now we demonstrate the existence of an optimal transport map in a special case that is a consequence of Brenier's theorem.

Corollary 2.7.1. Assume that $X = Y = \mathbb{R}^d$ and $C(x,y) = \frac{|x-y|^2}{2}$. Let $\int_{\mathbb{R}^d} |x|^2 d\mu(x) + \int_{\mathbb{R}^d} |y|^2 d\nu(y) < +\infty$. If $\mu \ll dx$, then

(i) There exists a unique optimal transport map $H : X \to Y$ such that $H_{\#}\mu = \nu$. Furthermore, there exists a convex function $\phi : X \to \mathbb{R} \cup \{+\infty\}$, $H = \nabla\phi$.

(ii) If for some convex function $\phi : X \to \mathbb{R} \cup \{+\infty\}$, $H = \nabla\phi$ and $H_{\#}\mu = \nu$, then H is the unique optimal transport map.

Proof. (i) By Brenier's theorem, there exists an optimal coupling $\tilde{\gamma}$ given by $(Id \times H)_{\#}\mu$ where $H : X \to Y$ is a transport map, i.e., $H_{\#}\mu = \nu$. Now we show that $H = \nabla\phi$, where $\phi : X \to Y$ is a convex function.
By using $C_M(\mu, \nu) \geq C_K(\mu, \nu)$ (See [1], Remark 2.2.1), we have

$$\inf_{H_{\#}\mu=\nu} \int_X |x - H(x)|^2 d\mu(x) \geq \inf_{\gamma \in \mathrm{Coup}(\mu,\nu)} \int_{X \times Y} |x - y|^2 d\gamma(x, y).$$

Assume that ϕ is a convex function such that $H = \nabla\phi$ and $\tilde{\gamma} = (Id \times H)_{\#}\mu$. Thus $\tilde{\gamma} = (Id \times \nabla\phi)_{\#}\mu$. Therefore

$$\begin{aligned}\int_X |x - H(x)|^2 d\mu(x) &= \int_X |x - \nabla\phi(x)|^2 d\mu(x) \\ &= \int_{X \times Y} |x - y|^2 d(Id \times \nabla\phi)_{\#}\mu \\ &= \int_{X \times Y} |x - y|^2 d\tilde{\gamma}(x, y) \\ &= \min_{\gamma \in \mathrm{Coup}(\mu,\nu)} \int_{X \times Y} |x - y|^2 d\gamma(x, y),\end{aligned}$$

where we used $H = \nabla\phi$ in the first equality and optimality of $\tilde{\gamma}$ in the last equality.
Therefore we get:

$$\int_X |x - H(x)|^2 d\mu(x) = \min_{\gamma \in \mathrm{Coup}(\mu,\nu)} \int_{X \times Y} |x - y|^2 d\gamma(x, y).$$

Hence $H = \nabla\phi$ is optimal.
Now we show that the solution of the Monge problem is unique:
For this purpose assume that H_1 and H_2 are optimal. Then by the previous part, we have

$$\gamma_1 = (Id \times H_1)_{\#}\mu \quad \text{and} \quad \gamma_2 = (Id \times H_2)_{\#}\mu$$

that are optimal for the Kantorovich problem.
To get the conclusion, by using the uniqueness of the optimal plan from Brenier's theorem, we get $H_1 = H_2$ μ-a.e. (Since $\gamma_1 = \gamma_2$).
(ii) By using Remark 2.6.2, we get the optimality and uniqueness follows from (i) (See [1], Corollary 2.5.12, pp. 38-39). □

Now the following question arises: what happens if $\nu \ll dx$ under the assumptions of Brenier's theorem?
To answer this, we establish the following corollary:

Corollary 2.7.2. Assume that $X = Y = \mathbb{R}^d$ and $C(x, y) = \frac{|x-y|^2}{2}$. Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and $\int_X |x|^2 d\mu + \int_Y |y|^2 d\nu < +\infty$. Suppose that $\mu \ll dx$ and $\nu \ll dx$. If $\nabla\phi$ is an optimal transport map from μ to ν and $\nabla\psi$ is an optimal transport map from ν to μ, then $\nabla\phi \circ \nabla\psi = \mathrm{Id}$ ν-a.e. and $\nabla\psi \circ \nabla\phi = \mathrm{Id}$ μ-a.e.

Proof. By Brenier's theorem, ϕ and ψ are convex functions such that $\nabla\phi$ is an optimal map from μ to ν, and $\nabla\psi$ is an optimal map from ν to μ.
Now we show that $\nabla\psi \circ \nabla\phi = \mathrm{Id}$ μ-a.e.

$$\begin{aligned}\int_X |x - \nabla\phi(x)|^2 \, d\mu(x) &= \int_{X\times Y} |x-y|^2 \, d(Id \times \nabla\phi)_{\#}\mu \\ &= \inf_{\gamma\in\mathrm{Coup}(\mu,\nu)} \int_{X\times Y} |x-y|^2 \, d\gamma(x,y),\end{aligned}$$

where we used the optimality of $\nabla\phi$ in the last equality.
We know that the cost function C is symmetric in x and y. Hence

$$\begin{aligned}\int_Y |\nabla\psi(y) - y|^2 \, d\nu(y) &= \int_{X\times Y} |x-y|^2 \, d((\nabla\psi \times \mathrm{Id})_{\#}\nu) \\ &= \inf_{\gamma\in\mathrm{Coup}(\mu,\nu)} \int_{X\times Y} |x-y|^2 \, d\gamma(x,y),\end{aligned}$$

where we used the optimality of $\nabla\psi$ in the last equality. Therefore, we get that $(Id \times \nabla\phi)_{\#}\mu$ and $(\nabla\psi \times Id)_{\#}\nu$ are optimal. Thus,

$$(Id \times \nabla\phi)_{\#}\mu = (\nabla\psi \times Id)_{\#}\nu. \tag{2.28}$$

Consider a test function $f : X \times Y \to \mathbb{R}$. Hence

$$(*) := \int_X f(x, \nabla\phi(x)) \, d\mu(x) = \int_X |x - \nabla\psi(\nabla\phi(x))|^2 \, d\mu(x) =: (**),$$

where we used $f(x, \nabla\phi(x)) = |x - \nabla\psi(\nabla\phi(x))|^2$ here, and

$$(*) = \int_Y f(\nabla\psi(y), y) \, d\nu(y) = \int_Y |\nabla\psi(y) - \nabla\psi(y)|^2 \, d\nu(y) = (**),$$

where we used (2.28) in the first equality.
Therefore,

$$\int_X |x - \nabla\psi(\nabla\phi(x))|^2 \, d\mu(x) = \int_Y |\nabla\psi(y) - \nabla\psi(y)|^2 \, d\nu(y) = 0,$$

and since $\nabla\psi(\nabla\phi(x)) = x$, we get

$$\nabla\psi \circ \nabla\phi = \mathrm{Id} \quad \mu\text{-a.e.}$$

Similarly, $\nabla\phi \circ \nabla\psi = \mathrm{Id} \quad \nu$-a.e. □

Now we show the existence and uniqueness of the optimal transport map for general cost functions. We use the proof of Brenier's theorem that includes the derivative of a convex function. Hence, we consider X with a differentiable structure. For simplicity, assume $X = Y = \mathbb{R}^d$. As we will show in Chapter 4, the same argument can be generalized for any arbitrary Riemannian manifold M.
In this case, Y does not need to be differentiable. For more general statements, refer to [6], Chapters 9-10.

Theorem 2.7.2. Assume that $X = Y = \mathbb{R}^d$ and $\mu \ll dx$ and $\mathrm{supp}(\nu)$ is compact. Let $C \in \mathcal{C}(X \times Y)$ be a continuous and bounded from below cost function with $\inf_{\gamma \in \mathrm{Coup}(\mu,\nu)} \int_{X \times Y} C \, d\gamma < +\infty$. If

(i) For all $y \in \mathrm{supp}(\nu)$, $x \mapsto C(x,y)$ is a differentiable map for $x \in X$;

(ii) For all $x \in X$, $y \mapsto \nabla_x C(x,y)$ is an injective map from $\mathrm{supp}(\nu)$ to Y,

(iii) For all $y \in \mathrm{supp}(\nu)$ and $r > 0$, $|\nabla_x C(x,y)| \leq M_r$ for any $x \in B_r$ (where M_r depends only on r and not on y).

Then there is a unique optimal plan $\tilde{\gamma}$ such that $\tilde{\gamma} = (\mathrm{Id} \times H)_{\#}\mu$ and H satisfies

$$\nabla_x C(x,y)|_{y=H(x)} + \nabla\phi(x) = \nabla_x C(x, H(x)) + \nabla\phi(x) = 0,$$

for some $\phi : X \to \mathbb{R} \cup \{+\infty\}$, a C-convex function.

Proof. See [1], Theorem 2.7.1, pp. 52-54. □

Remark 2.7.1. Theorem 2.7.2 covers Theorem 2.7.1 with one more assumption on the support of ν, which is compact. (See [1], Remark 2.7.2, pp. 52-53.)

Example 2.7.1. If $C(x,y) = |x-y|^n$ with $n > 1$, then Theorem 2.7.2 applies. See [1], Example 2.7.3.

Remark 2.7.2. Theorem 2.7.2 is not true for the cost function $C(x,y) = |x-y|$. (See [1], Remark 2.7.4 and [9], Proposition 2.7., and [10], Remark 2-19 (iii).)

3 Riemannian geometry

3.1 Riemannian geometry

In this chapter, we follow [1], Section 1.3, [12], Chapters 1 and 2, [13], Chapters 1 and 2, [14], Chapter 2, and [15], Section 1. Here, we aim to introduce some fundamental concepts of Riemannian geometry.
First, we construct the essential algebraic framework.

Definition 3.1.1 (Bilinear Forms)**.** Assume that V is a finite-dimensional vector space. We define a bilinear form on V by an $\mathbb{R}$-bilinear mapping $B : V \times V \to \mathbb{R}$, i.e.,
(i) $B(u+v, w) = B(u, w) + B(v, w)$ and $B(\lambda u, v) = \lambda B(u, v)$.
(ii) $B(u, v+w) = B(u, v) + B(u, w)$ and $B(u, \lambda v) = \lambda B(u, v)$ for all $u, v, w \in V$ and $\lambda \in \mathbb{R}$.

Definition 3.1.2 (Symmetric bilinear form)**.** A bilinear form $B : V \times V \to \mathbb{R}$ on a finite-dimensional vector space V is called symmetric if

$$B(u, v) = B(v, u) \quad \text{for all } u, v \in V.$$

Definition 3.1.3 (Definite bilinear form)**.** Let V be a finite-dimensional vector space and $B : V \times V \to \mathbb{R}$ be a symmetric bilinear form.
(i) If $B(v, v) > 0$ for all $v \in V$, B is called positive definite.
(ii) If $B(v, v) < 0$ for all $v \in V$, B is called negative definite.
We say B is definite if B is positive definite or negative definite. Alternatively, if not, B is called indefinite.

Definition 3.1.4 (Semidefinite bilinear form)**.** Let $B : V \times V \to \mathbb{R}$ be a symmetric bilinear form on a finite-dimensional vector space V. B is called positive (or negative) semidefinite if for all $v \in V$: $B(v, v) \geq 0$ (or $B(v, v) \leq 0$).

Definition 3.1.5 (Nondegenerate bilinear form)**.** A symmetric bilinear form $B : V \times V \to \mathbb{R}$ on a finite-dimensional vector space V is called nondegenerate if for all $v \in V$ $B(u, v) = 0$ implies $u = 0$.
If B is not nondegenerate, we say B is degenerate.

Definition 3.1.6 (Scalar product and inner product)**.** A nondegenerate symmetric bilinear form g on a finite-dimensional vector space V is called a scalar product.
A positive definite scalar product is called an inner product.

Definition 3.1.7 (Tangent space)**.** Let M be an n-dimensional $\mathcal{C}^\infty$ differentiable manifold, and assume that $p \in M$ is a point.
We define the tangent space T_pM of M at p by

$$T_pM := \{C'(0) \mid C : (-1, 1) \to M,\, C(0) = p\},$$

where C is a smooth curve.

Definition 3.1.8 (Tangent bundle)**.** Given an n-dimensional smooth manifold M. A tangent bundle consists of all tangent spaces for every point on the manifold M, organized in such a way that it itself forms a separate manifold.

$$TM := \bigcup_{p \in M} (\{p\} \times T_pM).$$

Definition 3.1.9 (Riemannian metric and Riemannian manifold)**.** Let M be an n-dimensional smooth manifold. A Riemannian metric is a symmetric positive definite scalar product $g_x : T_xM \times T_xM \to \mathbb{R}$, defined on each tangent space which depends on $x \in M$. $g = (g_x)_{x \in M}$ is called a Riemannian metric, and (M, g) is called a Riemannian manifold.

Note. A Riemannian metric defines a scalar product and a norm on each tangent space:

$$\langle u, v\rangle_x := g_x(u, v) \quad , \quad |v|_x := \sqrt{g_x(v, v)} \quad \forall u, v \in M.$$

Now we would like to construct a basis of T_xM:
Assume that U is an open subset of $\mathbb{R}^n$ and $\phi : U \to \phi(U) = V \subset M$ is a chart. For a given $x = \phi(x^1, \ldots, x^n) \in V$, the vectors

$$\frac{\partial}{\partial x^i} := \frac{\partial \phi}{\partial x_i}(x^1, \ldots, x^n) \quad \text{for} \quad i = 1, \ldots, n$$

form a basis of T_xM:
Let $v \in T_xM.v$ can be written as

$$v = \sum_{i=1}^{n} v^i \frac{\partial}{\partial x^i}.$$

By using this chart, we can express the metric g in coordinates inside V:

$$g_x(v, v) = \sum_{i,j=1}^{n} g_x \left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) v^i v^j = \sum_{i,j=1}^{n} g_{ij}(x) v^i v^j,$$

where

$$g_{ij}(x) := g_x \left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right).$$

Note that g^{ij} denotes the coordinates of the inverse of g; i.e., $g^{ij} = (g_{ij})^{-1}$.

More precisely,

$$\sum_j g^{ij} g_{jk} = \delta^i_k,$$

where δ^i_k denotes Kronecker's delta, i.e.,

$$\delta^i_k = \begin{cases} 1 & \text{if } i = k; \\ 0 & \text{if } i \neq k. \end{cases}$$

Definition 3.1.10 (Length)**.** Assume that M is a Riemannian manifold, and let $C : [a,b] \to M$ be a smooth curve from x to y with $C(a) = x$ and $C(b) = y$. Then $C'(t) \in T_{C(t)}M$. The length of C with respect to the Riemannian structure is given by

$$L(C) := \int_a^b |C'(t)|\, dt. \tag{3.1}$$

Note. The length of a curve remains invariant under reparametrization (see [12], Lemma 2.3.2), and for its definition, we only need to compute the g-norm of vectors tangent to M.
Note. In formula (3.1),

$$|C'(t)| := g(C'(t), C'(t))^{\frac{1}{2}}.$$

Definition 3.1.11 (Riemannian distance)**.** Let (M, g) be a Riemannian manifold and $x, y \in M$. We define the Riemannian distance by using the notion of norm of a tangent vector, given by

$$d(x,y) := \inf \left\{ \int_a^b |C'(t)|\, dt \;\middle|\; C : [a,b] \to M,\ C(a) = x,\ C(b) = y \right\}.$$

Definition 3.1.12 (Riemannian volume)**.** A Riemannian manifold (M, g) possesses an inherent reference measure known as the Riemannian volume:

$$\mathrm{vol}(dx) := \sqrt{\det(g_{ij})}\, dx^1 \dots dx^n.$$

This definition allows us to write a change of variable formula, exactly as in $\mathbb{R}^n$ (see [8], Chapter 1).

Definition 3.1.13 (Differential and gradients)**.** Let (M, g) be a Riemannian manifold and $\phi : M \to \mathbb{R}$ be a smooth map. Its differential $d\phi : TM \to \mathbb{R}$ is defined by

$$d\phi(x) \cdot v := \frac{d}{dt}\Big|_{t=0} \phi(C(t)),$$

where $C : (-a, a) \to M$ is any smooth curve such that $C(0) = x$ and $C'(0) = v$.
Note that this definition is independent of the choice of C.
Now, by using the Riemannian metric, we can define the gradient of ϕ, denoted by $\nabla\phi$, at any point $x \in M$ by

$$\langle \nabla\phi(x), v \rangle_x := d\phi(x) \cdot v.$$

By using coordinates induced by a chart, we obtain

$$g_{ij} (\nabla\phi)^i v^j = (d\phi)_i v^j \Rightarrow (\nabla\phi)^i = g^{ij} (d\phi)_j.$$

Definition 3.1.14 (Minimizing geodesic)**.** Let (M, g) be a Riemannian manifold. Assume that $C : [a, b] \to M$ is a smooth curve with constant speed, i.e., $|C'|$ is constant, and $C(a) = x$, $C(b) = y$ with $L(C) = d(x, y)$. C is called a minimizing geodesic.

Definition 3.1.15 (Geodesic)**.** Assume that (M, g) is a Riemannian manifold. We say that a curve $C : [a, b] \to M$ is a geodesic if its tangent vector C' is parallel along C. Equivalently, $C'' = 0$, i.e., C has constant speed.

Lemma 3.1.1 (Uniqueness of geodesics)**.** Let I be an interval and $C, \tilde{C} : I \to M$ be geodesics. If there exists $t \in I$ such that $C(t) = \tilde{C}(t)$ and $C'(t) = \tilde{C}'(t)$, then $C = \tilde{C}$.

Proof. See [13], Lemma 2.1.5. □

Definition 3.1.16 (Exponential map)**.** Assume that (M, g) is a Riemannian manifold. For any $p \in M$, let the set $D(p)$ be an open subset of T_pM defined by

$$D(p) := \{v \in T_pM \mid C_v(1) \text{ is defined}\},$$

where C_v is the unique maximal geodesic with initial condition $C_v(0) = p$ and $C_v'(0) = v$. We define the exponential map by

$$\begin{aligned} \exp_p &: D(p) \to M; \\ \exp_p(v) &:= C_v(1). \end{aligned}$$

4 Optimal transport on Riemannian manifolds

In this chapter, we follow [15], Section 2 and [17], Section 1, and [16], Chapter 3. Here, we introduce optimal transport on Riemannian manifolds. Next, we seek a solution to the Monge problem and ultimately characterize it.
Throughout this chapter, (M, g) denotes a smooth connected compact Riemannian manifold without boundary of class $\mathcal{C}^3$, i.e., $g \in \mathcal{C}^2$, where the metric tensor components $g_{ij}(x)$ are twice continuously differentiable functions of local coordinates.

4.1 The Monge Problem

Assume that $C : M \times M \to [0, +\infty)$ is a cost function and μ, ν are two probability measures on M, i.e., $\mu(M) = \nu(M) = 1$ for all $\mu, \nu \in \mathcal{P}(M)$. Consider measurable maps $H : M \to M$ pushing forward μ to ν, i.e., $H_{\#}\mu = \nu$, satisfying

$$H_{\#}\mu(A) = \mu(H^{-1}(A)) = \nu(A), \quad \forall A \in \mathcal{B}(M). \tag{4.1}$$

Alternatively, the Monge optimal transport problem from μ to ν with respect to the cost C involves minimizing the transportation cost among all H:

$$\inf_{H_{\#}\mu=\nu} \int_M C(x, H(x))\, d\mu(x). \tag{4.2}$$

Such maps $H : M \to M$ with $H_{\#}\mu = \nu$ are called transport maps from μ to ν.
We set

$$C_M(\mu, \nu) := \inf\Big\{ \int_M C(x, H(x))\, d\mu(x) \Big| H_{\#}\mu = \nu \Big\}. \tag{4.3}$$

Remark 4.1.1. Property (4.1) is equivalent to

$$\int_M f(H(x))\, d\mu(x) = \int_M f(y)\, d\nu(y),$$

for all $f : M \to \mathbb{R}$ ν-integrable.
Due to the highly nonlinear constraint $H_{\#}\mu = \nu$, the Monge optimal transport problem poses significant difficulty from an optimization viewpoint. Therefore, we will explore a concept of weak solution for this problem.

4.2 The Kantorovich relaxation

Given two probability measures $\mu, \nu \in \mathcal{P}(M)$, let $\gamma \in \mathcal{P}(M \times M)$ be a probability measure on the product space $M \times M$ such that

$$\pi_{x\#}\gamma = \mu \quad \text{and} \quad \pi_{y\#}\gamma = \nu, \tag{4.4}$$

where $\pi_x, \pi_y : M \times M \to M$ denote the projections on the first and second variables, respectively.
Consider the cost function $C : M \times M \to [0, \infty)$.
The Kantorovich optimal transport problem with respect to C involves minimizing

$$\int_{M\times M} C(x,y)\, d\gamma(x,y), \tag{4.5}$$

among all $\gamma \in \mathcal{P}(M \times M)$.
Any measure $\gamma \in \mathcal{P}(M \times M)$ is called a transport plan between μ and ν, denoted by $P(\mu, \nu)$.
We set

$$C_K(\mu,\nu) := \inf\left\{\int_{M\times M} C(x,y)\, d\gamma(x,y) \middle| \gamma \in P(\mu,\nu)\right\}. \tag{4.6}$$

Remark 4.2.1. For any measurable set A in M, the property (4.4) is equivalent to

$$\mu(A) = \gamma(A \times M) \quad \text{and} \quad \nu(A) = \gamma(M \times A).$$

Equivalently,

$$\int_{M\times M} (f(x) + g(y))\, d\gamma(x,y) = \int_M f(x)\, d\mu(x) + \int_M g(y)\, d\nu(y),$$

for all μ-integrable functions f and ν-integrable functions g.

Remark 4.2.2. Any transport map $H : M \to M$ from μ to ν induces a transport plan between μ and ν given by

$$\gamma := (Id \times H)_{\#}\mu,$$

where $\mu, \nu \in \mathcal{P}(M)$.

4.3 Optimal plans and Kantorovich potentials

4.3.1 Optimal plans

In this section, our aim is to demonstrate the existence of a solution to the Kantorovich optimal transport problem and establish specific properties of the optimal transport plans.

Theorem 4.3.1. Assume that μ and ν are two probability measures on M. Let $C : M \times M \to [0, +\infty)$ be a continuous cost function and $\mathrm{supp}(\mu)$ and $\mathrm{supp}(\nu)$ be compact. Then there exists at least one optimal transport plan $\tilde{\gamma} \in P(\mu, \nu)$ such that

$$C(\tilde{\gamma}) = \inf \left\{ \int_{M \times M} C(x, y) \, d\gamma(x, y) \,\middle|\, \gamma \in P(\mu, \nu) \right\},$$

i.e., $\tilde{\gamma}$ is a solution of the Kantorovich optimal transport problem.

Proof. Refer to [16], Theorem 3.2.1. □

Now, we recall the concept of C-cyclically monotone sets, which represents a specific property of the support of measures.

Definition 4.3.1 (C-Cyclically Monotone Set)**.** A subset $\mathcal{A} \subset M \times M$ is called C-cyclically monotone if, for every finite collection of points $(x_i, y_i) \in \mathcal{A}$, $i = 1, \dots, n$, the following condition is satisfied:

$$\sum_{i=1}^{n} C(x_i, y_i) \leq \sum_{i=1}^{n} C(x_{i+1}, y_i),$$

with $x_{n+1} \equiv x_1$.

Now, we demonstrate that optimal transport plans always possess C-cyclically monotone supports.

Theorem 4.3.2. Given two probability measures μ and ν on M with compact supports, and assuming the cost function $C : M \times M \to [0, +\infty)$ is continuous, then there exists a C-cyclically monotone compact set $\mathcal{A}$ such that the support of any optimal transport plan between μ and ν is contained in $\mathcal{A}$.

Proof. See [16], Theorem 3.2.5. □

4.3.2 Kantorovich potentials

The purpose of this section is to offer a more analytical description of C-cyclically monotone sets.

Definition 4.3.2 (C-transform)**.** Let $\phi : M \to \mathbb{R} \cup \{\pm\infty\}$. Its C-transform is the function $\phi^C : M \to \mathbb{R} \cup \{\pm\infty\}$ defined by

$$\phi^C(y) := \inf_{x \in M} \{C(x, y) - \phi(x)\}.$$

Definition 4.3.3 (C-concave function)**.** A function $\psi : M \to \mathbb{R} \cup \{-\infty\}$ is called C-concave if $\psi \not\equiv -\infty$ and there exists $\phi : M \to \mathbb{R} \cup \{-\infty\}$ such that $\psi = \phi^C$, i.e.,

$$\psi(y) := \inf_{x \in M} \{C(x, y) - \phi(x)\}.$$

Note 4.3.1. Observe that $\psi : M \to \mathbb{R} \cup \{-\infty\}$ is C-concave if and only if $\psi^{CC} = \psi$. This is the consequence of the fact that for any function $\phi : M \to \mathbb{R} \cup \{\pm\infty\}$, it holds that $\phi^C = \phi^{CCC}$:

$$\begin{aligned}
\phi^{CCC}(x) = (\phi^{CC})^C(x) &= \inf_{\tilde{y}\in M}\{C(x,\tilde{y}) - \phi^{CC}(\tilde{y})\} \\
&= \inf_{\tilde{y}\in M}\{C(x,\tilde{y}) - \{\inf_{\tilde{x}\in M}\{C(\tilde{x},\tilde{y}) - \phi^C(\tilde{x})\}\}\} \\
&= \inf_{\tilde{y}\in M}\{C(x,\tilde{y}) - \{\inf_{\tilde{x}\in M}\{C(\tilde{x},\tilde{y}) - \inf_{y\in M}\{C(\tilde{x},y) - \phi(y)\}\}\}\} \\
&= \inf_{\tilde{y}\in M}\sup_{\tilde{x}\in M}\inf_{y\in M}\{C(x,\tilde{y}) - C(\tilde{x},\tilde{y}) + C(\tilde{x},y) - \phi(y)\}.
\end{aligned}$$

If $x = \tilde{x}$, then we obtain $\phi^{CCC} \geq \phi^C$ and by setting $y = \tilde{y}$, we get $\phi^{CCC} \leq \phi^C$. Therefore, we conclude that $\phi^{CCC} = \phi^C$, and consequently $\psi^{CC} = \psi$.

The pair (ϕ, ϕ^C) is called a C-pair of potentials.
The following result indicates that the reverse of a C-concave function equals the C-transform of the reverse of its C-transform.

Proposition 4.3.1. Let $\phi : M \to \mathbb{R} \cup \{-\infty\}$ with $\phi \not\equiv -\infty$ be a C-concave function. Then the function $-\phi^C$ is C-concave, and

$$\phi(y) = \inf\{\phi^C(x) - C(x,y) \mid x \in M\} \quad \forall y \in M.$$

Proof. See [16], Proposition 3.2.8. □

The C-cyclically monotone sets are those sets that lie within the C-superdifferential of C-concave functions.
Now, let's define the C-superdifferential of C-concave functions.

Definition 4.3.4 (C-superdifferential)**.** Let $\phi : M \to \mathbb{R}\cup\{-\infty\}$ be a C-concave function. For every $x \in M$, its C-superdifferential at x is defined by

$$\partial^C\phi(x) := \{y \in M \mid \phi^C(y) = \phi(x) + C(x,y)\}.$$

The contact set of the pair (ϕ, ϕ^C) is defined by

$$\partial^C(\phi) := \{(x,y) \in M \times M \mid y \in \partial^C\phi(x)\}.$$

The C-subdifferential of a C-convex function can be defined analogously.

It is obvious that differentiability implies both C-subdifferentiability and C-superdifferentiability. Similarly, the converse holds true, as demonstrated by the following statement.

Proposition 4.3.2 (C-subdifferentiability and C-superdifferentiability imply differentiability)**.** Assume that O is an open subset of a smooth Riemannian manifold M, and let $\phi : O \to \mathbb{R}$ be a function. Then ϕ is differentiable at x if and only if it is both C-subdifferentiable and C-superdifferentiable there for $C : M \times M \to [0,\infty)$; and then

$$\partial^C\phi(x) = \partial_C\phi(x) = \{\partial\phi(x)\}.$$

Proof. See [4], Proposition 10.7. □

4.4 Kantorovich duality problem

In this section, we want to prove the existence and uniqueness of the solution to Monge's transportation problem and characterize it geometrically. For this purpose, we first define the infimal convolution of some C-concave function with C, and then state a dual problem of Kantorovich type on M.
Assume that (M, g) is a connected, compact, and $\mathcal{C}^3$-smooth Riemannian manifold without boundary. Let $d(x, y)$ denote the geodesic distance between $x, y \in M$, and let $d\text{vol}(x)$ denote the volume element. Set $C(x, y) = \frac{d^2(x,y)}{2}$.

Definition 4.4.1 (Infimal convolution)**.** Let $\phi : M \to \mathbb{R} \cup \{\pm\infty\}$ be a function. Its infimal convolution ϕ^C with C is defined by

$$\phi^C(y) := \inf_{x \in M} \{C(x, y) - \phi(x)\}. \tag{4.7}$$

Note that the infimal convolution, i.e., ϕ^C, acts as an involution $\phi^{CC} := (\phi^C)^C = \phi$ if and only if ϕ is itself an infimal convolution of some function with C (ϕ is C-concave).

Now we aim to state the dual problem of Kantorovich.
Let μ and ν be finite and nonnegative measures on M with the same total mass, i.e., $\mu(M) = \nu(M)$. Without loss of generality, normalize them so that $\mu(M) = \nu(M) = 1$, i.e., μ and ν are Borel probability measures on M.
The dual problem of Kantorovich is given by

$$\sup_{(\phi,\psi) \in Lip_C} I(\phi, \psi) = \inf_{H_\# \mu = \nu} \int_M C(x, H(x))\, d\mu(x) \tag{4.8}$$

where

$$I(\phi, \psi) = \int_M \phi(x)\, d\mu(x) + \int_M \psi(y)\, d\nu(y) \tag{4.9}$$

is a linear functional defined on a convex subset

$$Lip_C := \{(\phi, \psi) : M \to \mathbb{R} \text{ continuous} \mid \phi(x) + \psi(y) \leq C(x, y)\} \tag{4.10}$$

of continuous functions.
Now, we aim to demonstrate that the supremum in the Kantorovich duality formula is achieved by a function $\phi = \phi^{CC}$ (with $\psi = \phi^C$), which constructs a map minimizing Monge's cost. To proceed, we first require some preliminaries.

Definition 4.4.2 (Lipschitz continuity)**.** Let X and Y be two metric spaces with metrics d_X and d_Y, respectively. A function $f : X \to Y$ is called Lipschitz continuous if there exists a real constant $L \geq 0$ such that for all $x_1, x_2 \in X$:

$$d_Y(f(x_1), f(x_2)) \leq L d_X(x_1, x_2).$$

L is called the Lipschitz constant.

Lemma 4.4.1 (Lipschitz cost)**.** Assume that (M, d) is a metric space with diameter

$$\mathrm{diam}(M) := \sup\{d(x,z) \mid x, z \in M\}.$$

Let $\mathrm{diam}(M) < +\infty$. Then the function $\phi(x) = \frac{d^2(x,y)}{2}$ for any $y \in M$ is Lipschitz continuous, namely,

$$|\phi(x) - \phi(z)| \leq \mathrm{diam}(M) d(x,z). \tag{4.11}$$

Proof. Set $\psi(x) := d(x,y)$. By the triangle inequality, we have:

$$|\psi(x) - \psi(z)| = |d(x,y) - d(z,y)| \leq d(x,z) \quad \forall x, z \in M,$$

i.e., $\psi(x)$ has Lipschitz constant 1.
Also, $\psi(x) = d(x,y)$ is bounded:

$$\psi(x) = d(x,y) \leq \sup\{d(x,y) \mid x, y \in M\} = \mathrm{diam}(M) < +\infty.$$

Set $\phi(x) = \frac{\psi^2(x)}{2}$. Thus,

$$\begin{aligned}
2|\phi(x) - \phi(z)| &= 2\left|\frac{\psi^2(x)}{2} - \frac{\psi^2(z)}{2}\right| = |\psi^2(x) - \psi^2(z)| \\
&= |(\psi(x) - \psi(z))(\psi(x) + \psi(z))| \\
&= |\psi(x)(\psi(x) - \psi(z)) + \psi(z)(\psi(x) - \psi(z))| \\
&\leq |\psi(x)||\psi(x) - \psi(z)| + |\psi(z)||\psi(x) - \psi(z)| \\
&\leq \mathrm{diam}(M) d(x,z) + \mathrm{diam}(M) d(x,z) \\
&= 2\mathrm{diam}(M) d(x,z).
\end{aligned}$$

Therefore,

$$|\phi(x) - \phi(z)| \leq \mathrm{diam}(M) d(x,z).$$

□

Lemma 4.4.2 (Infimal convolutions are Lipschitz)**.** Let (M, d) be a metric space with finite diameter, i.e., $\mathrm{diam}(M) < +\infty$. Then any infimal convolution $\phi : M \to \mathbb{R} \cup \{\pm\infty\}$ given by $\phi = \phi^{CC}$ with $C(x,y) = \frac{d^2(x,y)}{2}$ is either identically infinite, i.e., $\phi = \pm\infty$, or Lipschitz continuous throughout M, and (4.11) is satisfied.

Proof. First consider the case $\phi = \pm\infty$.
Consider $\psi : M \to \mathbb{R} \cup \{\pm\infty\}$ such that $\phi = \psi^C$, i.e.,

$$\phi(x) = \inf_{y \in M} \{C(x,y) - \psi(y)\}. \tag{4.12}$$

Since $C(x,y) = \frac{d^2(x,y)}{2}$, hence $0 \leq C(x,y) \leq \frac{\mathrm{diam}^2(M)}{2}$ is bounded.
If ψ is unbounded from above, then since $C(x,y)$ is bounded, we get $\phi = -\infty$ in (4.12).
If not, ϕ is bounded from below.

Fix $z \in M$. If $\psi := -\infty$ everywhere, then $\phi(z) = +\infty$ in (4.12). Therefore, $\phi = +\infty$. Now we assume that ϕ is finite everywhere. We show that ϕ is Lipschitz continuous. Given $\varepsilon > 0$, there exists $y \in M$ such that

$$\phi(z) + \varepsilon \geq C(z, y) - \psi(y). \tag{4.13}$$

By (4.12), we have

$$\phi(x) \leq C(x, y) - \psi(y). \tag{4.14}$$

By subtracting (4.13) and (4.14), we obtain

$$\begin{aligned}\phi(x) - \phi(z) &\leq C(x, y) - C(z, y) + \varepsilon \\ &\leq \operatorname{diam}(M) d(x, z) + \varepsilon,\end{aligned}$$

where we used Lemma 4.4.1 in the last inequality. Therefore,

$$\phi(x) - \phi(z) \leq \operatorname{diam}(M) d(x, z) + \varepsilon$$

holds for arbitrary $\varepsilon > 0$, thus ϕ is Lipschitz and (4.11) is satisfied. □

Now we demonstrate that the supremum in the Kantorovich duality problem (4.8) is achieved by a function $\phi : M \to \mathbb{R}$, $\phi = \phi^{CC}$ (with $\psi = \phi^C$).

Proposition 4.4.1 (Dual potentials)**.** Let (M, d) be a separable metric space with finite diameter. Assume that μ and ν are Borel probability measures on M and $C(x, y) = \frac{d^2(x,y)}{2}$. Then the supremum in (4.8) is attained by some $\phi : M \to \mathbb{R}$ satisfying $\phi = \phi^{CC}$ and its infimal convolution $\psi = \phi^C$ with C.

Proof. Claim: If $(f, g) \in \mathrm{Lip}_C$, then $(g^C, g) \in \mathrm{Lip}_C$ and $I(f, g) \leq I(g^C, g)$. Moreover, $(g^C, g^{CC}) \in \mathrm{Lip}_C$ and $I(f, g) \leq I(g^C, g^{CC})$:
Proof of claim: Fix $(f, g) \in \mathrm{Lip}_C$, i.e., $f, g : M \to \mathbb{R}$ are continuous and $f(x) + g(y) \leq C(x, y)$. Therefore,

$$C(x, y) - f(x) \geq g(y) > -\infty.$$

Hence the infimal convolution $f^C(y) = \inf_{x \in M}\{C(x, y) - f(x)\} = g(y)$ must be finite:

$$f^C(y) \geq C(x, y) - f(x) \geq g(y).$$

Therefore,

$$f^C(y) \geq g(y), \tag{4.15}$$

hence by Lemma 4.4.2, Lipschitz continuous.
Now by (4.15) and (4.10), we have
(4.15): $f^C(y) = \inf_{x \in M}\{C(x, y) - f(x)\}$ Lipschitz continuous.

(4.10): $(f, f^C) : M \to \mathbb{R}$ continuous and $f(x) + f^C(y) \leq C(x,y)$. Therefore $(f, f^C) \in \mathrm{Lip}_C$. We have $f^C \geq g$. Hence

$$\begin{aligned} I(f,g) &= \int_M f(x)\, d\mu(x) + \int_M g(y)\, d\nu(y) \\ &\leq \int_M f(x)\, d\mu(x) + \int_M f^C(y)\, d\nu(y) \\ &= I(f, f^C). \end{aligned}$$

Therefore, if $g \leq f^C$, then $I(f,g) \leq I(f, f^C)$. Symmetry under $f \leftrightarrow g$ demonstrates that the first part of the claim is proved. Now we apply the first part of the claim to (g^C, g) and obtain $(g^C, g^{CC}) \in \mathrm{Lip}_C$ and $I(g^C, g) \leq I(g^C, g^{CC})$, and the claim is proved.
Now we complete the proof of the proposition.
Choose a sequence $(\phi_n, \psi_n) \in \mathrm{Lip}_C$ such that $I(\phi_n, \psi_n) \to \sup\{I(\phi, \psi) \,|\, (\phi, \psi) \in \mathrm{Lip}_C\}$.
By the claim, we conclude that the sequence (ψ_n^C, ψ_n^{CC}) also maximizes I on Lip_C.
Fix $z \in M$. Define a sequence

$$(f_n, g_n) := (\psi_n^C - \alpha_n, \psi_n^{CC} + \alpha_n) \in \mathrm{Lip}_C$$

which maximizes I:
We have $(\psi_n^C, \psi_n^{CC}) \in \mathrm{Lip}_C$ maximizes I on Lip_C, i.e.,

$$I(\phi_n, \psi_n) \leq I(\psi_n^C, \psi_n^{CC}).$$

Since $\mu(M) = \nu(M) = 1$; then for $(f_n, g_n) := (\psi_n^C - \alpha_n, \psi_n^{CC} + \alpha_n)$ we have:

$$\begin{aligned} I(\psi_n^C - \alpha_n, \psi_n^{CC} + \alpha_n) &= \int_M (\psi_n^C - \alpha_n)\, d\mu + \int_M (\psi_n^{CC} + \alpha_n)\, d\nu \\ &= \int_M \psi_n^C\, d\mu - \alpha_n \mu(M) + \int_M \psi_n^{CC}\, d\nu + \alpha_n \nu(M) \\ &= \int_M \psi_n^C\, d\mu + \int_M \psi_n^{CC}\, d\nu \\ &= I(\psi_n^C, \psi_n^{CC}). \end{aligned}$$

Here the normalization constants $\alpha_n := \psi_n^C(z)$ are chosen to make $f_n(z) = 0$.
Since $f_n = \psi_n^C - \alpha_n$ and $g_n = \psi_n^C + \alpha_n$ are infimal convolutions, by Lemma 4.4.2, they are Lipschitz continuous and uniformly bounded:

$$|f_n(x)| = |f_n(x) - f_n(z)| \leq \mathrm{diam}(M) d(x,z) \leq \mathrm{diam}^2(M). \tag{4.16}$$

$$\begin{aligned} |g_n(y)| = |f_n^C(y)| &\leq |C(x,y) - f_n(x)| \\ &\leq C(x,y) + |f_n(x)| \\ &\leq \frac{\mathrm{diam}^2(M)}{2} + \mathrm{diam}^2(M) \\ &= \frac{3\mathrm{diam}^2(M)}{2}, \end{aligned}$$

where we used (4.16) in the last inequality.
Hence, since (f_n, g_n) is a uniformly bounded Lipschitz continuous sequence that converges to (f, g), by the Arzelà-Ascoli theorem (see below), there exists a subsequence $(f_{n_k}, g_{n_k}) := (f_n, g_n)$ that converges pointwise to (f, g). Since μ and ν are finite measures, by the Dominated Convergence Theorem, we have

$$\int_M f_n(x)\,d\mu(x) + \int_M g_n(y)\,d\nu(y) \to \int_M f(x)\,d\mu(x) + \int_M g(y)\,d\nu(y).$$

Therefore, $I(f_n, g_n) \to I(f, g)$, i.e., $\lim_{n\to\infty} I(f_n, g_n) = I(f, g)$. Thus, the supremum is obtained at (f, g) and therefore by the claim, at $(\phi, \psi) := (g^C, g^{CC})$. By using Note 4.3.1, i.e., $g^{CCC} = g^C$, we conclude that $\phi^{CC} = \phi$. □

Theorem 4.4.1 (Arzelà-Ascoli)**.** If $f_n : [a, b] \to \mathbb{R}$ is a sequence of uniformly bounded functions such that each f_n is Lipschitz continuous with the same Lipschitz constant L, i.e.,

$$|f_n(x) - f_n(y)| \leq L|x - y| \quad \forall x, y \in [a, b],$$

then there exists a subsequence that converges uniformly on $[a, b]$.

Proof. See [18], Theorem 2. □

Theorem 4.4.2 (Dominated Convergence Theorem)**.** Assume that $(f_n)_{n\in\mathbb{N}}$ is a sequence of complex-valued measurable functions on a measure space (S, Σ, μ). Let the sequence converge pointwise to a function f, and for some integrable function g, we have $|f_n(x)| \leq g(x)$ for all $x \in \mathbb{R}$ with $g : \mathbb{R} \to [0, \infty]$. Then f is integrable as is f_n for each n, and

$$\lim_{n\to\infty} \int_{\mathbb{R}} f_n\,d\mu = \int_{\mathbb{R}} \lim_{n\to\infty} f_n\,d\mu = \int_{\mathbb{R}} f\,d\mu.$$

Proof. See [19], Theorem 3.3.1. □

Lemma 4.4.3 (Rademacher's Theorem)**.** Let M be a connected, $\mathcal{C}^1$-smooth Riemannian manifold and let $d(x, y)$ denote the geodesic distance on M. Then any function $\phi : M \to \mathbb{R}$ locally Lipschitz with respect to $d(x, y)$ is differentiable outside a set $N \subset M$ of zero volume. Its gradient, denoted by $\nabla\phi$, is a Borel map from $\mathrm{dom}(\nabla\phi) := M \setminus N$ into the tangent bundle TM.

Proof. See [16], Lemma 4 and [4], Theorem 10.8, (ii). □

Now we define superdifferentiability of a function on Riemannian manifolds. For this purpose, first we define the supergradient.

Definition 4.4.3 (Supergradient)**.** Let $U \subseteq \mathbb{R}^n$ be a convex open subset and $f : U \to \mathbb{R}$ be a concave function. Its supergradient at $x_0 \in U$ is defined as a vector $\vec{v}$ if for all $x \in U$:

$$f(x) - f(x_0) \leq \vec{v} \cdot (x - x_0).$$

Such pairs $(\vec{v}, x)$ form a subset of the tangent bundle, i.e., $\partial^- f \subset TU$.

We define the subgradient of a convex function f at $x_0 \in U$ as a vector $\vec{u}$ if for all $x \in U$:

$$f(x) - f(x_0) \geq \vec{u} \cdot (x - x_0).$$

Definition 4.4.4 (Superdifferential and Subdifferential)**.** Let (M, g) be a $\mathcal{C}^3$-smooth Riemannian manifold. Fix $x \in M$. A function $f : M \to \mathbb{R}$ is said to be superdifferentiable at x with supergradient $\vec{v} \in T_xM$ if

$$f(\exp_x \vec{u}) \leq f(x) + g\langle \vec{v}, \vec{u} \rangle_x + o(|\vec{u}|_x) \tag{4.17}$$

holds for small $\vec{u} \in T_xM$, where $\frac{o(\alpha)}{\alpha} \xrightarrow{\alpha \to 0} 0$.
f is said to be subdifferentiable with subgradient $\vec{w} \in \partial_- f_x \subset T_xM$, if

$$f(\exp_x \vec{u}) \geq f(x) + g\langle \vec{w}, \vec{u} \rangle_x + o(|\vec{u}|_x). \tag{4.18}$$

If (4.17) and (4.18) hold, then f is differentiable at x and we have $\vec{v} = \vec{w} = \nabla f(x)$.

This important observation inspires the chain rule for supergradients, helping to link minimal geodesics with the superdifferentiability of the cost.

Lemma 4.4.4 (Chain Rule)**.** Let M be a smooth Riemannian manifold. Assume that $\psi : M \to \mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$ have supergradients $\vec{v} \in \partial^- \psi(x)$ and $t \in \partial^- f_{\psi(x)}$ at some $x \in M$. If f is non-decreasing, then $t\vec{v} \in \partial^-(f \circ \psi)_x$.

Proof. We show that

$$(f \circ \psi)(\exp_x \vec{u}) \leq (f \circ \psi)(x) + g\langle t\vec{v}, \vec{u} \rangle_x + o(|\vec{u}|_x).$$

We apply Definition 4.4.3, (4.17) to f. Hence

$$f(\exp_{\psi(x)}(\varepsilon)) = f(\psi(x) + \varepsilon) \leq f(\psi(x)) + t\varepsilon + o(\varepsilon). \tag{4.19}$$

Also, $\vec{v} \in \partial^- \psi_x$. Hence

$$\psi(\exp_x \vec{u}) \leq \psi(x) + g\langle \vec{v}, \vec{u} \rangle_x + o(|\vec{u}|_x). \tag{4.20}$$

Since f is non-decreasing, set $\varepsilon = g\langle \vec{v}, \vec{u} \rangle_x + o(|\vec{u}|_x)$ in (4.19) and (4.20) to get

$$\begin{aligned} f(\psi(\exp_x \vec{u})) &= f(\psi(x) + g\langle \vec{v}, \vec{u} \rangle_x + o(|\vec{u}|_x)) \\ &\leq f(\psi(x) + tg\langle \vec{v}, \vec{u} \rangle_x + t \circ (|\vec{u}|_x) + o(g\langle \vec{v}, \vec{u} \rangle_x + o(|\vec{u}|_x))) \\ &= f(\psi(x)) + tg\langle \vec{v}, \vec{u} \rangle_x + o(|\vec{u}|_x). \end{aligned}$$

□

Proposition 4.4.2 (Superdifferentiability of Geodesic Distance Squared)**.** Assume that (M, g) is a $\mathcal{C}^3$-smooth Riemannian manifold with boundary. Let $C : [0, 1] \to M$ have minimal length among all piecewise $\mathcal{C}^1$-curves connecting $x = C(0)$ to $y = C(1) \notin \partial M$, parametrized with constant speed. Then the function $\phi(\cdot) = \frac{d^2(\cdot, y)}{2}$ has supergradient $C'(1) \in \partial^- \phi_x$ at x.

Proof. See [16], Proposition 6. □

The next lemma introduces a Young-like inequality and conditions for equality. These conditions help determine the optimal map's shape and are essential for a uniqueness argument based on Brenier's original idea. First, let's state the Young inequality.

Proposition 4.4.3 (Young Inequality)**.** Let $U = V = \mathbb{R}^n$ and $C(u, v) = \langle u, v\rangle$. Assume that $f : U \to \overline{\mathbb{R}}$. Then for all $u \in U$ and $v \in V$, we have:

(i) $f(u) + f^C(v) \geq C(u, v)$;

(ii) $f^C(v) + f^{CC}(u) \geq C(u, v)$.

Proof. See [17], Proposition 3.3.3. □

Lemma 4.4.5 (Tangency)**.** Assume that (M, g) is a connected, compact, $\mathcal{C}^3$-smooth Riemannian manifold without boundary. Let $\phi : M \to \mathbb{R}$ be an infimal convolution, i.e., $\phi = \phi^{CC}$ with $C(x, y) = \frac{d^2(x,y)}{2}$. Then

$$C(x, y) - \phi(x) - \phi^C(y) \geq 0 \tag{4.21}$$

for all $x, y \in M$.
If $x \in M$ is chosen where ϕ is differentiable, then in equation (4.21), equality occurs if and only if $y = \exp_x[-\nabla\phi(x)]$.

Proof. By the definition of infimal convolution of $\phi : M \to \mathbb{R} \cup \{+\infty\}$

$$\phi^C(y) := \inf_{x \in M}\{C(x, y) - \phi(x)\},$$

we conclude

$$C(x, y) - \phi(x) \geq \phi^C(y).$$

Therefore

$$C(x, y) - \phi(x) - \phi^C(y) \geq 0.$$

(See [16], Proposition 3.3.7).
Now fix $x \in M$ such that ϕ is differentiable at x.
$\Rightarrow$) Assume $y \in M$ satisfies $C(x, y) - \phi(x) - \phi^C(y) = 0$.
We show that $y = \exp_x[-\nabla\phi(x)]$:
Consider $z \in M$. Hence

$$C(z, y) - \phi(z) - \phi^C(y) \geq 0 = C(x, y) - \phi(x) - \phi^C(y).$$

Therefore,

$$C(z, y) - \phi(z) \geq C(x, y) - \phi(x). \tag{4.22}$$

Define $\psi(z) := C(z, y)$ and $z := \exp_x \vec{u}$, and we know ϕ is differentiable at x, i.e.,

$$\phi(z) = \phi(x) + g\langle\nabla\phi(x), \vec{u}\rangle_x + o(|\vec{u}|_x). \tag{4.23}$$

Thus

$$\begin{aligned}\psi(z) = \psi(\exp_x \vec{u}) &= C(z,y)\\ &\geq C(x,y) - \phi(x) + \phi(z)\\ &= \psi(x) - \phi(x) + \phi(x) + g\langle \nabla\phi(x), \vec{u}\rangle_x + o(|\vec{u}|_x),\end{aligned}$$

where we used (4.22) in the first inequality and (4.23) in the last equality. Hence we obtain

$$\psi(z) \geq \psi(x) + g\langle \nabla\phi(x), \vec{u}\rangle_x + o(|\vec{u}|_x);$$

i.e., $\psi(z)$ has subgradient $\nabla\phi(x) \in \partial_-\psi_x$ at x.
Now, by using Proposition 4.4.2, we want to show that ψ has a supergradient.
By the Hopf-Rinow theorem (see below), there exists a minimal geodesic $C : [0,1] \to M$ such that $C(0) = y$ and $C(1) = x$ given by: $C(1-t) = \gamma(t)$ where $\gamma(t) := \exp_x(-tC'(1))$ is a geodesic from x to y. Therefore

$$C(1-t) = \exp_x(-tC'(1)).$$

As the manifold M, C is $\mathcal{C}^3$-smooth, by Proposition 4.4.2, we have $C'(1) \in \partial^-\psi_x$. Thus, we conclude that ψ is both subdifferentiable and superdifferentiable at x. Hence ψ is differentiable with

$$C'(1) = \nabla\phi(x). \tag{4.24}$$

Therefore

$$y = C(0) = \gamma(1) = \exp_x(-C'(1)) = \exp_x(-\nabla\phi(x)),$$

where we used (4.24) in the last equality. Thus we obtain

$$y = \exp_x[-\nabla\phi(x)].$$

$\Leftarrow$) We show that if $y = \exp_x[-\nabla\phi(x)]$, then $C(x,y) - \phi(x) - \phi^C(y) = 0$.
Since $\phi = \phi^{CC}$ is real-valued, by Lemma 4.4.2, we have $\phi^C : M \to \mathbb{R}$ is Lipschitz continuous.
M is compact, therefore we define for $y \in M$:

$$\phi^{CC}(x) := \inf_{y\in M}\{C(x,y) - \phi^C(y)\}. \tag{4.25}$$

By (4.21), we have:

$$C(x,y) - \phi(x) - \phi^C(y) \geq 0 \tag{4.26}$$

and by (4.25):

$$\phi(x) = \phi^{CC}(x) \geq C(x,y) - \phi^C(y).$$

Thus

$$C(x,y) - \phi(x) - \phi^C(y) \leq 0. \tag{4.27}$$

By (4.26) and (4.27), we conclude $C(x,y) - \phi(x) - \phi^C(y) = 0$.
Therefore, we have shown that for fixed $x \in M$, at most one $y \in M$ satisfies inequality (4.21). $\square$

Theorem 4.4.3 (Hopf-Rinow)**.** Let M be a geodesically complete connected Riemannian manifold. Then any two points $x, y \in M$ can be connected by a minimal geodesic.

Proof. See [17], Theorem 10.9. $\square$

Theorem 4.4.4 (Unique Optimal Maps)**.** Let (M, g) be a connected, compact Riemannian manifold, $\mathcal{C}^3$-smooth, and without boundary. Assume that μ is a Borel probability measure on (M, g) such that $\mu \ll \text{vol}$. If $\phi : M \to \mathbb{R}$ is an infimal convolution, $\phi = \phi^{CC}$ with $C(x, y) = \frac{d^2(x,y)}{2}$, then $T(x) = \exp_x[-\nabla\phi(x)]$ minimizes the Monge's transportation cost $\int_M C(x, T(x))\, d\mu(x)$ on the class of push-forward measures $T_\#\mu$, and T is unique μ-almost everywhere.

Before we prove this theorem, we state the change of variables formula for push-forward measures $T_\#\mu = \nu$:
Change of Variables Formula. Let (M, g) be a connected, compact, and $\mathcal{C}^3$-smooth Riemannian manifold without boundary, and $T : M \to M$ be a Borel map and $h : M \to \mathbb{R} \cup \{\pm\infty\}$ a Borel map. Assume that μ and ν are probability measures on M such that $T_\#\mu = \nu$. Then

$$\int_M h\, d(T_\#\mu) = \int_M h(T(x))\, d\mu(x), \tag{4.28}$$

where, as in the case of $\mathbb{R}^n$, by the Riesz representation theorem, we can take (4.28) as the definition of $T_\#\mu$.

Proof of Theorem 4.4.4. Assume that $\phi : M \to \mathbb{R}$ is an infimal convolution, $\phi = \phi^{CC}$, and $T(x) = \exp_x[-\nabla\phi(x)]$. By Lemma 4.4.2, ϕ is Lipschitz continuous and by Lemma 4.4.3 (Rademacher's theorem), ϕ is differentiable μ-a.e., where $\nabla\phi$ is Borel, and consequently $T(x)$ is Borel.
Now, we consider the push-forward measure $T_\#\mu$ and $(0, 0) \in Lip_C$ such that $S(\mu, T_\#\mu) \neq \emptyset$, where $S(\mu, \nu)$ consists of equivalence classes of maps which agree up to sets on which μ vanishes, and $Lip_C \neq \emptyset$.
Assume $(f, g) \in Lip_C$ and $S \in S(\mu, T_\#\mu)$. Thus,

$$\begin{aligned} I(f, g) &= \int_M f(x)\, d\mu(x) + \int_M g(y)\, d(S_\#\mu)(y) && (4.29)\\ &= \int_M f(x)\, d\mu(x) + \int_M g(S(x))\, d\mu(x) \\ &\leq \int_M C(x, S(x))\, d\mu(x) && (4.30)\\ &= C(S), && (4.31) \end{aligned}$$

where we used the change of variables formula in the second equality, and (4.10) in the inequality, and $C(S) = \int_M C(x, S(x))\, d\mu(x)$ from the Monge's transportation cost. Hence, we obtain

$$I(f, g) \leq C(S),$$

where both sides are finite. Therefore, for some $r \in \mathbb{R}$,

$$\sup_{(f,g)\in Lip_C} I(f,g) \leq r \leq \inf_{S\in S(\mu,T_{\#}\mu)} \int_M C(x,S(x))\,d\mu(x). \tag{4.32}$$

Then by Lemma 4.4.5 for $y = \exp_x[-\nabla\phi(x)]$, we have equality. Therefore $(\phi,\phi^C) \in Lip_C$ implies

$$\phi(x) + \phi^C(T(x)) = C(x,T(x)) \qquad (\mu\text{-a.e.}).$$

Choose $(f,g) = (\phi,\phi^C)$ and $S = T$. Hence, we get

$$I(\phi,\phi^C) = \int_M C(x,T(x))\,d\mu(x)$$

and therefore

$$\sup_{(\phi,\phi^C)\in Lip_C} I(\phi,\phi^C) = \inf_{T\in S(\mu,T_{\#}\mu)} \int_M C(x,T(x))\,d\mu(x),$$

which proves the optimality of the map T.
Conversely, any optimal map $S \in S(\mu,T_{\#}\mu)$ must also satisfy $C(S) = r = I(\phi,\phi^C)$, ensuring that equality remains in (4.30).
From (4.21), i.e., $C(x,y) - \phi(x) - \phi^C(y) \geq 0$, we must have

$$\phi(x) + \phi^C(S(x)) = C(x,S(x))$$

pointwise μ-a.e., i.e.,

$$I(\phi,\phi^C) = \int_M (\phi(x) + \phi^C(S(x)))\,d\mu(x) = \int_M C(x,S(x))\,d\mu(x),$$

where we used (4.29) in the first and (4.30) in the last equality.
By Lemma 4.4.5, since ϕ is differentiable on the set of full measure, we conclude that

$$S(x) = \exp_x[-\nabla\phi(x)] = T(x).$$

□

Note that the Kantorovich duality (4.8) was proven for $\nu = T_{\#}\mu$ in the previous proof. Thus, it must continue to hold under the hypothesis of Theorem 4.4.5.

Theorem 4.4.5 (Existence and Characterization of Optimal Maps)**.** Assume that (M,g) is a $\mathcal{C}^3$-smooth, connected, compact Riemannian manifold without boundary. Let μ and ν be two probability measures such that $\mu \ll \mathrm{vol}$ and ν arbitrary on M, and fix $C(x,y) = \frac{d^2(x,y)}{2}$. Then for some potential $\phi : M \to \mathbb{R}$ satisfying $\phi = \phi^{CC}$, the map $T(x) = \exp_x[-\nabla\phi(x)]$ satisfies $T_{\#}\mu = \nu$.
Apart from differences in sets of μ-measure zero, only one map $T \in S(\mu,\nu)$ can result in this manner.

Proof. Choose $(\phi, \psi) = (\phi^{CC}, \phi^C)$. This maximizes

$$I(\phi, \psi) = \int_M \phi(x)\, d\mu(x) + \int_M \psi(y)\, d\nu(y)$$

on Lip_C, which is obtained by Proposition 4.4.1. By Lemma 4.4.2, since infimal convolutions are Lipschitz, hence ϕ and ψ are Lipschitz. Therefore by Lemma 4.4.3, $\nabla\phi : M \to TM$ and hence $T : M \to M$ are Borel maps defined almost everywhere with respect to measure μ.
To show that the map $T(x) = \exp_x[-\nabla\phi(x)]$ pushes μ forward to ν, we can integrate each continuous function $f \in \mathcal{C}(M)$ against both measures μ and $T_{\#}\mu$, confirming that the two integrals are equal.
We define the perturbation $\psi_\varepsilon(y) := \phi^C(y) + \varepsilon f(y)$ and $\phi_\varepsilon := (\psi_\varepsilon)^C$ for $x, y \in M$ and $|\varepsilon| < 1$, by

$$\begin{aligned}
(\psi_\varepsilon)^C(x) = \phi_\varepsilon(x) &:= \inf_{y\in M}\{C(x,y) - \psi(y) - \varepsilon f(y)\} \\
&= \inf_{y\in M}\{C(x,y) - \phi^C(y) - \varepsilon f(y)\} \\
&= \inf_{y\in M}\{C(x,y) - (\phi^C(y) + \varepsilon f(y))\} \\
&= \inf_{y\in M}\{C(x,y) - \psi_\varepsilon(y)\}.
\end{aligned}$$

Thus,

$$(\psi_\varepsilon)^C(x) = \phi_\varepsilon(x) := \inf_{y\in M}\{C(x,y) - \psi_\varepsilon(y)\}. \tag{4.33}$$

Fix $x \in M$ such that ϕ_0 is differentiable at x.
Since M is compact and continuous, then (4.33) is well-defined:
If $\varepsilon = 0$, then

$$\phi_0(x) = \inf_{y\in M}\{C(x,y) - \psi(y)\} = \inf_{y\in M}\{C(x,y) - \phi^C(y)\}.$$

Hence by Lemma 4.4.5 and (4.21), we have:

$$y = T(x) = \exp_x[-\nabla\phi(x)].$$

Now for some small $\varepsilon > 0$, we get the infimum at $y_\varepsilon = T(x) + o(1)$:
Define $F_\varepsilon : [0,1] \times M \to \mathbb{R}$; $(\varepsilon, y) \mapsto F_\varepsilon(y)$ where y_0 is the unique minimum of F_0 and y_ε is the unique minimum of F_ε.
We claim: $y_\varepsilon \to y_0 = y$ as $\varepsilon \to 0$.
We show this by contradiction: Let there exist a subsequence y_{ε_n} such that $y_{\varepsilon_n} \to \tilde{y} \neq y_0$ as $n \to \infty$ and $F_{\varepsilon_n}(y_{\varepsilon_n}) = \min F_{\varepsilon_n}(\tilde{y})$. Therefore, $F_{\varepsilon_n}(y_{\varepsilon_n}) \le F_{\varepsilon_n}(y_0)$. Then $F_0(\tilde{y}) \le F_0(y_0)$ as $n \to \infty$, i.e., $\tilde{y} \neq y_0$ is minimal of F_0, which contradicts y_0 being minimal of F_0. Therefore, $y_\varepsilon \to y_0 = y$ as $\varepsilon \to 0$.
Thus, since the infimum in (4.33) is attained for $\varepsilon = 0$ at $y = T(x)$ and for $\varepsilon > 0$ at $y_\varepsilon = T(x) + o(1)$, we have

$$\phi_\varepsilon(x) = C(x, y_\varepsilon) - \psi(y_\varepsilon) - \varepsilon f(y_\varepsilon). \tag{4.34}$$

If $\varepsilon = 0$, then

$$\begin{aligned}\phi_0(x) &= C(x, T(x)) - \psi(T(x)) \\ &\leq C(x, y_\varepsilon) - \psi(y_\varepsilon) \\ &= \phi_\varepsilon(x) + \varepsilon f(y_\varepsilon),\end{aligned}$$

where we used $y_\varepsilon = T(x) + o(1)$ in the first inequality and (4.34) in the last equality. Therefore

$$C(x, T(x)) - \psi(T(x)) - \varepsilon f(y_\varepsilon) \leq \phi_\varepsilon(x) \leq C(x, y) - \psi(y) - \varepsilon f(y),$$

for all $y \in M$, where we used (4.33) in the second inequality.
Choose $y = T(x)$. Then

$$\begin{aligned}\phi_0(x) - \varepsilon f(y_\varepsilon) &\leq C(x, y) - \psi(y) - \varepsilon f(y_\varepsilon) \\ &\leq \phi_\varepsilon(x) \leq C(x, y) - \phi(y) - \varepsilon f(y).\end{aligned}$$

Thus

$$\begin{aligned}&\phi_0(x) - \varepsilon f(y_\varepsilon) \leq \phi_\varepsilon(x) \leq \phi_0(x) - \varepsilon f(y) \\ &\Rightarrow \phi_0(x) - \varepsilon f(y_\varepsilon) \leq \phi_\varepsilon(x) \leq \phi_0(x) - \varepsilon f(T(x)) \\ &\Rightarrow -\varepsilon f(y_\varepsilon) \leq \phi_\varepsilon(x) - \phi_0(x) \leq -\varepsilon f(T(x)) \\ &\Rightarrow \varepsilon f(T(x)) - \varepsilon f(y_\varepsilon) \leq \phi_\varepsilon(x) - \phi_0(x) + \varepsilon f(T(x)) \leq 0 \\ &\Rightarrow \varepsilon\left(f(T(x)) - f(y_\varepsilon)\right) \leq \phi_\varepsilon(x) - \phi_0(x) + \varepsilon f(T(x)) \leq 0 \\ &\Rightarrow \phi_\varepsilon(x) = \phi_0(x) - \varepsilon f(T(x)) + O(\varepsilon(f(T(x)) - f(y_\varepsilon))) \\ &\Rightarrow \phi_\varepsilon(x) = \phi_0(x) - \varepsilon f(T(x)) + o(\varepsilon),\end{aligned} \tag{4.35}$$

where since $f \in \mathcal{C}(M)$ and the infimum is obtained at y_ε, we have $o(\varepsilon) \leq \frac{2}{\varepsilon}\|f\|_\infty$ that vanishes as $\varepsilon \to 0$.
Hence, we obtained $\phi_\varepsilon(x) = \phi_0(x) - \varepsilon f(T(x)) + o(\varepsilon)$. Since $(\phi, \psi) = (\phi^{CC}, \phi^C)$ maximizes (4.9) and $I(\phi_\varepsilon, \psi_\varepsilon)$ at $\varepsilon = 0$ obtains its maximum, therefore

$$\begin{aligned}0 &= \lim_{\varepsilon \to 0} \frac{I(\phi_\varepsilon, \psi_\varepsilon) - I(\phi_0, \psi_0)}{\varepsilon} \\ &= \lim_{\varepsilon \to 0} \left\{ \int_M \frac{\phi_\varepsilon(x) - \phi_0(x)}{\varepsilon}\, d\mu(x) + \int_M \frac{\psi_\varepsilon(y) - \psi_0(y)}{\varepsilon}\, d\nu(y) \right\} \\ &= \lim_{\varepsilon \to 0} \left\{ \int_M \frac{\phi_\varepsilon(x) - \phi_0(x)}{\varepsilon}\, d\mu(x) + \int_M f(y)\, d\nu(y) \right\} \\ &= \int_M -f(T(x))\, d\mu(x) + \int_M f(y)\, d\nu(y) \\ &= -\int_M f\, d(T_\# \mu) + \int_M f\, d\nu,\end{aligned}$$

where we used the definition of ϕ_ε in the third equality and (4.35), and the Dominated Convergence Theorem in the fourth equality, and the change of variables formula in the last equality.
Thus

$$\int_M f\, d(T_\# \mu) = \int_M f\, d\nu,$$

and by the Riesz-Markov theorem (see below), we conclude $T_\# \mu = \nu$, for arbitrary $f \in \mathcal{C}(M)$.
Now we show that $T(x) = \exp_x[-\nabla \phi(x)]$ with $T_\# \mu = \nu$, is unique:
Let $\Phi = (\Phi^C)^C$ be another potential map on M for which the map $S(x) := \exp_x[-\nabla \Phi(x)]$ satisfies $S_\# \mu = \nu$. By Theorem 4.4.4, T and S both minimize the Monge problem satisfying $T_\# \mu = \nu$ and $S_\# \mu = \nu$. Therefore $T = S$ μ-a.e. □

Theorem 4.4.6 (Riesz - Markov)**.** Let M be a locally compact Hausdorff space and ϕ a positive linear functional on $\mathcal{C}_c(M)$. Then there exists a σ-algebra $\mathscr{A}$ on M which contains all Borel sets in M, and there exists a unique positive measure μ on $\mathscr{A}$ such that

$$\phi(f) = \int_M f\, d\mu$$

for every $f \in \mathcal{C}_c(M)$.

Proof. See [20], Theorem 4.4. □

5 Polar factorization

In this chapter, we follow [21], [16], [22], Subsection 1.7.2, and [23] Section 2. Here, we first state the polar decomposition in Euclidean space and then generalize it to Riemannian manifolds.

5.1 Polar decomposition in Euclidean space

In this section, we follow [21]. We state the polar decomposition of a linear transformation on a finite-dimensional Euclidean space. For this purpose, we recall some basic notions.

Polar factorization. We know that from the transportation problem in certain situations, especially when dealing with a cost function $C(x, y) = |x-y|^2$, it may be proven that the optimal coupling γ does not allow for splitting masses. Brenier shows that particles at x are sent to a unique destination $T(x)$, solving Monge's problem, and also shows that this optimal map has a simple form $T(x) := \nabla u(x)$ where u is a convex function.
Brenier also states a theorem for an original polar factorization of vector maps: Vector fields can be written as the composition of the gradient of a convex function and of a measure-preserving map.
For this purpose, we introduce the notion of measure-preserving map, and then we state the polar factorization theorem.

Definition 5.1.1 (Measure-preserving map)**.** Let (X, μ) and (Y, ν) be two probability spaces and $f : X \to Y$ a measurable map. Then we say that f is measure-preserving if and only if $f_{\#}\mu = \nu$.

Theorem 5.1.1 (Polar Factorization)**.** Let Ω be a convex subset of $\mathbb{R}^n$. Consider a map $f : \Omega \to \mathbb{R}^n$ and a measure μ on Ω which is absolutely continuous. Assume that $f_{\#}\mu$ is absolutely continuous. Then there exists a convex function $\phi : \Omega \to \mathbb{R}$ and a measure-preserving map $S : \Omega \to \Omega$ (i.e., $S_{\#}\mu = \mu$) such that $f = (\nabla\phi) \circ S$.
Moreover, $\nabla\phi$ and S are almost everywhere uniquely determined.

Proof. See [22], Theorem 1.53, and [23], Theorem 2.1. □

In the next section, we will focus on the main result.

5.2 Polar factorization of maps on Riemannian manifolds

In this section, we follow [16], Section 3. We aim to state and prove that a Borel map, which never maps a positive volume into a zero volume, can be uniquely factorized almost everywhere into an exponential map and a measure-preserving map. First, we show that the exponential map achieved by Theorem 4.4.5 is invertible under more assumptions.

Corollary 5.2.1 (Invertibility)**.** Let (M, g) be a $\mathcal{C}^3$-smooth connected compact Riemannian manifold without boundary and $C(x, y) = \frac{d^2(x,y)}{2}$. Assume that $\mu, \nu \in \mathcal{P}(M)$ are both absolutely continuous with respect to the volume, i.e., $\mu \ll \text{vol}$ and $\nu \ll \text{vol}$. If for some potential $\phi : M \to \mathbb{R}$ with $\phi = \phi^{CC}$, $g(x) = \exp_x[-\nabla\phi(x)]$, then $g^*(y) := \exp_y[-\nabla\phi^C(y)] \in S(\nu, \mu)$, which is the inverse map and satisfies $g^*(g(x)) = x$ (μ-almost everywhere) and $g(g^*(y)) = y$ (ν-almost everywhere).

Proof. First, we show that $g^*(y) = \exp_y[-\nabla\phi^C(y)]$:
We use the symmetry $\mu \leftrightarrow \nu$ and $(\phi, \phi^C) \leftrightarrow (\psi, \psi^C)$ and the fact that $\psi := \phi^C$ is the infimal convolution of ϕ with $C(x, y)$, so that

$$\psi^C = \phi^{CC} = \phi \tag{5.1}$$

and

$$\psi^{CC} = \phi^C = \psi. \tag{5.2}$$

We define the set where ψ is differentiable as $\text{dom}(\nabla\psi) \subset M$. Since ψ is the infimal convolution, by Lemma 4.4.2, ψ is Lipschitz continuous. By using Rademacher's theorem (Lemma 4.4.3), ψ must be differentiable outside a set of zero volume, and $\nabla\psi$ gives a Borel map on $\text{dom}(\nabla\psi)$. Therefore, we have $\nu(\text{dom}(\nabla\psi)) = 1$, and since $g_{\#}\mu = \nu$ (by Theorem 4.4.5), this implies that $U := \text{dom}(\nabla\phi) \cap g^{-1}(\text{dom}(\nabla\psi))$, which is a Borel set of full measure, i.e., $\mu(U) = 1$.
Now consider $x \in U$. By Lemma 4.4.5, we have:

$$\begin{aligned} 0 &= C(x, g(x)) - \phi^C(g(x)) - \phi(x) \\ &= C(g(x), x) - \psi(g(x)) - \psi^C(x), \end{aligned}$$

where we used (5.2) and symmetry in the last equality. Thus, we get

$$C(g(x), x) - \psi(g(x)) - \psi^C(x) = 0.$$

Since $g(x) \in \text{dom}(\nabla\psi)$, by Lemma 4.4.5, we conclude that

$$x = \exp_{g(x)}[-\nabla\psi(g(x))].$$

Since $\psi = \phi^C$, thus $x = \exp_{g(x)}[-\nabla\phi^C(g(x))]$. We conclude that $x = g^*(g(x))$ on U and hence μ-a.e.:
We arrive at $x = \exp_{g(x)}[-\nabla\phi^C(g(x))]$.
Let $g(x) = y$, hence $x = g^*(y)$. Therefore, $g^*(y) = \exp_y[-\nabla\phi^C(y)]$, and $y = g(x)$. Thus

$$g^*(g(x)) = \exp_y[-\nabla\phi^C(y)] = \exp_{g(x)}[-\nabla\phi^C(g(x))] = x$$

on U (since $x \in U$). Therefore, $g^*(g(x)) = x$ on U, and hence μ-almost everywhere.

Now we show $g^* \in S(\nu,\mu)$, i.e., $g^*_{\#}\nu = \mu$:
Let $f \in \mathcal{C}(M)$ be any continuous function. Then:

$$\begin{aligned}\int_M f(x)\ d(g^*_{\#}\nu)(x) &= \int_M f(g^*(y))\, d\nu(y)\\ &= \int_M f(g^*(y))\, d(g_{\#}\mu)(y)\\ &= \int_M f(g^*(g(x)))\, d\mu(x)\\ &= \int_M f(x)\, d\mu(x),\end{aligned}$$

where we used the change of variables formula in the first equality, $\nu = g_{\#}\mu$ in the second equality, and $g^*(g(x)) = x$ μ-a.e. in the last equality. Therefore,

$$\int_M f(x)\, d(g^*_{\#}\nu)(x) = \int_M f(x)\, d\mu(x) \quad \forall f \in \mathcal{C}(M).$$

Since f is arbitrary, we conclude that $g^*_{\#}\nu = \mu$. Thus $g^* \in S(\nu,\mu)$.
Now, by symmetry, we can conclude that

$$V := \mathrm{dom}(\nabla\psi) \cap (g^*)^{-1}(\mathrm{dom}(\nabla\phi))$$

is a Borel set with $\nu(V) = 1$, while $g(g^*(y)) = y$ holds for each $y \in V$. $\square$

Now we state and prove the main result.

Theorem 5.2.1 (Polar factorization of maps)**.** Assume that (M,g) is a $\mathcal{C}^3$-smooth connected compact Riemannian manifold without boundary. Fix $C(x,y) = \frac{d^2(x,y)}{2}$ and let $f : M \to M$ be a Borel map and $\mu \ll \mathrm{vol}$ be a Radon measure on M. If $f_{\#}\mu \ll \mathrm{vol}$, i.e., if f does not map any set with positive measure to a set with zero volume, then $f = g \circ h$ (μ-a.e.) for some map $g(x) = \exp_x[-\nabla\phi(x)]$ with $\phi = \phi^{CC}$ and a measure-preserving map $h_{\#}\mu = \mu$. These two factors $g,h : M \to M$ are unique μ-a.e. and Borel.

Proof. First, we show that f can be factorized to g and h by $f = g \circ h$, where $g(x) = \exp_x[-\nabla\phi(x)]$ with $\phi = \phi^{CC}$ and $h : M \to M$ such that $h_{\#}\mu = \mu$.
Since by (4.1) we have $\nu(A) = \mu(f^{-1}(A))$ for all $A \in \mathcal{B}(M)$, and $f : M \to M$ a measurable map, we define $\nu := f_{\#}\mu$ as the push-forward of the measure μ through the map $f : M \to M$.
We know that M is compact and the Radon measure μ is locally finite, so we can define a normalization constant $n^{-1} := \mu(M) = \nu(M) < \infty$ such that $n\mu$ and $n\nu$ are probability measures, i.e., $n\mu(M) = n\nu(M) = 1$. Since by our hypothesis μ and ν are absolutely continuous, $n\mu$ and $n\nu$ are also absolutely continuous with respect to volume. Now, by using Theorem 4.4.5, for some potential $\phi : M \to \mathbb{R}$ satisfying $\phi = \phi^{CC}$, the Borel map $g(x) = \exp_x[-\nabla\phi(x)]$ pushes $n\mu$ forward to $n\nu$, i.e., $g \in S(\mu,\nu)$.

Now we construct h and show that $h_{\#}\mu = \mu$:
By Corollary 5.2.1, since $n\nu \ll \mathrm{vol}$, then the Borel inverse map g^* belongs to $S(\nu, \mu)$ and satisfies $g(g^*(y)) = y$ ν-a.e. (and $g^*(g(x)) = x$ μ-a.e.) for a Borel set $A \subset M$ of full ν-measure $n\nu(A) = 1$. Now we set $h := g^* \circ f$. Therefore $f(x) = g(h(x))$ on $f^{-1}(A)$, hence μ-almost everywhere.
Since $f \in S(\mu, \nu)$ and $g^* \in S(\nu, \mu)$, therefore $g^* \circ f \in S(\mu, \mu)$. Hence $h \in S(\mu, \mu)$.
Now we show that $h_{\#}\mu = \mu$:
For every $F \in \mathcal{C}(M)$, we have:

$$\begin{aligned}
\int_M F\, d(h_{\#}\mu) &= \int_M F\, d((g^* \circ f)_{\#}\mu) \\
&= \int_M F(g^* \circ f)\, d\mu \\
&= \int_M F \circ g^*\, d(f_{\#}\mu) \\
&= \int_M F \circ g^*\, d\nu \\
&= \int_M F\, d(g^*_{\#}\nu) \\
&= \int_M F\, d\mu.
\end{aligned}$$

Hence we get:

$$\int_M F\, d(h_{\#}\mu) = \int_M F\, d\mu \quad \forall F \in \mathcal{C}(M).$$

Thus $h_{\#}\mu = \mu$, i.e., h is a measure-preserving map.
Now we show that these factors are unique:
Let $f = g' \circ h'$ hold on $U' \subset M$ of full μ-measure, i.e., $n\mu(U') = 1$, where $g'(x) = \exp_x[-\nabla\phi(x)]$ with $\phi = \phi^{CC}$ and $h' \in S(\mu, \mu)$, i.e., $h'_{\#}\mu = \mu$.
Claim: $g' \in S(\mu, f_{\#}\mu)$:

$$g'_{\#}\mu = g'_{\#}(h'_{\#}\mu) = (g' \circ h')_{\#}\mu = f_{\#}\mu.$$

Therefore, we get $g'_{\#}\mu = f_{\#}\mu$. Thus $g' \in S(\mu, f_{\#}\mu)$. Hence, by Theorem 4.4.5, $g' = g$ on a set $U \subset M$ of full μ-measure, i.e., $n\mu(U) = 1$.
Consider $x \in f^{-1}(A) \cap h'^{-1}(U) \cap U'$, i.e., for μ-a.e. x, we have $f(x) = g'(h'(x)) = g(h'(x))$, where we used $g' = g$ in the last equality.
Thus, $f(x) = g(h'(x))$ and also we have

$$h(x) = g^*(f(x)).$$

Therefore,

$$h(x) = g^*(f(x)) = g^*(g(h'(x))) = h'(x).$$

So we conclude that $h = h'$ μ-almost everywhere. □

References

[1] A. Figalli, F. Glaudo. An invitation to optimal transport, Wasserstein distances, and gradient flows, 2021.

[2] L. Ambrosio, N. Gigli. A user's guide to optimal transport, Springer, 2009.

[3] H. Bercovici, A. Brown and C. Pearcy. Measure and integration. Springer ,Chem, 2016.

[4] C. Villani. Optimal transport old and new, 2009, Springer - Verlag Berlin Heidelberg.

[5] Van den Brink, CB. The Banach-Alaoglu theorem for topological vector spaces, master thesis Harvard, 2019.

[6] H. Hosseini Giv. Proving the Banach–Alaoglu Theorem via the Existence of the Stone–Čech Compactification, The American Mathematical Monthly, 2014.

[7] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs, CR Acad. Sci. Paris Sér. I Math. 305 (1987), 805-808.

[8] C. Villani. Optimal transport, old and new. Grundlehren Math. Wiss. 338, Springer, Berlin, 2009.

[9] F. Santambrogio. Optimal transport for applied mathematicians. Progr. Nonlin. Differential Equations Appl. 87, Birkhäuser/ Springer, Cham, 2015.

[10] C. Villani. Topics in optimal transportation. Graduate Stud. Math. 58, American Mathematical Society, Providence, RI, 2003.

[11] T. Fischer. Existence, uniqueness, and minimality of the Jordan measure decomposition, 1206.5449, arXiv, math.ST, 2012.

[12] M. Kunzinger. Riemannian geometry, Lecture Notes, University of Vienna, 2020.

[13] M. P. do Carmo. Riemannian Geometry, Birkhäuser, Boston, Basel, Berlin, 1992.

[14] M. Kunzinger. Differential geometry 1, Lecture Notes, University of Vienna, 2008.

[15] A. Figalli, C. Villani. Optimal Transport And Curvature, Lecture Notes, 2008.

[16] R. J. McCann. Polar factorization of maps on Riemannian manifolds, Birkhäuser-Verlag, Basel, 2001.

[17] S. T. Rachev, L. Rüschendorf. Mass Transportation Problems, Probab. Appl. Springer-Verlag, New York, 1998.

[18] D. Jekel. The Heine-Borel and Arzela-Ascoli Theorems, University of Washington, Math 334-336, 2016.

[19] D. Hong, R. Gardner. Real Analysis with an Introduction to Wavelets and Applications, ScienceDirect, 2005.

[20] S. Chavan. Riesz-Markov representation theorem: Problems, lecture notes, 2014,2015.

[21] A. L. Onishchik, V. S. Shul'man. Polar-decomposition, Encyclopedia of Mathematics, 2011.

[22] F. Santambrogio. Optimal Transport for Applied Mathematicians, Calculus of Variations, PDEs, and Modeling, Laboratoire de Mathématiques d'Orsay Université Paris-Sud Orsay, France, Springer, 2015.

[23] Y. Brenier. Polar Factorization and Monotone Rearrangement of Vector-Valued Functions, Universite' de Paris VI, 1991.

[24] J. Cerny. Measure and integration theory, Lecture Notes, University of Vienna, 2016W.